

PROCEEDINGS OF THE
MULTIDISCIPLINARY DOCTORAL SCHOOL
2008-2009 ACADEMIC YEAR
FACULTY OF INFORMATION TECHNOLOGY
PÁZMÁNY PÉTER CATHOLIC UNIVERSITY
BUDAPEST
2009

Faculty of Information Technology
Pázmány Péter Catholic University

Ph.D. PROCEEDINGS

PROCEEDINGS OF THE
MULTIDISCIPLINARY DOCTORAL SCHOOL
2008-2009 ACADEMIC YEAR
FACULTY OF INFORMATION TECHNOLOGY
PÁZMÁNY PÉTER CATHOLIC UNIVERSITY
BUDAPEST

June, 2009



Pázmány University ePress
Budapest, 2009

© PPKE Információs Technológiai Kar, 2009

Kiadja a Pázmány Egyetem eKiadó
2009
Budapest

Felelős kiadó
Dr. Fodor György
a Pázmány Péter Katolikus Egyetem rektora

Cover image by András József Laki, microfluidic focusing device
A borítón Laki András József mikrofluidikai fókuszáló eszköze látható

HU ISSN 1788-9197

Contents

INTRODUCTION	7
DÁVID CSERCSIK • Model Synthesis and Identification of a GnRH Neuron	9
DÁVID TISZA • Application of Software Defined Radio in Wireless Sensor Networks, Superposition Coding	13
VILMOS SZABÓ • Detection and Classification of Living Organism for Water Quality Monitoring	17
ÁDÁM FEKETE • A First-Principle Computational Model for Electronic Structure of Molecular or Atomic Media	21
LÁSZLÓ GRAND • Assessing Tissue Reaction Around Silicon-based Multielectrode Arrays with Different Bio-coatings	25
BALÁZS DOMBOVÁRI • Electrophysiological Recordings with Electronic Depth Controlled Intracortical Microprobe Arrays	29
FERENC LOMBAI • Improvement of Tactile Sensor Measurements for Biologically Motivated Robot Control	33
JÓZSEF VERES • An Improved Biped Actuation System Inspired by the Human Flexor-Extensor Mechanism	37
ÁKOS TAR • 3D Geometry Reconstruction using Large Infrared Proximity Array for Robotic Applications	41
LÁSZLÓ FÜREDI • Stream Processing Evaluation Platform and Some Application Results	45
ANDRÁS KISS • Mach 3 Flow Simulation on IBM Cell Processor Based Emulated Digital Cellular Neural Networks	49
LÁSZLÓ KOZÁK • Quasi Non-deterministic Turing Machine	53
ZOLTÁN KÁRÁSZ • Heuristic Optimization with Processor Array Architecture	57
GERGELY FELDHOFFER • A Comparison of Audio to Visual Speech Conversions	61
BÁLINT SASS • Automatically Creating a Frequency Dictionary of Verb Phrase Constructions	65
GYULA PAPP • Comparison of Unsupervised Word Sense Disambiguation Methods	69
ANDREA KOVÁCS • Local Contour Descriptors around Scaleinvariant Keypoints	73
NORBERT BÉRCI • A Visual Human-Machine Interface on Massively Parallel Computers	77
TAMÁS PILISSY • The Relation of Kinematic Movement Patterns and Muscle Synergies in Lower Limb Cycling	81
RÓBERT TIBOLD • Non-linear 3D Model of Muscle Forces and Kinematic Variances in Reaching Arm Movements	85
ÁDÁM BALOGH • Phonocardiography in Preterm Newborns with Patent Ductus Arteriosus	89

ANDRÁS GELENCSÉR • Colors and Color Perception	93
BALÁZS VARGA • Color Based Image Segmentation in a Water Supply Surveillance System	97
DÁNIEL SZOLGAY • Human Detection in Videos with Strong Camera Motion	101
KÁLMÁN TORNAI • Strategy Optimization on Financial Time Series	105
KRISTÓF TAHY • High Field Characteristics of Long and Short channel 2D Graphene FETs	109
GERGELY TREPLÁN • Efficient Routing and Communication in Wireless Systems	113
BALÁZS KARLÓCAI • Energy Balancing in Wireless Sensorial Network by Using Discrete Energies	117
ANDRÁS BOJÁRSZKY • Reliable Routing in Wireless Sensorial Network by Combinational Optimization	121
PÉTER VIZI • Mobile Platform for Testing Communication Protocols, Challenges in the Implementation of SP Receiver	125

Introduction

Since September 2000, our multidisciplinary Doctoral School, originally established in 1993 as a joint program of four Universities, has been operating at our Faculty of Information Technology.

It is our pleasure to publish this annual proceedings to demonstrate the genuine multidisciplinary research done at our Jedlik Laboratories. Thanks are also due to the supervisors and consultants, as well as to the five collaborating National Research Laboratories of the Hungarian Academy of Sciences and the Semmelweis Medical School. The collaborative work with the partner Universities, especially, Katolieke Universiteit Leuven, Politecnico di Torino, Technische Universitat in München, University of California at Berkeley, University of Notre Dame, Univetsidad Sevilla, Universita di Catania is gratefully acknowledged..

As a milestone of this special collaboration, we were able to jointly complete the first year with the Semmelweis Medical School a new undergraduate curriculum on Molecular Bionics, the first of this kind in Europe.

We acknowledge the many sponsors of the research reported here. Namely,

- the Hungarian National Research Fund (OTKA),
- the Hungarian Academy of Sciences,
- the National Office of Research and Development (NKTH),
- the Gedeon Richter Co.,
- the Office of Naval Research (ONR) of the US,
- the National Science Foundation (NSF) of the US,
- IBM Hungary,
- Eutecus Inc., Berkeley, CA,
- Morphologic Ltd., Budapest,
- Analogic Computers Ltd., Budapest,
- AnaFocus Ltd., Seville, and
- the Pázmány Péter Catholic University.

Budapest, July 2009.

TAMÁS ROSKA
Head of the Doctoral School

Model Synthesis and Identification of a GnRH Neuron

Dávid Csercsik

(Supervisor: Gábor Szederkényi)

csercsik@digitus.itk.ppke.hu

Abstract—As the first step to build a hierarchical model of the GnRH pulse generator, a one compartment Hodgkin-Huxley type electrophysiological model of the GnRH neuron was constructed. The parameters of the model were estimated using both voltage clamp and current clamp data in such a way that the model is able to reproduce the qualitative features (depolarizing afterpotentials and bursting) of the neuronal behavior relevant to neuroendocrine functions.

Index Terms—Computational neuroscience, Parameter estimation, Neuroendocrinology, Hodgkin-Huxley modelling

I. INTRODUCTION

This GnRH pulse generator governs the central control of reproduction in vertebrates. This pulse generator controls the activity of hypothalamic neuroendocrine cells that secrete GnRH in a pulsatile way, closely associated with concurrent increases in multiunit electrical activity in the mediobasal hypothalamus (MUA volleys) (See [1], [2]). The pulsatile release of GnRH is driven by the intrinsic activity of GnRH neurons, which is characterized by bursts and prolonged episodes of repetitive action potentials correlated with oscillatory increases in intracellular Ca^{2+} . Several in vitro experiments have shown that changes in cytosolic Ca^{2+} concentration determine the secretory pattern of GnRH - see [3] -, underlining that Ca^{2+} plays a central role in the signal transduction processes that lead to exocytosis. Furthermore, as described by [4], GnRH secretion from perfused GT-1 and hypothalamic cells is reduced by L-type Ca^{2+} channel inhibitors and augmented by activation of voltage-gated Ca^{2+} channels. These results underline the importance of modeling Ca^{2+} currents, because the model is presumed to be later extended with the description of intracellular Ca^{2+} levels to describe hormone secretion.

1) *Significance and aim:* The models of the GnRH pulse generator, which can be found in literature nowadays, use generalized and very simple neuron models and networks. Furthermore, they are neither based on the known membrane properties of GnRH neurons, nor are able to describe the effect of ovary hormones - see [5]. Nevertheless, these results can provide interesting insights into pulsatility and synchronization as described in [6], [7].

With the application of cell marking based on the green fluorescent protein (GFP) and transgenic mice, the targeted

measurements and experiments on GnRH neurons became available, see [8], [9]. Another possibility is the application of so the called "immortalized" GnRH neurons, as described by [10]. The new biological data originating from these measurements, together with the new (or appropriately reformulated and integrated) approaches in the neoclassical computational neuroscience by [11], [12] offer promising possibilities in the field of modeling and identification of GnRH neurons and the GnRH pulse generator network.

The work reported in this paper is intended to be the first step in a bottom-up procedure which aims to build a hierarchical model of the GnRH pulse generator that includes the effects of ovarian hormones. In order to reach this aim, GFP-based patch clamp recordings were done on mice GnRH neurons at the Institute of Experimental Medicine (KOKI). The measured data are used to identify a [13] type conductance-based model of membrane dynamics.

II. MATERIALS AND METHODS

The measurements, the model development and the parameter estimation method are described in this section.

A. Measurements

1) *Obtaining and preparing samples:* Mouse brain was used for obtaining GnRH neurons for measurements.

2) *Whole-cell recording of GnRH neurons:* In order to visualize GnRH neurons in the brain slices, GnRH-enhanced green fluorescent protein (GnRH-GFP) transgenic mice were chosen in which the GnRH promoter drives selective GFP expression in the majority of GnRH neurons.

B. Model development

Using literature data on the ion channels of the GnRH neuron and properties of ion channels, a simple GnRH neuron-model can be developed and identified involving further literature data, voltage clamp and current clamp measurements. Measurement data were available in the form of whole cell patch-clamp recordings.

1) *The suggested model framework of single cell models:* A single compartment Hodgkin-Huxley (HH) type model -see [13] - is suggested, which can be extended to a multicompartmental structure, if needed, for the description of bursting. The main benefits of this model class are the following.

- Each ion channel is represented by an element of the model (conductance), so different ion channels can be

Faculty of Information Technology, Pázmány Péter Catholic University H-1364 Budapest 4., P.O. Box 178, Hungary

Process Control Research Group, Systems and Control Laboratory Hungarian Academy of Sciences H-1518, P.O. Box 63, Budapest, Hungary csercsik@scl.sztaki.hu

taken into account separately, and in a *modular* way. This structure allows the integration of most available literature data into the model.

- The properties of specific ionic currents *can be measured separately* via voltage clamp (reversal potential-based) methods combined with pharmacological (TTX, TEA, etc. based) methods. These types of measurements can gather data corresponding to specific elements of the model. This implies the benefit of the opportunity, that various elements of the model can be identified separately, using different parameter estimation methods, if needed.
- Because the different ion channels are described by different elements of the model, the model can be extended with equations describing the effect of estradiol, acting on specific ion channels.

2) *Elements of the model:* According to the literature data, previous results point to the existence of the following conductance elements in the GnRH neuron:

- Na⁺ channel: According to [14], [15], a simple voltage gated inward rectifier Na⁺ channel can be assumed, with standard characteristics. The current related to this channel will be denoted by I_{Na1} .
- Based on the results of [15], [16], [17], [14], [8], a voltage gated transient or rapidly activating/ inactivating K⁺ conductance is also taken into account, responsible for the rapid, transient component of the outward K⁺ current (I_{K1}).
- A voltage gated delayed outward rectifier K⁺ channel can be assumed, which contributes to the more slowly activating, sustained component of the outward K⁺ current (I_{K2}) - see [15], [16], [17], [14], [8].
- According to [18], [19], [8], a low voltage gated (T-type) Ca²⁺ channel can be assumed, which is activated in earlier phases of depolarization (I_{Ca1}).
- Furthermore, based on [20], [18] we assume a high voltage gated Ca²⁺ channel representing R and N type conductances (I_{Ca2}).
- Lastly, a high voltage gated, long-lasting current(L-type) Ca²⁺ channel is modelled (I_{Ca3}) - see [4], [19].
- In addition, a leakage current with constant conductance is also taken into account (I_L).

The **equivalent electric circuit** of a one-compartment GnRH neuron model with all the above conductances is shown in Fig. 1.

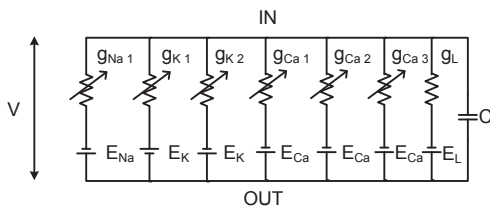


Fig. 1. Parallel conductance model, with conductances representing different ion channels

Literature data of qualitative features and parameters related to some of the above ion channels can be found in [21], [22], [18].

3) *Model equations:* The model depicted in Fig. 1 can be described by the following equations:

$$\frac{dV}{dt} = -\frac{1}{C}(I_{Na1} + I_{K1} + I_{K2} + I_{Ca1} + I_{Ca2} + I_{Ca3} + I_L) + \frac{1}{C}I_{ex} \quad (1)$$

$$\frac{dm_i}{dt} = (m_{i\infty} - m_i)/\tau_{mi}, \quad \frac{dh_i}{dt} = (h_{i\infty} - h_i)/\tau_{hi} \quad (2)$$

where V is the the membrane voltage, C is the membrane capacitance, I_{Na1} denotes the sodium current, I_{K_i} denotes the various potassium currents, I_{Ca_i} stands for the calcium currents, I_L for the leakage. The m_i and h_i variables are the activation and inactivation variables of the corresponding currents. $m_{i\infty}$, $h_{i\infty}$ and τ_{mi}/τ_{hi} denote the steady-state activation and inactivation functions, and the voltage dependent time constants of activation and inactivation variables, which are nonlinear Boltzmann and Gauss-like functions of the membrane potential:

$$a_{\infty i} = \frac{1}{1 + e^{\frac{V_{half_{ai}} - V}{K_{ai}}}} \quad (3)$$

$$a \in \{m, h\}, \quad i \in \{1, 2, 3, 4, 5, 6\}, \quad K_{mi} > 0, K_{hi} < 0 \quad \forall i$$

$$\tau_{ai} = C_{base_{ai}} + C_{amp_{ai}} e^{\frac{-(V_{max_{ai}} - V)^2}{\sigma_{ai}^2}} \quad (4)$$

Finally, I_{ex} refers to the external injected current, and the indices refer to: $i = 1 - I_{Na1}$, $i = 2 - I_{K1}$, $i = 3 - I_{K2}$, $i = 4 - I_{Ca1}$, $i = 5 - I_{Ca2}$, $i = 6 - I_{Ca3}$. The currents of ionic channels are given by

$$\begin{aligned} I_{Na1} &= \bar{g}_{Na1} m_1^3 h_1^2 (V - E_{Na}), & I_{K1} &= \bar{g}_{K1} m_2 h_2^2 (V - E_K) \\ I_{K2} &= \bar{g}_{K2} m_3 h_3 (V - E_K), & I_{Ca1} &= \bar{g}_{Ca1} m_4 h_4 (V - E_{Ca}) \\ I_{Ca2} &= \bar{g}_{Ca2} m_5 h_5 (V - E_{Ca}), & I_{Ca3} &= \bar{g}_{Ca3} m_6 h_6 (V - E_{Ca}) \\ I_L &= \bar{g}_L (V - E_L) \end{aligned} \quad (4)$$

where the E_{Na} , E_K , E_{Ca} and E_L denote the reversal potentials of the corresponding ions and the leakage current.

C. Voltage and current clamp recordings

The identification was based on both voltage clamp and current clamp measurements. In the case of voltage clamp, the term I_{ex} in eq. (1) was defined to force the clamping voltage to the membrane: $I_{ex} = p(V_{clamp} - V)$. Considering this modification, with a p big enough, the desired voltage step could be simulated with the same model used later for current clamp simulations.

D. Parameter estimation method

The estimated parameters were the membrane capacitance C in (1), the maximal conductances \bar{g}_i where $i \in \{Na1, K1, K2, Ca1, Ca2, Ca3, L\}$, in Eqs (4), and the activation parameters V_{half} , K , C_{base} , C_{amp} , σ , V_{max} in (3). An optimization based parameter estimation procedure was used to identify the model parameters. The objective function related to the VC recordings was based on the 2-norm of the difference of simulated and measured signals, and the objective function related to the CC recordings was based on the rate and time of action potentials, depolarization and repolarization values.

1) *Initial values for the optimization:* Before applying the optimization algorithm, intuitive rough-tuning of the activation parameters (parameters of the Boltzmann and Gauss functions) and conductance values was performed to capture some important features of the neural behavior (e.g. the model should fire action potentials as response to the exciting current, the firing frequency, the local maxima after the appearance of the exciting current should be similar). This preparation proved to be necessary for convergence to an acceptable optimum.

This laborous procedure was mainly based on qualitative considerations.

The determination of suitable initial values was decomposed into two phases. First, the activation parameters were chosen and then, the maximal channel conductances were estimated from VC and CC data. After a result for conductance values, the activation parameters were further tuned via numerical optimization.

2) *Optimization algorithm:* Since the activation variables can not be measured, a simulation based minimization of the objective function was performed. Because of the model nonlinearity, the objective function value can be a complicated function of the estimated parameters. Moreover, the precise simulation of the system dynamics for a given parameter set is computationally quite demanding, i.e. a few hundred evaluations of the objective function takes a couple of hours on a typical desktop PC. This also means that avoiding the numerical approximation of the gradients of the objective function was desirable in our case. A promising choice is the freely available Asynchronous Parallel Pattern Search (APPS) algorithm for parameter estimation. As described in [23], the parallellpattern search (PPS) is a useful tool for derivative-free optimization where the number of variables is not large (about fifty or less) and the objective function is expensive to evaluate, both of which hold in our case.

As described by [24], the APPS algorithm is an asynchronous extension of the PPS method that efficiently handles situations when the individual objective function evaluations may take significantly different time intervals and therefore it is very suitable to be implemented in a parallel or grid environment. Furthermore, recent implementations of the APPS method handle bound and linear constraints on the parameters. The global convergence of APPS under standard assumptions is also proved by [25].

III. RESULTS AND DISCUSSION

As a result of the initializing and the algorithm a parameter set was found, which provided good approximation of measured data and several qualitative features (e.g. firing only during current injected, firing frequency, resting potential, DPAPs) during the simulations.

It has to be noted that the simulations were started from the initial voltage corresponding to the observed resting potential, and all activation values were set to their steady-state values, corresponding to this voltage.

The measured and simulated results of the voltage clamp are depicted in Figs. 2 and 3. The measured and simulated results of current clamp are depicted in Fig. 4.

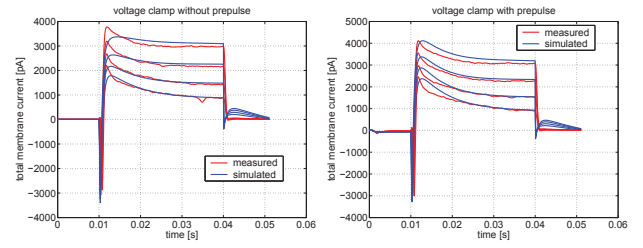


Fig. 2. Measured and simulated membrane currents in the case of VC without and with prepulse in the medium voltage range: red line - measured, blue - simulated

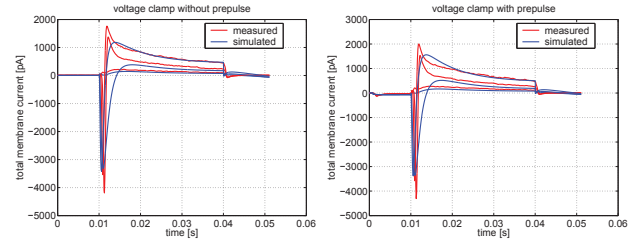


Fig. 3. Measured and simulated membrane currents in the case of VC without and with prepulse in the low voltage range

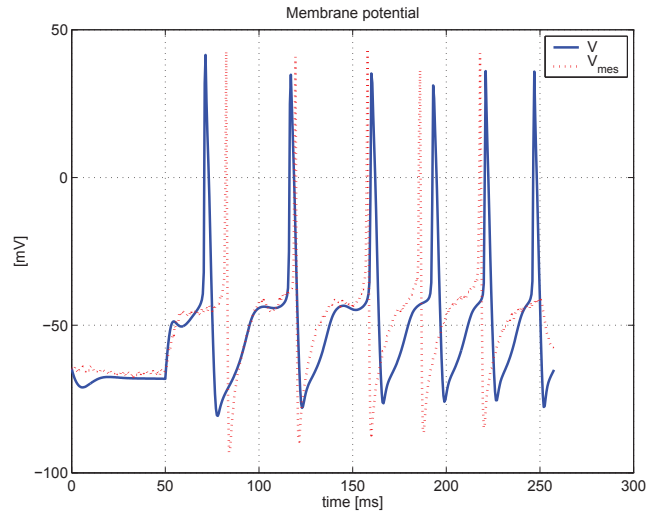


Fig. 4. Measured and simulated membrane voltage in the case of CC

It is visible in Figs. 2 and 3, that the model can not fully reproduce the high-frequency components in the CC and VC traces. Because of the noise and postsynaptic currents this is acceptable, but regarding the fast currents, the model can still be improved. This model error can be related to the smaller de- and repolarization amplitudes during the AP.

The amplitude and the qualitative features of the Ca currents during VC simulation (not depicted here) are in good agreement with the results of [20], [18].

A phenomena, which indicates a further model error, is the tail current in the VC simulations after 40 ms. These currents can not be observed in the measurements - in the simulations they can be related to the delayed-rectifier conductance.

IV. CONCLUSIONS AND FUTURE WORK

In this article, a Hodgkin-Huxley type model of the GnRH neuron and its parameter estimation procedure was proposed. An important aim during the work was to incorporate up-to-date biological literature data into the model. The initial values for parameter estimation were determined using known results and prior knowledge about the modeled system. It is emphasized, that the used mathematical model shows acceptable fit to the measurements with the same estimated parameter values both in the case of VC and CC that can be rarely found in the literature, although a high number of parameters had to be tuned to reach the appropriate behavior. This suggests that the model is possibly overparametrized. The resulting parameter set showed great sensitivity to the initial conditions of the optimization, which were tuned using qualitative considerations.

In the future, we intend to develop an at least semi-automatized identification method which would be based on both voltage and current clamp recordings. Since the objective function seems to be quite fragmented as a function of the parameters, the appropriate re-parametrization of the model and/or the application of other parameter estimation methods that do not require the simulation of continuous-time models may be necessary in the future.

ACKNOWLEDGMENT

The author would like to thank Imre Farkas and Zsolt Liposits at the Institute of Experimental medicine for the availability of whole-cell patch clamp recordings. GnRH-enhanced green fluorescent protein (GnRH-GFP) transgenic mice were a kind gift by Dr. Suzanne Moenter.

REFERENCES

- [1] E. Knobil, "The neuroendocrine control of the menstrual cycle," *Hormone Research*, vol. 36, pp. 53–88, 1980.
- [2] —, "The hypothalamic gonadotropin releasing hormone (GnRH) pulse generator in the rhesus monkey and its neuroendocrine control," *Human Reproduction*, vol. 3, pp. 29–31, 1988.
- [3] S. Stojilkovic, L. Krsmanovic, D. Spergel, and K. Catt, "GnRH neurons: intrinsic pulsatility and receptor-mediated regulation," *Trends in Endocrinology and Metabolism*, vol. 5, pp. 201–209, 1994.
- [4] L. Krsmanovic, S. Stojilkovic, F. Merelli, S. Dufour, M. Virmani, and K. Catt, "Calcium signaling and episodic secretion of gonadotropin-releasing hormone in hypothalamic neurons," *Proceedings of the National Academy of Sciences of the USA*, vol. 89, pp. 8462–8466, 1992.
- [5] D. Brown, A. Herbison, J. Robinson, R. Marrs, and G. Leng, "Modelling the lutenizing hormone-releasing hormone pulse generator," *Neuroscience*, vol. 63, pp. 869–879, 1994.
- [6] A. Khadra and Y. Li, "A model for the pulsatile secretion of gonadotropin-releasing hormone from synchronized hypothalamic neurons," *Biophysical Journal*, vol. 91, pp. 74–83, 2006.
- [7] J. Gordan, B. Attardi, and D. Pfaff, "Mathematical exploration of pulsatility in cultured gonadotropin-releasing hormone neurons," *Neuroendocrinology*, vol. 67, pp. 2–17, 1998.
- [8] A. Herbison, J. Pape, S. Simonian, M. Skynner, and J. Sim, "Molecular and cellular properties of GnRH neurons revealed through transgenics in mouse," *Molecular and Cellular Endocrinology*, vol. 185, pp. 185–194, 2001.
- [9] K. Suter, J. Wuarin, B. Smith, F. Dudek, and S. Moenter, "Whole-cell recordings from preoptic/hypothalamic slices reveal burst firing in gonadotropin-releasing hormone neurons identified with green fluorescent protein in transgenic mice," *Endocrinology*, vol. 141, pp. 3731–3736, 2000.
- [10] P. Mellon, J. Windle, P. Goldsmith, C. Padula, J. Roberts, and R. Weiner, "Immortalization of hypothalamic GnRH neurons by genetically targeted tumorigenesis," *Neuron*, vol. 5, pp. 1–10, 1990.
- [11] E. Izhikevich, "Neural excitability, spiking and bursting," *International Journal of Bifurcation and Chaos*, vol. 10, pp. 1171–1266, 2000.
- [12] —, *Dynamical Systems in Neuroscience*. 999 Riverview Drive Suite 208 Totowa New Jersey 07512: The MIT Press, 2005.
- [13] A. Hodgkin and A. Huxley, "A quantitative description of membrane current and application to conduction and excitation in nerve," *Journal of Physiology*, vol. 117, pp. 500–544, 1952.
- [14] M. Bosama, "Ion channel properties and episodic activity in isolated immortalized gonadotropin-releasing hormone (GnRH) neurons," *Journal of Membrane Biology*, vol. 136, pp. 85–96, 1993.
- [15] K. Kusano, S. Fueshko, H. Gainer, and S. Wray, "Electrical and synaptic properties of embryonic lutenizing hormone-releasing hormone neurons in explant cultures," *Proceedings of the National Academy of Sciences of the USA*, vol. 92, pp. 3918–3992, 1995.
- [16] J. Constantin and A. Charles, "Modulation of Ca^{2+} signaling by K^+ channels in a hypothalamic neuronal cell line (GT-1)," *Journal of Neurophysiology*, vol. 85, pp. 295–304, 2001.
- [17] J. Sim, M. Skynner, and A. Herbison, "Heterogeneity in the basic membrane properties of postnatal gonadotropin-releasing hormone neurons in the mouse," *The Journal of Neuroscience*, vol. 21, pp. 1067–1075, 2001.
- [18] M. Kato, K. Ui-Tei, M. Watanabe, and Y. Sakuma, "Characterization of voltage-gated calcium currents in gonadotropin-releasing hormone neurons tagged with green fluorescent protein in rats," *Endocrinology*, vol. 144, pp. 5118–5125, 2003.
- [19] F. Van Goor, L. Krsmanovic, K. Catt, and S. Stojilkovic, "Control of action potential-driven calcium influx in gtl neurons by the activation status of sodium and calcium channels," *Molecular Endocrinology*, vol. 13, pp. 587–603, 1999.
- [20] M. Watanabe, Y. Sakuma, and M. Kato, "High expression of the R-type voltage-gated Ca^{2+} channel and its involvement in Ca^{2+} -dependent gonadotropin-releasing hormone release in GT1-7 cells," *Endocrinology*, vol. 145, pp. 2375–2388, 2004.
- [21] H. Rehm and B. Tempel, "Voltage-gated k^+ channels of the mammalian brain," *FASEB J.*, vol. 5, pp. 164–170, 1991.
- [22] K. Talavera and B. Nilius, "Biophysics and structure-function relationship of T-type Ca^{2+} channels," *Cell Calcium*, vol. 40, pp. 97–114, 2006.
- [23] P. D. Hough, T. G. Kolda, and V. J. Torczon, "Asynchronous parallel pattern search for nonlinear optimization," *SIAM Journal on Scientific Computing*, vol. 23, pp. 134–156, 2000.
- [24] T. G. Kolda, "Revisiting asynchronous parallel pattern search for nonlinear optimization," *SIAM J. Optim.*, vol. 16, pp. 563–586, 2005.
- [25] T. Kolda and V. Torczon, "On the convergence of asynchronous parallel pattern search," *SIAM J. Optim.*, vol. 14, pp. 939–964, 2004.

Application of software defined radio in wireless sensor networks, superposition coding

Dávid Tisza

(Supervisor: Dr. János Levendovszky)

david.tisza@itk.ppke.hu

Abstract—In this paper a brief introduction is given about the Software Defined Radios and an existing rapidly developing platform the USRP. Application examples presented for the usefulness of the SDR in wireless sensor networks, and a theoretically well known multiuser technique the superposition coding is introduced and a design of a software defined radio-based implementation of superposition coding using the GNU radio architecture. In theory, multiuser techniques such as superposition coding are known to improve throughput in wireless networks. However, in order to understand their practical limitations, it is imperative to actually implement and test such techniques in a realistic setting. In the process, we also describe software and hardware issues associated with superposition coding implementation on a USRP.

I. INTRODUCTION

Software Defined Radio is not a new concept, it has roots way back to the late 1970's at the defense sector, but the first public attempt was the SPEAKEasy project [1] around 1991. The motivations for using software defined radios (SDR) were not changed dramatically. The SPEAKEasy project was targeted to use programmable processing to emulate more than 10 military radios and to be able to incorporate new modulation techniques. Even today primarily SDRs are used for rapid prototyping, testing of new ideas and building highly dynamic, complex systems which would be very expensive (however more effective) if it would be built in hardware. The universal purpose processors where the main processing takes place can be programmed and the program can be changed easily while their hardware counterpart not. With the vast increase of computing capability at the general purpose computers and the cheaper IC design, SDRs became more and more accessible to the public. These properties of the SDRs make them an ideal platform for academic research and for low cost testing, learning, debugging and interacting with other radios such as wireless sensor networks.

II. SOFTWARE DEFINED RADIO IN NUTSHELL

SDRs are mainly consist of three parts, see Figure 1.

- RF front end with antenna(s)
- A/D and D/A converter
- an universal purpose computer with high enough computing capability

The purpose of the RF front end is to “down or up convert” a signal to a specified center frequency. A typical RF front end can be viewed at Figure 2. The A/D converter is the connection

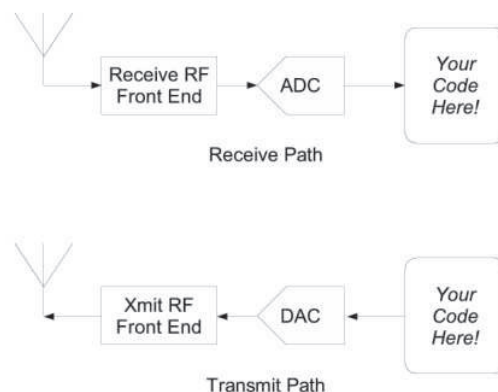


Fig. 1. SDR from 1000m away

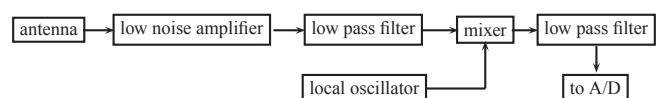


Fig. 2. A typical receiver RF front end

between the physical world of continuous analog signals and the world of discrete digital samples manipulated by software. Then the samples enter the well known digital domain, where the general purpose computer does all the processing, let it be demodulation, filtering, decoding, aligning, buffering, packetizing, etc. There exist and still under development a software radio platform and an open source program library maintained by a community, called USRP(Universal Software Radio Platform)[2] and GnuRadio[3].

The USRP is briefly consist of a changeable RF front end, an A/D converter and an FPGA that's purpose is to further narrow the bandwidth of the digitized stream's bandwidth so that it can be fit into the connection media between the USRP and the PC. For the first version it is an USB 2.0 link, and for the second generation USRP's that is a gigabit ethernet link. The USRP version 1 can be seen at Figure 3.

The GnuRadio itself is a program library consisting of signal processing blocks designed to be interconnected and there is a scheduler that is maintaining and ensuring the data flow between the signal processing blocks.

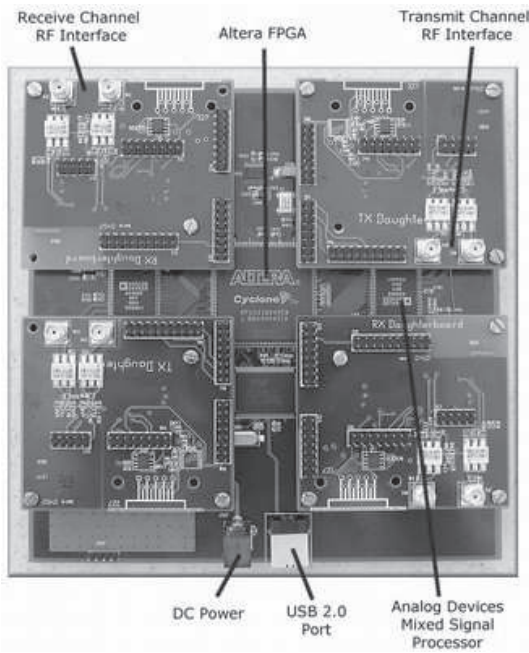


Fig. 3. The USRP version 1 - SDR platform

III. APPLICATION EXAMPLES FOR WIRELESS SENSORIAL NETWORKS

In wireless sensorial networks the underlying radio modules are not too complex due to the serious energy constrains. Thus one can easily implement a physical layer structure for these radios so our radio platform will be able to communicate with these radios as well. For example the Berkeley Mica2 mote uses the chipcon cc1000 radio chip, which uses a simple fsk modulation and differential coding. Physical layer software components were developed during my learning phase for the USRP and GnuRadio.

In addition that we can communicate to the motes we can run more sophisticated algorithms because our general purpose computer is much more powerful than the micro controllers on the motes, and one can exploit the fact that the USRP can “listen” at a much wider band than a mote, so it could listen to multiple parallel transmissions at different frequencies. This enables us to use the SDR as a gateway between mote groups that communicate at different frequencies and otherwise could not communicate with each other.

Furthermore if one takes into account that the SDR can run multiple algorithms on the same chunk of band or using two different RF front end cards, on different chunks of bands, one can easily bridge between two different mote types which use different physical access methods but speak the same higher hierarchy languages (eg. the frame and data structures are the same).

Or even an SDR can remain transparent for the network, and “intelligently” listen to transmissions, harness informations about the network, the links and if it decides that one communication channel needs to be helped for example because of a changed environmental fading, it can act as a repeater between the two communicating parties so that they won’t notice the change in the environment.

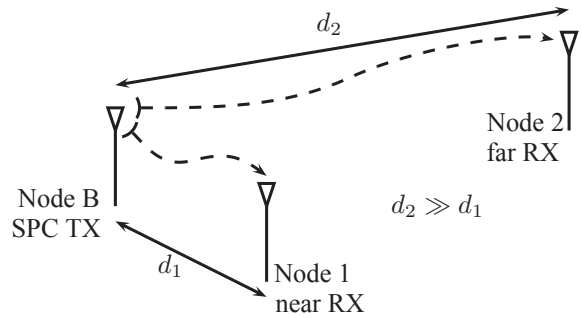


Fig. 4. Superposition Coding model

IV. AN IMPLEMENTED MULTIUSER TECHNIQUE - SUPERPOSITION CODING

In this section, a brief description of SPC is provided. Consider two users sharing a wireless channel with received $\text{SNR}_1 \gg \text{SNR}_2$. For the purpose of exposition, we assume each point-to-point channel to be AWGN. For a fixed transmit power, suppose that the messages of user 1 and user 2 are sent over the channel simultaneously, such that the SNR degradation at user 1 (due to decreased power allocation) is nearly equal to the SNR improvement at user 2 (due to increased power allocation). Assuming ideal decoding, the achievable rate at user 1 decreases logarithmically (high SNR regime) while that at user 2 increases linearly (low SNR regime), and hence the aggregate rate increases. This effect can also be seen in non-capacity approaching scenarios. Consider a scenario with a single base station B and two users, 1 and 2, as shown in Figure 4. User 1 (near user), being closer to the base station B (as depicted) will have a higher receiver SNR compared to user 2 (far user). Denote the modulated symbol streams of the two users by $\{x_1\}$ and $\{x_2\}$ respectively, each of unit power. The k^{th} symbol transmitted by the base station B can be expressed as

$$s_k = \sqrt{\alpha_1}x_{1,k} + \sqrt{\alpha_2}x_{2,k}$$

where the α_i is the fraction of the transmit power allocated to user $i, i = 1; 2$. Without loss of generality, we assume a unit power constraint ($\alpha_1 + \alpha_2 = 1$). The k^{th} received symbols at the near-user and the far-user as

$$\begin{aligned} r_{1,k} &= s_k + w_{1,k} \\ &= \sqrt{\alpha_1}x_{1,k} + \sqrt{\alpha_2}x_{2,k} + w_{1,k} \end{aligned}$$

and

$$\begin{aligned} r_{2,k} &= s_k + w_{2,k} \\ &= \sqrt{\alpha_1}x_{1,k} + \sqrt{\alpha_2}x_{2,k} + w_{2,k} \end{aligned}$$

where $w_{i,k}; i = 1; 2$ are unit variance circularly symmetric complex Gaussian variables with variance σ^2 . Denote the power of stream i observed at receiver j as P_{ij} . It is easy to see that $P_{ij} = \alpha_i$. Since the power allocated to user 2 acts as interference to user 1 and vice versa, we have

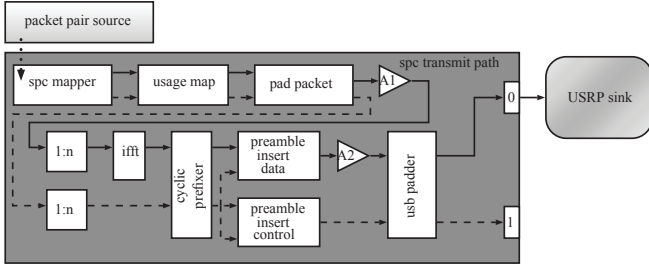


Fig. 5. Transmitter block diagram

$$\frac{P_{11}}{\sigma^2 + P_{21}} = \frac{\alpha_1}{\sigma^2 + \alpha_2} = \frac{\alpha_1}{\sigma^2 + 1 - \alpha_1}$$

$$\frac{P_{22}}{\sigma^2 + P_{12}} = \frac{\alpha_2}{\sigma^2 + \alpha_1} = \frac{1 - \alpha_1}{\sigma^2 + \alpha_1}$$

The above two equations represent the signal to interference and noise ratio (SINR) at user i to detect stream i : Suppose that the near-user is much closer to the base station than the far-user. Then, if both the users were to see near-equal link quality, most of the transmit power must be allocated to the far-user, i.e., $\alpha_2 \gg \alpha_1$: Together with the transmit power constraint, this implies that

$$\frac{P_{11}}{P_{21}} = \frac{\alpha_1}{\sigma^2 + 1 - \alpha_1} \approx \frac{\alpha_1}{\sigma^2 + 1} + \left(\frac{\alpha_1}{\sigma^2 + 1} \right)^2 \approx \frac{\alpha_1}{1 + \sigma^2}$$

and

$$\frac{P_{22}}{P_{12}} = \frac{1 - \alpha_1}{\sigma^2 + \alpha_1} \approx \frac{1}{\sigma^2 + \alpha_1}$$

where we have used $\alpha_1 \ll 1$ and a subsequent binomial approximation to obtain the first expression. So, we have $P_{22}/P_{12} > P_{11}/P_{21}$ which implies that the detection of far user symbols from the symbol stream $\{r_1\}$ will be more accurate than the detection of near-user symbols from symbol stream $\{r_2\}$. Therefore, user 2 symbols can be detected accurately from $\{r_1\}$, and their effect can be canceled to yield

$$\begin{aligned} \tilde{r}_{1,k} &= r_{1,k} - \sqrt{\alpha_2} \hat{x}_{2,k} \\ &= \sqrt{\alpha_1} x_{1,k} + \sqrt{\alpha_2} (x_{2,k} - \hat{x}_{2,k}) + w_{1,k} \\ &\stackrel{(a)}{=} \sqrt{\alpha_1} x_{1,k} + w_{1,k} \end{aligned}$$

where (a) holds for most of the symbols (there will be an occasional symbol error)

A. SPC transmitter design

The block diagram of the SPC transmitter is shown in Figure 5. The payload pair (near and far user data in bits) is provided to the physical layer by the higher layers and the SPC mapper multiplexes the bits intended for the two users, and maps it to the superposition constellation, shown by Figure 6, 7. At the figures it is assumed, that the near user uses a QPSK modulation and the far user uses a BPSK modulation, but this is just for illustrations. The super constellation and modulation types used by the SPC mapper can be changed dynamically. An OFDM modulation scheme was used atop

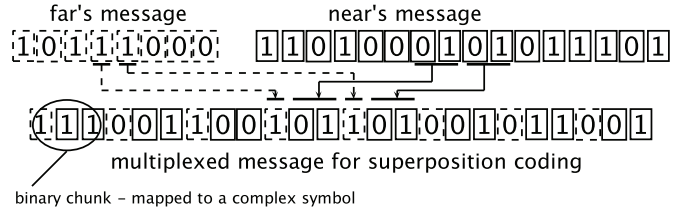


Fig. 6. Multiplexing the two user's bitstream

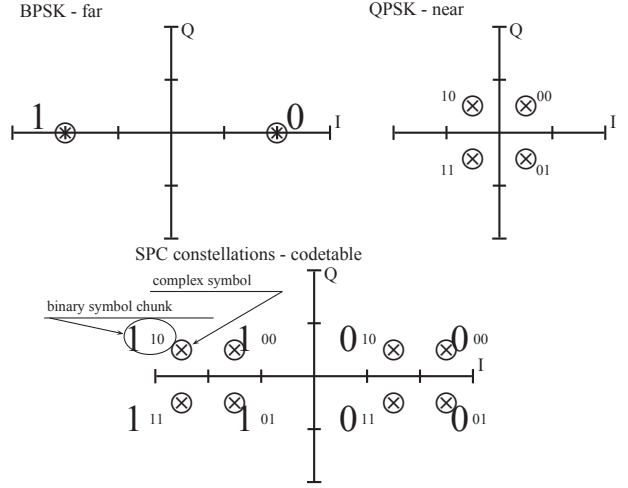


Fig. 7. SPC constellation mapping

of the SPC modulation, to have a flexible and easily expandable system. SPC modulation was used on each OFDM sub carrier. Since some of the tones in OFDM may be used as pilots (for frequency tracking), the usage map is employed to specify the tones over which data may be transmitted. The OFDM modulator, inserts the preamble (used for channel and frequency synchronization), modulates the SPC symbols and inserts the cyclic prefix. USRP is connected to the PC via a USB and this mandates that the frame transmitted via USB to USRP be a multiple of 512 bytes. We use the USB padder so as to satisfy this constraint. Since the GNU radio framework is a flow-based implementation, for a packet based system, it is imperative to clearly indicate the beginning and ending of packets and a control channel is used for this purpose.

An OFDM frame structure can be found at Figure 8

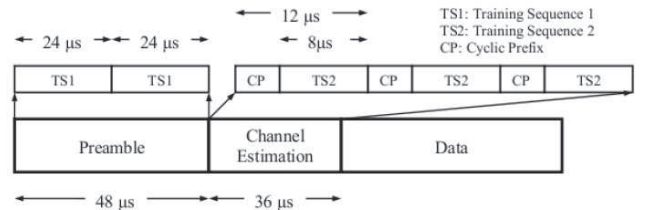


Fig. 8. Used OFDM-SPC frame

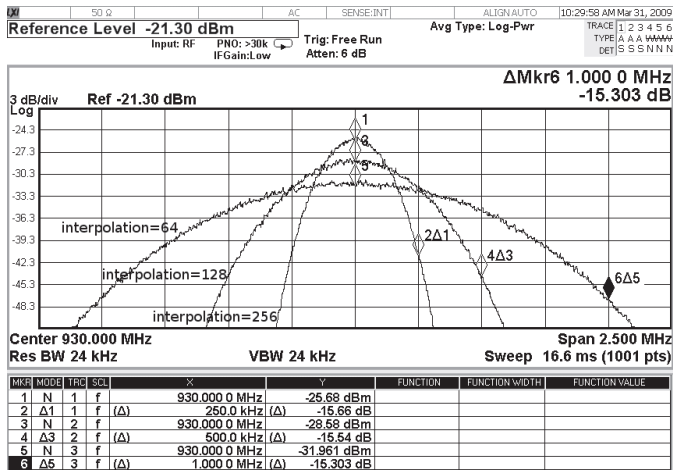


Fig. 9. Frequency response of the USRP TX path for different interpolation factors. For our SPC experiment, we used an interpolation factor of 128, which corresponds to a BW of 1 MHz. We observe that the 3-dB bandwidth is about 450 kHz for an interpolation factor of 128. We also observe that this frequency selectivity is consistent over different interpolation factors.

B. Measurement results

The effective bandwidth of the USRP is much smaller than that set by the user. The cause of this problem lies in the highly non-ideal transmit path implementation of the USRP. The DACs on the transmit path are designed to operate at a fixed frequency of 128 MHz [3]. Therefore, any digitally synthesized signal at a lower bandwidth to be input to the DACs must be interpolated to 128 MHz. However, we observe that the USRP uses a rather simplistic scheme to implement this interpolation. In Figure 9, we have plotted the frequency response of the USRP over different user bandwidths. Notice that the characteristics show a poor passband response. For example, for a user-requested bandwidth of 1 MHz, the 3-dB bandwidth as only 450 kHz. Indeed, this effect is seen in the channel response estimate at the receiver, shown in Figure 10. It is therefore not surprising that such a frequency response causes significant degradation of sub carrier SNR as one moves away from the DC sub carrier. Similar problems were reported recently [4]. In a SPC system, the near user typically enjoys a much greater SNR when compared to the far user. Therefore, distortions in near user's receive symbols due to RF impairments are more important since they fundamentally limit the performance of the near user, and in turn, the achievable gains from SPC. In this paper, we briefly describe the effect of one such RF non-ideality: the carrier frequency offset relative to the transmitter. For nonzero residual frequency offset, the received constellation will rotate over time. Given that the minimum distance of near user symbols is smaller than that of the far user, rotations will have a greater impact on near user performance. While this issue could be addressed by tracking the carrier frequency offset in time, such schemes would entail additional algorithmic complexity and/or throughput loss (owing to additional training symbols).

V. CONCLUSION

In this paper, we have investigated the basics of the software defined radios, some of their basic applicability in wireless

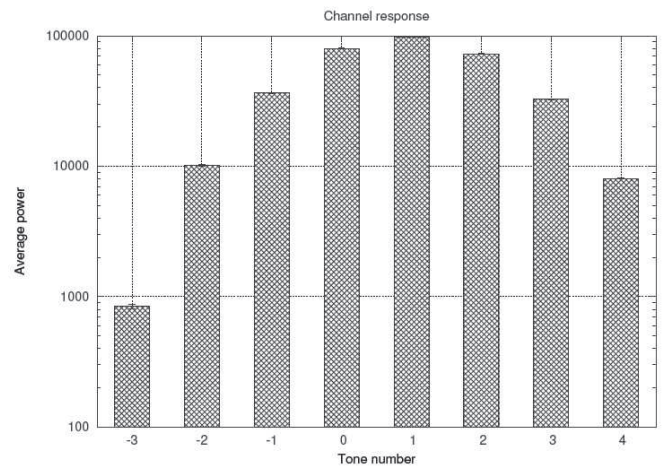


Fig. 10. Estimated channel response (subcarrier-wise) at the RX baseband. We observe that the relative power between the tones vary significantly and this is caused by the non-ideal frequency response of the USRP TX path.

sensory networks and an open source SDR platform called GnuRadio and the design of an OFDM based SPC transmitter on the GnuRadio platform. Potential design issues were identified in implementing an SPC transmitter on this platform. The next steps include improving the transmitter performance with FEC and implementing medium access and routing protocols. It would also be interesting to study the effect of node placements on SPC and to design medium-access schemes that exploit SPC.

REFERENCES

- [1] F. Torre, "Speakeasy-a new direction in tactical communications for the 21st century," in *Tactical Communications Conference, 1992. Vol. 1 Tactical Communications: Technology in Transition., Proceedings of the*, Apr 1992, pp. 139–142 vol.1.
- [2] "<http://www.ettus.com/>."
- [3] "<http://gnuradio.org/>."
- [4] K. Mandke, R. C. Daniels, S. M. Nettles, and R. W. H. Jr., "On the challenges of building a publications multi-antenna software defined packet radio," in *Proceedings of the SDR 08 Technical Conference and Product Exposition*, October 2008.

Detection and Classification of Living Organisms for Water Quality Monitoring

Vilmos Szabó

(Supervisors: Dr. Roska Tamás, Dr. Tőkés Szabolcs, Dr. Szatmári István)

szavi@digitus.itk.ppke.hu

Abstract—In this paper I present an algorithmic framework for automated detection, recognition and counting of living organisms in drinking water. The system consists of the following main sections: sensing images using normal light microscope, detecting and segmenting objects in the image, feature extraction and classifying individual organisms and particles in the water. The aim of the system is to continuously characterize living populations of algae in real-time. The current framework is able to process video files, detect objects, store them in a database, and classify them accordingly. Water quality analysis and monitoring is also performed.

Index Terms—Biological image segmentation, Classification, Feature extraction, Water quality, Living organisms

I. INTRODUCTION

Traditionally normal optical microscopes are used in identifying alga population in water [8], [9], [11]. These methods can only measure data from slices of specimen. The newest microscopes are digital holographic microscopes which capture data from a three dimensional volume. This data can be used to detect the axial and lateral position of an object, selecting only region-of-interest areas. This can greatly reduce the computational demand of the system. This paper is organized in the following manner: the next section gives an overview of the microscope system. Section three describes the image database. The segmentation algorithm is discussed in section four, while section five shows the feature extraction. Next is the classification, and the experimental results, followed by the conclusions and the future work plans.

II. OVERVIEW OF THE MICROSCOPE AND MEASUREMENT SYSTEM

Most simple and commonly used optical microscopes work in the spectrum of visible light (380-750 nm). These microscopes use refractive glass to focus light onto a CCD or CMOS image detector. The magnification of optical microscopes can be very high (1000-1500x), but there is a theoretical diffraction limit for optical magnification (assuming visible light spectrum) which is 200 nm. The use of shorter wavelength can increase the spatial resolution of the microscope system allowing the detection of smaller objects. More recently fluorescence is used in classification of particles containing chlorophyll which can highly increase the accuracy of detection, segmentation and identification.

The water for examination is pumped through a glass flow chamber by a peristaltic pump. The microscope objective (in our experiments: 10x, 20x magnification and a Numerical Aperture of 0.25, 0.4 were used) projects the image on a 2 mega pixel CCD colour or greyscale camera. The image is

recorded by the camera and the detected objects are further analyzed by the software.

In order to make a good training set of images a few considerations should be taken to account. The density of the algae should not exceed 5-10 samples per image otherwise touching algae will be segmented together. The illumination should be constant and homogeneous. The flow rate of the water should be set accordingly to the cameras frame rate. To get the actual size of an object the system needs to be calibrated.

III. IMAGE DATABASE

The images detected by the CCD are analyzed by the segmentation algorithm (see section IV. Segmentation algorithm). The detected object and its corresponding binary mask image are stored in a database for training and testing the classification algorithms. Both the original colour image and the output mask of the segmentation are stored in Portable Network Graphics format (png). The original video flows are stored in uncompressed video format. The training set is extracted from a video flow therefore it will contain multiple views, scales and rotations (three dimensional rotations) from the same organism. Figure 1 contains the four main alga classes (Selenastrum capricornutum, Scenedesmus obtusiusculus, Chlamydomonas, Scenedesmus armatus) that have been analyzed in this paper.

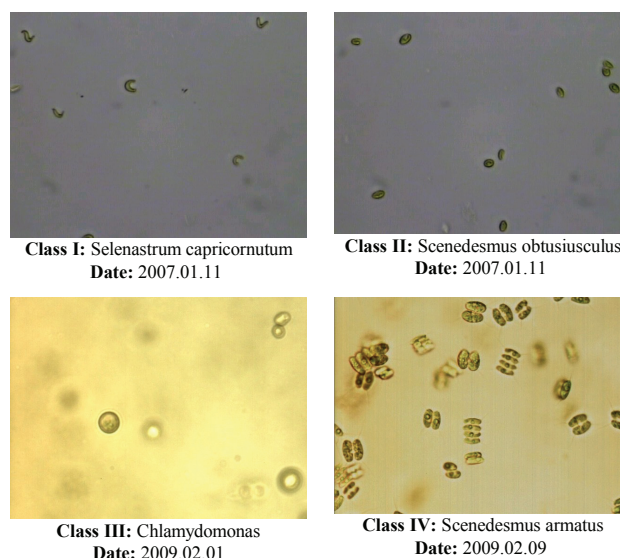


Fig. 1. Images of algae used for classification. Top left: Selenastrum capricornutum, Top right: Scenedesmus obtusiusculus, Bottom left: Chlamydomonas, Bottom right: Scenedesmus armatus

Table 1, contains the taxonomical scheme for these algae and the number of image samples that were used in the classification experiments.

TABLE I.
ALGAE SAMPLES IN THE DATABASE

Class Number	Taxonomy of Classified Organisms	Number of Samples
Class I	Phylum: Chlorophyta	286
	Class: Chlorophyceae	
	Order: Chlorococcales	
	Family: Oocystaceae	
	Genus: Selenastrum	
	Species: Selenastrum capricornutum	
Class II	Phylum: Chlorophyta	1176
	Class: Chlorophyceae	
	Order: Chlorococcales	
	Family: Scenedesmaceae	
	Genus: Scenedesmus	
	Species: Scenedesmus obtusiusculus	
Class III	Phylum: Chlorophyta	114
	Class: Chlorophyceae	
	Order: Volvocales	
	Family: Chlamydomonadaceae	
	Species: Chlamydomonas	
Class IV	Phylum: Chlorophyta	488
	Class: Chlorophyceae	
	Order: Chlorococcales	
	Family: Scenedesmaceae	
	Genus: Scenedesmus	
	Species: Scenedesmus armatus	
	Total number of image in database:	2064

IV. SEGMENTATION ALGORITHM

The first step in the pattern recognition is to separate the possible objects (e.g. alga, fungi, pollen) from the background. Through the segmentation process a mask image is created where the background has zero valued pixels and the foreground has one valued pixels. The background is usually not homogeneous therefore a simple global thresholding method fails to solve the segmentation problem. The current segmentation algorithm extracts high frequency (a) and low frequency (b) components separately.

A. High Frequency Component

The high frequency component is extracted with an edge detector. Possible edge detectors can be Sobel, Prewitt, Roberts or the Canny edge detector. The Canny edge detector achieved the best performance in our experiments.

B. Low Frequency Component

The low frequency components are extracted with Gaussian low pass filter. The sigma parameter of the Gaussian filter was determined empirically. The low and high frequency channels are fused in a linear α -parameter homotopy ($\alpha=0.5$ was used in our experiments).

After thresholding this fused image we get a binary image where each pixel unambiguously belongs to a foreground or to a background pixel. This binary image is further enhanced by binary morphological operations to join the fragmented object and to reduce the number of miss-detections. In the binary image connected pixels are grouped and every object from both gray scale image and the binary mask image is

“cut” out and stored in the image database for feature extraction.

V. FEATURE EXTRACTION

The input image is highly redundant. The transformation to reduce the dimensionality of input data while keeping relevant information content is called feature extraction. On each sample in the database a number of features are extracted from the original gray scale image and the mask image. Some simple features that are extracted include size, length, and width of the object. Some of the more complex features extracted include colour information, contour descriptors and statistical moments.

The size of the object is expressed in Equivalent Circle Diameter (ECD). It can be calculated from the area of the object with the assumption that every particle is circular. The length of the object is the fitted ellipses major axis length. The width of the object is the fitted ellipses minor axis length.

The colour data is very important in case of algae. The colour images are represented in YCbCr colour-space. For each colour channel the average value and standard variation is extracted for description of colour statistics within the sample. This helps to describe the texturedness of the object.

Describing the shape of the sample, first the contour is extracted from the mask image. The vector, made from the coordinates is represented in complex form. This vector is then Fourier Transformed. With proper normalizations this type of description can be made to be size, translation, and rotation invariant. The drawback of this method is that a very accurate contour must be found.

Statistical moments can describe the inner structure of intensity values. These moments are also size, translation, and rotation invariant. A sum of 20 features is stored for each individual sample which are summarized in Table 2.

TABLE II.
EXTRACTED FEATURES FOR CLASSIFICATION

	Name of the Feature	Abbrev.
1	Eccentricity	E
2	Equivalent Circle Diameter	ECD
3	Major Axis Length	MaAL
4	Minor Axis Length	MiAL
5	Edge Density	ED
6	Colour: Y channel average	YCA
7	Colour: Y channel std	YCS
8	Colour: Cb channel average	CbCA
9	Colour: Cr channel std	CbCS
10	Colour: Cb channel average	CrCA
11	Colour: Cr channel std	CrCS
12	Histogram: Y channel	HY
13	Histogram: Cb channel	HCb
14	Histogram: Cr channel	HCr
15	Fourier Descriptor	FD0
16	Fourier Descriptor	FD1
17	Fourier Descriptor	FD2
18	Invariant Moment	M0
29	Invariant Moment	M1
20	Invariant Moment	M2

VI. CLASSIFICATION

After the features are extracted, every sample is represented by a point in d dimensional vector space. The classification task is to find the decision boundaries which separate each class from the other. It can be viewed as an optimization problem where the goal is to minimize the intraclass variation while maximizing the interclass variations.

Mainly there are two ways a pattern recognition task can be solved [1], [2], [5], [6]. The first is called supervised learning where the class of the sample is known *a-priori*. The other case is unsupervised learning where the samples are assigned to unknown classes based on the similarity metric of patterns.

Each extracted feature has a different range of values it can take. Therefore a feature corresponding to a set of samples needs to be normalized so that the classification algorithm can converge. The ranges of all features are normalized between 0 and 1. Also the samples were randomized before the learning algorithm.

Today the most used classification algorithms include feed-forward neural networks (FFNN), Kohonen-networks (KN), radial basis functions (RBF), fuzzy clustering, support vector machine (SVM) [3], [4], [10] and nearest neighbour classifiers (NN). In my experiments I tested the k-means [7], decision tree, linear and quadratic discriminant analysis, and k nearest neighbour (KNN) classifiers.

A. K-Means

K-means is one of the simplest unsupervised learning algorithms to solve clustering problems. Assuming k number of clusters, k random samples are selected as the initial starting points. As the next step each point is associated with its nearest neighbour. A new k number of centroids are generated from the centre of the cluster. For each iteration the k centre points are moving toward an equilibrium point, where the algorithm stops. Finally, the algorithm minimizes a square error function to arrive at the last stage. The algorithm can be sensitive to the selection of the initial centre point. To reduce the probability of stopping in a local optimum, the algorithm usually evaluated a number of times and the best result is chosen as final clustering.

B. Decision Tree

Decision trees are tree-like graphs. The branches represent threshold values of a feature and the leaves are the discrete categories it classifies the samples into. The structure is very easy to interpret, understand and combine with other decision techniques. Synthesis and testing of decision tree are computationally inexpensive, but its drawback is that over fitting of data can happen which reduces its generalization capability.

C. K Nearest Neighbour Classifier

The KNN classifies a data sample to its closest learning data in the feature space. KNN sometimes referred as lazy learner meaning that the function is only approximated locally. An object is classified based on the majority of its k nearest neighbours. Usually Euclidean or Mahalanobis can be used as distance metrics. The major drawback of the KNN arises when we would like to classify a new vector to a class.

The classes with the more frequent examples can dominate the prediction of a new vector. A solution for this problem can be resolved by weighting the k nearest neighbours with their distance in the feature space. The choice of k depends on the data. Higher k values will result in smoother decision boundaries but decreases the feature's distinctiveness.

In the next section, these classifiers have been evaluated using the features explained in section IV.

VII. RESULTS AND DISCUSSION

The result of the different classifiers can be found in Table III. The average classification performance shows the probability for a sample getting correctly classified. The minimal classification performance may reflect the classifiers performance more accurately. The best average classification is reached by the k nearest neighbour. The best classifier according to the minimal classification performance is the quadratic classifier.

Figures 2-6 contains the feature maps of classifiers. The different colours correspond to the alga class they are identified as. The decision tree and the linear classifier resulted in straight decision boundaries. The quadratic and the KNN classifier adapted well to the shape of feature data.

All five type of classifiers have resulted adequate separation performance for reliable object recognition.

TABLE III.
CLASSIFICATION RESULTS

Classifier	Average Classification Performance [%]	Minimal Classification Performance [%]
1 K-means	92.90%	79.97%
2 Decision Tree	94.81%	91.61%
3 Linear Discriminant Analysis	92.80%	90.48%
4 Quadratic Discriminant Analysis	94.84%	93.54%
5 K Nearest Neighbor	95.59%	89.16%

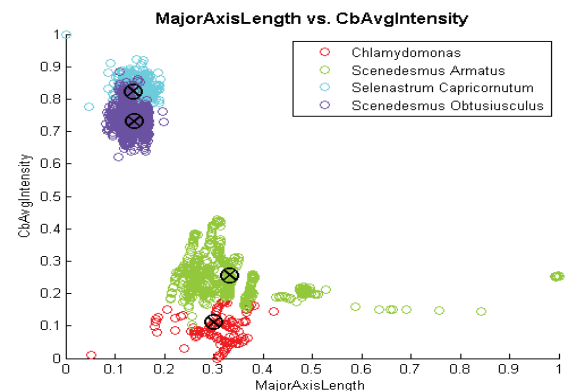


Fig. 2. The result of K-means clustering ($K = 4$). The x is showing the centre points of the clusters.

VIII. CONCLUSIONS

In this article an algorithmic framework was presented for detection and recognition of living organisms in water. After the segmentation of the original input image a total of 20 features are extracted. The algorithm was tested on 2,064 images which contained four different algae species. K-means, linear-, quadratic discriminant analysis, and k nearest neighbour classification have resulted adequate separation performance for reliable object recognition.

IX. FUTURE WORK

In the future, I would like to improve the segmentation algorithm, since the biological diversity of living organisms is very high. Develop an algorithm that is capable of separating touching algae. I also would like to enhance the segmentation algorithm by exploiting auto fluoresce of chlorophyll to detect only the living algae in drinking water. To increase the robustness of the classification system the number of samples per categories should be increased to assure statistical significance.

REFERENCES

- [1] M. Steinbach, G. Karypis, and V. Kumar, "A Comparison of Document Clustering Techniques," 2000.
- [2] B.E. Boser, I.M. Guyon, and V.N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers," *PROCEEDINGS OF THE 5TH ANNUAL ACM WORKSHOP ON COMPUTATIONAL LEARNING THEORY*, 1992, pp. 144–152.
- [3] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *DATA MINING AND KNOWLEDGE DISCOVERY*, vol. 2, 1998, pp. 121–167.
- [4] A.J. Smola, B. Schölkopf, and B.S. Gmd, "A Tutorial on Support Vector Regression," 1998.
- [5] Melia, M.L. Zhang, M.W. Edu, T. Zhang, T. Zhang, R. Ramakrishnan, R. Ramakrishnan, and M. Livny, "An Experimental Comparison of Several Clustering and Initialization Mehtods," *DATA MINING AND KNOWLEDGE DISCOVERY*, vol.1, 1998, pp.141-182.
- [6] Melia, M.L. Zhang, M.W. Edu, T. Zhang, T. Zhang, R. Ramakrishnan, R. Ramakrishnan, and M. Livny, "An Experimental Comparison of Several Clustering and Initialization Mehtods," *DATA MINING AND KNOWLEDGE DISCOVERY*, vol.1, 1998, pp. 141-182.
- [7] L. Bottou and Y. Bengio, "Convergence Properties of the K-Means Algorithms," *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 7*, vol. 7, 1995, pp. 585–592.
- [8] M.H.F. Wilkinson and J.B.T.M. Roerdink, "Diatom Contour Analysis using Morphological Curvature Scale Spaces," *IN: PROC. 15TH INTERN. CONF. ON PATTERN RECOGNITION (ICPR'2000, 2000*, pp. 3–7.
- [9] H.D. Buf, M. Bayer, S. Droop, R. Botanic, G. Edinburgh, S. Fischer, R. Head, and S. Juggins, "Diatom Identification: a Double Challenge Called ADIAC," 1999.
- [10] E. Osuna, R. Freund, and F. Girosi, *Improved training algorithm for support vector machines*, 1997.
- [11] T. Luo, K. Kramer, D. Goldgof, L.O. Hall, and S. Samson, "Learning to Recognize Plankton," *IN PROC. IEEE INT. CONF. SYSTEMS*, 2003, pp. 888–893.
- [12] A.C. Jalba, M.H. Wilkinson, and J.B. Roerdink, *Morphological Hat-Transform Scale Spaces and Their Use in Pattern Classification*.

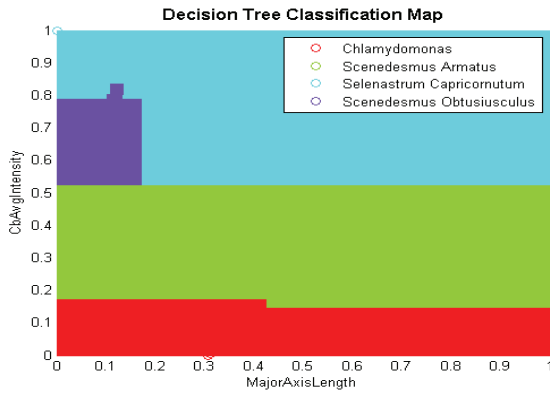


Fig. 3. Classification Map of Decision Tree. Coordinates are normalized to [0 1] range.

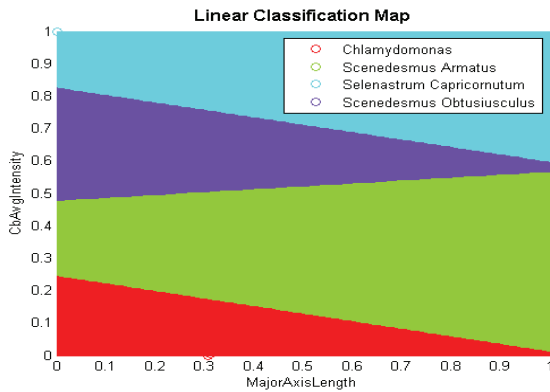


Fig. 4. Classification Map of Linear Discriminant Analysis. Coordinates are normalized to [0 1] range.

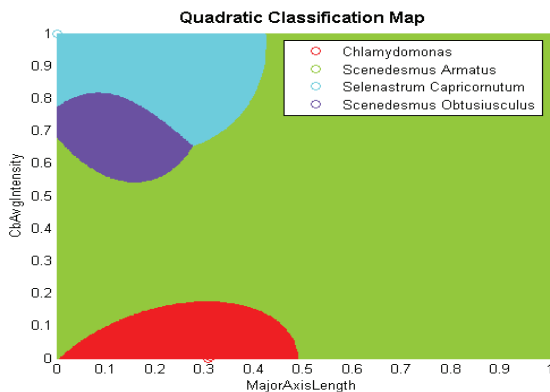


Fig. 5. Classification Map of Quadratic Discriminant Analysis. Coordinates are normalized to [0 1] range.

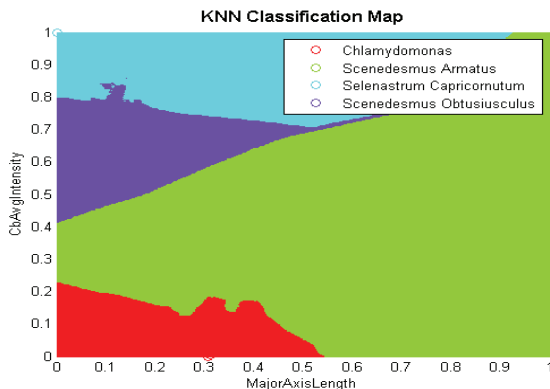


Fig. 6. Classification Map of K-Nearest Neighbour Classification (K = 5). Coordinates are normalized to [0 1] range.

A First-principle Computational Model for Electronic Structure of Molecular or Atomic Media

Ádám Fekete

(Supervisor: Dr. Árpád I. Csurgay)
fekad@digitus.itk.ppke.hu

Abstract — We started developing a simulator with the usage of finite difference method that is capable of modeling molecular constructions and processes.

I. INTRODUCTION

In the latter decade the molecular simulators in use performed weaker, or too inaccurate or too slow as expected. By the application of heuristics we may decrease the time of simulation but on the other hand we get suitable accuracy in special cases only. We recede from the real physical meaning increasingly due to the beauty of the formalism.

Current models (Hartree-Fock, Density Functional Theory (DFT), Car-Parrinello Molecular Dynamics) applied on the new computer architectures mean mathematical and programming difficulty primarily. Their development is going on continually. The DFT is capable to simulate bigger systems so this is the most preferred method. Its principal disadvantage, that it approaches to the electron exchange correlation function heuristically (magic function), which causes a considerable mistake already in the definition of the excited states. In case of the dynamic systems when the definition of states is not stationary then it cannot be applied.

Nowadays the sizes of the transistors are so small, that they consist of countable number of atoms and a multitude of electrons moves on the wires instead of a current. The new computer architectures insure the simulation of increasingly bigger and bigger systems. The opportunity is created at the same time for the new and more detailed modeling, with which we may plan new devices (it can be a nanosensor, a transistor or even a drug molecule). All this not only quantitative, but reports a qualitative improvement.

II. BASIC FRAMEWORK

My aim is to make a simulator that is able to imitate a realistic molecular environment and to plan the function of sensors. The molecular processes are redundant in time and space in the nature.

In the simulation similarly to the reality the decrease of the energy of a system only happen through the radiation of a photon, its increase happen by absorbance.

There is a need for numerical methods since the problem cannot be solved analytically in a general case. We may apply finite differences and numerical difference equations for the given models which already worked well with electromagnetic field simulations for example.

As opposed to the models until now, the state of the system will not be represented in the linear combination of base functions, but all the elements of the space will change according to the local relations and models. The task can be made parallel easily that is in accordance with the computer developmental guidelines (many-core architectures) in the immediate future.

A. Modeling of larger systems

In case of bigger systems the ab initio modeling is a hopeless challenge that is why we need hierarchies of models. One of the most important tasks is to link the quantum and the classic world.

Characteristic example is the double split experiment with a few electron or buckyball or the interaction of the classic electromagnetic space and the molecules. The framework system opens the door for us to use models with a different level inside the space structure.

We have the opportunity to simulate interfaces of metals and molecules when we measure the Current-Voltage characteristic of molecules with significantly bigger electrodes [5]. In this case we apply a more detailed model onto a region of interest (on the figure the central one).

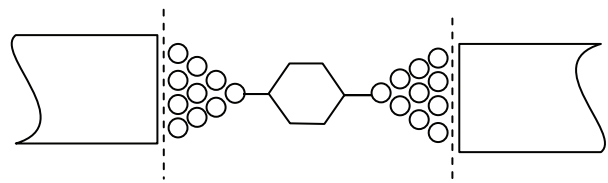


Fig. 1. Schematic model of measure a single molecule.

B. Main properties of a molecular system

First task in modeling a process is to determine the ground state. This is the stable state with the smallest energy which we calculate with the full neglect of the environment.

The next step is the interaction of the molecule and the environment that is a photon or classic electromagnetic space. The photons with various wavelengths affect different parts of the molecules that can be seen in 1. table in details.

TABLE I
ELECTROMAGNETIC WAVES

Wavelength (m)	Type of radiation	Site of interaction
10^{-10}	gamma-ray	nucleus
$10^{-9} - 10^{-7}$	X-ray	nucleus, inner orbitals
10^{-7}	ultra violet	inner orbitals, valence electrons
10^{-6}	visible light	valence electrons
$10^{-6} - 10^{-5}$	Infra red	molecules, chemical bonds
$10^{-5} - 10^{-4}$	microwave	molecules, rotation of molecules
$10^{-2} - 10^2$	radio wave	nuclear spin

This defines the measurable features that take place in the dynamics of time and space. The undermentioned cases are with emphasized significance:

- excited states
- molecular vibrations
- electron transport

C. Resolution of time and space

The reciprocal space would reduce the claim of the resource of a simulator significantly, but the simulation of the nanoelectronic devices needs spatial modeling, since the geometry of the devices defines their behaviour. [4]

The processes take place in virtual space and time. The time passes equally in all voxel and records discrete values that are defined by spatial resolution. Initially the space will be divided up uniformly, but with the help of multi-scale modeling better accuracy and faster convergence can be reached. It is possible to reach optional accuracy with the change of the spatial resolution.

