Research Programmes for Telecommunication Electronics

Research Reports
Contents

**Telectronics**

Integration of Microcircuits with Multilayer Substrates ................................................................. 5
Integrated Waveguide Bragg Gratings ............................................................................................... 9
Adiabatic Logic and its Integration with CMOS Technology .............................................................. 13
Blind Source Separation and Deconvolution in Telecommunications and Array Signal Processing ........ 19
Personal Access and User Interface for Multimodal Broadband Telecommunication ...................... 25
Techniques for the Third Generation of Wireless Systems ............................................................... 41
Fractal Processes in Telecommunications .......................................................................................... 45
Distributed Media Processing in Hybrid Networks ........................................................................... 47
Computational Methods for the Performance Analysis of Broadband Communication Networks ....... 51

**Telectronics II**

Future Internet - Traffic Handling and Performance Analysis .......................................................... 59
Blind Signal Separation in Communications Receivers and Antenna Array Systems ..................... 67
Concurrent Design and Fabrication of Integrated Module Board ..................................................... 79
Advanced Radio Channel Identification and Equalization ............................................................... 87
Development of the Second Generation Delta-sigma Frequency Synthesis Techniques .................. 93
On-Chip Communication Architecture for HW/SW Co-design ....................................................... 101
Integration of Microcircuits with Multilayer Substrates Using Advanced Thin-Film Processing

Jorma Kivilahti
Professor
Lab. of Electronics Production Technology
Helsinki University of Technology
P.O. Box 3000, FIN-02015 HUT, Finland
Jorma.Kivilahti@hut.fi

ABSTRACT

The project had two major objectives. The first one was to design and fabricate an effective diffusion barrier layer between copper and silicon on the basis of the theoretical and experimental study of Cu|M|Si interactions. Having the expertise needed for manufacturing functioning Cu-metallised IC's it was possible to realise the second objective that was to interconnect bare Cu-chips with the IMB multilayer build-up technology developed earlier in the laboratory. Both objectives were attained successfully in the project.

I. INTRODUCTION

Increasing capabilities of integrated circuits cannot be converted fully into component or system performance, because the interconnection and packaging technologies limit not only the technical performance but also contribute too much to cost (Fig. 1). Thus, a great challenge is related to advances in thin film processing, underlying electronic materials and, in particular, to fabricating reliably very small interconnections. Likewise, the integration of microcircuits and passive components into high density multilayer substrates should be realised cost-effectively.

These challenges can be met effectively by making use of IC's with Cu metallisation. But the utilisation of the Cu-metallised microcircuits provides essentially that the interactions between Si and Cu are understood thoroughly. Therefore, the objectives of the project were, firstly, to study theoretically and experimentally Cu/M/Si interactions and, secondly, based on the knowledge gained by the first objective to interconnect bare chips with the build-up technique. Cu and Si react strongly even at low temperatures and so a diffusion barrier layer is required between them. Tantalum and tantalum-based materials have been chosen as the diffusion barrier materials in the present study.

II. THEORETICAL CONSIDERATIONS

In the theoretical considerations the combined thermodynamic-kinetic analysis was employed. Phase diagrams can be used to have information about the phases that can exist in local equilibrium with each other at different temperatures. Furthermore, if the phase diagrams are coupled with the available kinetic data, it is possible to predict phase formation sequences in diffusion couples. However, this type of predictions must be executed with precautions for the thin film systems, since the use of phase diagrams for solve the phase formation sequence requires an assumption that the local equilibrium is attained at the interfaces, which is not always the case in the thin film systems. The assumption of local equilibrium demands that reactions at the interfaces are fast enough so that the atoms arriving in the reaction region are used immediately and the rate determining step is the diffusion. However, with very thin layers this requirement may not be fulfilled. The reasons for this originate mainly from the special conditions prevailing during thin film reactions: (a) relatively low reaction temperatures, (b) small dimensions, (c) high density of short-circuit diffusion paths, (d) relatively large stresses incorporated in thin
films, (e) relatively high concentration of impurities, (f) metastable structures, (g) large gradients, etc. However, despite the difficulties mentioned, it is expected that by employing this approach better understanding of the equilibria and reactions in the system is achieved.

In this study several important ternary phase diagrams were assessed. Among them are especially Si-Ta-Cu, Si-Ta-C, Ta-C-Cu, Si-Ta-N, Ta-N-Cu and Ta-C-O systems. The activity diagrams as well as several binary diagrams were evaluated. The information extracted from these diagrams together with the experimental results were used to discover the underlying mechanism(s) for the failure.

III. EXPERIMENTS

Tantalum and tantalum-based materials have been chosen as the diffusion barrier materials in this study. Ta, TaC, Ta2N and TaN diffusion barriers with thicknesses of 10, 50 and 100 nm were used in the investigations. The films were sputter-deposited either with inert carrier gas (Ta and TaC) or with reactive sputtering (Ta2N and TaN). The copper overlayer thickness was either 100 or 400 nm. The stacked sheet films on (100) Si substrates have been annealed at different temperatures between 400 and 800 °C for 30 min under the vacuum of 10^-4 Pa. The sheet resistance measurements at room temperature using the four-point probe have been used to detect interfacial reactions after each annealing step. The reaction products in the different metallisation schemes have been characterised by using the grazing incidence x-ray diffractometry (XRD), the Rutherford backscattering spectroscopy (RBS), the secondary ion mass-spectroscopy (SIMS), and the transmission electron microscopy (TEM). Surfaces of the samples were also examined with an optical microscope and scanning electron microscope (SEM). The atomic force microscopy (AFM) was used to monitor the surface of the sputtered films.

IV. RESULTS AND IMPACTS

The objectives of the project were attained very well. The project gave new insight and significant fundamental results on CuM/Si interactions. Furthermore, the project produced valuable practical results, which have been utilised successfully in developing the Integrated Module Board (IMB) technology in the ETX projects supported by the National Technology Agency and the Finnish Electronics Industry. The co-operation between the partners was very useful. By combining the thermodynamic-kinetic approach together with the detailed microstructural analyses (HUT/EPT) and by employing the novel thin film fabrication facilities (VTT), much better understanding of the diffusion barrier problem was achieved. In addition to greater expertise in fabricating high density build-up structures several scientific papers have been published in highly esteemed research journals during the project. The theoretical approaches employed in the project proved to be very useful, and they provided excellent basis for further investigations in the field.

The first objective of the present work was to obtain a deeper understanding of the failure mechanisms, microstructures and stabilities of the Ta-based barrier layers. The combined thermodynamic-kinetic approach, which has not been previously used in the investigation of thin film diffusion barriers, was utilized for analysing and explaining the experimentally observed reaction sequences. It has been shown that this approach can provide unambiguous knowledge of the reactions taking place during the annealing in the metallization structures.

The results demonstrate that Ta-based barriers offer a very feasible solution to the diffusion barrier problem. The failure mechanisms of the different barrier layers (e.g. Ta, TaC and Ta2N) have many similarities. Especially, TaC and Ta2N behave almost identically. With the help of thermodynamic evaluation of the corresponding ternary phase diagrams, we pointed out that the reason behind this similarity was the almost identical phase relationships found in both metallization systems. In the case of elemental Ta diffusion barriers, we were able to demonstrate that the failure mechanism was thickness-dependent. Using this knowledge, it was possible to solve many contradictions with respect to the first phase formation during the annealing in the Si/Ta/Cu metallization system, as reported in the literature. Furthermore, the crucial effect of oxygen on the reactions in all the metallization schemes with the different Ta-based diffusion barriers was demonstrated and the thermodynamic basis for understanding the origins of this behaviour was given (Figs. 2 and 3).

The ternary Ta-Si-Cu, Ta-Si-N, Ta-Si-C, Ta-N-Cu and Ta-C-Cu phase diagrams have been calculated from the assessed binary thermodynamic data. The binary Ta-O and ternary Ta-C-O phase diagrams (Fig. 3) have also been calculated in order to model thermodynamically the influences of oxygen on the reactions in different metallisation schemes.

The Integrated Module Board (IMB) technique developed in the laboratory of Electronics Production Technology is used to integrate the Cu-metallised IC’s being fabricated by the VTT Microelectronics. In this PWB-based technology both bare microcircuits and passive components are integrated successfully in conjunction with the fabrication of high density organic HDI substrates in large panels. This solderless, non-vacuum and fully additive technology is based on a
photodefinable epoxies and fully additive electroless plating process.

Fig. 2. Bright field TEM micrograph from the Si/TaC(70nm)/Cu(400nm) sample annealed at 600 °C for 30 min.

Fig. 3. Isothermal section from the evaluated metastable ternary Ta-C-O phase diagram at 600 °C under the external oxygen pressure of about 0.2×10^4 Pa. The tie-lines in the TaC-Ta2O3 two-phase region are shown in the diagram. The contact-line (C.L.) between the TaC film and oxygen indicating the initial unstable equilibrium as well as the approximate composition of the TaC[O]gb are also shown.

Metals such as copper or nickel are chemically deposited onto photodefined wiring tracks and I/O pads of embedded active components. The IMB technology enables short conductor line lengths, small line pitches (< 50µm) and extremely high component density. In this manner functional high density modules with good electrical performance and better reliability than with solder-based surface mount technology (SMT) can be achieved. However, in order to fully utilise the capabilities of the IMB technology, one needs Cu-metallised IC’s to get rid of the complex Under Bump Metallurgy (UBM) structures as needed currently with Al metallisation. The Cu-metallised IC’s enable the fabrication of Cu/Cu contacts throughout the whole functional module (Figs. 4 and 5). This is very beneficial both from the electrical and reliability point of view, and it will simplify the fabrication process considerably. Thus, the work performed in this project has supported the development of the IMB technology, which is presently being implemented into production by the Imbera Electronics established in 2002 jointly by Aspocomp Group and Elcoteq Networks. Further development of the IMB technology continues also in the "Interfacial Compatibility Between Dissimilar Materials in Ultra-High Density Electronics" project financially supported by Academy of Finland.

Fig. 4. Integrated Module Board (IMB) fabricated with the fully additive build-up process (www.ept.hut.fi)

Fig. 5 (a) Plain view of the IC with Cu contact pads and Ta diffusion barrier and (b) cross-sectional view of the same structure showing the Cu/Cu interconnections.
The results of the project are currently being used also in another project within the Telelectronics II program, which continues and enables the development of concurrent design of very-high density integrated electronics like IMB modules and so expands the work carried out in this project. Without any doubt the research in this field will continue in the future.

V. PUBLICATIONS

A. PEER-REVIEWED PAPERS, CONFERENCE PAPERS AND REPORTS


B. ACADEMIC DEGREES


C. SCIENTIFIC REPORTS

Integrated waveguide Bragg gratings

Matti Leppihalme, Timo Aalto, Päivi Heimala, and Sanna Yliniemi

VTT Centre for Microelectronics
P.O. Box 1208, FIN-02044 VTT
Matti.Leppihalme@vtt.fi

Jari Turunen, Markku Kuittinen, Janne Simonen, Jani Tervo, and Tuomas Vallius

University of Joensuu, Department of Physics
P.O. Box 111, FIN-80101 Joensuu
Jari.Turunen@joensuu.fi

ABSTRACT

In this project integration of Bragg grating structures to silicon-on-insulator (SOI) waveguides was studied. An efficient quasi-rigorous computational method was developed for analysis of waveguide gratings and applied to SOI structures. Phase masks for Bragg grating exposure experiments on the proximity mask aligner were designed and fabricated. A new method was found to eliminate the harmful zeroth diffraction order of a phase-mask grating. Investigations on photonic bandgap structures by rigorous diffraction theory were initiated during the project.

State of the art fabrication methods for both nanostructures and waveguides were developed. The methods were demonstrated by fabricating and characterizing several test structures and components. The integration of SOI waveguides with gratings was accomplished during the project. Optical measurements, however, showed that the process still needs some fine tuning.

I. INTRODUCTION

Broadband Information Technology is a major business area and optics is the enabling technology of choice needed for its realization. All-optical networks where the signal is transmitted, routed and switched optically from end to end have been intensively studied worldwide leading to a tremendous development of optical telecommunication. At present the data transmittance rates are on the level of 40 Gb/s for a single laser-fiber combination and wavelength division multiplexing (WDM) systems can have hundreds of channels. These multi-wavelength optical networks offer an extremely large bandwidth and are transparent to signal format and type and offer many advantages for future telecommunication and computer networks.

The rapid development in WDM optical communications has generated strong needs for novel planar multiwavelength devices, which can treat optical signals. One of the key components is a grating, which can be tailored directly in the waveguide structures. These structures will enable for instance such functions in optical telecommunications, as add-drop filters needed to accomplish a WDM circuit, notch band filters, dispersion compensation applying chirped Bragg gratings and external cavities for DFB-lasers etc. The grating structures could also be used in microsystems and in various sensing applications. The use of a silicon substrate as an “optical bench” is furthermore a very appealing approach to reduce the costs of the multifunctional optical systems by integration.