D. Natural laws

Initially we implement behaving according to Schrodinger equations only that is right between certain conditions. Next, we will construct a numerical model of Maxwell and Lorentz equations [1]:

$$\begin{aligned}\nabla \cdot \mathbf{E}(\vec{r}, t) &= \frac{1}{\epsilon_0} \rho(\vec{r}, t) \\ \nabla \cdot \mathbf{B}(\vec{r}, t) &= 0 \\ \nabla \times \mathbf{E}(\vec{r}, t) &= -\frac{\partial}{\partial t} \mathbf{B}(\vec{r}, t) \\ \nabla \times \mathbf{B}(\vec{r}, t) &= \frac{1}{c^2} \frac{\partial}{\partial t} \mathbf{E}(\vec{r}, t) + \frac{1}{\epsilon_0 c^2} \mathbf{j}(\vec{r}, t)\end{aligned}$$

$$m_\alpha \frac{d^2}{dt^2} r_\alpha(t) = q_\alpha [E(r_\alpha(t), t) + v_\alpha(t) \times B(r_\alpha(t), t)]$$

$$H = \sum_\alpha \frac{1}{2} m_\alpha v_\alpha^2(t) + \frac{\epsilon_0}{2} \int d^3r [E^2(\vec{r}, t) + B^2(\vec{r}, t)]$$

$$P = \sum_\alpha m_\alpha v_\alpha(t) + \epsilon_0 \int d^3r B(\vec{r}, t) \times B(\vec{r}, t)$$

$$J = \sum_\alpha r_\alpha(t) \times m_\alpha v_\alpha(t) + \epsilon_0 \int d^3r r \times [B(\vec{r}, t) \times B(\vec{r}, t)]$$

H is the total energy of the system, P is the total momentum and J the total angular momentum.

III. USING FINITE DIFFERENCE TO SOLVE THE SCHRODINGER EQUATION

The time-dependent Schrödinger equation for a particle having mass m in a potential $\Psi(\vec{r}, t)$ is:

$$j\hbar \frac{\partial \Psi(\vec{r}, t)}{\partial t} = \left[-\frac{\hbar^2}{2m_e} \nabla^2 + V(\vec{r}) \right] \Psi(\vec{r}, t)$$

The eigenfunction $\Psi(\vec{r}, t)$ and energy E is obtained by solving the time-dependent Schrodinger equations,

$$\hat{H}\Psi(\vec{r}, t) = E\Psi(\vec{r}, t)$$

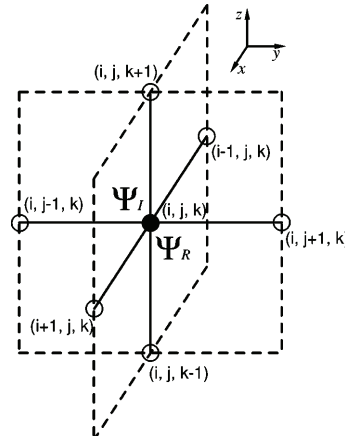
A. Ground state [2]

The solution of equation can be expanded in terms of eigenfunctions of the form

$$\Psi(\vec{r}, t) = \Psi(\vec{r}) e^{-\frac{iEt}{\hbar}}$$

By changing the real time to imaginary time, $\tau = it$, and using units with $\hbar = 1$ and $m = 1$ we then have

$$\frac{\partial}{\partial \tau} \Psi(\vec{r}, t) = \frac{1}{2} \nabla^2 \Psi(\vec{r}, t) - V(\vec{r}) \Psi(\vec{r}, t)$$



The time derivative is discretized by using the forward finite difference scheme given by

$$\frac{\partial}{\partial t} \Psi(x, y, z) \approx \frac{[\Psi^{n+1}(i, j, k) - \Psi^n(i, j, k)]}{\Delta \tau}$$

$$\Psi^n(i, j, k) = \Psi(i\Delta x, j\Delta y, k\Delta z, n\Delta t)$$

The second-order derivatives of space are discretized using centered differences.

$$\begin{aligned} & \frac{1}{2} \vec{\nabla}^2 \Psi(\vec{r}, t) - V(\vec{r})\Psi(\vec{r}, t) \\ & \approx \frac{1}{2\Delta x^2} [\Psi^n(i-1, j, k) - 2\Psi^n(i, j, k) + \Psi^n(i+1, j, k)] \\ & + \frac{1}{2\Delta y^2} [\Psi^n(i, j-1, k) - 2\Psi^n(i, j, k) + \Psi^n(i, j+1, k)] \\ & + \frac{1}{2\Delta z^2} [\Psi^n(i, j, k-1) - 2\Psi^n(i, j, k) + \Psi^n(i, j, k+1)] \\ & + \frac{1}{2} V(i, j, k) [\Psi^n(i, j, k) + \Psi^{n+1}(i, j, k)] \end{aligned}$$

Using previous equations becomes:

$$\begin{aligned} \Psi^{n+1}(i, j, k) = & \alpha \Psi^n(i, j, k) + \\ & \beta \frac{\Delta \tau}{2\Delta x^2} [\Psi^n(i-1, j, k) - 2\Psi^n(i, j, k) + \Psi^n(i+1, j, k)] \\ & + \frac{\Delta \tau}{2\Delta y^2} [\Psi^n(i, j-1, k) - 2\Psi^n(i, j, k) + \Psi^n(i, j+1, k)] \\ & + \frac{\Delta \tau}{2\Delta z^2} [\Psi^n(i, j, k-1) - 2\Psi^n(i, j, k) + \Psi^n(i, j, k+1)] \end{aligned}$$

where α and β are:

$$\alpha = \frac{[1 - \frac{\Delta \tau}{2} V(i, j, k)]}{[1 + \frac{\Delta \tau}{2} V(i, j, k)]}$$

$$\beta = \frac{1}{[1 + \frac{\Delta \tau}{2} V(i, j, k)]}$$

The condition of stability:

$$\Delta \tau \leq \frac{1}{\left[\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} + \frac{1}{\Delta z^2} \right]}$$

The potential field of a hydrogen atom is:

$$V(x, y, z) = -\frac{1}{\sqrt{x^2 + y^2 + z^2}}$$

An initial random wavefunction is used.

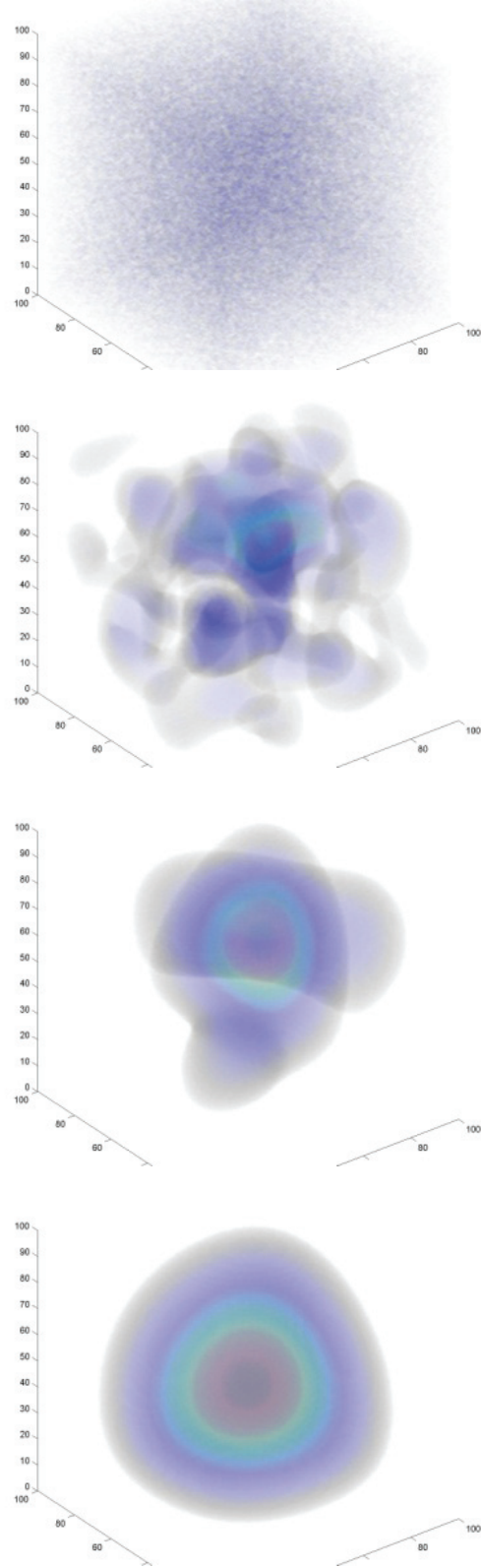


Fig. 2. The probability of an electron in Coulomb potential field that is started from random initial state. The snapshots represent four consecutive moments in time.

The complex wave function is separated into two real functions that correspond to its real and imaginary parts

$$\Psi(\vec{r}, t) = \Psi_R(\vec{r}, t) + i\Psi_I(\vec{r}, t)$$

Two equations involving real functions corresponding to Ψ_R and Ψ_I

$$\hbar \frac{\partial \Psi_R(\vec{r}, t)}{\partial t} = \left[-\frac{\hbar^2}{2m_e} \nabla^2 + V(\vec{r}) \right] \Psi_I(\vec{r}, t)$$

$$\hbar \frac{\partial \Psi_I(\vec{r}, t)}{\partial t} = \left[+\frac{\hbar^2}{2m_e} \nabla^2 - V(\vec{r}) \right] \Psi_R(\vec{r}, t)$$

The following equations are solved in an iterative process. Firstly calculate the imaginary part from the real part and use this new imaginary part to calculate the new real part.

$$\begin{aligned} & \Psi_R^{n+1}(i, j, k) \\ &= -\frac{\hbar \Delta t}{2m_e} \frac{\Psi_I^{n+\frac{1}{2}}(i-1, j, k) - 2\Psi_I^{n+\frac{1}{2}}(i, j, k) + \Psi_I^{n+\frac{1}{2}}(i+1, j, k)}{\Delta x^2} \\ & - \frac{\hbar \Delta t}{2m_e} \frac{\Psi_I^{n+\frac{1}{2}}(i, j-1, k) - 2\Psi_I^{n+\frac{1}{2}}(i, j, k) + \Psi_I^{n+\frac{1}{2}}(i, j+1, k)}{\Delta y^2} \\ & - \frac{\hbar \Delta t}{2m_e} \frac{\Psi_I^{n+\frac{1}{2}}(i, j, k-1) - 2\Psi_I^{n+\frac{1}{2}}(i, j, k) + \Psi_I^{n+\frac{1}{2}}(i, j, k+1)}{\Delta z^2} \\ & + \frac{\Delta t}{\hbar} V(i, j, k) \Psi_I^{n+\frac{1}{2}} + \Psi_R^n \end{aligned}$$

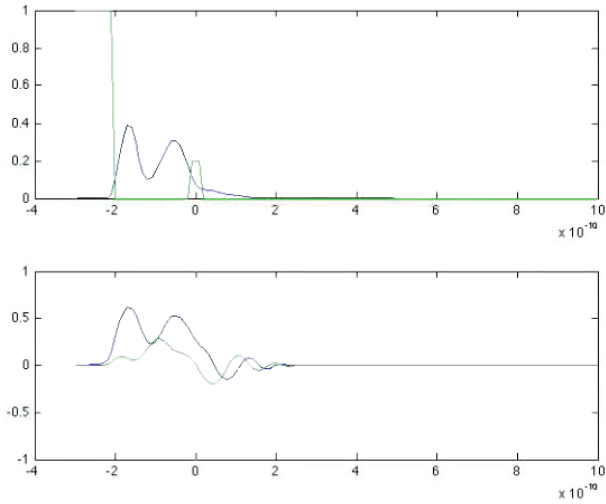


Fig. 3. Simulation of tunnel effect

- [1] Claude Cohen-Tannoudji, Photons and Atoms: Introduction to Quantum Electrodynamics, Wiley-VCH (February 1997), ch1-2
- [2] I Wayan Sudiarta et al, "Solving the Schrodinger equation using the finite difference time domain method", J. Phys. A: Math. Theor. 40 (2007) 1885-1896
- [3] Soriano et al, "Analysis of the finite difference time domain technique to solve the Schrodinger equation for quantum devices", J. Appl. Phys., Vol. 95, No. 12, 15 June 2004
- [4] Klimeck, G. et al. Atomistic simulation of realistically sized nanodevices using NEMO 3D: Part I — models and benchmarks. IEEE Trans. Electron Dev. 54, 2079-2089 (2007).
- [5] Time dependent transport phenomena, G. Stefanucci, S. Kurth, E.K.U. Gross and A. Rubio, Molecular and nano electronics: analysis, design and simulation, J. Seminario, ed(s), Elsevier Series on Theoretical and Computational Chemistry 17, p. 247-284 (2007).

Assessing Tissue Reaction around Silicon-based Multielectrode Arrays with Different Bio-coatings

László Grand

(Supervisors: Dr.George Karmos, Dr.Istvan Ulbert, Dr.Lucia Wittner)
grand@cogpsyphy.hu

Abstract— Biocompatibility and the impact of different coatings on four shank NeuroProbes silicon multielectrodes were investigated in vivo in rats. Effects of uncoated silicon (Si), hyaluronic acid (Hya), dextrane (Dex), dexamethasone (DexM) and Hya/DexM coatings on neuronal and glial densities were examined in the rat neocortex. Tissue reaction was explored with neuronal (NeuN) and glial (GFAP) staining at short (1 and 2 weeks) and long term (4, 8 and 12 weeks) survivals after implantation. Neuron counting was carried out by stereological methods around the probe tracks. Transmission electron microscopy (TEM) was used to verify tissue damage. Adhesive properties of the explanted probes with different coatings were assessed by scanning electron microscopy (SEM).

Neuronal loss and slight gliosis have been shown only within 100 μ m from the track. Neuron numbers were ~60% of the control after one week and ~90% after 8 weeks. Bleeding during implantation had a serious effect on cortical cell density. Comparing different coatings showed that cell numbers were the highest around DexM followed by Hya/DexM>Dex>Si>Hya coated probes, both at short and long-term. At Transmission Electron Microscopy level, healthy neurons and numerous synapses have been found in the vicinity of the track. Tissue residues were found on good adhesion molecule coated probes (Hya, Dex) at Scanning Electron Microscopy level.

Based on our results, tissue damage is the highest at short-term, and reduces over time. The different coatings have moderate impact on neural and glial densities. Probe insertion without bleeding highly increases cell and tissue preservation.

I. INTRODUCTION

1D, 2D and 3D brain multichannel microelectrode arrays are of great interest for neuroscience research and for the rehabilitation of sensory and/or motor functions in patients with neurological diseases. However, one important problem reported with all available microelectrodes to date, is long-term functionality and biocompatibility [1]. Indeed acute [3] and chronic inflammatory reactions [4] have been observed, which often result in the damage of the

surrounding tissue and the device itself compromising its functionality.

The challenge is to improve significantly the long-term biocompatibility of the implant by coating it in a proper way in order to suppress the immunological response, and optimize the neuron/probe interface [5]. We report the effects of different bio-coatings, ranging from a component of the extracellular matrix (Hyaluronic acid) to an anti-inflammatory substance, on the tissue response around the probes.

II. ELECTRODE IMPLANTATION

Biocompatibility of NeuroProbes silicon multielectrodes [2] was tested in vivo, in the neocortex of Wistar rats (n=13). All implanted electrode combs had four, 2mm long shanks, and were non-functional, since they had no output cables. Three different types of probes were implanted. 1) A3TB probes had shanks with a cross section of 100 μ m x 100 μ m in size, five platinum electrode contacts (20 μ m diameter, and 400 μ m distance between contacts) on each shank, and a thin probe base for platform-based assembly (thickness is 300 μ m, with four segments of 400 μ m x 300 μ m). 2) 3aWB probes had shanks of 100 μ m x 100 μ m in size, no electrode contacts on the shafts, and a wide probe base for cable assembly (thickness is 300 μ m, dimensions are 2440 μ m x 640 μ m). 3) 2WB probes had shanks of 150 μ m x 100 μ m in size, no electrode contacts, and the wide probe base for cable assembly.

The effect of four different coatings was tested. Five probes – of the same type – were implanted in each rat, one without coating (Si), and four with the following coatings: hyaluronic acid (Hya), dextrane (Dex), dexamethasone (DexM) and a mixture of hyaluronic acid and dexamethasone (Hya/DexM). Two probes (Si, Hya) were implanted in one of the hemispheres, and three (Dex, DexM, Hya/DexM) into the other, in the dorsal neocortical areas between the range of bregma and lambda (Paxinos rat atlas). The surgery was performed in ketamine/xylazine anesthesia (ketamine: 75mg/kg, xylazine 5mg/kg). Independent craniotomies were made for the 5 probes in each rat. The dura was perforated with the sharp tips of the probes. After implantation, the craniotomies were filled with dental acrylic.

L. Grand is with the Faculty of Information Technology, Péter Pázmány Catholic University, Budapest, Hungary (phone: +36-1-886-4700; fax: +36-1-886-4725; e-mail: grand@cogpsyphy.hu).

III. CELL COUNTING

In each animal, every 6th section was stained with the neuronal marker NeuN, and was used to estimate cell loss around the probe tracks, within the entire depth of the neocortex. The sections containing probe tracks were examined by light microscope, and digitized with high resolution at 10x magnification. These images were analysed with a manual method, as it follows. A grid of 100 μ m x 100 μ m was placed over the probe tracks, and neurons were counted in squares at a distance of 100, 200, 300 and 400 μ m from the side of the tracks. For each distance and coating, cells were counted in 40 to 66 squares. Since neuronal density is very variable in the different layers of the neocortex, control areas were determined in each animal, for each different coating. In all cases, two times 34 neighbouring squares (n=68), covering the entire thickness of the cortex were chosen as control, located at least 600 μ m from the probe tracks. Following the rules of stereological methods, cell bodies touching the top and left side of the grid were counted, but somata touching the bottom and right sides were not included. Cell loss within a given distance, at a given coating was determined as percentage of its control.

IV. EFFECT OF BLEEDING DURING PROBE INSERTION

During probe insertion procedure, there is a certain chance to hit superficial blood vessels causing minor or more serious bleeding. We examined the impact of bleeding on neuronal and glial cell densities at short and long term survivals. Both at one week and twelve weeks after surgery the signs of serious bleedings were visible. The tissue was damaged around the probe tracks, very small numbers of neurons or glial cells – if any – could be observed. Large holes and a dark unspecific staining characterized the cortical regions affected by the serious bleedings, masking any stained neural or glial cell (Fig. 1A, B, C). In some cases, but not always, patches with severe neuron loss could be observed next to the damaged tissue (Fig. 1C). Larger and darker glial cells were present in the vicinity of the damaged region, compared to control. In other cases, the severe neuronal loss was not accompanied by tissue damage and the dark unspecific staining (Fig. 2D).

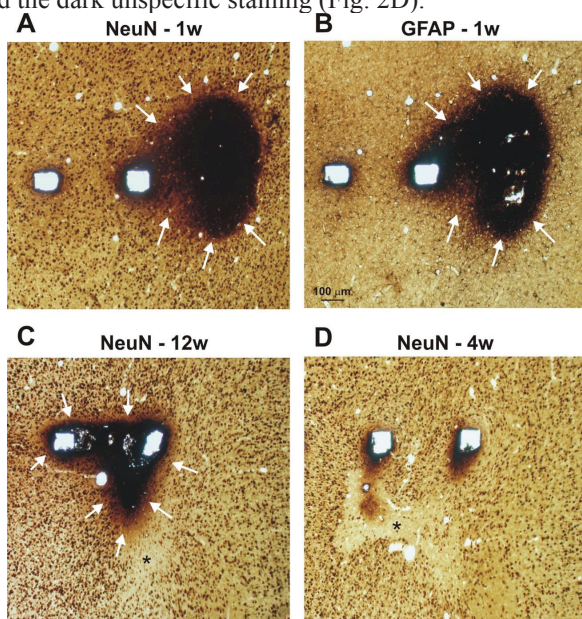


Fig 1. Effect of bleeding on neuronal (NeuN) and glial (GFAP) cells.

V. EFFECT OF DIFFERENT COATINGS ON NEURON SURVIVAL

Only probe tracks were included in this study, where no bleeding occurred during implantation. Control squares were defined in a line crossing the entire width of the cortex, at least 600 μ m from the probe track. In all cases, cell number is given as percentage of control cell numbers.

At one week survival, we examined neuron numbers relative to the different coatings in a rat implanted with 2WB type electrodes (n=21 950 cells in total). We found a considerable loss of neurons in the vicinity of all the tracks (within 100 μ m). The proportion of the surviving neurons varied from 49.7% to 76.0%, compared to control, with the lowest proportions around the Hya coated, and with the highest around the DexM coated probe. (Fig. 2A, 2B). At distances of 200, 300 and 400 μ m, neuron densities were higher, reaching 87.1% to 102.0% at 400 μ m from the track. The efficiency of the coatings to preserve as many neurons as possible – based on these percentages – was the following: Hya<Dex<Si<Hya/DexM<DexM (Fig. 2D).

At long term survival (8 weeks) we examined neuronal densities in a rat implanted with 3aWB probes (n=37 805 neurons). We found a minimal neuron loss from 100 to 400 μ m from the track. Neuron percentages varied from 87.9% to 98.4% at 100 μ m, while it was between 93.8% and 100.0% at 400 μ m (Fig. 2C). Although neuron loss is very little at long term, and the implanted probe is different in size (2WB vs. 3aWB), the efficiency of the different coatings was similar to that found at short term survival: Hya<Dex<Si<Hya/DexM<DexM (Fig. 2D).

At short term survival, we found a considerable neuron loss around 2WB probes, whereas neuron loss was minimal at long term survival, around 3aWB probes. To reveal the impact of time and electrode type on cell loss, we examined neuronal density around the same probe type at short and long terms as well (Fig. 2E). We counted neurons around A3TB uncoated Si probes at one, two, four and twelve weeks survival (n=38 773 cells). We found similar tendencies to the above described studies: lower neuron percentages at short term, and higher percentages at long term survivals. At all time points, notable neuron loss could be demonstrated only in the vicinity of the probe track (<100 μ m). The lowest neuron numbers were observed at one week survival (77.8%) and from the second week neuronal survival was about 85% of the control. At 200 to 400 μ m distances neuron percentages tended to reach the control values, they were higher than about 90%.

In summary, neuron loss is the highest at one week survival after probe implantation, at distances less than 100 μ m. At long term survival, the considerable neuron loss observed at short term survival decreases, and neuronal densities are close to control values. The impact of different coatings on cell numbers can be considered at one week survival, and becomes less significant at long term survivals.

VI. GLIAL REACTION AROUND THE PROBES

Glial reaction takes place in the brain tissue after insertion of foreign bodies [3,4]. This consists of changes in the quantity of glial cells, as well as in the length and number of their processes. Gliosis around the probe tracks was investigated in astroglial marker glial fibrillar acidic protein (GFAP)-stained sections. We made qualitative analysis, and

did not perform cell counting which would not show the increase in length and number of glial processes.

At short term survivals, large and strongly stained astroglial cells were visible around the probe tracks. At long term, glial cells were similar to the control, even in the vicinity of the track. Glial reaction was found to be relatively low, both at short term and long term survivals. By light microscopic examination no considerable differences could be found around the differently coated probe tracks. Higher and lower glial reactions could be observed at the different shanks of the same probe, sometimes related to a bleeding. When a serious bleeding occurred during probe insertion, and the tissue was considerably damaged, large and dark glial cells surrounded the injured area.

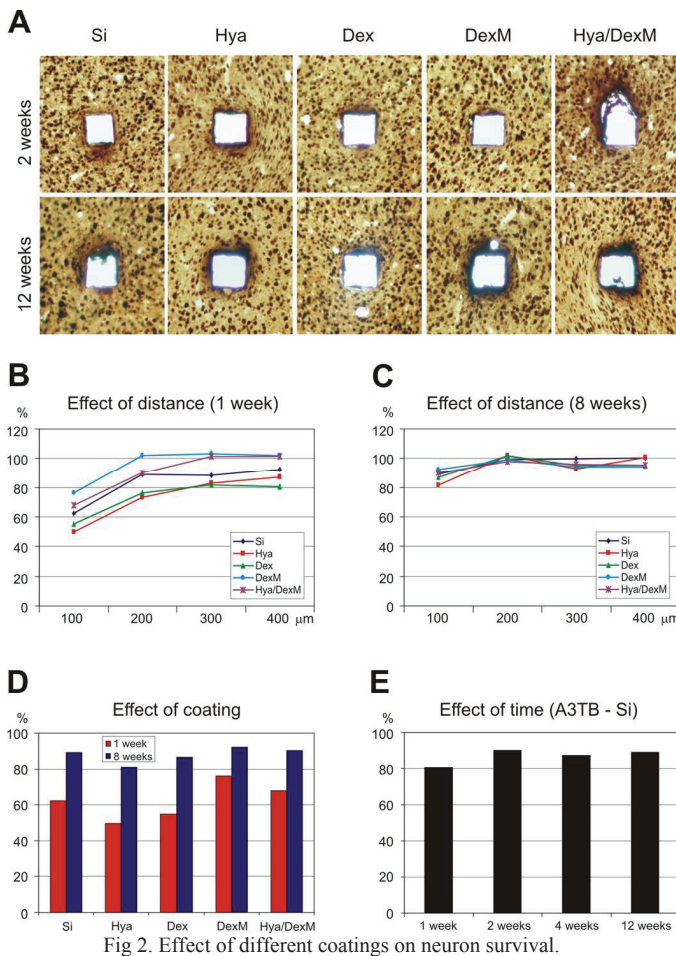


Fig 2. Effect of different coatings on neuron survival.

VII. TRANSMISSION ELECTRON MICROSCOPY OF THE TRACKS

Light microscopic examination showed a notable neuron loss in the vicinity of the probe track one week after probe implantation. This cell loss decreases with distance from the track, and with survival time of the animal. Glial reaction was found to be relatively low if bleeding did not occur during probe insertion. We wished to verify the presence or absence of neuronal and glial elements around the tracks at electron microscopic level. We investigated tracks of A3TB Hya coated probes in animals of one, two, four and twelve weeks survival in GFAP-stained sections. In the animal with one week survival, one shank of the probe hit a blood vessel making a serious bleeding, whereas no signs of bleeding

were observed around the other shanks. In the twelve weeks survival animal, high glial reaction was found around one shank and low around the others. This allowed us to examine tissue preservation around the same probe type with the same coating, in case of bleeding, as well as at high and low glial reaction.

VIII. NEURONAL CELL BODIES – TISSUE PRESERVATION

The presence of healthy neuronal cell bodies is essential to achieve a reasonable recording with any implanted electrodes. Light microscopy showed numerous NeuN-stained neurons around the tracks of the silicon probes. In all examined cases, healthy looking neuronal cell bodies were visible within 100μm from the track at electron microscopic level. Neuronal cell bodies were distinguished from glial somata based on the following criteria: the absence of the astroglial marker GFAP, light and homogeneous nucleus, electron-dense nucleolus (if visible), clearly visible cytoplasmic structure and organelles. Astroglial cells show positivity for GFAP, other glial cells, such as oligodendrocytes and microglia possess either very light cytoplasm with poor structure, or dark, electron dense homogeneous cytoplasm and nucleus. In addition, neuronal cell bodies are usually larger in size, 12-25μm, whereas glial cells are smaller, 5-10μm.

One week after probe implantation, brain tissue was considerably damaged around the tracks. Holes, degenerating structures and incomplete membranes could be observed around all shanks, including those showing no signs of tissue damage at light microscopic level. The tissue injury extended to about 30 to 40μm from the track at the shank with no bleeding. Signs of damage were observed in a considerably wider region around the shank with bleeding, extending to about 120-150μm from the border of the track. Glial cells and processes formed a layer of about 5-10μm at the border of all the tracks. They did not form a densely packed glial scar, but a mixture of glial and neuronal processes. Neuronal cell bodies were detected close to the track with no bleeding (~10μm distance), and farther at the track with bleeding (~50μm).

Two weeks after probe implantation, the tissue damage was observed around the tracks, but in a lesser extent than in the one week survival animal. Holes were present around the tracks, but only in patches. The number of incomplete membranes and degenerating structures was decreased. At four weeks of survival, tissue preservation was good, with no obvious signs of tissue injury. The glial margin was usually 5-10μm thick, but at one edge of one of the examined tracks glial cells were more numerous forming a densely packed scar of about 10-20μm. Healthy looking neuronal cell bodies were present all around the tracks, even at 10μm distance.

At 12 weeks survival, tissue preservation was similar to areas at least 300μm from the track: signs of damage could be hardly seen. Little structural difference was visible around the tracks with high and low glial reaction observed at light microscopic level. More GFAP-positive astroglial processes were dispersed in a 30-50μm thick region around the track with high glial reaction. This network of astroglial processes did not form a dense glial scar. Neuronal cell bodies were abundant around the tracks, sometimes in the very close vicinity (less than 10μm).

IX. SYNAPSES

One of the most important neuronal communications is based on synaptic connections between neurons. We examined the presence of asymmetric (presumably excitatory) and symmetric (presumably inhibitory) synapses around the tracks of the implanted silicon probes. In all cases – at 1, 2, 4 and 12 weeks survival, as well as in case of bleeding and high glial reaction – numerous synapses were present within 100 μ m of the tracks.

The presence of damaged synapses (mostly at one and two weeks of survival) was closely related to the extension of the tissue damage. Usually the outer membranes of pre- and postsynaptic elements were discontinuous, whereas the synaptic cleft and the group of vesicles seemed to be intact. Mostly asymmetrical (excitatory), and very few symmetrical (inhibitory) synapses could be observed around the tracks where tissue damage was considerable. In the brain areas, where the tissue is well preserved, numerous asymmetrical and symmetrical synapses were found, with intact ultrastructure. Many synapses of both types could be observed close to the track, less than 5 μ m distance in some cases.

X. SCANNING ELECTRON MICROSCOPY OF THE EXPLANTED PROBES

Brain tissue is a special environment, with intra- and extracellular compartments filled with different ionic solutions. In response to any injury, neural tissue shows a glial reaction. Light and transmission electron microscopy gave us valuable information about changes of the neural tissue caused by the probe implantation, but did not serve any data about modifications of the silicon probes provoked by the water based solutions and the cellular response of the brain. We examined the explanted probes at scanning electron microscopic level to see how the surface of the silicon was altered after spending short and long periods in the rat neocortex. A3TB probes were chosen for this study (1, 2, 4 and 12 weeks of survival, all different coatings), giving us the possibility to examine alterations of the platinum electrode contacts as well.

All examined explanted probes were partly covered by tissue residues and/or by a cell layer (Fig. 3A). We expected differences in the coverage, as a result of the different functions of our bio-coating molecules: more tissue residues/cells on the good neural adhesion surface Hya coated probe, than on the others. In contrast to this hypothesis, we could not find direct relation between the covered surface of the probes and the type of the coating.

Cells with different morphology could be observed on the surface of the explanted probes. Endothelium-like small, flat and round cells usually formed a densely packed, continuous layer (Fig. 3B). Large multipolar, amoeboid cells, resembling to fibrous astrocytes (Fig. 3C), as well as fusiform glial-like cells (Fig. 3D) could be also distinguished on the explanted probes. In some cases, red blood cells were also identified in the tissue residue.

The platinum electrode contacts were found to be covered by tissue residue at different levels, changing from total to no coverage (Fig. 3E, 3F). When not covered, small or larger extension damage of the silicon/platinum could be also observed (Fig. 3E).

ACKNOWLEDGMENT

The work was performed within the framework of the Information Society Technologies (IST) Integrated Project NeuroProbes of the 6th Framework Program (FP6) of the European Commission.

I thank the generous support and encouragement of my supervisors. My special thanks goes to Lucia Wittner for teaching me the NeuN and GFAP staining procedure as well as to Emmanuelle Göthelid for providing us coated probes.

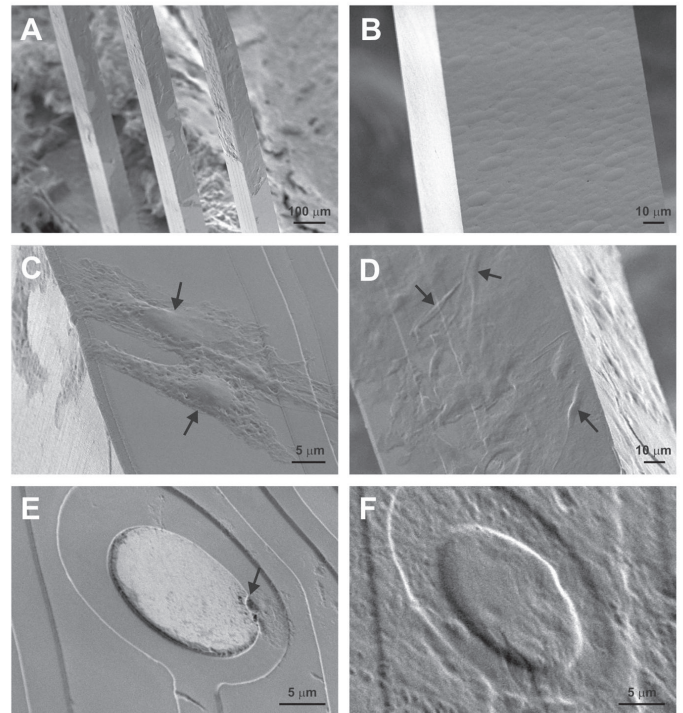


Fig 3. SEM pictures of explanted probes.

REFERENCES

- [1] R. Biran, D. C. Martin, P. A. Tresco, "Neuronal cell loss accompanies the brain tissue response to chronically implanted silicon microelectrode arrays", *Exp Neurol*, vol 195, pp. 115-126, 2005.
- [2] S. Kisban, S. Herwik, K. Seidl, B. Rubehn, A. Jezzini, M. A. Umiltá, L. Fogassi, T. Stieglitz, O. Paul, P. Ruther, "Microprobe array with low impedance electrodes and highly flexible polyimide cables for acute neural recording", *Proceedings of the 29th Annual International Conference of the IEEE EMBS*, Lyon, France, 2007.
- [3] D. M. Landis, "The early reactions of non-neural cells to brain injury", *Annu Rev Neurosci*, vol 17, pp. 133-51, 1994.
- [4] D. H. Szarowski, M. D. Andersen, S. Retterer, A. J. Spence, M. Isaacson, H. G. Craighead, J. N. Turner, W. Shain, "Brain responses to micro-machined silicon devices", *Brain Res*, vol 983, pp. 23-35, 2003.
- [5] Y. Zhong, R. Bellamkonda, "Dexamethasone-coated neural probes elicit attenuated inflammatory response and neural loss compared to uncoated neural probes", *Brain Res*, vol 1148, pp. 15-27, 2007.

Electrophysiological Recordings with Electronic Depth Controlled Intracortical Microprobe Arrays

Balázs Dombóvári

(Supervisors: Dr. György Karmos and Dr. István Ulbert)
dombaga@digitus.itk.ppke.hu

Abstract—This paper presents the first successful *in vivo* electrophysiological recording with electronic depth controlled cerebral microprobe and NeuroSelect software developed by NeuroProbes for managing the electronically switchable electrodes. These microprobes contain large number of electrodes (up to 500 per probe shaft) which enables electrodes to be appropriately selected with respect to specific neuron locations. The NeuroSelect software makes it possible to scan electronically within a group of electrodes and to select those electrodes with the best signal quality. Additionally, tracking of individual neurons is achieved by independently switching to neighboring electrodes in the course of an experiment to compensate for micromotion of the probe in brain tissue. This depth control constitutes a significant improvement for multielectrode probes, given that so far the only alternative has been the fine positioning by mechanical probe translation. However, in multielectrode configurations this positioning can only optimize the position of a single electrode at a time. Electrodes are selected in a manual or semi-automatic mode with NeuroSelect based on visualized signal quality. The used metric is the signal-to-noise ratio. Besides managing the communication with the hardware controller of the probe array, the software also controls the acquisition, processing, display and storage of the neural signals for further analysis. Furthermore, here we report about the first successful *in vivo* acute recordings with depth controlled CMOS probe, where we could prove the advantages of the possibility of electrode selection.

Index Terms—3D probe arrays; electronic depth control; neural recording; silicon microprobes

I. INTRODUCTION

NeuroProbes is a European Project aiming at developing a system platform for the scientific understanding of cerebral systems, and for the treatment of the associated diseases [6]. The aim of the proposed research is to develop a system platform that will allow an extremely wide series of innovative diagnostic and therapeutic measures for the treatment and for the scientific understanding of cerebral systems and associated diseases.

Neural recordings with high spatial resolution are required for a basic understanding of neural processes. This goal is currently achieved only by silicon based MEMS arrays realized as one-dimensional (1D), two-dimensional (2D) or three-dimensional (3D) electrode configurations. Additional electronic circuitry has been implemented for signal buffering, multiplexing, amplification, processing and

telemetry. So far, electronics have only been integrated in the larger connecting areas of the probes or on a separate chip attached to the backbone of the probe arrays as specific process technologies applied to realize these silicon-based probes are often incompatible with the integration of electronics on the probe shaft. In addition, space constraints, i.e. minimal line width and spacing of internal leads defined by lithography, limit the number of electrodes per shaft. Therefore, in order to increase the number of electrodes, the integration of electronics on the probe shaft itself is mandatory.

Aside from the large number of electrodes required for high density recordings, long-term recording is often inhibited by micromotions of the recording probe in neural tissue which can increase the distance between a recording electrode and the neuron of interest. Since close proximity between electrode and neuron is mandatory for the discrimination of single action potentials, these micromotions can hinder the quality of recorded signals. In contrast with single wire electrodes, manual adjustment of the probe position is not an option in case of multielectrode arrays since the position of only one electrode can be optimized each time. In addition, signal quality may degrade over time due to apoptosis, tissue drift, relaxation, inflammation and reactive gliosis, among other reasons. Hence, there is a need for (re)selecting the electrodes which are richest in information about the firing activities of the cells.

Recently, the new concept of electronic depth control with CMOS-based hardware has been presented. Restrictions of existing systems, namely the limited number of electrodes on microfabricated probes and the required mechanical position control to compensate for micromotions, are circumvented. The slender, silicon based probe shafts contain integrated CMOS circuitry which allows to minimize the number of connecting lines and at the same time enables the selection of a subset of recording sites from an unprecedented number of electrodes.

The task of finding high quality signals in neural recording typically depends on the operator's intuition and subjective assessment with the aid of oscilloscopes and loudspeakers. In case of multielectrode arrays with a large number of recording channels as proposed in, a manual selection is tedious and rather impracticable. By contrast, (semi-)automatic selection is required that aims to identify the best recording channels out of a set of electrodes. This selection is based upon a signal quality metric, i.e. the signal

to noise ratio (SNR). The NeuroSelect software presented in this paper has been developed in order to (i) control the innovative CMOS-based neural probes with electronic depth control, (ii) process the recorded neural signals, (iii) select appropriate electrodes with optimized signal quality based on the data processing, and (iv) display and (v) store the neural information

II. RECORDING SYSTEM OVERVIEW

A. CMOS-based neural probe array

As described in detail in [1], the CMOS-based neural probes developed within the NeuroProbes project [7] comprise slender probe shafts (Figure 1) with integrated circuitry making it possible to switch electronically between different electrodes. The neural probes are realized in a post-CMOS compatible process combining deep reactive ion etching with sputter deposition of electrode arrays having an electrode pitch of $40\ \mu\text{m}$ with an exposed electrode diameter of $20\ \mu\text{m}$. Besides the probes with a shaft-length of 4 mm introduced in [1], systems in a different CMOS technology with shaft lengths of 2 and 8 mm are currently fabricated. The probes are being implemented in 1D-, 2D-, and 3D-arrays.

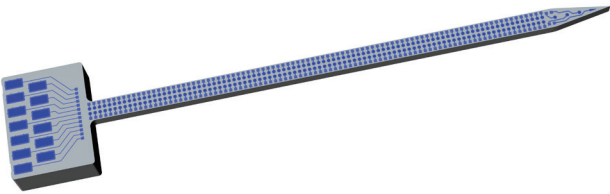


Figure 1: Schematic of the active probe shaft with a length of 4 mm comprising rows of electrodes with a pitch of $40\ \mu\text{m}$.

The integrated circuitry of the CMOS-based probes comprises a switch matrix to select simultaneously eight recording sites per shaft from a total of N ($N = \max 8102$) electrodes (in our in vivo experiment, we used the 4mm long single shaft probe on PCB with 188 electrodes, Figure 2). With the selection, each electrode can be switched to one of two possible lines out of the total eight analog output lines A0 to A7, as illustrated in Figure 3. The switching matrix itself contains a shift register formed by a chain of D-type flip-flops which allows the serial programming of the switches using two control lines (data input, DIN, and clock, CLK) in combination with two lines for power supply (VSS and VDD) [1]. Details on the integrated circuitry are given in [1].

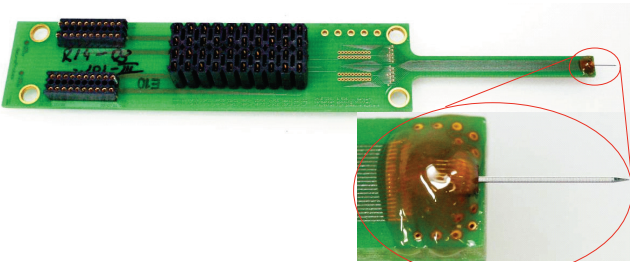


Figure 2: Assembled probe on PCB and close-up of probe bonded to PCB encapsulated by glob top.

As shown in Figure 3, the electrode selection code is sent from the host computer via a controller to the microprobe. Neural signals are acquired, displayed and saved. Adding signal processing functionality to rate the signal quality and to (re)select the best electrodes results in a closed-loop design of a neural acquisition system.

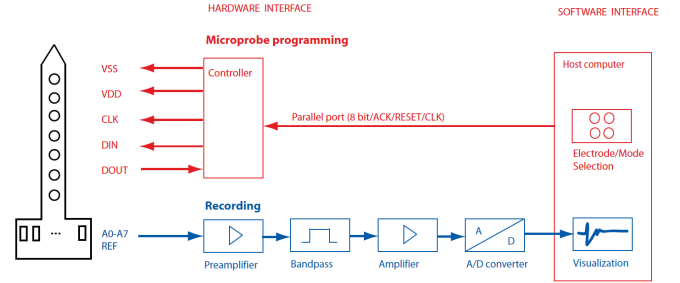


Figure 3: Electrode selection is transferred from the host computer via a controller to the microprobe. Neural signals are recorded and visualized. Based on the computed signal quality the electrodes are (re)selected. Selection is achieved via a shift register comprising flip-flops

III. DEPTH CONTROL SOFTWARE

A. Graphical User Interface

The NeuroSelect software provides a Graphical User Interface (GUI) that integrates the components for data acquisition (DAQ), signal processing and communication with the hardware controllers. As shown in Figure 4, the GUI is split into different windows that can be resized individually. The upper left pane of Figure 4A is used to control the data acquisition from the DAQ card PCIe 6259 from National Instruments as detailed later and to define the file name for the recorded signals. The left center pane is used to configure the plot settings, i.e. data scaling and selection of electrode signals to be displayed. The neural signals acquired from the DAQ card are visualized in the main window in the center. The bottom pane gives feedback about the current status of the software as well as the hardware. The right side window shows the control panel for electrode selection and settings. When maximizing the right pane (Figure 4B), one is able to select the electrodes of the different probe types in a manual or semi-automatic mode as described in the section on electrode selection.

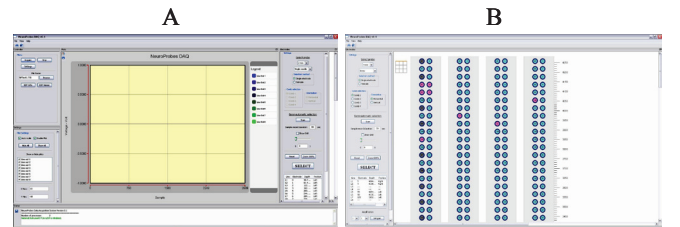


Figure 4: Graphical User Interface of NeuroSelect software: (A) GUI-overview of software for electronic depth control with sub-windows for electrode selection, control for data acquisition, data visualization, plot settings and status information. (B) Manual electrode selection: Electrodes can be selected by clicking with the mouse on the electrode. Selected, deselected and non-selectable electrodes are differently color-coded.

B. Data acquisition

As each CMOS-based neural probe shaft provides eight analog output channels, 32 and 128 signals are provided by the 4-comb and 4×4 platform arrays, respectively. These

signals are pre-amplified using a custom-made CMOS amplifier and fed to the data acquisition cards (PCIe 6259 DAQ, National Instruments) with 16-bit resolution, sampling rate of 31.25 kHz per channel and up to 32 analog inputs. Four of these cards are required to acquire all signals from a full 4×4 platform; one is sufficient for a single comb. The auxiliary digital inputs (available on a connector to the interface electronics) can be treated like an additional analog channel. Data are acquired in blocks of 4096 samples, which correspond to the block size in the stored file.

The software part for the data acquisition is included in a multithreaded way to avoid a bottleneck within NeuroSelect. Moreover, the processing of the GUI is slower and less important than the data acquisition process. There is a double buffer object that is used for visualizing the acquired data on the screen (Figure 4A, central window). Thanks to the applied multithreading, gapless data acquisition is guaranteed during the computations.

C. Electrode selection

Prior to any data acquisition, the user has the choice between different probe types. Probes with different shaft lengths (2 mm, 4 mm and 8 mm) and different probe configurations (single-shaft probes, probe combs with four probe shafts or 3D probe arrays comprising 4×4 probe shafts) are currently fabricated within the NeuroProbes project and can be selected within the NeuroSelect software. Furthermore, the gain factor for the CMOS based pre amplifier is set in this window of the GUI (Figure 4B). The probe type selection is followed by the electrode selection mode. The user is offered the choice between a manual electrode selection mode and a semi-automatic mode for electrode selection. In the manual mode, the user is not supported by data analysis and corresponding calculation and sorting of the quality metric as in the semi-automatic mode. In the future, this might be extended with a fully automatic mode which selects the electrodes with the best signal quality using pre-defined selection constraints, i.e. single electrode or tetrode configurations and spacing of the electrodes.

D. Programming environment

The software NeuroSelect has been written in C++ and uses the multiplatform framework wxWidgets distributed under free software license for GUI implementation [8]. It has been developed for Windows, but can be ported to Linux, MacOS or other platform environments. Visual Studio is used as integrated development environment (IDE) and compiler, but the Microsoft foundation classes (MFC) are not used. The design of the graphical user interface was developed using the DialogBlocks editor from Anthemion [9]. The signal analysis package is developed in C/C++ and uses the OpenMP library [11] in the parallel-processing version. This allows to use all available processors (and cores) of the computer to process multiple signals in parallel. The source code to control the data acquisition cards (PCIe 6259 DAQ, National Instruments) and to visualize the neural data has been generated using LabWindows/CVI from National Instruments. The development of the NeuroProbes DAQ software is controlled using the version control software Subversion [12].

IV. ACUTE IN VIVO RECORDINGS

A. Insertion procedure

In our experiment the goal was to test the stability of the probe from insertion to recording. We used a single shaft CMOS probe with a length of 4 mm and 188 electrodes. The probe had 7 output lines assembled on PCB and encapsulated with EPO-TEK (Figure 2). The 7 output lines were bonded to two PRECI-DIP connectors, which are compatible with our system for the first test recording.

The CMOS probe was implanted in the primary motor cortex (2 mm in lateral direction, aiming M1/M2, Figure 5) of a Long-Evans rat under ketamine/xylazine anesthesia (0.2 ml/100g). The probe assembled on the PCB was fixed to a manipulator which drove the probe deeper into the cortex. The probe was carefully inserted into the cortex through the dura. Dimpling was visible but no critical bleeding. Perhaps the brain was influenced by the insertion procedure, i.e. neurons were damaged especially in the upper region of the insertion.

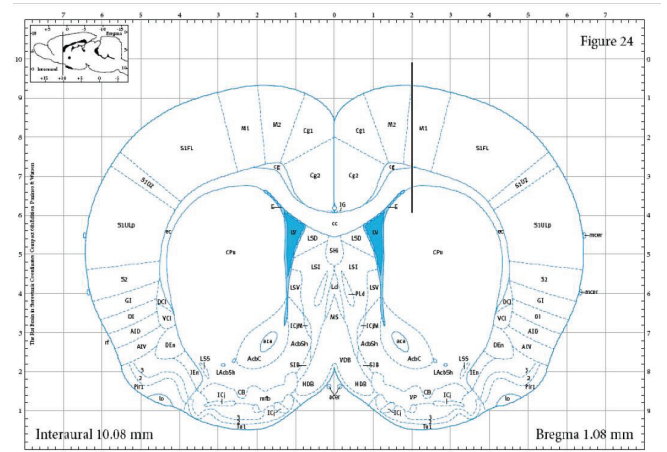


Figure 5: Cross section of the located area of implantation (based on [10]). The probe was inserted 2 mm in lateral direction aiming M1/M2 region indicated by the black line.

B. In vivo recordings

After insertion we recorded the neural activity at 20 kHz sampling frequency and with a total gain of 1000. The 5 electrodes were selected to output line 3, 4, 5, 6 and 7 (output line 3-7 were bonded to one PRECI-DIP connector; simultaneously we used only one of the two connectors). The whole probe was scanned by manual switching the electrodes. Local field potentials, multiple units and single unit activity of different units could be observed on different output lines.

62nd recorded block showed nice single unit activity (SUA) on line 5 (Figure 6) with a bursting unit at 500Hz and single firing unit with amplitude of about 100 μ V. Therefore, several single units could be recorded.



Figure 6: Sample recorded single unit activity on line 5 (Bandpass filtered multiunit activity between 0.5-5 kHz)

In the next step the electrode configuration was moved down by 80 μ m therefore switching the electrode from line

5 to line 7. The unit could be recorded from the same electrode but on different output lines. Switched back to the previous configuration the bursting unit is still present on line 4 and a firing unit on line 6. Units could be recorded on different electrodes.

Characterization of crosstalk is only possible by measuring signals on a blocked electrode. But, it seems not to be crosstalk since units could be observed on line 6 and 7 and other units on line 5 and 6. Single unit activity was present on one, two or three adjacent electrodes with decreasing spike amplitudes with the distance.

C. Impedance measurement

This section describes the impedance characterization of the platinum electrodes prior and during implantation. The impedances of the electrodes were measured before and during acute implantation in physiological saline solution (0.9% NaCl). A two electrode setup was used in which the working electrode was connected to the Pt electrode and the counter electrode was short-connected to the reference (stainless steel). The same measurement was also performed in vivo with the electrodes positioned in the cortex of the rat. The impedance was measured at 1 kHz.

The electrodes of line 7 showed an impedance of more than $2M\Omega$ which was higher than the electrode impedance of the other lines. Perhaps this is one reason for the higher noise level. Line 4, 5 and 6 had an impedance of about $0.5M\Omega$. The impedance in the off-state of electrode 4, 5, 6 and 7 was greater than $2M\Omega$. In general, line 3 did not perform well.

D. System calibration and noise measurement

Before insertion we tested the probe in saline solution to measure the noise level of the system. The recording inside the saline solution showed low noise level with white characteristics. Perhaps saline moves along the probe shaft covering the electrodes outside of the solution. For electrodes outside of the solution measured signals are caused by capacitive coupling. A higher noise level is measured for the electrodes outside from the saline solution compared to the electrodes inside the solution which is expected. Output line 3 showed a much higher noise level than the other lines.

V. DISCUSSION

The NeuroSelect software enables the experimenter to visualize the recorded signals, the spikes, as well as the calculated metrics like the SNR value per electrode and their relative ordering with respect to each other. This tool also manages the recording of the signals through the DAQ PCIe 6259 card from National Instruments, the storage of these signals into EDF files, execution of the SNR calculation algorithms, and steering the electronic circuitry to record from the electrodes as selected manually by the user or by the semi-automatic methods.

In an acute in vivo experiment we tested a 4mm long single shaft probe. We found that the insertion of the probe through the dura was trouble free with the manipulator. The noise level of the probe was low. We recorded some nice single unit activity, but later on we have to test the stability of probes in chronic experiments.

VI. ACKNOWLEDGEMENT

The work was performed within the framework of the Information Society Technologies (IST) Integrated Project NeuroProbes of the 6th Framework Program (FP6) of the European Commission. Furthermore I would like to express my gratitude to all those who gave me the possibility to complete this work.

REFERENCES

The depth control software part of this paper is an excerpt of reference [2].

- [1] Seidl K, Herwik S, Nurcahyo Y, Torfs T, Keller M, Schuettler M, Neves HP, Stieglitz T, Paul O and Ruther P. CMOS-based high-density silicon microprobe array for electronic depth control on neural recording. Proc Int MEMS Conf 2009, pp 232-235.
- [2] Seidl K, Torfs T, De Mazière P A, Van Dijck G, Csercsa R, Dombovari B, Nurcahyo Y, Ramirez H, Van Hoof C, M Van Hulle M, A Orban G, Paul O, Ulbert I, Neves H, Ruther P (2009). Control and data acquisition software for microprobe arrays with electronic depth control. Journal Biomedical Engineering, submitted.
- [3] Kisban S, Herwik S, Seidl K, Rubehn B, Umilta MA, Fogassi L, Stieglitz T, Paul O, Ruther P. Microprobe array with low impedance electrodes and highly flexible polyimide cables for acute neural recording. 29th Int IEEE EMBS Conf, Lyon, Aug 23- 27, 2007, pp 175-178.
- [4] Herwik S, Kisban S, Aarts AAA, Seidl K, Girardeau G, Benchenane K, Zugaro M, Wiener S, Neves H, Paul O, Ruther P. Fabrication technology for silicon based microprobe arrays used in acute and subchronic neural recording. Techn Dig 19th MicroMechanics Europe Workshop, Aachen, Sept 28-30, 2008, pp 57-60.
- [5] Aarts A, Neves HP, Ulbert I, Wittner L, Grand L, Fontes MBA, Herwik S, Kisban S, Paul O, Ruther P, Puers RP, Van Hoof C. A 3D slim-base probe array for in vivo recorded neuron activity. Proc 30th Annual Int Conf of the IEEE Eng in Medicine and Biology Society (EMBC), Aug 20-24, 2008, pp 5798-5801.
- [6] Neves HP, Torfs T, Yazicioglu RF, Aslam J, Aarts AAA, Merken P, Ruther P, Van Hoof C. The NeuroProbes project: A concept for electronic depth control. 30th Int IEEE EMBS Conf, Vancouver, Aug 20-24, 2008, p 1857.
- [7] Ruther P, Aarts A, Frey O, Herwik S, Kisban S, Seidl K, Spieth S, Schumacher A, Koudelka-Hep M, Paul O, Stieglitz T, Zengerle R, Neves H. The NeuroProbes project multifunctional probe arrays for neural recording and stimulation. Biomed Tech 2008; 53(1): 238-240.
- [8] Smart J, Hock K and Csomor S. Cross-platform GUI programming with wxWidgets. Prentice Hall PTR 2005.
- [9] Anthemion Software Ltd, Midlothian, Edinburgh, UK.
- [10] Paxinos G and Watson C The Rat Brain in Stereotaxic Coordinates, 6th Edition London, Academic Press, 2008.
- [11] Chapman B, Jost G and van der Pas R. Using OpenMP: Portable Shared Memory Parallel Programming. The MIT Press 2007.
- [12] Subversion open source version control system <http://subversion.tigris.org/>

Improvement of Tactile Sensor Measurements for Biologically Motivated Robot Control

Ferenc Lombai

(Supervisors: Tamás Roska and Gábor Szederkényi)

lomfe@digitus.itk.ppke.hu

Abstract—This paper presents a planned control architecture for a 7-degree-of-freedom (DOF) rigid robot arm using three-axial tactile-sensor arrays development at the TactoLogic Ltd. The goal of the system is to be able to stabilize a flexible pendulum mounted as the end effector of the robot arm. The work aiming is to uncover the possibilities to develop practical robotic control methods on the neural bases of tactile and visual modalities found in different species. The document introduces a Motor Map neural controller and outlines problems related to the measurements of the tactile sensor in dynamic environment. Measurements presented from the tactile arrays and simulations are performed to validate the neural controller.

I. INTRODUCTION

The main aspect of the recent work is to develop suitable control method that is capable to benefit from the rich information measured by the three-axial tactile-sensor arrays [1]. More precisely a 7 DOF robot arm is planned to stabilize the vibrating movement of an elastic rod mounted at the end of the manipulator.

This task can be interpreted as a control problem that attenuates the resonances caused by any disturbance and should have useful practical aspects for both serving robots and industrial material handling tasks. But it can be translated as a force follower that should be useful also for haptic interfaces or for other human-machine interactions.

Most of such problems are solved using six axial force/torque wrist sensors. Although these solutions are appropriate to measure the effect of the manipulated object or investigate the interaction between the end effector (EF) and the environment, but from biologically aspects the tactile sensing at the contact area much more fit to the natural solution one should used to. More over equipping a wrist force/torque sensor should significantly change the kinematic properties of the robot, thus altering it's work space and abilities for a particular task. There are several solutions using force/torque sensors.

In [2] an unknown flexible payload manipulation task was solved for a one-link machine. The controller worked in an adaptive manner utilizing wrist torque feedback from the sensor. Other one DOF example [3] was show that the damping of undesired oscillations is necessary while handling deformable linear objects. Even in the case of absolute measurements of forces and torques acting on the EF, further investigations made to compensate the effects of the robot self movements and prepare dynamic models for the wrist sensors [4] to speed up the interacting control loop.

II. SYSTEM DESCRIPTION

The robotic arm contains seven electrical servo motors mounted on each other in a chain like topology. The Dynamixel RX-28 type servos shipped by the Robotis Ltd. Figure 1 c.) shows the picture of the robot at The University of Catania during my stay there in collaboration with Professor Paolo Arena and his team. On figures 1 a.) and b.) the robot collision detecting model and the computer aided design (CAD) model can be seen, respectively.

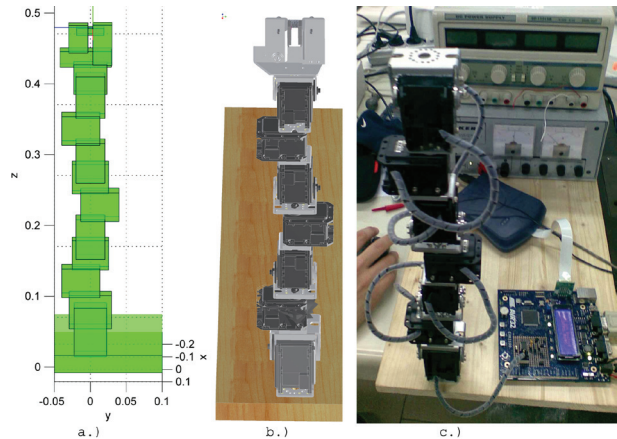


Figure 1. 7 DOF robot arm a.) collision detecting model, b.) CAD model, c.) real system.

Table I lists the modified Denavit-Hartenberg (MDH) parameters according to [5] and joint angle limits for the built topology. The joint limits are due to wiring along the robot.

Table I
MODIFIED DENAVIT-HARTENBERG NOTATION AND ROTATIONAL LIMITS

i	α_{i-1}	a_{i-1}	d_i	$q_{i_{min}}$	$q_{i_{max}}$
1	90°	0	0	-90°	90°
2	-90°	0	96 mm	-128°	172°
3	-90°	0	0	-90°	90°
4	90°	0	107.5 mm	-171°	129°
5	90°	0	0	-90°	90°
6	-90°	0	217 mm	-126°	174°
7	-90°	0	217 mm	-90°	90°

The inertia properties of the links were computed based on CAD, assuming uniform density distribution inside the motors and neglecting wires. Tables II and III list the main dynamic parameters for the 7 DOF robot.

Table II
LINK MASS(kg) AND CENTER OF MASS (cm)

i	mass	cg_{x_i}	cg_{y_i}	cg_{z_i}
1	107.64	0	40.166	7.076
2	15.12	0	2.046	-13.757
3	163.54	0.053	-36.5	5.223
4	13.58	0.087	-2.278	-13.897
5	163.79	-0.057	36.506	5.224
6	83.66	0.105	-0.24	-13.901
7	53.41	-4.85	-44.846	-1.055

Table III
LINK INERTIA PROPERTIES (cm^2kg)

i	I_{xx_i}	I_{xy_i}	I_{xz_i}	I_{yy_i}	I_{yz_i}	I_{zz_i}
1	65394.8	0	0	36352.2	-10435.5	46295.1
2	6226.9	0	0	2651.1	-425.4	5196.9
3	145042.6	25.6	45.7	39390.6	19768	127978.8
4	5576.6	23.1	-16.4	2512.8	404	4634.5
5	145047.2	27.9	-47.4	39403.5	-19768.7	127995.5
6	29423	-2.1	223.3	24217.1	-45.3	15648.6
7	48129.4	-2458.9	148.3	18392.9	1514.1	36455.9

The available tactile sensors are two four-by-four tactile arrays with elastic cover that forms half spheres over the taxels. The mechanical properties of the coverage strengthen the sensor response as described in [1].

III. DESCRIPTION OF THE CONTROL ARCHITECTURE

The bioinspired aspect of the control method lies in the so called motor-map neural controller. It proved to be a robust input-output mapping for unrevealed nonlinear tasks due to its self organizing capabilities. The structure is capable to learn high level control problems in an unsupervised manner [6], [7]. This net is an extension of the well known Kohonen network [8] and performs topology-preserving mappings between sensory inputs and motor functions. The flow chart of the planned control architecture is shown on figure 2.

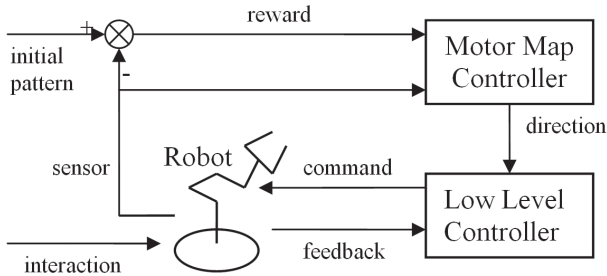


Figure 2. Planned control scheme.

At the beginning of the task the flexible stick is gripped by the robot and the sensors measure this initial pattern as further reference to reproduce. Based on the difference from the given desired state of sensor data a reward function teaches the Motor Map Controller (MMC) to produce appropriate directional output for the Low Level Controller (LLC) that minimizes the effect caused by any unknown environmental interaction. The MMC should adopt its output to achieve the best performance with respect to the reward function and

receives also the measurements directly from the sensors. The LLC uses faster feedback loop to measure the position and velocity of the Robot and command safety actions if necessary.

IV. DESCRIPTION OF THE CONTROLLER

The motor map controller performs mapping from the input layer $\mathbf{W}^{in} \in \mathbf{R}^{m \times n}$ to the output layer $\mathbf{W}^{out} \in \mathbf{R}^{k \times n}$, where n is the number of neurons, m is the input dimension and k is the output dimension of the system. For every input vector \mathbf{v} the neuron r is selected which weight vector \mathbf{w}_r^{in} best matches it. Then the network output is the winner neurons weight vector \mathbf{w}_r^{out} .

This mapping basically corresponds to a lookup table. The usefulness of this system is the topology preserving teaching mechanism, with attenuated interaction ratios over the neighboring cells that produces a smooth control surface that span the available storage space. Other useful aspect using such neural structure is its adaptation rate over time, when the controlled object or sensor mechanism changes due to environmental effects such that the temperature.

Control action during teaching performs a random search $\mathbf{w}_r^{out} = \mathbf{w}_r^{out} + a_r \eta$, where a_r is the winning neurons random search rate and η is a Gaussian random variable with zero mean and unit variance. An other learning related variable is b_r , that is the mean increase in the reward function $Re w(\omega) = -\|\omega^{des} - \omega\|$, where ω is a system state related variable in connection with the desired control state ω^{des} .

A detailed description of the teaching algorithm can be read in [6] but a short summary is made below:

- 1) store control action (\mathbf{v}, \mathbf{u})
 - store input \mathbf{v}
 - find winning neuron r as $\min_{r \in n} \|\mathbf{w}_r^{in} - \mathbf{v}\|$
 - store and perform control action $\mathbf{u} = \mathbf{w}_r^{out} + a_r \eta$
- 2) wait for control action to effect the system
- 3) perform adaptation in reward rate:
$$b_r^{new} = b_r^{old} + \gamma(\Delta Re w - b_r^{old})$$
- 4) perform adaptation in random search rate:
$$a_s^{new} = a_s^{old} - \epsilon'' h''_{(r,s)} a_s^{old}$$
- 5) perform learning if $\Delta Re w > b_r$
 - teach input weights:
$$\Delta \mathbf{w}_s^{in} = \epsilon h_{(r,s)} (\mathbf{v} - \mathbf{w}_s^{in})$$
 - teach output weights:
$$\Delta \mathbf{w}_s^{out} = \epsilon' h'_{(r,s)} (\mathbf{v} - \mathbf{w}_s^{in})$$

where ϵ , ϵ' and ϵ'' are learning step width parameters while $h_{(r,s)}$, $h'_{(r,s)}$, $h''_{(r,s)}$ are related to the shape and time evolution of the interaction between neighboring neurons.

Drawback of the system is the random output thus some kind of heuristic method should be applied for the particular problem or the weight initialization must be done in a simulation environment.

An open question is the generalization of the proposed method for higher dimensional reward functions or investigate the preferred topology of the neurons. Later is an exciting question mostly in case when it is necessary to provide the derivatives

of the sensor signals to the controller, thus the input vector size would lead to a 64 element one (2 forthcoming measurement, 2 arrays, 4 taxels and 4 voltages at each taxel). Finally a suitable reward function must be found.

V. THE IMPLEMENTED MOTOR MAP CONTROLLER SIMULATION

On figure 3 the output of the motor map controller is shown by the red line, interacting an inverted pendulum with force outputs in every 0.3 sec. The blue line is the angle of the balanced rod measured clockwise from the vertical direction. The green line is the angle of the same rod controlled by a known good stabilizing reference signal.

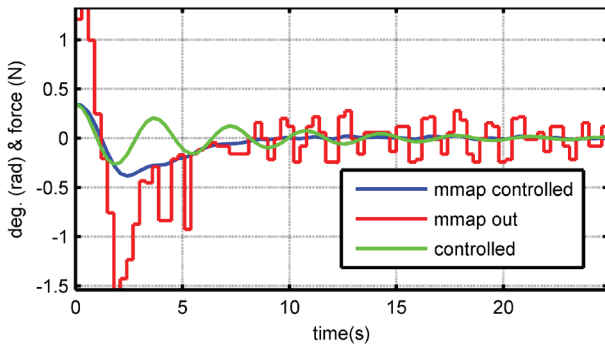


Figure 3. MMC balancing a planar inverted pendulum

VI. UPGRADE PLANS FOR TACTILE SENSING

The most conspicuous problem is that the reference zero position considerably changes due to the displacement between the contact surfaces. Measurements showed that the difference between two samples corresponding to unexcited elastic rods could lead to a same magnitude signal as the measurements for excited rod states. Such a phenomenon can be seen on figure 4. This fact is mostly due to pure mechanical aspects of the contact behavior, thus some kind of simulation is necessary to implement the sensors virtually. Further more a suitable model (probably a linear time invariant (LTI) system) inversion could let the incoming time series of the measured data to be translated as the time series of pressure and displacement profiles over the active sensing surface.

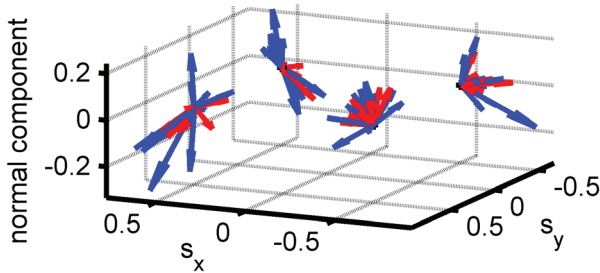


Figure 4. Measurements from 4x4 tactile array holding an elastic rod. Blue and red arrows corresponds to deformed and unexcited rod states, respectively. s_x and s_y are the shear force components.

The finite element simulations done on the elastic cover in [1] are stabilized strain distributions assuming infinite friction between the contact surfaces. The dynamic simulation of the sensors must be validated through intensive dynamic measurements of the given sensors in different well defined environments. Identifying the frequency attenuation profile of the cover could simplify the usage of these sensors in most dynamic applications [9]. The test environments should include measurable loads with point-, line- and plane contact surfaces and surface displacement and rotational observations for constant planar pressure with known friction parameters.

VII. CONCLUSIONS AND FUTURE WORKS

A control architecture has been presented that is using tree-axial tactile input to stabilize a flexible pendulum. The corresponding hardware elements have been introduced. The validation of the motor map controller algorithm has been done through inverted pendulum balancing task simulation. The sensor measurements were analyzed and the necessary measurement setup was proposed to extend the usage of the tactile sensors in dynamic applications. Further works are to complete the simulation of the sensors, and record measurements to validate their behavior.

REFERENCES

- [1] Éva Vázsonyi István Bársony Gábor Vásárhelyi, Mária Ádám and Csaba Dűcső. Effects of the elastic cover on tactile sensor arrays. *SENSORS AND ACTUATORS A: PHYSICAL*, 132:245–251, November 2006.
- [2] S. Jain and F. Khorrani. Positioning of unknown flexible payloads for robotic arms using a wrist-mounted force/torque sensor. *IEEE Transactions on Control Systems Technology*, 3(2):189–201, June 1995.
- [3] A. Schlechter and D. Henrich. Manipulating deformable linear objects: manipulation skill for active damping of oscillations. In *Intelligent Robots and System, 2002. IEEE/RSJ International Conference on*, volume 2, pages 1541–1546, 2002.
- [4] Ke-Jun Xu, Cheng Li, and Zhi-Neng Zhu. Dynamic modeling and compensation of robot six-axis wrist force/torque sensor. *IEEE Transactions on Instrumentation and Measurement*, 56(5):2094–2100, October 2007.
- [5] John J. Craig. *Introduction to robotics: mechanics and control*. Pearson Prentice Hall, Berlin, Germany, third edition, 2005.
- [6] Helge Ritter, Thomas Martinetz, and Klaus Schulten. *Neural Computation and Self-Organizing Maps; An Introduction*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1992.
- [7] P. Arena, L. Fortuna, M. Frasca, and G. Sicurella. An adaptive, self-organizing dynamical system for hierarchical control of bio-inspired locomotion. *Systems, Man, and Cybernetics, Part B, IEEE Transactions on*, 34(4):1823–1837, August 2004.
- [8] Jacek M. Zurada. *Introduction to Artificial Neural Systems*. PWS Publishing Co., Boston, MA, USA, 1999.
- [9] M. Shimojo. Mechanical filtering effect of elastic cover for tactile sensor. *IEEE Transactions on Robotics and Automation*, 13(1):128–132, February 1997.

An Improved Biped Actuation System Inspired by the Human Flexor-Extensor Mechanism

József Veres

(Supervisors: Tamás Roska, György Cserey and Gábor Szederkényi)

verjo@digitus.itk.ppke.hu

Abstract—In this paper an improved method of a biped [1] actuation system is presented. The nowadays biped robots are mainly actuated by rotational mechanisms. These electronic drives have a lot of drawbacks compared to the human muscular systems. The in question biped system also has these weaknesses, for example the backlash of the joints. To reduce this effect, a joint level tightness control has been developed that was inspired by the human flexor-extensor mechanism. To achieve that the closed-loop stepper motor commutation has also been created. In the following sections one can find the detailed steps of the above mentioned methods. And at the end of the article the experimental results will also be showed.

I. INTRODUCTION

In the last few decades humanoid robots furthermore the bipeds has become a highly investigated area. There are numerous articles about the different approaches. But we can state most of them uses rotational forces to move their joints [2],[3],[4],[5]. Principally this means a gear driven by an electric motor. Advantages and disadvantages varying by the actually applied technology, for example the cost effectiveness of the DC motors, or the high efficiency of the brushless DCs, etc. But in generally speaking these solutions show significant weaknesses compared to the human muscular system. Among other aspects let me focus on the powertrains. On the robotic side, to produce enough torque for example to lift a leg segment, gear reduction is used. In many cases this can cause a special uncertainty that is absolute unfamiliar to the muscular system, namely the backlash. Of course there are solutions, like the harmonic drive, to minimize this but that would significantly raise the total cost. In our case we have faced with exactly the same problem.