In this project, our major goal was to develop and demonstrate Bragg grating structures integrated with SOI waveguides and to study their properties and applicability in optical telecommunication systems. As the integration of nanoscale gratings with micron-scale waveguides proved to be very challenging we concentrated on solving the technical problems and put less effort on the applications. Due to the process synergy some studies on photonic crystal structures were also made during the project.

This project was a joint research between VTT Centre for Microelectronics and University of Joensuu. University of Joensuu has been responsible for theoretical modeling and direct e-beam writing while VTT has been responsible for grating and waveguide processing and fabrication and optical characterization of accomplished structures.
II. RIGOROUS DIFFRACTION THEORY IN GRATING MODELLING

a) Bragg grating structures in SOI waveguides

The analysis method of waveguide gratings based on rigorous diffraction theory, developed earlier at the University of Joensuu, was applied to modeling of Bragg gratings in SOI waveguides [1]. The structure parameters were provided by VTT Centre for Microelectronics. In particular, the effect of the weak effective index modulation on the grating length required was analyzed to achieve strong distributed feedback. The results were compared to the stratified medium theory, which was shown to predict the spectral position of the resonance peak with inadequate precision.

As a continuation to the work reported in [1] an efficient quasi-rigorous computational method was developed for the analysis of waveguide gratings and applied to SOI structures [2,3]. This method is based on rigorous electromagnetic diffraction theory of gratings and on certain assumptions appropriate for waveguide structures.

b) Phase masks for grating structures

Rigorous diffraction theory was applied to design phase masks for Bragg grating exposure experiments on the proximity mask aligner at VTT Centre for Microelectronics. The method was verified experimentally by fabricating a test phase mask by electron beam lithography.

A new method was found to eliminate the harmful zeroth diffraction order of a phase-mask grating: the grating was coated with a thin film of dielectric material having a refractive index higher than that of the substrate material. This method was demonstrated experimentally [4] using electron beam lithography, reactive ion etching, and vacuum deposition.

c) Photonic crystal structures

Investigations on photonic bandgap structures by rigorous diffraction theory were initiated during the project. The intention was to fabricate such structures in SOI waveguides using the same technology as in the fabrication of waveguide gratings. During the project a novel electromagnetic approach to photonic crystals was developed. The basic idea of the method is the reformulation of the so called C method with adaptive spatial resolution enabling the use of non-identical trapezoidal profiles [5]. The developed method allows the efficient numerical analysis of photonic crystal structures consisting of polygonal rods.

III. FABRICATION PROCESS

The goal of the project was to fabricate integrated optical components based on Bragg gratings with periods on the order of 200 nm. In a waveguide grating the corrugation was intentionally extended beyond the ridge in order to enhance the refractive index modulation in the waveguide region. A schematic of the Bragg grating integrated with a silicon waveguide and a possible application is shown in Figure 1.

Figure 1 – A schematic of a targeted waveguide structure with a grating on top (left) and its integration to a functioning waveguide component (right). The illustrated structure can perform as an add-drop multiplexer by reflecting a selected wavelength and passing through the rest of the wavelengths.

a) Bragg grating structures

According to modeling the depth of the gratings over the waveguide should be in the order of one micron to achieve sufficiently strong reflection from the grating. The fabrication procedure consisted of three critical steps to be optimized: e-beam resist patterning, oxide mask etching, and silicon etching [6]. The direct electron beam resist patterning was performed at the University of Joensuu and the rest of the processing was done at VTT Centre for Microelectronics.

Bragg grating test structures with lattice periods of 225 and 450 nm with intended air-to-dielectric filling ratio of 50% were fabricated on ordinary n-type silicon substrates. A thin oxide hard mask was used in silicon etching in inductively coupled plasma (ICP) to reach the required etch depth of over 1 µm. In order to inhibit the strong underetching in the beginning of the etching and also the sideward etching during the process an etching procedure with a linear passivation was used.

In Fig. 2 scanning electron microscopy (SEM) pictures of the fabricated nanogratings are presented. Due to an artefact of the e-beam writing the filling ratio of the smaller grating could not be precisely controlled. This artefact can be eliminated by optimizing the exposure parameters. The variation of the etch depth followed from the aspect ratio dependent etching (ARDE) effect during the ICP etching. The aspect ratio for the 225 nm period grating varied between 6 and 10 and for the 450 nm
period grating it was 5.3. The dip in the top of the grating structure followed from slight underetching in the beginning of the silicon etching.

**b) SOI waveguides and their integration with gratings**

The ridge waveguide was processed by depositing first a plasma enhanced chemical vapor deposited (PECVD) oxide layer on the wafer. The waveguide structure was patterned by standard UV-lithography and the pattern was transferred into the PECVD oxide hard mask by dry etching. Finally, the waveguide structure was etched into silicon using ICP. Etching process started with a similar continuously passivating etch process as was used for the grating formation and the process was continued with pulsed etching process [7]. After etching, the waveguide structure was covered with an oxide buffer layer. We measured propagation losses less than 0.5 dB/cm at 1.55 µm wavelength.

After the grating and waveguide processes were tested separately, they were combined to form a waveguide grating structure. First, the gratings were processed and then the waveguide process was performed as described above but no oxide buffer layer was grown on top. The PECVD oxide layer used to mask the waveguides covered also the small grating grooves and the nanogratings on the ridge remained intact during the ICP etching of the waveguide. The gratings outside the ridge were exposed for one more oxide and silicon etching steps while forming the waveguides.

In Fig. 3 SEM micrographs of sideview profiles in corrugated waveguide grating test structures are shown. The grating was intentionally turned 90 degrees with respect to the real optically functioning position in order to reveal processing details in SEM-analysis. The period and the depth of the grating were 675 nm and 1.1 µm, respectively. The dimensions of the silicon ridge were 4.5 µm and 7.5 µm, respectively. In the waveguide grating test structure shown in Fig. 3 a) the air filling ratio was approximately 60%. The grating on top of the ridge succeeded well while the grating aside the ridge showed some unintentional unidealities. High, narrow peaks on top of the grating corners resulted from unperfect dry oxide etching when the waveguide etching mask was formed. Plasma etching removed the oxide efficiently from the grooves but in the walls of the groove part of the oxide was left and acted as a mask during the subsequent ICP etching of the silicon ridge structure.

For slightly higher air-to-dielectric filling ratios or slightly smaller periods the situation was even worse. In these cases the oxide in the bottom of the groove was thicker and was not totally removed in the oxide etching. The remaining oxide layer in the bottom behaved as an inverse mask in the silicon etching. The aspect ratio in trenches aside the ridge in Fig. 3 b) was on average 12.

We completed the integration of SOI waveguides and gratings during the project, but optical measurements showed that the process still needs fine tuning. There are two potential modifications to the process that could lead to functioning waveguide gratings, i.e. to replace the oxide dry etching step with wet etching (very straightforward to implement), or to limit the grating only on top
of the waveguide ridge (processing sequence must be modified).

c) Photonic crystal test structures

The processes, which were developed for Bragg gratings, are also applicable to other photonic nanostructures, such as photonic crystals. Photonic crystal test structures were fabricated on silicon wafers with only slightly modifying the process described above. The results were very promising and optically functional photonic crystal structures will be realized by applying the same processes to SOI wafers.

IV. RESULTS AND IMPACTS

Rigorous diffraction theory was applied to model Bragg gratings in silicon-on-insulator (SOI) waveguides and developed for efficient analysis of waveguide gratings. In particular, the effect of the weak effective index modulation on the grating length required was analyzed to achieve strong distributed feedback. The results were compared to the stratified medium theory, which was shown to predict the spectral position of the resonance peak with inadequate precision.

An efficient quasi-rigorous computational method was developed for analysis of waveguide gratings and applied to SOI structures. The developed method allows efficient rigorous analysis of corrugated waveguide structures without any limitations for the corrugation depth. Furthermore, the method also facilitates the analysis of coupling of light from the fundamental mode into higher waveguide modes. So far, this coupling has not been possible to analyze in waveguide gratings with widely used thin-film stack methods.

Rigorous diffraction theory was applied to design phase masks for Bragg grating exposure experiments on the proximity mask aligner. It was found that if a phase mask is coated with a thin film of dielectric material having a refractive index higher than that of the substrate material the harmful zeroth diffraction order of a phase-mask grating can be eliminated. The method was verified experimentally by fabricating a test phase mask by electron beam lithography.

Investigations on photonic bandgap structures by rigorous diffraction theory were initiated during the project. We believe that the computational breakthroughs already obtained will open new avenues in the analysis of photonic bandgap structures in waveguides and permit the analysis of periodic structured with defects, a problem that has appeared almost impossible to handle rigorously in the past.

State of the art fabrication methods for both nanostructures and waveguides were developed. The methods were demonstrated by fabricating and characterizing several test structures and components. The depth and aspect ratio of the nanostructures were much better than what is generally reported in literature. The measured propagation losses of the realized SOI waveguides were one of the lowest reported in the world.

The integration of SOI waveguides with gratings was accomplished during the project. Optical measurements, however, showed that the process still needs fine tuning to eliminate the problems caused by the unintended oxide masking during the waveguide etch.

In addition to the knowledge about the theory and fabrication of SOI waveguides, silicon Bragg gratings and photonic crystals, many fabrication processes and procedures were developed, as well as general knowledge, relating to the use of silicon as a core material in integrated optics. We believe that these will be very valuable during the next few years, both scientifically and commercially.

REFERENCES

Adiabatic logic and its integration with CMOS technology

Markku Åberg, Veikko Porra, Pekka Kuivalainen
Jouko Marjonen, Jacek Flak, Arto Rantala

VTT Centre for Microelectronics
P.O.Box 1208, 02044 VTT
<markku.aberg@vtt.fi>

ABSTRACT

The theme of this study is low power logic design by applying adiabatic logic. Adiabatic logic is based on ramp clocks to reduce thermal dissipation and on charge recycling from capacitive loads. Adiabatic static and pass-transistor logic with applications (multiplication circuit, bus driver) have been studied. Clock generators have been optimized. The applicability of adiabatic principle in driving floating gate neural circuits has been examined.

I. Introduction

Power consumption of digital circuits has become a critical issue in high speed and portable applications. Several means have been proposed to reduce power consumption. On software level for instance:
- Energy efficient algorithms
- Minimalisation of logic
  - active time of circuits/blocks is minimised
  - unnecessary bit processing removed
And on circuit (hardware) level:
- Development of (CMOS) technology
  - lower capacitance and supply voltage;
  \[ P = kNfCV^2 \]
  - V supply voltage, C gate capacitance, f frequency, N number of gates, k activity coefficient
- Shrinking of energy/computational operation
- Adiabatic logic/circuit
- New operation principles for circuits
  - e.g. neural structure instead of digital structure

"Adiabatic circuit" is a fuzzy concept that covers (some of) several methods to reduce power consumption. Basic types are [1]:
- Energy recovery circuits; that recover a substantial portion of the \( CV^2 \) energy invested in logic signals.
- Asymptotically isentropic circuits; that in some appropriate limit (low speed/rise & fall time, low temperature) generate asymptotically zero entropy per operation.
- Time-proportionally reversible circuits; entropy generation per operation approximately inversely proportional to the length of time over which operations are performed.
- Ballistic circuits (today hypothetical); entropy coefficient so low (e.g. superconducting devices) that entropy generation per operation is practically zero.

Practical circuits usually combine a couple of types.

The basic operation principles of adiabatic logic considered here are:
- Voltage or current ramps are used to prevent resistive dissipation in parasitic resistance (asymptotically isentropic). The ramp rise and fall times \( t_r, t_f \gg \tau = RC \) time constant of the circuit or block.
- Charge stored in the gate capacitors of the circuit is collected back to the power supply (energy recovery). This needs an oscillating power supply that is in proper phase to data load/discharge. The power supply can be realized either by LC-type resonance circuits, where C is gate capacitance of logic blocks’ transistors and L an inductor at the power supply, or by many phase circuits, where the charge is circulating from phase to phase.

II. Adiabatic STATIC LOGIC

We have studied both a dynamic-operation-based adiabatic dynamic logic (ADL) [2] and a static-operation-based static logic (ASL). This report with concentrate on the former.

Adiabatic Static Logic (ASL)

In ASL the adiabatic operation have been achieved by avoiding a DC current path from SPS to ground. The circuit is basically a full-wave rectifying circuit. Because the output of the SPS is above the zero voltage through the period it is possible to achieve a full-wave rectification by means of two diodes only. One of the diodes is forward biased toward the source and it provides
the logical ‘1’ state. The other diode is reverse biased and provides the logical ‘0’ state. The purpose of the CMOS circuitry connected between these two states is to allow one of the states to appear on the output. Using this principle the functionality of several logic gates (NOT, NAND, NOR, AND, OR, SR flip-flop and D flip-flop) was verified in measurements. In an inverter chain measurement a 77% power saving was achieved compared to CMOS implementation.

8x8-Multiplier based on the Adiabatic Charge Recycling
In this chapter an 8x8 binary multiplier based on the adiabatic static logic (ASL) is presented. The purpose of the fabricated chip was to research the useability of the ASL gates on a larger system. Because the ASL gates are very similar to CMOS gates it was possible to fabricated a single chip which contained both the CMOS multiplier and the ASL multiplier. Using a prober wiring it was possible to use the chip either in CMOS or in adiabatic mode.

Measurements
The functionality of the circuit was confirmed. In the power consumption test a continuous multiplication was performed between items of a random vector [3]. The vector length was 31 items. The resonator frequency was kept constant whereas the logic’s sampling frequency was varying between 4kHz and 40MHz. There was no synchronization between the input data and resonator signal in the adiabatic power measurement. A layout error in the output buffer made exact power measurements impossible in the first prototype. A corrected version is still in process at the time of writing this report.

III. Adiabatic Bus-Driver
This chapter describes the researches exploring the properties of adiabatic logic on a system-level including design, simulations and prototype tests. The bus-driver was chosen as a system due to its major power-consumption of the application such as microprocessor, display controller, etc.

The driver was designed with the following goal specifications:
- number of lines: 8
- signal frequency: 10MHz
- load capacitance: 10pF per line
- voltage swing: 3.3V

The heart of the system is dual-rail line-driver based on the proposal of Athas et al. in [5]. It has 2 very appealing features: simplicity of the structure and ability to cooperate with standard CMOS receivers.

The circuits were redesigned for the 0.6µm technology chosen for fabrication, and during the HSpice simulations the sizes of the components were obtained as depicted in Figure 1.
For logic 1 (high level) at the input $X$ the driver repeats power-clock signal $\phi$ (from supply in Fig.1) at the corresponding output $Y$. Resistance of the switches together with inductance, tank and load capacitance form an RLC resonant circuit. It generates quasi-sinusoidal slopes utilized for charging and discharging of the load capacitance in adiabatic mode. Switches of the power-clock supply need two control signals $V_{Ma}$ and $V_{Mb}$.

Simulation proved proper system operation and power-saving of 66%. Also several requirements for the timing of the clock signals appeared. To meet them the control logic circuit (Fig.3) was designed. In addition to that logic, the D flip-flops were placed at the input of system, to correct the phases of input signals and provide the inverted signals for line-drivers. They are also clocked from the designed control logic.

![Figure 4. Components’ sizes in the power-clock supply (upper) and the line driver (lower).](image)

**Measurement results**

Waveforms observed during the measurements showed that the pulse shape for the 5MHz operating frequency is the same as simulated. At 7.5MHz it is still close to ideal, but at 10MHz a rather small part of the charge is being recovered. Also, the amplitude is lower than predicted. The explanation lies in imperfection of components of the resonant circuit (poor Q-factor of the available tunable inductors and large capacitor) as well as from the implementation problems (inner lines, package and test board parasitics). Since the parasitic parameter values are not included in the circuit models, and they are very difficult to predict, the imperfections were included in the circuit afterwards. The resulting simulated waveforms come very close to the observed ones (Fig.5).

![Figure 5. Adiabatic Bus-Driver simulation waveforms. From top: $V_{Ma}$, $V_{Mb}$, $\phi$, $X$ for bit sequence [0,1,…,0,1], $Y$, power-dissipation.](image)

![Figure 6. Scheme of the clocks generator.](image)

![Figure 7. Chip microphotograph.](image)

![Figure 8. Output waveforms.](image)

Power consumption of entire system (coming from measured current flow of the voltage sources) was 5.99mW, compared to 9.04mW calculated for standard CMOS approach. That means power saving slightly over 33%, and could be further improved when applied in particular device (replacing the tunable inductor with the high Q-value one after tuning in tests, lines and board optimization, etc.).
Conclusions

Presented system – bus-driver gains significant power-saving over the standard CMOS approach, especially when applied for driving large load capacitance at relatively low frequencies. The simulations of the system showed that over 66% of the energy dissipated by conventional circuits may be saved. Measurements of the prototype chip proved the efficiency of 33%.

It seems to be reasonable to integrate some processing (adiabatic) logic together with the driver, because that would gain better usage of chip area and further reduction of the power consumption (one power-clock may supply the logic and driver).

IV. NEURON MOS CIRCUITS

Introduction: A neuron MOS transistor has a floating gate (FG) and a multiple number of the input gates capacitively coupled to the floating gate as shown in Fig. 1 (a) [6][7]. The accumulating weight of the input couplings controls the channel current in the neuron MOS transistor. The floating gate potential is determined as a linear sum of all the input signals weighted by the capacitive coupling coefficients \( C_i/C_{total} \). The voltage signals are directly summed in the floating gate without any DC-power dissipation. In reality there is always an initial charge in the floating gate due to the processing of the chip. This extra charge is possible to remove by UV erasing. In Figure 1 (c) a schematic cross-section of a neuron MOS transistor implemented with a double polycrystal CMOS process is illustrated. The upper polysilicon layer forms the floating gate and the bottommost polysilicon layer composes capacitively coupled input gates. It should be noted that in addition to the designed capacitances (the coupling capacitances of the input gates) some unwanted parasitic capacitances appear. The dominant parasitic capacitance \( C_{fc} \) in Figure 1 (c)) is between the floating-gate and the substrate. If there are any additional structures, e.g. metal layer above the floating-gate, also other parasitic capacitances are formed. In this study these parasitic capacitances were used to provide calibration i.e. make more accurate neuron MOS structures.

Neuron MOS D/A: Neuron MOS transistors acts like an artificial neuron, i.e., the weighted sum of the input voltages controls the current of the transistor. By adding a specific weights to the coupling capacitors more complex analog computation functions are achieved. A simple digital-to-analog conversion can be obtained if the weights are designed as binary weighted \( i = 1, 2, \ldots, n \), and the neuron NMOS is connected as a source follower as shown in Figure 1 (b). An enhanced version of the D/A converter was designed so that the voltages driving the lowest bits are also weighted by using a voltage division. The neuron D/A converters were processed with a 0.8 mm CMOS technology provided by VTT Electronics. Measurement results show that the presented ideas provide an efficient way to implement DAC circuits even at very high frequencies with low power and minimal silicon area [8].

Neuron A/D converters and calibration: In order to demonstrate how the neuron MOSFETS reduce the total number of transistors in a circuit, and especially how the proposed new calibration structures improve the circuit performance, we have designed and fabricated simple A/D converters utilizing the advantages offered by the proposed neuron transistor structures. A 3-bit A/D converter was implemented by using only 18 transistors (an ordinary CMOS implementation would require 174 transistors). Also a special calibration topology was introduced, which makes it possible to implement more accurate A/D-converters (up to 6 bits) and an artificial neurons circuit.

![Figure 9. a) A schematic diagram of a neuron (a) NMOS (b) a simple neuron MOS D/A-converter (c) a schematic cross-reference of a neuron MOS implemented with CMOS](image)

V. Optimization of adiabatic clock generators

Clocking of the adiabatic logic is done with adiabatic clock generators that are of three main types: charge recycling [9], tank capacitor [9] and resonant clocks. The efficiency of the clock generators for charging and discharging a capacitive node is compared to complementary or CMOS charging of similar node. Charge recycling clocks recycle the charge between phases and the efficiency compared to complementary charging is \( \frac{(M-1)}{M} \times 100\% \) \( (M=\text{number of phases}) \). If number of phases is increased efficiency is increased but control cycle is longer too, which lowers the maximum usable frequency. Tank capacitor clocks store the charge in tank capacitors and their efficiency is \( \frac{N}{(N+1)} \times 100\% \) \( (N=\text{number of tank capacitors}) \). Again higher efficiency causes slower operation because of longer control cycle. Resonant clocking is based on natural resonance of a LC -
circuit. Its efficiency can be even over 97%. The series resistance of the circuit can dampen the resonance and with resonant clocking as small series resistance as possible should be designed. The power consumed at the line and coil series resistances is (1).

when both ends of the coil are oscillating and have a load capacitance C. Resonance efficiency is thus \((1-\pi^2R_c/C/2)*100\%\). Also the MOSFET channel resistance could be added to the series resistance - of course only some approximation could be used because channel resistance is varying all the time as well as the load capacitance C. Power is consumed also at the switching MOSFET channel-resistances, which reduces the efficiency. Shorter control pulse duration decreases MOSFET channel power consumption, but no significant difference has been detected with pulses that have duty cycle under ~25%. Highest adiabatic clocking frequency can be achieved with the resonant clocking because it has the shortest controlling cycle. Also the control logic consumes power which means that no efficiency will be get if the adiabatic circuit is too small. This means that the controller should be optimized to get a maximum efficiency and also the switching MOSFETs should be as small as possible so that they would have a small gate capacitance.

Conclusions

Adiabatic clock power generators, especially resonant clocking can generate high efficiency for charging and discharging a capacitive load. Adiabatic clocking provides both clock and power for the logic circuit which can be a great advantage compared to CMOS where both clock and signal charging are usually nonadiabatic. Because of the power consumption of the CMOS clock power generator adiabatic circuit should be that large that this can be overcome. With a capacitive load a power saving of 92% have been achieved when also the clock generator power consumption was included to the total power consumption. With other clock generators total power saving has been about 50% in maximum. Clock generator should be optimized by the load it runs, because a too large clock generator causes extra power consumption. Also integration of the clock generator and adiabatic chip to a minimum area increases power saving as the parasitic capacitive loads decrease.

References


List of publications:

Diploma theses
2. Kosonen, Jari: Power saving in Adiabatic circuits and implementation of Adiabatic Dynamic Logic (ADL) in 0.8 µm CMOS process, HUT, Espoo, 1999

Theses on preparation
1. Rantala, Arto; PhD thesis on Neuron floating gate MOSFET circuits
2. Marjonen, Jouko; licenciate thesis on Adiabatic static logic
3. Kosonen, Jari; licenciate thesis on Adiabatic dynamic logic and clocking

Journal papers
2. A. Rantala, P. Kuivalainen and M. Åberg. Low power high-speed neuron MOS digital-to-analog converters


Conference papers


5. A. Rantala, P. Kuivalainen and M. Åberg Low power high speed neuron MOS digital-to-analog converters with minimal silicon area NORCHIP’99, Oslo 1999


7. Kosonen, Jari & al.: Implementation of ADL gates with CMOS 0.8 µm process, NORCHIP’99, Oslo 1999


Other activities

Adiabatic logic - seminar at Munich University of Technology (Prof. Nossek) 14 - 14 December, 2001.
ABSTRACT

In blind source separation and deconvolution, several instantaneous or convolved mixtures of source signals are observed and the sources are reconstructed from the mixtures without knowing the mixing system. In this Teletronics I project with three partners, these techniques were developed both theoretically and for applications in antenna array processing and in CDMA telecommunication systems.

I. INTRODUCTION

To state the problem briefly, suppose we have a set of discrete time signals (e.g., signals arriving at a portable phone by various routes from a number of other portable phones within the same cell, or signals arriving at a sensor array). We know or assume that each of the received signals is some linear mixture of a number of unknown source signals (e.g. actual output signals from portable phones, or the signals whose superpositions are measured by the sensor array). Also the coefficients in the mixtures are unknown. Yet, the problem is to find out the source signals and the coefficients, given only samples of all the mixture signals. This is called blind source separation (blind, because we do not know the sources nor the mixing system).

The problem is clearly impossible unless some restrictions are imposed. It turns out that if the number of mixtures is at least as large as the number of sources, and the sources are statistically independent and non-gaussian, then the problem can be solved by the technique of Independent Component Analysis (ICA).

Deconvolution refers to the problem of determining the input signal or the impulse response of a system. Typically either the system or the input is known. The need for deconvolution arises typically from distortions caused by interference and multipath propagation. In blind deconvolution, the system is unknown and its input is unobservable. Consequently, it is a more difficult task. There is a clear connection to blind source separation, too: in that technique we are dealing with multiple inputs and multiple outputs (MIMO), we observe an instantaneous mixture of source signals and a matrix of mixture coefficients have to be estimated, whereas in multichannel blind deconvolution a matrix of FIR filters have to be estimated. Often, both these problems need to be solved in order to recover the original source signals.