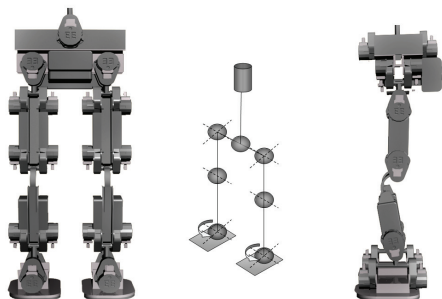


Fig. 1. This figure shows our biped called EE [1]. On the left the front and on the right the side view is placed. In the middle the DOF distribution can be found.

Our robot [1] is a 13 DOF (degree of freedom) biped. It is 50 cm tall and has 3.5 kg total weight. Each joint consists of two 300:1 gear each driven by a stepper motor as you can see on the Figure 1. This paired type joint mechanism is unconventional, since in most of the cases only one electric motor unit is used. The idea is to utilize the possibility of the separate drive of the joints. Like during the leg flexion and extension I would use a dedicated drive for both these two movements. Since these would be antagonists, as the biceps and triceps muscles in the upper arm, I expected that with a proper regulation, tension control could be achieved. And with a controllable level of tightness, we would be able to reduce the effect of the joint's backlash. In order to do that first we had to create a precise individual stepper drive control, that required the use of closed-loop commutation. In the following section the detailed steps of this process will be covered.

II. THE IMPROVED ACTUATION SYSTEM

Since this section will deal with the above mentioned drive method, the stepper type in-detail, thus first the basic fundamentals will be covered.

A. Stepper motor basics

Stepper type motors [6] are belonging to the electronically commutated ones like the BLDCs. These are usually having high pole numbers that gives extra torque during low speed rotation. Principally stepper motors convert digital signals (stepping impulses) into mechanical shaft rotation. The number of step impulses required for one turn is exactly the pole number of the construction. For characterizing it let's denote the step angle with α , the position with θ and the speed with ω then the following holds: (Equations 1.)

$$\alpha = \frac{2\pi}{spr} [rad] \quad \theta = n\alpha [rad] \quad \omega = \frac{\alpha}{\delta t} [rad/sec] \quad (1)$$

Where spr is the steps per round, n is the number of steps and δt is the time difference between the step impulses. The change in θ requires a pull-in torque to overcome the friction and the inertia. And also exists a pull-out torque where the stator pulls out of synchronism. Normally the applied is below the pull-out torque so it never happens, but in our cases that can not be fulfilled all the time. So we can introduce the following errors: (Equations 2.)

$$e_\alpha = \alpha - \hat{\alpha} \quad e_\theta = \theta - \hat{\theta} \quad e_\omega = \omega - \hat{\omega} \quad (2)$$

Where $\hat{\alpha}$ is the real step angle that is taken, $\hat{\theta}$ is the real position and the $\hat{\omega}$ is the real speed.

B. Closed-loop commutation

Originally the drive method was a simple open-loop commutation. There was no back EMF, hall sensor or encoder based feedback. We used full stepping method to create the stepping impulses. In order to achieve more torque micro stepping has been introduced. In the following table, the main differences are summarized. (Table I.)

TABLE I
THE STEPPER MOTOR'S CHARACTERISTICS

	Steps per round	Step angle	Maximum stepping frequency
Full stepping	24	15°	600 Hz
Micro stepping	384	0.94°	9.6 kHz

To create the closed-loop system the feedback of the stator position had to be measured. For this purpose a magnetic rotary encoder is used. This is a contactless way of sensing the stator movement. This was done by placing a cylinder shaped two-pole rare earth magnet on the free end of the rotor's axis. Since the applied magnet's magnetic force is relatively high and the encoder is based on a hall sensor array, the motor's varying magnetic field significantly does not disturb the operation. The major properties are listed in the Table II.

TABLE II
THE ROTOR ANGLE FEEDBACK'S CHARACTERISTICS

Resolution (per turn)	Smallest slew value	Maximum Feedback rate	Maximum allowed rotational speed
12 bit	0.088°	2.6 kHz	9440 RPM

Principally in a closed-loop commutation system the stator moving impulses are controlled by the feedback sensor's signal. In this case this means briefly we can not step over a position if the stator has not arrived there yet. Formally speaking consider $\theta_d = \infty$ is the desired stator position, and let the initial values θ and $\hat{\theta}$ equal to 0, then the step function is the following: (Equation 3.)

$$s(k+1) = \begin{cases} 1 & \text{if } |e_\theta| \leq D \\ 0 & \text{if } |e_\theta| > D \end{cases} \quad (3)$$

Where D is a position error threshold level, $k \in N$ the time instance identifier and θ estimated position is the following: (Equation 4.)

$$\theta(k+1) = \begin{cases} \theta(k) + 1 & \text{if } s(k+1) = 1 \\ \theta(k) & \text{if } s(k+1) = 0 \end{cases} \quad (4)$$

For a better understanding Figure 2. shows a flowchart of the above mentioned feedback loop. As you can see it begins with an initialization phase where a θ initial value is read from the sensor to fulfill the $e_d \leq D$ condition. The $t(\omega_d)$ is the step impulses δt time in millisecond that creates the given ω_d angular velocity.

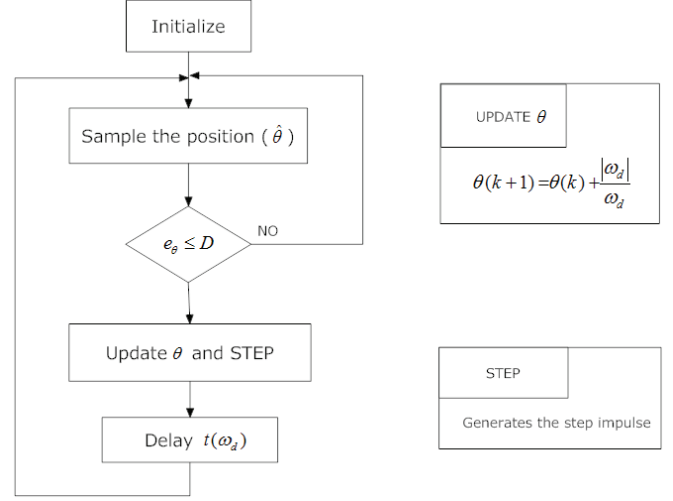


Fig. 2. This figure shows the simplified flow chart of the rotor movement's feedback loop. It begins with an initialization phase where a θ initial value is read from the sensor to fulfill the $e_d \leq D$ condition. The $t(\omega_d)$ is the step impulses δt time in millisecond that creates the given ω_d angular velocity.

Under some circumstances this loop can cause a dead-lock. Since if the $e_\theta \leq D$ can not be reached it will wait for the infinity. To handle this we must integrate the time we spent in this status. After an appropriate time we must equalise θ with θ_d to force the system to over step on this and of course reset the integration. Let's denote this time by t_i , then we can introduce an other error and a new step impulse delay time: (Equation 5.)

$$e_t = t(\omega_d) - t_i \quad t_a = t(\omega_d) + e_t * P \quad (5)$$

If e_t becomes greater then a predefined difference threshold value (I) the new t_a replaces the $t(\omega_d)$ in the *Delay* subroutine to assist the acceleration and de-acceleration phase.

So the loop could be controlled by ω_d from a higher level and there are three different parameters that bias the operation. D is responsible for the hold in torque or in angular velocity. If you increase the value of D the more you get a constant angular velocity and the decreasing means the more constant torque you have. I sets the time when your basic loop depicted on Figure 2. stops waiting and tries to overstep the lacked state. And the P coefficient can be used to set the gradient of that time acceleration and de-acceleration phase.

There are also output values for the higher level control. The $\hat{\theta}$ stores the actual position value of the stator and the $\hat{\omega}$ that gives the actual angular velocity.

C. Flexor-Extensor Model

In the human muscular system there is no rotating forces, since muscles are only able to pull. Due to this, muscles needs

antagonists, like the flexors and the extensors or the abductors and the adductors. This is a totally different approach compared to what we used to in the field of the bipeds. In a healthy case, tension of the antagonist pairs are always controlled to fit the actual need. One of the simplest model of this tightness control is the reciprocal innervation of the antagonist muscle's α motoneurons. On the Figure 3. this type of collateral inhibition is illustrated.

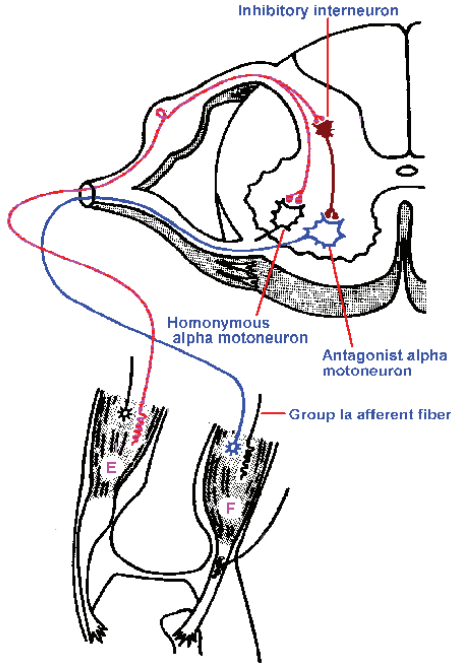


Fig. 3. An illustration [7] of a collateral inhibition of the antagonist muscle. The extensor muscle's spindle innervates an inhibitory interneuron that inhibits the antagonist flexor muscle.

The effected muscle groups are the quadriceps and the hamstrings, that extends and flexes the leg respectively. In this case the extensor muscle's spindle innervates an inhibitory interneuron that inhibits the antagonist flexor muscle.

This scheme could be applied in our actuation system in the following way. Consider the two across placed close-loop commutated stepper motors with their gear boxes. Denote by $\hat{\theta}_1, \hat{\omega}_1$ and $\hat{\theta}_2, \hat{\omega}_2$ the left and the right ones position and angular velocity, respectively. Let the main desired angular velocity be ω_{12} . Then introduce the following errors: (Equation 6.)

$$e_1 = \omega_{12} - \hat{\omega}_1 \quad e_2 = \omega_{12} - \hat{\omega}_2 \quad (6)$$

Then as we have calibrated the initial values, the following high level control values are generated: (Equation 7.)

$$\omega_{d1} = \omega_{12} - e_2 * S_1 \quad \omega_{d2} = \omega_{12} - e_1 * S_2 \quad (7)$$

That follows the above mentioned model where the antagonist error changes the agonist behaviour. With the S_1 and the S_2 parameters we can tune the strength this effect. Of course by using more sophisticated algorithms instead of the cross connection, this can be improved.

III. EXPERIMENTAL RESULTS

To see how the new actuation system's performs, first I set to the maximum thightend level. This was done by measuring the position difference that I could cause by pushing and pulling a joint manually. The measured position came from the very same magnetic rotary encoder that was used to create the closed-loop commutation. The backlash was in the range of 0.1° - 0.15° . Originally it was in the range of 1.5° - 2° , which was at least a 10 times greater.

The next experiment was made to test it in a probable situation. It had to follow a sine wave mimicking a typical extension and felxation period. Figure 4 shows the result of the old actuation system. As you can see it tries to follow but the continuous impacts caused by the large backlash creates significant position errors.

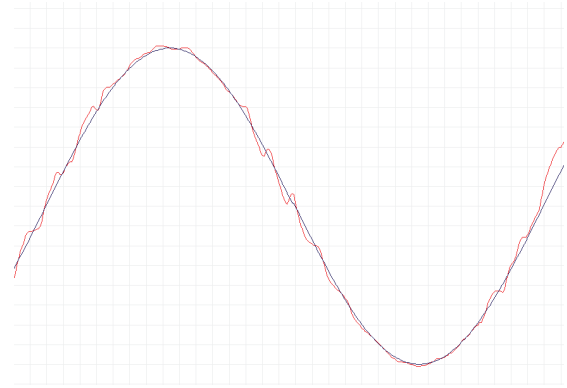


Fig. 4. This figure shows the old actuation system while trying to follow the desired sine wave. The high impacts caused by the large backlash creates significant position errors.

Then this experiment was repeated with the improved one. The Figure 5. shows that's result. The motion was apparently smoother, there was no large impacts. This suggest us that during the motion the joint keeps thightend that significantly reduce the effect of the backlash.

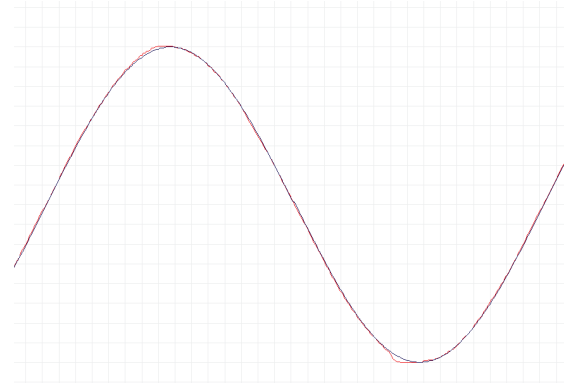


Fig. 5. In this figure the improved actuation system sine wave following result can be seen. The motion is apparently smoother, there is no large impacts.

IV. CONCLUSION

An improved biped actuation system was presented. The original full stepping method has changed to a better micro stepping. A new closed-loop commutation is also introduced that significantly increased the system reliability. The previous ones and the new improved speed control gave me the ability to create flexor-extensor like bio-inspired mechanism. This was significantly reduced the rate of the backlash, and apparently created a much smoother operation.

ACKNOWLEDGEMENTS

The Office of Naval Research (ONR), the Operational Program for Economic Competitiveness (GVOP KMA), the NI 61101 Infobionics, Nanoelectronics, Artificial Intelligence project, and the National Office for Research and Technology (NKTH RET 2004) which supports the multidisciplinary doctoral school at the Faculty of Information Technology of the Pázmány Péter Catholic University are gratefully acknowledged. The author is also grateful to Professor Tamás Roska, György Cserey, Gábor Szederkényi and the members of the Robotics lab for the discussions and their suggestions.

REFERENCES

- [1] A. Tar, J. Veres and Gy. Cserey, "Design and Realization of a Biped Robot Using Stepper Motor Driven Joints", IEEE Proceedings International Conference on Mechatronics, 493-498, 2004
- [2] M.A. Lewis, F. Tenore and R. Etienne-Cummings, "CPG Design using Inhibitory Networks", Proceedings of the 2005 IEEE International Conference on Robotics and Automation, pp. 3682-3687, April 2005
- [3] C. Chevallereau and P. Sardain, "Design and Actuation Optimization of a 4 axes Biped Robot for Walking and Running", Proc. IEEE Int. Conf. on Robotics and Automation, pp. 3365-3370, San Francisco, CA-USA, April 2000.
- [4] J. Shan and F. Nagashima, "Neural Locomotion Controller Design and Implementation for Humanoid Robot HOAP-1", Proceedings of The 20th Annual Conference of the Robotics Society of Japan, Osaka, October 2002.
- [5] A. L. KUN and W. T. MILLER, "Adaptive Static Balance of a Biped Robot Using Neural Networks." Fifth IASTED International Conference on Robotics and Manufacturing, Cancun, Mexico, May 1997.
- [6] P.P. Acarnley, "Stepping Motors: A guide to modern theory and practice.", Stevenage, UK., Pregrinus, 1982
- [7] J.P. Schadé, D.H. Ford, "Basic Neurology", Amsterdam, Elsevier, 1965

3D Geometry Reconstruction using Large Infrared Proximity Array for Robotic Applications

Ákos Tar

(Supervisors: Tamás Roska, György Cserey and Gábor Szederkényi)

Email: tarak@itk.ppke.hu

Abstract—In this paper, we propose a novel Large Infrared Proximity Array (LIPA), which is capable of reproducing 3D images of the target object. An IPA uses infrared sensors and infrared emitters to accurately measure distance and thus, creates 3D monographic geometry of the sensed objects in real time. In the current setup infrared emitters (LEDs) and photo transistors are used, placed in one package. In many applications determining an obstacle’s height, orientation, or distance with a very high resolution is not needed (less than 1-2mm is enough) but the sensor often should have a big sensing area. Contrary to nowadays techniques we tried to form a rather large but sparse sensor array to cover as big part of the surrounding area as possible. In addition it has the detection range of 1cm to 15cm (it can be increased depending on the application requirements). 64 LED and photo transistor pairs were used and placed in an 8×8 matrix order creating a $9cm \times 9cm$ sensor array. We also demonstrate how the resolution of these kind of systems can be increased. The concept of the proposed sensor array spans a wide range of potential applications, i.e. in robotics.

Keywords: Infrared sensing, mobile robots, LIPA, IPA

I. INTRODUCTION

One of the most challenging areas in robotics is autonomous navigation in an unknown environment. In order to achieve this, the robot must detect obstacles within its vicinity as fast as possible. The traditionally used methodology is sensor based motion planning. Nowadays’ common solution is to equip the robot with laser scanner, camera, ultrasonic range finder and infrared proximity sensors [1]. Ultrasonic range finders are widely used but due to their wide sensing area, sensitivity to special surfaces and the minimum distance needed to be measured (approximately 5-10cm), its usage is restricted to object detection or helping to guide such tasks as docking or wall following. 3D camera and laser scanner are widely used to reconstruct the sensed object’s 3D geometry ([2]–[4]), but they are very computation and power consuming solutions and both solutions have its own drawbacks. In 3D camera systems the obstacle must be further than a few centimeters, and many mobile robots can not handle computation in real time. Laser scanners are very accurate but they only create a 2D image (3D with a built in motorization) and nowadays’ commercial versions are relatively large and still expensive [5]. Infrared sensors are widely used in many kinds of robotic applications. It’s easy to use, just a few components are needed for operation, and it has a relatively small size. There are also surface mount versions available. Its measurement range can be from millimeters to meters depending on the amplification, but it is typically used in 1 to 30 cm range. Unfortunately,

the sensor is very sensitive to the reflectance properties of the object. This is the reason why an infrared sensor is typically used for in or out range detection, meaning obstacle or no obstacle. Although there are some techniques to validate the measurements, like using the Phong illumination model to measure accurate distance [6], or using sensor fusion to come over this problem [7], [8]. The IPA sensors based on CCD technology meet the sensor requirements to be small, lightweight and portable and to produce real time information. In applications where short distance (1-15cm) and large area coverage are both needed and the sensed resolution does not need to be very high, simple infra LEDs and photo transistors could be used, i.e. Modularized Sensitive Skin [9], [10]. One typical application could be detecting the environment under the feet of a humanoid, bipedal or any kind of legged robot, where very high accuracy of detecting the orientation, size and distance of the obstacle is not needed, 1-5mm is enough, but the whole area under the robot’s feet need to be sensed. With LIPA [11] we can create a 3D image of the area and use this information in motion planning algorithms. In the case of sparse sensor arrays, the computation power needed for the measurement can be decreased. By using the fact that in our LIPA, each infra LEDs can be independently switched on and off and the reflection according to one infra LED is sensed by multiple photo transistors, it can be used to increase the resolution of the system by using a multiple reflection model. So optimalization can be made in the number of sensors in the array and the desired resolution.

This paper is organized as follows. First in Section I we describe the Sensor Array where the infra LEDs and photo transistors are placed. Section II describes a method of how spatial resolution can be increased. In Section III the new Tactile Sensor master board capabilities can be seen. Section IV present some measurement results. Conclusions are presented in Section V.

II. SENSOR ARRAY

Due to the high variety of available infra LEDs and photo transistors on the market, different measurement ranges, sensor integrity and power consumption can be achieved, thus the sensor array is separated as a dedicated board. This makes it much easier to test several types of infra LEDs and photo transistors. In the current setup we are using TCRT1000 sensors, which have an operation range from 0.2mm up to 4 mm. We extended this range to 2mm - 4cm. The reason of the sensor

choice is that we could get the requested quantity of only this kind of sensor. The sensor's functional diagram can be seen in Figure 1a. Each infra LED can be controlled independently by the appropriate I/O pin of the microcontroller. Each photo transistor has its own resistor (as shown in Figure 1a/R). This resistor acts as a current to voltage converter and amplifier. In Figure 1b, the sensor distance vs. output voltage diagram can be seen using a cartonboard as the reflection object. Because of the standard deviation in the manufacture process, each sensor is identical, and before the measurements we calibrated each sensor with a 0.1mm precision step.

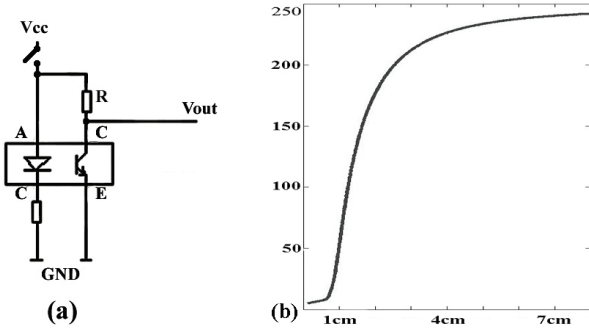


Fig. 1. (a) shows the sensor functional diagram. Each photo transistor has a resistor (R) that acts as a current-to-voltage converter and amplifier. (b) shows the used sensor voltage vs. distance characteristic in the range of operation.

These sensors are 7mm wide, 4mm thick and were placed in an 8×8 matrix order, trying to keep equidistance between the infra LEDs and photo transistors in the array. The sensor grid size is about 9×9 cm.

III. IMPROVING THE RESOLUTION

The used sensor is a reflective type so the infra LED signal reflects to its appropriate photo transistor placed in the same package. Hence in the sparse sensor array the native resolution is the distance of two sensors. In our case it is 1 cm. To translate the photo transistor output into distance, first we must solve Equation (1) for native points.

$$h_{s,d} = LUT(O(S_{i,j})) \quad (1)$$

where $O(S_{i,j})$ is the binary output of the sensor, placed in row i and column j when only its LED was on.

With the LIPA we can create subpixel resolution without moving the sensor plane. We are using the fact that the obstacle reflects the sensor infra LED signal not only to the corresponding photo transistor but to its surrounding area too. So lighting with one infra LED and measuring with other photo transistors in a radius, we can create higher spatial resolution. The new point, called imaginary point is defined between two sensors' native points so the resolution is doubled, in our case instead of 1 cm it is 5 mm, as it can be seen in Figure 2.

The sensor output does not give the imaginary point distance directly, because it is not in front of the sensor but it measures

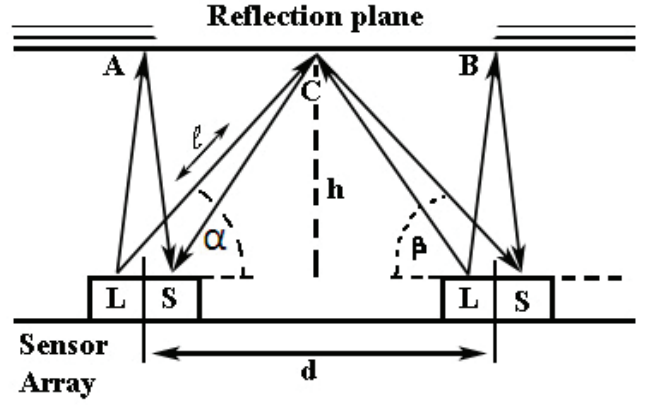


Fig. 2. Using the fact that when an infra LED is switched on, the reflection object not only reflects the light to the appropriate photo transistor but the other nearby sensors photo transistors as well, the resolution can be increased. This image shows two sensors and their native pixels (A) and (B). Using one sensor infra LED and measuring with the other sensor photo transistor, a new point (imaginary point C) of the obstacle can be defined. With this method the spatial resolution can be doubled as can be seen in Figure 3a. The new point C' measured distance is h , d is the distance of two sensors, the sensor output is the length of l .

the distance l showed in Figure 2. Thus the imaginary point distance has to be calculated. Given the used infra LED and the used photo transistor distance (d), the reflection angle (α) and the photo transistor output (l) as they form a rectangular triangle by using the Pythagoras Equation (2) we can calculate the imaginary point distance from the sensor plane.

$$k = \sqrt{l^2 - \left(\frac{d}{2}\right)^2} \quad (2)$$

where d is the distance of two sensors, k is the height of the imaginary point.

There is high correlation between the infra LED intensity and the reflective angle as can be seen in Figure 3b. If the reflective angle is greater than 10 degrees, an output correction has to be made. Equation (4) gives a good approximation in the change of illumination with respect to the reflective angle. To get an accurate distance, we have to take into account that in the case of an imaginary point, the reflection angle is much larger than for native points, hence the saved LUT cannot be used for linearization. This relationship is also true in the case of the photo transistors. If the angle of incidence is greater than 10 degrees output correction also has to be made as shown in Equation (5).

The correction can be made by multiplying the sensor output with the infra LED intensity, as shown in Equation (3), and photo transistor sensitivity. After the correction, the appropriate previously saved LUT can be applied.

$$l = LUT[O(L_{i,j}, P_{o,t}) \cdot I_{LED} \cdot S_{Photo}] \quad (3)$$

$$I_{LED} = \cos(0.74 \cdot (90 - \alpha)) \quad (4)$$

$$S_{Photo} = \cos(0.88 \cdot (90 - \beta)) \quad (5)$$

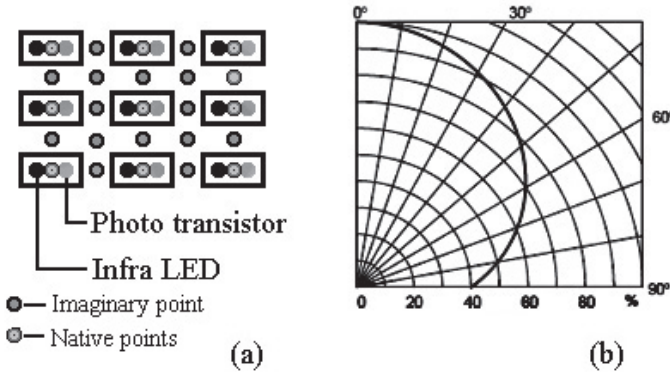


Fig. 3. (a) shows the native and the imaginary point in a 3X3 sensor array, (b) is the infra LED characteristic, it can be clearly seen if the reflective angle is greater than 10 degrees, an output correction has to be made.

where $90-\alpha$ denotes the reflective angle, $90-\beta$ is the angle of incidence, $O(L_{i,j}, P_{o,t})$ is the binary output, when the photo transistor in row o and column t was measured and the infra LED in row i and column j was emitted, l is the measured distance shown in Figure 2.

So in the case of imaginary point the appropriate measured distance ($h_{s,d}$), can be given as shown in Equation (6).

$$h_{s,d} = \sqrt{LUT[O(P_{i,j}, L_{o,t}) \cdot I_{LED} \cdot S_{Photo}]^2 - \left(\frac{d(P_{i,j}, L_{o,t})}{2}\right)^2} \quad (6)$$

where $h_{s,d}$ is the measured distance for an imaginary point in row s and column d , and $d(P_{i,j}, L_{o,t})$ is the distance between the measured photo transistor and the emitted LED.

Unfortunately as it can be seen in Equation (6), α and β are undefined. To get a relevant distance, somehow we have to make a good estimation for α and β . To do that, first in Equation (7) we calculate a distance using the sensor output without the correction terms denoted as $\hat{h}_{s,d}$.

$$\hat{h}_{s,d} = \sqrt{LUT[O(P_{i,j}, L_{k,t})]^2 - \left(\frac{d(P_{i,j}, L_{k,t})}{2}\right)^2} \quad (7)$$

We assume that the imaginary point is right in the middle of two native points, so α and β are part of a rectangular triangle and they are equal. Using this assumption α can be calculated in Equation (8), and can be used in Equation (6) to calculate an estimated distance.

$$\alpha = \arctan\left(\frac{\hat{h}_{s,d}}{\frac{d(P_{i,j}, L_{k,t})}{2}}\right) \quad (8)$$

The notation between $\hat{h}_{s,d}$ and $h_{s,d}$ is shown in Equation (9).

$$\hat{h}_{s,d} \geq h_{s,d} \quad (9)$$

where $\hat{h}_{s,d}$ is the measured distance without correction, $h_{s,d}$ is the measured distance of the imaginary point with correction. It can be seen in Equation (6) that $h_{s,d}$ is not so sensible to the small variation of α (maximum ± 5 degree). With a good approximation we can create an accurate distance estimation of the imaginary point.

IV. MEASUREMENTS

In this section we would like to present some hardware experiment results using the presented LIPA. The infrared distance measurement is very sensitive to various measurement parameters like absorption and the different reflection property of the obstacle. For instance, a shiny surface will reflect more energy than a matte surface, thus in the test we tried to use objects with homogeneous texture. After the calibration process, LIPA can approximate the 3D geometry of an object in real-time. Unfortunately in our case, because of the used sensor properties, it is less than real-time.

For 3D visualization, Equation (1) and (6) are applied to get the depth of an object. An example of 3D geometry reconstruction can be seen in Figure 4.

In the first test we placed a small wrench over the sensor array, in an approximately 4cm high. Figure 4a shows the native points in the array. Figure 4b show, the native points and the mean of the neighboring native points. Figure 4c shows the improved resolution image using the introduced method. Figures 4d,e,f shows a 2D slice of the above images reconstructing the object geometry placed on the sensor array. The demonstrated method gives a more realistic image of the obstacle, as can be seen in Figures 4c,f, whereas using the native resolution, shown in Figures 4a,d, the object can not be reconstructed.

V. CONCLUSION AND FUTURE WORK

In this paper a LIPA (Large Infrared Proximity Array) for 3D object geometry reconstruction has been presented and proved to be useful for quick scan and coarse 3D geometry reconstruction. The sensor array is $9 \times 9cm$ and containing 64 sensors placed in an 8×8 matrix order. The native resolution of these kinds of systems is the nominal distance of two sensors. We investigated a method which can increase it without adding more sensing elements. In this way, the spatial resolution is doubled. This result suggests that if the achieved resolution is higher than needed, sensor elements can be abandoned.

Several test cases show the application capabilities. As it is well known, the infrared sensors are very sensitive to measurement conditions but they can be used in a certain environment after calibration and additional sensor fusion can improve distance measurement.

The presented system's structural complexity yields to be an affordable, easy-to-use and well scalable sensor board for any kind of robotics application, whereas traditional solutions would be too complicated. It could be useful in object geometry reconstruction, collision detection,

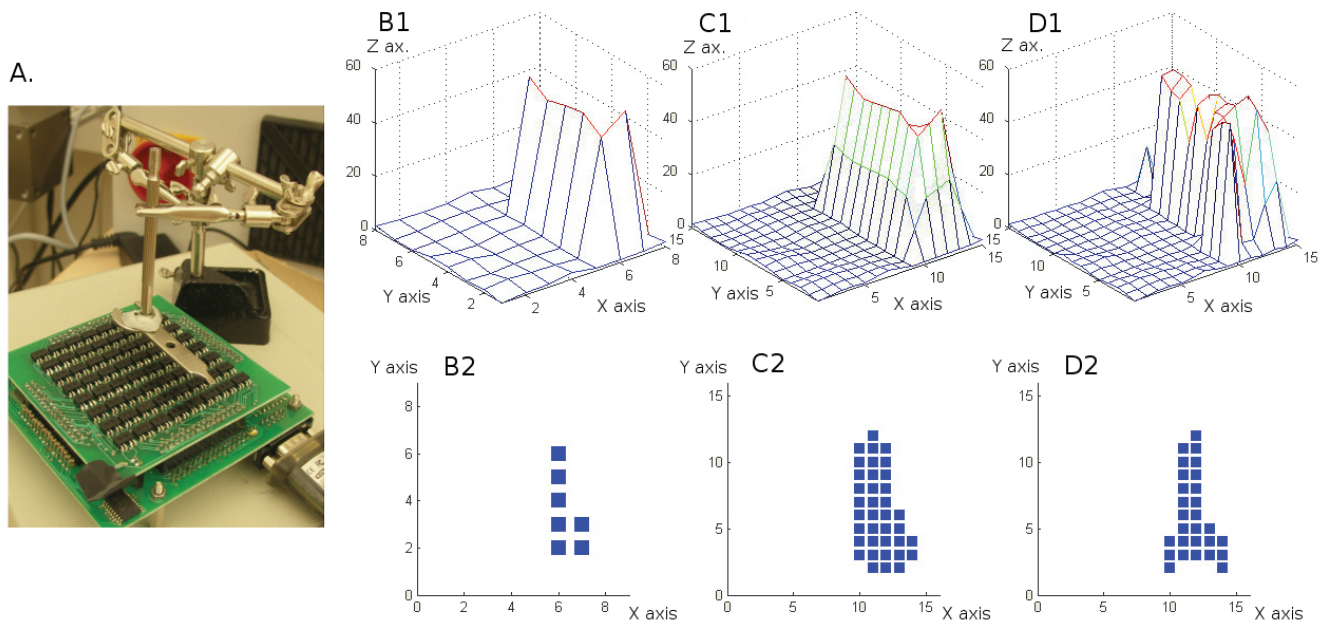


Fig. 4. Measure setup with a little wrench. On each picture the X and Y axes indicate the measured points in a row and column order above the ($9\text{cm} \times 9\text{cm}$ large) sensor array. The Z axes represent the measured distance from the surface of the panel in millimeters. (A) shows the photographic documentation of the setup. (B1) shows the picture constructed from native points. (C1) is an interpolated image using the native points and their means. (D1) shows the result image using our measurement method. (B2),(C2),(D2) show horizontal slices of the above 3-dimensional surfaces. Here can be seen that with our method the measured object shape is well recreated and the object can be recognized.

movement planning applications in unknown environment for manipulators or any kind of legged robots, typically object position and orientation detection under a biped feet.

In the future, we would like to change the infra LEDs and photo transistors used in the sensor array and realize on board ambient light correction.

We also would like to perform several tests with the LIPA mounted on a bipedal robot foot.

ACKNOWLEDGMENT

The Office of Naval Research (ONR) and the Operational Program for Economic Competitiveness (GVOP KMA) which supports the multidisciplinary doctoral school at the Faculty of Information Technology of the Pázmány Péter Catholic University is gratefully acknowledged. And special thanks goes to Miklos Koller.

REFERENCES

- [1] H. R. Everett, *Sensors for mobile robots: theory and application*. Natick, MA, USA: A. K. Peters, Ltd., 1995.
- [2] D. Murray and J. Little, "Using Real-Time Stereo Vision for Mobile Robot Navigation," *Autonomous Robots*, vol. 8, no. 2, pp. 161–171, 2000.
- [3] Y. OIKE, M. IKEDA, and K. ASADA, "A 375 x 365 high-speed 3-D range-finding image sensor using row-parallel search architecture and multisampling technique," *IEEE journal of solid-state circuits*, vol. 40, no. 2, pp. 444–453, 2005.
- [4] J. Martínez, A. Pozo-Ruz, S. Pedraza, and R. Fernández, "Object following and obstacle avoidance using a laser scanner in outdoor mobile robot auriga- α ," in *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 204–209, 1998.
- [5] J. Guivant, E. Nebot, and S. Baiker, "Localization and map building using laser range sensors in outdoor applications," *Journal of Robotic Systems*, vol. 17, no. 10, pp. 565–583, 2000.
- [6] P. Novotny and N. Ferrier, "Using infrared sensors and the Phong illumination model to measuredistances," in *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, vol. 2, 1999.
- [7] A. Sabatini, V. Genovese, E. Guglielmelli, A. Mantuano, G. Ratti, and P. Dario, "A low-cost, composite sensor array combining ultrasonic and infrared proximity sensors," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, vol. 3, pp. 120–126.
- [8] A. Flynn, "Combining Sonar and Infrared Sensors for Mobile Robot Navigation," *The International Journal of Robotics Research*, vol. 7, no. 6, p. 5, 1988.
- [9] D. Um, B. Stankovic, K. Giles, T. Hammond, and V. Lumelsky, "A modularized sensitive skin for motion planning in uncertain environments," in *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, vol. 1, 1998.
- [10] E. Cheung and V. Lumelsky, "Proximity sensing in robot manipulator motion planning: system and implementation issues," *Robotics and Automation, IEEE Transactions on*, vol. 5, no. 6, pp. 740–751, 1989.
- [11] A. Tar, M. Koller, and G. Cserey, "3d geometry reconstruction using large infrared proximity array for robotic applications," in *IEEE Proceedings International Conference on Mechatronics*, 2009.

Stream Processing Evaluation Platform and some Application Results

László Füredi

(Supervisor: Dr. Péter Szolgay)

furla@digitus.itk.ppke.hu

Abstract—Stream processor, Inc. (SPI) has introduced the Storm-1 stream processing System-on-a-Chip (SoC) which contains 16 data-parallel processing arrays with five Very Long Instruction Word (VLIW[1]) Arithmetic Logic Units (ALUs) in each lane (all together 80 processing units). In the paper some of basic operation is described with simulated tradeoffs. The implementation of emulated-digital Cellular Neural Networks Universal Machine (CNN-UM) simulation kernel on the Storm-1 Stream Processing Platform (SPP) is described as well and optimized according to the special requirements of SPP. The area/speed/power tradeoffs of my solution and other emulated digital CNN implementations are also compared.

Index Terms—Stream processing, Stream processor, emulated digital CNN-UM.

I. INTRODUCTION

Nowadays computer system architectures and Integrated Circuit (IC) development trends are forecasting a new kind of Moore's law, namely the number of processors are doubled in each 12-18 months on the same silicon area[2]. The different kind of array computers can effectively solve computation intensive problems such as image and signal processing, or Partial Differential Equations (PDEs). The question is how should we select from the different array architectures to solve a given problem. A new kind of stream processor architecture was used to simulate a Cellular Neural Network (CNN) with linear and non-linear weights/templates [3],[4],[5]. The results were able to be used in solution of other spatiotemporal problems. In section II the Stream processor architecture is defined, in section III the Stream processor programming language specialities and some basic operation is described and in section IV the emulated digital CNN model is derived. The CNN model implementation on stream processor array is discussed in section V and performance comparisons are given in section VI.

II. STORM-1 STREAM PROCESSOR ARCHITECTURE

The SPI's Stream Processor Architecture (Figure 1.) with a high-performance data-parallel unit (DPU). The Storm-1 SP16-G160 is able to sustain 160 billions of operations per second (GOPS) and 80 billions of multiply-accumulate operations per second (GMACS) at 500 MHz. The architecture includes two industry-standard CPU cores a host CPU (System Multi-Instruction Processing System (SMIPS)) for system-level tasks (runs Linux and handles I/Os) and a digital signal processing (DSP) Coprocessor Subsystem where the DSP SMIPS runs the main threads and off loads processing of

compute-intensive kernel functions to the DPU. Both of the SMIPS runs only on 250 MHz clock frequency [6],[7]. The SoC designed with TSMC 0.13 μ m CMOS technology and packed to a 896-pin, 1.0 mm ball pitch heat spreader ball grid array (HSBGA). A key feature of the architecture is its compiler-managed memory hierarchy that leverages the data-parallelism and locality characteristics of signal processing applications. The simple C programming model allows the specification of compute-intensive kernel functions that process streams of data records, enabling the compiler and hardware to efficiently manage on-chip memory and synchronize runtime Direct-Memory Access (DMA). The internal DMA controllers has 28 channels, and through this DMA the whole SoC uses the external memory, which is a 16- to 128-bit DDR2 with 16 MB to 2 GB capacity. The connection bandwidth with the memory is 6.4 GBytes/s with 400 Mbits/s rate. This approach eliminates the need for a cache and greatly increases predictability of throughput, simplifying the overall programming task[8].

The architecture exploits multiple levels of parallelism:

- task-level parallelism between the system processor, DSP processor and DPU,
- data-level parallelism (DLP) with multiple lanes executing the same instructions on different data in parallel,
- instruction-level parallelism (ILP) via very long instruction word (VLIW) driving multiple arithmetic logic units (ALUs) per lane,
- sub-word single instruction multiple data (SIMD) in which each ALU can operate on multiple operands.

The DPU Dispatcher receives kernel function calls to manage runtime kernel and stream loads. On the DPU, a kernel function runs identically on every lane processing different data, for example ALU 0 in each lane executes the same operation in the same cycle but on different data elements. The local stream data stored in the Lane Register File (LRF) of each lane, which size is 256 KBytes, so each lane register file is 16 KBytes. Kernel functions are executed one at a time until completed across the lanes with a static VLIW schedule. Data can be shared across different lanes via the InterLane switch under the guidance of the compiler. Each lane has a set of distributed Operand Register Files (ORF) allowing for a large working data set and processing bandwidth exceeding 1 TeraByte/s. Each ALU contains a MAC unit and is capable of subword parallelism for combinations of 8, 16, and 32-bit op-

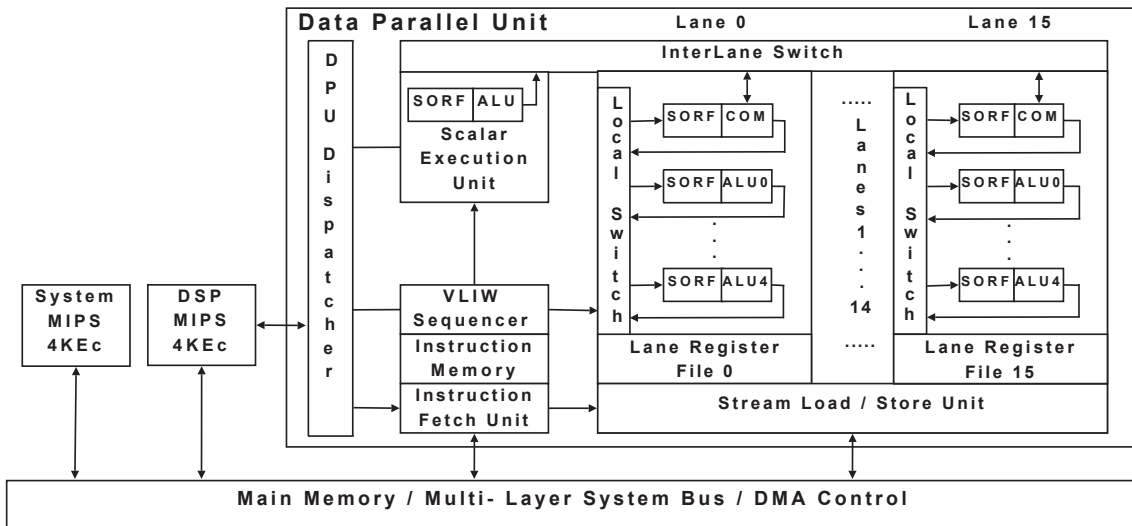


Fig. 1. Block diagram of the Stream processor

erations. This high-speed inter-lane communications provides more versatility than conventional SIMD architectures. The ORF has 19 KBytes size. The Stream Load/Store Unit provides gather/scatter with a wide variety of access patterns. The Inter-Lane Switch is a compiler-scheduled, full crossbar for high-speed access between lanes. For the internal communication uses a 24 GBytes/s interconnect bus. The single-threaded execution model provides inherent load-balancing, eliminating the need for code partitioning across multiple cores. Another advantage to SPIs architecture is the ability to easily scale to higher levels of performance by adding more or less lanes without the need to restructure my software.

III. STREAM PROCESSOR PROGRAMING AND EXAMPLES

The stream processor like architecture model organizes an application program into streams and kernels. A stream is a sequence of data items; the kernel keeps intermediate data locally on-chip to minimize external memory transactions and it is a function that performs computationally intensive operation on one or more input streams to produce one or more output streams. The flow of streams through kernels is expressed in an extended version of C called StreamC, described in the RapiDev Reference Manual[9]. The general purpose processing unit that executes a StreamC program is a conventional scalar processor, while the DPU executes the kernel functions as a SIMD processor. The data transfer between kernel and DPU is defined as VLIW fashion way.

The VLIW sequencer manages the execution of instruction in the DPU kernel functions. The sequencer branches accordingly based on changes to the instruction stream. The sequencer gets commands from the DPU dispatcher to start execution at a given instruction memory address. The sequencer then proceeds to fetch the VLIW instructions out of the instruction memory. The instructions are decoded and sequenced accordingly until an end is reached. An end is

TABLE I
STREAM PROCESSOR BASIC OPERATION THROUGHPUT TABLE

Example	Description	Throughput (MPixels/s)
3×3 Convolution	Performs a 3×3 convolution on 256 input pixels at a time.	2020
3×3 Convolution using 8-bit coefficients	Performs a 3×3 convolution on 256 input pixels at a time, using 8-bit coefficients.	3610
5×5 Convolution	Performs a 5×5 convolution on 64 pixels of input row at a time.	1123
16×16 SAD	Performs a 16×16 SAD on 256 pixels (a 16×16 block) of input per lane.	106

encoded in the instruction to signal to the sequencer to pass control back to the DPU dispatcher.

IV. CNN MODEL

My primary goal is to get an efficient CNN [3] implementation on the stream architecture. Consider the CNN model and its hardware effective discretization in time.

A. Linear templates

The state equation of the original Chua-Yang model [3] is as follows:

$$\dot{x}_{ij}(t) = -x_{ij} + \sum_{C(kl) \in N_r(i,j)} A_{ij,kl} y_{kl}(t) + \sum_{C(kl) \in N_r(i,j)} B_{ij,kl} u_{kl} + z_{ij} \quad (1)$$

where u_{kl} , x_{ij} , and y_{kl} are the input, the state, and the output variables. A and B are the feedback and feed-forward

templates, and z_{ij} is the bias term. $N_r(i,j)$ is the set of neighboring cells of the $(i,j)^{th}$ cell. The discretized form of the original state equation (1) is derived by using the forward Euler form and the Full Signal Range (FSR) model used to simplify the computation. The discretized version of the CNN state equation with FSR model is as follows:

$$x_{ij}(n+1) = \begin{cases} 1 & \text{if } v_{ij}(n) > 1 \\ v_{ij}(k) & \text{if } |v_{ij}(n)| \leq 1 \\ -1 & \text{if } v_{ij}(n) < -1 \end{cases}$$

$$v_{ij}(n) = (1-h)x_{ij}(n) +$$

$$+h \left(\sum_{C(kl) \in N_r(i,j)} \mathbf{A}_{ij,kl} x_{kl}(n) + \sum_{C(kl) \in N_r(i,j)} \mathbf{B}_{ij,kl} u_{kl}(n) + z_{ij} \right) \quad (2)$$

Now the x and y variables are combined by introducing a truncation, which is simple in the digital world from computational aspect. In addition, the h and (1-h) terms are included into the A and B template matrices resulting templates $\hat{\mathbf{A}}$, $\hat{\mathbf{B}}$.

B. Nonlinear templates

In general the nonlinear CNN template values are defined by an arbitrary nonlinear function of input variables (nonlinear B template), output variables (nonlinear A template) or state variables and may involve some time delays. The survey of the nonlinear templates shows that in many cases the nonlinear template values depend on the difference of the value of the currently processed cell (C_{ij}) and the value of the neighboring cell (C_{kl}). The Cellular Wave Computing Library [4] contains zero- and first-order nonlinear templates. Therefore I focused to implement these two types of templates, my solution can be extended to handle any type of nonlinearity[10].

In case of the zero-order nonlinear templates, the nonlinear functions of the template contains horizontal segments only as shown in Figure 2(a). This kind of nonlinearity can be used, e.g., for grayscale contour detection [4].

In case of the first-order nonlinear templates, the nonlinearity of the template contains straight line segments as shown in Figure 2(b). The application of a nonlinear template is not a problem in the emulated digital realization. This type of nonlinearity is used, e.g., in the global maximum finder template [4]. Naturally, some nonlinear templates exist in which the template elements are defined by two or more nonlinearities, e.g., the gray-scale diagonal line detector [4].

V. EMULATED-DIGITAL CNN IMPLEMENTATION ON STREAM PROCESSOR ARRAY

The emulated digital CNN implementation assumes unsigned 8-bit character input data and unsigned 8-bit character for state variables. The simulator uses signed 16-bit filter coefficients (in Q14 fixed point format) for feed-back and feed-forward templates, and signed 16-bit for bias (as well in Q14 fixed point format). The output is unsigned 8-bit data which is clipped. For better running time we can use only 8-bit weights as well. The input image needs to be padded on all four sides

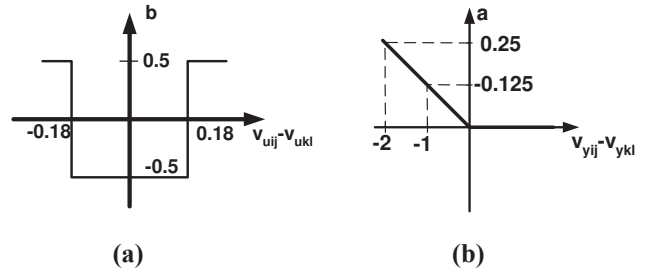


Fig. 2. Zero- (a) and first-order (b) nonlinearity

with (filter size/2) pixels in order to calculate the result for the border pixels. The most commonly used boundary condition is specified in the simulator. For optimal solution the padded image width must be adjusted to be a multiple of 64, because all of the lanes get 4 pixels in a cycle and this implementation uses 16 lanes for calculation. First of all the constant g_{ij} needs to be calculated, and for this calculation the CNN equation input, the bias, the step size and the Feed-forward coefficient matrix are needed.

Each lane receives one 32-bit word at a time as input. This 32-bit word packs four, 8-bit input pixels internally, so each lane processes four pixels at a time. Therefore, processing a total of 256 pixels requires four internal processing stages because the architecture contains 16 lanes.

The implementations get their input through input streams, use intermediate streams for temporary storage, and deliver the output in output streams. The algorithm used in this implementation requires three input streams. The first stream consists of the 12 words corresponding to padded 3×3 filter coefficients that are duplicated in each lane. The duplication is needed because all of the lanes only can read their own local memory. The second stream consists of the input pixels to be convolved and the third consists of the constant what is calculated before. Similarly, the output stream stores the convolved output of 48 rows of the image. Two intermediate streams store the partial sums and one for the addition of the constant. The simulation needs to be running to the specified number of iteration.

To make use of zero- and first-order nonlinear templates, the nonlinear functions which belongs to the templates should be stored in a Look Up Table (LUT), so in nonlinear template simulation is needed to store in each LRF this LUT and it is resulted a new input stream. This type of nonlinearities should be partitioned into segments. The parameters of the nonlinear function and the boundary points should be stored in LUT for each nonlinear elements. In case of the first-order nonlinear template, additional computation is required, because after the identification of the right interval of nonlinearity, the difference should be multiplied and added to the constant[15].

VI. PERFORMANCE COMPARISONS AND CONCLUSION

To measure the performance of the linear simulation a 256×256 sized input cell array was used and 10 forward Euler iterations with 0.1 timestep were computed, using a diffusion

TABLE II
COMPARISON OF DIFFERENT CNN IMPLEMENTATIONS: CORE 2 DUO T7200@2GHZ PROCESSOR, Q-EYE ANALOG VLSI CHIP, FALCON EMULATED DIGITAL CNN RUNNING ON XILINX VIRTEX-5 FPGA(XC5VSX95T)@550MHZ 1PE (MAX. 71 PE), CELL PROCESSORS (8 SPEs)@3.2GHZ, GRAPHICS PROCESSING UNIT (GPU) NVIDIA 8800GTX@1350MHZ AND STORM-1 STREAM PROCESSOR@500MHZ

Parameters	CNN Implementations					
	Core 2 Duo[11]	Q-Eye[12]	FPGA[13][14]	CELL (8 SPEs)[15]	GPU[16]	Stream
Execution time of linear template (μ s)	4092.6	250	737.2	111.8	3030	986
Execution time of nonlinear template (μ s)	84691.4	-	737.2	197.33	-	2624
Power (W)	65	0.1	20	85	160	25
Area (mm ²)	143	25	~ 389	253	334	~ 196
Silicon technology (nm)	65	180	65	90	90	130
Input precision (bit)	32	8	32	32	16	8
Weights precision (bit)	32	32	32	32	32	16

(CNN cell array size: 176×144, 10 forward Euler iterations for linear template and 16 forward iterations for non-linear template)

template (as shown in Figure 3). For nonlinear simulation as well a 256×256 sized input cell array was used, but a 16 forward Euler global finder template was computed. By using the MIPSsim simulator detailed statistics can be obtained about the operation of DSP SMIPS while executing a stream program. Comparison of the performance of the stream processor to a high performance microprocessor showed that about 4 times speedup can be achieved for linear template running. When using nonlinear templates the performance of the stream is higher, in this configuration 32 times speedup can be achieved. A single layer CNN model was successfully implemented on a VLIW architecture based stream processor with 80 ALUs + two SMIPS processors (SP16-G160). Using this kernel both linear and nonlinear CNN arrays can be simulated. The computational speed, dissipated power, silicon area, silicon technology and CNN model precision of the different array processors were compared (see Table II). The proposed solution is a low cost one. Based on this knowledge we are looking for special class of problems, which fits to the architecture.

REFERENCES

- [1] W. F. Lee, *VLIW Microprocessor Hardware Design (For ASIC and FPGA)*. New York: MC Graw Hill, 2008.
- [2] P. Szolgay, "Emulated digital CNN-UM on different kind of array processors," *Proceedings of the 11th International Workshop on Cellular Neural Networks and their Applications(CNNA2008)*, pp. 154–156, 2008.
- [3] T. Roska and L. O. Chua, "The CNN Universal Machine: an Analogic Array Computer," *IEEE Transaction on Circuits and Systems-II*, vol. 40, pp. 163–173, 1993.
- [4] "Cellular Wave Computing Library," [Online] <http://cnn-technology.itk.ppke.hu/>.
- [5] T. Roska and L. O. Chua, "Cellular Neural networks with nonlinear and delay-type template elements and non-uniform grids," *International Journal of Circuit Theory and Applications*, vol. 20, pp. 469–481, 1992.
- [6] *Storm-1 SP16-G160 Stream Processor Data Sheet*, Stream Processors, Inc., 2007.
- [7] "Stream processing: Enabling the new generation of easy to use, high-performance dsps," White Paper, Stream Processors, Inc., June 2008.
- [8] U. Kapasi, S. Rixner, W. J. Dally, B. Khailany, J. H. Ahn, P. Mattson, and J. D. Owens, "Programmable Stream Processors," *IEEE Computer Society, Computer Magazine*, pp. 54–62, 2003.
- [9] *RapiDev Reference Manual*, Stream Processors, Inc., 2008.

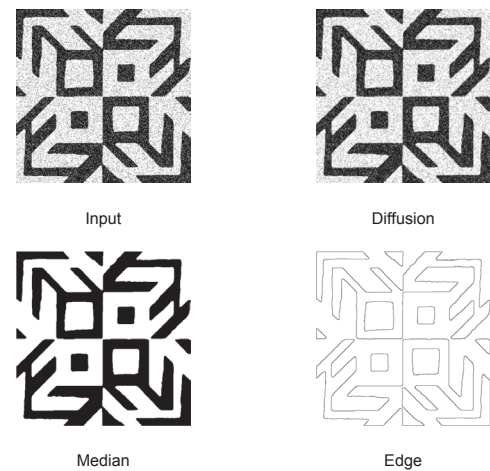


Fig. 3. Results of kernel simulations

- [10] Z. Kincses, Z. Nagy, and P. Szolgay, "Implementation of nonlinear template runner emulated digital CNN-UM on FPGA," *Proceedings of the 10th IEEE international workshop on cellular neural networks and their applications(CNNA 2006)*, pp. 186–190, 2006.
- [11] Z. Nagy, L. Kk, Z. Kincses, A. Kiss, and P. Szolgay, "Toward exploitation of cell multi-processor array in time-consuming applications by using CNN model," *International Journal of Circuit Theory and Applications*, vol. 36, pp. 605–622, 2008.
- [12] *eye-RIS v1.0/v2.0 Datasheet*, AnaFocus Ltd, [online] <http://www.anafocus.com>, 2004.
- [13] Z. Vörösházi, A. Kiss, Z. Nagy, and P. Szolgay, "Implementation of embedded emulated-digital CNN-UM global analogic programming unit on FPGA and its application," *International Journal of Circuit Theory and Applications*, vol. 36, pp. 589–603, 2008.
- [14] Z. Nagy and P. Szolgay, "Configurable Multi-layer CNN-UM Emulator on FPGA," *IEEE Transaction on Circuit and Systems I: Fundamental Theory and Applications*, vol. 50, pp. 774–778, 2003.
- [15] Z. Nagy, Z. Kincses, L. Kék, and P. Szolgay, "CNN Model on Cell Multiprocessor Array," *Proceedings of the European Conference on Circuit Theory and Design (ECCTD'2007)*, pp. 276–279, 2007.
- [16] B. G. Soós, A. Rák, J. Veres, and G. Cserey, "GPU boosted CNN simulator library for graphical flow based programmability," *EURASIP Journal on Advances in Signal Processing*, 2008.

Mach 3 Flow Simulation on IBM Cell Processor Based Emulated Digital Cellular Neural Networks

András Kiss
(Supervisor: Péter Szolgay)

Abstract—It is hard to get the solutions of partial differential equations (PDEs) fast and accurate for today's real world simulation. One exciting and important part of this area is the Computational Fluid Dynamics (CFD), which involves the problem of gas or fluid flow over different obstacles, e.g., air flow around vehicles, buildings, or the flow of water in the oceans. In engineering applications the temporal evolution of non-ideal, compressible fluids is quite often modeled by the system of Navier-Stokes equations. They are a coupled set of nonlinear hyperbolic partial differential equations and form a relatively simple, yet efficient model of compressible fluid dynamics. In the paper the implementation of a CFD solver on the Cell Broadband Engine (Cell BE) based Emulated Digital Cellular Neural Networks is described. The kernel is optimized according to the special requirements of the Cell BE and may implement on a set of QS22 racks. Our solutions performance is measured and the speed, power, area parameters are compared with different hardware implementations.

I. INTRODUCTION

In the paper I applied the finite volume Lax-Friedrich scheme for solving 2D Euler equations over uniformly spaced rectangular meshes. However, most real life applications of CFD require handling more complex geometries, bounded by curved surfaces. A popular and often an efficient solution to this problem is to perform the computation over non-uniform, logically structured grids. Technically, this idea can be exploited either by employing body fitted grids or by performing the computation in a curvilinear coordinate frame following the curvature of the boundaries. However, the standard 2D scheme over Cartesian geometry can be put to work.

Performance of the general purpose computing systems is usually improved by increasing the clock frequency and adding more processor cores. However, to achieve very high operating frequency very deep pipeline is required, which cannot be utilized in every clock cycle due to data and control dependencies. If an array of processor cores is used, the memory system should handle several concurrent memory accesses, which requires large cache memory and complex control logic. In addition, applications rarely occupy all of the available integer and floating point execution units fully.

Array processing to increase the computing power by using parallel computation can be a good candidate to solve architectural problems (distribution of control signals on a chip). Huge computing power is a requirement if we want to solve complex tasks and optimize to dissipated power and area at the same time.

In this work the topographic IBM Cell heterogeneous array processor architecture was used, because it has high peak

computing performance and supports double-precision floating point numbers. Additionally the Cell Software Development Kit is open source. Cell based clusters are providing a scalable environment to increase computing performance using standard multiprocessing libraries such as OpenMP or MPI. On the other hand development time of an optimized software solution is much shorter than designing a reconfigurable architecture [1] [2], however its computational efficiency is smaller in terms of area and power.

II. CELL PROCESSOR ARCHITECTURE

A. Cell Processor Chip

The Cell Broadband Engine Architecture (CBEA) [3] is designed to achieve high computing performance with better area/performance and power/performance ratios than the conventional multi-core architectures. The CBEA defines a heterogeneous multi-processor architecture where general purpose processors called Power Processor Elements (PPE) and SIMD processors called Synergistic Processor Elements (SPE) are connected via a high speed on-chip coherent bus called Element Interconnect Bus (EIB). The first implementation of the CBEA is the Cell Broadband Engine (Cell BE or informally Cell) designed for the Sony Playstation 3 game console, and it contains 1 PPE and 8 SPEs. The block diagram of the Cell is shown in Figure 1.

The PPE is a conventional dual-threaded 64bit PowerPC processor which can run existing operating systems without modification and can control the operation of the SPEs. The EIB is not a bus as suggested by its name but a ring network which contains 4 unidirectional rings where two rings run counter to the direction of the other two providing 200GB/s bandwidth between the elements. The dual-channel Rambus XDR memory interface provides very high 25.6GB/s memory bandwidth. I/O devices can be accessed via two Rambus FlexIO interfaces where one of them (the Broadband Interface (BIF)) is coherent and makes it possible to connect two Cell processors directly.

The SPEs (Figure 2) are SIMD (Single Instruction Multiple Data) only processors which are designed to handle streaming data. Therefore they do not perform well in general purpose applications and cannot run operating systems. Data for the instructions are provided by the very large 128 element register file where each register is 16byte wide. Therefore SIMD instructions of the SPE work on 16byte-wide vectors, for example, four single precision floating point numbers or eight 16bit integers. The SPEs can only address their local 256KB

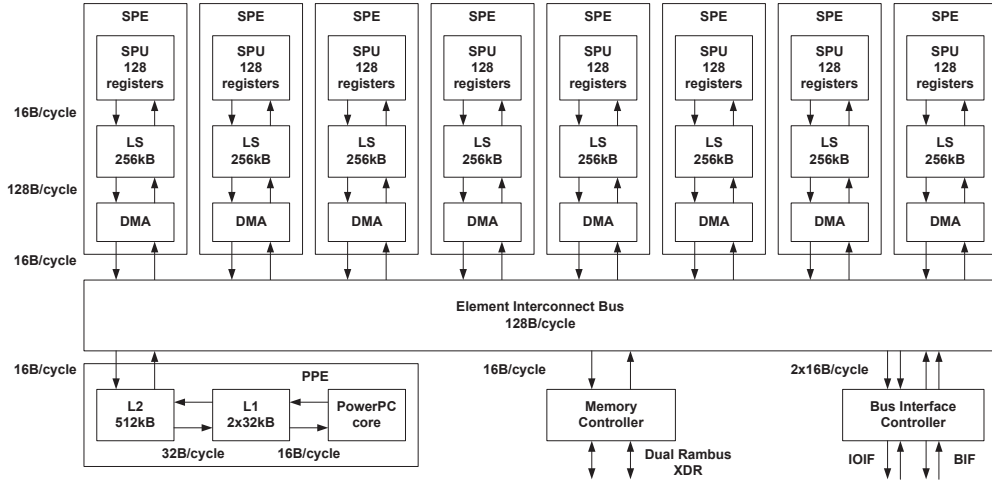


Fig. 1. Block diagram of the Cell processor

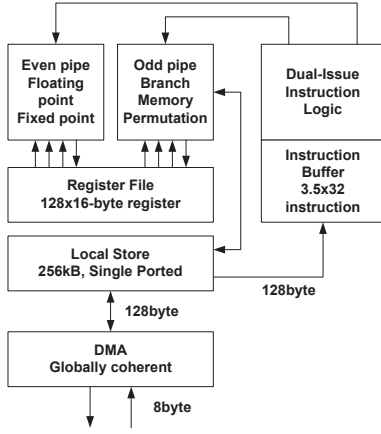


Fig. 2. Block diagram of the SPE co-processor

III. FLUID FLOWS

A wide range of industrial processes and scientific phenomena involve gas or fluids flows over complex obstacles, e.g. air flow around vehicles, buildings, or the flow of water in the oceans. In engineering applications the temporal evolution of non-ideal, compressible fluids is quite often modeled by the system of Navier-Stokes equations. It is based on the fundamental laws of mass-, momentum- and energy conservation, extended by the dissipative effects of viscosity, diffusion and heat conduction. By neglecting all these non-ideal processes and assuming adiabatic variations, we obtain the Euler equations [4], [5], describing the dynamics of dissipation-free, inviscid, compressible fluids. They are a coupled set of nonlinear hyperbolic partial differential equations, in conservative form expressed as

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \quad (1)$$

$$\frac{\partial (\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v} \mathbf{v} + \hat{I} p) = 0 \quad (2)$$

$$\frac{\partial E}{\partial t} + \nabla \cdot ((E + p) \mathbf{v}) = 0 \quad (3)$$

where t denotes time, ∇ is the Nabla operator, ρ is the density, u , v are the x - and y -component of the velocity vector \mathbf{v} , respectively, p is the pressure of the fluid, \hat{I} is the identity matrix, and E is the total energy density defined as

$$E = \frac{p}{\gamma - 1} + \frac{1}{2} \rho \mathbf{v} \cdot \mathbf{v}. \quad (4)$$

where γ is the ratio of specific heats. It is convenient to merge (1), (2) and (3) into hyperbolic conservation law form in terms of $U = [\rho, \rho u, \rho v, E]$ and the flux tensor

$$\mathbf{F} = \begin{pmatrix} \rho \mathbf{v} \\ \rho \mathbf{v} \mathbf{v} + \hat{I} p \\ (E + p) \mathbf{v} \end{pmatrix} \quad (5)$$

SRAM memory but they can access the main memory of the system by DMA instructions. The DMA engine is part of the globally coherent memory address space but we must note that the local store of the SPE is not coherent.

B. Cell Blade Systems

The IBM Cell blade systems are the main building blocks of the world's fastest supercomputer at Los Alamos National Laboratory which first break through the "petaflop barrier" of 1,000 trillion operations per second. These blades are built up from two Cell processor chips interconnected with a broadband interface. They offer extreme performance to accelerate compute-intensive tasks. The third generation blade system is the IBM Blade Center QS22 equipped with new generation PowerXCell 8i processors manufactured using 65nm technology. Double precision performance of the SPEs are significantly improved providing extraordinary computing density – up to 6.4 TFLOPS single precision and up to 3.0 TFLOPS double precision in a single Blade Center house.

as:

$$\frac{\partial U}{\partial t} + \nabla \cdot \mathbf{F} = 0. \quad (6)$$

IV. DISCRETIZATION OF THE GOVERNING EQUATIONS

For the sake of simplicity, in this paper we only consider rectangular computational domains. Following the finite volume methodology, we store all components of the volume averaged state vector $U_{i,j}$ at the mass center of cell (i,j) . The simplest algorithm we consider is first-order both in space and time [1] [2]. Application of the finite volume discretization method leads to the following semi-discrete form of governing equations (5)

$$\frac{dU_{i,j}}{dt} = -\frac{1}{V_{i,j}} \sum_f \mathbf{F}_f \cdot \mathbf{n}_f, \quad (7)$$

where the summation is meant for all four faces of cell (i,j) , \mathbf{F}_f is the flux tensor evaluated at face f and \mathbf{n}_f is the outward pointing normal vector of face f scaled by the length of the face. Let us consider face f in a coordinate frame attached to the face, such, that its x -axes is normal to f . Face f separates cell L (left) and cell R (right). In this case the $\mathbf{F}_f \cdot \mathbf{n}_f$ scalar product equals to the x -component of $\mathbf{F}(F_x)$ multiplied by the area of the face. In order to stabilize the solution procedure, artificial dissipation has to be introduced into the scheme. According to the standard procedure, this is achieved by replacing the physical flux tensor by the numerical flux function F^N containing the dissipative stabilization term. A finite volume scheme is characterized by the evaluation of F^N , which is the function of both U_L and U_R . In this paper we employ the simple and robust Lax-Friedrichs numerical flux function defined as

$$F^N = \frac{F_L + F_R}{2} - (|\bar{u}| + \bar{c}) \frac{U_R - U_L}{2}. \quad (8)$$

In the last equation c is the local speed of sound, $F_L = F_x(U_L)$ and $F_R = F_x(U_R)$ and bar labels speeds computed at the following averaged state

$$\bar{U} = \frac{U_L + U_R}{2}. \quad (9)$$

The last step concludes the spatial discretization. Finally, the temporal derivative is discretized by the first-order forward Euler method:

$$\frac{dU_{i,j}}{dt} = \frac{U_{i,j}^{n+1} - U_{i,j}^n}{\Delta t}, \quad (10)$$

where $U_{i,j}^n$ is the known value of the state vector at time level n , $U_{i,j}^{n+1}$ is the unknown value of the state vector at time level $n+1$, and Δt is the time step.

A vast amount of experience has shown that these equations provide a stable discretization of the governing equations if the time step obeys the following Courant–Friedrichs–Lewy condition (CFL condition):

$$\Delta t \leq \min_{(i,j) \in ([1,M] \times [1,N])} \frac{\min(\Delta x, \Delta y)}{|u_{i,j}| + c_{i,j}}. \quad (11)$$

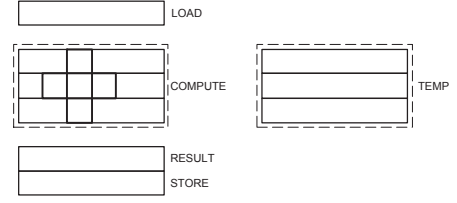


Fig. 3. Local store buffers

V. IMPLEMENTATION

By using the previously described discretization method a C based CFD solver is developed which is optimized for the SPEs of the Cell architecture.

Since the relatively small local memory of the SPEs does not allow to store all the required data, an efficient buffering method is required to save memory bandwidth. In our solution a belt of 9 rows is stored in the local memory from the array: 3 rows are required to form the local neighborhood of the currently processed row, one line is required for data synchronization, and two lines are required to allow overlap of the computation and communication as shown in Figure 3. Additionally 3 rows are required to temporarily store the primitive variables (u, v, p) computed from the conservative variables $(\rho, \rho u, \rho v, E)$. During implementation the environment of the CNN simulation kernel was used [6]. Template operations are optimized according to the discretized equations (6) to improve performance. The optimized kernel requires about 32KB memory from the local store of the SPE leaving approximately 224KB for the row buffers. Therefore the length of the buffer is maximum 1430 grid points while the number of rows is only limited by the size of the main memory.

To utilize the power of the Cell architecture computation work should be distributed between the SPEs. In spite of the large memory bandwidth of the architecture the memory bus can be easily saturated. Therefore an appropriate arrangement of data between SPEs can greatly improve computing performance. One possible solution is to distribute grid data between the SPEs. In this case each SPE is working on a narrow horizontal slice of the grid as shown in Figure 4(a). Communication between the SPE is required only during the computation of the first and last row of the slice (gray areas), which can be efficiently carried out by a single DMA transaction.

However the above data arrangement is well suited for the architecture of the array processors and simplifies the inter-processor communication, the SPEs are accessing main memory in parallel which might require very high memory bandwidth. If few instructions are executed on large data sets then memory system is saturated resulting in low performance. One possible solution for this problem is to form a pipeline using the SPEs to compute several iterations in parallel as shown in Figure 4(b). In this case continuous data flow and synchronization is required between the neighboring SPEs but this communication pattern is well suited for the ring

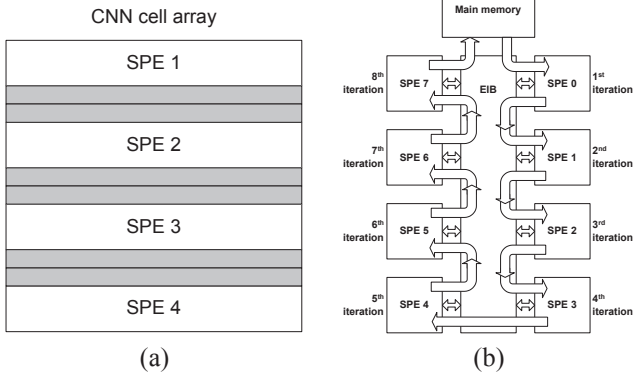


Fig. 4. Data distribution between SPEs, (a) slicing, (b) pipeline

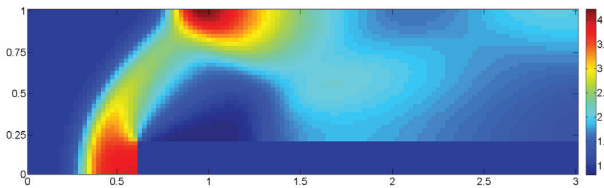


Fig. 5. First-order solution of the Mach 3 flow on a 128×512 array

structure of the EIB. Static timing analysis of the optimized CFD solver kernel showed that a grid point can be updated in approximately 200 clock cycles. Each update requires movement of 36byte data (4x4byte conservative state value, 4x4byte updated state value, 1x4byte mask value) between the main memory and the local store of the SPE. The Cell processor is running on 3.2GHz clock frequency therefore in an ideal case the expected performance of the computation kernel using one SPE is 64 million update/s. The estimated memory bandwidth requirement is 2.3GByte/s which is less than $1/10^{th}$ of the available memory bandwidth. Therefore grid data can be distributed between the SPEs and each of them can work on its own slice in parallel without running into a memory bottleneck.