The problems of blind source separation / deconvolution have several practical applications in telecommunications and array signal processing including user separation in CDMA, separating signal sources by a collection of antennas without calibrating the array, and deconvolving channel distortions. These were the topic of research in this project.

There were three project partners, initially at Helsinki University of Technology (prof. Oja), Tampere University of Technology (prof. Koivunen), and Jyväskylä University (prof. Joutsensalo). Prof. Koivunen got a full professor's chair at HUT and moved from Tampere in 1999 together with his Statistical Signal Processing research group.

---

1 Leader of the Consortium. Helsinki University of Technology, Laboratory of Computer and Information Science, P.O.Box 5400, 02015 HUT, Finland; email erkki.oja@hut.fi
2 University of Jyväskylä, Faculty of Information Technology, Telecommunications, P.O.Box 35, 40351 Jyväskylä, Finland; email jyrkij@infoma.jyu.fi
3 Helsinki University of Technology, Signal Processing Laboratory, P.O.Box 3000, 02015 HUT, Finland; email visa.koivunen@hut.fi
In the following chapters, the outcomes of the project are outlined.

II. BLIND AND NONPARAMETRIC STATISTICAL METHODS FOR ANTENNA ARRAYS

a) Background / Organizational information

In this subtask, led by Prof. Visa Koivunen, blind and nonparametric statistical methods for antenna arrays were studied. The program provided funding for one full-time researcher over 3 year period. Prof. Koivunen's group is included in center of excellence (SMARAD) in research nominated by the Academy of Finland.

b) Performance of the project

The subtask by prof. Koivunen's research group at HUT was addressing the problem of blind and nonparametric statistical methods for antenna arrays. The nature of the work was basic research. This is attested by a good number of publications: total of 3 journal papers and 12 international conference papers were published (see the reference list at the end of this report). In addition, a book chapter to the Review of Radio Science 1999-2002 was written based on the research done in this subtask. In terms of academic degrees one doctoral degree (D.Sc. (Tech)) and two M.Sc. degrees were earned in this subtask. This research was among the first ones to consider Multiple-Input Multiple-Output (MIMO) models in communications that are now core ideas in future (beyond 3G) wireless systems.

c) Impact

We intended that our BSS methods would be adopted by some application fields, in particular, radio engineering (antenna array signal processing) and wireless communications. We intended that we would have a strong impact in basic research (citations, invitations etc.) The actual impact was largely as expected: some of our papers have been very well cited, we have given invited talks in conferences. Our work on blind receivers and smart antennas was included as a chapter in the Review of Radio Science 1999-2002 that defines the state of the art in radio engineering. The impact on telecom industry depends on the future in standardization. Currently, evolution of 3G is under way and the group member are active in beyond 3G air interface research in co-operation with Nokia.

d) Future expectations

This is a very fruitful research area, attracting growing interest world-wide. We are certainly going to pursue this research, with more flexible and complex models and computational techniques and with more realistic applications. In wireless communications, MIMO systems are among the hottest topics right now and we got a head start to that area thanks to Teletronics program.

III. SYNCHRONIZATION AND BSS IN CDMA SYSTEMS

a) Background

In this subtask, synchronization and blind signal processing in Code Division Multiple Access (CDMA) systems was studied by Prof. Jyrki Joutsensalo and prof. Tapani Ristaniemi at University of Jyväskylä, Department of Mathematical Information Technology, in cooperation wit Prof. Juha Karhunen’s and Prof. Erkki Oja’s group at Helsinki University of Technology.

b) Objectives

The group concentrated on two subproblems of the project: the development and application of blind source separation and timing estimation algorithms to wireless communications.

CDMA (Code Division Multiple Access) technology is a strong candidate for the evolving wireless communications systems. Wideband CDMA (WCDMA) has already been selected for an air interface solution e.g. in UMTS, which will provide a multitude of services, especially multimedia, and high bit rate packet data.

b.1. Application of ICA in wireless communications

In CDMA systems the users share the same frequency band, and thus good care must be taken to limit mutual interference. Multiuser detection (MUD) is a technique which tries to exploit the structure of interference to be able to suppress it. Optimal MUD, however, is computationally exhausting, and requires several system parameters to be known. As a consequence, many suboptimal multiuser receivers and adaptive multiple access interference (MAI) suppression techniques have been studied extensively during the past ten years.

Recently, independent component analysis (ICA) and the closely related blind source separation (BSS) problem have attracted a lot interest both in statistical signal processing and neural network communities. ICA can also be applied to the interference suppression
problem in CDMA. There exist many motivating reasons to use the means of ICA in the reception of a CDMA system. First of all, ICA provides a near-far resistant receiver, being able to resist strong interferences. Resistance is achieved by ICA quite naturally, since ICA only requires the source signals to be statistically independent, but their strengths are allowed to differ. In CDMA the sources are, roughly speaking, users’ symbol streams, and it is hence reasonable to assume that they are independent. Near-far resistance is one of the key requirements of a receiver, and it becomes even more important as there is a demand for higher data rates. This is because many solutions for higher data rates, e.g. smaller spreading factors, and higher symbol constellations, tend to worsen the near-far situation either directly or indirectly via e.g. power control imperfections. Secondly, the propagation delay and the state of the channel should be estimated prior to actual symbol estimation. For subspace-type receivers also the estimation for the model order should be available. The estimation of these parameters will always include some measurement errors, which degrade the accuracy of symbol estimation. ICA, on the other hand, doesn’t need that precise knowledge of the system's parameters, since the estimation is based purely on the (higher order) statistical properties of the signal. Therefore, with ICA we should expect some robustness against erroneous parameter estimation. Thirdly, an ICA block can be used as an add-on feature, to be attached to any existing receiver structure. This makes it possible to consider hybrid receiver structures, in which the ICA block could be intelligently activated only when it is expected to improve performance.

b.2. Code timing estimation in direct sequence (DS) CDMA systems

The final objective in the reception of a DS-CDMA system is to estimate the symbols which carry the data, but a prerequisite task is to get the local code generator synchronized to that of received signal. This means estimation of the propagation delay, which gives the required knowledge to the receiver about the phase of the spreading code.

Matched filter is the traditional method to the problem. Although simple, it is inadequate if the code orthogonality conditions are perturbed. This happens even in synchronous system with orthogonal codes due to the existence of multipaths with different delays. Moreover, if the desired signal is much weaker than the interfering signals, which is commonly known as a near-far problem, matched filter techniques can totally collapse. Most promising performance with low computation can be achieved by so called differentially coherent/non-coherent algorithms, where the effects of interference are suppressed by correlating the matched filter output with a delayed version of it. The goal in this project was to further development of differentially coherent algorithms, in order to gain more interference mitigation capabilities and higher resolution.

c) Results

Regarding the application of ICA to wireless communications, many ICA/BSS methods were successfully applied in this Teletronics project to blind multi-user detection in DS-CDMA. Especially, FastICA was found the most suitable due to its simplicity and fastness. The performance of ICA/BSS methods was compared to existing methods, and also some theoretical results on the convergence were made. In addition, new receiver structures were proposed. Recall that in blind source separation there is no control which source is separated. A CDMA application in mind it is thus not meaningful to use ICA on its own, since the desired signal is well identified by the user-specific spreading code. Therefore, hybrid receiver structures were proposed, where ICA was considered as a post-processing tool for conventional detection (RAKE). By doing this, also the statistical independence of the users’ signals can be exploited. In addition, it makes it possible to alleviate the performance drop due to the erroneous parameters estimation in the receiver. The research resulted in one journal article, one patent pending, a number of conference papers and two invited international talks.

Regarding code timing estimation, many new algorithms were proposed and compared to existing ones. In addition, the average time used to delay estimation was analyzed analytically. Improved interference mitigation and higher resolution was achieved with so called DC-MUSIC algorithm, in which traditional multiple signal classification (MUSIC) was modified into differential mode. In this way, traditional MUSIC no more suffer from high system loads, since differential correlations effectively filters interference prior to actual delay estimation. Also another method for fine estimation was proposed, in which differential correlations enabled the estimation of the fractional part of the delay by solving a system of two second-order polynomials for each code chip interval. The research resulted in one journal article, one patent pending, one book chapter, and a number of conference papers.

REFERENCES

1) Chapter 23: ‘‘Telecommunications”, pp. 417-440 in the book Hyvärinen, A., Karhunen, J., and


Cristescu, R., Ristami, T., Joutsensalo, J., and Karhunen, J., "CDMA Delay Estimation Using a Fast ICA Algorithm", in Proc. of the IEEE Int. Symp. on Personal, Indoor, and Mobile Communications (PIMRC’00), London, United Kingdom, September 17-19, 2000, pp. 1117-1120.


ABSTRACT

Advances in broadband telecommunications, multimedia and virtual reality are expanding human-computer interaction. In addition to normal text and multimedia, increasing levels of telepresence and immersion, including 3D vision, 3D sound, touch and gestures, will be brought to our consciousness through our senses. Although there exist well-developed single modalities for communication, we do not really understand the general problem of designing integrated multimodal telecommunication systems. Yet, managing these modalities is needed for successful implementation of future broadband telecommunication products and services. Recent advances in mobile communication based on picocellular technologies allow the transmission of high-bandwidth of data over Personal Surrounding Networks (PSN). The technology offers more freedom for the design of mobile multimodal 3D user interfaces but does not solve the design problem.

In the PAULA project, we study and apply new and advanced approaches of mobility and augmented reality to multimodal user interfaces, developed essential technology, general architecture and integration framework. We experiment with and demonstrate potential future application by building prototypes and scenarios. The research team is formed among the most advanced researchers of telecommunication electronics, digital media, user interface and virtual reality in Finland. The research results help to scientifically prepare way for the Finnish telecommunication industry’s continued growth in the next millennium with new products and services for mobile broadband telecommunication.

II. Research Approach

Scientifically we are utilising two major approaches to extend the use of computing and communication resources: "ubiquitous computing" and "augmented reality". "Ubiquitous computing" is a term coined by Weiser [2] to mean a situation, where small computational devices are embedded into our everyday environment in a way that allows them to be operated seamlessly and transparently. These devices are active and aware of their surroundings so that they can react and emit information when needed. One implementation of ubiquitous computing are active badges, which can trigger automatic doors and give information about the location of a person. Weiser's team and others at Xerox have experimented the idea by using several types of devices, like small pager-sized "Tabs", notebook-sized "Pads" and whiteboard-sized "Boards" [3].
"Augmented reality" [4, 5, 6, 7] is a research approach that attempts to integrate some form of computer media with the real world. When in ubiquitous computing there are many different active devices, in many cases each of them having their own display and interaction devices, the augmented reality approach usually uses much fewer devices and aims at a seamless integration between real and digital. The integration may be between paper and electronic documents, like in DigiDesk [4], or even more commonly overlaying digital information (as a non-immersive virtual reality) on real world images [5]. The overlaying of images may take place in several ways, like by using video projection [4, 8], by the means of small, hand-held video screens or palmtop computers [6], by mixing surrounding reality with non-immersive VR by using head-mounted see-through displays [4, 9] or through haptic displays [10].

The core idea of our approach has been to use very short distance broadband wireless communication network to mix these two approaches. Thus we suggest, "ubiquitous computing" in the form where different devices in spaces and places we move around are computationally active and can recognise our presence and identity. But instead of a multitude of different displays and interaction devices that the interaction with all devices would take place in an "augmented reality", for example by using a head-mounted see-through display and a mobile phone/remote controller.

Several networked virtual reality environments [11, 12, 13, 14] exist today, but none of them supports mobility of users. Most of the networked environments are based on Internet, which could be easily replaced by a mobile multimedia wireless network like SWAN [15] providing mobile connections. However, typical currently available applications and their interfaces, based on immersive virtual reality and heavy desk-top computers, would still restrict the user’s ability to move and access services in a natural and convenient way. In the Nara Institute of Science and Technology in Japan [16] an experimental mobile virtual reality system is being developed. This system, like our experimental system [17] is based on augmented reality merging both real and virtual environments to provide totally new telepresence services and interfaces to mobile users.

The backbone of the mobile virtual reality is a wireless broadband picocellular personal surrounding network (PSN). The PSN network connects user’s personal mobile terminals like a head-mounted-display or a pen-shaped input device and provides mobile access to other mobile and fixed networks as depicted in Fig 1.

![Diagram of different mobile networks](image)

**Figure 1.** The hierarchy of different mobile networks.

The benefits of using very small cells in mobile virtual reality are obvious. The smaller the cell size, the higher the throughput, because there are fewer users in each cell and higher transmission frequencies can be used. Usually very high frequencies are not used in mobile networks, because of quick signal attenuation, but if the transmission range is just a few meters, the effect of attenuation is almost negligible. In addition, smaller cell size enables greater frequency reuse. The diameter of a PSN cell in our system is going to be some three meters, which enables the construction of small very low-powered hand held or wearable terminals still capable of transmitting high-bandwidth (> 1Mbit/s) multimedia data required by broadband services such as virtual reality applications.

Conventional mobile networks consisting of very small cells have two serious drawbacks: The number of base stations and handovers will be enormous. In our system adjacent PSNs can change information directly without using a fixed base station. This does not only enable wireless communication between user’s personal terminals but makes it possible for two users to transmit data to each other directly too. In fact each user’s personal surrounding network constitutes a mobile base station, which can forward traffic packets between a fixed base station and some user outside the cell around the base station. In this way the number of expensive base stations needed can be greatly reduced. The number of handovers cannot be reduced, unless users’ ability to move is
restricted. But handovers can be made more transparent to the user by e.g. multicasting same data packets to adjacent cells [18]. When a handover occurs, i.e., the user changes a cell, new data packets are already in the new cell and the user does not have to wait the time it takes to forward the data packets from the old cell to the new one. Another problem related to handovers is the availability of services. When a user changes a cell, does the new cell provide the same services as the previous one. In spite of implementing the same services in each network node including the mobile PSNs, it may be more reasonable to get services on demand using e.g. mobile agents [19]. So in addition to data, knowledge about the services required by the user should be multicast to adjacent cells to provide transparent handovers.

In Fig. 2 a block diagram of the mobile virtual reality system is given. The main components of the system are a fixed base station providing access to fixed networks and their services, personal mobile terminals including virtual reality devices and a mobile base station providing connections between mobile terminals and a fixed network. User’s virtual reality devices contain input and output devices and an advanced position tracking system. As an output device we use a see-through high resolution head-mounted-display (HMD), with which we augment the real world electronics and telecommunication products with virtual properties and interfaces. As an input device we use a 3D pen mouse, which provides very natural way of entering data e.g. editing a text file. In order to provide a reasonable augmented reality we should track both the output and input devices very precisely. Several methods for tracking the position exist. From these an electromagnetic method based on spread-spectrum communication providing an accuracy of a few millimetres [20] seems to be most promising for our purposes.

**Figure 2.** The components of the proposed experimental research environment for personal broadband telecommunication services.

***III. RESULTS***

**AD-hoc networks & accurate Indoor tracking**

Living Room is a project that concentrates on the user interface design in the home environment. The living room itself is a laboratory room of about 40 m² that has been converted into a typical single room apartment. The room is instrumented with pressure sensors under the floor, 20 individually controllable light fixtures, computer controllable curtains, electronic locks, etc. The infrastructure for two-way communication between the smart objects is based on in-house developed IR and RF communication systems. The principle is that all objects that are brought into the smart room are instrumented and intelligent.

**Display for a Wearable Computer**

The purpose of this work was to design and construct a small and light head-mounted monocular display, which could be used with wearable computers. The original plan was to make the display device as general-purpose as possible. During the design process it was decided, that the display would be used with a wearable computer system being built at TUT electronics lab. This wearable computer is designed to be used in industrial environments.
TELECONTRONICS Research Programme Final Report.

As the starting point of the display system a miniature display component was selected. The resolution of this display, the Kopin CyberDisplay 320C, is 320 x 240 pixels. The optics designed for the display component was based on the principle of the simple magnifier. In order to minimize distortions the optics consists of two achromatic lenses. The head-mounting system of the display is a folding frame going around the back of the head. This enables the use of eye glasses simultaneously with the display system. The designed display driver circuit accepts industry standard 640 x 480 pixels at 60 Hz analog VGA-signal. Central sections of the electronics are an AD-converter, buffer memories, a DA-converter, an analog video amplifier and a programmable logic device, which controls the digital part of the system.

The constructed display driver board fulfils all set requirements, and the picture quality of the display system is good.

**Differential GPS in Wearable Computers**

Wearable computers and other portable devices are becoming more popular every day. These appliances allow the development of location sensitive applications, which require an accurate and reliable positioning method. The objective of this work was to implement a positioning method for wearable computers to serve a research project on location aware applications and wearable computing at TUT Electronics Laboratory. The work describes methods for improving Global Positioning System (GPS) performance using differential GPS (DGPS) methods. A sophisticated method was developed for improving DGPS performance by aiding positioning with carrier phase measurement processing during poor satellite geometry. The DGPS system was developed purely for research purposes and as such it serves the research of wearable computing and positioning well.

**Deviceless User Interfaces**

In the first PAULA project at the University of Oulu, research concentrated on aspects of the user-interfaces for mobile users based on Augmented Reality technology and equipment, the problems associated with these devices and appropriate telecommunication services for these mobile users. During the course of our research on these topics, the results suggested that a gradual shift of focus was needed. In terms of the User Interface (UI), less emphasis should be placed on the devices needed to provide a UI, but more so on the UI itself. This lead to the concept of 'deviceless' User Interfaces, whereby a non-obtrusive device can provide various UI according to the desired situation. Work on this concept is well under way and the first prototype of the system, called MARISIL [http://marasil.org] is complete and shown in figure 3.

**Visual Cells**

New telecommunications services suitable for a mobile user were also investigated during the PAULA project. Experiments were carried out using a Telepresence system to enable a mobile user attend a meeting while on the move. The results of these tests demonstrated that not only is current technology insufficient to support the mobile user, but that enabling the technology to support the mobile user could open up possibilities in the manner in which future Virtual Enterprises can work. To increase the power of the mobile user a concept called 'Visual Cells' is under development. The Visual Cells concept is based on Augmented Reality technology and aims to provide a highly interactive local environment for a mobile user, allowing mobile users to 'carry' their virtual office with them. Additionally, it also allows the mobile user to be 'teleported' virtually to a remote office and environment while still retaining not only the resources of their own local environment, but also access to the remote environments resources.

**Spatial Augmented Reality**

Accurate registration of objects in AR is difficult so in the PAULA project a dual registration system, using magnetic and visual tracking was implemented. This system provides an improved registration compared to magnetic tracking alone.

**3Dcity Info.**

TUT Digital Media Institute (DMI) has built a 3D City Info system, which combines databases, 2D maps and 3D visualizations. The usability studies show that users greatly prefer 3D views over symbolic 2D maps, although both may be needed. The system uses Java 2 and Java JDBC API and is thus platform-independent. Also a steering wheel navigation and an immersive Java3D user interface has been developed for VRML scenes.

![Figure 3. MARISIL concept for a deviceless user interface.](image-url)
In Paula-project, the system was adapted for mobile use by integrating a GPS, a digital compass, a laptop PC, and optionally a wireless broadband connection to the system.

IV IMPACT

The major impact of the project has affected a number of research fields for mobile Augmented Reality systems. An improved Tracking and registration system will improve the accuracy and operation of mobile Augmented reality systems, using both differential GPS and visual image recognition tracking. The 3d City Info system for mobiles has developed new 3D interfaces and outdoor tracking and mobile services. A new interface hand gesture language called MARISIL was designed and patented. A Visual cell metaphor was developed that can be used for designing better presence services.

Many of the research results for the PAULA project were directly impacted in the EU Wireless Strategic Initiative (WSI) Book of Visions 2000 and WWRF book of Visions 2001. Both of these books are seen as a precursor to research activity for 4th generation mobile systems.

REFERENCES

Teletronics IC Solutions
for Adaptive and Wideband
Wireless Communication Systems

1. Adaptive RF, Anal. and Mixed IC's
2. 217 pmo

Integrated 40-60GHz
circuit blocks, 36 pmo

Academy of F.
1999-2002
CNN Integrated parallel processors multimedia
50 pmo (+40 pmo/02)

60GHz Wireless Modems

70 pmo

TEKES-LALAMO

TEKES-ORAVAT-SLAMU etc.

INWITE-HIGH SPEED A/D
Adaptive radio receivers

>300 pmo

ESPRI-T-INSPECT2: FM-DCSK

36 pmo

CNN/CHAOS

113 pmo

I. INTRODUCTION

The project is an extension of an earlier project supported by the Academy of Finland (SAAR, project Nr. 37713): Adaptive radio architectures and technologies for new radio systems, sub-project 3: Design and modeling of adaptive low-voltage integrated analog circuits, A/D-circuits and RF-circuits for wireless communication.

The objectives listed in the original research program were the following:
1. Adaptive RF circuits for 'smart' antennas
2. Integrated 60 GHz circuit blocks for wireless multimedia systems
3. Integrated implementation of Cellular Neural Networks and chaos dynamics

Already before beginning of the project, some re-adjustment of the individual objectives turned out to be necessary. Making a wider circuit integration possible in wideband radio systems requires deep architectural modifications. These modifications change the borderline between analogue and digital circuits, and between hardware and software blocks. Therefore, the first objective: Adaptive RF circuits for 'smart' antennas had to be replaced by a more system-oriented consideration of the adaptability of a radio circuitry to different specifications. The key words were found to be 'software radio', the direct conversion receiver and use of mixed analogue and digital circuits in the synthesis of modulated signals. The original idea to control the antenna matching was found problematic and was finally abolished because of the intermodulation problems connected with the non-linearities of the impedance matching circuits. The final objectives were selected to
3. New nonlinear circuit applications: Cellular nonlinear networks (CNN) for telecommunication applications and image processing; use of chaos dynamics in spread-spectrum communication systems

These sub-project names will be used in the following project summary.

Other projects on the same topic

During 1998-2001, the following other projects were carried out partly or fully in the same subject area (The amount of research work carried out within the Telectronics project objectives is given in person months, pmo):

1. TEKES- projects, more than 300 person months (pmo) in the area of Telectronics (Kari Halonen)
   - High speed A/D Conversion in INWITE program
   - ORAVAT - "Integrated circuits for adaptive radio receivers" in ETX program and
   - SUMU - "Direct conversion radio receivers" in ETX

2. LALAMO, TEKES, 70 pmo, (Veikko Porra)
   Wideband Wireless Modems, Integrated Millimeter Wave Circuit Blocks

3. INSPECT, ESPRIT/LTR, 36 pmo (Veikko Porra),
   Innovative Signal Processing using Chaos Theory, circuit techniques for a 2.4GHz radio system based on frequency-modulated chaos-shift keying (FM-DCSK)

4. CNN, Academy of Finland, 50 pmo 2000-2002 (Kari Halonen)
   Integration of parallel processing circuits for future multimedia and telecommunication systems and image processing (Academy project Nr. 45796, 1999-2002, 40 pmo in 2002)

From these projects, the total research work within Telectronics project area exceeded 650 pmo (annually 12-14 person-years), and made the project three times larger than expected. Furthermore, the supplementary financing given by HUT: research by staff members, administration, rents and design/testing laboratory costs etc. sum up to 50% of project salary costs or 100 pmo.

The projects of TEKES were more industrially oriented, but they had a central role in achieving close contacts with industry necessary for understanding the practical problems and challenges in the field. The research team could be made much larger than expected. The Telectronics research could be focused on deepening the understanding of the subject area and for writing the doctoral dissertations of the senior team members at a late phase of their doctoral studies.

Co-operation in Finland and abroad

In all sub-projects, there were several co-operating partners in Finland and abroad. Some of them are listed here:

Finland: Industrial enterprises: Nokia Research Center (NRC), Nokia Mobile Phones (NMP), Nokia Networks (NN), Electrobiz Inc., MicroAnalog Systems (MAS) Inc., Ylinen Electronics Inc. (Objectives 1 and 2)

University of Turku, objective 3

International co-operation:

Chalmers University of Technology, Sweden, Prof. Herbert Zirath, Processing of 60GHz PHEMT chips, joint seminars (Objective 2)

Hungarian Academy of Science, design and testing of CNN circuits, Prof. T. Roska, (Objective 3)

University of Frankfurt, CNN Image Processing, prof. Tezlaff, (Objective 3)

Caltech/Jet Propulsion Laboratory, Millimeter-wave integrated circuits design, Prof. Sandy Weinreb, postdoc exchange, Pekka Kangaslahti at JPL 1999-2001

(Planned visiting professorship of Prof. J. Choma from the University of Southern California had to be cancelled because of personal reasons of Dr. Choma)

Postgraduate student exchange: Universities of Warsaw, Cracow and Granada

II. SUMMARY OF RESEARCH RESULTS

a) Adaptive RF, analogue and mixed analogue and digital integrated circuit design for wireless communication systems

The main research topics in this sub-project were design of tunable RF and analogue circuit blocks: low-noise amplifiers, oscillators, mixers etc., and fast high-resolution A/D and D/A converters, and mixed analogue-digital direct digital synthesizer (DDS) circuits for flexible and programmable wideband modulation and non-linearity error correction. The research subjects of the individual researchers were:


Jarkko Jussila, Mikko Waltari 2001, Direct conversion

Olli Viäännänen, 2001, Reducing the crest factor of CDMA downlink signal by adding unused channelization codes, effect of clipping in wideband CDMA system and algorithms for peak windowing mixed A-D design

Jonne Lindeberg, 2001, FIR filters for compensating D/A converter frequency response distortion

b) Integrated 40-60 GHz circuit blocks for wireless multimedia systems

This project was started as doctoral thesis study of Pekka Kangaslahti on 40 GHz wave frequency multiplication. The thesis was examined in Aug 1999. Frequency multiplication is beneficial in millimetre wave signal sources to obtain low phase noise. Furthermore, high efficiency frequency doublers reduce power consumption and need for additional amplification and their small area requirement enables integration of several doublers and other circuit functions on one MMIC -chip. Pekka Kangaslahti then continued as a postdoc team manager. During 1999-2001 he has been an exchange postdoc researcher (financing from HUT-IDC/ECDL) with prof. S. Weinreb at Caltech/JPL, and studying 70 GHz monolithic integrated circuit design. The researchers at HUT have been first Jan Riska 1999-2000, (Licentiate thesis and examination in 2000), and after him Mikko Kärkkäinen. The financing for both of them was arranged via a supporting TEKES project LALAMO, but the work has been supervised from Telecommunications (Kangaslahti, Porra).