VI. RESULTS AND PERFORMANCE

To show the efficiency of our solution a complex test case was used, in which a Mach 3 flow over a forward facing step was computed. The simulated region is a two dimensional cut of a pipe which has closed at the upper and lower boundaries, while the left and right boundaries are open. The direction of the flow is from left to right and the speed of the flow at the left boundary is 3-time the speed of sound constantly. The solution contains shock waves reflecting from the closed boundaries. This problem was solved on a 128×512 sized grid with 10^{-4} s timestep. Result of the computation after 4s of simulation time is shown in Figure 5. Experimental results of the average computation time on different architectures are summarized on Table I.

Compared to a high performance microprocessor the Cell based solution is 33 times faster even using a single SPE dur-

TABLE I
COMPARISON OF DIFFERENT HARDWARE IMPLEMENTATIONS

	Implementations			
	FPGA SX240T	Intel Core2Duo	Cell Processor	
			1 SPE	8 SPEs
Clock Frequency (MHz)	500	2000	3200	3200
Million cell iteration/s	2500	1.3004	44.089	313.0319
Computation Time on 128×512 1 step (μ s)	13.11	25197.92	743.22	104.68
Computation Time on 128×512 65536 steps (s)	0.86	1651.37	48.71	6.86
Speedup	1922.448	1	33.90353	240.7151
Power Dissipation (W)	~ 30	65	85	85
Area (mm^2)	389	143	-	253

ing the computation. Utilizing all SPEs of the Cell architecture the computation can be carried out two orders of magnitude faster while the power dissipation of the architectures are in the same range.

VII. CONCLUSION

The governing equations of two dimensional compressible Newtonian flows were solved using the IBM Cell architecture. The first-order Lax-Friedrichs scheme was used during the solutions. The main advantage of this method over the forward Euler method, which is used extensively in the computation of the CNN dynamics, is that this approximation is more robust in case of complex computational geometries and in presence of shock waves in the solutions.

The solution was optimized according to the special requirements of the Cell architecture. Performance comparison showed that about 34-time speedup can be achieved with respect to a high performance microprocessor in the single SPE solution, while the speedup is 240-time higher when all the 8 SPEs are utilized.

In the future I am planning to extend the solution to use multiple Cell processors and to use a more accurate second order method during the simulation.

REFERENCES

- [1] S. Kocsárdi, Z. Nagy, Á. Csík, and P. Szolgay, "Simulation of two-dimensional inviscid, adiabatic, compressible flows on emulated digital CNN-UM," *International Journal of Circuit Theory and Applications*, vol. DOI:10.1002/cta.565, 2008.
- [2] —, "Two-dimensional compressible flow simulation on emulated digital CNN-UM," in *Proc. IEEE 11th International Workshop on Cellular Neural Networks and their Applications (CNNA'08)*, Santiago de Compostella, Spain, July 2008, pp. 169–174.
- [3] J. A. Kahle, M. N. Day, H. P. Hofstee, C. R. Johns, T. R. Maeurerand, and D. Shippy, "Introduction to the Cell multiprocessor," *IBM Journal of Research and Development*, 2005.
- [4] J. D. Anderson, *Computational Fluid Dynamics - The Basics with Applications*. McGraw Hill, 1995.
- [5] T. J. Chung, *Computational Fluid Dynamics*. Cambridge University Press, 2002.
- [6] Z. Nagy, L. Kék, Z. Kincses, A. Kiss, and P. Szolgay, "Toward exploitation of Cell multi-processor array in time-consuming applications by using CNN model," *Int. J. of Circuit Theory and Applications*, vol. 36, no. 5-6, pp. 605–622, 2008.

Quasi Non-deterministic Turing Machine

László Tamás Kozák
 (Supervisors: Tamás Roska, Péter Földesy)
 kozla@digitus.itk.ppke.hu

Abstract— In this paper a procedure is presented where the time complexity of NP problems is transformed to space complexity by a parallel multiprocessor system. The size limitation of the device, where the architecture give result in polynomial time, and what we should do when the input set exceeds this limit are also addressed.

I. INTRODUCTION

We have known for a long time the relation between the problems, which can be solved in polynomial time (P), and the others, which can not be (non-polynomial, NP), is $P \subset NP$, but one of the most important questions of the mathematics whether $NP \subset P$ is also true. The most received theory derives from Mihalis Yannakakis who showed a certain approach to settling this question will never work out [1]. Mathematically the NP problems such problems that can be decided by a Non-deterministic Turing Machine (NTM) in polynomial time [2]. So theoretically we do not have to do other then make a NTM. In the kilo core systems' world we have the opportunity to make architecture similar to NTM. This paper gives a possible solution to make a quasi NTM which can be applied on general NP problems. It's worth to mention the device does not solve the P, NP question just give a general procedure to reduce the Time Complexity with more order of magnitude. The main property of the NP problems (or NTM) is that from one state we can proceed to more other states. Hence these problems can be represented by trees where the possible solutions or results are the leaves of these trees. Let us suppose that we have a device that is able to follow all branches. Precisely each processing unit follow different path and makes the proper operations step by step.

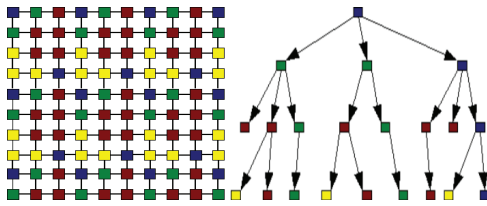


Fig. 1. Tree and multicore representation of an NP problem

The Fig. 1. shows how the processor units follow branches.

II. THEORY

When we try to create hardware architecture to solve an NP problem two cases can come up. There is a simple case when the size of the input set is small enough, the device does not run out from processing units (PUs) so it acts as a real NTM.

In the other case the input set and/or its complexity is too large to follow as many branches as it should.

Let us suppose there is a set with size M, a device contains P processors, additionally the maximum number of branches from one state is denoted by E.

With these parameters the time complexity of the problem according the first case is $O(M)$ but the space complexity is $O(M^E)$

The examination of the second case is much more difficult. If the graph representation of the problem is an N deep tree the following inequality is true.

$$\lfloor \log_E P \rfloor < N$$

When we reach the state where we do not have more PUs we must store the actual points and results of the tree and start a "Modified Depth First Search" (MDFS). The difference between the DFS and the MDFS is the MDFS [3] algorithm each $\lfloor \log_E P \rfloor$ steps generates additional P intermediate results which are denoted by different colors as Fig. 2. shows.

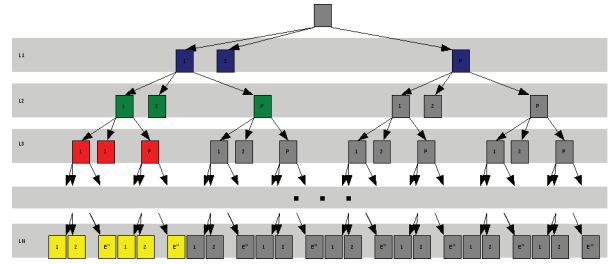


Fig. 2. Modified Depth First Search

The device reaches the yellow leaves (according to the second case) under $(S_{N-1} \cdot \lfloor \log_E P \rfloor + \lfloor \log_E E^M \rfloor)$ i.e. $(S_{N-1} \cdot \lfloor \log_E P \rfloor + M)$ steps, while we need $(P \cdot E^M)$ space. Where M is the remain part of the input set in the last step has less deepness then $(\lfloor \log_E P \rfloor)$. The space and time complexity in worst case are worked out as follows.

$$SC = P \cdot \frac{N}{\lfloor \log_E P \rfloor} + E^{(N \bmod \lfloor \log_E P \rfloor)}$$

Where the first part has to be stored in external memory while the second or last part contains the result(s).

$$TC = \lfloor \log_E P \rfloor + P \cdot \lfloor \log_E P \rfloor + P^2 \cdot \lfloor \log_E P \rfloor \dots \\ \dots + P^{S-1} \cdot \lfloor \log_E P \rfloor + P^S \cdot M$$

or simpler

$$TC = \lfloor \log_E P \rfloor \cdot \left(\sum_{i=0}^{S-1} P^i \right) + M \cdot P^S$$

where

$$S = \lfloor \frac{N}{\log_E P} \rfloor$$

The most critical part of the sum is the (P^S) .

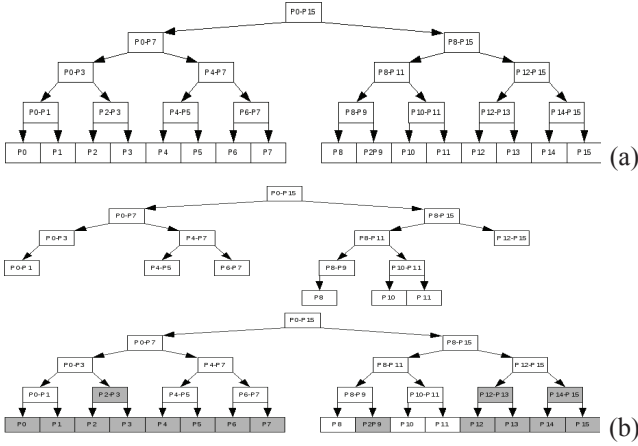


Fig. 3. Special case (Binary Tree) (a), General case (b)

Consequently if $P \Rightarrow \infty$ than $S \Rightarrow 0$, hence in case of $P_0 \ll P_1 \Rightarrow S_0 > S_1$ therefore $P_0^{S_0} > P_1^{S_1}$ so the larger the P the quicker the execution.

In Fig. 3.a. there is a special case where the graph is a binary tree ($E=2$) and the device has 16 PUs. It can be seen how the PUs follow the branches. However generally during the execution there are branches that certainly cannot perform good solution or no way to other points like the gray points on Fig. 3.b. In this case the PU, which deals with this branch, sends a signal to the MU and halts.

III. DEVICE ARCHITECTURE

All multicore system have a serious problem. How do the cores communicate each other? In most multi core architecture each core needs to know the state of every other cores. In some case this makes the design impossible. If we want the cores step forward simultaneously we have to reduce the communication radically between them.

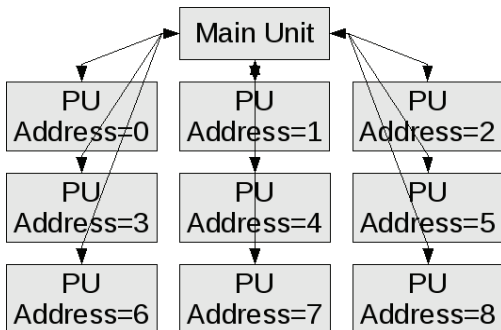


Fig. 4. Architecture (no connection between PUs)

A. Processing Unit

Each core roams a path depends on their core address and the parameters of the actual NP problem of course. So we must create a function which determines the next state from these given data. In this way the processors do not need to change their data so the PUs step forward in the tree without any information about each other. To do this the PUs need the same data. Since the memory, which can be accessed from more direction at the same time, does not exist therefore each PU needs own memory array with the same data. The function, calculates the next state, detailed later.

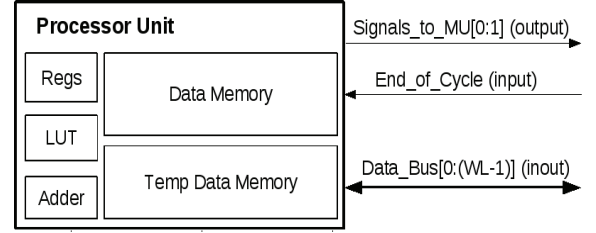


Fig. 5. PU block

1) *Algorithm*: Every PU stores the start point (regardless to how many devices is connected to cascade mode), the roamed points (this is done with the help of additional registers) and the actual result. If a PU comes to a halt sends a signal to the MU about this. A PU comes to a halt if:

- It didn't roam all points but got back to the start point again.
- It get back to an earlier roamed point again.
- There is no way from the actual point.

The exact data representation is discussed in the environment section. The PU memory is build up two main part. The first part contains the input data while the second is used to store the temporary data.

2) *The definition of the PATH() Function*:

$$S_i(x, y) = PATH(S_{i-1}(x, y), E, A)$$

where

- A is the physical address of the PU
- S(x,y) is the state $S(\lfloor A/E \rfloor, A \bmod E)$
- E is the maximum branch from a point

First of all we need to determine a deepness where all the branch can be roamed. This deepness, is denoted by C, must satisfy the $E^C \leq P = E^C + M$ inequality, where P is the number of PUs, E is the maximum branch from one point and M is a remain part which is not interesting. Since $0 \leq A \leq (P - 1)$ where A is the physical address of the PUs, $A = (E^C - 1) + M$.

The addresses of PUs are serial continuous and it can be written by the help of E, C and M.

$$A = x \cdot E + y + M$$

M is irrelevant, so

$$A = x \cdot E + y$$

where $0 \leq x \leq (E^{C-1} - 1)$ and $0 \leq y \leq (E - 1)$

C is the deepness or in other words the cycle number where the PUs are able to follow each branch so we need to use $\text{PATH}()$ function C times. The definition of $\text{PATH}()$ function is the next:

```

 $S_i(x, y) = \text{PATH}(E, A, S_{i-1}(x, y))$ 
begin
  if  $i = 1$  then
     $x := \lfloor A/E \rfloor$ ;
     $y := A \bmod E$ ;
  else
     $x := \lfloor (x + y)/E \rfloor$ ;
     $y := (x + y) \bmod E$ ;
  end if;
end

```

If $C > 1$ in the first cycle each branch is followed by more PUs. Each PU, which had the same y value (i.e. same branch), had different x value therefore $y := (x + y) \bmod E$ ensures that these PUs will follow different branches. The expression $x := \lfloor (x + y)/E \rfloor$ ensures that in the next cycle each y value will have different x value. An example is depicted in the next figure.

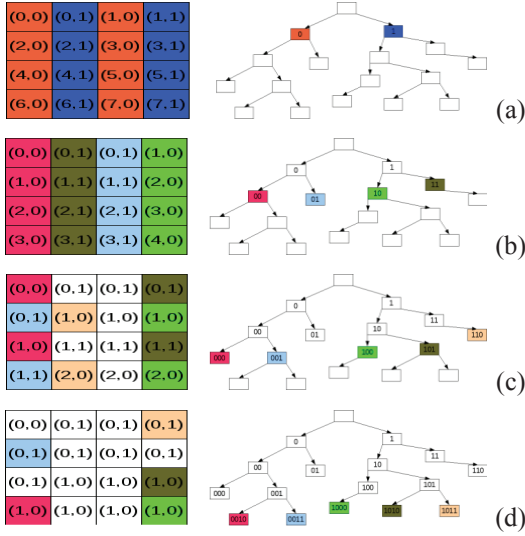


Fig. 6. Example: $P=16, E=2$ therefore $C=4$

In the example above there are some points where there are no paths to other points from. These PUs, which came to a halt, were depicted by white. Additionally it can be seen during the determination of the new state the PU makes a division. This is a very time or/and space demanding operation. To do this during one clock cycle the PUs build up a look-up-table contains the proper fraction and modulo values depends on E . This LUT is written into the beginning of temporary data memory of PUs at the beginning of the algorithm. The LUT stands for as many rows as many PUs are.

$\lfloor 0/E \rfloor$	$0 \bmod E$
$\lfloor i/E \rfloor$	$i \bmod E$
$\lfloor (P-1)/E \rfloor$	$(P-1) \bmod E$

Fig. 7. Look Up Table

3) *Temporary Data*: On the end of the actual cycle, the PUs must store their actual data in order to they can continue the process from the state where they stopped. Without this procedure they cannot return to that branches which has not processed yet.

Cycle number	Actual result
List of roamed points	

Fig. 8. Temporary data structure

Every two words mean a temporary cycle result. The first word means the roamed branches while the second contains the actual result. The size of first word is equal to the number of memory address bits.

B. Main Unit

The MU is an interface makes connection with other MUs of other devices or the PC. At the beginning of the algorithm it receives data from the PC and write it to the memory of PUs and at the end from PUs to the PC of course.

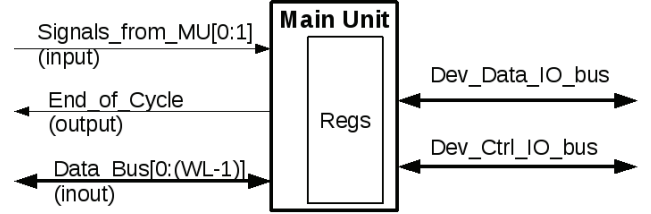


Fig. 9. MU block

At the end of every cycle, when the device runs out from PUs, the MU downloads their actual results and gives them new data. It has separate connection with each PU that is used to inform the MU which PU comes to a halt. Additionally the MU manages the data transport and the cascade communication respectively.

IV. ENVIRONMENT

The device needs an environment where the user can describe the input or/and the actual problem and which transform the input to processable form to the device. It must determine the parameters of the problem mentioned before, build up the data word by word, send it to the device and receive the result from it.

A. Memory Representation

Every graphs can be represented by a matrix so we just have to build up this matrix. After that we can make the proper data structure easily.

Data of POINT _{i-1}	
POINT _i	Branch number from this point
Identification of the connected points	Distans to the connected points
Data of POINT _{i+1}	

Fig. 10. Memory structure

V. CASCADE MODE

If the task or the application has bigger or/and more complex input more device can be connected to each other to reduce the processing time. In this operation mode the MUs choose a master MU (MMU) from themselves and this MMU gives the MUs a serial number. The MU with number 0 is the master and the others are the slave (SMU). The master manage the all process while the slaves deal only with the data transfer.

VI. CONCLUSION

Today's technology gives us opportunity to place millions of transistor into a single chip and hundreds of chips into a supercomputer. This theory can be implemented as a hardware device. So if we manage to connect more processing units into cascade mode for an NP problem we are able to solve bigger and bigger input sets of it however, since each PU has own memory with the same data, more PU means much more memory. This is one big disadvantage of it. The other essential problem of this architecture we must determinate the maximum number of path from one point but generally a point does not have as many neighbours. So that PUs, which would deal that branch, stops and they cannot go on.

REFERENCES

- [1] Mihalis Yannakakis: "Node- and Edge-Deletion NP-Complete Problems", Proc. 10th Annual ACM Symposium on Theory of Computing, pp. 253-264, San Diego, California, 1978..
- [2] Algorithms and Theory of Computation Handbook, CRC Press LLC, 1999, Nondeterministic Turing Machine, Paul E. Black, ed., U.S. National Institute of Standards and Technology. 21 November 2005.
- [3] Tom Bohman, Colin Cooper, Alan M. Frieze, Ryan Martin, Mikls Ruszink: On Randomly Generated Intersecting Hypergraphs, Electr. J. Comb. 10, 2003.
- [4] Kamrul Islam, Henk Meijer, Yarai Nnez, David Rappaport, Henry Xiao: Hamilton Circuit in Hexagonal Grid Graph, Queen's university (Canada), Roosevelt Academy, The Netherlands', 2007.

Heuristic Optimization with Processor Array Architecture

Zoltán Kárász

(Supervisor: Dr. Tamás Roska, Dr. Péter Földesy)

karzo@digitus.itk.ppke.hu

Abstract — Thereinafter I introduce a new processor array architecture, which can allocate the optimum of an energy functions with stochastic method in a large parameter space. This method based on the evaluation the function values in random locations, where every processor works with different part of the co-domain. After defined number of iteration, which depends on the task, the best domain re-divided between the processors used the WTA principle. I'll compare the results with the conventional solution like the Simulated Annealing algorithm on FPGA. In addition I show the latest questions in the area of the processor design, by-pass the different solutions including 3D-IC technology.

I. INTRODUCTION

The key issue of the microchip design in nowadays is not simply how we can accelerate the transistors. In the 80' years the scientist managed an exchange from the bipolar technique to the CMOS transistors, but in the last five years this technology reached its limit and there wasn't such important technology breakthrough as before. Therefore increasing the clock frequency and supply voltage cause heat dissipation, which can be handle only transistor scaling-down. Till that time in the Speed x Power x Area (SPA) trio we deal with only the speed, but now the others becomes also important. Solving different tasks could be allow different power and area:

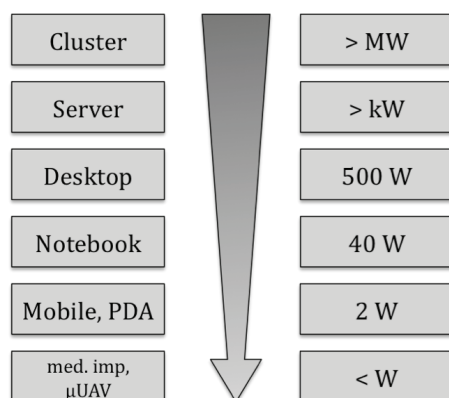


Figure 1 – Different architecture power consumption

The Fig.1 goes to show the importance of the ASIC (Application Specific Integrated Circuits) design, especially in the low power segment. Mainly in the industry were the design cost is negligible according to chip working parameters. The different applications have to fit for different criterions. Already in the architecture design phase, it's important to considering the bottleneck of the algorithm.

At these days good example the HP Oracle Database Machine which designed to act database transactions. I emphasize this example, because this is the first application specific mainframe computer. Above all, considering the signal propagation delay required a new design approach. As we started to use the latest CMOS technologies like 90 nm or 65 nm beyond the advantages we have to face the new challenges. Namely, the distance between the units which able to communicate to each other within one clock cycle is firmly limited (Fig.2).

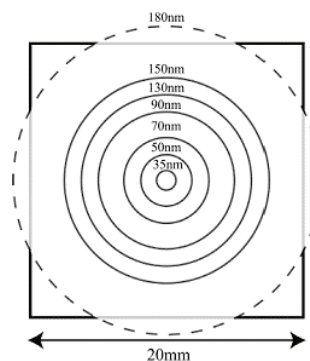


Figure 2 – Signal propagation under one clock

Other way to reserve this pace of development in near future is the 3D integration technologies. Basically the idea behind three-dimensional integration comes from the observation that only the surface of an integrated circuit is electrically active. Therefore thinning the silicon chips and stacking many of them one on top of the other would increase the circuit density. There are many methods to connect inter-chip, such as wire bonding, edge connect, capacitive or inductive coupling method as standard thinned chips in a single package (system in package), that connections from one chip to another are still obtained with standard bondings to the common package substrate. And direct contact with the most innovative technology that use through silicon vias (TSV). With opportunities also come risks, among which the most obvious is the thermal management of the stack.

Each 3D IC application imposes different requirements on the integration design and its process implementation. TSVs can be vary in size and the materials they pass through, they may be created at various points in the manufacturing sequence-in the frontend wafer fabrication, or in the assembly and packaging facility (before or after bonding). When viewed this way, the integration schemes under consideration can be classified as via-first or via-last, depending on when the vias are created.

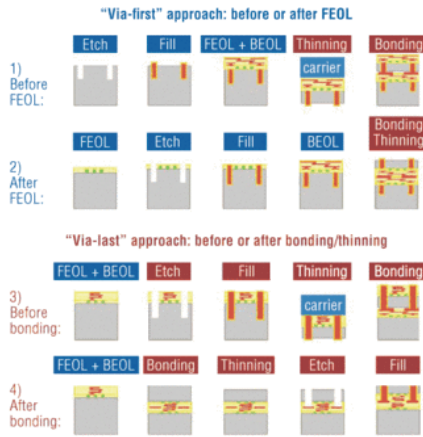


Figure 3 –Various TSV integration scheme

II. HEURISTIC OPTIMIZATION

In the real life many optimization problems that become unmanageable using combinatorial methods as the number of objects becomes large. A typical example is the travelling salesman problem, which belongs to the NP-complete class of problems. For these problems, there is a very effective practical algorithm called *simulated annealing* (thus named because it's alike the process undergone by misplaced atoms in a metal when its heated and then slowly cooled). While this technique is unlikely to find the optimum solution, it can often find a very good solution, even in the presence of noisy data.

The travelling salesman problem can be used as an example application of any optimization. In this problem, a salesman must visit some large number of cities while minimizing the total travelled distance. If the salesman starts with a random itinerary, he can then pair wise trade the order of visits to cities, hoping to reduce the mileage with each exchange. The difficulty with this approach is that while it rapidly finds a local minimum, it cannot get from there to the global minimum.

Simulated annealing improves this strategy through the introduction of two tricks. The first is called Metropolis algorithm, in which some trades that do not lower the mileage are accepted when they serve to allow the solver to explore more of the possible space of solutions. Such bad trades are allowed using the criterion that

$$e^{-\Delta D/T} > R(0,1),$$

where ΔD is the change of distance implied by the trade (negative for a good trade; positive for a bad trade), T is a temperature, and $R(0,1)$ is a random number in the interval $[0,1]$. D is called a cost function, and corresponds to the free energy in the case of annealing a metal. In which case the temperature parameter would actually be the kT , where k is Boltzmann's Constant and T is the physical temperature, in the Kelvin absolute temperature scale. If T is large, many bad trades are accepted, and a large part of solution space is accessed. Objects to be traded are generally chosen randomly, though more sophisticated techniques can be used.

The second trick is, again by analogy with annealing of a metal, to lower the temperature. After making many trades

and observing that the cost function declines only slowly, one lowers the temperature, and thus limits the size of allowed bad trades. After lowering the temperature several times to a low value, one may then cool the process by accepting only good trades in order to find the local minimum of the cost function. There are various annealing schedules for lowering the temperature, but the results are generally not very sensitive to the details.

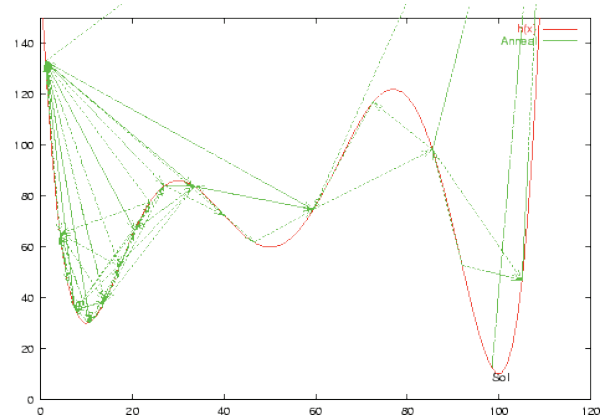


Figure 4 –Minimization with simulated annealing

There is another faster strategy called threshold acceptance. In this strategy, all good trades are accepted, as are any bad trades that raise the cost function by less than a fixed threshold. The threshold is then periodically lowered, just as the temperature is lowered in annealing. This eliminates exponentiation and random number generation in the Boltzmann criterion.

I laboured a simple algorithm by the pattern of above. In the first step, it divides the parametric space as much part as processors we want to use. In the next step the algorithm evaluate the function in random points. When it's found the local optimum in every part just choose the best, and re-divide the winner part of the space.

On the grounds of simplicity I made a simulator in MATLAB to compare the different heuristics algorithms like genetic algorithm, simulated annealing and my naive random algorithm, which is optimized for processor arrays. I found that, if I can set the correct learning parameters get same or better result then the others. The principal assumption for the algorithm it is have to calculate enough points in the parametric space that's depends of the frequency.

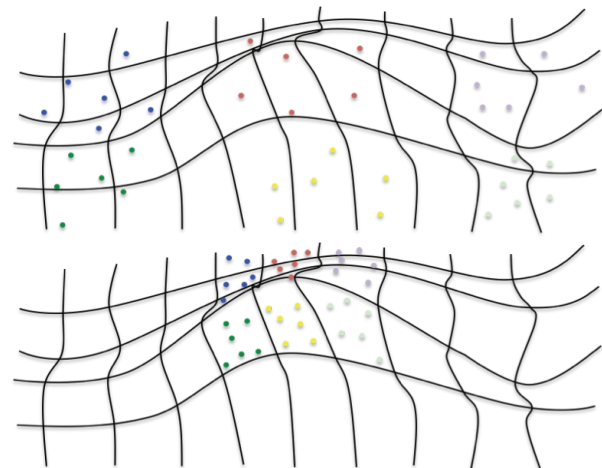


Figure 5 – Consecutive iteration steps

III. ARCHITECTURE

If minimised the communication between each module already in the design phase, and these modules capable to parallel processing the task, then the system will work with the smallest delays. The maximum number of the Processing Units (PU) determined by the size of the Arithmetic and Logical Unit (ALU). That affected by the complexity of the function. The OSU (Optimum Selector Unit) chooses the best PU after several thousand iteration, and divides its space between every processor.

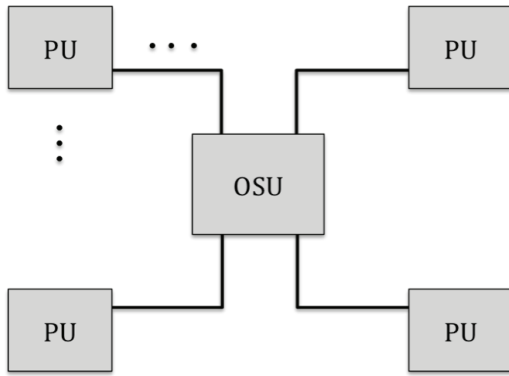


Figure 6 – Main architecture

Every PU contains a random number generator to define the next position, ALU for a function value calculation. And a comparator to collate the actual value with the best one, and stores it in every iteration cycle.

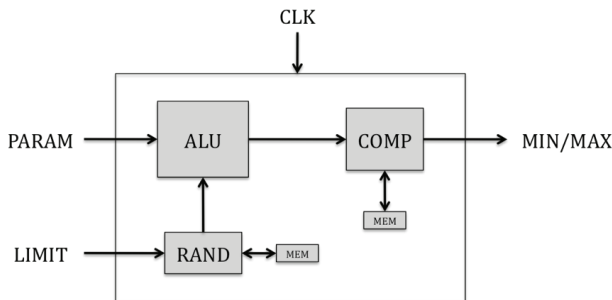


Figure 7 – Architecture of the PU

A random number generator is also necessary for the working. The basic principle behind is to extract the random from the jitter of the clock signal synthesized in the embedded analog PLL. The jitter is detected by the sampling of a reference (clock) signal using a rationally related (clock) signal synthesized in the on-chip analog PLL. The fundamental problem lies in the fact that the reference signal has to be sampled near the edges influenced by the jitter.

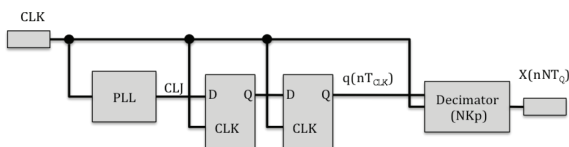


Figure 8 – Basic structure of random bitstream generator

Because there is a probability that the first flip-flop could become metastable, the second flip-flop is cascaded. In case

the first flip-flop produces a metastable output, it can resolve until its output is clocked by the second flip-flop. This flip-flops connection does not assure that only stable signal is clocked, but the probability that the output $q(nT_{CLK})$ will get a valid state is much higher.

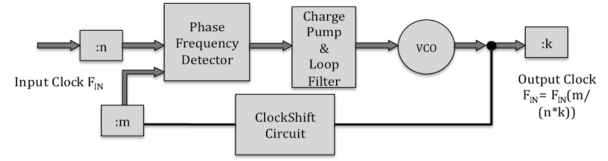


Figure 9 – Block diagram for an embedded PLL

The PLL block that can provide at least one synthesized clock signal with frequency:

$$F_{OUT} = F_{IN} \frac{m}{n * k} = F_{IN} \frac{K_M}{K_D}$$

where F_{IN} is the frequency of the external input clock source. Reference-, feedback- and post-divider values n , m and k can vary from one to several hundreds in FPGAs or to several thousands in ASICs.

IV. ESTIMATION

I mentioned how important the Speed x Power x Area in nowadays. I did some estimation with CADENCE. This estimation supposes a standard size ALU for 20 bit wide words. I designed the ALU takes about 1000 gate and the whole PU includes 1500. The OSU takes about 2000, which depends on teaching possibilities in the OSU.

Array Size	I/O type	Freq [MHz]	Techn. [μm]	Die Area [mm ²]	Power [mW]
4x4	Wirebond	50	0.18	4.1	5
12x12	Stragger	300	0.13	6.1	630
32x32	FlipChip	300	0.065	12.9	3260

The proposed computation power with the smallest one is about 5 MEPS (mega evaluation per second) while largest one 1.5 TEPS.

Other relevant phase of the design, how we can solve with another existing equipment. I compare this data with the Xilinx Cool Runner-II series. It also works on same frequency range, but it has lower computational speed with one order of magnitude. Of course this series are generic and not ASIC, so the compare is not appropriate.

If the power consumption is becoming the more important it can be possible to use power saving techniques like subthreshold working or clock gating which means a trade-off between the speed and power.

V. DISCUSSION

One of the most important parts is the utilization area. This architecture has a very low power consumption rate with an acceptable computational power, or outstanding computation power with higher consumption. The difficulty how to define the concrete function class, but not too large because the ALU will be takes too much space. And it wouldn't be too small also, because then the whole chip is

useless. The main area supposed be in low power segment, where require to define an optimum of a system periodically and the parameters changes constantly. It seems to be the computer controlling robotics could use that in the μ UAV where the power consumption is primary. If this is a right way further, I can design the fitting ALU for the system, and run the desired simulations on a given technology.

REFERENCES

- [1] Dueck, G. and Scheuer, T. "Threshold Accepting: A General Purpose Optimization Algorithm Appearing Superior to Simulated Annealing." *J. Comp. Phys.* 90, 161-175, 1990
- [2] Ingber, L. "Simulated Annealing: Practice Versus Theory." *Math. Comput. Modelling* 18, 29-57, 1993
- [3] Kirkpatrick, S.; Gelatt, C. D.; and Vecchi, M. P. "Optimization by Simulated Annealing." *Science* 220, 671-680, 1983
- [4] Szecsi J, Fiegel M, Krafczyk S, Straube "A Smooth pedaling of the paraplegic cyclist – a natural optimality principle for adaptation of tricycle and stimulation to the rider", *J.Rehabil. Res. Dev.*, 41 Supp.2: 30, 2004
- [5] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M.; Teller, A. H.; and Teller, E. "Equation of State Calculations by Fast Computing Machines." *J. Chem. Phys.* 21, 1087-1092, 1953
- [6] Otten, R. H. J. M. and van Ginneken, L. P. P. P. *The Annealing Algorithm*. Boston, MA: Kluwer, 1989
- [7] Flynn P Carson, Young Cheol Kim, In Sang 3-D Stacked Package Technology and Trends IEEE Proceeding 2009 vol. 97 (3-D Integration Technologies) pp. 31-42
- [8] Makoto Motoyoshi Through-Silicon Via IEEE Proceeding 2009 vol. 97 (3-D Integration Technologies) pp. 43-48
- [9] Mitsumasa Koyanagi, Takafumi Fukushima, Tetsu Tanaka High-Density Through Silicon Vias for 3-D LSIs IEEE Proceeding 2009 vol. 97 (3-D Integration Technologies) pp. 49-59
- [10] Jian-Qiang Lu 3-D Hyperintegration and Packaging Technologies for Micro-Nano Systems IEEE Proceeding 2009 vol. 97 (3-D Integration Technologies) pp. 18-30
- [11] CoolRunner-II CPLD Family Datasheet 2008 Sept.
- [12] Raychowdhury, A. Xuanyao Fong Qikai Chen Roy, K. Analysis of Super Cut-off Transistors for Ultralow Power Digital Logic Circuits Low Power Electronics and Design, 2006. ISLPED'06. Proceedings of the 2006 International Symposium

A comparison of audio to visual speech conversions

Gergely Feldhoffer
(Supervisor: György Takács)
flug@digitus.itk.ppke.hu

I. INTRODUCTION

The audio to visual speech (ATVS) conversion targets to convert audio speech into visual speech. There are different concepts, some of them use automatic speech recognition (ASR) to extract phonetic information from the signal and on the phoneme string use some kind of visual coarticulation rule set or model in a range from simple viseme interpolation to phoneme-viseme cross influencing sophisticated models[1]. One of the main properties of ASR based solution is the possibility of using language models. There are semi-ASR based approaches also like [2] using phoneme level probabilities without language level. Other approaches among others use direct conversion between the modalities by a learning system[3], [4] without phoneme level.

Research laboratories develop solutions, and the evaluations of the solutions are intelligibility tests and/or opinion score tests. The results are independent from each other so it is hard to tell which approach is better than the other. Now we describe a comparative evaluation which is performed between different conversion approaches by keeping all the other components of the workflow to be the same. See Fig 1.

We also introduce a hybrid method of different concepts which performs well on subjective opinion tests.

There are different aspects of quality of ATVS systems. The best possible conversion regarding intelligibility makes lip-reading possible. We traditionally work with hearing impaired, so intelligibility will be tested in this term. The best possible conversion concerning naturalness makes output which can not be distinguished from a record of original facial motion. In our task we start from a natural acoustic speech signal, and for all speech qualities the best possible visual speech should be generated. In other words, translation between the modalities should be done, independently of the speech quality in terms of articulation, coding and noise ratio.

II. CONTESTANTS

Basically five kinds of approaches will be evaluated:

- a reference natural facial motion
- a direct conversion system
- an ASR based cartoon industry ad-hoc standard method
- a modular system of ASR and text based sophisticated visual coarticulation modeling
- and the hybrid of modular ATVS and direct conversion.

The frontend of the methods up to the ATVS conversion is common. The same voice data is processed. Two speakers are in the database. One of the speakers is used for training, the other is for testing.

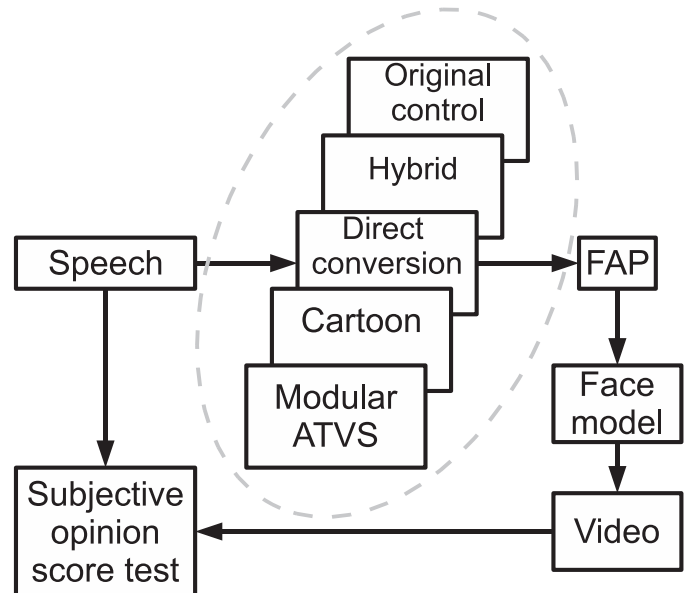


Fig. 1. Multiple conversion methods were tested in the same environment

The visualization of the output of the ATVS methods is also common. The results are represented by facial animation parameters (FAP). The FAP is a part of MPEG-4 standard. Each FAP data flow is animated on the same head model. There are better facial descriptors than MPEG-4 but our motion capture system could not give more detail than MPEG-4, so we used this widely popular system, and simplified the more sophisticated methods to this common space. MPEG-4 FAP is a facial animation coding standard, it represents normalized facial feature point displacements.

The videos used in the tests are created from FAP sequence by an Avisynth[5] 3D face renderer plugin, which provides a convenient way of testing control parameters and handling multiple videos from multiple sources with cross-referencing trimming intervals, which is the usual task of subjective test compilation.

A. Direct conversion

Our research group developed a direct conversion system[3] published in 2006. The direct conversion takes actual voice segment and by a machine learning component estimates the best articulated facial parameters for the voice. It does not use phoneme or viseme level nor language-dependent elements. The machine learning method is usually regression by examples of audio and video data pairs. See Fig 2.

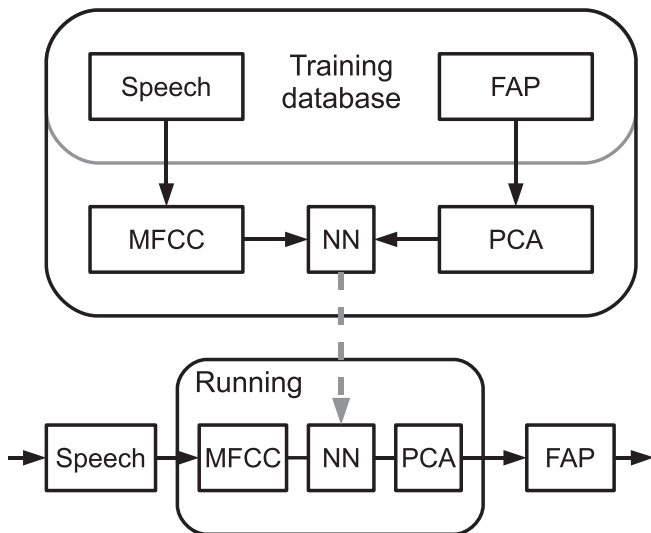


Fig. 2. Structure of direct conversion.

In our case a backpropagation neural network was trained between carefully chosen representations of the audio and the video data. MFCC was performed on the audio data for each frame of the database. The facial data was expressed with Principal Component Analysis (PCA) of the FAP. The neural network used 11 frame long window on the input side (5 frames to the past and 5 frames to the future), and 4 principal component weights on the output.

The trained neural network was used on the test set with a different speaker. Choosing training speaker is an important detail. Our speaker is a professional lip-speaker who works with deaf and hard of hearing people.

B. ASR based solutions

For the ASR based approaches we used the best available speech recognition system for Hungarian. This is a piece of work of Mihajlik et al. The system is capable to use language model or vocabulary, and during the test we used both informed and uninformed recognitions. Uninformed recognition uses only general properties of the language, informed recognition uses vocabulary with the words occur in the test material.

In the ASR, a standard frame synchronous Weighted Finite State Transducer - Hidden Markov-Model (WFST-HMM) decoder called as VOXerver [6] was applied to obtain the phonemic segmentation of input waveforms. MFCC based feature vectors were computed with delta and delta-delta components. Blind channel equalization was used in the cepstral domain to reduce linear distortions. Speaker independent cross-word decision-tree based triphone acoustic models were applied, trained previously on the MRBA Hungarian speech database [7].

In the uninformed ASR system, phoneme-bigram phonotactic model constrained the decoding process. The phoneme-bigram probabilities were estimated on the MRBA database. In the informed ASR system a zerogram word language model

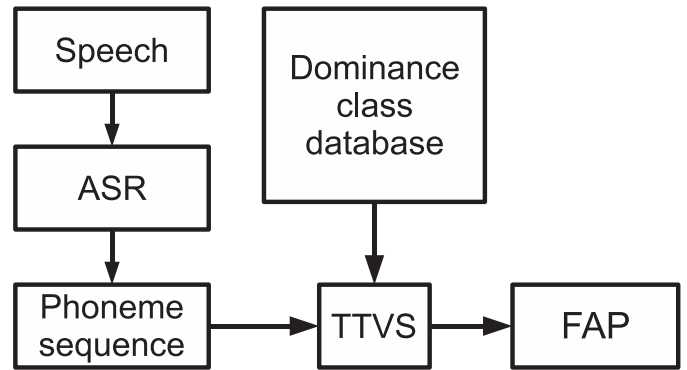


Fig. 3. Modular ATVS consists of an ASR subsystem and a text to visual speech subsystem.

was used with a vocabulary size of 120. Pronunciation of words were determined automatically as described in [8].

In both type of speech recognition approaches the WFST-HMM recognition network was constructed offline using the AT&T FSM toolkit [9]. In the case of the informed system, phoneme labels were projected to the output of the transducer instead of word labels.

The system gives 10 ms precision of segmentation and the most probable phoneme for each segment. This data will be used to create visual speech.

1) *Cartoon control*: The baseline solution is the animation industry ad-hoc standard phoneme-to-phoneme interpolation with directly linked visemes. This approach is particularly popular among cartoon animators since the viseme count can be taken into consideration, the variety of the used visemes is scalable. The animator uses a table to connect phonemes to visemes. In the best case all the visemes for the given language is present. The interpolation of the visual states is uniform in this case. The sophisticated visual blending is supported by the next contestant.

In this solution the viseme set was the full value set of the ASR, the sample visemes were extracted from the original facial motion and linear interpolation was used.

2) *Modular ATVS*: We call an ATVS system modular ATVS (MATVS) if it consists of a separable ASR subsystem and a phoneme string to visual speech synthesizer subsystem. This is a particularly popular approach, since ASR technologies are well developed, standalone trainable and testable. See Fig 3.

As of 2009 the best available ASR is combined with the best available text based visual coarticulation system for Hungarian by Czap et al called TTVS [10]. This is a text to visual speech conversion system, a part of a text to audiovisual system without the voice synthesizer component. The system's workflow consists of a text preprocessor, a phoneme-to-viseme mapping with phoneme neighborhood dependent effect ratio, filtering and other post-processing steps. We hijacked the system in the text preprocessor step by injecting readily time-aligned data.

TTVS features dominant, uncertain and mixed dominance

classes according to the level of influence by the neighborhood, and uses a database of mixed and uncertain class members' behavior in different neighborhoods.

The ASR system gives an output of phoneme strings with timing information. TTVS produces a high quality video from this data using Poser which was processed as an original recording, and FAPs were extracted. To test the methods we had to use the same virtual head, please note that some of the valuable information was lost during the conversion, so the test result may not show the real quality of the whole TTVS system, only the conversion part and only for those parameters which can be handled in our MPEG-4 subset.

C. Hybrid control

Results of different systems are in a common linear space, since all of them are represented in FAP. Therefore the average of different controls is an appropriate control also. We used an inverse amplitude-weighted mean of the output of the direct conversion and the output of uninformed modular ATVS. Weighting is to equalize the different articulation amplitudes in the result. The only synchronization between the control parameters is the common voice source.

III. TESTS AND RESULTS

We made subjective opinion score tests and lip reading test with the video material created by each of the methods and the original audiovisual recording. For modular ATVS tests we used both informed and uninformed recognition results, which are detailed below.

A. ASR subsystem

The quality of the recognition has two aspects. One of them is the precision of the assumed phoneme string. This is 100% at the informed run since the test set consists of small set of words as names of months or digits. The uninformed run falsely recognizes phonemes in 25.21% of the video frames. This may seem too high error ratio, but an ATVS using this input performs surprisingly well. The reason of this phenomenon may be the special confusion pattern which makes the error small if it is expressed with the resulting visual data. The speech recognizer confuses phonemes with visemes closer to each other more frequently than with others. The error expressed in relative viseme distance using quadratic metrics in FAP space is only 9.6% compared to random confusions for the whole data, or if we count only the falsely recognized frames it is still 52% of the random confusions.

The other point of view is the precision of the segmentation. This was a bit harder task to the speech recognizer system. The uninformed run was more precise on the average than the informed. This makes a very heavy impact on the subjective opinion scores.

B. Opinion scores

The 34 test subjects evaluated the naturalness of the mouth motion. One of the original facial motion driven face models was shown as one of the bests and one of the most unaligned

TABLE I
RESULTS OF OPINION SCORES, FIRST ROUND

Method	Average score
Uninformed MATVS	3.82
Original facial motion	3.79
Hybrid	3.72
Cartoon control	3.17
Direct conversion	3.02
Informed MATVS	2.85

TABLE II
RESULTS OF OPINION SCORES, SECOND ROUND

Method	Average score
Original facial motion	3.73
Direct conversion	3.58
Hybrid	3.48
MATVS	3.43
Hybrid-2	2.97
Cartoon control	2.73
MATVS-2	2.67

recognition based linearly interpolated motion as one of the worst. The test subjects were instructed to give scores between 1 and 5 according to the presented videos.

The test material consisted of 7 videos of 5 methods. Another 5 videos synthesized from original facial motion recordings were added. The videos were presented in random order. Each video contained 2-4 separated words, started and ended in closed mouth state. The voice source of the direct conversion and the hybrid method was recorded with a person whose voice is not included in the training set of the neural network.

The results (Table I) are unexpected, the informed ASR based sophisticated visual coarticulation system was expected to be one of the bests. Closer investigation on the lower scores shows that the audiovisual synchrony of the informed recognition is worse than to uninformed.

The results show two groups. The lower group around 3.0 points is the time alignment problem oppressed modular ATVS, the direct conversion and the linear interpolation of correctly segmented ASR results. The higher group around 3.7 points is the uninformed but correctly segmented modular ATVS, the original facial motion and the hybrid control.

We made a second round for the measurements because of the unexpected results. We decreased the articulatory amplitude of the direct conversion which was used for deaf people and has bigger opened mouth than the average. As table I shows it is irrelevant, whether the ASR used a vocabulary or not, in the point of view of naturalness the impact of segmentation precision is much stronger. Therefore in the second round we do not use the denomination "informed". Let us call the MATVS with segmentation errors "MATVS-2" and also introducing another hybrid, now with MATVS-2, marked as "Hybrid-2".

The second test was done with 58 test subjects.

The results of the second round (Table II) confirm the theory of the beneficial properties of hybridization. The improvement of the subjective opinion score average between MATVS-2

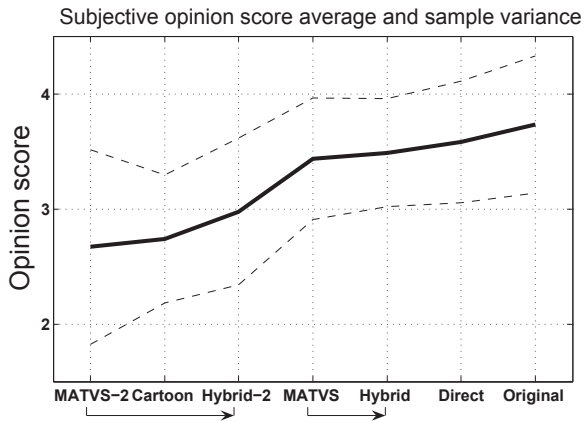


Fig. 4. Direct conversion is the closest to the original recording in the aspect of naturalness. Modular ATVS systems are vulnerable to synchronization errors, hybridization can help this issue.

TABLE III
RESULTS OF RECOGNITION TESTS

Method	Precision
Informed MATVS	61%
Uninformed MATVS	57%
Original motion	53%
Cartoon	44%
Hybrid	42%
Direct conversion	36%

and Hybrid-2 is significant with $p = 0.00026$ according to two-sample t-test. The advantage of direct conversion against MATVS is on the edge of significance with $p = 0.0512$ as well as the difference between the original speech and the direct conversion with $p = 0.06$ but MATVS is significantly worse than original speech with $p = 0.00029$. The results compared to the first round also show that in aspect of naturalness the excessive articulation is not eligible.

C. Intelligibility

Intelligibility was measured with a test of recognition of video sequences without sound. This is not the popular modified rhyme test but for our purposes with hearing impaired is more relevant. The 58 test subjects had to guess which word was said from a given set of 5 other words. The sets were numbers, names of months and the days of the week. All the words were said twice. The sets were intervals to eliminate the memory test from the task. This task models the situation of hearing impaired or very noisy environment where an ATVS system can be used to. It is assumed that the context is known, so the keyword spotting is the closest task to the problem.

The performance of the conversion methods changes to reverse in this task compared to naturalness. The main result here is the dominance of ASR based approaches (Table III), and the insignificance of the difference between informed and uninformed MATVS results ($p = 0.43$) in this test which may deserve further investigation. Note that as the synchrony is not an issue without voice, the informed MATVS achieves best results.

IV. CONCLUSION

A comparative study was proposed. Our direct conversion system was not compared to conceptually different conversion solutions before.

In the subjective tests we have the following definite results. We observed higher importance of the synchrony over phoneme precision in ASR based ATVS systems. There are publications on the high impact of correct timing in different aspects [10], [11], [12], but our result show explicitly that more accurate timing achieve much better subjective evaluation than more accurate phoneme sequence.

Also, we have shown that in the aspect of subjective evaluation, direct conversion is a method which produces the highest opinion score of 95.9% of an original facial motion recording with less computational complexity than ASR based solutions. We showed that hybridization can be used as a technique to significantly improve naturalness of segmentation problem oppressed ASR based ATVS systems. See Fig 4.

For tasks where intelligibility is important (support for hearing impaired, visual information in noisy environment) modular ATVS is the best approach among the presented. Our mission of aiding hearing impaired people call upon us to consider using ASR based components. Hybridization worse intelligibility significantly, so only entertaining applications should use it. For naturalness (animation, entertaining applications) direct conversion and hybridization is a good choice.

REFERENCES

- [1] J. Beskow, I. Karlsson, J. Kewley, and G. Salvi. Synface - a talking head telephone for the hearing-impaired. *Computers Helping People with Special Needs*, pages 1178–1186, 2004.
- [2] S. Al Moubayed, M. De Smet, and H. Van Hamme. Lip synchronization: from phone lattice to pca eigen-projections using neural networks. In *Proceedings of Interspeech 2008*, Brisbane, Australia, sep 2008.
- [3] György Takács, Attila Tihanyi, Tamás Bárdi, Gergely Feldhoffer, and Bálint Srancsik. Speech to facial animation conversion for deaf customers. In *4th European Signal Processing Conf.*, Florence, Italy, 2006.
- [4] Gregor Hofer, Junichi Yamagishi, and Hiroshi Shimodaira. Speech-driven lip motion generation with a trajectory hmm. In *Proc. Interspeech 2008*, pages 2314–2317, Brisbane, Australia, 2008.
- [5] <http://avisynth.org>.
- [6] Péter Mihajlik, Zoltán Tüske, Balázs Tarján, Botyán Németh, and Tibor Fegyó. Improved recognition of spontaneous hungarian speech morphological and acoustic modeling techniques for a less resourced task. *IEEE Transaction on Audio Speech and Language Processing (submitted)*, 2008.
- [7] <http://alpha.tmit.bme.hu/speech/hdbMRBA.php>.
- [8] P. Mihajlik, T. Révész, and P. Tatai. Phonetic transcription in automatic speech recognition. *ACTA LINGUISTICA HUNGARICA*, pages pp. 407–425, 2003.
- [9] Mehryar Mohri, Fernando C. N. Pereira, and Michael Riley. Weighted finite-state transducers in speech recognition. *Computer Speech and Language*, pages pp. 69–88, 2002.
- [10] L. Czap and J. Mátyás. Virtual speaker. *Hiradätechnika Selected Papers*, Vol LX/6:pp. 2–5, 2005.
- [11] Gérard Bailly, Oxana Govokhina, Gaspard Breton, and Frédéric Elisei. A trainable trajectory formation model td-hmm parameterized for the lips 2008 challenge. In *Proceedings of Interspeech 2008*, Brisbane, Australia, sep 2008.
- [12] Gergely Feldhoffer, Tamás Bárdi, György Takács, and Attila Tihanyi. Temporal asymmetry in relations of acoustic and visual features of speech. In *15th European Signal Processing Conf.*, Poznan, Poland, 2007.

Automatically Creating a Frequency Dictionary of Verb Phrase Constructions

Bálint Sass
(Supervisor: Gábor Prószték)
sass.balint@itk.ppke.hu

Abstract—The way of dictionary creation is changing. There are large corpora and specific algorithms to deal with them, and most of the process is becoming more and more automatic. We show that a monolingual dictionary of verb phrase constructions can be created fully automatically. Meaning of constructions are not defined here, they are represented by a suitable corpus example instead. The methodology is not just automatic but also language independent: it is built on a simple, unified language independent corpus representation. A general algorithm works with this representation and collects all verb phrase constructions. The resulting dictionary can be useful as a learners' dictionary and also as a rich lexical resource in different natural language processing tasks.

I. INTRODUCTION

We always need new dictionaries to keep up with the constantly changing vocabulary of languages. Two key factors of modern data-driven lexicography are: large corpora and natural language processing algorithms which are able to extract lexicographically relevant information from them. Algorithms which are language independent are still more important, they give a unified solution for dictionary building. Using them we can build specific dictionaries *dynamically* that means by loading our specific corpus (i.e. in a foreign language or with special terminology) into the dictionary creation system and waiting for the output. Research in this field can turn the traditionally very labour-intensive process of dictionary creation to fully automatic in the next decade. This automatic process will be more flexible and rapid keeping the dictionaries more up-to-date than they were in the past.

Present-day dictionaries are in fact electronic lexical databases. They can be transformed into a classical dictionary (even into printed form) on the one hand, but can also be used as an important resource in several natural language processing applications on the other.

II. AIM OF RESEARCH

Our research fits into the research field outlined above, and this shows what kind of dictionary we originally planned to create.

The aim is to develop a language independent corpus-driven [1] methodology which is capable of building a dictionary of verb phrase constructions (VPCs) automatically.

We chose verb phrase constructions, because these constructions cover a substantial part of all natural language utterances. Actually every clause consists of a verb and some dependents of it. Collecting these constructions provides a detailed picture of the whole language. We obtain an almost error free result in a fully automatic way. If we want to publish the dictionary in printed format, some manual lexicographic work (error correction) is needed. We will see an example entry later in a stage before the manual work. We achieve language independency by creating a unified, language independent corpus representation from language dependent corpora and then running our single general dictionary building algorithm which takes this representation as input.

Before going into details let us see the whole research path “from raw text to dictionary” from a bird’s-eye view.

- 1) text
→ segmentation, morphological analysis and disambiguation
- 2) annotated corpus (Hungarian National Corpus)
→ clause boundary detection, shallow syntactic parsing
- 3) parsed corpus in unified representation
→ discovery/collecting of important/frequent verb phrase constructions
- 4) dictionary of verb phrase constructions

III. PREVIOUS WORK

As a first step we worked out the unified corpus representation, which uses dependency grammar as theoretical framework. Basic units are clauses consisting of one verb and its dependents. Dependents are represented by their head and their relationship to the verb. Towards this representation we need to detect clause boundaries, normalize verbs and identify dependent phrases by chunking. Using a large POS-tagged corpus (namely the 190 million word Hungarian National Corpus [2]) we obtain a large shallow parsed corpus that way.

1st prospective thesis. POS-tagged corpora can be reliably converted to our unified representation with a rule-based approach using a fairly simple rule set [3].

We created a corpus query system – the Verb Argument Browser – specifically for investigation of corpora represented in the above manner. It answers the following research ques-

tion: what are the most important collocate words of a given verb in a particular dependent position.

2nd prospective thesis. The Verb Argument Browser is a useful tool in during manual building of lexical databases (e.g. Hungarian WordNet or the database of a machine translation system) [4], [5].

Using the Verb Argument Browser we can study particular frequent patterns of a verb manually. Here we arrived at the most important task of our research: how to identify all important combinations, how to collect all important constructions of a verb automatically? The developed algorithm is flexible in two ways: it detects automatically the number of elements in a construction; and it adapts to the fact that only the dependency relationship is relevant in some constructions while some other determines also the dependent lemma.

3rd prospective thesis. Our algorithm – based on cumulative frequency and treating free and fixed dependent lemmas properly – is capable of collecting all typical, frequent VPCs from a corpus with unified representation [6].

IV. THE DICTIONARY

Using the frequent VPCs provided by our algorithm we compile a special dictionary which ...

- is a corpus-driven, frequency dictionary;
- has *verb phrase constructions* (VPCs) as entries, not simple words;
- is a *meaningless* dictionary in the sense of [7];
- is for Hungarian but the core methodology is *language independent*;
- is created in a *mostly automatic* way with less manual lexicographic work;
- can be created with a low budget;
- is hoped to be useful in language teaching and natural language processing both.

We follow the Sinclairian approach of corpus-driven lexicography. We take a corpus and “jettison ruthlessly” [8] all verbs and constructions which have zero or low frequency in our corpus. In other words, we take the data from the corpus as is, and we do not allow the lexicographer to add any “missing” constructions. Knowing that “authenticity alone is not enough: evidence of conventionality is also needed” [8] we take the most frequent VPCs into account and record and display their frequency in the dictionary. We focus on frequent patterns and do not “seek to cover all possible meanings and all possible uses” [9].

The target is all Hungarian VPCs (consisting of a verb plus some noun phrase or postpositional phrase dependents) from verb subcategorization frames through light verb constructions to completely rigid figures of speech. The so called *multiword verbs* (e.g. *to take sg into account* or *to get rid of sg*) are at the heart of our approach. Having *fixed* (like the object *rid* above) and *free* (like the *of* dependent above) dependents both they are borderline cases between verb subcategorization frames and real multiword expressions. Contrary to common intuition, they are expressly frequent, they can not be treated marginally. If we take the fixed dependent as a part of the multiword verb

Hungarian National Corpus →

- 1) chunking to have verbs and noun phrase or postpositional phrase dependents;
→ corpus with unified representation →
- 2) an algorithm based on cumulative frequency of corpus patterns to collect frequent VPCs, with appropriate treatment of fixed and free dependents (details and evaluation can be found in [6]); and another algorithm to collect suitable examples for VPCs;
→ frequent VPCs with examples →
- 3) manual lexicographic work: error correcting and example selecting
→ dictionary

Fig. 1. Dictionary creation process.

itself, we can treat simple and multiword verbs the same way: both can have some free dependents beside. Entries in the dictionary are VPCs, the microstructure apparently integrates phraseology as the a basic units are phrases. We arrange the VPCs around a verb in a subsequent step to form more traditional dictionary entries.

On the one hand, our dictionary can be called a “meaningless dictionary”. It does not contain explicit definitions, just enumerates the frequent VPCs together with corpus frequencies. Most dictionary users are looking up only basic information, for these tasks meaningless dictionaries are efficient [7]. On the other hand, it contains also a corpus sentence exemplifying the meaning. Furthermore, this meaning is fairly concrete, as VPCs (being collocations) usually have one and only one meaning [10]. In fact, most VPCs are real constructions, “form and meaning pairings” [11], as they cannot be broken down into smaller units without loss of meaning. Each VPC represent a pattern of use, and can be paired with one sense of its main (simple or multiword) verb.

The dictionary creation process is mostly automatic: starting from the morphosyntactically tagged and disambiguated Hungarian National Corpus (HNC) [2] we obtain raw dictionary entries using some NLP tools; only the final editing step is manual lexicographic work. The whole process is shown in Fig. 1.

It should be emphasized that step 2 (in Fig. 1) supersedes a substantial amount of manual lexicographic work. As a result of this step VPCs (arranged around verbs) are presented in XML form, so the lexicographer can edit the entries in an XML editor. In step 3 he/she basically has to check if the patterns suggested by the program are correct, and to choose among the example sentences the most appropriate ones. When doing this, the suggestions made in [12] are taken into account (choosing full-sentence examples, or at least clauses with full predicate structure, avoid personal names etc.). Sometimes none of the example sentences are correct or appropriate for illustration, in this cases other ones are retrieved from the HNC by the previously described corpus query system [5]. In this form, the task of the lexicographer is considerable easier and

elver [744]
elver -t [284] hogy minap elvertelek azért, ...
elver jég -t [36] Már ahol a jég nem verte el
a termést!
elver -n **por**-t [95] vagy hogy egy pár túlbuzgó
helyi tanácselnökön verjék el a port.

Fig. 3. Dictionary entry of the verb *elver* (to beat).

beat [744]
beat OBJECT [284] that I beat you yesterday,
because ...
beat ice OBJECT [36] Just where the hail did
not destroy the crop!
beat ON dust-OBJECT [95] or to blame some
overzealous local mayors.

Fig. 4. English translation of the entry in Fig. 3 Verb phrase constructions are translated word by word while example sentences have overall translations, so it can be seen that when *hail destroys* something Hungarians say *the ice beats* it; and *to blame sy* is put in Hungarian something like *to beat the dust on sy*.

the result needs much less corrections than before.

This methodology allows creating smaller budget dictionaries as the programming and support costs (step 1 and 2) are estimated to 1 man-year, and the lexicographic work (step 3) is also about 1 man-year for a dictionary containing about 3000 verbs and 8000 VPCs altogether.

Beside the traditional (alphabetically ordered by verb) presentation we plan to have several indexes. All of them can be generated automatically:

- aggregated list of all VPCs sorted by frequency – in fact this is the true frequency dictionary;
- an index by general patterns (i.e. VPCs without the verb);
- an index by number of fixed/free dependents;
- a frequency list of verb stems;
- an index by lemmas in fixed dependents.

Figure 2 shows an example entry for the verb *elver* (to beat) in XML form. It is in the stage after step 2 (amended by manually choosing one corpus example from the auto-generated ten for each VPC). The corresponding dictionary entry of this verb is shown in Fig. 3. It contains the most important three verb phrase constructions, from which two are multiword verbs. English translation of the entry is shown in Fig. 4.

4th prospective thesis. The dictionary compiled from automatically collected frequent verb phrase constructions contains those expressions which are typical of the language [13]. Using a purely automatic methodology we obtained a learners' dictionary which shows the most frequent verb meanings and "helps students to write and speak idiomatically" [8].

V. LANGUAGE INDEPENDENCY

The algorithms in step 2 (in Fig. 1) are language independent: having a corpus with the unified representation we

can use our general VPC collecting algorithm regardless of language.

The question is whether we can work out this representation for languages different from Hungarian. We chose the Danish language as testbed because its structure is considerably different: while Danish has fixed word order and a system of prepositions, Hungarian has a rich case system and its word order is relatively free.

To demonstrate language independency we converted the Danish Dependency Treebank [14] to the unified representation. Using this richly annotated resource we extracted the clauses identified the dependents, their heads and their relations to the verb straightforwardly. Main result which supports language independency is that a Verb Argument Browser built on the top of this corpus shows the same properties as the original Hungarian version: it can be used to collect multiword verbs (e.g. *være i tvivl om* 'to be in doubt about', *få lov til* 'to allow') and other important verb frames of the language.

5th prospective thesis. Our unified corpus representation can be worked out presumably for a broad class of languages. In essence it only relies on the existence of clauses consisting of a verb and its dependents, and on the fact that a relationship between the verb and a particular dependent can be defined [15].

6th prospective thesis. We saw that chunked corpora can be converted to the unified representation (*5th th.*), and our general algorithm can run on any corpus which is represented in this manner (*3rd th.*). Accordingly, the frequency dictionary of verb phrase constructions (*4th th.*) can be created for a number of languages. All we need is a chunked corpus or a POS tagger and a suitable chunker.

VI. CONCLUSION

We described the whole research line to create a corpus-driven frequency dictionary of verb phrase constructions. It is for the Hungarian language, but we showed that our methodology is language independent in great part due to the unified corpus representation. The general main algorithm runs on this representation and collects automatically all VPCs from corpus. Using this methodology we can obtain a lexical database, which is at first an almost-ready learners' dictionary which lists all frequent VPCs and "helps students to write and speak idiomatically" [8]. Beyond that, it is a rich lexical resource from which many natural language processing tasks could benefit from information retrieval to machine translation.

REFERENCES

- [1] E. Tognini-Bonelli, *Corpus Linguistics at Work*. John Benjamins, 2001.
- [2] T. Váradi, "The Hungarian National Corpus," in *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC2002)*, Las Palmas, Spain, 2002, pp. 385–389.
- [3] B. Sass, "Igei vonzatkeretek az MNSZ tagmondataiban [Verb frames in the clauses of the Hungarian National Corpus]," in *Proceedings of the 4th Magyar Számítógépes Nyelvészeti Konferencia [Hungarian Conference on Computational Linguistics] (MSZNY2006)*, Szeged, Hungary, 2006, pp. 15–21.

```

<entry>
<verb lemma="elver" freq="744"/>
<pattern freq="284">
<frame><p c="-t"/></frame>
<type str="1:01" len="1" fixed="0" free="1"/>
<cits>
  <cit>hogy minap elvertelek azért,</cit>
</cits>
  <pattern freq="36">
  <frame><p c="" l="jég"/><p c="-t"/></frame>
  <type str="3:11" len="3" fixed="1" free="1"/>
  <cits>
    <cit type="sentence">Már ahol a jég nem verte el a termést!</cit>
  </cits>
  </pattern>
</pattern>
<pattern freq="95">
<frame><p c="-n"/><p c="-t" l="por"/></frame>
<type str="3:11" len="3" fixed="1" free="1"/>
<cits>
  <cit type="sentence">vagy hogy egy pár túlbuzgó
    helyi tanácselnökön verjék el a port.</cit>
</cits>
</pattern>
</entry>

```

Fig. 2. Example entry for the verb *elver* (to beat) in XML form.

- [4] —, „Mazsola” – eszköz a magyar igék bővítményszerkezetének vizsgálatára [Mazsola – a tool for investigating argument structure of hungarian verbs],” in *Proceedings of the 1st Alkalmazott Nyelvészeti Doktorandusz Konferencia [Hungarian Student Conference on Applied Linguistics]*, MTA Nyelvtudományi Intézet, Budapest, 2007, pp. 137–149.
- [5] —, “The Verb Argument Browser,” in *Sojka P. et al. (eds.): 11th International Conference on Text, Speech and Dialogue. LNCS, Vol. 5246.*, Brno, Czech Republic, 2008, pp. 187–192.
- [6] —, “A unified method for extracting simple and multiword verbs with valence information and application for Hungarian,” in *Proceedings of RANLP 2009*, Borovets, Bulgaria, 2009, p. (accepted).
- [7] J. Maarten, “Meaningless dictionaries,” in *Proceedings of the XIII EURALEX International Congress*, Institut Universitari de Linguística Aplicada, Universitat Pompeu Fabra, Barcelona, 2008, pp. 409–420.
- [8] P. Hanks, “The lexicographical legacy of John Sinclair,” *International Journal of Lexicography*, vol. 21, no. 3, pp. 219–229, 2008.
- [9] —, “The probable and the possible: Lexicography in the age of the internet,” in *Proceedings of AsiaLex 2001*, Yonsei University, Seoul, Korea, 2001.
- [10] D. Yarowsky, “One sense per collocation,” in *Proceedings of the workshop on Human Language Technology*, Princeton, New Jersey, 1993, pp. 266–271.
- [11] A. E. Goldberg, *Constructions at Work*. Oxford University Press, 2006.
- [12] A. Kilgarriff, M. Husák, K. McAdam, M. Rundell, and P. Rychly, “GDEX: Automatically finding good dictionary examples,” in *Proceedings of the XIII EURALEX International Congress*, Institut Universitari de Linguística Aplicada, Universitat Pompeu Fabra, Barcelona, 2008, pp. 425–432.
- [13] B. Sass and J. Pajzs, “FDVC – creating a corpus-driven frequency dictionary of verb phrase constructions,” in *Proceedings of eLexicography in the 21st Century Conference*, Louvain-la-Neuve, Belgium, 2009, p. (accepted).
- [14] M. Trautner Kromann, “The Danish Dependency Treebank and the DTAG treebank tool,” in *Proceedings of the 2nd Workshop on Treebanks and Linguistic Theories (TLT 2003)*, Växjö, Sweden, 2003.
- [15] B. Sass, “Verb Argument Browser for Danish,” in *Proceedings of the 17th Nordic Conference of Computational Linguistics, NoDaLiDa 2009*, Odense, Denmark, 2009, pp. 263–266.

Comparison of unsupervised word sense disambiguation methods

Gyula Papp
(Supervisor: Gábor Prószéky)
gyupa@digitus.itk.ppke.hu

Abstract—This paper presents and compares different methods of vector- and graph-based text representation created from corpus-based word co-occurrences. Both types of techniques are language independent and they don't need any other resources apart from a large amount of text. The methods were compared by applying them for the unsupervised word sense disambiguation task. The performance of the different representations was measured on a standard dataset: the nouns of the Senseval-2 English lexical sample data were used for this task. The results show that vector-based methods are more appropriate for the representation of larger text units, e.g. paragraphs.

Index Terms—word sense disambiguation, unsupervised WSD, natural language processing, co-occurrence vector, vector-based co-occurrences

I. INTRODUCTION

Semantic analysis of texts is one of the biggest challenges in natural language processing (NLP). The goal of semantic analysis is the computer-based representation of the text sense. A lot of NLP applications need an efficient and good quality solution of the problem. This could improve for example the quality of web searching applications and machine translation systems.

As mentioned before the basic question is how to represent the meaning of texts. The analyzed text units can differ in size: they can be sentences, paragraphs or larger texts. These text units are called *contexts* in general. This paper presents text representation methods based only on corpus-based co-occurrences. These techniques are able to represent any of the previously mentioned context types.

It is not easy to compare different sense representation methods. One solution is to test their performance in an application. The word sense disambiguation (WSD) task can be used for this. Its goal is to decide the actual word sense of a certain word in the analyzed context.

There are more types of WSD: supervised and unsupervised machine learning approaches and lexical resource-based systems.

Supervised WSD methods need hand-tagged data to learn how the actual word sense of an ambiguous word (usually called target word) in a specific context can be decided.

Unsupervised WSD algorithms don't need any tagged data. They don't even try to decide from a previously given list of senses the actual sense of the target word. Their goal is to separate the different uses of the ambiguous word. (The

term word use is preferred to word sense in cases of unsupervised WSD approaches.) They use clustering techniques to group similar contexts together. Hopefully the induced clusters represent the different uses of the ambiguous word. Each new context of the target word will be compared to the clusters and the most similar cluster will be the selected word use.