All costly RF circuits of a 60GHz front end except one at the end of 2001 were processed and tested as a part of LALAMO using test equipment of MILLILAB at VTT/HUT. The last chip will be processed as a part of the Telecommunications-project. The circuits were processed in a GaAs 0.2, 0.1 mm PHEMT Technology in co-operation with prof. H. Zirath, Chalmers University of Technology, Gothenburg, Sweden.

c) New nonlinear circuit applications: Cellular nonlinear networks, signal processing using chaos dynamics

Since 1990, the laboratory has developed novel integrated CMOS implementations for cellular nonlinear networks (CNN, also called cellular neural networks due to the similarity with Hopfield networks). These are dense analog processors with only local connectivity between the processing units. The processing is based on analog nonlinear dynamics of the cells. It has been shown, that this architecture will allow highly parallel teraflop speed real time processing of moving images to be applied in such demanding applications as in writing pads of personal communicators. All largest and fastest CNN circuits for black and white image processing, especially an 128 by 128 cell universal machine chip and an 176 by 144 pixel video processor chip with teraflops speed, have been developed at HUT-ECDL. The doctoral thesis of Ari Paasio was examined in January 1999.

The research of CNN circuits has led the laboratory in close contact with an international network of research groups in the field of nonlinear system dynamics. The major partners in this network have been prof. Leon Chua, the originator of the idea of CNN from the University of California, Berkeley, and prof. Tamas Roska from Hungarian Academy of Sc. Using the CNN principle as a part of a computer system, Professor Roska's group has developed a Universal CNN Machine, and a programming language to write CNN applications. When equipped with a CNN chip this computer will reach teraflop speed. During fall term 2002, Prof. Chua will be a guest professor at HUT/ECDL.

Analogue integrated circuits of this size (about 1 million transistors) have never earlier been made, and many properties of circuits of such complexity (propagating offsets and disturbances, maximizing yield, minimizing power, adaptive dynamics etc) will need large amount of basic and applied research. This will be the topic of one doctoral thesis (Asko Kananen).

The studies of nonlinear dynamics of CNN's and related structures especially by Prof. Chua at UCB have lead to other application areas of nonlinear dynamics. The chaotic behavior of nonlinear electronic circuits has been shown to lead to spread-spectrum signal processing capability which can be utilized in communication systems. ECDL has been a partner in a European Union ESPRIT long term research project (IT/LTR) INSPECT – Innovative Signal Processing using Chaos Theory. and developed a 500 kb/s 2.4 GHz radio system based on Frequency-Modulated Differential Chaos Shift Keying (FM-DCSK). This is the first experimental ‘chaos radio’ reported so far. The measured performance was verified to follow in multipath propagation conditions closely the theoretical predictions. A prototype chaos oscillator based on CNN architecture (T. Huhtanen) was developed as a part of Telecommunications-project.

III. IMPACTS

Because of the parallel industrially oriented projects all ideas have been effectively adopted by Finnish industrial enterprises, and the results have influenced the product design in industry. The extensive academic cooperation has led to a wide exchange of ideas for future research. As an example, Prof. Leon Chua from UC Berkeley will supervise the doctoral students in the area
of CNN research as a guest professor during the fall term of 2002.

REFERENCES

Only publications of each author during the Telectronics research year period are included.

Book:


Refereed articles in international journals:

1998:


1999:


2000:


2001:


Papers submitted in 2001 to International Journals:


International conference papers:
1998:


1999:


2000:


2001:


Papers in National Conferences and Workshops


Patents:


Doctor's theses


Lic. Tech theses:


Master's theses:


Techniques for the Third Generation of Wireless Systems

Jukka Saarinen, Jarmo Takala, and Jarkko Niittylahti
Institute of Digital and Computer Systems, Tampere University of Technology
P.O.Box 553, 33101 Tampere
Jukka.Saarinen@tut.fi, Jarmo.Takala@tut.fi, Jarkko.Niittylahti@tut.fi

Markku Renfors, Nikolay Tchamov
Institute of Communication Engineering, Tampere University of Technology
P.O.Box 553, 33101 Tampere
Markku.Renfors@tut.fi, Nikolay.Tchamov@tut.fi

Tapio Saramäki
Institute of Signal Processing, Tampere University of Technology
P.O.Box 553, 33101 Tampere
Tapi.Saramaki@tut.fi

Visa Koivunen
Signal Processing Laboratory, Helsinki University of Technology
P.O.Box 3000, 02015 TKK
Visa.Koivunen@hut.fi

ABSTRACT

Recently the most popular techniques used to exploit the limited bandwidth of the radio channel as efficiently as possible have been TDMA and CDMA methods. Also the popularity of multicarrier techniques has increased. In this project, several important topics in the design cycle of a third generation’s receiver design have been addressed including multicarrier modulation techniques, adaptive and blind equalization methods, structures for digital ASIC, and RF-ASIC implementations.

I. INTRODUCTION

Throughout the history, there has been interest to exploit the limited bandwidth of the radio channel to be used as efficiently as possible. Several principal techniques have been used to achieve this goal and lately the popular techniques have been both TDMA and CDMA techniques, as well as multicarrier techniques (OFDM). The role of digital signal processing (DSP) is increasing and there have been great challenges in developing efficient implementations. Since the DSP part is becoming more and more complicated, advanced design tools and methodologies have an important role in making the design task manageable. Another important factor in wireless systems is the analog design of the high frequency front-ends.

Successful design of wireless systems requires careful selection of algorithms, architectures, implementation styles, and technologies. In this project, several important topics in the design cycle of a third generation’s receiver design have been addressed. The contents of the project had two main subjects: a) algorithmic developments for broadband wireless systems and b) hardware implementations for broadband wireless systems. The objective was to develop new algorithms and hardware implementations for the third generation of wireless systems.

II. MULTICARRIER MODULATION TECHNIQUES

In this task, simple multicarrier CDMA scheme for down-link was developed. The scheme is based on differential modulation with favourable performance compared to direct sequence CDMA [13] [14]. In order to design efficient filter banks for the multicarrier scheme, new design methods were developed. Further more design tools supporting the design method were developed.

During the project, also new fundamental results on complex modulated critically sampled filter banks were obtained [15] [16] [17]. These results have enabled in-depth understanding of filter banks in channel equalization tasks. Based on this knowledge, efficient schemes for channel equalisation in filter bank based systems and for complex modulated critically sampled filter banks were developed.

III. ADAPTIVE AND BLIND EQUALIZATION METHODS

Conventionally, equalization is seen as an inverse-filtering task. Here however, we have considered equalization as a classification problem. The idea is to map the received signal into desired binary values by partitioning the signal space into some decision regions. For this purpose, we have studied multilayer perceptron (MLP) neural network equalizers, clustering algorithms...
and decision trees for equalization in a GSM-type environment, where data is transmitted as bursts.

Conventional MLP networks obtained very good bit error rates (BER) in our simulations, but they also needed a lot of computation compared to traditional equalization techniques, such as Viterbi and decision-feedback equalizer. To address this problem, we have used maximum covariance weight initialization technique, which speeds up the convergence and thus less training is needed. Also the use of cascade-correlation learning method decreases the amount of computation needed. There, we start with a network, which has no hidden units (corresponds to a linear equalizer), and then add hidden units to it one by one, if needed. This way, the network has a suitable size for each received data burst and unnecessary computation can be avoided. By combining the weight initialization method and cascade-correlation learning, and by applying RPROP-algorithm for training, we can achieve a significant improvement compared to a conventional MLP network trained with BP-algorithm. Competitive bit error rates compared to Viterbi equalization have been achieved.

The studied clustering algorithm and decision-tree techniques carry out the classification using similar idea, but they are implemented in a different way. The clustering method finds cluster centers from the known training sequence in the signal space and classifies the received data sequence to the binary values represented by the nearest center. The decision-tree partitions the signal space by placing axis-parallel decision boundaries and thereby separates the clusters formed by the set of training examples. Both methods are computationally very light, even though their BERs were not as good as obtained with Viterbi or MLP networks.

In this research, also blind communication receiver structures were derived. Blind receivers do not require training sequences or pilot signals in mitigating inter symbol interference (ISI) caused by multipath propagation. Consequently, information symbols can be transmitted instead and improved effective data rates and spectral efficiency is achieved. In this task, single-input single-output (SISO) and single-input multiple-output receivers were developed. In particular blind equalizers based on cyclostationary statistics of communication signals and prediction error filtering were derived and their performance analyzed analytically as well as in simulation studies [15] [19] [20].

Least-square lattice algorithm was derived for blind equalization based on prediction error filtering. This type of Lattice filtering algorithm lends itself to efficient implementation because of its modular nature. Theoretical results on the performance of the proposed methods were derived in a form of a performance bound. We also showed that most channel coding methods fulfill the requirement of uncorrelatedness of the sequence for blind equalization purposes. The validity of the concept was shown in realistic simulations.

**IV. STRUCTURES FOR DIGITAL ASIC**

In this task, low-power arithmetic structures based on residual number systems were developed [21] [22] [23] [24]. These units are carry-free structures based on modulo arithmetic. The results also include new advanced algorithms and area-efficient architectures for numerical controlled oscillators and especially the results include new record in lookup table compression ratio in NCOs [25].

In this task, also two levels of interconnections for ASIC implementations were considered, local interconnections within computational resources and global interconnections between different computational units. The local interconnections were studied with emphasis on array processors, where data-dependencies of operations require input data to be reordered between the intermediate computational stages. In this project, low-latency interconnection structures have been developed, which can realize special class of reorderings, stride permutations. Such a reorderings are present in several important algorithms, e.g., in discrete trigonometric transforms and Viterbi decoding. A general factorialization of stride permutation matrix has been developed [26] [27]. The resulting decomposition contains sparse matrices, which can be mapped onto simple shift-exchange structures containing only registers and multiplexers. The developed method is general, i.e., reordering networks can be designed for sequence sizes of powers of two and the permutation can be performed over number of ports, which is also any power of two.

In this task, also two levels of interconnections for ASIC implementations were considered, local interconnections within computational resources and global interconnections between different computational units. The local interconnections were studied with emphasis on array processors, where data-dependencies of operations require input data to be reordered between the intermediate computational stages. In this project, low-latency interconnection structures have been developed, which can realize special class of reorderings, stride permutations. Such a reorderings are present in several important algorithms, e.g., in discrete trigonometric transforms and Viterbi decoding. A general factorialization of stride permutation matrix has been developed [26] [27]. The resulting decomposition contains sparse matrices, which can be mapped onto simple shift-exchange structures containing only registers and multiplexers. The developed method is general, i.e., reordering networks can be designed for sequence sizes of powers of two and the permutation can be performed over number of ports, which is also any power of two.

Global interconnections were studied with emphasis on block transfers between coarse grain computational resources. As the complexities of ASICs increases there is need to develop design methods to create global interconnections between heterogeneous resources. For this purpose a parameterisable interface for bus based systems is developed [29] [31] [32]. The interface to the actual communication bus contains a low-level, low-overhead protocol, which allows multiple masters to share the communication medium. The communication unit contains another interface to computational resource but this interface is parameterisable, thus several different units can easily be connected to a shared bus with the aid of the parameterisable interconnection unit.
V. RF-ASIC IMPLEMENTATIONS

In this task, monolithic CMOS/BiCMOS voltage controlled oscillator (VCO) designs satisfying the demanding phase-noise specifications of mobile communication systems were developed[33]. The obtained results have been extended to frequency bands utilized in Bluetooth and WLAN applications[34].

VI. RESULTS AND IMPACTS

The main results of the project have been published in international conferences and scientific journals. The impacts of the research include citations and invited papers and presentations. Several professionals with expertise in the field have been produced, i.e., the project has produced five D.Sc. dissertations [1-5] and seven graduate degrees [6-12]. The project has created new international collaboration, e.g., with Univ. CA, Berkeley; IMEC; Delhi Univ. Tech., etc.

The research performed in this project has been carried out in collaboration with other research projects and the developed technology has been transferred to industry through collaboration. Close co-operation has been with the following projects: VDSL modem design based on filter banks in the TEKES-TLX project "Fast DSL technologies in broadband transmission", filter bank based narrowband interference cancellation approach for CDMA and VCO design developed in the TEKES-ETX project "Digital and Analog Techniques for Flexible Radio", “RF-ASIC” project with Bell Labs, development of array processing architectures in the TEKES-ETX project "System Design in Electronics", and Multiprocessor systems for baseline processing in the Tekes-ETX project "Intelligent and configurable multimedia systems".

The basic studies on filter bank based modulation techniques in wireless communications continue in the Academy of Finland funded project “Advanced Multicarrier Techniques for Wireless Communications”. Application oriented studies on this topic, as well as on MC-CDMA techniques are carried out in the project “Beyond 3G Multidimensional Air Interface” in the Tekes NETS program.

REFERENCES


Fractal processes in telecommunications

Ilkka Norros

VTT Information Technology
P.O. Box 1202, 02044 VTT
ilkka.norros@vtt.fi

ABSTRACT

Stochastic processes with fractal paths have been encountered in telecommunication research in several contexts during the 1990’s. Some mathematical problems related to them were studied in this project.

I. INTRODUCTION

The project “Fractal processes in telecommunications” was set up in order to develop mathematics needed to working with traffic processes with a fractal character. Its objectives were to achieve internationally recognised results on some of the following areas: stochastic analysis of fractal processes, multifractal analysis, fractional Levy processes, fractal aspects of very large networks, and statistics of fractal processes. The educational aims were to create expertise and to produce two PhD theses on these topics. The project resulted in considerable development of expertise, manifested in several journal papers, and in the extension of international research contacts, including real collaboration with joint papers. The two theses are underway and expected to be finished within the year 2002.

II. OVERVIEW OF THE RESULTS

a) Queueing theory

The project produced two kinds of results in queueing theory. Both grew out from earlier work done at VTT on a queue where the input process is a fractional Brownian motion (fBm). The fBm is a zero-mean Gaussian process with stationary increments such that the variance of an increment on an interval of length \( t \) is \( t^{2H} \), where \( H \) is a number between 0 and 1. The usual Brownian motion is obtained as the case \( H=1/2 \). When \( H>1/2 \), the increments of the process are equally dependent at all timescales – a feature that was observed to be typical for data traffic and prompted the study of new kinds of traffic models.

The paper [3] analyses a queue (storage), where the input is Gaussian but the variance is a power only asymptotically as \( t \) grows to infinity. It is shown that, under mild conditions, the large deviations of the queue are similar to those of the queue fed by an fBm.

Papers [1] and [2] have a more applied character, generalising techniques learnt in the fBm context to general Gaussian input processes with stationary increments. The papers show how reasonably accurate estimates of the queue length distribution can be obtained with a straightforward method that also yields the most probable paths along which the rare events happen. Figure 1 shows an example of such a path.

\[ \text{Figure 1} – \text{The most probable path of the input rate process that produces a queue of size 1, when the input process is a superposition of a fractional Brownian motion and a periodic Brownian bridge, and the service rate is 1.} \]

b) Stochastic analysis of Fractional Brownian motion

Several other types of problems related to the fBm were also studied in the project. Paper [8] obtains the fBm as a limit process of a special type of binary trees. The classical binary trees have the geometric Brownian motion as the limit process. The practical motivation of these problems comes from mathematical finance.

Papers [5], [6] and [7] investigate general analytical properties of the fBm. Paper [5] develops an original approach to integration with respect to a fBm. In the paper
[7], maximal inequalities valid for Brownian motion are shown to hold for fBm, whereas in the paper [6] it is shown that the classical Itô isometry of Brownian motion is not valid for fBm even as an inequality.

c) Multifractal analysis
The analysis of data traffic traffic at small timescales has revealed that the traffic processes are inherently multiscale objects with strong rescaling properties. Roughly speaking, this means that the process path has fractal-like burstiness at each timepoint but the local fractal scaling may be different from point to point. The mathematical analysis of truly multifractal measures is a highly sophisticated mathematical discipline. Our contribution, paper [4], introduces a stationary multifractal measure that is constructed by multiplying independent copies of one simple “mother process” so that the time run faster and faster for each copy. With finitely many factors one obtains realisations that resemble measured data traffic quite a lot, whereas the limit process is a true multifractal whose existence conditions and basic properties are found in the paper.

Figure 2 – An example of a realisation of a product of seven similar but time-scaled two-state Markov processes.

III. RESULTS AND IMPACTS
The project gave a strong impact to the study of fractal stochastic processes in Finland, in particular as regards topics arising from applications. The project also invited several foreign visitors with whom many of the articles were written, and organised an international symposium in May 1999.

REFERENCES

Distributed Media Processing in Hybrid Networks

Matti Pietikäinen
University of Oulu
Department of Electrical Engineering
mkp@ee.oulu.fi

ABSTRACT

This research was concerned with processing and refining of multimedia information in hybrid network environments. The transmission and consumption of bandwidth intensive media objects were considered and service presentation techniques were developed for low-to-high bandwidth scalable service units, such as intelligent wireless terminals. Experimental validation of the solutions was done in complete end-to-end systems developed in projects funded by the National Technology Agency (Tekes) and industry.

I. INTRODUCTION

Different types of media are growing in importance in telecommunication network service techniques. The value-added services offered on the top of network infrastructure(s) are one of the main focus of the teleoperators functions in the future, e.g. video, documents, and mixed media conferencing. The network types or elements are connected using interface techniques and different handover/switching mechanisms. Using so called hybrid architecture, having many different connected network elements in delivering service to telecommunications access and content consumers, place high requirements for media provider and delivery through the network used.

This joint research project, carried out by groups from the University of Oulu, University of Maryland (USA) and National Institute of Standards and Technology - NIST (USA), contributed to the areas of media processing, transmission and distribution in varying network elements and end terminals.

In Oulu, the project was aimed to strengthen the basic research component of our emerging research program on distributed multimedia. An important goal was also to strengthen the collaboration with our American research partners.

The responsible reader of the research in Oulu was Prof. Matti Pietikäinen and the activities within and between this project and related Tekes funded projects were coordinated by Prof. Jaakko Sauvola and Dr. Timo Ojala.

The leaders of the research in US groups were Dr. David Doermann (U. of Maryland) and Dr. Omid Kia (NIST).

The total funding obtained for the research from the Teletronics programme was 1 375 000 FIM. The US partners funded their research from their own sources. The project was carried out during the period 1.6.1998-31.12.2001.

II. TOPICS OF RESEARCH

This research was concerned with processing and refining of multimedia information in hybrid network environments. The transmission and consumption of bandwidth intensive media objects were considered and service presentation techniques were developed for low-to-high bandwidth scalable service units, such as intelligent wireless terminals. The main challenge centered around analyzing the media and identifying underlying physical and semantic characteristics that can be utilized in optimizing the service and network performance. The type of network, load, usability characteristics and application processing profiles in complex human-terminal-network service environments were targeted. The hybrid network elements offer the basic control for media domain preparation and affect the final object package delivered through distributed processing units. The user terminal dictates the control parameter profiles for media preparation and transmission that is used to interpret and optimize the media profile characteristics in transmission/consumption events.

Experimental validation of the solutions was done in complete end-to-end systems developed in projects funded by the National Technology Agency (Tekes) and industry. The most closely related Tekes projects led by Prof. Sauvola and Dr. Ojala, and carried out in parallel with this Teletronics project, were:

- Princess - Mobile Media Adaptation
- CTI - Computer Telephony Integration
- Duchess - Wireless Media Telephony Services in Office Environment
Descriptions of the results obtained in the Academy funded project and related Tekes projects are presented in references [1-15].

A part of the project funding was also used to support our research on media analysis methods dealing with document images and videos, see e.g. [16]. This created basis for a deeper understanding of the media characteristics related to this project.

III. RESULTS AND IMPACTS

The research produced new scientific results with high industrial relevance.

In mobile multimedia adaptation, a new multimedia content model was introduced. An architecture called Princess was developed for providing mobile services in a hybrid network environment. Some pilot services, like news, video surveillance and sports events, were implemented and demonstrated to verify the usefulness of the developed approach.

In distributed computing, a mobile agent platform with built-in security capabilities was developed. Among its advantages are dynamic distribution and load management, and service scalability. The Distributed Agent Network (DAN) architecture was developed, providing interconnectivity of a hybrid network and service structure. Implemented pilot systems include Internet Video Call Center and Internet Shopping Center.

In media telephony area, frameworks for building H.323 and SIP compliant client applications with rich multimedia content were developed. The pilot applications included multimedia conferencing and remote application control.

The results of research were reported in over 20 refereed scientific publications. About 30 articles aimed for general public and popularization of science appeared in newspapers and in business/professional magazines. The industrial partners participating in the related Tekes projects were informed about the progress of research by technical reports, workshops and experimental pilots.

The results of research are being exploited in R&D projects and commercial products of industrial partners. A successful spin-off company (Icecom) was established by some students participating in this research.

The research also promoted the development of a new research group at the University of Oulu, MediaTeam Oulu (http://www.mediateam.oulu.fi), and promoted careers of its young directors (J. Sauvola, T. Ojala). The research has also promoted new research projects in this area in Oulu. The key members of this project played key roles in the founding of Mobile Forum (http://www.mobileforum.org).

REFERENCES


ABSTRACT

In the Com² project traffic related problems emerging in broadband multiservice networks and the Internet were studied. Several new results were obtained and algorithms developed. Notably the results include an algorithm for calculating blocking probabilities in multicast networks, an analysis of the performance of the RED buffer management mechanism, and a novel application of importance sampling technique for very efficient variance reduction in Monte Carlo simulations. Further, optimal link control of multibitrate networks was studied and an efficient approximation method developed for the problem. Optimal marking mechanism for congestion pricing in the Internet was developed. In traffic modelling, a new way of estimating the Hurst parameter of self-similar traffic was presented and implemented. A program library of developed algorithms is being maintained.

I. INTRODUCTION

The goal of the project was to develop mathematical and computational methods for the performance analysis of broadband multiservice networks, and, in particular, the Internet. Such methods are needed to provide the basic tools for developing appropriate dimensioning methods and traffic and congestion management mechanisms for the networks.

The topics were broadly divided into three areas: 1) queueing models, 2) simulation methods, and 3) traffic models. All three areas were addressed in the project.

In the area of queueing models, four different problems were studied. First, a queue management mechanism, called RED, was modelled and analysed. The model, consisting of a set of coupled ordinary differential equations, captures the most important features of the system and allows us to analyse the stability conditions of the system. This kind of analysis gives guidelines for setting the parameter values of the RED controller. Second, multicast networks were studied and new algorithms were developed for calculating the blocking probabilities of end users of such a system. Third, optimal link control of multibitrate systems were analysed within the framework of Markov decision processes. This problem appears as a sub-problem in optimal routing of connections in the network. A new approach for the problem was proposed and successfully solved. Fourth, optimal marking of packets in the congestion pricing scheme was studied. Use of expected shadow prices for marking was suggested, and these expected shadow prices were explicitly worked out for some model systems.

In the area of simulation methods, the work focused on developing efficient speed-up methods for the Monte Carlo simulation of blocking probabilities of multibitrate systems. A series of progressively better methods were proposed. The ultimate result was a new importance sampling based method, which in an advantageous way combines analytical results and the actual simulation. Sample generation in this method is done applying so-called inverse convolution. The performance of this method is far better than that of the previously known methods.

The traffic in the Internet is long-range dependent. A popular model for this kind of traffic is fractional Brownian motion, proposed by I. Norros. The estimation of the three parameters of the model on the basis of measured traffic, however, is not trivial. In particular, the estimation of the Hurst parameter, describing the scaling behaviour of the traffic, may require a very large number of sample points if done in a straight forward way. In this work the use of geometric sampling grid, along with Maximum likelihood method, was suggested and the method was fully developed. The advantages of the method were demonstrated; it yields as good or better results than the wavelet based estimation method. As spin-off of this work a new method for generating 2-dimensional fractional Brownian motion was developed.