One of the advantages of unsupervised WSD over supervised WSD is that it doesn't require any sense-tagged corpora. We tried to exploit this advantage by using a lot of contexts to discriminate the different word uses of a target word.

There are two main types of unsupervised word sense disambiguation: graph-based and vector-based techniques. Most of the unsupervised WSD systems are based on the vector space model ([9] and [10]). In 2004, Véronis published HyperLex [1], a graph-based algorithm for unsupervised WSD. The goal of this paper is to compare the performance of the different representation types in case when corpus-based co-occurrences are used for text representation.

This paper is structured as follows. At first the unsupervised WSD task is presented. Section III and IV describe the main ideas of the graph-based and vector-based WSD systems. After that the method comparison experiment and its result are briefly summarized. Finally, the conclusions and plans for the future are presented.

II. THE UNSUPERVISED WSD TASK

The word sense disambiguation task can be summarized in the following way:

- 1) A word with multiple senses is given. This word is usually called *target word*.
- 2) Context containing the target word are also given. These are usually either paragraphs or word windows of size n , where n is usually between 1 and 100.
- 3) A learning algorithm is applied on the contexts in order to learn how the sense of the target word in a previously unseen context can be decided.
- 4) The performance of the applied method can be decided through evaluating it on the test data. This contains contexts not used during training. The proportion of the correct answers on the test data is the precision value of the algorithm.

A WSD algorithm is called unsupervised when it doesn't take the actual sense of the target word in each context into account. This means that there is no need for contexts tagged with the senses of the target word.

This is an advantage over supervised methods because sense-tagging a corpus is a very time-consuming task. However, this is also a disadvantage because unsupervised methods don't know the different senses of the target word so they can't assign them to the different contexts. This is why unsupervised systems only group the similar contexts together – they don't assign word senses to the groups. This process is called *clustering*.

However, in order to be able to measure performance of unsupervised WSD systems the different senses of the target word need to be assigned to the different clusters. For this task sense-tagged contexts are needed although their number is much less than the number of context used for clustering.

So 3 corpora are needed to measure performance of unsupervised WSD systems [3]:

- 1) *Base corpus*: Contexts containing the target word without any sense-tags.
- 2) *Mapping corpus*: Contexts containing the target word tagged with its senses. This corpus is used to assign word senses to the clusters.
- 3) *Test corpus*: Contexts containing the target word tagged with its senses. This corpus is used to measure performance of the applied algorithm.

III. GRAPH-BASED METHODS

Graph-based methods represent all the contexts of the base corpus in one graph. This means that they build one graph from the whole base corpus. The vertices of the graph are words except the target word occurring at least $p1$ times, where $p1$ is a parameter of the algorithm. An edge connects two nodes if both nodes occur in at least $p2$ contexts together where $p2$ is also a parameter of the algorithm. The edge weights are integer number: they correspond to the number of contexts in which both words occur. This kind of graphs is called *co-occurrence graph*.

[1] proved that co-occurrence graphs have the *small world property*. This means they contain strongly connected components and between any two of these components there is only loose connection.

The graph-based unsupervised WSD algorithms try to explore the different strongly connected components (clusters) of the co-occurrence graph. The best performing graph-based unsupervised WSD algorithm is the HyperLex [1]. This algorithm was re-implemented and used in the performance measurement experiment. The next subsection briefly presents its main ideas.

A. The HyperLex algorithm

At first, the HyperLex algorithm creates the co-occurrence graph of the base corpus in the above mentioned way. After this step new weights are assigned to the edges; the following formula presents the computation of the edge weight between nodes i and j :

$$w_{ij} = 1 - \max[P(i | j), P(j | i)], \text{ where } P(i | j) = \frac{freq_{ij}}{freq_j}$$

After the re-weighting of the graph the nearer the edge weight is to 0 the tighter the connection is between the connected nodes. This happens when at almost every occurrence of one word the other occurs as well. (There is a minimal edge weight close to 0.) If the edge-weight is near

to 1 then the connection is loose between the nodes. In this case the words rarely occur together.

At the next step the most frequent nodes of each component in the graph need to be selected. These are called *root hubs* and can be collected with a simple iterative algorithm.

After this step the target word is added to the graph. It is connected with 0 weights with the root hubs. Finally, an arbitrary minimal spanning tree (MST) algorithm is executed on the graph. The root node of the result tree is the node of the target word. (This is sure because only the edges between the target word and the root hubs have 0 weights so these edges are part of the minimal spanning tree.) For the later steps of the algorithm it is enough to store the computed minimal spanning tree.

In the minimal spanning tree the different root hubs correspond to the different word uses of the target word. These can be assigned to the real word senses with the help of the mapping corpus. This can be seen on Fig. 1. in case of the ambiguous Hungarian word *levél*. It has 2 main senses: (*levél1*: *leaf* and *levél2*: *letter*). The component on the left side corresponds to the sense *leaf*; the cluster on right belongs to the sense *letter*. The root hubs of the components are the words *növény* (*plant*) and *postás* (*postman*).

In case of a previously unseen context HyperLex can decide the actual sense of the target word by finding the words of the context that also occur in the minimal spanning tree. After this step points are assigned to the clusters depending on the distance of the found nodes in a component from its root hub. The nearer a node to its root hub is the higher number is added to the components points. Finally, the component with the most points is word use guessed by HyperLex. This word use can be transformed to a dictionary sense with the help of the word use – word sense mapping that has been learnt from the mapping corpus. For each context of the test corpus the word sense computed the previous way is compared to the sense tag of the target word in the context. This is the way how the precision and recall measures of the algorithm can be computed.

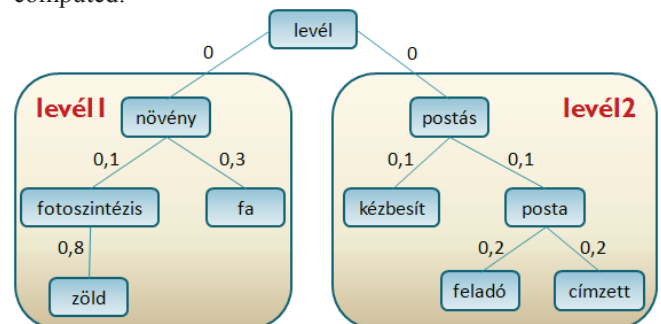


Fig. 1. Generated MST for target word *levél*. The *levél1* and *levél2* senses can be assigned to the clusters with the help of the mapping corpus.

IV. VECTOR-BASED METHODS

Vector-based methods are different from graph-based techniques in the way they represent the contexts of a corpus. While graph-based algorithms create a single co-occurrence graph for the whole corpus vector-based methods represent each context separately with a vector. After context representation the similar context are grouped

together into a cluster. Finally, these clusters are mapped to the dictionary senses of the target word.

At first the dimensions of the representation vectors are collected from the base corpus. These are called features. They can be for example the most frequent words, co-occurrences (unordered word pairs) or bigrams (ordered word pairs) of the corpus.

Then for each context a representation vector is created based on the number of occurrences of the features in the context. In this way k vectors of n dimensions are computed where k is the number of contexts in the base corpus and n is the number of selected features.

The grouping of similar vectors is called *vector clustering*. There are several algorithms for solving this task. The groups after clustering are the different word uses. They can be mapped to the real word senses with the help of the mapping corpus. The clustering and sense-mapping process is presented on Fig. 2. It is enough to store the centroids of the clusters and the earlier extracted features for the later steps of the algorithm.

In case of a previously unseen context its representation is computed. (This step requires the features to be stored.) The guessed word sense is the one that belongs to the nearest centroid. Performance of an algorithm can be measured by calculating the proportion of its correct answers over the test corpus.

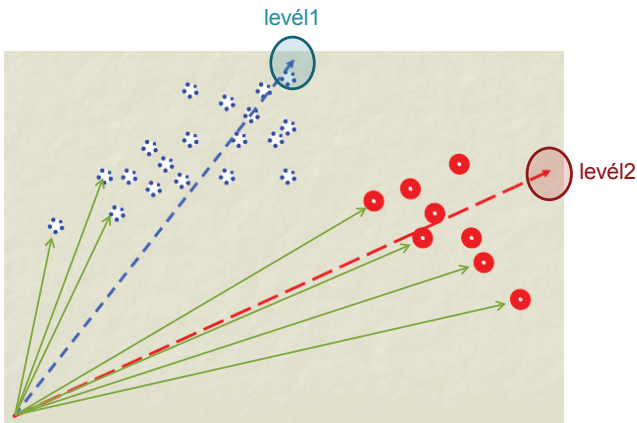


Fig. 2. Context representations for target word *levél*. These belong to 2 groups (marked with red and blue). The dictionary senses are assigned to the centroids of the clusters.

A. The SenseClusters software package

SenseClusters [10] is a free software tool for vector-based context representation. It was implemented in Perl. This tool was used for vector-based unsupervised WSD in the experiments.

It contains a module for *feature extraction* that collects features from an XML file containing contexts in a special format.

The *context representation module* computes the vectors for the contexts of an input file. This program gets the earlier collected features also as an argument.

The *clustering module* is able to group together the similar context representations. It supports several vector clustering algorithms. These algorithms require the number of clusters as an input parameter but SenseClusters provides a tool for predicting this number automatically.

Performance measurement is also supported by the

system but in our case it couldn't be used so a new evaluation module had to be implemented.

V. THE EXPERIMENT

An experiment has been carried out in order to compare the performances of unsupervised WSD systems using different context representations.

The experiment used target words for which there are publicly available sense-tagged corpora. The nouns of the Senseval-2 competition were the target words. This was a WSD competition on 57 previously selected target words. For each target word a corpus was provided by the organizers tagged with the senses of the target word.

The British National Corpus (BNC) was used as the base corpus in the experiment. For each target word about 2000-3000 paragraphs were collected that contain the target word. The HyperLex algorithm and different vector-based unsupervised algorithms were applied on this dataset. The minimal spanning tree (in case of HyperLex) and the cluster centroids (in case of vector-based methods) were in this way created. The re-implementation of HyperLex and the SenseClusters toolkit were used in order to perform the experiment.

The mapping corpora consisted of the training parts of the Senseval-2 data for each target word. The mapping between word uses and dictionary senses of the target words was computed based on these corpora.

By using the previously described mapping it was possible to decide the actual sense of the target word in the contexts of the test corpora (the test part of the Senseval-2). The precision value of a method is the proportion of correct answers over all test contexts. Both the graph-based and the vector-based methods give an answer for all test contexts so their recall values are always the same as precision.

Performance of both types of methods depends on several free parameters. These are usually optimized on an independent dataset. In case of this experiment the nouns of the Senseval-3 competition were used for optimization. The best performing set of parameters was chosen for inputs of the experiment over the Senseval-2 data.

VI. RESULTS

Table 1. shows the results of the experiments. It can be seen that the baseline method that always chooses the most frequent sense of the target word is both by the graph-based and the vector-based highly outperformed.

The optimized vector-based algorithm exceeds the optimized graph-based method with 4 per cent. In cases of the most words this solution gives the best result.

Performance measures of the optimized vector-based algorithm are competitive with the state of the art unsupervised WSD algorithms. Although it is not as good as the best supervised WSD systems but this statement is true for all unsupervised methods because they lose a lot of information at the sense mapping step.

WORD	MFS	GRAPH	VECTOR
art	0,44	0,46	0,46
authority	0,39	0,44	0,52
bar	0,43	0,56	0,59
chair	0,85	0,80	0,82
channel	0,30	0,52	0,64
child	0,59	0,65	0,63
church	0,57	0,70	0,71
circuit	0,27	0,40	0,63
day	0,63	0,63	0,62
facility	0,52	0,63	0,68
feeling	0,63	0,63	0,63
holiday	0,89	0,84	0,78
feeling	0,71	0,63	0,67
material	0,42	0,50	0,54
mouth	0,48	0,53	0,59
nation	0,85	0,77	0,74
nature	0,48	0,51	0,53
post	0,39	0,41	0,52
sense	0,33	0,45	0,43
stress	0,55	0,55	0,56
Average	0,509	0,561	0,603

Table 1. Precision values of the different methods for each target word. The columns correspond to the analyzed words, the baseline method (using the most frequent sense heuristics), the optimized graph-based algorithm and the optimized vector-based system.

VII. CONCLUSIONS

The paper examined graph- and vector-based sense representations of written texts. The main advantage of these methods is that they don't need any preprocessing of the training corpus. (In some cases the corpus is preprocessed but the expensive sense-tagging operation is unnecessary.) They only need large amount of data in the training corpora.

It is difficult to measure the usefulness of the methods. This is why the task was simplified into the unsupervised WSD problem. For measuring the precision and recall values of unsupervised WSD algorithms contexts tagged with the senses of the target word are needed. This is why the data of the Senseval-2 competition were used.

The results show that the optimized vector-based method outperformed the best graph-based algorithm. Its other advantage over the graph-based solution is that it uses the whole base corpus for only the feature selection step while the other technique builds the whole co-occurrence graph of the base corpus during training and this takes more time.

REFERENCES

- [1] J. Véronis, "HyperLex: lexical cartography for information retrieval," *Computer Speech & Language*, 2004, 18(3), pp. 223-252.
- [2] R. Mihalcea, T. Chklovski, and A. Kilgarriff, "The senseval-3 English lexical sample task," in *Senseval-3 proceedings*, 2004, pp. 25-28.
- [3] E. Agirre, D. Martinez, O. Lopez de Lacalle, and A. Soroa, "Evaluating and optimizing the parameters of an unsupervised graph-based WSD algorithm," *Workshop on TextGraphs, at HLT-NAACL*, 2006, pp. 89-96.
- [4] E. Agirre, D. Martinez, O. Lopez de Lacalle, and A. Soroa, "Two graph-based algorithms for state-of-the-art WSD," in *2006 Proc. EMNLP Conf.*, pp. 585-593.

- [5] D. Jurafsky, J. H. Martin, *Speech and Language Processing; An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Second Edition*, 2007, Chapter 20.
- [6] C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*, MIT Press. Cambridge, MA: May 1999, pp. 178-183.
- [7] D. A. Cruse, *Polysemy: Theoretical and Computational Approaches*, chapter Aspects of the Microstructure of Word Meanings.
- [8] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, June 1998, pp. 440-442.
- [9] H. Schütze, "Automatic word sense discrimination," *Computational Linguistics*, 1998, 24(1):97-123.
- [10] A. Purandare and T. Pedersen, "Word sense discrimination by clustering contexts in vector and similarity spaces," in *Proc. of CoNLL-2004*.
- [11] P. Edmonds, S. Cotton: "SENSEVAL-2: Overview," in *Proceedings of The Second International Workshop on Evaluating Word Sense Disambiguation Systems (SENSEVAL-2)*, 2001.

Local Contour Descriptors Around Scaleinvariant Keypoints

Andrea Kovács

(Supervisors: Dr. Tamás Szirányi, Dr. Zoltán Vidnyánszky)
kovacs.andi@gmail.com

Abstract—A new local feature descriptor is defined by generating active contours around keypoints. This new feature set introduces an additional local descriptor of low dimensionality. Active contours are characterized by a small number of Fourier descriptors. Similarity among keypoints in different images are searched through these descriptor sets. Due to promising results the local descriptors were also used for texture classification.

Index Terms—Local features, Active Contour, Fourier descriptor, SIFT, texture classification.

I. INTRODUCTION

DESCRIBING local patches to register image keypoints is an important task for building a huge database from video frames. It might be very time consuming if the dimensionality of features is high. Nowadays quite stable features, so called descriptors are used, to exploit a considerable amount of usable information from an image area. Although now we have very efficient local features, we should continuously search for more applicable solutions. Task is twofold: features must describe the featuring patches at a high efficiency, while the dimensionality should be kept at a manageable low value.

The goal of this paper is to introduce an additional local descriptor of low dimensionality. Presently used descriptors, see [1] for details, are based on frequency analysis, scale invariant feature transforms, steerable filters, local derivatives, spot detections give some statistics or integral values around a salient point. The main assumption in finding local descriptors is the defect of continuity in the discrete neighborhood or the imperfectness of local shape formats. The definition of local shapes is efficiently substituted by a bag of features, described in statistical sets. However, the dimensionality of these descriptors (e.g. SIFT) is quite large resulting in time consuming search process for image/video database.

There are different dimension reduction techniques for data of SIFT or similar approaches: PCA-SIFT [2], GLOH [1], LocMax SIFT [3]. However, the valid interpretation of data content may be loosed during compression. To avoid it, we should looking for methods where some formal meaning of the local properties can be maintained at a reduced dimension. The saliency point (detected corner etc.) can be considered as peak/valley in 3D over the 2D image plain. Different feature collection methods scan the neighborhood of the saliency point to describe the microstructure somehow. We may also consider the shape of this peak/valley in a contour-map description, but in this low-resolution local neighborhood inside a radius of cc. 5 pixels a definite shape cannot be found.

Curve fitting methods for noisy shapes may be used: active contours. The limited number of pixel values in a local neighborhood is surrounded by a curve. This curve might be distorted due to the transformed image contents, but its main characteristic can remain nearly constant through several frames of videos or among geometrical/lighting transformation of images from similar scenes. This paper is about the possible outcome of a hypothetic approach: can active contour be applied instead or together with other local featuring techniques for a better local description of image content? This paper is a first step to this goal: I examine here the possible solutions, regarding the local features, and I will give some measures about this additional approach.

Finding a good salient point is a crucial step of any local featuring processing. For this process several solutions are known, like Harris [4] corner detection or DoG [5] in SIFT. This method can be efficient if the corner-detection is a good peak/valley definition method as well. For this reason, I remain to the very efficient salient point detection of SIFT.

Shape definition around a small salient blotch (centered by a keypoint) is not easy since the limited neighborhood in the raw grid-resolution. Canny edge detection [7] could be efficient, but a closed curve around the keypoint is needed. Active contour can be used as an interpolation filter for the given blotch. Conventional shape definition and comparison methods suffer from the limited amount of information around the blotch. For example, in [8] the dissimilarity between two shapes is computed as a sum of matching errors between corresponding points, together with a term of measuring the magnitude of the aligning transform given by similar shape contents. In case of SIFT corner points we can hardly define shape descriptors around the keypoints in a limited neighborhood. For this reason we should fine an interpolation-based closed-curve detector, what is the active contour for example.

II. MAIN COMPONENTS OF THE FEATURE DESCRIPTION

In this section the main steps necessary to estimate the local curve characteristics are shown:

- 1) Finding keypoints as defined for SIFT [5]
- 2) Generating Active Contours around the given keypoint [9]
- 3) Calculating the Fourier Descriptor for the estimated closed curve [10]
- 4) Finding similar curves counting a limited set of components of Fourier Descriptors [11]

A. SIFT descriptors

Some basic issues in computer vision, named image segmentation, object recognition and picture comparison, are based on significant local features. The segmentation and comparing needs to be efficient and real-time. Robustness is also inevitable in computer vision, similar objects must be found under variant circumstances (like translation, rotation, image scaling, illumination, etc).

SIFT (Scale Invariant Feature Transform) is one of the widely used methods for object recognition. It was published in 1999 by David G. Lowe. According to [5] SIFT uses a class of local image features. This approach transforms an image into a large collection of local feature vectors each of which is invariant to image scaling, translation, and rotation, and partially invariant to illumination changes, noise and affine or 3D projection. Following are the major stages of computation used to generate the set of image features:

- 1) Scale-space extrema detection: The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential interest points that are invariant to scale and orientation.
- 2) Keypoint localization: At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability.
- 3) Orientation assignment: One or more orientations are assigned to each keypoint location based on local image gradient directions. All future operations are performed by providing invariance to scaling and orientation transformations.
- 4) Keypoint descriptor: The local image gradients are measured at the selected scale in the region around each keypoint.

All the keypoints are described with a 128-long-vector.

The disadvantage of SIFT is its high dimensionality. It is known that this efficient data content cannot be compressed [2] without important information loss [3]. Similar local descriptors [1] also give a batch of collected features. Some of them is defined as scale-invariant features, where zoomed attributes describe larger scale connectedness (e.g. scaling in [6]).

Instead of the batch of features, here I examine closed curves around keypoints with an algorithm resulting in a meaningful low dimensional descriptor, and for using it in the selection of similar keypoints in different images.

The main idea for retrieving further description is analyzing the contour features in the neighborhood of the keypoint. If significant features are extracted, these can be added to any keypoint descriptor.

B. Active Contour

The method used for contour analysis is Active Contour (AC) [12]. Active contour is used in computer vision especially for locating object boundaries. It is an energy minimizing algorithm, where the spline (also known as snake) is guided and influenced by different forces. These forces include external constraint forces (like the initial contour given by the user)

and image forces (like the edges and ridges in the image, which moves the snake). The basic version had limited utility as the initialization should have been near the real contour of the object. Problems also occurred while detecting concave boundaries. To eliminate these problems, a new external force, GVF (Gradient Vector Flow) was introduced [9]. Using GVF the initial shape of the snake is almost arbitrary.

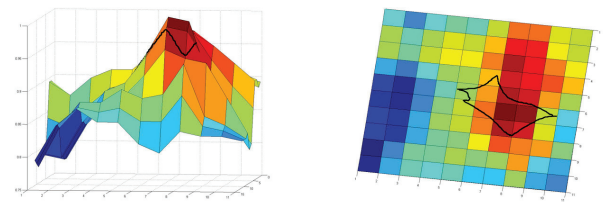
In my work I used the active contour with GVF (called as GVF snake). The snake was computed in a 11×11 size area, where the keypoint was in the middle.

C. Fourier Descriptor

I applied a boundary-based method for the classification [11]. Fourier descriptors [13] are widely used for shape description. In this method, the local contours are classified by the nearest-neighbor rule and the distance metric based on the modified Fourier descriptors [10] (MFD) that is invariant to translation, rotation and scaling of shapes. The method calculates the discrete Fourier transform (DFT) of this complex sequence, which measures magnitude values of the DFT coefficients that are invariant to rotation. MFD method should be extended to get symmetric distance computation. Denoting the DFT coefficients of the compared curves by F_1 and F_2 , the standard deviation function by σ , and n is the cut-off frequency. The 0-th component is discarded to remove the positional sensitivity. The distance metric between two curves is as follows:

$$Dist(F_1, F_2, n) = \sigma\left(\frac{|F_{1,1}|}{|F_{2,1}|}, \dots, \frac{|F_{1,n}|}{|F_{2,n}|}\right) + \sigma\left(\frac{|F_{2,1}|}{|F_{1,1}|}, \dots, \frac{|F_{2,n}|}{|F_{1,n}|}\right) \quad (1)$$

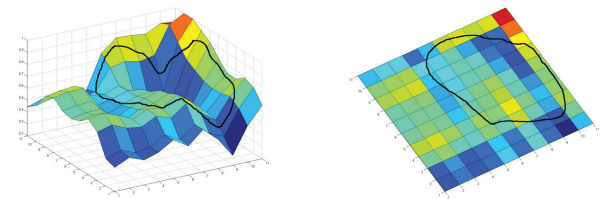
I examined how many Fourier descriptors (**FD**) should be used in the distance computation. For optimizing the cut-off frequencies to maximize the recognition accuracies, the first ten coefficients (excluding the DC component) were chosen.



(a) Contour in 3D

(b) Contour from top view

Fig. 1. Significant contour



(a) Contour in 3D

(b) Contour from top view

Fig. 2. Not significant contour

III. MATCHING SEQUENTIAL FRAME PAIRS

After having the FDs for a keypoint, the possibly outlier FDs must be filtered out. Similar selection is also applicable in other descriptor methods, like ratio of the best two pairing for SIFT. The goal is to filter out the less characteristic points.

Local contours (LC) are not so general features when searching in a large set of characteristic points, since many of them can be easily confused when close to some average contour shape, like circles (Figure 1 and 2).

Therefore a *Filter* function is defined to filter out the not significant curves, which are close to the average meaningless cases:

$$Filter(F_i, n) = \{Th(\frac{|F_{i,1}| - M_1}{V_1}) + \dots + Th(\frac{|F_{i,n}| - M_n}{V_n})\} \quad (2)$$

where M_j and V_j are the mean and the variance of the j th component of $F_{i,j}$ Fourier descriptors over the sample features $\{i\}$, and $Th(\cdot)$ is a threshold function:

$$Th(x) = \begin{cases} 1 & \text{if } |x| > t \\ 0 & \text{otherwise} \end{cases}$$

Choosing $t = 1.2$, the less significant keypoints are eliminated having $Filter(F_i, n) < 3$, resulting in a selection of the most representative points, about 30% of the initial set.

The method was tested on 22 real-life video frames made by an outdoor surveillance camera of a city police central. The algorithm was tested on sequential frame pairs. The filtered, significant keypoints were matched with a simple tracker based on only the location of the keypoints. Then every keypoint was compared with every tracked pair. The distance metric between two curves was measured with eq.(1). The results were evaluated in two ways:

- The j th closest *LC* in Frame 2 to that of i in Frame 1: $\min_j(Dist(F_{1,i}, F_{2,j}, n)|F_{1,i})$,
- Checking the five closest *LC*s for fitting.

As it can be seen in Table I the real pair was the best match approximately in the 40% of the cases, and in 73% of the pairs it was in the best five matches. A typical result is on Figure 3, where the found keypoint pairs are shown on one frame. The original keypoint is in red, and its tracked pair is in blue.

IV. TEXTURE CLASSIFICATION

The above results showed that *LC*s can be comparable features against compressed descriptors. So, due to the promising result on image matching, I aimed at another challenging field of image segmentation, the texture classification.

A well-known and widely-used dataset for texture classification is the Brodatz-textures, which consists of 112 different natural textures, like brickwall, grass, etc. This dataset was used to test the SIFT-method and the *LC*-method, introduced above, for classifying textures. The major stages of the classification algorithm were the following:

- 1) Generating the descriptors (either SIFT or *LC* features) for training and test images
- 2) Clustering the descriptors of all images with k-means algorithm



Fig. 3. Keypoint pairs (red/blue) on the video frame, plotted to the first frame while the second moved

- 3) Classifying the test image according to the clustering results

In both cases 10 differing images for training and 5 images for testing were chosen. These images can be seen in Figure 4 and 5. In one iteration, all the 10 images were used for training and one image was classified.

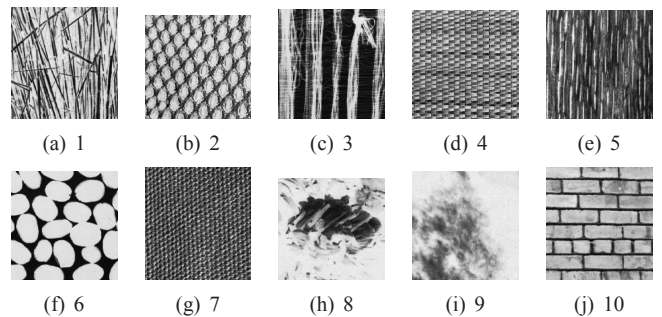


Fig. 4. Textures for training

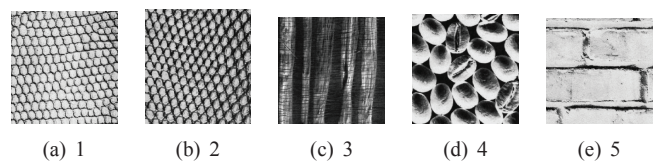


Fig. 5. Textures for testing

The first step, in case of SIFT, was to generate the 128-long SIFT descriptors (Section II-A). While in case of *LC*s, the first three step, mentioned in the beginning of Section II, were needed.

The second step was the clustering, where a simple k-means algorithm, with $k = 15$ was used. The clustered results were summarized and normalized. Therefore every image had a value for every cluster, which meant the ratio of the descriptors rated into the actual cluster: $R(i, 1), \dots, R(i, k)$, where i is the image and k is the number of the clusters.

The last step was to evaluate the clustering and classify the test image. The classification was based on the normalized

Frame pairs	Nr. of keypoints pairs	Real pair is best match (%)	Real pair in best 5 matches(%)
1-2	10	6 (60%)	8 (80%)
8-9	20	8 (40%)	14 (70%)
18-19	22	7 (32%)	16 (73%)

TABLE I
THE NUMBER OF MATCHES FOUND IN THE IMAGES.

Test images	Similar pair	SIFT	LC with static radius(r=20)	LC with dynamic radius (r)
1	2	4 (80%)	4 (80%)	5 (100%) (20)
2	2	5 (100%)	4 (80%)	4 (80%) (20)
3	3	5 (100%)	5 (100%)	5 (100%) (20)
4	6	0 (0%)	0 (0%)	5 (100%) (25)
5	10	5 (100%)	0 (0%)	5 (100%) (30)

TABLE II
RESULT OF TEXTURE CLASSIFICATION: THE NUMBER AND RATE OF THE CORRECT CLASSIFICATIONS

results of the clustering. For each training image the value of each cluster were compared with the test image and the difference was summarized. In Equation 3, $Diff(tr, te, i)$ means the difference of training image tr , test image te for the cluster i . Therefore the total difference:

$$Diff(tr, te) = \sum_{i=1}^k Diff(tr, te, i) = \sum_{i=1}^k |R(tr, i) - R(te, i)| \quad (3)$$

After counting the difference for all training images, the lower value means the higher similarity.

The results can be seen on Table II. The test textures are shown in Figure 5. Every test image was classified 5 times. The numbers of correct classification are in the table. The similar pair from the training images is also indicated. The first column contains the classification results of the SIFT-method.

The second column represents the results of the LC -method, where LC s generated in a fixed 20×20 -sized neighborhood of the keypoint. The 4th and 5th test images showed poor results. In these two cases the iterative pattern is larger, than 20×20 , so the fixed, 20 radius is too small and should be extended. To solve this problem, dynamic radius (**DR**) was introduced. DR is counted based on the variation of the FDs. If the iterative pattern is large, extending the radius results the increasing of the main FD components. This means, the contour of the pattern is more characteristic. Similarly, if the iterative pattern is small, extending the radius has no effect on the FDs, as the contour of the pattern does not change. Therefore, the radius should be extended until the main FD components are increasing. The last column shows the results of the LC -method with DR. The main difference is in case of the 4th and 5th test images, where using the DR obtains better results.

V. CONCLUSION

A new, contour based local descriptor was introduced. The primary aim was not to substitute sophisticated descriptors, like SIFT, just to give lower dimensional additional information with meaningful interpretation. The results shows that LC s can be useful features with meaningful interpretation

and can help for image matching and texture classification. In my future work, I will concentrate on classification of „real” textured images, such as airborne images.

ACKNOWLEDGEMENTS

I would like to thank Dr. Tamas Sziranyi and Dr. Zoltan Vidnyanszky for all the help and support.

REFERENCES

- [1] Krystian Mikolajczyk and Cordelia Schmid, “A performance evaluation of local descriptors,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [2] Y. Ke and R. Sukthankar, “PCA-SIFT: A more distinctive representation for local image descriptors,” in *CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference*, 2004, pp. V.2 506–513.
- [3] D. Losteiner, L. Havasi, and T. Sziranyi, “LocMax SIFT: Non-statistical dimension reduction on invariant descriptors,” in *Int. Conference on Computer Vision Theory and Applications, Lisbon*, 2009.
- [4] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Proceedings of the 4th Alvey VisionConference*, 1988, pp. 147–151.
- [5] David G. Lowe, “Object recognition from local scale-invariant features,” in *International Conference on Computer Vision*, Corfu, Greece, Sept 1999, pp. 1150–1157.
- [6] David G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International Journal of Computer Vision*, vol. 60, pp. 91–110, 2004.
- [7] J. Canny, “A computational approach to edge detection,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.
- [8] S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape contexts,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.
- [9] Chenyang Xu and Jerry L. Prince, “Gradient vector flow: A new external force for snakes,” in *IEEE Proc. Conf. On*, 1997, pp. 66–71.
- [10] Y. Rui, A. She, and T.S. Huang, “A modified fourier descriptor for shape matching in MARS,” in *Image Databases and Multimedia Search*, 1998, p. 165180.
- [11] A. Licsar and T. Sziranyi, “User-adaptive hand gesture recognition system with interactive training,” *Image and Vision Computing*, vol. 23, no. 12, pp. 1102–1114, 2005.
- [12] Michael Kass, Andrew P. Witkin, and Demetri Terzopoulos, “Snakes: Active contour models,” *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.
- [13] C.T. Zahn and R.Z. Roskies, “Fourier descriptors for plane closed curves,” *IEEE Transactions on Computers*, vol. C21, pp. 269281, 1972.

A Visual Human-Machine Interface on Massively Parallel Computers

Norbert Bérci

(Supervisor: Dr. Péter Szolgay, Dr. Tamás Roska)

berci.norbert@itk.ppke.hu

Abstract—Current multimedia and 3D user interfaces need new methods for efficient input however they are controlled by conventional interfaces like the keyboard and the mouse nowadays. Touch based ones are emerging recently both with average and small screen sizes, but not with giant displays which are used mainly in virtual reality systems. The current devices has some properties which make them unusable (wires attached or gloves or other markers must be worn) and lacking some others in need (touchless and distant control). In this paper we present a 3D finger tracking algorithm and its implementation properties on massively parallel camera computers.

I. INTRODUCTION

A. Motivation

The user interfaces in the industrial automation, vehicle control, surveillance, medical application fields are in urgent search for a new human-machine interface which is able to represent the complexity of commands but without the complexity of use and the hassles associated with learning a new system. It would be beneficial also if it could operate in a touchless way. The main characteristics of such a system today is that it requires prerequisite installments on the object tracked, or the object itself should be special compared to the surrounding environment (for example: shape, color, dynamics). With these capabilities in mind our solution is to follow the fingers and define the commands of the human-machine interface as hand gestures.

B. Brief Description

The system runs on the Bi-i smart camera computer [1] [2], and its logical components are depicted on Fig. 1. One part consists of the analogic algorithm running on a massively parallel computing engine (in this case it is a CNN [3] based ACE16k [4] image sensor-processor analogic chip), the hand model and error recovery running on a DSP and finally the communication layer on the communication chip. All of these are in the Bi-i hardware platform physically.

In order to have 3D capabilities two Bi-i is utilized, each with the just mentioned hardware-software architecture. The new part of the system which serves as a glue between the two Bi-is and providing the extra functionality needed includes the 3D reconstruction and OS driver interface, both run on the

This work was supported by the National Research Fund of Hungary (OTKA, Doctoral School Support program), and the Regional University Knowledge Center program of the National Office of Research and Technology, Hungary.

host computer. The communication channel between the parts is TCP/IP over Ethernet, so the system can be easily connected to any machine which needs 3D command input, and it also allows remote placement.

C. The Tracking Algorithm

Before the explanation of the algorithm, we should draw a distinction between postures and gestures: as it is clearly explained in [5] a posture is a state of an object of which the meaning can be fully extracted. In contrast, a gesture has only meaning when we examine (almost) all the states the gesture consists of, which include not only successive image frames but their temporal distribution also. This way complex dynamics could play a role in the recognition process.

We have been dealing with gestures only, so the goal of the algorithm is to follow the objects instead of recognize key frames. Conventional tracking systems search for the tracked object in every frame, which is very hard in certain environments. The main disadvantage of those methods is that they “forget” information gathered in the previous steps.

During the algorithm design process we have chosen to track the fingertips (other suitable features are under investigation) which is done in four phases:

- A. preprocessing
- B. segmentation
- C. skeletonization
- D. masking
- E. feature extraction

The flowchart summary of the analogic algorithm is depicted on Fig. 2. More detailed description of the algorithm and its properties can be found in [6] and [7].

II. 3D RECONSTRUCTION

The two cameras has been set up such that two of their axes are horizontal and vertical, and their third axes are perpendicular to each other. Each of the cameras see a projection of the whole 3D scene, and due to this setup these projection planes are also perpendicular.

The 3D reconstruction software component receives the detected 2D coordinates of the object projected onto their sight-planes. We then construct the two sight-lines which are defined with four points: two of them are fixed (the positions of the cameras) the other two are the tracked 2D coordinates.

In a perfect scenario the sight-lines intersect and the intersection point is the 3D coordinate we were looking for. Due to detection errors this case almost never occurs in practice, so the sight-lines are actually skew lines. It seems feasible to define the tracked 3D coordinate as the middle point between the two sight-lines.

Formally: the direction vectors (\vec{s}_1, \vec{s}_2) of the two sight-lines are the difference of the camera positions (C_1, C_2) and the detected 2D points (D_1, D_2) on the sight-planes:

$$\vec{s}_1 = D_1 - C_1 \quad \vec{s}_2 = D_2 - C_2$$

In order to compute the middle point we have to compute the points which are the closest to the other sight-line. They can be computed as any \vec{c} vector connecting the two lines projected onto the normed direction vectors:

$$\vec{c}_1 = \left\langle \vec{c}, \frac{\vec{s}_1}{|\vec{s}_1|} \right\rangle \frac{\vec{s}_1}{|\vec{s}_1|} \quad \vec{c}_2 = \left\langle \vec{c}, \frac{\vec{s}_2}{|\vec{s}_2|} \right\rangle \frac{\vec{s}_2}{|\vec{s}_2|}$$

where the \vec{c} vector could be for example the vector connecting the camera positions: $\vec{c} = C_1 - C_2$. Having these results at hand, the detected 3D coordinate can be computed by:

$$D = \frac{1}{2}((C_1 - \vec{c}_1) - (C_2 + \vec{c}_2)) + (C_2 + \vec{c}_2)$$

III. ROBUSTNESS

To qualify our proposed system we have done some measurements according to the accuracy of the computed 3D coordinates. The error measure we have chosen is the distance of the two (skew) sight-lines, which is the length of the transversal. Since the transversal is perpendicular to both lines, its normal vector \vec{n}_t can be computed by the cross product of the two direction vectors:

$$\vec{n}_t = \vec{s}_1 \times \vec{s}_2$$

If the \vec{c} vector obtained by connecting any two points on the lines is projected on the normed \vec{n}_t , we obtain the transversal vector, which has got length equal to their inner product:

$$d = \left| \left\langle \vec{c}, \frac{\vec{n}_t}{|\vec{n}_t|} \right\rangle \right|$$

This formula also verifies that this error measure is zero if and only if \vec{c} and \vec{n}_t are perpendicular (the inner product vanishes) which only happens when the lines intersect.

This error measure is in coordinate units which itself is not so meaningful, but if we consider that the system outputs 3D coordinates, this measure can be interpreted as the measure of uncertainty of that output coordinates (something very similar to the notion of the mean and the variance of a random variable). It is also just an approximation since we do not know the exact coordinates of the tracked objects. Over all, this measure has an appropriate level of error representation in our usage scenario. Actual measurements can be seen on Fig. 3.

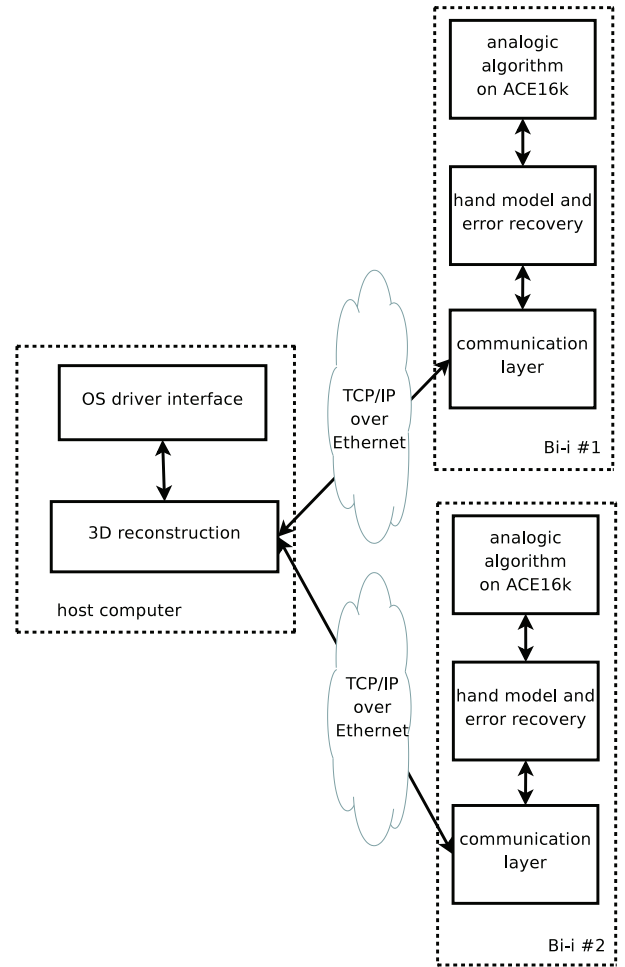


Figure 1. The main components of the system and their relationship

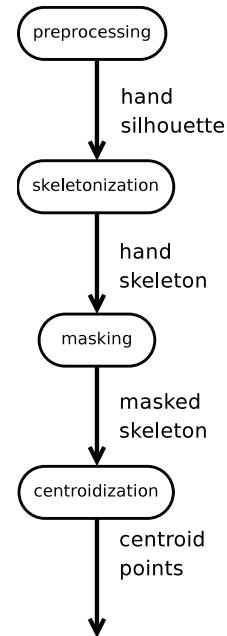


Figure 2. The flowchart of the analogic algorithm.

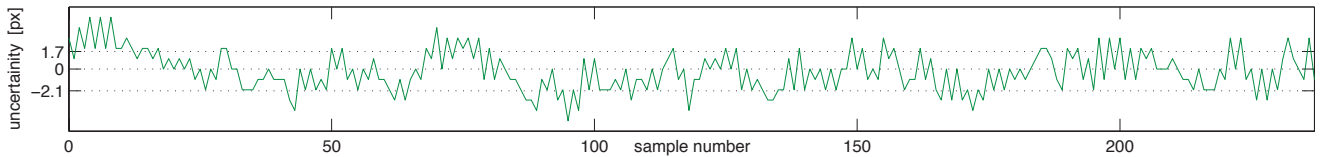


Figure 3. The computed length of the transversal (the uncertainty measure) in pixels (the sign means direction relative to one of the cameras) of almost 250 samples. The biggest is a 5 pixel difference, the data is centered approximately around 0 (the mean is -0.2) and the standard deviation is 1.9. The dotted lines bound the one sigma interval around the mean.

A. Camera Placement and Its Effect on Precision

In order to have precise 3D coordinates, the computation needs precise camera placement as well. Fortunately, misplacements make difference only in the mean of the error, which can be seen in Fig. 4. This way, instead of placing the cameras in the exact position, we have installed them roughly and later adjusted it manually to have the error mean around zero. This phenomena can be used for automatic camera calibration as well.

Perpendicular camera placement could be unfeasible because the cameras should be placed apart roughly 1.5 times the distance of the tracked object and the camera. A more usable setup could be when the distance is greatly reduced, and the whole setup is much more similar to that of our vision system: our eyes are very close but we can even detect 3D scenes. This has a tremendous impact on the precision of the computed coordinates, but aligns well with the experience: we can tell the distance less precisely if the object is farther away.

While in a 2D scenario focusing does not make things harder, since the objects are moving in the same (focused) plane, in 3D we have to overcome the problem of defocused grabbed images. We used zoom objectives in a high zoom setting, so the object could move farther and closer in some interval with being minimally out of focus. We currently do not deal with the non-linear optical distortions of the lenses and their effect on the precision.

B. Skeletonization Algorithm and Its Effect on Precision

Unless it is mainly one of the algorithmic part of the 2D tracking, we have to revisit the skeletonization process, since it has a major impact on the precision. There are two kinds of artifacts:

First, in each point in time the skeletonization algorithm makes a skeleton tree which seems adequate in that time step but if we watch the frames as an animation, there could be huge differences between them (even one pixel difference in the input could make tremendous difference in the resultant skeleton).

Secondly, the end of the skeleton branches apart depending on direction.

The artifacts make the coordinate detection more uncertain and we believe that most of the measured error is originated in these phenomena.

We are heavily executing research on how other skeletonization algorithms could be implemented or a new one developed

with these properties fulfilled as much as possible. For a current survey see for example [8].

C. Failure Handling

Using more than one camera not only enables us to move from 2D to 3D but enables us to recognize object lost errors. If the error measure grows beyond a limit, we can consider the tracked object lost, and refuse to send the coordinates to the PC. In this case the hand model assisted failure recovery is activated, and in most cases it enables the system to return to a normal working state.

D. Frame rate, delay, jitter

As can be seen on Table I the frame rate of the system is about 51-52 FPS, which is enough to track normal human finger movement. It is quite stable also, the differences are due to the error recovery algorithm part. The frequent run of the error recovery signs that the analogic algorithm should be more robust which we believe mainly caused by the skeletonization artifacts. There are plenty algorithmic optimization possibilities ahead.

According our tests, delay and jitter is not measurable and is negligible to the achieved precision.

IV. CONCLUSIONS

We have successfully designed and built a system capable of 3D tracking based on our previously published algorithm. The physical setup and 3D recovery are presented in this paper. We have computed the uncertainty of it by defining an error measure which is a well established analytic formula and we have also dealt with solving misplacement errors.

We have to emphasize it again that the system described here is just the first stage of the vision based gesture interface. It serves as a solid base for the next layers to come. Understanding the meaning of some gestures may involve further studies, hopefully achievable with a neural network implementation.

The algorithm is an analogical algorithm utilizing the computing power of a non-linear cellular wave computer. It is also platform independent, we have implemented it on not just the Bi-i platform used here for the measurements, but on another CNN based device, the Eye-RIS smart camera computer. Other implementations could be straightforward also.

The algorithm is based a on grayscale video stream which is on one hand a definite disadvantage because we can not

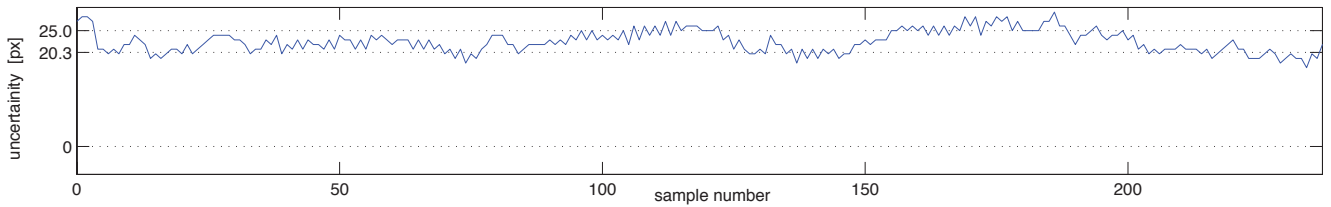


Figure 4. The computed length of the transversal (the uncertainty measure) in pixels of almost 250 samples with rough camera placement. Compared to Fig. 3, the difference is just an offset in the mean (it is 22.6) while the standard deviation remains almost the same (it is 2.4). The dotted lines bound the one sigma interval around the mean.

Table I
FRAME RATE OF TRACKING ONE FINGER

time from start [sec]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
frame rate [FPS]	54	49	48	53	51	51	54	49	55	49	55	48	54	53	52

separate the hand by the skin color, but on the other hand it makes it available to other visual systems (IR cameras, etc.).

The application areas include motorized wheelchair control for disabled people, biohazardous environments for touchless human-machine interface for medical devices, deaf sign language translation for hearing people, to name a few. Currently we are focused on hand gestures, but further research has been planned to enable the successful usage of face gestures (mimics).

ACKNOWLEDGMENT

The author wish to express his gratitude to professor Tamás Roska for his supporting and inspiring role here at the Faculty of Information Technology of the Péter Pázmány Catholic University. His thoughts, reasoning and care gives us the motivation.

REFERENCES

- [1] Á. Zarándy and C. Rekeczky, "Bi-i: a standalone cellular vision system, Part I. Architecture and ultra high frame rate processing examples," in *Proc. 8th Int. Workshop on CNNs and their Appl. (CNNA)*, Budapest, Hungary, July 22–24, 2004, pp. 4–9.
- [2] —, "Bi-i: a standalone cellular vision system, Part II. Topographic and non-topographic algorithms and related applications," in *Proc. Int. Workshop on CNNs and their Appl. (CNNA)*, Budapest, Hungary, July 22–24, 2004, pp. 10–15.
- [3] L. O. Chua and T. Roska, *Cellular Neural Networks and Visual Computing*. Cambridge, UK: Cambridge University Press, 2002.
- [4] G. Liñán, R. Domínguez-Castro, S. Espejo, and Á. Rodríguez-Vázquez, "ACE16k: A programmable focal plane vision processor with 128x128 resolution," in *Proc. Eur. Conf. on Circ. Theory and Design (ECCTD)*, Espoo, Finland, Aug. 28–31, 2001, pp. 345–348.
- [5] J. J. LaViola, "A survey of hand posture and gesture recognition techniques and technology," Department of Computer Science, Brown University, Providence, Rhode Island, USA, Tech. Rep. CS-99-11, 1999.
- [6] N. Bérci and P. Szolgay, "Vision based human-machine interface via hand gestures," in *Proc. Eur. Conf. on Circ. Theory and Design (ECCTD)*, Seville, Spain, Aug. 26–30, 2007, pp. 496–499.
- [7] —, "Application of cellular wave computers in high speed real-time processing: Measurements of a finger tracking algorithm for human-machine interface," in *Proc. Int. Symp. on Nonlinear Theory and its Applications (NOLTA)*, Budapest, Hungary, Sept. 7–10, 2008, pp. 184–186.
- [8] P. Golland and W. E. L. Grimson, "Fixed topology skeletons," Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, Tech. Rep., 2000.

The Relation of Kinematic Movement Patterns and Muscle Synergies in Lower Limb Cycling

Tamás Pilissy

(Supervisor: Dr. József Laczkó)

piltom@ieee.org

Abstract — In the beginning of our functional electrical stimulation (FES) project the functionality of our current stimulator was sufficient but last year we decided to develop an own stimulator that will dispense with all the limitations and faults of the old one. However the development of this new device is still in progress, we have prepared to set it up with new stimulation patterns. It means not only a program which could generate input parameters, but also some new approaches that would be the source of new stimulation methods. For this reason we investigated angular changes (velocity and acceleration) of knee joint, relation between cycling cadence and angular velocities in joints, muscle attachment sites and muscle lengths. Originating from the latter a 3D kinematical model of the lower limb was also developed.

Index Terms — 3D kinematical model, EMG, muscle length, muscle stimulation

I. INTRODUCTION

In 2007 we began our functional electrical stimulation (FES) related project with a standard stimulator made by Krauth+Timmermann Ltd. Although this device is still in use, we have started to develop a completely new stimulator which will be much more customizable. This development is vital since the functionality of the currently used stimulator is very limited; e.g. most of the stimulation parameters cannot be changed during the stimulation procedure. Beside, the amplitude of the electrical signal which is the only parameter that is changing as a function of pedal angle, remains constant in that time interval in which the actual muscle is stimulated. The issue is that the output channels cannot be programmed to give increasing current to the muscles at the beginning of activation. To overcome this limitation I specified a new stimulator in which each channel is planned to be controlled. The planned control parameters are amplitude, frequency and pulse width.

The proper control of the electrical signals requires knowledge about the relation of desired kinematic movement patterns and underlying muscle activities.

I processed and analyzed a large dataset derived from healthy subjects' lower limb cycling. The measurement of healthy movement kinematics and muscle activities has been described in my previous report. This year I further examined these data taking into account individual muscle geometry. I paid special attention to changing muscle synergies while the speed of the movement increased.

Meanwhile I continued FES-cycling measurements in the National Institute for Medical Rehabilitation with my three extant and some new paraplegic patient. They cycled twice a week, while physiological parameters (blood pressure, heart rate) were measured [1].

II. STIMULATOR SPECIFICATION

Basically the device will have two inputs: a *stimulation pattern (1)* on a memory card that will contain all physical parameters of the stimulation as functions of *pedal angle (2)*.

The stimulator pattern will be composed of a header and three tables. The header will contain a name for identification and all of the parameters (neuromechanical delay, starting parameters) that are remaining constants during the stimulation procedure, so this part is independent from the pedal angle. Contrarily, the second part of the file will contain three lookup tables that will define the percentage values of amplitude, frequency and pulse width on each of the eight output channels as functions of pedal angle.

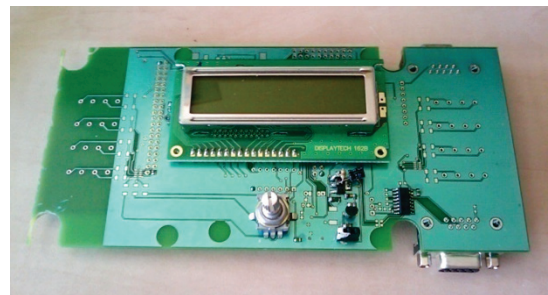


Fig. 1. Mainboard and LCD display of the first prototype of the stimulator.

Since different values can be assigned to every degree of the pedal angle, there are unlimited possibilities to generate custom stimulation pattern. In the beginning of the stimulation the actual stimulation parameters will be the product of the percentage values (taken from the lookup tables) and the corresponding starting parameters that will be changeable manually on the stimulator itself. As for adjustment 8 channel-selector button and special knob will be placed on the stimulator. During the stimulation actual values of the parameters will be shown on a dot-matrix display. For experimental purposes, the stimulation will also be controllable externally from a PC via RS232 or USB port.

There will be numerous features that will serve safety reasons such as stopping of the stimulation when the cadence of cycling drops below a predefined value. User friendliness will be provided by sound signals and simple menu structure displayed on the LCD display.

Beyond the above mentioned, the stimulator will have even one large advantage over the current one. Namely it would be usable on every bicycle-like device (e.g. ergometer, stationary bike, tricycle) because encoding of the pedal angle will be performed by a tiny device which is practically applicable to every rotating shaft. This pedal angle encoder will use serial communication protocol to send pedal position in every millisecond with an accuracy of 0.36° which is far above our requirements.

III. PROCESSING OF MEASURED HEALTHY MOVEMENT PATTERNS

A. Angular acceleration and muscle activities of knee

Earlier the stimulation pattern was based simply on pedal angle which is common method in FES-cycling systems [2,3]. Now we are working on a more physiological approach of stimulation focusing on angular changes of the knee [4].

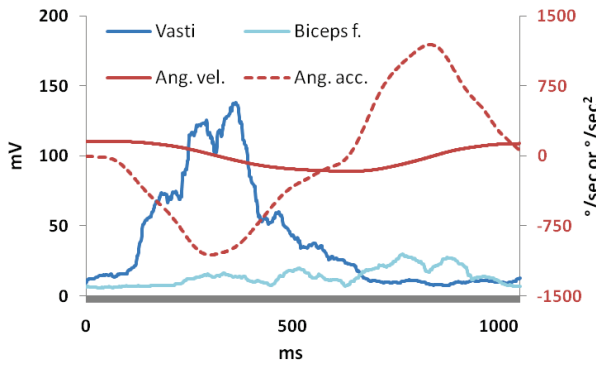


Fig. 2. Angular velocity (solid, red) and acceleration (dashed, red) of the external angle of knee joint and muscle activities are presented. Vasti: knee extensor (blue); Biceps femoris: knee flexor (light blue). At 0ms the pedal is in the lowest position. Knee angle means outer angle of the knee joint. A sample cycle from one subject.

Muscle activities and angular velocity and acceleration of knee joint of a typical cycle are presented on Fig. 2. It can be clearly seen that activity of the knee extensor muscle (vasti) is highest at highest negative acceleration of the knee joint. Activity of the knee flexor muscle (biceps femoris) was usually very low because the feet were not fastened to the pedal (so the subjects could not pull it), but a small increase experienced in flexor activity around the highest positive acceleration of the knee joint.

B. Cycling cadence and angular velocities in joints

The speed of the cycling (pedal velocity) depends on angular velocities in the joints. The question is which of these angular velocities have to be changed mostly to reach higher pedaling speed.

As a first step toward the answer every angular velocity for the whole database of 41 subjects, 6 condition and approximately 10 cycles per condition was

computed. After the pedal speed and angular velocities at hip, knee and ankle joints were retrieved, I used these data to define average angular velocities for every subject's 6 different cycling measurements. Since the angular velocities at the joints have a sinusoidal shape, the average of the momentary velocities would be a nearly zero. Therefore I rather computed average values from absolute values of velocities. In (1) slow and fast refers to cycling speeds of 45 and 60 rpm respectively, while j means joints.

$$P_j = \frac{\frac{1}{n} \sum_{i=1}^n |\omega_i|_{j,slow}}{\frac{1}{n} \sum_{i=1}^n |\omega_i|_{j,fast}} \quad (1)$$

Then I divided these velocity ratios of hip, knee and ankle by the ratio of the pedal, hereby I could analyze if angular velocities at the joints changed together with that of the pedal. Table I presents the average ratios (\pm standard deviation) of the 41 healthy subjects at three different load of pedal (L1, L5, L8).

TABLE I
ANGULAR VELOCITY CO-CHANGES OF PEDAL AND JOINTS BETWEEN SLOW AND FAST CYCLING

Level	Hip	Knee	Ankle
L1	1.016 ± 0.157	1.097 ± 0.281	1.293 ± 0.467
L5	0.994 ± 0.150	1.028 ± 0.206	1.246 ± 0.315
L8	1.023 ± 0.265	1.025 ± 0.223	1.182 ± 0.346

Mean values \pm standard deviations of the 41 healthy subjects in each level of pedal resistance are presented.

Angular velocities of hip and knee joints changed almost equally with that of the pedal, but the angular velocity of ankle joint increased at a higher rate compared to the pedal. Additionally, standard deviations were also greater in the case of ankle joint.

This investigation revealed that among the joints of the lower limb, ankle can rotate the most ways during cycling, which is in accordance with our previous results concerning the role of ankle joint [5].

C. Muscle geometry

In cooperation with the Semmelweis University, Faculty of Physical Education and Sport Sciences we developed a 3D geometric model for determining muscle attachment sites [6]. Based on this I interpreted a MATLAB program to compute and visualize muscle length changes for all the subjects.

I would like to emphasize that we did not use invasive or imaging techniques, thus we searched methods for muscle attachment determination in the literature.

First of all we had to convert the measured marker positions to actual coordinates of the center of joints. For this reason the coordinates of hip, knee and ankle had to be decreased by 60, 45, 25 mm respectively in the direction that is perpendicular to plane of motion.

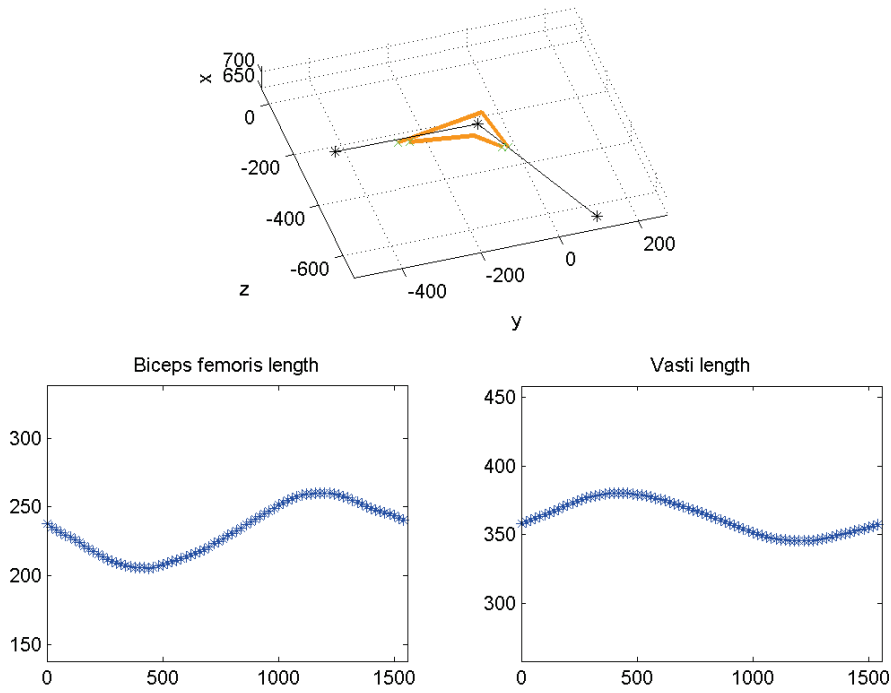


Fig. 3. Screenshot from the animation that shows muscle attachment sites (upper) and muscle lengths (lower). Upper part shows hip, knee and ankle marker locations and lines between them (black), locations of muscle origins and insertions (green) and line of muscles (orange). Units are in millimeters. Lower part shows muscle lengths in millimeters as functions of time (milliseconds).

Then we defined the first base vector (Y_{femur}) of a new frame of reference by subtracting the new knee coordinates from the new hip coordinates and divided the obtained vector by its length. To compute the second base vector the coordinates of the center of the ankle joint were subtracted from the knee coordinates. Then the cross product of this vector and the first base vector was divided by its length which resulted in the second base vector (Z_{femur}). Finally the third base vector (X_{femur}) was computed as the cross product of the previous two. The introduction of this new reference frame, originated from the center of knee, was necessary because in this way we could compute muscle attachment sites on the femur using the coordinates given by Hoy et al. [7]. The product of each new base vector and the adequate coordinate resulted in three vectors. The sum of these vectors was added to the coordinates of the knee to get the actual location of the muscle origin in 3D space. The computation of muscle insertions was very similar. The whole algorithm was implemented by a MATLAB program which computed muscle origin and insertion of biceps femoris and vasti in every millisecond. Later the program was expanded with another features, like muscle length computation and 3D plotting and animation which helped us to check the validity of our model. After the first run of the algorithm we got ambiguous results; muscle attachment sites were realistic and impossible among the subjects. The source of the problem was that Hoy et al. derived the coordinates of muscle attachment sites from an anatomical study based on three cadavers by Brand et al. [8]. Since the three cadavers had different heights (ranged from 163 to 183cm) than our subjects

(ranged from 160 to 195cm) we had to insert a height related factor to the algorithm.

$$f = \frac{\text{subject height}}{\text{average cadaver height}} \quad (2)$$

Then the vectors pointing from the knee to the muscle attachment sites were multiplied by the factor defined in (2). This modification resulted in realistic location of muscle origins and insertions (Fig. 3.).

To obtain muscle length we had to define two other points, called via points on which the lines of muscles brake. Via points were defined as the points on the angle bisector of the knee that are 35mm (vasti) and 45mm (biceps femoris) far from the knee point. Obviously these distances were measured in opposite direction.

Finally, muscle lengths were obtained by computing the length of the broken lines (Fig. 3.).

IV. FES-CYCLING PROJECT

As far as the practical application of FES is concerned I continued the measurements with slightly more spinal cord injured patients than last years. At the moment we have five spinal cord injured subjects who are training with FES twice a week. We further monitored heart rate and blood pressure, and measured blood-gas, spiroergometry and breath-function from time to time. The new patients' results are promising but here I present only the performance graph of a patient who has been attending FES-cycling trainings very regularly in the last 18 months and now he had the highest performances

among all of the patients during the three years of the FES-cycling program.

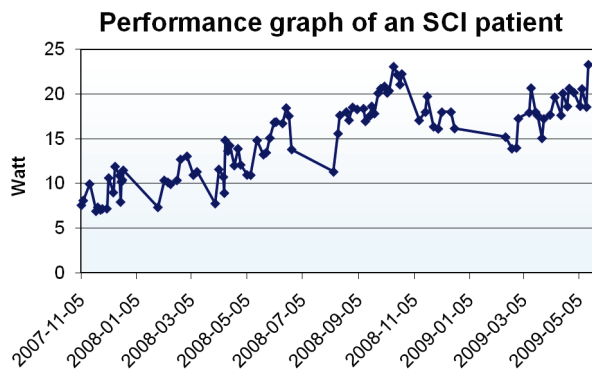


Fig. 4. Performance graph of our strongest patient during 18 months of regular FES-cycling. Drops in performance caused by longer intermissions are clearly visible.

All of our spinal cord injured patients are waiting the completion of the new stimulator since they will probably get one and will be able to use it in their own homes. Realization of this would be helpful not only for them but for us too, since they will be able to train their paralyzed muscles even everyday which undoubtedly resulted in a great increase in their performance and state of health as well.

V. CONCLUSION

We made important steps towards the implementation of a new stimulator that will be capable to accurately apply the results that we acquired from the processing of healthy movement patterns. It seems that all the requirements written in my specification will be realized and clinical application of some new stimulation patterns will be performed in the next months.

The software that is capable to provide adequate commands to the hardware has already been finished.

It is an interesting theoretical question that what kinematic parameter is controlled during neuromorph control of limb movements? Without dealing with high level neural structures, without invasive investigations, with appropriate processing and evaluation of externally measured movement patterns I presented how to discern which physical variables have to be controlled to mimic natural human like movements.

According to inverse kinematics it is not obvious that all of the three joint angular velocities must be increased by the same rate in order to reach higher pedaling speed. Our investigation showed that this is happening at hip and knee in human cycling, but the ankle is behaving differently, since we experienced a higher rate of velocity increase than that of the pedal. In former study [5] I proved that the ankle should play an important role. Now we proved that by changing the velocity of pedal.

A computer model has been also developed in which muscles were presented with their midlines (directions of muscle force) that connect muscle attachment sites through via points near the rotated joint. We managed to adjust muscle attachment parameters of cadavers to our

subjects. The resulted muscle attachment sites were plotted in 3D animation, besides muscle lengths were computed.

VI. FURTHER PLANS

In the near future we will try the new stimulator firstly on a rat, then in case of success, on our patients. After the first prototype is proved to be reliable, we would like to make some and give them to our patients for use at home.

As for the kinematical model, it will be improved to dynamical by adding forces to it.

Muscle length is only one component in the determination of muscle force, therefore we will compute muscle contraction velocity and acceleration too, and compare the resulted force with the measured muscle activities (EMG).

ACKNOWLEDGEMENT

We express our thanks to Attila Tihanyi and Dávid Kormos for building the stimulator, to Dr. András Klauber, Dr. Gábor Fazekas and Dr. Jenő Malomsoki for providing the clinical guidance and environments, to the subjects for their participation and to Imréné Szanyi and Györgyi Stefanik for their help in the measurements. Our research is supported by a Grant of the Scientific Council of Healthcare No. ETT 363/2006.

REFERENCES

- [1] Pilissy T, Klauber A, Fazekas G, Laczkó J, Szécsi J: Improving functional electrical stimulation driven cycling by proper synchronization of the muscles, *Ideggyogy Sz* 2008;61(5-6)
- [2] Szécsi J, Fiegel M, Krafczyk S, Straube A(2004) Smooth pedaling of the paraplegic cyclist – a natural optimality principle for adaptation of tricycle and stimulation to the rider, *J.Rehabil. Res. Dev.*,41 Suppl.2: 30
- [3] Pilissy T, Lábmozgások kinematikai és bioelektromos jellemzőinek modellezése gerincsérültek és egészségesek kerékpározó mozgásánál, Diploma work, 2007
- [4] Katona P, Pilissy T, Fazekas G, Laczkó J: Ízületi szögsebességek, izomaktivitások és a pedálhajtás sebességének kapcsolata kerékpározó mozgáskor. (poster) VII. Országos Sporttudományi Kongresszus, 2009, Budapest
- [5] Pilissy T, Pad K, Fazekas G, Horváth M, Stefanik Gy, Laczkó J: The role of ankle-joint during cycling movement task. Proceedings of the 9th Congress of European Federation for Research in Rehabilitation, Budapest, Aug 26-29. 2007. *Int J Rehabil Res* 2007;30 (Suppl 1):58-59.
- [6] Katona P: Gerincsérültek terápiájában alkalmazott Funkcionális Elektromos Stimuláció hatékonyságának növelése, Diploma work, 2009
- [7] Hoy M G, Zajac F E, Gordon M E: A musculoskeletal model of the human lower extremity: the effect of muscle, tendon, and moment arm on the moment-angle relationship of musculotendon actuators at hip, knee, and ankle, *J. Biomechanics*, Vol. 23, No. 2, pp. 157-169, 1990
- [8] Brand R A, Crowninshield R D, Wittstock C E, Pederson D R, Clark C R, van Krieken F M: A model of lower extremity muscular anatomy, *J. Biomechanical Engineering*, Vol. 104, pp. 304-310, 1982

Non-linear 3D Model of Muscle Forces and Kinematic Variances in Reaching Arm Movements

Róbert Tibold
(Supervisor: Dr. József Laczkó)
tibro@digitus.itk.ppke.hu

Abstract— A three-dimensional (3D) kinematic model that is able to simulate kinematic properties and muscle forces in reaching arm movements with different loads in the hand is presented. Healthy subjects performed reaching arm movements repetitively. 3D coordinates of anatomical landmarks and electromyograms of arm muscles were measured. 3D angular changes, angular acceleration, muscle moment arms, inertial parameters of the upper extremity were calculated to determine 3D muscle torques and forces. Hand position, joint angle and predicted muscle force variances, furthermore ratios of movements performed with load to without load cases were also determined, respectively. We found that the load stabilized the movement. Although kinematic variances were smaller for movements with load than without load, there were no significant differences between them. Since muscle force variances were increasing significantly by moving the heavier object, the stabilization phenomena is derived from raising muscle synergies.

I. INTRODUCTION

The number of handicappeds is increasing. It's important to help these people to improve their social life after being handicapped. Our aim is to develop a special limb controlling method [1] for different rehabilitation techniques such as functional electrical stimulation (FES). In tetraplegics and paraplegics there are some FES method applied during cycling. [2],[3],[4] However, in hemiplegics such as stroke individuals and in tetraplegics, we have lack of information whether FES would be applied to stimulate muscles to get their upper extremities moved. To get the whole arm moved by an artificial method is rather complicated because the complexity of the shoulder mechanism. Furthermore, the three dimensional (3D) inverse kinematic problem must be solved in order to get muscle forces needed to reach one selected point in the 3D space. The two dimensional (2D) inverse kinematic problem is well studied and applied for instance in graphical programming and in a neuro-mechanical transducer model that is able to determine muscle forces and firing frequencies of flexors and extensors [5]. This neuro-mechanical model is also applied for determining movement patterns in rats during swimming by Laczko et al. [6] Furthermore, a commercially available product *FES Hand Grasp System* was introduced in Cleveland F.E.S Center, [7] in the end of the 80's that has been used for stimulating mostly the hand, fingers by exciting muscles of the wrist,

and rarely the forearm.[8] Controlling the lower segments (wrist, fingers) of the upper extremity by using the FES is well studied. Electrical stimulation has been used to provide grasp and release tasks in handicapped individuals. Different systems have been developed using surface electrodes [9],[10],[11], percutaneous electrodes [12] and implanted systems [13].

Our aim is to develop a musculoskeletal model that is able to 1) simulate arm movements by using measured kinematic data as the inputs of the model; 2) determine 3D muscle torques and muscle forces, assuming individual muscle attachments and biomechanical characteristics 3) generate muscle stimulation patterns and strategies from the computed muscle forces.

In the current study we only deal with 1-2 by presenting an ultrasound measurement technique and protocol; and the results of a special simulation method written in MATLAB environment based on [5],[6] but in the 3D space contributing to the complex shoulder mechanism.

II. METHODS

II.1. EXPERIMENTAL METHODS

Twenty healthy subjects were measured experimentally according to [14] with the configuration presented in (Table 1); data were processed by the methods based on [14].

TABLE I
THE MEASUREMENT CONFIGURATION OF THE UPPER EXTREMITY.

MARKER CHANNEL		EMG CHANNEL	
1	Prox. Clavicle	1	Biceps (BI)
2	AC joint	2	Triceps (TR)
3	Prox. Humerus	3	Delta anterior (DA)
4	Dist. Humerus	4	Delta posterior (DP)
5	Dist. Ulna		
6	Dist. Radius		
7	Little finger prox. Metacarp.		
8	Moving object point		

Marker channel numbers 1-7 marked anatomical landmarks of the upper extremity while EMG channel numbers 1-4 measured the activity of the muscles rotating the elbow joint and shoulder complex respectively.

II.2. SIMULATION METHODS

A 3D simulation model is presented in this section. The model applies the following coordinate system: a) x-axis is horizontal in the frontal plane directed outward b) y-axis is horizontal in the sagittal plane directed forward c) z is perpendicular to the x-y plane directed upward.

Input parameters of the model are 3D coordinates of anatomical landmarks (Table 1), calculated 3D shoulder, elbow and wrist angles. Joint angular changes were obtained according to [14] by using the cosine rule. Beside 3D joint angles summarized in [14], shoulder angular changes respect to the frontal and sagittal plane were also determined. Further input parameters of the model were arm segment masses calculated according to Zatsiorsky [15] and arm segment lengths.

The torque generated by a single muscle is computed as follows:

$$F_m(t) * R_m(t) = \beta(t) * \Theta^{(joint)}(t) - T_g(t) \quad (1)$$

where $F_m(t)$ is the force, $R_m(t)$ is the moment arm, $\beta(t)$ is the acceleration of the joint that is spanned by the muscle, $\Theta^{(joint)}(t)$ is the moment of inertia of the rotated body part, $T_g(t)$ is the gravitational torque. $F_m(t)$, $R_m(t)$, $\beta(t)$ and $T_g(t)$ are vectors in the 3D space as a function of time. The direction of the force and torque exerted by a muscle; the direction of the moment arm are perpendicular to each other. Thus, if we know the direction of the torque of a muscle and the moment arm of the same muscle then the direction of the muscle force is known.

The magnitude of the angular acceleration vector of the rotated joint $\beta(t)$ is calculated as $\beta(t) = \frac{d^2\alpha(t)}{dt^2}$ where

$\alpha(t)$ is the joint angle of the rotated body segment. The direction of the angular acceleration vector is perpendicular to the plane in which the rotation takes place. The axis of the rotation was calculated by the cross product of the actual unit segment vectors. Resulted vectors are perpendicular to the plane of rotation if they point from the joint of interest. When the internal joint angle was decreasing (flexion) and $\beta(t) > 0$ the acceleration vector points toward the left of the plane of rotation (according to the right-hand rule). In flexion with $\beta(t) < 0$ the acceleration vector points toward the right of the plane of rotation. In the case of increasing joint angle (extension) with $\beta(t) < 0$ the acceleration vector points toward the right of the rotation plane. During extension with $\beta(t) > 0$ the angular acceleration vector points toward the left.

Muscle moment arm vectors, $R_m(t)$ were determined by 3D muscle attachment sites based on the study of Veeger et al. [16]. A subject with virtual body heights, segment lengths and muscle attaches was created. All muscle attachments were assumed as points on the given bone segment. 4 arm muscle attachments were determined by generating virtual ratios of the distance of the selected muscle attachment from the nearest anatomical landmark on the given bone to the length of the bone segment that contained the muscle attachment point and the anatomical landmark as well. The moment arm calculation method is presented in (Figure 1) in the shoulder extensor muscle.

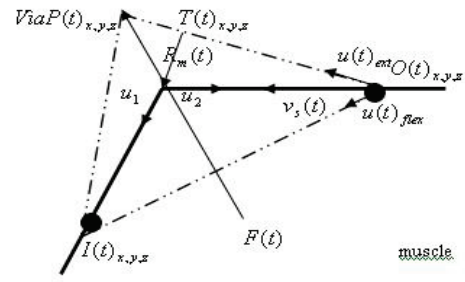


Figure 1 - The moment arm vector calculation in shoulder's extensor

First $ViaP(t)_{x,y,z}$ via point was determined that contains the muscle. If we know the 3D coordinates of point $T(t)_{x,y,z}$ and the coordinates of the joint of interest the moment arm vector of the muscle is given by subtracting the coordinates of $T(t)_{x,y,z}$ from the coordinates of the joint. $ViaP(t)_{x,y,z}$ is on the bisectrix vector $F(t)$ of the joint angle. If the distance of $ViaP(t)_{x,y,z}$ and $I(t)_{x,y,z}$ (insertion) is greater than its distance from the origin $O(t)_{x,y,z}$, $T(t)_{x,y,z}$ is determined as the sum of the dot product of the unit vector $u(t)_{ext}$ and the segment vector $v_s(t)$ directed from the origin to the joint multiplied by $u(t)_{ext}$ and the coordinates of the origin itself. Moment of inertia (MoI) $\Theta^{(joint)}(t)$ of all arm segments was obtained by using the parallel axis theorem. Arm segments were considered as uniform cylinders with different thickness.

$$\Theta_{parallel} = \Theta_{cm} + Md^2 \quad (2)$$

where Θ_{cm} is the MoI at the center of mass (CoM), M is the mass of the body, d is the distance between the axis through the CoM and the axis through the point of rotation. MoI about the center of mass of the segment is calculated:

$$\Theta_{center}^{(segment)} = \frac{1}{4}Mr^2 + \frac{1}{12}ML^2 \quad (3)$$

MoI about the end of the rotated body part is obtained:

$$\Theta_{end}^{(segment)} = \frac{1}{4}Mr^2 + \frac{1}{3}ML^2 \quad (4)$$

where r is the radius, L is the length of the segment. The object as a solid cylinder was determined as follows:

$$\Theta^{(mass)} = \frac{1}{2}M_{mass}r_{mass}^2 \quad (5)$$

If the rotation center is in the shoulder and the load is not held the MoI is calculated:

$$\Theta(t) = \Theta_{end}^{(U)} + \Theta_{center}^{(L)} + M_L A^2(t) + \Theta_{center}^{(H)} + M_H B^2(t) \quad (6)$$

(U), (L) and (H) refer to the upper arm, the lower arm and the hand respectively. If the object is held the MoI is calculated as the sum of (5),(6) and the product of the mass of the object and the distance square of its CoM from the rotation center of the rotated body part.

MoI of the elbow when the load is not held:

$$\Theta(t) = \Theta_{end}^{(E)} + \Theta_{center}^{(L)} + M_H C^2(t) \quad (7)$$

When the load is held it is calculated as the sum of (5),(7) and the product of the mass of the load and the distance square of its CoM from the rotation center.

$A(t), B(t), C(t)$ are distances between the rotation axis of the proximal and distal limb segments and the CoM of the rotated body part as a function of time.

Gravitational torque $T_g(t)$ was calculated as follows:

$$T_g(t) = R_g(t) * m * a_g \quad (8)$$

where $R_g(t)$ is the gravitational moment arm; m is the sum of rotated limb segment's masses and a_g is the gravitational acceleration. To determine the gravitational moment arm, the CoM of the entire upper extremity was calculated. The gravitational moment arm is the distance of the CoM of the rotated segment from the vertical line passing through the rotation center. Muscle cooperation is not implemented in the presented model. Hence, muscle forces were determined separately by using equation (1).

III. RESULTS

In kinematics and muscle forces time normalized variances of the performed ten movements per load condition 1)CD case (0.06 kg) 2) load (2kg) during uplifting and putting down were calculated to A) the whole interval of the movement and B) to the interval of holding. Thus, every movement was divided into 3 parts. After leaving the initial position was considered as the pre-holding part while the load was in the hand was considered as holding; while moving back to the initial position was considered as the post-holding part. In A) because of time normalization we calculated variances in every time step of ten trials. As a result, we got 20 variance vectors as a function of normalized time. Vectors were averaged and assigned to the actual load condition in both directions. (Figure 2) In B) holding interval was determined by assuming that it was started when the distance of 7th marker and 8th marker was smaller than the threshold. It was assessed to the minimal distance between the two markers (7-8) plus 25mm to avoid any measuring inaccuracy. Holding was finished as the distance was greater than the threshold after detecting a start of holding. For all detected holdings time normalization was made because of varying duration.

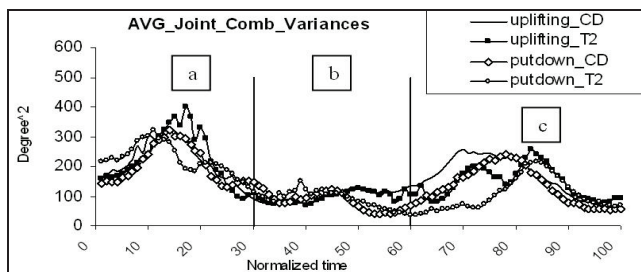


Figure 2 – Averaged joint configuration variances of 20 subjects for the whole movement interval. In a) and c) during pre-holding and post-holding higher variances were observed than during holding (b).

After that variances in each normalized time step of the ten trials per load condition were calculated and averaged for 20 subjects. Based on the methods of A) and B) we composed averaged variances of the end point (7th marker) (Figure 3); arm configuration variances (Figure 4); computed muscle force variances of 4 arm muscles (BI, TR, DA, DP) during holding. (Figure 5) End point variances

were determined by composing the sum of measured 3D coordinates in every normalized time step, while joint configuration variances were obtained by computing the sum of separately determined variances of joint angular changes respect to the normalized time. In A) and B) a two-sample t-test of the null hypothesis was performed at a 5% significance level ($p=0.05$), because data in the vectors were independent random samples from normal distributions. Results of A) didn't show any significance either in kinematics or in muscle forces. In the end point variances (Figure 3) during holding there was significant difference between the two object conditions. End point variances during uplifting in movements performed with 2 kg were significantly smaller $t=2.24$, $p=0.05$ than movement variances performed with CD case. However, during putting down there were no significant difference $t=0.9853$, $p=0.05$ between the two conditions but variance ratios of 2 kg load to CD case mostly remained smaller than 1. In joint configuration variances (Figure 4) there was no significant difference between movements with load or without load either during uplifting or putting down. Ratios of load to CD case remained only in half of the subjects less than 1. In computed muscle forces in flexors (BI, DA) (Figure 5) significant differences (Table 2) were observed between the two conditions either during uplifting or putting down while in extensors (TR, DP) (Figure 5) there was no significant deflection between the two conditions. Ratios of load to CD case were higher than 1 in all muscles.

TABLE 2
T PARAMETERS OF TWO-SAMPLE T-TEST ON MUSCLE FORCES

FLEXOR FORCE				EXTENSOR FORCE			
BI		DA		TR		DP	
U	D	U	D	U	D	U	D
3.50	4.59	2.31	2.44	0.92	1.55	0.51	0.10

Table 2 - During uplifting (U) and putting down (D). In values signed by bold the null hypothesis were rejected so there was significant difference between the variances of 2 kg object and CD case

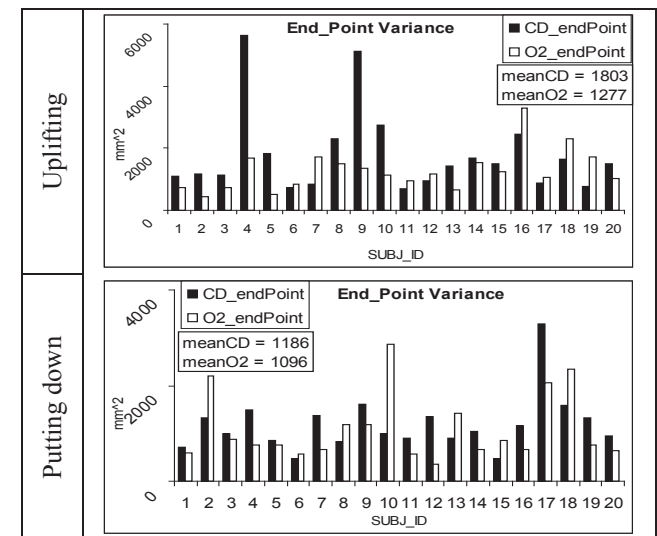


Figure 3 – (upper) End point (7th marker) variances under both conditions for all subjects (20) during uplifting. (lower) End point (7th marker) variances under both object conditions of 20 subjects during putting down.

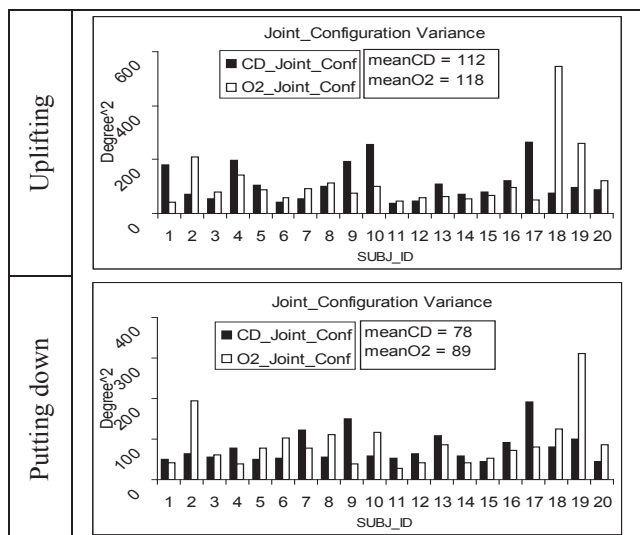


Figure 4 - (upper) Arm configuration variances under both object conditions for 20 subjects during uplifting. (lower) Arm configuration variances under both object conditions for 20 subjects during putting down.

IV. DISCUSSION

In arm configuration variances when the whole movement interval variances were averaged to all subjects (20) statistical methods didn't prove any significant difference between the two object conditions. This may be because averaged variances in the pre and post-holding parts of the actual movement during both uplifting and putting down showed much higher variability than during holding. (Figure 2) This suggests that movements performed with load varied less than movements performed without load. In muscle force variances (Figure 5) by the increase of load variances increased in both flexors and extensors. The range of increment of extensors is higher than in flexors suggesting that extensors generate greater force with higher variability than flexors. In the presented study the effect of fatigue was not investigated.

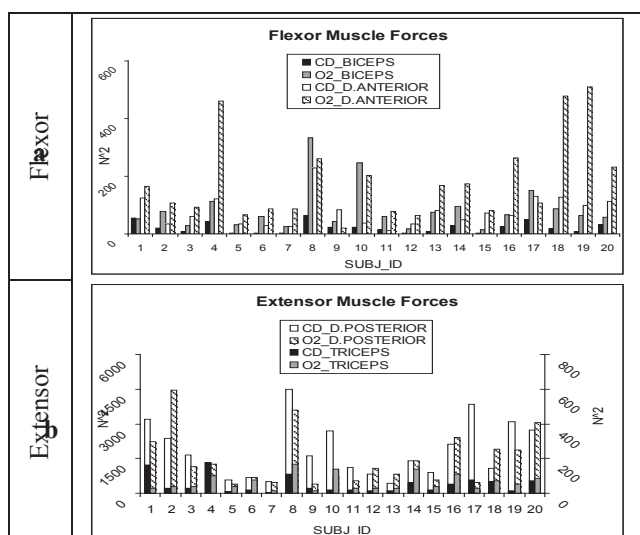


Figure 5 – In a) holding variances of flexors for all subjects under the two load conditions during uplifting. In b) holding variances of the extensors for all subjects during uplifting. In the left axis variances of TR, DP are presented under 2 kg condition; in the right axis CD case condition is presented.

During the measurements the influence of fatigue was avoided by randomly selecting the object to be moved. Furthermore the difference between uplifting and putting

down may be neglected though results of this are not presented in the current study. We conclude that the load helps to stabilize the movement at kinematic level of controlling mostly through by the hand position and less through by joint configurations. However, the stabilization effect of joint rotations is not the result of changing individual muscle activities produced by the 4 arm muscles separately but it's due to the increasing muscle synergies. The kinematic stability of the human arm movements executed with load is the consequence of neural control of multi-muscle systems. Such control principles would be advantageous in artificial control of neuroprostheses and robot arms.

REFERENCES

- [1] Laczko J, Walton K, Watanabe H, Llinas R "Modeling of limb movements to compute and transfer stimulation trains to spinal motoneuron pools". In: *12th annual Conference of the Intl. FES Society*. Philadelphia PA. Session 10.,2007
- [2] Pilissy T., Klauber A., Fazekas G., Laczko J.,Szécsi J., "Improving functional electrical stimulation driven cycling by proper synchronization of the muscles", *Clinical Neuroscience/Idegyogy Szle*. Vol. 61(5-6),pp. 162-167, 2008.
- [3] Szécsi J, Fincziczki Á, Laczko J, Straube A. "Elektrostimuláció segítségével meghajtott (háromkerekű) kerékpár: Neuroprotézis harántsérült páciensek mindennapos használatára." *Rehabilitáció*, 15. Évf. pp.9-14., 2005
- [4] Szécsi J,Fiegel M, Krafczyk S, Straube "A Smooth pedaling of the paraplegic cyclist – a natural optimality principle for adaptation of tricycle and stimulation to the rider", *J.Rehabil. Res. Dev.*,41 Supp.2: 30,2004
- [5] Laczko J., Walton K., Llinas R , "A neuro- mechanical transducer model for controlling joint rotations and limb movements," *Clinical Neuroscience/Idegy Szle*, 59(1-2), 32-43, 2006
- [6] Laczko J., Walton K., Llinas R., "A model for swimming motor control in rats", *Abstract Viewer Program No. 493.11.*. Washington, DC: Society for Neuroscience, 2003
- [7] The website of the U.S based company Cleveland F.E.S Center, <http://fescenter.case.edu/site2/index.php>
- [8] Ronald L. Hart, Kevin L. Kilgore, and P.Hunter Peckham,"A comparison between control methods for implanted FES hand-grasp systems", *IEEE Trans.Rehabil.Eng.*, vol.6., no 2, 1998
- [9] M.Wieler, Z.Kenwell, M.Gauthier, G.Isaacson, and A. Prochozka, "Electronic glove" augments tenodesis grip and hand opening in people with quadriplegia," *Physiother.Canada*,vol.46, pp. 94, 1994.
- [10] S.Saxena,S.Nikolic, and D.Popovic, "An EMG-controlled grasping system for tetraplegics," *J.Rehab. Res. Deve-lopment*, vol. 32, pp.17-24
- [11] R.H. Nathan, "Control strategies in FNS systems for the upper extremities," *Crit.Rev.Biomed.Eng.*, vol. 21, pp. 485-568, 1993
- [12] P.H. Peckham and J.T.Mortimer,"Restoration of hand function in the quadriplegic through electrical stimulation.Principles and preliminary experience, " *Functional Electrical Stimulation: Applications in Neural Prosthesis*, pp. 83-95, 1977
- [13] B.Smith, P.H. Peckham, M.W.Keith, G.B.Thrope and D.D.Roscoe "An externally powered, multichannel, implantable stimulator for versatile control of paralyzed muscle", *IEEE Trans.Biomed.Eng.*, Vol. 34, pp. 499-508, 1987
- [14] R.Tibold, "Relation of Muscle Activities and Joint Rotations in Reaching Arm Movements",*Proceedings of the Multidisciplinary doctoral school Faculty of Information Technology2007-2008*,pp 45-48,2008
- [15] Vladimir Zatsiorsky, "Kinematics of Human Motion", Publ Human Kinetics, Champaign IL, 2008
- [16] HEJ.Veeger, Bing Yu, Kai-Nan An, "Parameters for modeling the upper extremity", *J.Biomechanics*, Vol 30. No.6 pp647-652,1997

Phonocardiography in Preterm Newborns with Patent Ductus Arteriosus

Ádám Balogh

(Supervisor: Dr. Ferenc Kovács, Dr. Tamás Roska)

baladta@digitus.itk.ppke.hu

Abstract—This paper presents a pilot study investigating the usefulness of phonocardiography in assessing the hemodynamics of the patent ductus arteriosus (PDA) in preterm newborns. Nineteen infants have been examined, 11 with hemodynamically significant PDA verified by echocardiography. Four of these newborns have been measured every day until the clinically verified closure of the PDA occurred. Continuous murmur, as the hallmark of PDA in adults was not found but some low frequency ejection-like murmur was found which might indicate the state of the ductus arteriosus. Comparison of the heart sounds (mainly the first heart sounds) before and after the closure of the PDA is presented. Further sophisticated analysis, especially regarding the second heart sounds is recommended for statistical verification of the usability of phonocardiography concerning PDA in preterm newborns.

Index Terms—phonocardiography, patent ductus arteriosus, preterm newborns

I. INTRODUCTION

The ductus arteriosus is an essential fetal vascular structure which connects the main pulmonary artery with the descending aorta. It shunts the blood coming from the right ventricle into the aorta due to the high resistance of the pulmonary circulation. Closure during pregnancy may lead to right heart failure. After birth, with the first intake of breath the lungs expand and the resistance of the pulmonary circulation decreases greatly allowing the development of the normal human circulation. Apparently the ductus arteriosus loses its purpose, moreover, the persistence of the ductal patency is abnormal. Under normal conditions functional complete closure occurs within 24 to 48 hours after birth [1].

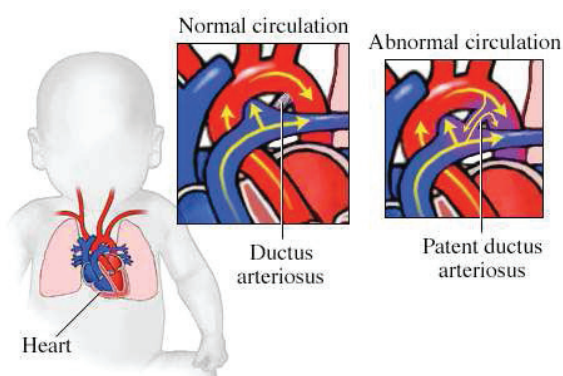


Fig. 1. Heart of a newborn with and without patent ductus arteriosus. In normal condition there is no blood flow through the ductus arteriosus after the first 24 to 48 hours of life [2].

The physiological impact and clinical significance of a patent ductus arteriosus depends mainly on its size and the status of the underlying cardiovascular system. After birth, in case of PDA a left-to-right shunt evolves due to the higher pressure in the aorta. This means an increased pulmonary fluid volume which may cause respiratory problems. Also the left atrium and ventricle have to compensate the increased fluid volume returning from the lungs and the "pressure leakage" in the aorta which may cause hypertrophy of the left atrium and ventricle.

Symptoms for physical examination are continuous murmur, located at the upper left sternal border, and bounding peripheral pulses due to the rapid decrease of the diastolic pressure through the ductus. That means that there is a greater difference between the systolic and diastolic blood pressure (eg. in case of neonatals 2:1 ratio instead of 3:2).

The PDA can be "silent" and depending on the size of the PDA, it can be classified as small, moderate or large. Very small, tiny PDAs may result in no abnormal physical findings but may cause problems in later age with other malformations.

The closure of the PDA may occur spontaneously or due to a surgical or transcatheter intervention. In case of preterm infants pharmacological closure is also possible [3][4].

In case of preterm infants the risk of PDA is clearly much greater [1] which is due to physiological factors related to prematurity. Unfortunately the diagnosis and the assessment of hemodynamical significance is not obvious [5]. This paper addresses exactly that point, namely whether phonocardiography is a useful tool in diagnosing and monitoring PDA in preterm newborns especially in the first days of life.

II. MATERIALS AND METHODS

In this section the measurement scenario and the analysing methods are described.

A. Measurements

In this pilot study 19 preterm newborns have been examined, with an average of 2 measurements per infant, but with large deviations: only those newborns were examined several times which were diagnosed with PDA, those without PDA or with other cardiac malformations only once. Hemodynamically significant PDA was verified by echocardiography in case of 11 infants but only 4 of those were examined over several days because the others had either also some other malformations or some other circumstances made further measurements not

possible. In case of the 4 newborns mentioned above the PDA was closed by means of pharmacological treatment.

Each measurement consisted of about 3 one minute long phonocardiographic records. These infants, except one, all weighted less than 2300 g at birth, with an average weight of 1600 g. Except one, all of them were less than 33 weeks of gestation, with an average of 30. They were examined on average on their 4th day after birth and those with PDA then every day until the complete closure of the PDA verified by echocardiography (the maximum was 8 measurements on one infant). Three measurements had to be posteriorly excluded from the study because of the poor quality of the records since the measuring equipment was also developed during the study.

B. Measurement equipment

The first measurements were recorded with a simple electret microphone capsule placed on the thorax of the infants and connected to a laptop. The different types of microphones used are listed in Table I. The main advantage of this scenario is the small size of these microphones as it is easy to place them on different places on the infant's thorax.

TABLE I
PARAMETERS OF USED MICROPHONES

	Dynamic range (Hz)	Sensitivity (mV/Pa/1kHz)	Diameter (mm)	Height (mm)
MCE-100	20-10k	5.6	9.7	6.5
MCE-4000	20-20k	5	6	5.8
MCE-2002	20-16k	5.6	6	2.5

Unfortunately no electronic stethoscope for infants can be found on the market but it became obvious later that measurements of much better quality can be achieved even with a microphone placed in the bell of a conventional stethoscope for infants. Unfortunately, with this configuration it is often difficult to place the bell of the stethoscope at the main auscultation sites due to the small size and big curvature of the thorax of the preterm newborns. All the measurements were made with this stethoscope except for the first 6 infants. The heart sound was recorded with the bell placed on the aortic area.



Fig. 2. The bell of the stethoscope for infants

The measurements were recorded at 44 kHz and with a resolution of 16 bits. After prefiltering with a 2nd order Butterworth bypass filter with a bandwidth from 30 to 400 Hz,

the data was resampled at 1200 Hz and only the useful part of the record (length varied from 10 s to 1 min) was kept for further analysis.

C. Analysing methods

The main goal was to investigate certain parameters of the heart sounds and, if present, of murmurs which could be related to some attributes of the ductus arteriosus (eg. width, velocity of blood flowing through it, etc.). It seemed obvious to investigate a *typical* heartcycle for the given record. This was achieved by calculating a weighted average heart cycle assuming the following:

- an average heart sound can be calculated from similar heart cycles for at least the time of the measurement if no big change occurs in the cardiovascular system (often the same envelope was present over several days)
- murmur caused by a given anomaly will appear approximately at the same point in the heart cycle with similar envelope
- according to these measurements the usually occurring noise (eg. from breathing machines) can be regarded as white noise in the investigated frequency range and thus averaging will suppress it

Due to the above mentioned aspects averaging seemed useful in enhancing the characteristics of the heart sounds and murmurs. The method is outlined in Fig. 3.

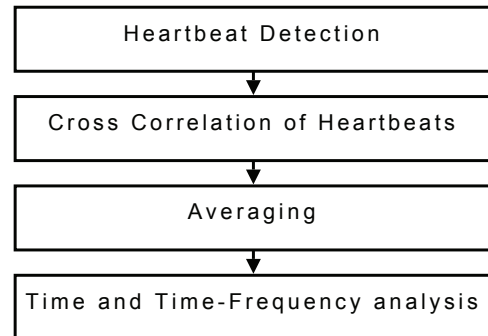


Fig. 3. General scheme of the analysing method

1) *Heartbeat detection*: The detection of the heart cycles and heartbeats was done with a heuristic method developed for fetal phonocardiography [6] with very good accuracy.

2) *Cross Correlation of the Heartbeats*: For the comparison of the different heartbeats cross correlation was used. According to our measurements this was calculated with a 100 ms window for S1-s and with a 66 ms window for S2-s (Fig. 4). For each heartbeat the normalized cross correlation coefficient with all the other heartbeats was calculated. A threshold of 0.9 (in case of poorer quality data 0.8 or 0.85) was used for selecting the most typical heartbeat for the given record: that heartbeat was selected which had normalized correlation coefficients greater than the mentioned threshold with the greatest number of other heartbeats.

Temporal alignment was also performed with cross correlation, namely by maximizing the cross correlation between each heartbeat.

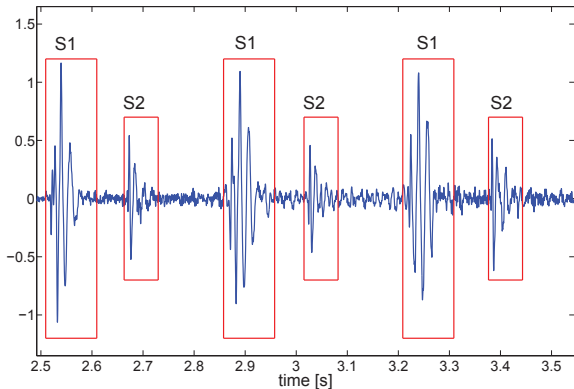


Fig. 4. Time windows used for extracting heart sounds for cross correlation

3) *Averaging*: The weighted average was calculated not only for the heartbeats but for the whole heart cycle, revealing potential murmurs, in the following manner:

$$\hat{b}_j[k] = \sum_{i=1}^J R_{ji} \cdot b_i[k - o_{ji}] \quad (1)$$

where \hat{b}_j is the average beat for the selected beat j , J are the number of beats similar to beat j with a normalized correlation coefficient greater than the given threshold, R_{ji} is the maximal normalized correlation coefficient between beat j and beat i , producing the offset of o_{ji} .

4) *Analysis*: The analysis of the typical heartbeats for a given record was done in the time and in the time-frequency domain due to the nonstationary nature of heart sounds. Short-time Fourier Transform was used to compute the time-frequency representation:

$$S_{\hat{b}_j}[k, f, h] = \sum_{l=-\infty}^{\infty} \hat{b}_j[l] \cdot h[l - k] \cdot e^{-j2\pi fl} \quad (2)$$

where h is the short-time analysis window. The optimal duration of the time-window used to compute time-frequency representation of phonocardiograms is between 16 and 32 ms [7]. Here a 27 ms long Hamming window was used, window shifting was 1 ms.

The analysis was mainly done by visual inspection of the envelope and the spectrogram looking for components which could be related to some attributes of PDA. This means that the length, the envelope and the frequency components of the heartbeats, presence of split between the aortic and pulmonary heart sounds, murmurs, abnormal sounds and their frequency relationship were investigated.

III. RESULTS AND DISCUSSION

As described in [8], preterms with PDA develop murmur only from the third day after birth and a continuous murmur like that of adults with PDA is also rare. Thus the close

inspection of the first and second heart sounds is of great importance. According to the advise of the paediatrician *blowing* heart sounds were searched and investigated which might be caused by a weak ejection murmur or due to some other consequence of PDA. Such heart sounds can be observed in Fig. 5.

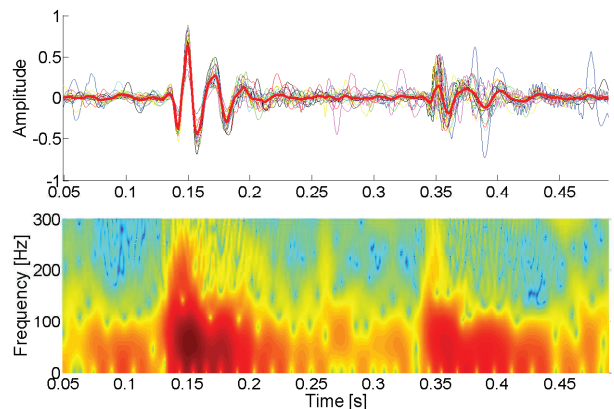


Fig. 5. Average heart sounds of an infant with PDA (thick red line in the top figure, average of 21 heart cycles). These are typical *blowing* heart sounds characterized by the slow attenuation especially seen in case of the second heart sound.

First of all, the consecutive measurements of infants with PDA were investigated to find those properties of the heart sound which disappear with the closure of PDA. For this the average beat of selected records from different measurements were compared.

Unfortunately, with the method described in section II. usually only average first heart sounds (S1) are extracted because they are less suppressed by noise due to their greater amplitude compared to second heart sounds (S2). For calculating average second heart sounds an improvement of the heartbeat detection algorithm is needed.

The comparison of S1-s before and after the closure of the ductus arteriosus revealed that the first heart sounds became more concentrated (faster attenuation) and the amplitude of the negative half waves increased (see Fig. 6). Unfortunately this is not always true. This investigation also revealed that heart sounds remained similar for several days, thus comparison is possible. Good results have been achieved by low-pass filtering the S1 heart sounds (Fig. 7), but further investigations are needed for this analysis.

From the time-frequency analysis one could make the assumption that the systole becomes more quiet after closure of PDA (Fig. 8), but it is difficult to introduce an objective measure for this due to averaging of different numbers of heart cycles.

The results so far indicate that the assessment of PDA is not a trivial signal processing task. Heart sound parameter extraction is needed (eg. [9]) for statistical verification of the investigation method and further analysis is recommended with a model based analysis, eg. [10], or with an analytic approach, eg. [11].

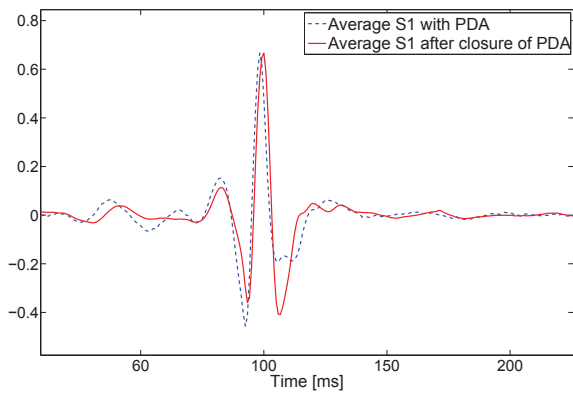


Fig. 6. Average S1 of an infant with PDA (dashed blue line, average of 82 heart sounds) and after closure of the PDA (red line, average of 69 heart sounds). The heart sound became more concentrated and faster attenuation is observable.

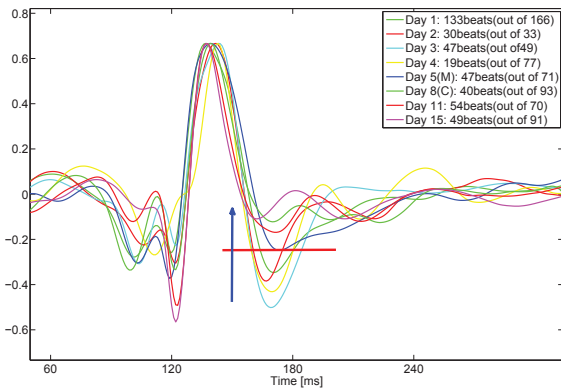


Fig. 7. Average S1-s of another infant, calculated after low-pass filtering the data with a 2nd order Butterworth filter, cutoff frequency was 30 Hz. Heart sounds above the thick red line are from records after the closure of the PDA.

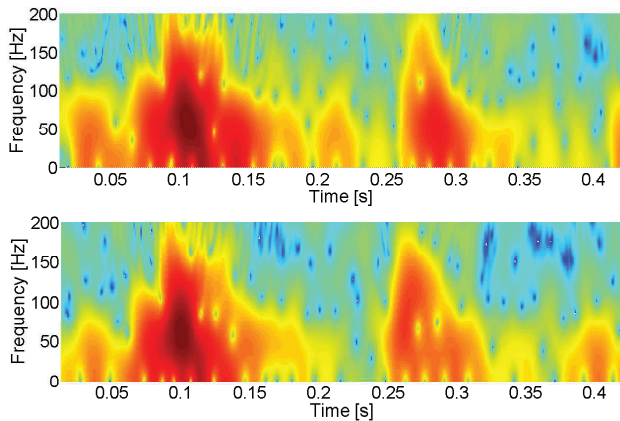


Fig. 8. Spectrogram of average heartcycle of an infant with PDA (top, average of 127 heart cycles) and after closure of the PDA (bottom, average of 74 heart cycles). It is observable, that the systole became more quiet after closure. Unfortunately not always true.

IV. CONCLUSION

The main goal of this pilot study was to investigate whether a satisfactory phonocardiography method can be developed for

sensitive and specific assessment of PDA in preterm newborns in the first days after birth. From the measurements so far it can be concluded that there is a difference between the heart sounds with PDA and the ones recorded after the closure of the ductus arteriosus. For reaching the main goal of this study further sophisticated analysis is needed, especially regarding the second heart sound.

ACKNOWLEDGMENT

The author would like to thank Dr. Zoltán Molnár and Dr. Miklós Szabó from the I. Department of Pediatrics, Budapest, for the measurements and for their assistance and advice.

REFERENCES

- [1] D. J. Schneider and J. W. Moore, "Patent ductus arteriosus," *Circulation*, vol. 114, pp. 1873–1882, 2006.
- [2] Nucleus Communications, "Illustration of the patent ductus arteriosus," 2000. [Online]. Available: <http://64.143.176.9/library/healthguide/en-us/support/topic.asp?hwid=tp12700>
- [3] B. van Overmeire et al., "A comparison of ibuprofen and indomethacin for closure of patent ductus arteriosus," *The New England Journal of Medicine*, vol. 343, no. 10, pp. 674–681, 2000.
- [4] D. B. Knight, "The treatment of patent ductus arteriosus in preterm infants, a review and overview of randomized trials," *Seminars in Neonatology*, vol. 6, no. 1, pp. 63–73, 2001.
- [5] N. Evans, "Diagnosis of patent ductus arteriosus in the preterm newborn," *Archives of Disease in Childhood*, no. 68, pp. 58–61, 1993.
- [6] E. Kósa, A. T. Balogh, B. Üveges, and F. Kovács, "Heuristic method for heartbeat detection in fetal phonocardiographic signals," *Signals and Electronic Systems, 2008. ICSES '08. International Conference on*, pp. 231–234, 2008.
- [7] G. Jamous, L.-G. Durand, Y. E. Langlois, T. Lanthier, P. Pibarot, and S. Carioto, "Optimal Time-Window duration for computing time frequency representations of normal phonocardiograms in dogs," *Medical and Biological Engineering and Computing*, vol. 30, no. 5, pp. 503–508, 1992.
- [8] R. Skelton, N. Evans, and J. Smythe, "A blinded comparison of clinical and echocardiographic evaluation of the preterm infant for patent ductus arteriosus," *Journal of Paediatrics and Child Health*, vol. 30, no. 5, pp. 406–411, 1994.
- [9] A. Voss, A. Mix, and T. Hübner, "Diagnosing aortic valve stenosis by parameter extraction of heart sound signals," *Annals of Biomedical Engineering*, vol. 33, no. 9, pp. 1167–1174, 2005.
- [10] M. C. Agostinho and M. N. Souza, "A new heart sound simulation technique," *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE*, pp. 323–326, 1997.
- [11] X. Jingping, L.-G. Durand, and P. Pibarot, "Nonlinear transient chirp signal modeling of the aortic and pulmonary components of the second heart sound," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 10, pp. 1328–1335, 2000.

Colors and Color Perception

András Gelencsér

(Supervisor: Dr. Tamás Roska)

gelan@digitus.itk.ppke.hu

Abstract— Every day we meet the colors of our world. They are everywhere, on our clothes, on books, on the traffic signs, on the monitor screen etc. Most of the people find it evident. But sense and understand the colors are not a trivial task. To reproduce a true world color with artificial means(e.g. with screen, printer, camera, projector), and to detect, classify colors, process colored pictures could be very hard, if we want natural committed result. In this paper I try to explain the working of the mammal retina. Only after we understand our own vision, especially our color vision, can we construct proficient instruments and develop efficient algorithms, which works with colors.

Index Terms — color vision, color principles, mammal retina, receptive field, single-opponent, double-opponent

I. INTRODUCTION

At the first step of the extensive world of colors I try to collect the basics of the science region. I find very emphasis to understand the functionality of our eye, so I make a relative big stress in this paper to analyze the mammal vision system, particularly considering the color vision. In the third section I give a brief summary of color principles to understand how can we describe a color, what is a color after all.

II. THE RETINA

The developed species collect information from the surrounding world with the five sensory organs: eyesight, hearing, taste, smell and touch. The most important perception of the majority of animals (aside from several e.g.: moles, bats) is the visual sensation. In this section I make a review of the main features of the mammal retina.

A. The Eye

The first station of the visual information towards the brain is the eye. This is the first building block of the visual system. It's task to collect and project the light to the retina, which detects the incoming photons, and generates optic sense. The light is focused by two lenses and two fluid mediums. The iris regulates the amount of the incident light to prevent the saturation of the light sensitive cells due to too much photon. The iris also can improve the acuity of the image on the retina. Similar thing happens when someone take a picture with a camera. (figure 1.)

B. The build-up of the retina

The retina is part of the central neural system that sends visual information from the eye to the brain. It is impressive how efficient it is in capturing and relaying as much visual information from the world as possible, and how under great

range of conditions can it work. We can see from dazzling sunlight to dark star lighted night.

The output of the retina consists of many parallel purpose-specific signals of the same visual input. The primate retina can be divided for two distinct areas. The central part is the fovea. This has high spatial acuity and is responsible for most visual tasks. The second part is the peripheral part, the peripheral retina. This is important for the vision of night time. The diameter of the fovea is very small, only a few degrees of the visual angle.

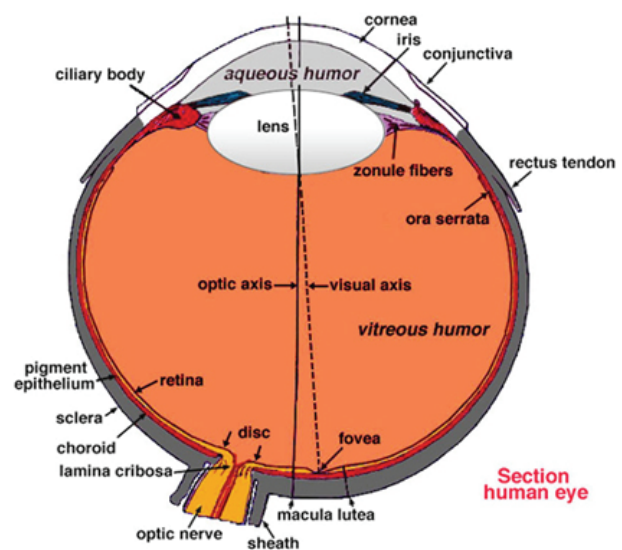


Figure 1. The detailed cut of the human eye.

Biological researches found out in the mammal's retina contain 50 anatomical cell types. All of them have different function. The neurobiologists speak about 10 different anatomically layers in the retina. We don't need such a detailed picture from this organ. So we use only a simplified view of the retina, with 5 cell class and 5 level layers. To understand the main functioning and the system of the retina it is more than enough.

The layers determine the similar parts of the retina. Each layer consists of the same cells or the similar connections of the cells. It is a sort of classification of the structure. The main layers are the followings: ganglion cell layer, inner plexiform layer, inner nuclear layer, outer plexiform layer, outer nuclear layer. There is another important layer the pigment trace layer. The photoreceptor cells are in the outer plexiform layer. It should be noted that the first layer from the view of incident light is the ganglion cell layer. Although the first four layers are transparent the light must travel through them, to reach the receptors. I will give a little detailed description while I display the cells of the retina. (figure 2.)

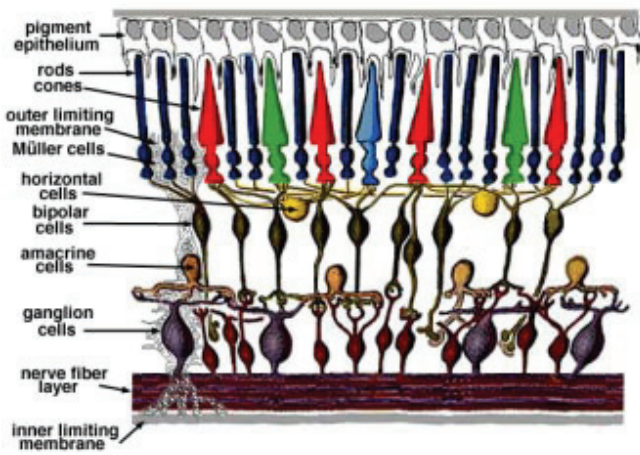


Figure 2. The detailed cut of the human retina.

The main cells in the retina are the followings: ganglion cells, amacrine cells, bipolar cells, horizontal cells, photoreceptor cells. The ganglion cells are the main and only outputs of the retina. Their axons create the optic nerve, which is heading to the lateral geniculate nucleus. The amacrine cells converges signals from the peripheral rod cells (a sort of receptor cell) and propagate it to the ganglion cells. Bipolar cells are the connections between the photoreceptor cells and the ganglion cells. Horizontal cells pool signals from the cones (a sort of receptor cell) and give it to the bipolar cells. They determine how many receptors each ganglion cell can see. The photoreceptor cells collect the incoming photons; their photo pigment is rhodopsin. Roughly the tierce of the incident light absorbs here. (If the receptor cells were longer or more receptor layers had existed, then more light could absorb. But such a structure cause more difficulties in extracting and propagating the correct visual signals. In nocturnal animals there is a reflective layer behind the photoreceptors to help capture even more photons). The photoreceptor cells have two main types: the cones and the rods. As stated above, there is another cell, the pigment cell. Their spurs penetrate among the photoreceptors. If the light is too strong, then pigment granules migrate into the spurs, so they can lower the amount of the incident light.

C. The photoreceptor cells

Cones operate under photopic conditions. That means they work only in bright, when much light is available. So the cones are “blind” in dark, in night. They can sense the colors of the world. Cones are divided into L, M, S cones, each being most sensitive respectively to long-, medium- and short-wavelength light. Other way they are divided to R, G, B cones. In humans the three cones are most sensitive to the 565nm, 530nm or 420nm wavelength light. So they are called Red, Green or Blue cones due to the wavelengths. The cones are less sensitive then rods, but send signals more often.

The rods operate under scotopic conditions. That means they work in dim light or in dark. Rods are insensitive for colors; they can sense the brightness of the light (in grayscale). The rods can detect even one photon.

The cones are dense at the retinal centre (approx. $160,000/\text{mm}^2$), but become sparse with eccentricity. The same types of cones are hexagonally close packed in the retina. But the instance of the cones doesn't follow any regular pattern. Generally, in the fovea the red cones are in majority and we can hardly find blue cones. The concrete cone ratio is different in every people, so everybody see the colors of the world a little bit different.

Rods don't exist in the fovea, but with eccentricity their number grows. The more the distance is from the fovea the more the diameter of the cone and rod cells become bigger – so the cell density decreases. In the peripheries the cone cells are bigger than the rod cells. (figure 3.)

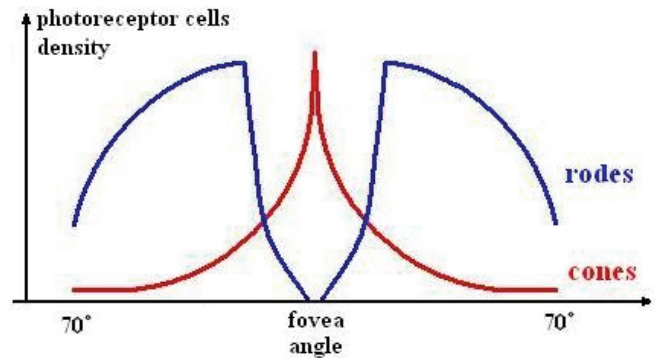


Figure 3. Photoreceptor density in the fovea

In the fovea the retina tails away. The ganglion and the bipolar cells stand not like columns, but lean out radial wise from the center. So the light can directly hit the cone cells, which results more clean and fine picture from the vision field. Because of the above described structure, contrast perception, acuity decreases with eccentricity. It is important to mention, acuity shows an increase with logarithmic light intensity of course with saturation.

The center of the retina is the foveola. The diameter of the foveola is approximately 100 cone cells (approx 1 degree in visual angle). It is interesting, that the B cones are totally missing here, only red and green light detection happens. The R and G cones not arranged regularly in the 2-D foveal space, they separately create clusters instead. The population rate of the cells in the foveola in primates is cone:horizontal:bipolar:amacrin:ganglion = 1:1:1:1:1.

D. Information flow in the eye

There are three main information directions in the retina: feed backward a feed forward and a lateral. We see an example for the first one, when the light propagates from the inner retina to the outer retina. The second one stands for the way of the detected visual information from the receptors, through the bipolar cells to the ganglion cells. Amacrine and horizontal cells pool information laterally from specific layers.

The cells in the retina use two modes to transmit their signals. One mode is the analog way. The photoreceptors, bipolar cells and horizontal cells produce graded changes in their potential under stimulus. This permits fast and continually signal flows. (E.g. there is a change in the brightness.) All of the ganglion cells and several amacrine cells produce action potentials under stimuli. This mode is much slower, but more robust, allows long-term information

travel. The ganglion cells carry the analog signals of the bipolar cells with discrete frequency coding, to the lateral geniculate nucleus and to the superior colliculus. Individual amacrin cells collect information from relative large area of the retina, accordingly using the second mode. Every cell has a base potential. It is interesting, that, if the photoreceptors detect light, they hyperpolarize (the voltage inside the cell drops), if they are in darkness, they depolarize (the voltage inside the cells rises). Thus dark acts like stimulus.

There are two physiological types of bipolar and ganglion cells. ON cells show stronger activity in case of stimuli, OFF cells reverse. They drop their activity if they are stimulated. These cells play a significant role in the preprocess of visual sensation.

E. The receptive field

The retina contains roughly 100 million photoreceptor cells, but only 1 million ganglion cells. Only a small part of the eye can produce detailed vision, because of this reduction.

The ganglion cell is only activated, if the light stimulus on the retina is near to the position of the ganglion cell. With other words, there is an area on the retina, which in case of getting light stimuli, then a particular ganglion cell changes its activity. This area is called the cells receptive field. This field shows which cones and rods are connected to a ganglion cell. The ganglion activity can be increased (excitatory) or decreased (inhibitory change). The receptive field of the ganglion cells is like a circle. It has two part: a central and a peripheral portion. Two major types of receptive field exist in the ganglion cells: ON centre, OFF surround. It can measure relative brightness. (How much brighter an object is in contrast to the background.) OFF centre, ON surround. It can measure relative darkness. (How much darker an object is in contrast to the background.) ON centre cells increase their fire rate if light touches the RF center, and decrease if light touches the RF surround. An OFF cell works reversed. This structure is called antagonistic surround. The ganglion cell gives the greatest response if the light completely fills the ON region. These ganglion cells get their inputs from ON/OFF bipolar cells. The bipolar cells produce signals from a small patch of the retinas photoreceptors. (In the fovea of primates one bipolar cell contacts only one photoreceptor). Central receptors give the ON area, surround receptors the OFF area. (Horizontal cells collect the surround information.) This structure can emphasize the significant point of the visual field, which differs from the surroundings or undergoes some changes.

There is a special type of ganglion cell. If its receptive field centre detects light, then the cell begin fire rapidly for a short period of time. If the light breaks off, then the cell is blocked. These cells are ideal for detecting quick changes, like flashing lights.

The antagonist surrounds of the ganglion cells have some more very important function. For example the edge detection (the retina watch the contrasts) or the constancy. Constancy means that the retina can ignore the change of the brilliance. For example black letters on a page should look black both indoors or outside. The white paper indoors reflects less light, so it is darker, than the black letters

outdoors, yet the white page does not turn black, because of constancy.

II. THE COLOR VISION

What is the advantage of color vision? Why do we see colors after all?

Cones respond better a particular wavelength of light than to others. With good approach, it is like a Gaussian function.

Some of the species see colors others not. In the animal world there are two, three and four color vision and color blindness too. The primates have trichromatic vision, some birds have tetrachromatic vision - they can see even ultraviolet.

Monochromatic vision system is not always enough, to distinguishing an object from its background. For example: we have two different colors and one cone. These colors not the specific colors of the cone, but each color gives the same response on the receptor. (figure 4.)

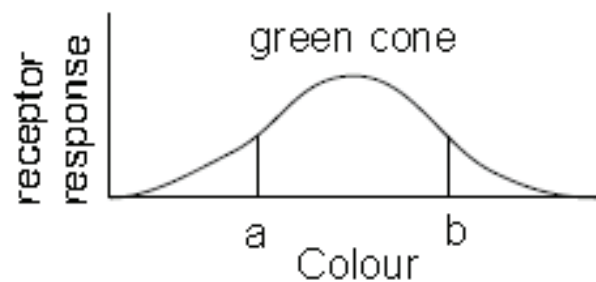


Figure 4. Spectral sensitivity of one cone. One photoreceptor type is not enough for distinguish obviously the color space. "a" and "b" are different colors, but they give the same receptor response.

The light, which reaches our eyes, is dependent on two things: the spectrum of the illuminant (e.g. white light), and the selective reflectance properties of the surface.

Most species have evolved at least a two cone system, but in some cases, it can be fooled too. Dichromatic vision could work well to uniquely define every colored surface, if these surfaces were always illuminated by light of constant spectral composition. Two different wavelength of the arriving light could be used to distinguish all intrinsic color. Because natural daylight varies slightly from composition of sunlight (being greenish in under the forest canopy, and reddish at sunset) one extra color dimension suffices. However, humans have invented light sources which have spectral compositions that vary greatly from sunlight, so we have difficulty in recognizing the true colors in some cases.

A. Opponent cells

We can distinguish three main types.

Non-opponent cells: The center and surround contain the same combination of cone types. The influence of the center is the opposite of the surround. These cells are good at detecting a change in brightness between the center and surround, but they insensitive to a change in color.

Single opponet cells: Here the center and surround contain different cone types and their influence is opposite.

Double opponent cells: This cell type is not found uoin he eyebut in the part of the cerebral cortex that receives information from the eye. Its receptive field is formed by combining information from non-opponent ganglion cells. These cells are more selective for a particular color.

B. Color blindness

As stated before each cone type contains a different light sensitive photo pigment. Color blindness in humans occurs when there is a defect in the genes that produce these photo pigments. Various combinations of defects can occur. Missing one or more cones, or a photo pigment can be different than normal (shifts the cone sensitivity toward a different color).

At animals we speak about color blindness if they absolutely can't see colors. (For example, there is a myth about the dogs. They have two color vision, but most of the people think they are colorblind.)

C. The human brains color distinguish

With the three cone types human can distinguish 200 hues, 20 levels of saturation and 500 brightness levels. So we can see $500 \times 200 \times 20 = 2$ million gradations of color.

III. EXTRACTS FROM COLOR PRINCIPLES

Color is an extremely important part of most visualization. Choosing good colors for our visualizations involves understanding their properties and the perceptual characteristics of human vision. It is also important to understand how computer software assigns colors and various hardware devices interpret those assignments.

There is a narrow range of electromagnetic energy from the sun and other light sources which create energy of wavelengths visible to humans. Each of these wavelengths, from approximately 400 nm to 700 nm, is associated with a particular color response. (For example 400 nm are violet in color 700 nm are red.)

The solar radiation, which reaches the Earth surface, is the strongest in the visible light range. That explains, why the vision systems evolved this way. (figure 5.)

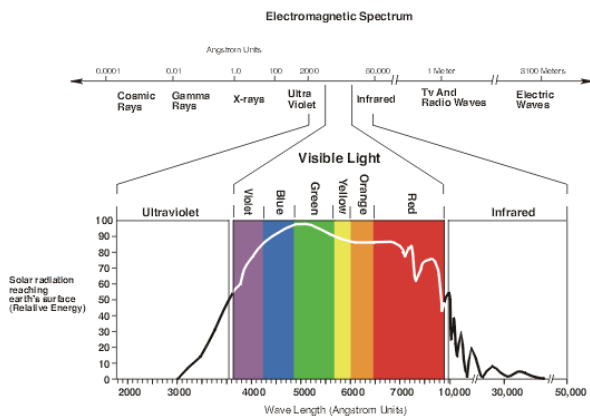


Figure 5. The electromagnetic spectrum. The curve demonstrates the power of the wave from the Sun, what reach the Earth surface.

Hue: The brain transforms the single wavelengths of light seen in a rainbow into a color circle. The opposite sides of the circle are complementary colors.

Saturation: refers to the dominance of hue in the colors. On the edges of the hue wheel are the pure hues. Moving towards the center of the wheel, the hue that describe the color dominates less and less. In the center the colors are desaturated, they become gray.

Brightness or value: How light or dark a color is. This parameter describes the overall intensity or strength of the light. (figure 6.)



Figure 6. Visual aid for the color wheels.

IV. CONCLUSION

The retina is not only a passive visual field acquisition and projection to the brain, but the first step of the light stimulus processing. It is an important station, where significant biological information is extracted from the world for the living being. The eye has three main tasks: it focuses the clear and fine image to the retina, it detects the light of various colors and intensities, it compresses the gained information in order to send it through the small bandwidth of the optic nerve.

Compression occurs in two ways: a detailed image is sent only from the fovea, a small part of the eye and only the most significant information are transmitted, like the edges, changes in color, brightness etc.

We can say that, only a poor reflection of the original image is sent to the brain, yet we perceive the world around us in extreme clarity, and this is a wonder of our brain!

REFERENCES

- [1] Semir Zeki, A Vision of the Brain, Blackwell scientific publications, Oxford
- [2] H. Momiji, A.A. Bharath, M.W. Hankins & C. Kennard (2006). Numerical study of short-term afterimages and associate properties in foveal vision, Vision Research Imperial College – London
- [3] K. Bumsted, C. Jasoni, A. Szél, A. Hendrickson (1997), Spatial and Temporal Expression of Cone Opsins During Monkey Retinal Development
- [4] Funkcionális anatómia III., Szentágothai János, Réthelyi Miklós, Medicina könykiadó rt. Budapest 2007
- [5] Megmagyarázzuk az emberi értelmet, Susan Greenfield Magyar Könyvklub Helikon kiadó
- [6] CIE standarts (International Commission on illumination)

Color Based Image Segmentation in a Water Supply Surveillance System

Balázs Varga

(Supervisors: Dr. Tamás Roska, Dr. Gábor Szederkényi)
varba@digitus.itk.ppke.hu

Abstract—In this paper I present a computational framework which is capable of performing image segmentation based on the color information of the input. The system uses mean shift segmentation which is followed by morphological post-processing in order to produce a binary mask which uniquely distinguishes the background from the foreground object present on the image. Numerical analysis has been done to evaluate the performance of the constructed algorithm and further adaptability for motion-based segmentation and image stabilization is considered.

Index Terms—Image color analysis, Image segmentation, Object detection

I. INTRODUCTION

Drinking water demands increase every year [1] making it more and more crucial to ensure mass production of water free from bacteria, dust and diseases. The current project aims to design an automatic water biology surveillance system which is capable of threat indication through real-time water analysis. The main blocks of this system are explained briefly in Figure 1.

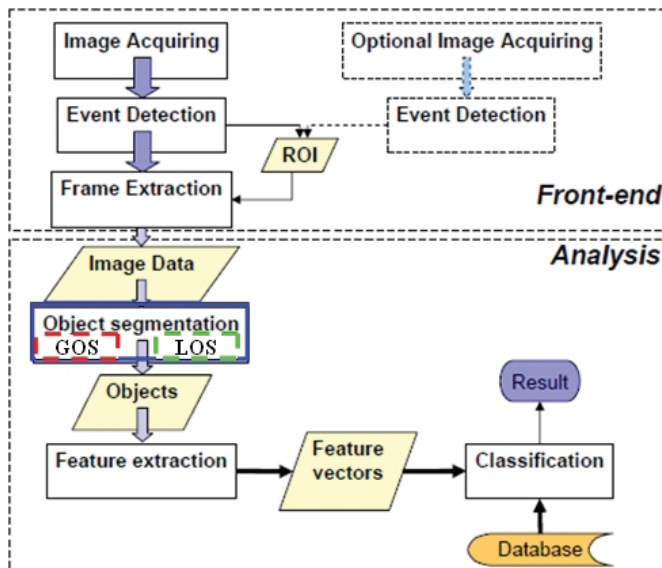


Fig. 1. Block diagram of the automatic water biology surveillance system (Thickness of the arrows represent amount of dataflow.)

Water monitoring is done in the domain of micrometers therefore a microscope is used to perform high-speed capture of the water flow (image acquisition). Next event detection is performed to indicate whether there are any *objects* (living organisms, such as algae, fungi, bacteria, etc) on a particular frame. Only those frames are selected and

considered as *image data* which contain information i.e. objects. Chosen frames go through segmentation which consists of two levels. First *global object segmentation* (GOS) is done during which each object on a given frame gets individually “cut” out by a rectangular *bounding box*. Next the exact boundaries of each individual object have to be calculated so that the image could be sent for feature extraction. This is the task of *local object segmentation* (LOS). LOS generates an n -by- m *binary mask* the exact size as the bounding box image in which a *zero-valued* (black) *pixel* unambiguously identifies a *background pixel*, and *one-valued* (white) *pixels* do likewise with the *foreground pixels*, which form the object (see Figure 2). This enables the feature extraction phase to consider only the object’s features without the background properties. This way the system can recognize the amount of certain objects in the water trough classification and e.g. indicate if any of them exceeds a given threshold.

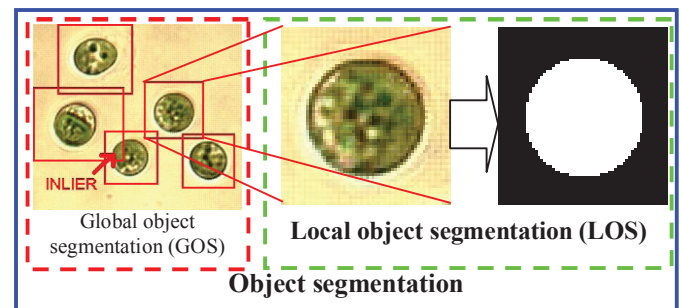


Fig. 2. Global- and Local object segmentation

GOS: The input frame (containing multiple objects) on left is segmented with bounding boxes. Notice that there’s another *inlier* object on one of the bounding box images.

LOS: An instance of a bounding box object is shown in the middle. The corresponding mask generated by the local object segmentation is shown on the right.

My task was to implement local object segmentation based on the color information of the input for which I’m using mean shift segmentation followed by morphological post-processing.

II. SYSTEM DESCRIPTION

A. Input image characteristics

Assumptions made concerning the input of the LOS (i.e. for the output of the GOS):

1. an object is present;
2. the background is quasi-homogeneous (due to ambient light setting of the microscope);
3. no. of background pixels is bigger than no. of

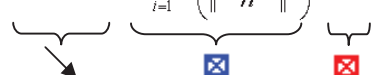
foreground pixels.

Main causes of possible errors:

1. multiple objects may be present (error of GOS);
2. *inlier object(s)* may be present (result of rectangle-shaped bounding box and small distance between objects – see Figure 2: GOS);
3. input may be noisy (microscopic measurement was incorrectly set).

B. Mean shift segmentation

Mean shift segmentation is a nonparametric clustering approach based on density estimation [2]. It considers the *feature space* as an empirical probability density function. Regions in which the density function is highly populated are called *modes* [3] and can be considered as the local maxima of the probability density function. The main idea of the system is briefly explained in the following: let $X = \{x_i \mid i = [1, n]\}$ be a set of samples taken from the feature space and let $G(x)$ denote the Gaussian kernel with profile $g(x)$. Also in the first iteration let x be a random element of X . In every iteration the algorithm calculates the $m_{h,G(x)}$ mean shift (a distance) as follows:

$$m_{h,G}(x) = \frac{\sum_{i=1}^n x_i g\left(\frac{\|x - x_i\|^2}{h}\right)}{\sum_{i=1}^n g\left(\frac{\|x - x_i\|^2}{h}\right)} - x \quad (1)$$


where h is the size of the Gaussian kernel, the fraction is the new mean, and x is the old mean. If $m_{h,G(x)}$ distance is smaller than a given threshold the iteration terminates and returns x which is a mode. The feature space *position* of the mode determines a *class* or cluster [4]. All other instances of the feature space which are in the ε region of this mode and yet are not in a class are associated with the mode's class, i.e. the ε parameter controls the smoothness of the algorithm. The iteration generates classes until all elements in the feature space are put into a class. The number of parameters the system directly utilizes is 3 (number of x_i samples, a threshold for $m_{h,G(x)}$, and parameter h controlling the size of the Gaussian kernel – also known as the bandwidth).

The main advantages [5] of this approach are that

- it needs no *a-priori* information about the number of classes making classification dynamical;
- sensitivity is controlled by basically one parameter (the bandwidth), which is a massive advantage;
- the feature space can be defined by almost any type of property if it is individual for every feature space element, making the feature space application dependent. (This is a very important property which I plan to exploit later on – see VI. Further work.)

The main disadvantages of this approach are:

- features with lesser support in the feature space may not be detected;
- the number of generated classes may blow up if the system gets stuck in false local maxima (detailed later in III. System evaluation, section C, remark 1).

The feature space in the case of this study was the *color*

space of the input image which adds interpretation-specific advantage: the segmentation is color-, shape-, orientation- and contrast independent.

The color space interpretation is shown on Figure 3 through an example. The left picture displays the representation of the middle image in Figure 2 in YCbCr color space; where each black element's spatial position is given by the corresponding pixel's color value. In this example it is straightforward that there are two dense regions: the yellowish background and the greenish region of the alga.

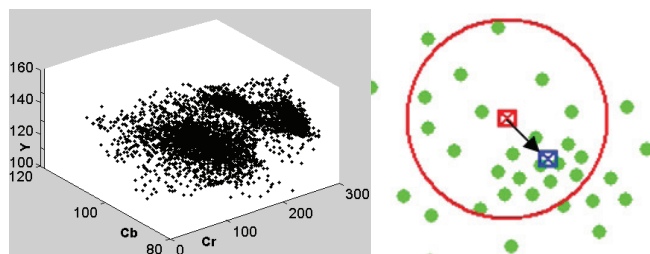


Fig. 3. Image representation and mean shift iteration
Left: Representation of a bounding box image in YCbCr color space (3D).
Right: A single mean shift iteration example in 2D. Spatial position of green dots is given by the color values of represented pixels. The red circle indicates scope of Gaussian kernel, red rectangle is the old position of the mean, blue rectangle is the new position of the mean, arrow indicates shift.

The right side indicates a mean shift iteration where the new mean position is calculated by (1). Pixels put into a class inherit color value of the class defining mean's color; which procedure can be considered as a quantization. Based on its result a binary mask is generated by exploiting assumptions about the properties of the input image.

The procedure is shown on Figure 4 (organism: *Codonella*).

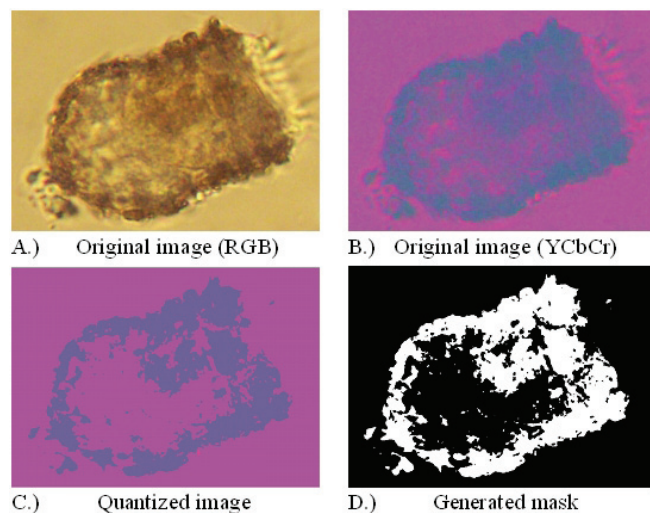


Fig. 4. Steps of the mean shift segmentation procedure
RGB image: given for illustrational reasons. **YCbCr image** is the input of the segmentation, image C.) displays the result of mean shift segmentation. Pixels of the image are put into different classes and from here a binary **mask** is generated.

The *generated mask* in Figure 4 clearly shows that the result of the segmentation cannot be used directly because several pixels in the objects body had been classified as background pixels; furthermore additional white blobs are present around the object. These errors are corrected by morphological operators during a post-processing phase.

C. Morphological post-processing

Morphological post-processing with basic-, [6] and complex

operators (filling and border cleaning [7]) is used to generate a homogeneous mask based on the result of the mean shift segmentation in order to obtain a proper LOS mask.

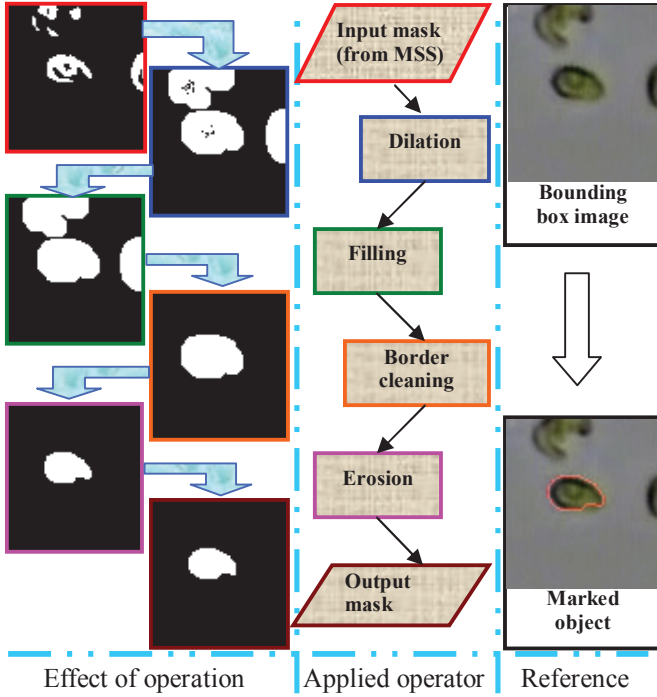


Fig. 5. Steps of the morphological post-processing procedure. The **Effect of operation** column shows the evolution of the input mask (which is the output of the mean shift segmentation) during the sequential application of the morphological operators (see column **Applied operator**). The **Reference** is only to show the original bounding box image based on which the mask was created, and an image on which the found object proportional to the generated mask is marked.

III. SYSTEM EVALUATION

A. Evaluation design: motivation

As it has been pointed the disadvantage of the mean shift approach is that features with lesser support in the feature space may not be detected. The representation of a given feature is a feature space dependent property meaning that the same image transformed into a different color space gives different results when analyzed with the system. The aim of the evaluation of the designed LOS method is to find the color space in which the best results can be achieved i.e. the most accurate mask can be generated.

B. Evaluation design: description

Five main color spaces had been selected for the analysis: *Lab*, *Luv*, *RGB*, *YCbCr* and *YUV* [8], [9].

Each input of the system is an n -by- m color image represented in Portable Network Graphics (.png) format. Its advantage is that it is lossless and has support up to a 24 bit color palette.

20 different sized bounding box images of 19 different algae types from 3 different realistic microscopic environment settings were selected and randomly divided into a *training set* and to an *evaluation set*; each carrying 10 images. The described LOS algorithm and the test environment was implemented in MATLAB 7.7.0 (MathWorks, Inc., Sherborn, MA), numerical analysis was done with MATLAB and with Microsoft Excel 2003 (Microsoft Corp., Redmond, WA).

C. Evaluation method

For all selected images a binary mask was made by hand in order to provide a *reference mask*.

10 masks were generated for each image in each color space. The reason multiple masks were made is that the LOS algorithm selects X samples arbitrarily. This results (5 color spaces \times 10 runs \times 10 images) = **500 generated masks in the training set** and (5 color spaces \times 10 runs \times 10 images) = **500 generated masks in the evaluation set**.

The system parameters for the different color spaces were set based on the training set's empirical analysis.

The output masks of the system were compared with the corresponding reference masks using the symmetric Hamming metric [10]. A measure of accuracy was calculated the following way: let $H_{I,CS,L}$ denote the *Hamming distance* for the L^{th} generated mask of image I in color space CS , where $L=[1\dots 10]$. Let $H_{I,CS}$ denote the *average Hamming distance* which is calculated as:

$$H_{I,CS} = \frac{\sum_{L=1}^{10} H_{I,CS,L}}{10} \quad (2)$$

From here $A_{I,CS}$ *accuracy measure* (measured in percentage) for the corresponding $H_{I,CS}$ average Hamming distance is given as:

$$A_{I,CS} = (1 - \frac{H_{I,CS}}{F_I}) * 100 \quad (3)$$

where F_I is the number of foreground pixels of the mask of image I . Also for a given CS color space $A_{CS} = \sum A_{I,CS} / 10$. Two remarks are made here:

1.) *Penalty mask*: in some cases the mean shift iteration resulted such classes into which little to non pixels were put simply because no other pixel of the image had similar color to the class defining mode. As a result the number of classes exceeded the manually set boundary of 9999 (note: regular no. of classes is approximately between 4 and 100, but a class number of 1000 might still give fair results). In this situation a completely black mask (*penalty mask*) was saved as the result to indicate failure and the algorithm stepped to the next mask's calculation procedure.

2.) *False identification*: a perfect match among the hand made-, and the generated mask gives 100% accuracy; and comparison of a hand made mask and a penalty mask gives 0% accuracy. But also negative A_{CS} percentage is possible, when the generated mask falsely identifies a set of pixels from the background region as the foreground object and these two regions doesn't overlap (are disjunctive). This phenomenon is referred as *false identification*.

IV. RESULTS

Table I summarizes the calculated values of performance measure A_{CS} concerning the different color spaces. The best results of a certain category are printed with bold letters. The most important column is the *Average* accuracy which represents the overall performance of the LOS procedure. The stability of the system is characterized by the standard deviation where a smaller deviation indicates how close the accuracy on individual masks was to the average. Results are displayed for both the test set and the training set and also the difference is given.

TABLE I
SYSTEM ACCURACY [A_{cs} (%)]

Name of color space	Minimum			Average			Maximum			Standard deviation		
	Training set	Test set	Difference	Training set	Test set	Difference	Training set	Test set	Difference	Training set	Test set	Difference
Lab	28,13	48,6	20,47	65,92	69,26	3,34	88,37	93,26	4,89	21,8	15,66	-6,14
Luv	-12,44	10,25	22,69	66,38	64,99	-1,39	96,58	87,61	-8,97	35,43	30,2	-5,23
RGB	23,7	15,06	-8,64	70,45	68,69	-1,76	96,91	92,27	-4,64	25,96	22,6	-3,36
YCbCr	-6,88	14,86	21,74	61,09	68,12	7,03	89,94	93,38	3,44	32,22	23,09	-9,13
YUV	-4,5	-0,1	4,4	61,61	51,97	-9,64	92,51	87,44	-5,07	39,63	29,9	-9,73

Differences close to zero indicate balanced system performance.

Table II displays the numbers of penalty masks and false identifications for the different color spaces.

TABLE II.
SUMMARY OF MAIN ERROR OCCURANCES

	No. of Penalty masks		No. of False identifications	
	Test set	Evaluation set	Test set	Evaluation set
Lab	1	1	4	0
Luv	0	0	8	1
RGB	0	0	2	0
YCbCr	0	0	14	2
YUV	0	0	12	11

The generation of penalty masks is proportional to the adjustment of the bandwidth parameter. The tradeoff of a bigger bandwidth is the better accuracy versus the increasing number of classes which might result an instable system and the generation of penalty masks. In the case of the Luv and the YCbCr color spaces there was one image respectively which the system identified falsely in at least 9 of the 10 cases of mask generation, while the YUV had two of these images. High numbers of false identification explain weaker performance (compared to the Lab color space).

A note that one should not forget is that the results represent how accurate the *object* was found on the input picture pixel-wisely in the given color space. For example if the generated mask's silhouette is even only one pixel thicker/thinner than the hand made reference; the difference will be highly reflected by the metric making the 90% < accuracy in a real life environment somewhat a real challenge.

V. CONCLUSION

A segmentation algorithm using color information and mean shift has been implemented and evaluated on numerous color spaces. The overall quality award should be given to the Lab color space as it not only provided balanced results (as the relatively low standard deviation shows) but also even in the case of the image for which the masks with poorest average quality were generated the algorithm was able to reach an average accuracy of 28.13% which means that it found the object and was able to determine its boundaries with a fair accuracy.

VI. FURTHER WORK

The direction of further tasks can be divided into two main classes. Concerning system improvement the selection of the X samples, additional information gained with preprocessing should be used instead of random sampling. Possibilities of

dynamical bandwidth parameter adjustment should also be investigated. Fluorescence property of the objects should be exploited later on; and performance tests should be made with fluorescent algae and adjusted settings.

Concerning the further usage of mean shift it has been stated that the feature space of this algorithm can be almost any type of data if it is individual for every feature space element. This gives the idea that instead of color information the information about velocity should be considered as the feature space. This way the output of an optical flow algorithm could be segmented; resulting classes which group pixels with similar velocities together. As a result of connecting these two structures together it would be possible to perform foreground-background segmentation and also real-time image stabilization based on movement information. Although the calculation optical flow has a high complexity [11] the procedure could be made in real time by the usage of cellular neural networks which can work parallel. Both of these applications could be well employed by the Bionic Eyeglass Project [12].

ACKNOWLEDGEMENTS

The author would like to thank Tateyama Laboratory Hungary Ltd for their support.

REFERENCES

- [1] D. W. Seckler, D. Molden, R. de Silva, R. Barker, 1998., "World Water Demand and Supply, 1990 to 2025: Scenarios and Issues" International Water Management Institute. Available: http://www.iwmi.cgiar.org/Publications/IWMI_Research_Reports/PD/F/PUB019/REPORT19.PDF (accessed 20-06-2009.)
- [2] D. Comaniciu, P. Meer, "Mean shift analysis and applications," in *The Proceedings of the Seventh IEEE International Conference on Computer Vision Vol. 2, 1999*, pp.1197-1203.
- [3] Y.Cheng, "Mean Shift, Mode Seeking, and Clustering," in: *IEEE Trans. Pattern Anal. Machine Intell.* Vol. 17. No. 8, august 1995. pp. 790-799.
- [4] A. Touzani and J. G. Postaire, "Clustering by mode boundary detection," *Pattern Recog. Letters*, vol. 9, 1989. pp. 1-12.
- [5] D. Comaniciu, V. Ramesh, P. Meer, "Real-Time Tracking of Non-Rigid Objects using Mean Shift," *IEEE Computer Vision and Pattern Recognition, Vol II, 2000*, pp.142-149.
- [6] E. R. Dougherty, R. A. Lotufo, "Hands-on morphological image processing", SPIE Press, 2003 pp. 1-90, 129-147.
- [7] P. Soille, , *Morphological Image Analysis: Principles and Applications*, Springer-Verlag, 1999, pp. 173-174.
- [8] R.G. Kuehni, Color space and its divisions: color order from antiquity to the present, John Wiley and Sons, 2003. pp. 19-63, 229-232.
- [9] S. J. Sangwine, R. E. N. Horne, "The colour image processing handbook", Springer, 1998. pp. 27-73.
- [10] I. N. Bronstejn, D. Musiol, H. Mühlig, K. A. Szemengyajev, "Matematikai kézikönyv", Tipotex, 2006, pp. 615.
- [11] S. S. Beauchemin, J. L. Barron, „The computation of optical flow”, *ACM Computing Surveys*, Vol. 27, Issue 3, Sept. 1995. pp. 433-466.
- [12] K. Karacs, A. Lazar, R. Wagner, D. Balya, T. Roska, M. Szuhaj, „Bionic eyeglass: An audio guide for visually impaired” in *Biomedical Circuits and Systems Conference, 2006. BioCAS 2006*. pp. 190-193.

Human Detection in Videos with Strong Camera Motion

Dániel Szolgay

(Supervisors: Prof. Jenny Benois-Pineau and Prof. Tamás Szirányi)

szoda@digitus.itk.ppke.hu

Abstract— In this paper we focus on the problem of human detection in a specific videos acquired with wearable cameras. These videos are characterized by a strong camera motion and parallax. Hence the various pre-processing steps such as camera motion compensation, orientation correction are needed in order to isolate foreground areas where humans may appear.

The proposed method detects foreground areas and then recognizes human silhouettes according to the trained model in shape-description space. The application is in the field of monitoring the behavior of people attained by dementia diseases.

Index Terms— foreground/background segmentation, kernel methods, SVM classifiers

I. INTRODUCTION AND MOTIVATION

Wearable video capture has been recently gaining popularity due to the availability of new low weight and low energy consuming hardware. From the pioneering works of Steve Mann [1] in the domain of wearable computing, the technology has evolved to allow autonomous devices with a long battery life and image capture capabilities.

This type of device produces a new kind of visual data, which brings new possibilities [2] and requires automatic analysis approaches that are adapted to the new circumstances (e.g. unpredictable motion of the camera).

In this paper we present a system that covers all the processing steps from acquiring videos with wearable cameras to human detection in the foreground areas.

A. Related Works

Foreground/background segmentation has been one of the most researched areas for a long time and still the problem has not been solved. In case of fixed cameras the work of Stauffer and Grimson [3] showed a way to deal with difficulties like illumination changes, repetitive motions, and long-term scene changes. Many papers were inspired by their method: [4].

To handle periodically changing dynamic background elements, the authors in [5] propose to use not only colour components but optical flow parameters also for building a kernel density estimation based background model.

The task is even more challenging with moving cameras. In some restricted case (predictable, slow camera movement, no parallax), the methods that work with still cameras can be applied after motion compensation: [6].

In [7] the image pixel intensities are first treated as independent random variables but the existing spatial

correlation is exploited also. Here both background and foreground models are built and used competitively to detect pixels belonging to the foreground.

But the extraction of foreground areas themselves is still not sufficient if the specific objects have to be recognized. In this paper we are interested in human detection. Hence the most powerful techniques today such as SVM classifiers can be deployed in previously detected foreground areas. Nevertheless adequate “human” features have to be proposed to build the description space. In this paper, we resort to the edge orientation histograms [8] proposed for human detection in stills with specific adaptation of descriptor computation in the wearable camera setting.

The paper is organized as follows. In section 2 we introduce the general scheme of human detection. Section 3 presents our method for foreground-background separation. The approach for “human features” computation is presented in section 4. Section 5 contains the results of the method on wearable camera devices and section 6 concludes this paper.

II. GENERAL SCHEME OF HUMAN DETECTION ON WEARABLE CAMERA SEQUENCES

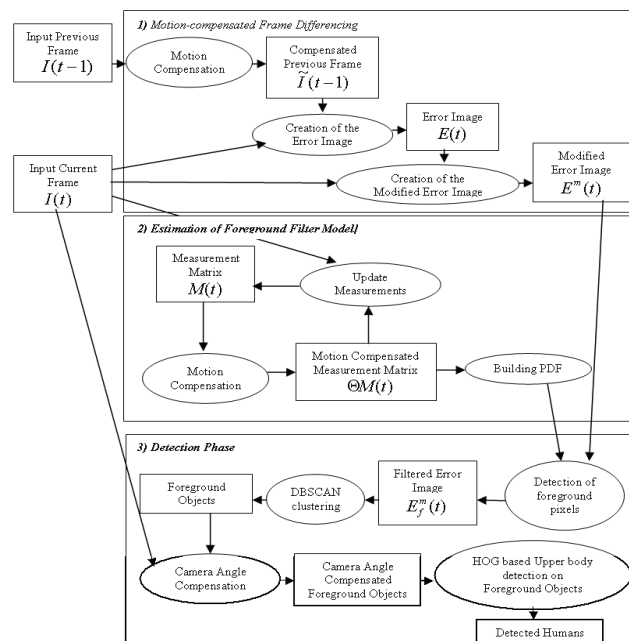


Fig. 1. General Scheme of the proposed method

The approach we propose consists of 3 main steps as depicted in Figure 1. It starts with camera motion compensation and creation of a modified error image, which is a motion-compensated difference frame with the colour

information of the original frame.

The next step is building a Probability Density Function (PDF) based foreground filter model. It is used to filter the modified error image from false positive pixels.

The two steps together form the background-foreground separation in case of strong camera motion and parallax.

The last step is the detection of humans in the foreground areas. For human detection the method developed by Dalal and Triggs [8] was used with a specific tuning for wearable camera setting.

In the following sections a more detailed description of these steps will be given.

III. FOREGROUND-BACKGROUND SEPARATION IN CASE OF STRONG CAMERA MOTION AND PARALLAX

We are working with moving cameras, hence the compensation of the camera motion is the first step towards moving foreground detection. In case of strong motion magnitude, the only possibility to estimate motion model θ close to the projected 3D motion is a hierarchical multi-resolution scheme. Hence we perform a hierarchical block-matching [9] and affine global motion estimation [10] and build the error image which should be strong on the pixels corresponding to the foreground objects with ego-motion.

$$E(t) = |\tilde{I}(t-1, \theta) - I(t)| \quad (1)$$

where $\tilde{I}(t-1, \theta)$ a compensated frame according to the estimated model θ .

The error (1) indeed represents a Motion Compensated Frame Difference (MCFD).

Nevertheless, due to the aliasing and high frequency noise, the MCFD cannot be used as such for foreground object detection: background contours are present in this signal as well. Hence we form a modified MCFD:

$$E^m(x, y, t) = \begin{cases} I(x, y, t) & \text{if } E(x, y, t) > th_E \\ 0 & \text{else} \end{cases} \quad (2)$$

where I is a 3 channel colour frame at time t , E is the grey scale motion compensation error at time t and th_E is a threshold.

A. Foreground-background separation

To separate pixels of moving objects from pixels in static contours present in modified error image due to the noise, we use the PDF estimation of the background and probabilistic decision rule.

In order to estimate the PDF for background pixel values, we build a *measurement matrix* M . This matrix contains the information of the original frames, in n consecutive time instances and it is continuously updated along the time.

Updating at time t means, that we add the information of the frame at time t , to the measurement matrix, of the previous time instance, $t-1$.

$$\begin{aligned} M(x, y, t) &= \Theta_t M(x, y, t-1) \cup I(x, y, t) \\ M(x, y, 1) &= I(x, y, 1) \end{aligned} \quad (3)$$

Here the operator \cup means adding new frame of measurements to the measurement matrix, while the oldest frame is being removed. The operator Θ_t stands for the affine transformation with the estimated camera model θ .

Applying this transformation we compensate all images of the measurement matrix to the reference frame, the current one.

The measurements, stored in M , will be used for estimating the probability that an incoming pixel belongs to the background. Here we use a kernel-based density estimation method [11].

1) Estimation of Foreground Filter Model

Density estimation is the construction of an estimate of an unobservable probability density function, based on observed data.

One of the most popular approaches for density estimation is kernel-based estimation. In this work we use K_n nearest neighbour approach (see [12], p174 for the detailed description of the method). In particular we use sample-point density estimator [11], with Gaussian kernel, which can be defined as follows:

$$f_s(v) = \frac{1}{n\sqrt{2\pi}} \sum_{i=1}^n \frac{1}{\sigma(v_i)} e^{-\frac{(v-v_i)^2}{2\sigma^2(v_i)}}, \quad (4)$$

where n is the number of measurements, and $\sigma(v_i)$ is the sample-point bandwidth associated to the i th sample point v_i .

A common choice for the bandwidth calculation is using the distance of the sample point v_i from its k th nearest neighbour. However in our case the number of sample points is strongly limited and this kind of calculation might give false result, as it is pointed out in [13].

As a solution to this problem we propose to use the distance from all the k nearest neighbours, instead of the distance from the k th alone. The σ_i parameter is calculated from the variance of the k nearest neighbours around the measurement v_i :

$$\sigma_i^2[c] = \frac{1}{k} \sum_{j=1}^k (v_i[c] - v_j[c])^2 \quad (5)$$

where v_j is the j th nearest value to v_i in the measurement matrix c is the index of colour channel. For the estimate $\tilde{f}(v)$ to converge to the true unknown PDF, the following should be satisfied: $k(n)/n \rightarrow 0$ when $n \rightarrow \infty$. We use $k = \sqrt{n}$, where n is the number of available measurements.

We also propose a new concept of a spatio-temporal PDF of the background. Indeed we call it spatio-temporal according to the choice of sample points: we use both spatial neighbourhood of a pixel and its temporal history.

The number of sample points in our approach is strongly limited, as we pointed out, and the effect of noise can be very strong. Therefore, the method for selecting the sample points has key importance. In [5] the sample points at given (x, y) coordinates are the n previous measurements taken at the same (x, y) position: $(v(x, y, 1), v(x, y, 2), \dots, v(x, y, n))$.

When the camera is moving the case is different. Even after motion compensation the real background scene position that corresponds to the (x, y) pixel in one frame, might move a little, due to minor errors of camera motion compensation, or quantization. Assuming that this spatial error is random, we use the values selected in a small patch centred on the (x, y) .

Based on the values of the measurement matrix, PDFs are built for each c channel of each non zero (x, y) pixel of the

current modified error image (2):

$$f_{x,y}(v[c]) = \frac{1}{n \cdot \sqrt{2\pi}} \cdot \sum_{i=1}^n \frac{1}{\sigma_i[c]} e^{-\frac{(v[c]-v_i[c])^2}{2\sigma_i^2[c]}} \quad (6)$$

where v_i is the a previously measured value of the (x,y) point, obtained from the M matrix through spatio-temporal selection, n is the number of measurements, and $\sigma_i=(v_i)$ (see (5)) is the bandwidth parameter of the Gaussians. The value \bar{v}_i is a vector, which contains the YUV colour components. In our case the PDF (6) is calculated for each colour component independently with different σ values (5).

IV. DETECTION OF FOREGROUND OBJECTS

Once the PDF has been built for each pixel in the current frame, we can proceed to the detection of foreground moving objects. Here the pixels will be first classified as belonging to the foreground or background on the basis of the PDFs characteristics. Then the detected pixels will be grouped into clusters (moving objects) on the basis of their motion, colour and spatial coordinates in the image plane.

A. Classification of foreground/background pixels

The $f_{x,y}$ function is a probability distribution function that shows how likely the pixel (x,y) takes a v value. Based on this likelihood the domain R of all possible v values is divided into two parts: R_1 and R_2 . R_1 is associated to the background colours and R_2 to the foreground colours.

Let $P(2|1, R)$ be the probability of false detection of an object and $P(1|2, R)$ is the probability of miss detection.

The decision rule consists in comparison of P with a threshold. Hence we propose the following choice of this threshold.

The goal here is to keep $P(2|1, R)$ small, while ensuring that the territory of the background (R_1) remains as small as possible as well. To achieve this, a heuristic and adaptive way for the calculation of the threshold can be considered:

$$T_{x,y}[c] = \lambda \cdot \frac{\sum_{i=1}^n f_{x,y}(v_i[c])}{n}, \quad (7)$$

where λ is a constant, n is the number of the available measurements and $v_i \in M(x, y)$.

B. Foreground Pixels Clustering with DBSCAN

To find moving objects' silhouettes and eliminate the remaining noise, we cluster detected foreground pixels based on their compactness in image plane, common motion and colour. Hence we used DBSCAN [14] clustering algorithm for its capacity to filter noise in data in a mixed feature space R^l . In this space with $l=7$ dimensions, each foreground pixel is described with a feature vector $X=(x, y, C_1, C_2, C_3, dx, dy)^T$ which contains the x, y coordinates, the colour coordinates C_1, C_2, C_3 in YUV space and the coordinates of the displacement vector dx, dy expressing pixel motion.

C. Human Detection on the Foreground

After foreground clustering, we got limited areas in image plane and for each area we need to decide if it is a human or

not. The inspiration of our work is the method of Dalal and Triggs [8] using Histograms of Oriented Gradients (HOG)

Our goal is to detect persons in their home environment, where the lower body parts are often occluded by furniture. Hence we modified the original detector [8] into an upper body detector, by training the SVM model on upper body part with a smaller window (80x64).

At the present state our database does not contain enough positive examples for training therefore we used the database [15] and added negative examples from indoor environment.

As the original, this modified detector is sensitive to camera tilt, which is common with wearable cameras. To handle this problem we perform a geometric correction of camera compensating its orientation.

To determine the orientation of the camera we make the assumption that a picture showing an ordinary living environment contains mainly horizontal and vertical edges.

Using this assumption, an edge orientation histogram $H(k)$ of 32 bins in the range of 0-180° is calculated on each frame $I(k)$.

The angular offset is then defined as principal mode of $H(k)$ and compensated. With this method we can estimate the angular offset (α) and correct camera tilt up to $\alpha=45^\circ$ with a simple affine transformation.



Fig. 3. From left to right: Original frame, edge orientation histogram of the original frame and orientation corrected frame.

V. RESULTS

The results presented here were obtained on aged healthy volunteers participating in a Dementia study. As they usually stay alone at home, the sequences containing persons are very rare. In the corpus of duration of 9.3 hours, we can hardly find a few seconds with the presence of persons. Hence to construct the ground truth for the tests of our method, we mainly used these short sequences.

The proposed *foreground extraction* method was compared to the one described in [4] which is an enhanced version of [3]. While [4] is originally working with fixed cameras, we applied the above used motion estimation method for camera motion compensation.

Table 1 shows the results of our method (with $n=10$ and $th_E=10$) and the method [4] (Gaussian Mixture Model based method) on videos acquired with a wearable standard button camera. The precision and recall rates were calculated based on the overlap between the pixels annotated as foreground in the ground truth and the estimated foreground image in case of both methods. Our method performs significantly better both in recall and precision metrics. Some results are given in Figure 4.



Fig. 4. Detection of foreground objects with the proposed method. (On Francois1 and Daniel1 sequences)

In order to measure the efficiency of HOG-based upper body detector and of the whole method we tested the detector on the whole frame and on the foreground areas only. Since the HOG filter is a scan window based method we defined efficiency measures as follows: we have a true positive if the SVM classifies a feature vector as human and the corresponding scan window has more than 50% overlap with the ground truth. A false positive is when the SVM classifies the window as human but the overlap is less than 50%. We have a missed detection in case a human is present on the image but no true positives were detected. Another important parameter of an SVM based method is the number of decisions it has to make (and inherent computational time). It is obviously drastically lower on foreground areas only. Table 2 and Figure 5 show the results of the above described tests.

TABLE I
PRECISION AND RECALL RATES OF FOREGROUND DETECTION FOR THE PROPOSED AND A CONCURRENT METHOD

Precision/Recall			
Sequence	#Frames	GMM	Prop. Meth
Francois 1	62	0.05/ 0.33	0.40 / 0.84
Francois 2	143	0.24/ 0.33	0.74 / 0.82
Daniel 1	92	0.26/ 0.43	0.62 / 0.54
Daniel 2	30	0.23/ 0.30	0.45 / 0.68

TABLE II
NUMBER OF SVM DECISIONS FOR THE FULL FRAME AND FOR THE FOREGROUND

Number of Scan Windows		
Sequence	On Full Frame	On Foreground
Francois1	607260	97525
Francois2	1437182	37608
Daniel1	931132	150695

As it can be seen from Figure 5 despite the area of human search is limited in our method, the number of false negatives (c) does not rise significantly while the number of false positives (a) drops drastically.

VI. CONCLUSION

In this paper we proposed a full method for detection of

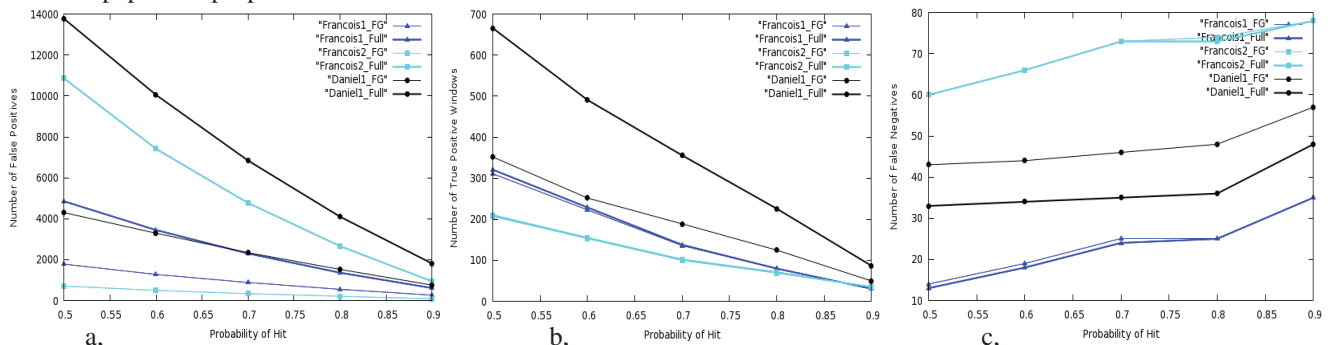


Fig. 5. Results of Human Detection for 3 different sequences on the full frame and on the foreground only. (Sequences: Francois 1, 2 and Daniel)

humans in videos with wearable cameras. The method proceeds first by the foreground object extraction and then by recognition of humans.

We show that we overcome strong constraints due to the camera motion. The presented foreground detection is much better than in greedy method and promising results are obtained for SVM based human detector. More features like HOG are now being explored.

The presented work has been submitted to IEEE International Conference on Image Processing 2009.

REFERENCES

- [1] S. Mann, Wearable computing: a first step toward personal imaging, *Computer* 30 (2), pp. 25-32, 1997
- [2] Personal and Ubiquitous Computing, special issue on Memory and Sharing of Experiences. Springer 11 (4), pp. 213-328, 2007.
- [3] C. Stauffer and E. Grimson, "Learning patterns of activity using real-time tracking", *IEEE Trans. PAMI*, 22 (8), pp. 747-757, 2000.
- [4] Z. Zivkovic, F. Van Der Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," in *Pattern Recognition Letters*, 27 (7), pp. 773-780, 2006.
- [5] A. Mittal, N. Paragios, "Motion-Based Background Subtraction Using Adaptive Kernel Density Estimation," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 302-309, 2004.
- [6] G.L. Foresti and C. Micheloni, "A robust feature tracker for active surveillance of outdoor scenes," in *Electronic Letters on Computer Vision and Image Analysis*, 1(1), pp 21-34, 2003.
- [7] Y. Sheikh and M. Shah, "Bayesian Modelling of Dynamic Scenes for Object Detection," *IEEE Transactions on PAMI* Vol.27, issue 11, Nov 2005.
- [8] N. Dalal and B. Triggs. "Histograms of Oriented Gradients for Human Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, USA, June 2005. Vol. II, pp. 886-893.
- [9] M. Bierling, "Displacement estimation by hierarchical block matching," in *Proc. SPIE Visual Communications and Image Processing*, Vol. 1001, pp. 942-951, 1988.
- [10] M. Durik, J. Benois-Pineau, "Robust motion characterisation for video indexing based on MPEG2 opticalflow," in *Proc. of the International Workshop on Content-Based Multimedia Indexing*, pp. 57-64, 2001.
- [11] M.P. Wand, M.C. Jones, *Kernel Smoothing*, Chapman and Hall, 1995.
- [12] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, 2nd ed. John Wiley & Sons, Inc., N.Y, 2001.
- [13] A. Bugeau, "Détection et suivi d'objets en mouvement dans des scènes complexes, application à la surveillance des conducteurs," Thèse de l'université de Rennes 1, Mention Traitement du Signal et des Télécommunications, 2007.
- [14] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. of Second Int. Conf. on Knowledge Discovery and Data Mining*, Portland, OR, pp. 226-231, 1996.
- [15] Dalal, N., "INRIA Person Dataset," Online, 2005. <http://pascal.inrialpes.fr/data/human/>

Strategy optimization on financial time series

Kálmán Tornai
(Supervisor: János Levendovszky)
tornai.kalman@gmail.com

Abstract—In the financial world, time series for different products are often linked or correlated to each other. The correlation of prices can be quite strong for products falling into the same product category. The changes in the spread and the prices of the corresponding products are influenced by events. An event can be some prior news about an unexpected happening, or concerning a company announcement on earnings/acquisitions, or securities issuance by governments ...etc. As result, there is a need for efficient and low complexity strategies in the financial markets to exploit and to draw maximum profit from the changes brought about by the events. Consequently, the aim of this summary is to introduce and to describe a formal model which can serve as a basis for developing optimal or near-optimal strategies exploiting the changes.

I. THE MODEL

Let two financial products be considered, the first one is referred to as Bond₁, while the second is Bond₂. The associated *ask* (sell) and *bid* (buy) prices are denoted by $x_s^{(1)}(t), x_s^{(2)}(t), x_b^{(1)}(t)$, and $x_b^{(2)}(t)$, respectively. As a result, the time series available about the prices can be represented as a 4D vector $\mathbf{x}(t) := (x_s^{(1)}(t), x_s^{(2)}(t), x_b^{(1)}(t), x_b^{(2)}(t))$, where t is assumed to be a discrete variable being a non-negative integer. Furthermore, there are some events $\xi(T_i)$ at time instants $T_i, i = 1, \dots, L$ are given, where ξ is regarded as a random variable taking its value from a discrete set. In order to turn the information on the event into profit there is a time interval $[T_i^b, T_i^e]$ (where $T_i^b := T_i - \Delta_i$ and $T_i^e := T_i + \Delta_i$) during which some actions can be carried out (e.g. buying or selling).

The investor is characterized by a state vector $\mathbf{y}(t) := (c(t), n_1(t), n_2(t))$, where $c(t)$ denotes the available cash, while $n_1(t)$ is the number of Bond₁ and $n_2(t)$ is the number of Bond₂ in possession at time instant t . There is reward associated with $\mathbf{y}(t)$, which can be calculated as turning all the assets into cash at the given instant t , defined as

$$r(t) = \Psi(\mathbf{y}(t)) := c(t) + n_1(t) \cdot x_s^{(1)}(t) + n_2(t) \cdot x_s^{(2)}(t) \quad (1)$$

During the intervals allocated for carrying out a strategy, the actions are characterized by vector $\mathbf{a}(t) := (s_1(t), b_1(t), s_2(t), b_2(t))$ $t \in [T_i^b, T_i^e]$, where $b_1(t)$ is the number of Bond₁ acquired and $s_1(t)$ is the number of Bond₁ sold. The same holds for the actions concerning Bond₂ described by the last two components of vector $\mathbf{a}(t)$.

Each action is constrained by the following bounds

$$\begin{aligned} b_1(t) \cdot x_b^{(1)}(t) + b_2(t) \cdot x_b^{(2)}(t) &\leq c(t) \\ s_1(t) &\leq \min\{A_1, n_1(t)\} \\ s_2(t) &\leq \min\{A_2, n_2(t)\}, \end{aligned} \quad (2)$$

where the quantities A_1 and A_2 are imposed on the investor by the market indicating that these are the maximal number of bonds which one can sell at the given time instant.

A. State transitions

Due to an action $\mathbf{a}(t)$ the state vector is updated as follows

$$\mathbf{y}(t+1) := \mathbf{a}(t) \Downarrow \mathbf{y}(t) \quad (3)$$

where the symbolic operation denoted by \Downarrow is defined component-wise as follows:

$$\begin{aligned} c(t+1) &:= c(t) - b_1(t) \cdot x_b^{(1)}(t) - b_2(t) \cdot x_b^{(2)}(t) + \\ &+ s_1(t) \cdot x_s^{(1)}(t) + s_2(t) \cdot x_s^{(2)}(t); \end{aligned} \quad (4)$$

$$\begin{aligned} n_1(t+1) &:= n_1(t) - s_1(t) \\ n_2(t+1) &:= n_2(t) - s_2(t) \end{aligned} \quad (5)$$

A sequence of actions operating on a state vector is denoted as

$$\mathbf{a}(t_1)\mathbf{a}(t_2)\dots\mathbf{a}(t_M)\Downarrow\mathbf{y}(t) \quad (6)$$

indicating the state has been changed first at time instant i_1 , then at i_2 and then finally at i_M .

B. The optimal strategy

The optimization task can be formulated: Given action $\xi(T_i)$, search for the optimal $T_i^b \leq t_{i_1, opt} \leq \dots \leq t_{i_M, opt} \leq T_i^e$ and $\mathbf{a}_{opt}(t_{i_1, opt})\mathbf{a}_{opt}(t_{i_2, opt})\dots\mathbf{a}_{opt}(t_{i_M, opt})$ for which

$$\begin{aligned} \max r(T_i^e) &\sim \max \Psi(\mathbf{y}(T_i^e)) \sim \\ &\sim \max \Psi(\mathbf{a}_{opt}(t_1)\mathbf{a}_{opt}(t_2)\dots\mathbf{a}_{opt}(t_M)\mathbf{y}(t)). \end{aligned} \quad (7)$$

This strategy can be made adaptive: Given action $\xi(T_i)$, the current time t_j , the actual state vector $\mathbf{y}(t_j)$, and $(x_s^{(1)}(t_q), x_s^{(2)}(t_q), x_b^{(1)}(t_q), x_b^{(2)}(t_q))$ for $q = 1 \dots j - 1$, search for the optimal next action $\mathbf{a}_{opt}(t_{j+1})$ for which

$$\begin{aligned} \max r(T_i^e) &\sim \max \Psi(\mathbf{y}(T_i^e)) \sim \\ &\sim \max \Psi(\mathbf{a}_{opt}(1)\mathbf{a}_{opt}(2)\dots\mathbf{a}_{opt}(M)\mathbf{y}(t)). \end{aligned} \quad (8)$$

C. Model of prices

Each Bond _{i} has two associated prices, denoted $x_s^{(i)}$ and $x_b^{(i)}$, *ask* and *bid* price. Each x considered as an random variable ζ with normal distribution. The variance σ^2 is constant, but the mean of ζ may change in time. The change is described with a function $z_x(t)$, which is a linear function of t ($z_x(t) = c_g \cdot t + c_i$). Therefore:

$$x \sim \zeta \sim \mathbf{N}(z_x(t), \sigma^2)$$

Basically the impact of ξ event is the change of $z_x(t)$ function and σ .

D. Probability of event

Consider one ξ event for time instant T , where $\xi \in \mathcal{P}(H)$, where H is a set of possible price changes. Due to ξ event the value of gradient of $z_x(t)$ function may change as follows:

- No change: $z_x(t > T) = z_x(t < T)$
- The sign of ‘trend’ changes: $z_x(t > T) = -1 \cdot z_x(t < T)$
- Raises, without sign change $z_x(t > T) = a \cdot z_x(t < T)$, $a > 1$
- Lowers, without sign change $z_x(t > T) = a \cdot z_x(t < T)$, $1 > a > 0$
- Raises, with sign change $z_x(t > T) = a \cdot z_x(t < T)$, $a < -1$
- Lowers, with sign change $z_x(t > T) = a \cdot z_x(t < T)$, $-1 < a < 0$

Considering $H = \{x_s^{(1)}, x_b^{(1)}, x_s^{(2)}, x_b^{(2)}\}$, the number of possible instances of ξ event is 6^n . (In our case $n = 4$, then $6^n = 1296$) Every possible ξ event has a probability denoted by p_i . One of possible events must occur, therefore

$$\sum_{k=1}^K p_k = 1 \quad K = 6^n$$

The gain after making and performing optimal decision is the difference between reward at begin of time interval (T^b) and at the end of time interval (T^e): $g_k = r(T^e) - r(T^b)$ where $r(T^e) = a_{opt}(t) \cdot y(T^b)$. t is the time instant where optimal action has to be made. This gain corresponding to one possible event also have a p_i probability indicating the chance of selected event occurs, therefore one can introduce the probability biased gain:

$$\hat{g}_k := g_k \cdot p_k \quad (9)$$

Because of large possible event set the optimal decisions searching space is extremely large. First of all this set should be narrowed by reasonable restrictions. First, one Bond’s ask and bid price changes together, therefore we consider only the bid prices. Second, if one Bond’s price does not rise over the time window, there is only one option for optimal strategy: sell that Bond at best price and buy another if it worths.

Then the possible events are the following

- The spread between the prices does not change significantly
- The spread between the prices grows
- The spread between the prices decreases

In the following the possible strategy actions and corresponding gains will be formalized. One action can be characterized by an action vector $\mathbf{a}(t) := \{s_1(t), b_1(t), s_2(t), b_2(t)\}$ where $s_i(t)$ is the amount of Bond _{i} to sell and $b_i(t)$ is the amount of Bond _{i} to acquire. The trivial assumption is the following: if $s_i(t) > 0$ then $b_i(t) = 0$ for $\forall i$ and vice versa.

For each possible event there is an optimal decision with the greatest amount of gain. For all possible gain the biased gain can be calculated, therefore the biased optimal strategy can be found by selecting the maximal biased gain.

For example if the optimal strategy is to sell every piece of Bond₂ and buy as much Bond₁ as possible, the gain at the end

of time period is (t_{tr} has the time instant when transaction is to be performed.)

$$g = \left[n_1(T^e) \cdot x_s^{(1)}(T^e) + n_2(t_e) \cdot x_s^{(2)}(T^e) + c(T^e) - n_1(T^b) \cdot x_s^{(1)}(T^b) - c(T^b) \right] \quad (10)$$

The cash after transactions is

$$c(T^e) = c(t_{tr}) - \left(x_b^{(1)}(t_{tr}) \cdot \left[\frac{x_s^{(2)}(t_{tr}) \cdot n_2(t_{tr}) + c(t_{tr})}{x_b^{(1)}(t_{tr})} \right] \right) \quad (11)$$

The amount of Bond 2 to sell is $n_1(t_b)$. And the amount of Bond 1 to buy (is the maximal amount):

$$buy = \left[\frac{x_s^{(2)}(t_{tr}) \cdot n_2(t_{tr}) + c(t_{tr})}{x_b^{(1)}(t_{tr})} \right] \quad (12)$$

Naturally this can be written if the optimal strategy is to sell Bond₁ and buy Bond₂.

E. Impact of actions

The previous chapters did not deal with the impact of the actions. Suppose that the decision is to sell. If the holder of bonds is selling some bonds it may decrease the bonds selling (and also buying) rate. In the following this symptom will be integrated to the model.

The impact of one transaction should be a function of the amount of bond participated in the transaction. First of all this function is considered as a linear function, therefore the impact of transaction can be defined as following when transaction is about to sell:

$$\text{impact}_b^{(i)}(s_i(t_j)) := -c_{s_i} \cdot s_i(t_j) \quad (13)$$

where c_{s_i} is the constant is peculiar to Bond _{i} selling price, and $s_i(t_j)$ is the amount of sold asset at time instance t_i . The impact of transaction on bid price is the following:

$$\text{impact}_b^{(i)}(b_i(t_j)) := c_{b_i} \cdot b_i(t_j) \quad (14)$$

Then the price of Bond₁ changes as following:

$$x_s^{(i)}(t) = \begin{cases} x_s^{(i)}(t) & \text{if } t < t_1 \\ x_s^{(i)}(t) - \text{impact}_s^{(i)}(s_i(t_j)) & \text{otherwise} \end{cases} \quad (15)$$

Therefore the possible reward is affected and also the reachable gain. With exhaustive search of optimal solution the computation algorithms should be modified only.

II. DETECTING THE TIME INSTANT OF ξ EVENT

A. Problem

The exact time instant when ξ event occurs cannot be determined in advance. Previously it was presumed that $\xi(T_i)$ and T_i is exactly known in advance.

Suppose that the type of event is known, the values of time series are also known and an algorithm must figure out the exact T_i value.

B. Statistical approach

This problem can be treated as an outlier detection problem. The ξ event has some influence on the distribution of Bond₁ and Bond₂ prices' as defined above. The changes of mean value and deviation may cause outliers in the time series. One can define the following differential time series of one Bond's price:

$$\tilde{\zeta}(n) = x(n) - x(n-1)$$

Using standard outlier detection algorithms on $\tilde{\zeta}$ with window $1 \dots n$, $n = 2, 3, \dots$ the possible T_i time instance could be found.

1) *Grubb's test*: Let us define the following test using the Student's distribution's inverse. The selected x_e value is outlier if the corresponding $G(x_e)$ value is greater than the critical value: G_c . G_c computed as follows:

$$G_c = \frac{n-1}{\sqrt{n}} \sqrt{\frac{t^2_{\left(\frac{\alpha}{2n}, n-1\right)}}{n-2 + t^2_{\left(\frac{\alpha}{2n}, n-1\right)}}}$$

where α denotes the significance level, and n the length of time series.

$$G(x_e) = \frac{x_e - \bar{x}}{\sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}}$$

where \bar{x} denotes the average of time series x .

2) *Algorithm*: Grubb's test can be used directly on time series and on the differential time series. Numerical test showed that working on differential time series produces better accuracy.

A differential time series $\tilde{\zeta}$ computed as defined preceding. The algorithm works on growing window, the first detected outlier is equivalent to the time instance of the event ξ .

C. Analysing performance of detection

Each Bond's price considered as a random variable $\zeta \sim \mathbf{N}(z_x(t), \sigma^2)$. This time series can be unfold to two time series as following $x_1 = x(i)$ with $i = 1 \dots n-1$ and $x_2 = x(i)$ with $i = 2 \dots n$. $x_1 \sim x_2 \sim \mathbf{N}(z_x(t), \sigma^2)$. The differential time series can be computed

$$\tilde{x}(i) = x_2(i) - x_1(i)$$

Outlier detection performed on differential time series $\tilde{\zeta}$, therefore to analyse performance the distribution of $\tilde{\zeta}$ must be calculated. According to the properties of normal distribution $\tilde{\zeta}$ is the following:

$$\begin{aligned} \tilde{\zeta} &= \mathbf{N}(z_x(t), \sigma^2) - \mathbf{N}(z_x(t+1), \sigma^2) = \\ &= \mathbf{N}(z_x(t) - z_x(t+1), \sigma^2 + \sigma^2) = \mathbf{N}(\dot{z}_x(t), 2 \cdot \sigma^2) \end{aligned} \quad (16)$$

The deviation of $\tilde{\zeta}$ is $\sqrt{2} \cdot \sigma$ and the mean is the gradient of $z_x(t)$ ($=c_g$).

In this case the detection problem is a two-class classification problem. c_{g_b} denotes the gradient of $z_x(t)$ where $t < T$ and c_{g_a} where $t > T$. Similarly σ_b denotes the deviation of price before the ξ event occurs and σ_a otherwise. There is C

where the p.d.f of price pre-event and after-event equals. This C should be used as the parameter of optimal decision.

$$C = x \longrightarrow PDF[\mathbf{N}(c_{g_b}, 2 \cdot \sigma_b^2)] = PDF[\mathbf{N}(c_{g_a}, 2 \cdot \sigma_a^2)] \quad (17)$$

Solving the equation:

$$\begin{aligned} C &= \pm \frac{\pm c_{g_b} \sigma_a^2 \mp c_{g_a} \sigma_b^2}{\sigma_a^2 - \sigma_b^2} * \\ &* \sqrt{\sigma_a^2 \sigma_b^2 \left((c_{g_a} - c_{g_b})^2 + 4(\sigma_a^2 + \sigma_b^2) \log\left(\frac{\sigma_b}{\sigma_a}\right) \right)} \end{aligned} \quad (18)$$

If $\sigma_b = D \cdot \sigma_a$:

$$\begin{aligned} C &= \pm \frac{\pm c_{g_b} \mp c_{g_a} D^2}{-1 + D^2} * \\ &* \sqrt{D^2 \left((c_{g_a} - c_{g_b})^2 - 4 + (2D)^2 \sigma_a \log(D) \right)} \end{aligned} \quad (19)$$

If $D = 1$:

$$C = \frac{c_{g_a} + c_{g_b}}{2} \quad (20)$$

The probability of misclassification: $c_{g_b} < c_{g_a}$

$$\int_{-\infty}^C \frac{e^{-\frac{(c_{g_a} + x)^2}{4\sigma_a^2}}}{(2\sqrt{\pi}\sigma_a)} dx + \int_C^{\infty} \frac{e^{-\frac{(c_{g_b} + x)^2}{4\sigma_b^2}}}{(2\sqrt{\pi}\sigma_b)} dx \quad (21)$$

1) *Numerical results*: The following figure shows the actual performance of the outlier detection algorithm described before. The first event is the 'Nothing happens' event.

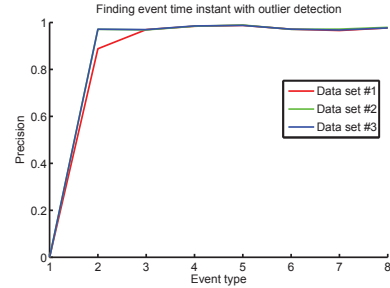


Fig. 1. Performance of outlier detection

2) *Numerical results with different time series model*: In this section the simulation are based up on a slightly different price model. The $z(t)$ function responsible for the slope of price is constant. The ξ event only changes the expectation value of price. The following figure shows a full performance analysis. The event (outlier) detection works perfectly on time series with small amount of noise. (Except the 'nothing happens' event.) Raising the deviation of noise on price reduces the performance of detection. The score is average of multiple tests. 1 is the best, when the algorithm correctly detected the T time of event ξ . The score is less than 1 when not exactly the correct T was found, but $|T - t| < 20$.

III. NUMERICAL SIMULATIONS

A. Environment

To test the model some numerical simulations were executed. The time series was generated with random generator, according to the time series model. The possible changes discussed above. Every ξ event had possibility and the strategy

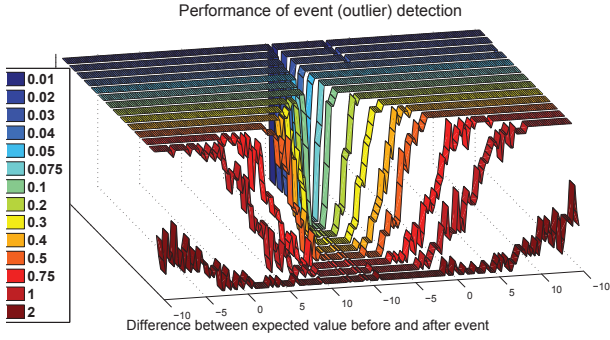


Fig. 2. Performance analysis of outlier detection (different price models)

corresponding to maximal biased gain was the optimal solution. To obtain the solution exhaustive search was executed.

The exhaustive search does not permit the following actions

- There is no loan, therefore the money cannot be smaller than zero at any time.
- Cannot sell more than actually have.
- There is a limit for sell and buy - the market has a maximal available Bond, and maximal acceptable Bond amount.

B. Results

All of simulation started with having equal amount of Bond₁ and Bond₂. The prices of Bonds are generated with ζ normally distributed random variable. The expected value of ζ changes slightly with time therefore the prices' timelines have a gentle slope. This slope is dramatically changed by the ξ event. The task is to find an optimal strategy to obtain most possible gain.

Figure 3 shows an outcome of exhaustive search of one optimal strategy, when one of Bonds' price lowers after the ξ event. In this case the possible maximal gain is 27.7% of the

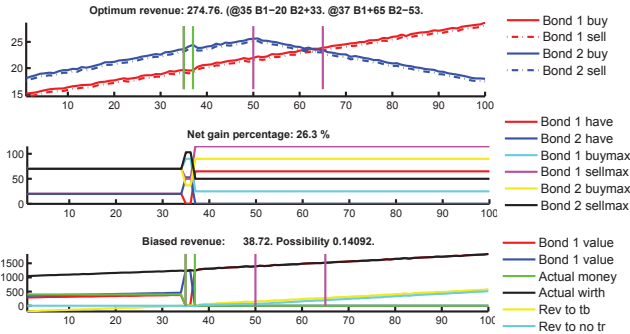


Fig. 3. Result of simulation - Bond 2's prices fall fortune at start point of time frame.

Previously we introduced the term of biased gain. When instead of most probable ξ event other event occurs the optimal strategy is different than we obtained. Therefore the gain is less than the optimal. Figure 4 shows an average of reachable gains when other than optimal strategy is used. In this figure clearly shows that in some cases instead of profit a loss appears.

IV. STRATEGY OPTIMIZATION BY LEARNING ALGORITHMS - FUTURE DEVELOPMENT

A. General idea

After posing the optimization problem, one can raise the concern about the fact that seeking the optimum is of enor-

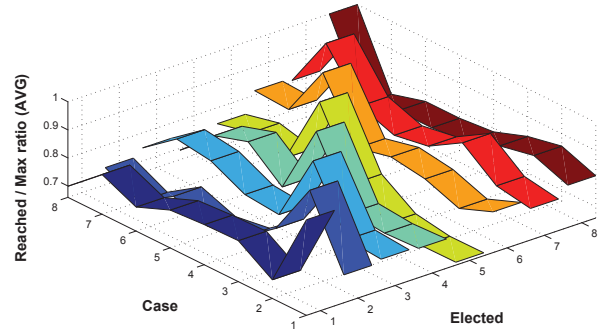


Fig. 4. Result of simulation - Using different strategies

mously large complexity due to the huge search space owing to the large number of free parameters. As a result, taking into account limited timespan and computational resources, the optimal strategy may only be found for a couple of events. Let vector \mathbf{d}_i include all the parameters of the optimal strategy

$$\mathbf{d}_i := (t_{i_1,opt}, \dots, t_{i_M,opt}, \mathbf{a}_{opt}(t_{i_1,opt}), \dots, \mathbf{a}_{opt}(t_{i_M,opt}))$$

corresponding to an event characterized by the parameters $\mathbf{z}_i = (T_i, \Delta_i, v_i)$. Then one can form a training set $\tau_K := \{(\mathbf{z}_i, \mathbf{d}_i), i = 1, \dots, K\}$. The size of this set is determined by the available computational resources. Based on τ_K any learning architecture having universal approximation capabilities can be trained to learn the optimal strategy on the training set. Once the free parameters of the corresponding architecture has been adopted (once the learning is over), the architecture can be used to generalize the results, i.e. an input vector \mathbf{z} is presented and the corresponding strategy \mathbf{d} can be obtained. Furthermore, the resulted reward can also be evaluated by using formula (1). Based on the rewards scored by a specific algorithm, the different strategy optimization methods can be compared and ranked.

V. SUMMARY

An efficient but simple model has been developed for two or more financial products (e.g. Bonds). Using this model algorithms can be implemented to obtain the optimal strategy (buy or sell) these products, when an event occurs. Also event detection algorithms (to determine the exact time when the event occurs) has been examined. The basis of event detection problem has been reduced to outlier detection problem, thus simple statistical methods can be used. The analytical and numerical comparison also performed of outlier detection method. For further developments the basic idea how to use learning algorithms has been posed.

REFERENCES

- [1] V. Barnett and T. Lewis: Outliers in Statistical Data, Wiley and Sons, 1994
- [2] Grubbs, f. e: Procedures for detecting outlying observations in Samples, Technometrics
- [3] Simon Haykin: Neural networks, Prentice Hall; 1998

High Field Characteristics of Long and Short channel 2D Graphene FETs

Kristof Tahy
(Supervisor: Dr. Árpád I. Csurgay, Debdeep Jena)
ktahy@nd.edu

Abstract □ The high-field transport properties of long and short channel 2D graphene FETs is investigated. Saturation current densities in the 1 □ 1.5 A/mm range have been measured, comparable to Si-MOSFETs and III-V Nitride HEMTs. It is found that the back-gate modulation of high bias currents near current saturation is strongly affected by the graphene channel length. Short channel effects cause a strong degradation of the gate modulation of the drain current. It is also found that for matched carrier concentrations, higher drive currents are possible in graphene compared to other semiconductors.

Index Terms □ graphene, FETs, current density, saturation, gate modulation

I. INTRODUCTION

THE recent discovery of graphene [1], a single atomic sheet of graphite, has ignited intense research activities to explore the electronic properties of this novel two-dimensional (2D) electronic system. Charge transport in graphene differs from that in conventional 2D electronic systems as a consequence of the linear energy dispersion relation near the charge neutrality (Dirac) point in the electronic band structure [2]. Field-effect mobilities as high as 15000 cm²/V·s and carrier velocity of ~10⁸ cm/s have been demonstrated at room temperature [3]. These impressive properties make graphene a possible candidate for electronic devices in the future. However, graphene has been studied mostly at low biases till date. For the application in practical devices, it is essential to investigate the high-field transport properties. In this letter, high field characteristics of 2D exfoliated graphene are reported on both short-channel and long-channel back gated FETs. High current drives were measured by pushing the devices up to breakdown. The saturation current density for many samples has been measured to be in the 1 – 2 A/mm range. Gate modulation of the drain current is found to depend strongly on the channel length; possible reasons are outlined. For comparable 2D carrier concentrations, saturation currents in 2D graphene is found to be higher than Si MOSFETs and III-V Nitride HEMTs – a feature that is attractive for various applications.

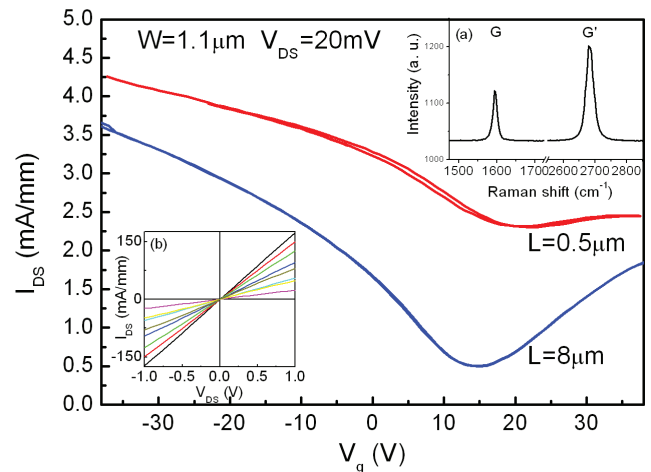


Fig. 1 (color online): I_{DS} as a function of V_g at $V_{DS}=20$ mV for short-channel (0.5 μ m) and long-channel (8 μ m) back-gated graphene FETs. Inset (a) Raman spectrum of single layer graphene. Inset (b) Low field I-V curves of long-channel FETs with $V_{DS}=-35$ V, $\Delta V_{DS}=+10$ V.

II. EXPERIMENT

Graphene flakes on heavily n-type doped silicon wafers with $t_{ox}=300$ nm thick thermal oxide from Graphene Industries [4] were used for the experiments. Single layer graphene flakes were identified by Raman spectroscopy (inset (a) of Fig. 1) [5]. The wafers were backside-metalized after oxide removal in HF to form back-gate contacts. E-beam evaporated Cr/Au (2/200 nm) was used to define the drain and source contacts by E-beam lithography. The source-drain separations ranged from $L = 250$ nm - 8 μ m. After metal deposition and liftoff, the samples were annealed in forming gas at ~350 °C for ~2 hours to remove the e-beam resist residue [6]. The graphene flakes were then patterned by O₂ plasma reactive ion etch with PMMA masks to widths ranging from $W = 1 - 10$ μ m. The SEM image (inset of Fig. 2) shows a typical FET. High current annealing [7] was performed to drive off impurities for some FETs to recover their intrinsic performance. The devices were measured using a semiconductor parameter analyzer in ambient environment and in vacuum (5×10^{-5} Torr), at room temperature and at 77 K.

III. RESULTS AND DISCUSSION

The drain currents of the FETs were first measured at a low bias of $V_{ds}=20$ mV, while the gate voltage was varied over a wide range. The gate leakage current was many orders of magnitude lower than the drain current. Fig. 1

shows representative characteristics of a long- and a short-channel FET. Over the same range of gate overdrives, the long channel FET was observed to exhibit higher gate modulation ($\sim 8x$) than the short-channel FETs ($\sim 2x$). The field-effect mobilities were calculated to be $\sim 2000 - 4000 \text{ cm}^2/\text{V}\cdot\text{s}$ for long-channel FET and $\sim 200 \text{ cm}^2/\text{V}\cdot\text{s}$ for short-channel FET respectively. These characteristics remained similar at different pressures, as well as at lower temperatures.

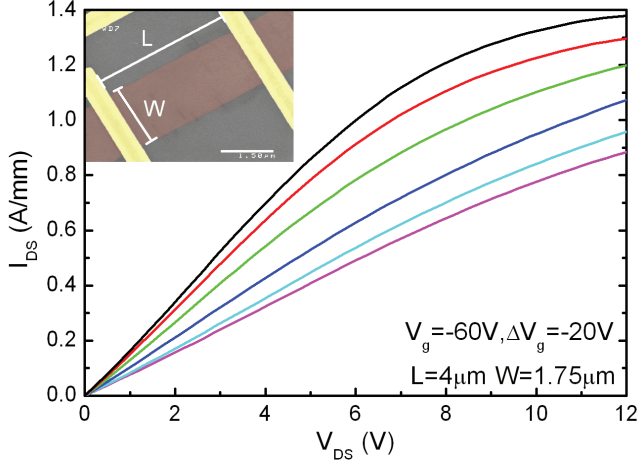


Fig. 2 (color online). I-V curves of a long-channel FETs with varied gate bias. Inset is a SEM image of this FET. The scale bar is $1.5 \mu\text{m}$.

Recently, saturation of drain current at high source-drain biases has been reported, and theory of optical phonon emission limited saturation current has been proposed [8, 9]. In particular, back-and top-gated 2D graphene FETs showed saturation current densities in the $0.5 - 0.7 \text{ A/mm}$ range [9]. In Figure 2, the high bias characteristics of a back-gated FET ($W/L = 1.75/4 \mu\text{m}$) are shown, along with an SEM image of the device. A saturation current of 1.38 A/mm is measured for this device. Saturation current densities in the $1.0 - 1.4 \text{ A/mm}$ range were measured on a number of devices, which is $\sim 2x$ higher than earlier reported values. The effective 2D carrier densities were in the $2-4 \times 10^{12}/\text{cm}^2$ range, as estimated from effective gate voltage on the back gate. The $\sim 0.5 - 0.7 \text{ A/mm}$ saturation current reported earlier [9] has been attributed to substrate-induced optical phonon scattering, but the $\sim 2x$ higher current drives measured here on identical SiO_2 substrates indicate that further investigation is necessary to clarify the mechanisms responsible for current saturation in graphene.

Such current drives can be obtained in Si MOSFETs and III-V HEMTs, but only with a) high 2DEG densities ($>10^{13}/\text{cm}^2$), and/or b) very short channel lengths taking advantage of ballistic transport. Our measurements show that Graphene FETs are able to achieve these high current densities without these criteria, implying that even higher current drives are possible when they are met in the future. For a rough comparison, a carbon nanotube with diameter $\sim 2 \text{ nm}$ and a saturation current $\sim 25 \mu\text{A}$ has an effective current per the circumferential width $\sim 4 \text{ A/mm}$ [10], much higher than either graphene or other semiconductors.

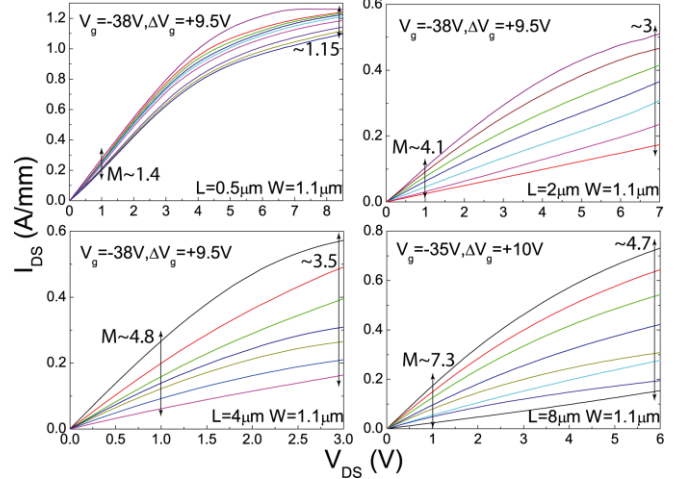


Fig. 3 (color online). Current density as a function of drain biases for FETs with different channel lengths varied from $0.5 \mu\text{m}$ to $8 \mu\text{m}$. Clear loss of gate modulation is observed with the shrinking of the graphene channel length.

The dependence of high bias drain currents and their modulation efficiency with the back-gate is observed to depend strongly on the channel length of graphene. In Fig. 3, the FET characteristics are shown for four different channel lengths, ranging from 0.5 to $8 \mu\text{m}$. Two features are observed – a) for the same channel length, the gate modulation at higher drain biases is lower, as is expected from the saturation of the current, and b) for the same bias voltages (or fields), the gate modulation efficiency decreases sharply as the channel length reduces. As shown in Fig. 3, the gate modulation reduces from ~ 7 to ~ 1.4 as the source-drain separation is scaled from $8 \mu\text{m}$ to $0.5 \mu\text{m}$. In this letter, we provide a qualitative explanation for this behavior; a detailed model will be presented in a more comprehensive later work.

Due to the absence of a bandgap in graphene, charge exchange is expected at the metal-graphene S/D contacts owing to the difference in the work functions. Considering the work function of Cr ($W_{\text{Cr}} \sim 4.5 \text{ eV}$), and that of pristine graphene ($W_{\text{Gr}} \sim 4.5 \text{ eV}$) one would expect them to form a flat-band (or ‘neutral’) contact without any charge transfer. However, recent calculations have shown that the necessary condition for the formation of a neutral contact with a metal (work function W_{M}) in intimate contact with graphene is $W_{\text{M}} - W_{\text{gr}} \sim 0.9 \text{ eV}$ [11]. Therefore, we expect that graphene region adjacent to the Cr contact pad is effectively doped with excess carriers. This excess charge region extends over an effective Debye length from each contact, and it results in S/D extension regions. When the S/D separation is smaller than twice the Debye length, the channel conductivity is controlled by the S/D contacts, and the back-gate gradually loses the capability to modulate the current, as is seen in the FET characteristics in Figs. 1 & 3.

Occasionally, naturally occurring thin parallel strips of graphene were observed. An SEM image of the device is shown in inset (a) of Fig. 4. Inset (b) of the same figure shows the consecutive burnout of the strips at high bias conditions after reaching clear current saturation; the saturation current in each strip scaled to the widths is also shown as a function of an effective channel electric field ($\sim V_{\text{ds}}/L$). Saturation current densities exceeding 1 A/mm are

measured before burnout of the strips, and the current saturation is observed at effective channel fields of ~ 15 - 20 kV/cm, in agreement with the long-channel FETs (for example shown in Fig. 2) with much wider channels. After the burnout, the devices were re-examined, and the recently reported ‘memory’ effect was observed (not shown), confirming such behavior [12, 13].

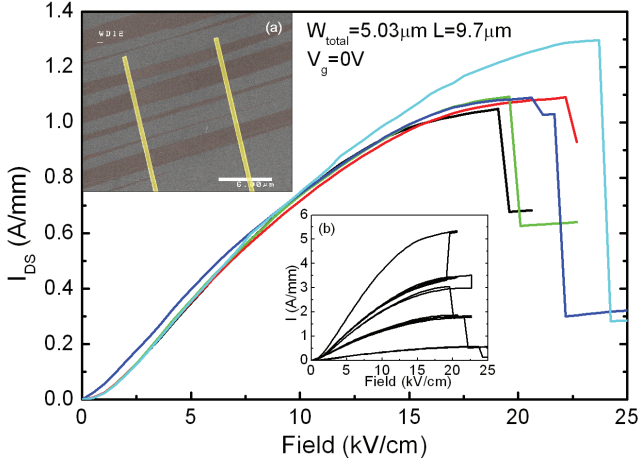


Fig. 4 (color online). Consecutive breakdown of a naturally multi-ribbon device. The breakdown occurs at >20 kV/cm acceleration field following saturation at >1.0 A/mm current density. Inset (a) SEM image of the multi-ribbon device. The scale bar is $6 \mu\text{m}$. Inset (b) the breakdown process without scaling the currents by the actual width.

IV. TRANSPORT IN NANORIBBON FETs

We have patterned graphene flakes successfully forming nanoribbons down to 30 nm. To achieve this we used 20 nm thick Al as a mask patterned by e-beam lithography using PMMA and lift-off. Temperature dependent measurements confirmed that >26 meV band-gap opened in the devices depending on the width and as a result we observed $10\times$ top gate modulation at room temperature and $10^4\times$ modulation at 4 K by varying the top gate potential between ± 2.5 V.

Operating the device at high source-drain potential we observed nonlinearity in the source-drain current but the detailed analysis showed that this is not due to saturation but rather because of the relative band alignment of the p and n regions controlled by top- and back-gates. The doping of the graphene at the contact regions of the devices are possibly

determined by the work function difference of the contact metal and the graphene [14]. If one chooses a metal with work function very similar to graphene this effect can be minimized and we can model the carrier density solely as a function of electrostatics. Cr was our choice ($W_{\text{Cr}} \sim W_{\text{gr}} = 4.5$ eV) for ohmic metallization.

We observed two distinct operational modes of the device. In the cases when both the back and the top gate bias have the same sign the GNR is homogeneously doped n or p type and this corresponds to a high conductivity state. If the top gate bias has an opposite sign the GNR is doped npn or pnp and this corresponds to a low conductivity state due to the barrier. In conventional semiconductor materials this would be the off-state, but due to the tiny bandgap of the GNR this barrier is just a double tunnel junction. At high enough source-drain bias the lateral field starts to compensate the vertical field of the gate and pulls back the band to homogenous state (entirely n or entirely p) and switch back the device to the high conductance state. This is a completely new phenomenon attributable to the band gap opening in the ribbons which cannot be observed in 2D graphene FETs. We weren’t able to drive the GNR FETs to the same saturation regime which is observed in 2D GFETs due to high gate leakage. But we can conclude that the observed sublinear behavior is not due to phonon limited saturation, but rather due to the onset of tunneling through the GNR bandgap, a first step towards realizing GNR-based tunnel FETs [15].

V. CONCLUSIONS

In conclusion, saturation current densities in the 1 - 1.5 A/mm range have been measured in exfoliated 2D graphene FETs, comparable to Si-MOSFETs and III-V Nitride HEMTs. Contact-induced short channel effects cause a strong degradation of the gate modulation of the drain current, which should have implications on the scaling of graphene devices. The high current drives in graphene are achieved at far lower 2D carrier concentrations and longer channel lengths, implying that the current drives in graphene FETs is superior to conventional semiconductors. This can be a clear advantage for a number of applications, including those for RF amplifiers and circuits where the zero bandgap is not of the utmost concern.

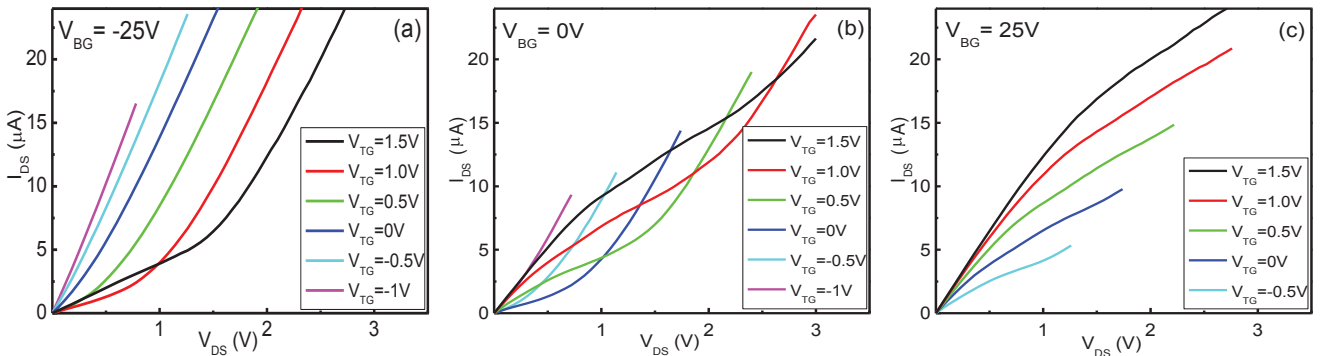


Fig. 5 (color online). Output characteristics of the GNR FET ($L = 4 \mu\text{m}$, $W = 75$ nm) at different top gate biases. (a) If both gate is negative the resistance is low. In case of positive top gate bias the resistance is high until the drain bias compensate the top gate bias. (b) At zero top gate bias the transition between case (a) and (c) can be observed. (c) High to low resistance transition occurs when the gate to drain potential is about zero.

ACKNOWLEDGMENT

The author thank Rachel Rasmussen and Rachel Masyuko for Raman spectroscopy measurements. Discussions with Xiangning Luo, Huili Xing and Gregory Snider are gratefully acknowledged. The authors also gratefully acknowledge financial support from NSF Award Nos. DMR-06545698 and ECCS-0802125 and from the Nanoelectronics Research Initiative (NRI) through the Midwest Institute for Nanoelectronics Discovery (MIND).

REFERENCES

- [1] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva and A. A. Firsov, "Electric field effect in atomically thin carbon films," *Science*, vol. 306, pp. 666-669, Oct. 2004.
- [2] Y. Zhang, Y.-W. Tan, H. L. Stormer and P. Kim, "Experimental observation of the quantum Hall effect and Berry's phase in graphene," *Nature*, vol. 438, pp. 201-204, Nov. 2005.
- [3] A. K. Geim and K. S. Novoselov, "The rise of graphene," *Nat. Mat.*, vol. 6, pp. 183-191, March 2007.
- [4] <http://www.grapheneindustries.com>. (For preparation of samples, see K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim, "Two-dimensional atomic crystals," *PNAS*, vol. 102, pp. 10451-10453, July 2005.)
- [5] A. C. Ferrari, J.C. Meyer, V. Scardaci, C. Casiraghi, M. Lazzeri, F. Maure, S. Piscanes, D. Jiang, K. S. Novoselov, S. Roth, and A. K. Geim, "Raman spectrum of graphene and graphene layers", *Phys. Rev. Lett.*, vol. 97, no. 18, p.187401, Nov. 2006.
- [6] M. Ishigami, J. H. Chen, W. G. Cullen, M. S. Fuhrer, and E. D. Williams, "Atomic structure of graphene on SiO₂", *Nano Lett.*, vol. 7, no. 6, pp.1643-1648, Apr. 2007
- [7] J. Moser, A. Barreiro, and A. Bachtold, "Current-induced cleaning of graphene", *Appl. Phys. Lett.*, vol. 91, no. 16, p. 163513, Oct. 2007.
- [8] X. Luo, Y. Lee, A. Konar, T. Fang, G. Snider, H. Xing, and D. Jena, "Current-carrying capacity of long and short-channel 2D graphene transistors," *IEEE DRC Tech. Digest*, p 29, June 2008.
- [9] I. Meric, M. Y. Han, A. F. Young, B. Oezylmaz, P. Kim and K. L. Shepard, "Current saturation in zero-bandgap, top-gated graphene field-effect transistors" *Nature Nanotech.*, vol 3, pp. 654-659, Nov 2008.
- [10] Y-F. Chen and M. Fuhrer, "Electric-Field-Dependent Charge Carrier Velocity in Carbon Nanotubes", *Phys. Rev. Lett.*, vol 95, pp. 236803, Dec 2005.
- [11] G. Giovannetti, P. A. Khomyakov, G. Brocks, V. M. Karpan, J. v d Brink, and P. J. Kelly, "Doping Graphene with Metal Contacts", *Phys. Rev. Lett.*, vol 101, pp. 026803, July 2008.
- [12] B. Standley, W. Bao, H. Zhang, J. Bruck, C. N. Lau, and M. Bockrath, "Graphene-based atomic-scale switches", *Nano Lett.*, vol 8, pp. 3345-3349, Aug 2008.
- [13] Y. Li, A. Sinitskii, and J. M. Tour, "Electronic two-terminal bistable graphitic memories," *Nat. Mat.*, vol 7, pp. 966-971, Nov 2008.
- [14] G. Giovannetti, P. A. Khomyakov, G. Brocks, V. M. Karpan, J. van den Brink, and P. J. Kelly, "Doping graphene with metal contacts," *Phys. Rev. Lett.*, vol. 101, p. 026803, July 2008
- [15] Q. Zhang, T. Fang, A. Seabaugh, H. Xing, and D. Jena, "Graphene nanoribbon tunnel transistors," *IEEE Electron Device Lett.*, vol. 29 (12), pp. 1344-1346, Dec. 2008.

Efficient Routing and Communication in Wireless Systems

Gergely Treplán
(Supervisor: Dr. János Levendovszky)
trege@digitus.itk.ppke.hu

Abstract—In this paper a reliability based routing algorithm is proposed for Wireless Sensor Networks (WSNs). The evaluation metric of links is calculated by using an arbitrary (e.g. Rayleigh, Rice) fading model. The proposed scheme minimizes the energy consumption and ensures reliable packet transmission to the Base Station (BS) at the same time. Reliability is guaranteed by selecting the path over which the probability of correct packet reception at the BS will exceed a predefined threshold. Meanwhile energy efficiency is reached by energy balancing (i.e. minimizing the energy consumption of the bottleneck node of the routing path). It will be derived that reliable and energy efficient packet forwarding by WSN can be reduced to a constrained optimization problem. By using a specific link metric this problem can then be mapped into a shortest path problem solved in polynomial time. Thus the obtained results make possible reliable path selection with minimum energy consumption in real time.

Index Terms—wireless sensor networks, reliability, generic fading model, bottleneck node, energy awareness

I. INTRODUCTION

Due to the recent advances in electronics and wireless communication, the development of low-cost, low-energy, multifunctional sensors have received increasing attention [3]. These sensors are compact in size and besides sensing they also have some limited signal processing and communication capabilities. However, these limitations in size and energy make the WSN-s different from other wireless and ad-hoc networks [4]. As a result, new protocols must be developed with special focus on energy effectiveness in order to increase the lifetime of the network which is crucial in case applications, where recharging of the nodes is out of reach (e.g. military field observations, living habitat monitoring etc., for more details see [4]). This paper addresses reliable packet transmission in WSN when packets are to be received on the Base Station (BS) with a given reliability in terms of keeping the error probability under a given threshold. Since the success of every individual packet transmission depends on the distance and the energy of transmission, the probability of correct reception will diminish exponentially with respect to the number hops, in the case of multi-hop packet transfers. In the paper a new approach is introduced to minimize the energy consumption of the bottleneck node on the routing path subject to the constraint of guaranteeing reliable packet transfer to the BS. The optimal path from source node i_1 to BS is represented by a set of indices $\mathfrak{R}_{opt} = (i_1, i_2, \dots, i_L)$ where the nodes $i_j, j = 1, \dots, L$ are selected for packet forwarding from node i_1 (where the packet is originated) to the BS. The

reliability of this packet transfer is $Reliab = \prod_{j=1}^L P_{i_{j-1}i_j}$, where $P_{i_{j-1}i_j}$ denotes the probability of successful packet transfer from node i_{j-1} to node i_j . The reliability constraint imposes that $Reliab = \prod_{j=1}^L P_{i_{j-1}i_j} \geq 1 - \varepsilon$ for a given ε .

In this paper we will demonstrate that the selection of \mathfrak{R}_{opt} can be carried out in polynomial time by using our proposed algorithm. We compare the achieved lifespan and reliability to the corresponding parameters of the traditional protocols by performing extensive simulations.

II. RELATED WORK

Many research works have been published in the topic of reliable data transport in WSNs recently. These proposed solutions can be classified into two groups: (i) guaranteed delivery approaches; and (ii) stochastic delivery approaches [9].

In guaranteed delivery (or in other words, packet-loss recovery[8]) approaches, one must guarantee the successful arrival of the packet at the destination. Hence, lost packets are recovered by retransmissions. Considering the technique of recovery, these retransmissions can be end-to-end recovery or per-hop recovery. Recently, per-hop recovery was advocated in many research works[10], since it is easy to manage. On the other hand, end-to-end recovery is believed to be not suitable for WSNs due to the big latency and the large energy cost. In stochastic delivery (or in other words, packet-loss avoidance[8]) approaches, one must choose routing paths such that the occurrences of packet loss on those paths are minimized. In these methods the possible forwarding nodes are carefully evaluated and the node of a higher probability of delivery is then selected as a forwarding node. However, the applied evaluation metrics vary in different approaches. For instance, in GeRaF [11] the geographic distance and a loss-aware metric in ETX [13] was used. Beside previous approaches, another way of packet-loss avoidance is to use multi-path transmissions, such as Directed Diffusion[1] and ReINFORM[12]. Our work is similar to those ones belonging to the latter group. We propose a stochastic delivery based method which guarantees the probability of successful delivery is at least $1 - \epsilon$ for any packets. Moreover, we use the more realistic generic fading model which could be any arbitrary (e.g. Rayleigh and Rice fading models[5], [15]) to calculate the evaluation metric of links in our model. In our work, we

aim to maximize the life span of the WSN taking energy balancing into account as well. To achieve this, our goal is to minimize the energy consumption of the bottleneck nodes on the forwarding paths to avoid the fast depletion of important forwarding nodes in the network.

III. THE MODEL

WSN is perceived as an arbitrary 2D topology of nodes where packet is forwarded from a source node to the BS in a multihop fashion. The selected path can be represented by a 1D chain and described by a set of indices $\mathfrak{R} = (i_1, i_2, \dots, i_L)$ as shown in Figure 1.

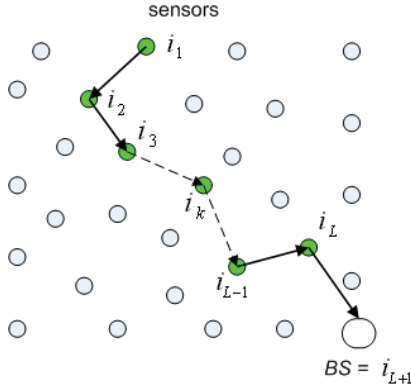


Fig. 1. Packet forwarding path from node i_1 to the BS in WSN

In the generic fading model the probability $P^{(r)}$ of correct reception of transmitting the packet to distance d with the given transmission power g can be given as

$$P^{(r)} = \Psi(d, g) \quad (1)$$

Furthermore, when the packet transfer takes place between two nodes i and j in the chain, then the corresponding reliability is $P_{ij}^{(r)} = \Psi(d_{ij}, G_{ij})$, where G_{ij} denotes the transmission energy consumption on node i sending a packet to node j and d_{ij} is the distance between those two nodes. We assume that $\Psi(d, g)$ is strictly monotone increasing as g is increased. For instance, we have the following formula by using the Rayleigh fading model:

$$P_{ij}^{(r)} = \exp\left\{\frac{-d_{ij}^\alpha \Theta \sigma_Z^2}{G_{ij}}\right\} \quad (2)$$

where Θ is the modulation constant, σ_Z^2 denotes the energy of noise and α depends on the propagation type, respectively. The range of α is usually $2 \leq \alpha \leq 6$.

In the case of the Rice fading model, the Rice probability density function (PDF) of the receiving power is given by:

$$f_{ij}^{(r)}(p) = \frac{(K+1)e^{-K}}{R} e^{\{-\frac{p(K+1)}{R}\}} I_0\left(\sqrt{\frac{4K(K+1)p}{R}}\right) \quad (3)$$

where $I_n(\cdot)$ is the n^{th} order modified Bessel function of the first kind, K is the measure of fading which is the ratio of the

power received via the direct line-of-sight(LOS) path to the power contribution of non-LOS paths and R is the average received power that can be calculated as follows:

$$R = (K+1)G_{ij}d_{ij}^{-\alpha} \quad (4)$$

To have the $P_{ij}^{(r)}$ in the Rice fading model, one must calculate $P_{ij}^{(r)}(G) = 1 - F_{ij}^r(N_0\Theta)$ where $F_{ij}^r(\cdot)$ is the cumulative distribution function(CDF) of formula (3) and N_0 is the Gaussian noise energy value. Let us note that in the case of $K=0$ we get the Rayleigh fading model. The estimation of K for can be found in [16]. For more details see [15].

For transmitting packets over a distance d , beside the transmitting energy consumption given in equation (1) there are also energy consumptions of the sensors electronics at both packet transmission and packet receiving. However, these energy consumptions is independent from the distance between the communicating nodes, therefore they can be assumed as constants (for more details, see [6]). Let G_T denote the energy consumption of the electronics at transmitting, and G_R denote the energy consumption at receiving. Without the loss of generality, we can assume that the transmission of each packet needs the same $\Delta T = 1$ time. Therefore the transmission energy consumption of a single packet is

$$G_{ij}\Delta T = G_{ij} \quad (5)$$

Hence G_{ij} can denote both the transmission power and transmission energy consumption of a single packet as well, which may be a slight abuse of notation. However, in this paper it will be clear that which one of the meanings above is used in each cases. Let $c_j(k)$ denote the residual energy of the j^{th} node at time instant k .

The residual energy level of the bottleneck node after a packet transfer described by the set of indices $\mathfrak{R} = (i_1, i_2, \dots, i_L)$ is characterized by $\min_{i_i} c_{i_i}(k+1)$ where

$$c_{i_i}(k+1) = c_{i_i}(k) - G_{i_i i_{i+1}} - (G_R + G_T) \quad (6)$$

if i_i is not the source node and

$$c_{i_1}(k+1) = c_{i_1}(k) - G_{i_1 i_2} - G_T \quad (7)$$

for the case of source node.

Let $G_{i_L BS}$ denote the last transfer from node i_L to the BS. For the sake of the brevity, let the BS be the node i_{L+1} . Our objective is to find the optimal path $\mathfrak{R}_{opt} = (i_1^{opt}, i_2^{opt}, \dots, i_L^{opt})$ which maximizes the remaining energy of the bottleneck node of the packet transfer from the source to the BS. This optimization is expressed as:

$$\mathfrak{R}_{opt} : \max_{\mathfrak{R}} \min_{i_i} c_{i_i}(k+1) \quad (8)$$

and is subject to guaranteeing that the packets arrive at the BS with a given reliability $1 - \varepsilon$, in terms of enforcing the condition

$$P(\text{Reliab}) = \prod_{l=1}^L \Psi(d_{i_l i_{l+1}}, G_{i_l i_{l+1}}) \geq 1 - \varepsilon \quad (9)$$

One must note that this problem depends not only on the set of paths from node i and ending at the BS but also on the corresponding transmission energies. Thus, we solve this problem in two phases. In the first phase, we assume that the path $\mathfrak{R} = \{i_1, i_2, \dots, i_L\}$ is given. In this case, we are only concerned with finding the optimal transmission energies which maximizes $\min_{i_i} c_{i_i}(k+1)$ subject to the reliability criterion. In the second phase, we determine not only the optimal energies but \mathfrak{R}_{opt} as well, i.e. the optimal packet forwarding route that guarantees the $1 - \varepsilon$ reliability and maximizes the remaining energy of the bottleneck node.

IV. MINIMUM ENERGY PATH SELECTION WITH RELIABILITY CONSTRAINTS

In this section, the minimum energy path selection with reliability guarantees is mapped into a constrained optimization problem and then a polynomial complexity solution is proposed. This goal is achieved in the two phases. First, we assume that the optimal path has already been selected and we are only concerned with minimizing the transmission energies over this path. Secondly, we extend the optimization to identifying the optimal path.

A. Energy minimization over a given routes

First we assume that the routing path \mathfrak{R} is already given. Our goal is to determine the energies by which the nodes must forward a packet to the BS in order to achieve minimal overall energy consumption on the paths subject to the reliability constraint. Since the energy consumption of the electronics are constants at all nodes, we can ignore them by doing the following modifications. Let $c_{i_1}(k) = c_{i_1}^0(k) - G_T$ and $c_{i_l}(k) = c_{i_l}^0(k) - (G_T + G_R)$ for each $l > 1$, where $c_{i_l}^0(k)$ is the original energy level of node i_l and $c_{i_l}(k)$ is the modified one at time instant k . Hence, one can state the following.

Theorem 1: *Assuming that the packet transmission path $\mathfrak{R} = \{i_1, i_2, \dots, i_L\}$ from node i_1 to the BS is given, under the reliability parameter $(1 - \varepsilon)$, $\min_{i_i} c_{i_i}(k+1)$ can only be maximal if the residual energy of each node is the same, expressed as $c_{i_i}(k) - G_{i_i i_{i+1}} = A$, and A satisfies the following equation:*

$$\prod_{l=1}^L \Psi(d_{i_l i_{l+1}}, c_{i_l} - A) = 1 - \varepsilon \quad (10)$$

It is easy to note that the left had side of formula (10) is monotone decreasing with respect to parameter A . Thus (10) will have a unique solution over the interval $\left(0, \min_{i_i} c_{i_i}(k)\right)$. If there is no solution then there is no such energy set $(G_{i_1 i_2}, G_{i_2 i_3}, \dots, G_{i_L i_{L+1}})$ which could fulfil the reliability constraint. Due to its monotonicity, one can develop fast method to solve (10) as the Newton-Raphson algorithm. Based on this theorem, in the case of predefined routes we can determine the optimal transmission energies which yield maximal lifespan by using (10).

B. Finding the optimal path

Now we investigate the path selection and point out that the minimum energy path subject to reliability constraint can be found in polynomial time. Having A at hand we can search for the most reliable path when the maximization of is equivalent with the minimization of $-\sum_{i_i \in \mathfrak{R}} \log(\Psi(c_{i_i}(k) - A))$.

This formula reduces the search for the most reliable path into a shortest path optimization problem where the measure $-\log(\Psi(c_{i_i}(k) - A))$ is assigned to each link. The task

$$\mathfrak{R}_{opt} : \min_{\mathfrak{R}} \sum_{i_i \in \mathfrak{R}} -\log(\Psi(d_{i_i i_{i+1}}, c_{i_i} - A)) \quad (11)$$

can be solved in polynomial time by any shortest path routing algorithm. (Note that $-\log(\Psi(d_{i_i i_{i+1}}, c_{i_i} - A)) \geq 0$). Let us note that by applying the Rayleigh model (as described by equation (2)) and with setting $A = 0$, expression (8) is equivalent with the optimization task solved by the PEDAP-PA algorithm [2].

It is easy to see that the solution of (11) depends on the value of A , however the optimal value of A depends on the path itself. Therefore, let us solve (11) and (10) recursively, one after another. This implies that we search from the most reliable path and then for the path found we make sure that the reliability constraint holds obtaining the value of A which belongs to the reliability parameter $(1 - \varepsilon)$. This algorithm will have a fix point and will stop when there are no changes in the obtained paths any longer. The convergence to the optimal solution is stated by the following theorem:

Theorem 2: *Let $A(k)$ indicates the series obtained by recursively solving (11) and (10) one after another. $A(k)$ is monotonically increasing and will converge to the fix point of (11) and (10). Furthermore (11) and (10) has a unique fix point. Hence the algorithm described above and depicted by Algorithm 1. converges to the global optimum.*

Algorithm 1 Calculate $[A, path]$

Require: $c_i > 0, \forall i$

Ensure: $[A, path]$

$A \leftarrow 0$

while $path \neq path_{old}$ **do**

$path_{old} \leftarrow path$

$E_{l_j} \leftarrow -\log(\Psi(d_{l_j}, c_l(k) - A)), \forall l_j$

$path \leftarrow DIJKSTRA(E)$

$A \leftarrow OptResEnergies(path)$

end while

Furthermore it can be shown the convergence speed of this algorithm is independent of the network size N , hence the complexity of our algorithm is still $O(N^2)$.

V. NUMERICAL RESULTS

In this section the performance of the new reliability based routing algorithm is analyzed and compared with the standard WSN routing algorithms. For the sake of brevity, let us denote the proposed algorithm as BERA (**B**ottleneck **E**nergy

Aware Reliability based Algorithm). Beside this, we have assigned values to the remaining parameters (e.g. G_R and G_T) based on the widely used RF module of the CC2420 (the specification of the mote can be found in [7]). We have compared the performance of BERA to such commonly used routing protocols in WSN like LEACH[14], PEDAP-PA[2] and Directed Diffusion(DD)[1]. Figure 2 indicates the number the percentage of the operational node as a function of time. One can see that in the case of the novel algorithm the nodes go flat more or less at the same time. Furthermore, the longevity of the first node going flat has been significantly improved compared with the traditional methods.

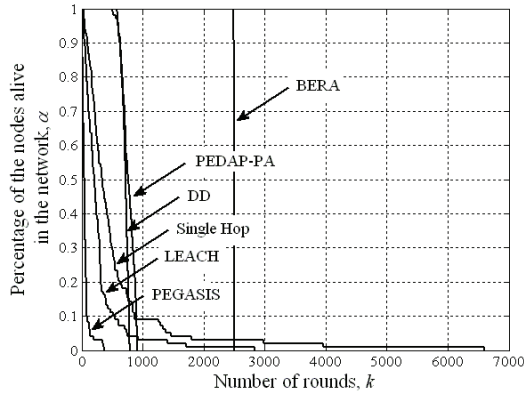


Fig. 2. The percentage of the operational node as a function of time

So far, we adopted the Rayleigh model for performance analysis, which is valid if the sensor antennas do not "see" each other. This typically occurs in indoor applications. If the antennas can see each other then the Rice fading [16] describes the radio channel better. Figure 3 indicates, how the lifespan increases when the amplitude of the dominant wave grows and it also demonstrates the effect of errors in the fading parameter estimation on the lifespan. Based on the numerical results, the new protocol can outperform the traditional ones and it can be applied in any application when longevity and reliability are of major concerns.

VI. CONCLUSIONS

In this paper, a generic fading model based approach has been introduced for reliable energy aware routing in WSNs. We also proposed a routing scheme to find the optimal path. It has been shown, that this algorithm gives the globally optimal solution which yields minimum energy consumption on the bottleneck node of the paths used for packet transfer. At the same time, the constraint of reliability is also satisfied. The performance of the protocol has been tested by extensive simulations which also demonstrated the improvement on the lifespan.

ACKNOWLEDGMENT

The research reported here was supported by the National Office for Research and Technology (Mobile Innovation Cen-

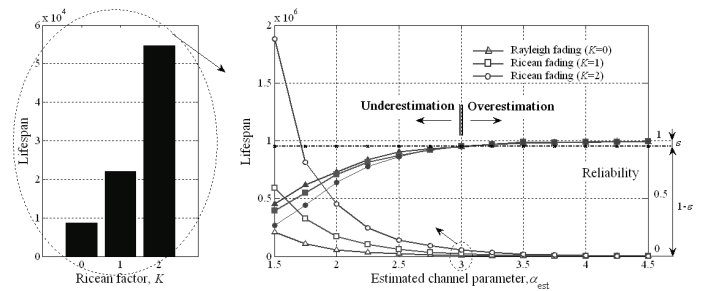


Fig. 3. The impact of fading parameter estimation on the lifespan and reliability (the reliability parameter was set as $\epsilon = 0.05$, while the exponent of fading attenuation was set as $\alpha = 3$). As can be seen by underestimating the fading parameters the lifespan will increase but it puts the reliability in jeopardy. In the case of overestimation the reliability criterion is satisfied but with excessive energy consumption (too large transmission energies are selected) and the life-span is decreased.

tre 2.1.3 project). The author would like to thank András Oláh Ph.D. for his assistance and advice.

REFERENCES

- [1] C. Intanagonwiwat, R. Govindan and D. Estrin., "Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks." *ACM MOBICOM 2000*.
- [2] H. O. Tan and I. Korpeoglu, "Power Efficient Data Gathering and Aggregation in Wireless Sensor Networks." *ACM SIGMOD Record*, vol. 32, No. 4, pp. 6671, December 2003.
- [3] C.Y. Chong and S.P. Kumar, "Sensor networks: Evolution, opportunities, and challenges." *IEEE Proceedings*, pp. 12471254, August 2003.
- [4] A. Goldsmith and S. Wicker, "Design challenges for energy-constrained ad hoc wireless networks." *IEEE Wireless Communications Magazine*, vol. 9, pp. 827, August 2002.
- [5] M. Haenggi, "Analysis and Design of Diversity Schemes for Ad Hoc Wireless Networks." *IEEE journal on selected areas in communications*, vol. 23, no. 1, 2005.
- [6] Q. Wang, M. Hempstead and W. Yang, "A Realistic Power Consumption Model for Wireless Sensor Network Devices." *Third Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, Veston, VA, September 2006.
- [7] Chipcon, SmartRF CC2420, 2.4GHz IEEE 802.15.4/ZigBee-ready RF Transceiver
- [8] Y. Liu, Y. Zhu and L. Ni, "Reliability-oriented Transmission Service in Wireless Sensor Networks." *The Fourth IEEE International Conference on Mobile Ad-hoc and Sensor Systems*, Pisa, October 2007.
- [9] H. Karl and A. Willig, "Architectures and Protocols for Wireless Sensor Networks, Chapter 13." *Chichester: John Wiley & Sons*, 2005.
- [10] Q. Cao, T. He, L. Fang, T. Abdelzaher, and J. S. Son. "Efficiency centric communication model for wireless sensor networks." *INFOCOM*, 2006.
- [11] M. Zorzi and R. R. Rao. "Geographic random forwarding (GeRaF) for ad hoc and sensor networks: Multihop performance." *IEEE Transactions on Mobile Computing*, 2, 2003.
- [12] B. Deb, S. Bhatnagar, and B. Nath. "Reinform: Reliable information forwarding using multiple paths in sensor networks." *ACM MobiCom*, pp. 406415, 2001.
- [13] D. S. J. D. Couto, D. Aguayo, J. Bicket, and R. Morris. "A high-throughput path metric for multi-hop wireless routing." *ACM MobiCom*, 2003.
- [14] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocols for Wireless Microsensor Networks," *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, pp. 1-10, January 2000.
- [15] T.S. Rappaport, "Wireless Communications, Principles and Practice, 2nd ed.," Prentice Hall, Englewood Cliffs, NJ. 2001.
- [16] A. Abdi, C. Tepedelenlioglu, M. Kaveh and G. Giannakis, "On the estimation of the K parameter for the Rice fading distribution." *Communications Letters, IEEE Volume 5, Issue 3, Mar 2001 pp. 92-94*

Energy Balancing in Wireless Sensorial Network by Using Discrete Energies

B. Karlócai

(Supervisor: Dr. János Levendovszky)
bkarlo@digitus.itk.ppke.hu

Abstract— In this paper some novel protocols are developed for wireless sensor networks (WSNs) considering that the real applied sensors have discrete energy values in order to ensure reliable packet transmission and maximize lifespan at the same time. Former optimizations in this subject had some restriction due to the complexity of the analytical, but in this case we can use a real 2D model.

The optimal transmission energies are derived which guarantee that the packets are received by the Base Station (BS) with a given probability subject to achieving the longest possible lifespan. The algorithm is based on a single iteration loop, which can be with a single Dijkstra or Bellmann-Ford algorithm. The new results have been tested by extensive simulations which demonstrated that the lifespan of WSN can significantly be increased by the new protocols.

Index Terms— wireless communication, routing algorithm

I. INTRODUCTION

Due to the recent advances in electronics and wireless communication, the development of low-cost, low-power, multifunctional sensors have received increasing attention [1]. These sensors are compact in size and besides sensing they also have some limited signal processing and communication capabilities. However, these limitations in size and energy makes WSNs different from other wireless and ad-hoc networks [4]. As a result, new protocols must be developed with special focus on energy balancing in order to increase the lifetime of the network which is crucial in case applications, where recharging of the nodes is out of reach (e.g. military field observations, living habitat monitoring ...etc., for more details see [6]).

The paper addresses reliable packet transmission in WSN. Since packets can be forwarded to the BS in multihop manner, the probability of reliable packet transfer will diminish in exponential order with respect to the number hops (i.e. assuming independent packet losses, the probabilities of correct packet reception must be multiplied hop by hop).

As a result, if uniform reliability is to be ensured for each packet (independently of their source node), either the transmission energies must be adjusted or packets must be multiplied to increase their chance to get to the BS without error if they are generated by a far-away node.

In this paper, the optimal energies are derived which yield the longest lifespan of WSN subject to guaranteeing uniform reliability in the case chain and other protocols.

II. THE MODEL

Our model is based on a single 2D (100x100) homogenous area. The base station is positioned at the center of the rectangle. The nodes are put into the model randomly. The optimal route can be characterized as follows:

- we assume the following relation,

$$P_{u,v} = \Psi(d_{u,v}, g),$$

where P is the succeed packet transmission probability, d denote for the distance between node u and v, and g is the gain of the transmission

- the mentioned function can be a Rayleigh fading model, where the energy needed to transmit packet over

$$g = \frac{d^\alpha \Theta \sigma_z^2}{-\ln p_r} + g_{elec}$$

distance d is given as , where d is the distance, α depends on the propagation type, p_r is the reliability of correct reception, Θ is the threshold, σ_z^2 is the noise energy, while g_{elec} represents the consumption of the electronics during transmitting and receiving;

III. TRANSMISSION ROUTE CALCULATIONS WITH DISCRETE ENERGIES WITH ITERATIONS

The goal is to keep the overall arrival probability above a predefined critical level:

$$\prod_{u,v} \Psi(d_{u,v}, g) \geq 1 - \varepsilon .$$

The form can be converted to a logarithmic scale:

$$\sum_{(u,v) \in R} -\lg(\Psi(d_{u,v}, g)) \leq -\lg(1 - \varepsilon)$$

Let's define the goal as minimizing the transmission energy

$$\min(g), \quad g \in \{G_1, G_2, \dots, G_N\}$$

The problem can be derived into iteration steps

- at the first phase the energy can be any random value
- at this point the problem can be derived as

$$g_0 \rightarrow R_0 : \min \sum_{(u,v) \in R} -\lg(\Psi(d_{u,v}, g))$$

- this problem can be solved very fast with Dijkstra or Bellmann-Ford algorithm
- the second phase is a simple criteria analysis, determining if we reached our goal or not

$$\sum_{(u,v) \in R} -\lg(\Psi(d_{u,v}, g)) \leq -\lg(1 - \varepsilon)$$

If it does, we decrease the energy

$$g_1 := g_0 - \Delta,$$

if it does not we increase it

$$g_1 := g_0 + \Delta.$$

As a result, we get an algorithm we can use in polynomial time.

IV. SIMULATION ENVIRONMENT

For comparison I have implemented the well-known Leach algorithm for the same environment.

For the numerical results I have used the following parameters for the Rayleigh-fading model:

$$g_{elec} := 5$$

$$\Theta := 0.1$$

$$\sigma_z := 0.1$$

$$\alpha := 2$$

The Base Station (BS) has been put into the middle of the simulation area (50,50). The starting energies of the nodes was 20k, and the required arrival probability (P_{req}) 0.9.

The simulation has been programmed with Matlab, and ran in a 3.2 Ghz Core2 Duo processor.

V. SIMULATION ENVIRONMENT

The simulation has been prepared with a 100x100 homogenous area, and a fully random wireless network. The random network had been copied as many times as many test cases I had. During the simulation I have paid attention to generate the new package on the very same node in the network, to guarantee the equal chances.

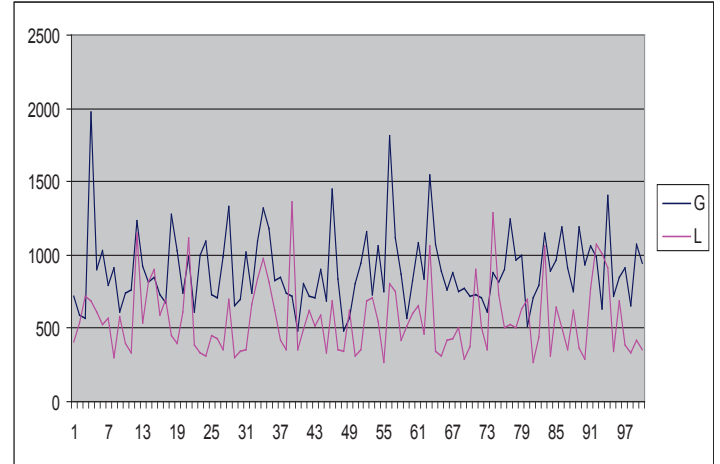
The stop criterion of the simulation was the total exhausting of any of the nodes in the network.

I have generated 100 total different networks and ran the simulation until the stopping criterion.

VI. SIMULATION RESULTS

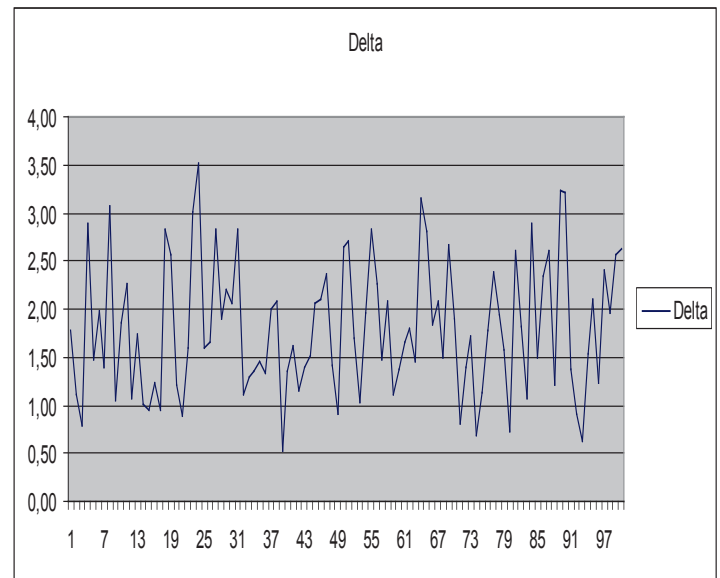
In this section we can see the actual results of the algorithms.

At the following chart we can see the number of packets (x1000) we could transfer until the exhausting of the network



1. Figure number of the transmitted packets (x1000) before the network exhausting

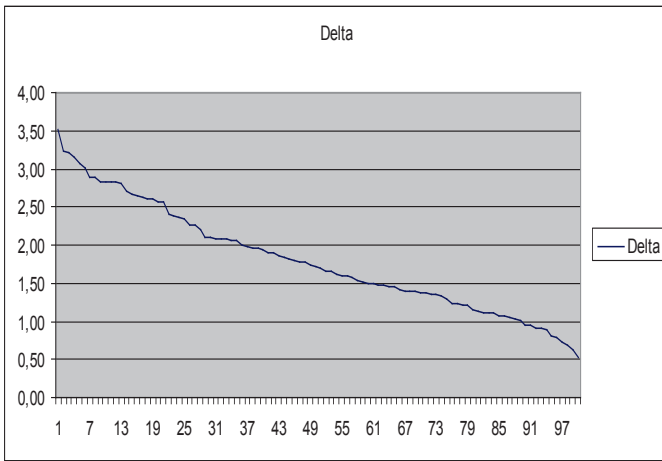
As we can see the new algorithm – denoted by “G” – showed much more efficiency compared to Leach. To see it better we should have a look at the following chart where we compare the two algorithm run-by-run by its ratio:



2. Figure the ratio of the two algorithms G divided by L

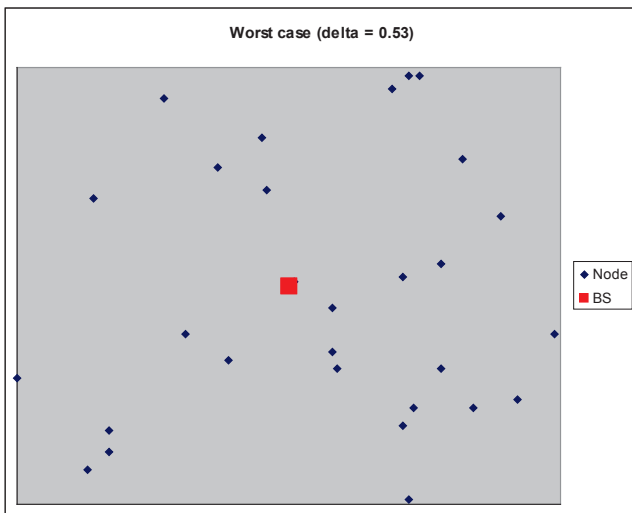
We can see that the most of the cases the chart is above 1, which means the new algorithm have better performance.

Now we can order it by its value, so we can see the point where it crosses the 1 point.

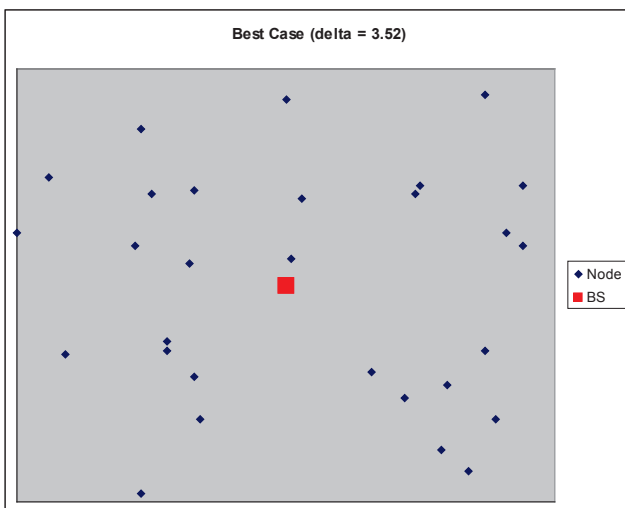


3. Figure the results crosses the 1 line at 90, which means at the 90% of the cases the new algorithm perform better

If we have a look at the worst and the best case we can say more about the boundaries



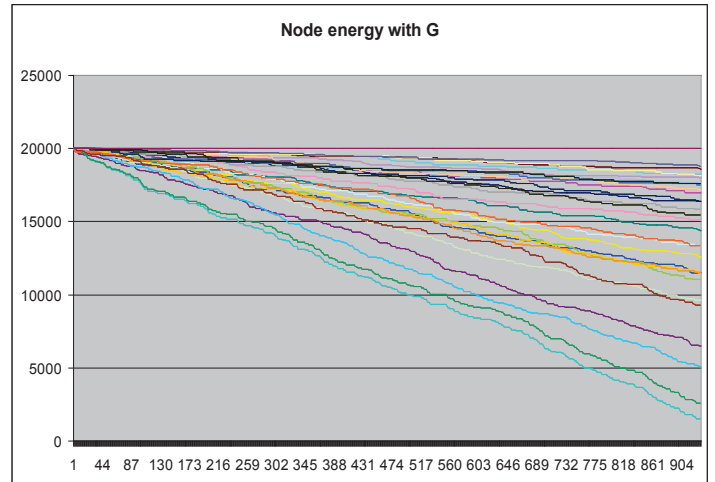
4. Figure worst case, the Leach performs better



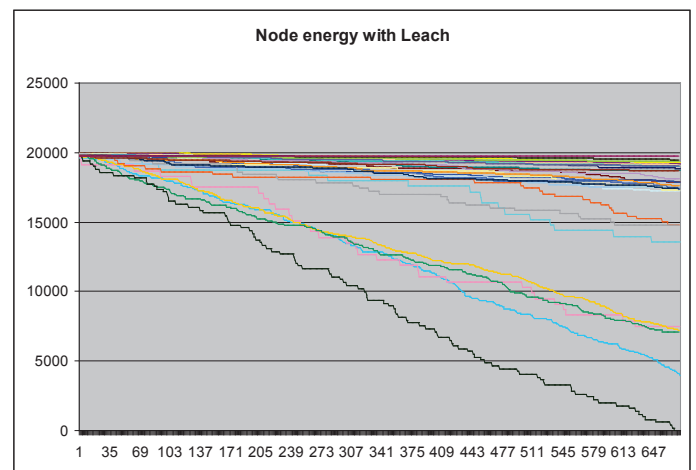
5. Figure best case: our new algorithm performs better

As we can see, the worst case had a lot of nodes around each other, while in the best case the nodes are equally distributed.

At the following two figures, we can see the energy each of the nodes.

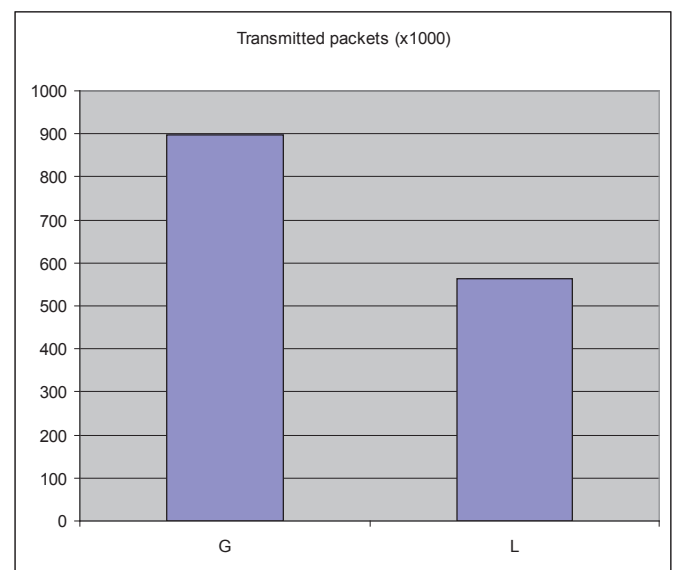


6. Figure each nodes energy in a random network, showed with the transmitted packets (x1000) with the new algorithm



7. Figure each nodes energy in a random network, showed with the transmitted packets (x1000) with the Leach algorithm

So we can see the overall results:



8. Figure overall result of the comparison of Leach and the new algorithm

As we can see the new method performed better.

VII. CONCLUSIONS

In this paper, novel energy balancing packet forwarding methods have been developed to maximize the lifespan of WSNs and to ensure reliable packet transfer at the same time. We have optimized the transmission energies of the nodes depending on the source of the packet in order to minimize the energy consumption of the bottleneck node with discrete energies, subject to satisfying a given reliability constraint. The algorithm has been developed to use the discrete energy values as an iteration base. The algorithm runs in P time complexity. The given method is able to find a good route and the energy optimization strategy is quite good as well.

The underlying protocol optimization was reduced to a Dijkstra or Bellmann-Ford algorithm, which can be solved fast and effective.

The performance of the protocol has been compared with a well known route find algorithm – the Leach algorithm. The results showed us that the new method has a good efficiency. The very same network exhausted 2-3 times later with the new algorithm than with the former Leach. We have made a long test with 100 random generated network, and we saw that the new algorithm perform in average 1.8 times better than the Leach.

At the same simulation we saw that some of the nodes still kept some energy, so there are also some opportunities to continue the research in this direction.

References:

- [1] J. Levendovszky, A. Olah, A. Bojarszky, B. Karlocai, “Energy balancing by combinatorial optimization for wireless sensor networks,” IWDN07 Conference, pp. 1–6., September 2007.
- [2] Y. Yun, R. Govindan, D. Estrin, “ Geographical and energy aware routing: a recursive data dissemination protocol for wireless sensor networks,” Technical Report, UCLA/CSD-TR-01-0023, 2001.
- [3] C.Y. Chong and S.P. Kumar. August, 2003. Sensor networks: Evolution, opportunities, and challenges. *IEEE Proceedings:* 1247–1254.
- [4] A. Goldsmith and S. Wicker. August, 2002. Design challenges for energy-constrained ad hoc wireless networks. *IEEE Wireless Communications Magazine* 9: 8–27.
- [5] M. Haenggi, “On Routing in Random Rayleigh Fading Networks,” *IEEE Transactions on Wireless Communications*, vol. 4, pp. 1553–1562., 2005.
- [6] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson. September, 2002. Wireless sensor networks for habitat monitoring. *First ACM Workshop on Wireless Sensor Networks and Applications*, Georgia: Atlanta.
- [7] D. Puccinelli and M. Haenggi. August, 2005. Wireless Sensor Networks-Applications and Challenges of Ubiquitous Sensing. *IEEE Circuits and Systems Magazine* 5: 19-29.
- [8] Wendi Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan. January, 2000. Energy-Efficient Communication Protocols for Wireless Microsensor Networks. *Proc. Hawaaiian Int'l Conf. on Systems Science*.
- [9] W. Heinzelman, A. Sinha, A. Wang, A. Chandrakasan. June, 2000. Energy-scalable algorithms and protocols for wireless microsensor networks. *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00)*.
- [10] Wendi Heinzelman, Anantha Chandrakassan and Hari Balakrishnan. 2002. An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communitacions* 1 (4).
- [11] Huseyin Ozgur Tan, Ibrahim Korpeoglu. December, 2003. Power Efficient Data Gathering and Aggregation in Wireless Sensor Networks. *ACM SIGMOD Record* 32 (4): 66-71.
- [12] V.O. Li and J.A. Silvester. October, 1984. Performance Analysis of Networks with Unreliable Components. *IEEE Transactions on Communications, COM-32* 10: 1105-1110.

Reliable Routing in Wireless Sensorial Network by Combinational Optimization

András Bojársky

(Supervisor: Dr. János Levendovszky)

bojan@digitus.itk.ppke.hu

Abstract— In this paper some novel unicast protocols are developed for wireless sensor networks (WSNs) in order to ensure reliable packet transmission and maximize the lifespan at the same time. The optimal transmission energies and the optimal nodes to cooperate in the packet transfers are derived which guarantee that the packets are received by the Base Station (BS) with a given probability, subject to achieving the longest possible lifespan. These protocols can be applied in biomedical applications where energy consumption and longevity are of crucial importance. The optimization has been carried out for the LEACH protocol (when nodes are forwarding the packets toward the BS via cluster head node). The reliability of information transfer is enhanced by repeated low-energy packet transmissions. The new results have been tested by extensive simulations which also demonstrated that the lifespan of WSN can significantly be increased by the new unicast protocols.

Keywords— Communication systems, Communication system routing, Network reliability, Protocols, Cooperative protocols

I. INTRODUCTION

In most of the applications the WSN must convey the collected data to a single Base Station. In this set-up we assume that each node generates packets randomly and these packets are then transferred to the BS by multihop transmissions as indicated by Figure 1.

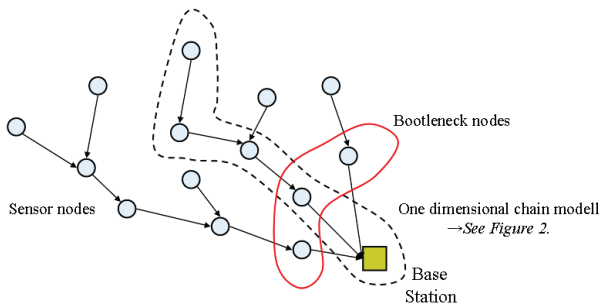


Figure 1. Spanning tree for the unicast flat routing

In this case the nodes being closer to the BS can get quickly overloaded as they have to retransmit almost all the packets sent to the BS. This results in very short longevity. In order to circumvent this, we propose novel unicast energy aware protocols which maximize the lifespan of the network. These novel methods assume that the distances between the nodes are known. Based on this information minimal energy consumption is achieved together with a guaranteed level of reliability (the probability of each packet reaching the BS properly is beyond a given threshold).

There are four energy balancing methods proposed by Haenggi [6,10]: (i) distance variation, (ii) balanced data

compression, (iii) shortcut routing, and (iv) equalization end-to-end reliability. The latter satisfies reliability constraints by using uniform transmission energies and no packet transmission repetition. In our proposed methods reliability constraints are met by using lower transmission energies.

The paper addresses reliable packet transmission in WSN when packets are to be received on the Base Station (BS) with a given reliability in terms of keeping the error probability under a given threshold [1]. Since the success of every individual packet transmission depends on the distance and the energy of transmission, the probability of correct reception will diminish exponentially with respect to the number hops, in the case of multi-hop packet transfers [2]. This effect can be compensated by setting the maximum number of hops, in which a packet should arrive to the Base Station.

We investigate this scheme in the case of LEACH protocol:

- *LEACH protocol* when the packet travels in the chain directly down to a certain node l which then sends the packet directly to the BS (shortcut). In this protocol the l node is the nearest cluster head node. Cluster head nodes are chosen randomly from the nodes in the network. This means, that in case of LEACH protocol a packet travels through only one node, which shortcuts the packet directly to the BS.

Our concern is to derive the appropriate transmission energies to achieve a given reliability and to minimize the overall energy consumption at the same time [6]. This problem leads to a constrained optimization which finally yields one optimal vector characterizing the packet transfer from a source node i to the BS as follows:

the optimal transmission energy vector $\mathbf{g}_{opt}^{(i)} = (g_{1,opt}^{(i)}, 0, \dots, g_{l,opt}^{(i)}, 0, \dots, 0)$ where component $g_{j,opt}$ describes the energy needed to transmit the packet from node j to node $j-1$;

In the case of the shortcut protocol (a packet originating from node i handed down to neighbours in the chain till node l from where it is directly sent to the BS) the optimal energy vector is described as $\mathbf{g}_{opt}^{(i)} = (0, \dots, 0, g_{l,opt}^{(i)}, 0, \dots, 0, g_{l,opt}^{(i)}, 0, \dots, 0)$, where $g_{l,opt}^{(i)}$ denotes the shortcut energy to BS from node l . In the latter case the optimization also involves to find the nearest cluster head l .

In this paper we will optimize the variables described above and compare the achieved lifespan to the longevities of repetition-free protocols.

II. THE MODEL

After the routing protocol has found the path to the base station, the subsequent nodes participating in the packet transfer can be regarded as a one dimensional chain labeled by $i = 1, \dots, N$ and depicted by the Fig. 1.

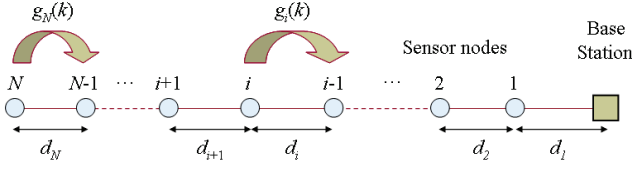


Figure 2. One dimensional chain model for WSN

The system is characterized as follows:

The topology is uniquely defined by a distance vector $\mathbf{d} = (d_1, \dots, d_N)$, where $d_i, i = 1, \dots, N$ denotes the distance between node i and $i-1$, respectively, whereas the energy needed to transmit packet over distance d is given as $g = \frac{d^\alpha \Theta \sigma_Z^2}{-\ln P} + g_0$ dictated by the Rayleigh model [11] (where d is the distance, α depends on the propagation type, P is the reliability of correct reception, Θ is the threshold, σ_Z^2 is the noise energy, while g_0 represents the consumption of the electronics during transmitting and receiving). For simplicity, this expression connecting the reliability parameter P of the transmission with the transmission energy g is denoted by $P = \Psi(g)$. The initial battery power on each node is the same and denoted by C and the energy state of the nodes at the k th time instant is expressed by vector $\mathbf{c}(k) = (c_1(k), \dots, c_N(k))$;

As a result, a WSN is fully characterized by vectors \mathbf{g} , \mathbf{p} , and \mathbf{c} respectively.

III. RELIABLE PACKET TRANSFER USING THE RANKING OPTIMIZATION ALGORITHM

In the case of the chain protocol, the end-to-end probability of correct packet reception at the BS is

$$P_{EE} = \prod_{j=1}^i P_j, \text{ and the associated energies are } g_j = \frac{d_j^\alpha \Theta \sigma_Z^2}{-\ln P_j} + g_0, j = 1, \dots, i. \text{ As a result, a given } P_{EE} \text{ can}$$

be achieved by several choices of P_j -s (i.e. several factorizations) yielding different energy consumptions.

A. Reliable packet transfer by using a game theoretic approach

As we described above, from the Rayleigh model one can see, that increasing the number of hops will exponentially increase the energy to deliver the packet to the Base Station with given reliability. The strength of the LEACH protocol is that it uses only one hop, to deliver the packet, of course from the parameters of the network and the nodes ability of transferring this can be less effective, but in most cases it works well.

As the node in the network serves their duty, each one of them is using their battery. Our goal is to optimize their

battery usage, so none of them is going flat to early. If one of the nodes goes flat, we assume that the communication in the network is not reliable any more. to evade this we have to create an algorithm, which invites the nodes, with relative high battery level to cooperate in the communication, to spare those which are at low battery.

We have to calculate the position of the nodes as well. The density of the nodes in a network isn't always equal. And the ones on the border naturally don't cooperate so much with the others as the ones in the center of the network. We have to create the optimization algorithm to involve the all nodes equally in the packet forwarding to reach maximum life span.

We created two parameters to classify the nodes in the network:

- One comes from the position of the node [7]. We can gather data for this classification from the topology of the network, or we can run an initial phase, with the help of Dijkstra algorithm, to find out which nodes are positioned on the best routes. From this we can create a parameter for each node described by the vector $\mathbf{T} = (t_1, t_2, \dots, t_N)$. We chose the parameters value from 1 to N , the node which has the best position will have 1, and the node with the worst will have N .
- The other parameter comes from the battery status of the nodes. Like above each node will have its own value according to the others battery status. We described this parameter with the vector $\mathbf{B} = (b_1, b_2, \dots, b_N)$. We chose the parameters value from 1 to N , the node which has the best position will have 1, and the node with the worst will have N . This parameter will change according to the communication in the network, so the parameter of the nodes, has to be updated frequently.

By combining the two parameters we can create a gradation of the nodes in the network. The nodes with higher rank will have more battery, and better position at the same time, or much more energy and a bit worse position, or much more position and a bit less energy than the others with lower rank. Our goal is to force the nodes with higher rank to participate in the communication, and to spare those with lower rank. We described this parameter with the vector $\mathbf{R} = (r_1, r, \dots, r_N)$. We can chose to weight to two vectors before combining, at the numerical result we shall show how this will affect the life span of the network.

B. The ranking optimization algorithm

The optimization above requires the following steps described in Table 1.

TABLE I. THE STEPS OF THE RO ALGORITHM

Steps	Action
step 0.	we set the batteries to maximum, and vector \mathbf{T} and \mathbf{B} and \mathbf{R} to zero.
step 1.	run an initial phase on the network. Every node has to send a packet to the base station, using to Dijkstra algorithm to specify the route. Each node counts how many packet it transferred in this phase. The node with the higher count gets a higher rank. We set vector \mathbf{T} .

Steps	Action
step 2.	we calculate the battery of a node. If a node has higher battery status than the others, gets a higher rank. We set vector B .
step 3.	we combine vector B and T , to evaluate vector R .
step 4.	when a node wants to send a packet to the Base Station, it searches in a q radius to see, which node i has the highest rank nearby, and forwards the packet, if none sends directly to the BS
step 5.	the node i who receives the packet forwards to the Base Station
step 6.	the Base Station collects the battery status of the nodes, who participated in the transmission, and updates vector B and R .

The optimization need three parameters, which can lead to further results, parameter q is the area in which a node search for a higher rank node to forward the packet to. If the result of the search is that none of the nodes in the q area has higher rank, the node sends the packet directly to the BS. If it finds more than one node with higher rank, it chooses the one with the highest rank to cooperate in the transmission. It can happen that a node sends a packet to a higher distance from the BS than it already are, but this serves the goal to squeeze all the nodes before one of them is going flat.

The other two parameters c and u are the weights for vector T and B . We can use these vectors to set the rank vector more accurately.

We set the maximum hop number no higher than two, because of the reliability constrain $P_{EE} = \prod_{j=1}^i P_j$. If we raise the maximum hop number the lifespan of the network will decrease.

The optimization is carried out for all possible source nodes $i=1, \dots, N$. The algorithm is represented by the following block diagram:

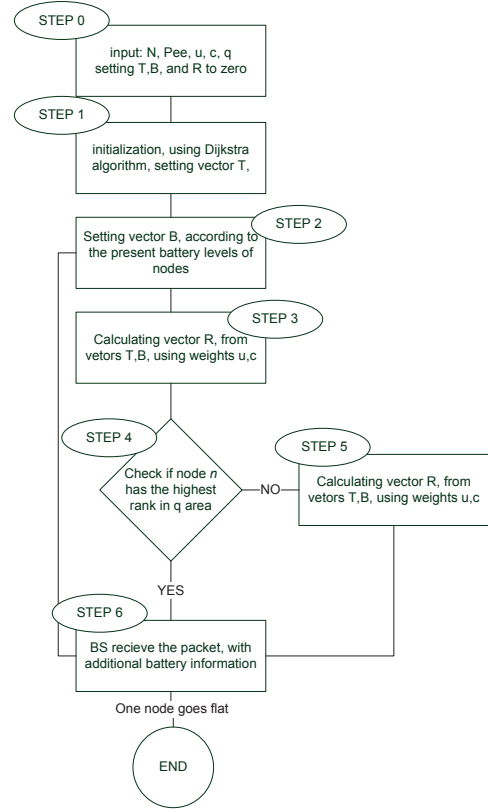


Figure 3. The flow diagram of the ranking optimization algorithm

IV. NUMERICAL RESULTS

In this section the performance of the protocols described above are investigated by extensive simulations.

The lifespan was defined as the number of steps until which each node has the energy to transmit packets complying with the given reliability parameter. As soon as, a node (the bottleneck node) goes flat (being not able to participate in the reliable packet transfer, because of falling short of the required energy), then network is considered dead. Fig. 4 depicts the lifespan achieved by the LEACH protocol compared with the ranking optimization algorithm.

The investigated WSN contained 30, 50, 80 nodes the locations of which were subject to Gaussian and the required reliability P_{EE} was set 0.7, and 0.9. The distances were set to 50m in all directions from the BS. The q , c , and u parameters were set to 1. The α parameter of Rayleigh model was set to 4.

The figure exhibits the number of sent packets.

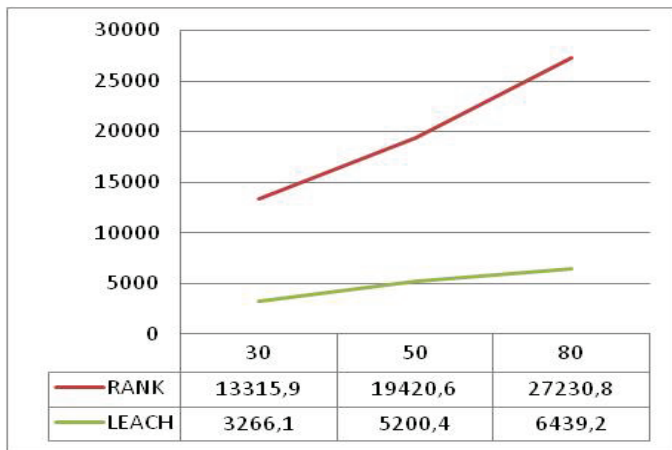


Figure 4. Number of sent packets $P_{EE}=0.7$

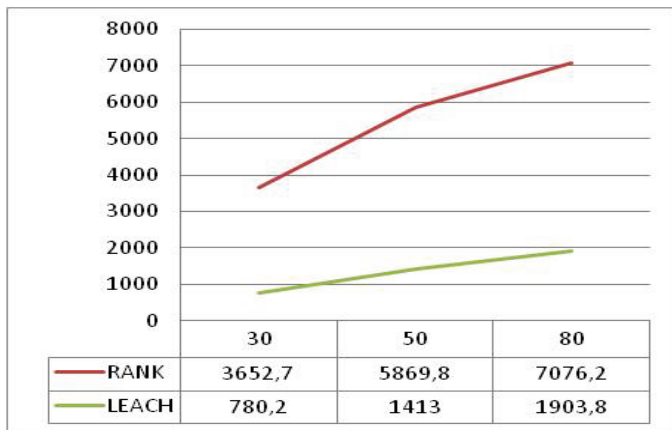


Figure 5. Number of sent packets $P_{EE}=0.9$

One can see that the protocol with the ranking vector achieves better longevity, than the LEACH protocol. The gain is approximately 400%. If we set the α parameter of Rayleigh model to 3 or 2, the gain decreases to approximately 300%.

The next figure shows the gain in the examined cases.

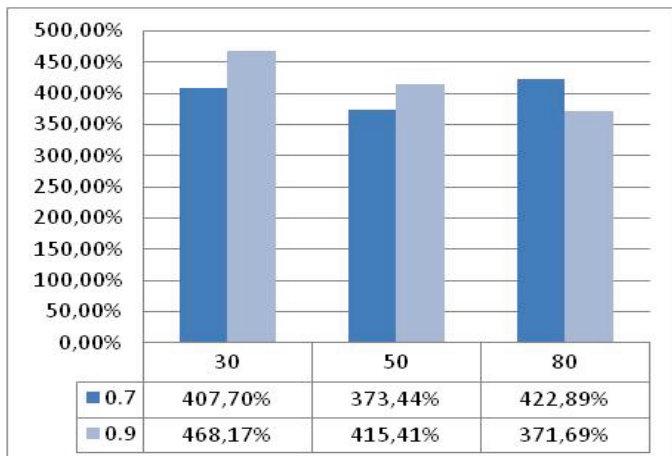


Figure 6. The gain in percentage by using the ranking algorithm instead of LEACH

V. CONCLUSION

In this paper, novel routing protocols have been developed using ranking between the nodes in the network.

We optimized the routing of the network, so the nodes with higher battery status, or better position cooperated with the others more frequently. One of the real strengths of the algorithm is the balancing between the position, and the battery status. If the nodes are in the same battery level, their position will decide which one to forward to, if one of them has significantly more battery than the others, it will be the one to cooperate with, regardless its position.

One possible extension of the method is to create a ranking without the BS. A node can calculate its own parameter, and broadcast to anyone who has higher one. With this the packet would multiply in the network, which is expensive, but in this case we could decrease the sending energy, of course with fulfilling the reliability constraint.

It has been demonstrated that the novel optimization can significantly increase the lifespan of WSN and enhance its ability to carry data to BS. This enhancement in lifespan is crucial in biomedical applications [11,12].

REFERENCES

- [1] J. Levendovszky, A. Olah, A. Bojarszky, B. Karlocai, "Energy balancing by combinatorial optimization for wireless sensor networks," IWDN07 Conference, pp. 1–6., September 2007.
- [2] A. Goldsmith and S. Wicker. August, "Design challenges for energy-constrained ad hoc wireless networks," IEEE Wireless Communications Magazine, Vol. 9, pp. 8–27., 2002.
- [3] C. Intanagonwiwat, R. Govindan, D. Estrin, "Directed Diffusion: a scalable and robust communication paradigm for sensor networks," Proc. 6th Annual Int. Conf. Mobile Computing and Networking, Boston, August 2000.
- [4] D. Bragonsky, D. Estrin, "Rumour routing algorithm for sensor networks," First ACM Int. Workshop on WSN and Applications, pp. 22–31., Atlanta, September 2002.
- [5] J. Kulik, W. Heinzelman, H. Balakrishnan, "Negotiation-based protocols for disseminating information in wireless sensor networks," ACM Wireless Networks Journal, Vol. 8, pp. 169–185., 2002.
- [6] M. Haenggi, "Analysis and Design of Diversity Schemes for Ad Hoc Wireless Networks." IEEE journal on selected areas in communications, vol. 23, no. 1, 2005.
- [7] Y. Yun, R. Govindan, D. Estrin, "Geographical and energy aware routing: a recursive data dissemination protocol for wireless sensor networks," Technical Report, UCLA/CSD-TR-01-0023, 2001.
- [8] W. Heinzelman, A. Chandrakasan, H. Balakrishnan. "Energy-Efficient Communication Protocols for Wireless Microsensor Networks," Proc. 33rd Hawaii Int. Conf. on Systems Sciences, pp. 223–234, Hawaii, January 2000.
- [9] H. Ozgur, Tan, I. Korpeoglu, "Power Efficient Data Gathering and Aggregation in Wireless Sensor Networks," ACM SIGMOD Record, Vol. 32, Issue 4, pp- 66–71., 2003.
- [10] M. Haenggi, "On Routing in Random Rayleigh Fading Networks," IEEE Transactions on Wireless Communications, vol. 4, pp. 1553–1562., 2005.
- [11] C.Y. Chong and S.P. Kumar, "Sensor networks: evolution, opportunities and challenges," IEEE Proceedings, Vol. 91, No. 8, pp. 1247–1254., 2003.
- [12] D. Puccinelli, M. Haenggi, "Wireless sensor networks – applications and challenges of ubiquitous sensing," IEEE Circuits and Systems Magazine, Vol. 5, pp. 19–31., 2005

Mobile Platform for Testing Communication Protocols, Challenges in the Implementation of SC Receiver

Péter Vizi

(Supervisor: Dr. János Levendovszky)

peter.vizi@itk.ppke.hu

Abstract—In this paper the design of a mobile platform is introduced, which is aimed for testing communication protocols in a more realistic scenario. The idea behind this is that including mobile nodes in a network can be used to create events that are natural and controlled - meaning that it can be reproduced -. Since these cases can be reproduced different communication protocols can be tested and compared for robustness. The second part of the paper is concerned with the challenges in implementing a Superposition Coding receiver on the GNU Radio platform.

I. INTRODUCTION

Advantages of mobility in a Wireless Sensorial Network (WSN) has been analyzed from different perspectives. To prolong the lifetime and increase robustness propositions have been made including random mobility, predictable mobility, and controlled mobility.

Our work puts mobility into a different perspective. Instead of trying to increase the performance of the network we use mobile nodes, similar to [1], to create an environment where the performance of communication protocols can be evaluated in a realistic setting. The idea is that we can control the position of the mobile nodes by observing Received Signal Strength Indicator (RSSI) values of incoming packages. If we want to simulate the outage of a node, we can move it into a position where there is no reception. In case we want to observe the effects of fading, we can move the node accordingly. The advantage of this platform is that we can reproduce these experiments, and put different protocols into the same environment, this way we can get a more realistic comparison.

The second part of the paper introduces design challenges in the implementation of a Superposition Coding (SC) receiver. For this implementation we used the GNU Radio Software Defined Radio (SDR) platform [2]. This platform is flexible, suitable for rapid development, and this is exactly what we need when experimenting with new communication techniques. SC is theoretically well studied, however there is lack of implementation in hardware. During the development a lot of interesting design choices had to be made, regarding for example the synchronization or the hardware in-perfections of the Universal Software Radio Peripheral (USRP) board.

The rest of the paper is organized as follows. In Section II we elaborate the design of the previous mobile platform. Section III includes more details about the implementation of

the SC receiver. We conclude the work in Section IV, finally plans for the future work are presented in Section V.

II. MOBILE PLATFORM FOR TESTING COMMUNICATION PROTOCOLS

This section elaborates on how mobility can be used for evaluating communication techniques.

A. Introduction

In a dynamic network where all the nodes are moving collecting measurement data is a hard challenge. In this case we consider a set of randomly moving entities, such as a pack of zebras [3], where even the Base Station (BS) is mobile. Mobility can aid the network in preserving energy [4] by having a small number of mobile nodes, which are performing a random walk, collect and deliver data. This way only short range communication is necessary, since the nodes do not talk directly to the BS, only to the mobile nodes.

One can exploit the predictability of the movement [5], which will increase the lifetime by putting nodes into sleep mode, and also give a bound on the transmission delay. If the trajectory of the data collecting node is known a more optimal routing scheme can be computed, as in [6].

In [7] researchers showed that by having controllable mobile nodes, the energy consumption of the network can be reduced: moving data physically can be advantageous over transmitting it over large distances.

Our mobile platform is built up from two off the shelf components: the widely used Mica motes for data acquisition and radio communication and the Lego NXT type of robot, see Figure 1.

The first task was to create the hardware and software environment, then we have performed the first measurements to support that we can control the node to our needs.

B. Hardware and Software Environment

The Mica mote is built up from a microcontroller, some flash memory and a tunable radio. In addition to this different sensor boards can be attached to the mote for data acquisition. Our Lego NXT robot contains the NXT brick as the controlling element and two motors. The communication between the two microcontrollers is over the I²C bus. This enables us to transfer



Fig. 1. Mica2 mote mounted on the Lego NXT robot.

data reliably between the two devices. The two equipment is connected through the Mica mote's 51-Pin extension and the NXT's 6-position modular connector for sensors.

TinyOS operating system is run on the Mica motes, which is a standard operating system for WSNs. On the Lego NXT we use the leJOS firmware, that includes a Java virtual machine. The challenge here was that the type of ARM microcontroller used in the NXT brick is not capable for I²C slave operation, and for the TinyOS there is only implementation for I²C master mode. Because of this the first task was to write the required software components for TinyOS that enables the Mote to operate in slave mode.

During operation the NXT brick polls the Mica mote, and when from the application layer there is a new command it is carried out, and the result is reported back to the Mote. In the current implementation we use the TachoNavigator from the leJOS library, which uses the feedback from the NXT's precise motors. With TinyOS commands we can direct the robot to turn with a given angle, travel a certain distance, or by using the navigator's internal reference go to a given coordinate.

C. Test Measurements

In the first measurement we simulated the effect of Rayleigh fading, by moving the robot on a line path and putting the receiver behind a metal cabinet, this way creating a non-line of sight communication. As Figure 2 shows the measured data fits well with the theoretical Rayleigh curve.

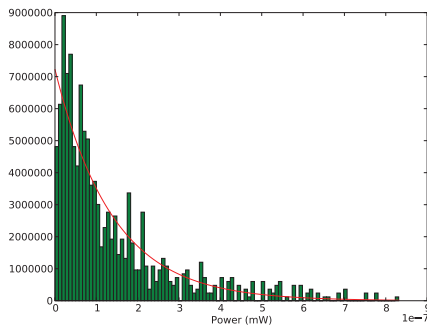


Fig. 2. The distribution of RSSI with green bars, compared to the Rayleigh distribution with red line.

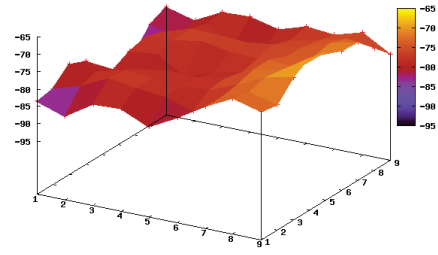


Fig. 3. RSSI measurements on a 1.5 m by 1.5 m square.

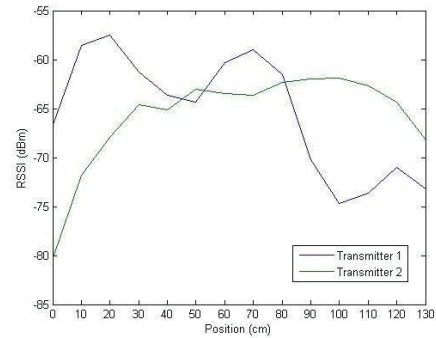


Fig. 4. The RSSI values in the case of two transmitters.

Using a transmitter and the Mica mote on the robot as the receiver we have also recorded RSSI values while the robot was moving on a grid. Figure 3 shows the results of one of these measurements. It can be seen that there is a decrease in the RSSI - the transmitter was position on the upper right corner of the square -, while there are also dips, introduced by fading. These dips can be exploited if we want to position the mobile node into a bad spot.

In the last set of experiments we put the robot between two transmitters, and it was relaying data between them while moving on a straight line. The recorded RSSI values are on Figure 4. In this case we can come up with different optimization criteria: we can position the robot where we minimize the maximum of the RSSI or we can minimize the sum of RSSI values. This will simulate the asymmetry of the wireless link.

III. SUPERPOSITION CODING RECEIVER IMPLEMENTATION

This section introduces the design and challenges in implementing an SC receiver on the GNU Radio platform.

A. Motivation

Multiuser techniques such as SC [8] in theory improves the throughput of wireless networks. Although these theories prove significant performance gain sometimes include assumptions for analytical tractability. These assumptions - perfect synchronization, immediate feedback - might not be available in practice, so experimental studies are required.

SC is particularly interesting because it achieves capacity in degraded Gaussian channel, its architecture is scalable, can be used for sophisticated Medium Access Control (MAC) protocols.

For testing new communication techniques we need a flexible platform, that support short development cycle. The GNU Radio architecture suits our needs, because it provides signal processing blocks implemented in software, this way it is easily extensible. The USRP board (see Figure 5) serves as an RF front-end, and with the help of replaceable daughter boards it gives us the opportunity to experiment on different frequency bands.

The Physical (PHY) layer of our design uses Orthogonal Frequency Division Multiplexing (OFDM) as its transmission scheme. This system transmits data over multiple orthogonal subcarriers [9]. With this scheme it is relatively easy to implement channel estimation, and the bandwidth scalability is also a positive factor.

In the following we describe the challenges encountered during the implementation of an SC based receiver.

B. Superposition Coding

In this section a brief summary of the SC theory is given. Consider two users sharing an Additive White Gaussian Noise (AWGN) wireless channel, their Signal to Noise Ratio (SNR) are $SNR_1 \gg SNR_2$ respectively. If we denote the the modulated symbol stream with destination 1 as $\{x_1\}$, destination 2 as $\{x_2\}$, the simultaneously transmitted symbol at time k is

$$s_k = \sqrt{\alpha_1}x_{1,k} + \sqrt{\alpha_2}x_{2,k}, \quad (1)$$

where α_i is the fraction of transmit power allocated to user $i = 1, 2$.

The k^{th} is

$$r_{i,k} = s_k + w_{i,k}, \quad (2)$$

where $w_{i,k}$ is the unit variance circularly symmetric complex Gaussian variables with variance σ^2 . With the appropriate adjustment of parameters α_i both users can detect their packets.

C. Receiver Design

The GNU Radio architecture is built up from a set of signal processing blocks, which can be connected into a flow graph. During execution the scheduler's task is to provide enough data for each of the blocks to operate on.

In the case of the receiver the source of data is the USRP board, and after a sequence of operations the decoded data

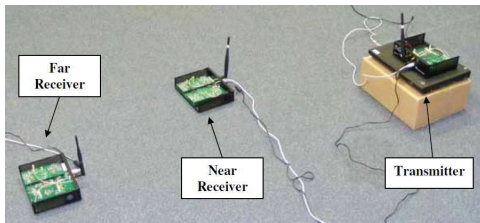


Fig. 5. Three USRPs in the experimental setup.

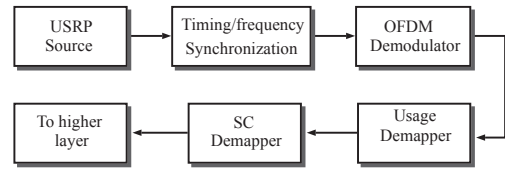


Fig. 6. Receiver flow graph.

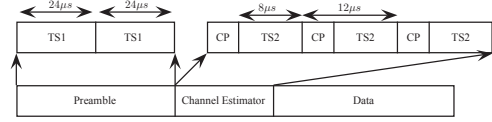


Fig. 7. The packet structure.

can be consumed by the next layer of the protocol stack, see Figure 6.

The first operation on the incoming stream is the Timing and frequency recovery, with the Schmid-Cox algorithm [10]. For this algorithm we use 48 OFDM tones as preamble (see Figure 7), this is enough for a coarse estimate, however for long packets a fine frequency tracking algorithm also had to be introduced.

After the start of the packet has been found the channel estimator algorithm can equalize the received stream, which is calculated base on a given channel estimator sequence.

In the following the OFDM demodulator can demodulate the frames. From this stream the tones which are not used for storing data - for example the ones that are reserved for the fine frequency tracking - have to be removed by the Usage Demapper. After this the SC demapping operation can be carried out, which will decode both the first and the second user's data. In the end of the flow graph the received data can be directed to the higher layer, for example using sockets.

D. Bit Error Rate Measurements

To characterize the performance of our implementation we first carried out Bit Error Rate (BER) measurements. With different received SNR we record the incoming data, and calculate the BER. Refer to Table I for the set of parameters in this experiment.

The OFDM packet structure used is depicted in Figure 7. The preamble sequence is used primarily for frequency and timing synchronization, and is designed by repeating a pseudo-random training sequence of length 24 symbols (TS1) twice. The channel estimation symbols are used for performing equalization and is generated by repeating a pseudo-random sequence of length 8 symbols thrice. The length of the cyclic prefix was chosen to be 4 symbols ($4\mu s$). In the paper, we report results for an uncoded system.

Bandwidth	1 MHz
Fraction of power allocated to the near user: α_1	0.8
Carrier frequency	930 MHz
No of tones per symbol	8

TABLE I
SYSTEM PARAMETERS FOR BER MEASUREMENTS.

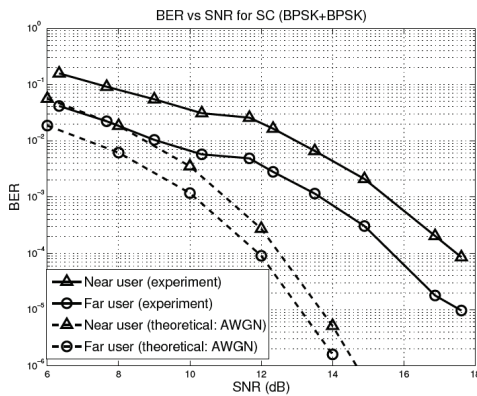


Fig. 8. BER versus SNR

Figure 8 plots the BER versus SNR performance of the SC system for both the near and far users. The dashed lines indicate the BER values for an ideal system (with perfect timing and frequency synchronization and without fading) in an AWGN channel, and were obtained via simulation. The solid lines represent the experimental results. The experiment was conducted indoors in our laboratory. In order to evaluate the BER, the received bits were fed back to the transmitter via Ethernet. Each point on the BER curve was obtained by averaging over 60000 packets.

E. Design Issues

During the implementation we had to solve some issues that are mainly caused by the restrictions of the USRP board.

The effective bandwidth of the USRP is much smaller than that set by the user. It is mostly because of the interpolation implementation, and this forced us to ignore some of the carriers, which are too much distorted.

The frequency offset between the transmitter and the receiver also had to be compensated. For this purpose we could use the channel estimator sequence.

IV. COLLUSION

In this paper we have introduced a novel application of mobility in WSNs. Our mobile platform can be used to more realistically compare communication techniques. This way we can benchmark these methods in an environment which is hard to achieve with simulation.

During the implementation of a SC receiver we have explored the limitations of the USRP board, and proved that it is realizable with an SDR platform.

V. FUTURE WORK

There are still interesting questions about mobility. Cooperation between mobile nodes can be studied with this platform. Positioning the mobile node where it can perform optimal relaying is also a challenging task.

For our SC transmitter the next step is to introduce coding into the system. This will increase the reliability of the communication significantly. After this is implemented experiments with novel MAC protocols on top of this architecture can be started.

ACRONYMS

AWGN Additive White Gaussian Noise
BER Bit Error Rate
BS Base Station
MAC Medium Access Control
OFDM Orthogonal Frequency Division Multiplexing
PHY Physical
RSSI Received Signal Strength Indicator
SC Superposition Coding
SDR Software Defined Radio
SNR Signal to Noise Ratio
USRP Universal Software Radio Peripheral
WSN Wireless Sensorial Network

REFERENCES

- [1] G. Sibley, M. Rahimi, and G. Sukhatme, "Robomote: a tiny mobile robot platform for large-scale ad-hoc sensor networks," in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, vol. 2, 2002, pp. 1143–1148.
- [2] I. Mitola, J., "Software radios: Survey, critical evaluation and future directions," *Aerospace and Electronic Systems Magazine, IEEE*, vol. 8, no. 4, pp. 25–36, Apr 1993.
- [3] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein, "Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with zebnet," *SIGPLAN Not.*, vol. 37, no. 10, pp. 96–107, 2002.
- [4] "Data mules: modeling and analysis of a three-tier architecture for sparse sensor networks," *Ad Hoc Networks*, vol. 1, no. 2-3, pp. 215 – 233, 2003, sensor Network Protocols and Applications.
- [5] A. Chakrabarti, A. Sabharwal, and B. Aazhang, "Using predictable observer mobility for power efficient design of sensor networks," in *The second International Workshop on Information Processing in Sensor Networks (IPSN, 2003)*, pp. 129–145.
- [6] E. Lee, S. Park, F. Yu, Y. Choi, M.-S. Jin, and S.-H. Kim, "A predictable mobility-based data dissemination protocol for wireless sensor networks," in *Advanced Information Networking and Applications, 2008. AINA 2008. 22nd International Conference on*, March 2008, pp. 741–747.
- [7] A. Somasundara, A. Kansal, D. Jea, D. Estrin, and M. Srivastava, "Controllably mobile infrastructure for low energy embedded networks," *Mobile Computing, IEEE Transactions on*, vol. 5, no. 8, pp. 958–973, Aug. 2006.
- [8] T. M. Cover and J. A. Thomas, *Elements of information theory*. New York, NY, USA: Wiley-Interscience, 1991. [Online]. Available: <http://portal.acm.org/citation.cfm?id=129837>
- [9] T. Hwang, C. Yang, G. Wu, S. Li, and G. Ye Li, "Ofdm and its wireless applications: A survey," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 4, pp. 1673–1694, May 2009.
- [10] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for ofdm," *IEEE Trans. Commun.*, no. 45, pp. 1613–1621, 1997.