In parallel with the Com² project, some other projects have been carried out in the teletraffic group of the Networking Laboratory. In particular, COST257 project, funded by Tekes and the industry, dealt with partly the same problems, and in some cases it is difficult to tell exactly which result is due to which project. An attempt, however, is made to give correct indication of this.
II. ANALYSIS OF AN ACTIVE QUEUE MANAGEMENT MECHANISM RED

Random early detection or RED [1] is a widely accepted mechanism for the buffer management in Internet routers. It works by randomly dropping packets even before the buffer is full, thus spreading out the phases of sources that use the TCP protocol and breaking up their tendency to synchronize themselves. Dropping is done with a probability that depends on the current value of the so-called exponentially weighted average queue length.

TCP sources are controlled by their internal flow control mechanism. Packet loss acts as an implicit signal of existing congestion, causing the sources to back off, i.e. to reduce their sending rate.

The TCP source population and a RED controlled queue constitute a closed loop control system whose dynamics is analysed in a series of papers, [2]-[6]. Our central modelling assumption is that the system can be described in terms of the expected values of the random variables. This amounts to saying that variances of the variables are unimportant in comparison to the dynamic behaviour of the expected values. This assumption can be justified when the time constant of the averaging is long, as is the case in practice.

In its full-blown form [5], our model comprises three coupled, retarded ordinary differential equations for the expectations for the instantaneous queue length, the averaged queue length and the aggregate packet rate of the source population. Retardation here means that the propagation delays of the packets in the network are taken into account. The dynamics of the system as described by these equations is found to agree reasonably well with results obtained by simulations.

A slightly simplified model is used in [6] to analyse the stability of the system. Three intrinsic parameters (denoted \( a_0 \), \( a_1 \), and \( a_2 \)) determine the stability region of the system, as depicted in Figure 1.

III. MULTICAST SYSTEMS

Multicasting means a mode of transmission where certain flow on information is sent simultaneously to a group of recipients (also called users). Instead of sending the same stream separately to each recipient, in multicasting a multicast tree is formed, having the origin of the information as the root and the recipients as leaves. The root sends only one stream, which is then copied to each direction at a branching point of the tree. Since in each branch only a single copy of the stream is being carried, multicast transmission has a bandwidth saving nature.

In a more general case, the root sends several multicast transmissions, called channels (c.f. TV program) and for each channel there exists a separate multicast tree. In a dynamic setting the users may join and leave any of the trees as they wish. Sometimes a request of a user to join a tree may be blocked because some of the links along the branch that has to be created may not have sufficient resources for the new stream. The problem addressed in our work is how to calculate the probability of blocking for a user request in such a setting. Prior to our work this problem was largely unexplored.

The problem was solved piece by piece in a series of papers. In [7] and [8], the single link problem is studied and reduced to so-called generalized Engset system (the single link problem means that there is only one bottleneck link, all the other links having, in effect, infinite capacities). The blocking problem in a network was first solved in [9], which resulted from the COST257 project. The algorithm developed, is an extension of the known convolution-truncation algorithm for tree-like access networks. The novel feature is that in the multicast

---

Figure 1 – A cross section of the stability region (constant \( a_2 \)).

Figure 2 – Stability region in the physical parameter space.
application the ordinary convolution operation has to be replaced by a new construct, so-called OR-convolution. The convolution-truncation algorithm remarkably reduces the complexity of computing the blocking probabilities, rendering the problem essentially independent of the network size (as opposed to an exponential growth of the complexity in a more direct approach).

The computational load remains heavy when the number of channels grows (to more than, say, 10 channels). Assuming that the channels can be divided into groups of statistically identical channels, the algorithm can be further simplified by defining a combinatorial convolution operation \([10], [11]\), which can be used to efficiently solve even systems with large number of channels.

The work is further extended to cases where each multicast stream uses layered coding (i.e., higher bit rates for better quality, and lower rates for lower quality) in \([12]\) and \([13]\). The importance sampling simulation methods developed in the project (see Chapter V) are applied to the multicast problem in \([14]\) and \([15]\). A related problem of the signalling load for multicast group management is studied in \([16]\).

In summary, the work carried out in the Com² project has considerably advanced the state of the art of calculating the basic performance characteristics of multicast networks.

IV. RESOURCE ALLOCATION

a) Optimal link control in multiservice network

Optimal routing based on the theory of Markov Decision Processes (MDP) is a well-studied area in traditional telephone networks. Assuming the links to be independent, routing decisions can be made based on optimal control of the resources of a separate links \([17]\). In multiservice network carrying connections with different bandwidth requirements, however, even the single link control problem becomes difficult. An exact analysis is not feasible due to the huge size of the state space.

An approximate approach was presented in \([18]\). The idea was to describe the system in terms of an aggregate process (describing the total usage of link capacity) and approximate this by a Markov process. In our work \([19]\), \([20]\), the idea is not to approximate the actual process (multidimensional Markov process) but approximate the so-called relative cost of states as a function in the state space. Experiments with different sets of base functions were done, e.g. piecewise polynomial (linear, quadratic, cubic) functions with or without cross-terms. The approach in \([18]\) can be viewed in this framework as an approximation of the cost function by piecewise constant base functions.

The parameters of the approximate functions, i.e. coefficients of the base functions, were determined by a least squares procedure, i.e. by maximizing the degree the exact equations are satisfied, leading to so-called normal equations. Efficient recursive algorithms were developed to calculate the matrix elements between different base functions. The results obtained with our method compare very favourably with the published results of other researchers. In particular, piecewise linear approximation worked well. Despite of the efficiency of the developed algorithms, the problem, however, remains hard for large systems.

b) Congestion pricing

The congestion pricing scheme proposed by Gibbens and Kelly \([21]\) has appealing simplicity. It is based on the idea that rational Internet users maximize the difference of their own utility and the cost of network usage, and their behaviour can be affected by economical incentives. More precisely, rational users accept that their terminal devices run algorithms that do this maximization on their behalf. Basically, the users are allowed to send traffic to the network as they wish, but provided that the network gives appropriate feedback to them, the system as a whole will tend towards a state where the total welfare is maximized.

A crucial question here is what is the correct feedback information. It turns out that the so-called shadow price of the resource usage is what should be signalled to the users (e.g. in the form of marked packets). For an unbuffered system the shadow price can easily be worked out. In case of a buffered system, the problem is that the shadow price is not known at the moment a packet leaves the buffer. Some suggestions to overcome this difficulty have been made, such as the use of a virtual queue or time reversal.

The idea in our work \([22]\) is to mark each packet with the expected shadow price, which turns out to be the same as the probability that after the departure of the packet the buffer overflows before becoming empty. This quantity is closely related to the so-called relative costs of states studied in the theory of Markov Decision Processes (MDP) and can be calculated for an M/M/1/K queue with the tools provided by the theory. An alternative derivation for the expected shadow price is also given.

Our analysis gives the best marking strategy given the available information. Unfortunately, the results depend on the modelling assumptions of the queue. As a more robust approach, also approximation of the queue by a diffusion process is studied. This approach calls for online estimation of the diffusion parameters. As only two parameters are involved, the task should not be too difficult.

V. SIMULATION METHODS

Exact calculation of end-to-end blocking probabilities in a network is in general a difficult problem. The difficulty is even more pronounced when the traffic is composed of different traffic classes, distinguished by
attributes such as the bandwidth. In principle, it is easy to write down the expression for the blocking probability: it is the ratio of two sums. The problem is the overwhelming size of the sums. There are a few approximation methods, e.g. the reduced load method, that allow computation to some accuracy. The accuracy, however, is out of the users control.

Monte Carlo (MC) method, based on sampling the terms, suits well to the estimation of the kind of sums appearing here and was adopted in our work. But even MC method is inefficient when very small blocking probabilities are to be estimated. The problem addressed in our work is how the efficiency of the MC method can be improved by variance reduction methods, and, in particular, by the importance sampling (IS). In importance sampling, the more interesting samples are made more probable, and this is compensated for in the statistics collection by giving them a smaller weight.

A series of progressively more efficient methods was developed in [23]-[28] and [14]. The method of conditional means, combined with sample generation by Gibbs sampling, was successfully applied (in COST257 project) in [23] and [24]. Importance sampling was first applied in [25], where the existing sample biasing methods was significantly improved by using composite distributions.

The main achievements are represented by the method described in [26]-[28]. The new method by far outperforms all the previously known methods in variance reduction. Two key components in our method are decomposition and inverse convolution. The method combines in a serendipitous way the power of analytical approach with actual simulations; the simulation starts where the theory leaves us.

Roughly speaking, decomposition means that we break down the end-to-end blocking problem into a set of independent single link blocking problems as depicted in Figure 3. In cases where several links block simultaneously, there is some ambiguity in determining which link the blocking should be attributed to. With a consistent convention, however, we make this assignment unambiguous.

Inverse convolution is the crucial element in our method. It is based on the observation that the steps in calculating the distribution of the sum of independent random variables can be reversed in such a way that each component, one by one, can be drawn a value on the condition that their sum has a given value. This method allows one to directly generate sample points of the blocking states for a system where just a single link has finite capacity and all others are have an infinite capacity (with some precomputed tables the generation of the samples is very fast). With this device at hand, the MC simulation of the interesting probability (the probability of the set of actual blocking states of the single link problem) becomes very efficient; for each generated sample one just records whether it is an actual blocking state or not.

Speed-up factors in the range $10^3$-$10^5$ were obtained. A comparison of the efficiency of the method with other methods is shown in Figure 4 (the Gaussian method is another variant of our method, which the space does not allow to describe in detail here). The inverse convolution method is best suited for high-accuracy calculations, where a large number of samples need to be generated; otherwise the overhead of preparations needed to generate the tables may not be justified.

VI. TRAFFIC MODELING

a) Estimation of the parameters of FBM model

Traffic in the Internet has been found to exhibit long-range dependent behaviour. Roughly speaking this means that the traffic fluctuates on all time scales, a fact that has some implications with regard to how efficiently different congestion management methods may work. In order to study these, it is necessary to develop traffic models that capture long-range dependence. The fractional Brownian motion (fBm), suggested by Norros [30] as such a model, has increasingly gained popularity in recent years. The advantages of the model are its parsimony - the model is basically simple mathematical properties - fBm traffic is self-similar (which, however, leads to many challenging problems).

One of the three parameters is the Hurst parameter, $H$, which describes the scaling behaviour of the traffic, i.e. how fast the covariance function decreases. It turns out that the estimation of the three parameters on the basis of traffic measurements is not trivial. In particular, the estimation of $H$ is difficult. With a naive approach even modest estimation accuracy requires a very large number of sampling points.

In [31]-[33] we present a new approach to the estimation problem. The idea in the method is to use a geometric sampling grid (sampling intervals increase in geometrical progression) instead of an evenly spaced grid. Intuitively, the aim is to sample the traffic at different time scales with few sample points and to avoid collecting redundant information. Also the self-similar nature of the grid seems to fit nicely with that of the traffic model. Indeed, the covariance matrix of the traffic over the geometric grid has a simple structure.

In this setting, we apply the well-established method of maximum likelihood estimation (MLE). This method calls for the inversion of the covariance matrix, which is a major difficulty in the method. For small number of points this can be done by direct numerical inversion. For larger number of points, to obtain a higher accuracy, we develop various approximation methods.

The results show that the goal of reducing the number of sampling points was indeed achieved. It turned out that the method also outperforms the wavelet approach, which is today very popular. Our method may, however, be more sensitive to measurement accuracy at short time scales.

In the hindsight, it can be realized that in essence we transformed the original process to another one by using logarithmic time, and that this transformed process is short-range dependent (to which we apply the standard MLE method). This explains why the method works well.

b) Generation of 2-dimesional fractional Brownian motion

For many purposes, e.g. for simulations, it is desirable to be able to generate sample paths from a given traffic process. Norros et al. [34] presented a top-down approach for generating sample paths from 1-dimensional fBm, a variant of the so-called random midpoint displacement method. The idea of this method is extended to a 2-dimensional fBm process in [35] (most of the work was done in COST257 project). A square is divided in 4-ary fashion into ever-smaller sub-squares. The 2-dimensional fBm process is generated by conditioning the value of the process in each sub-square on the values of the process already generated on the larger scale, i.e. on the environment of the sub-square (the process starts at some scale, where the number of sub-squares is small enough to allow transforming the process into white noise process by the square root of the covariance matrix). With a sufficiently large environment, the generated process very closely resembles a true 2-dimensional fBm.

c) Traffic measurements

Though traffic measurements were not specifically on the agenda of the project, some work in this field was also carried out. In particular, traffic statistics of the modem pool at the ingress of the HUT network were collected. Samples of the results can be found at http://keskus.hut.fi/tutkimus/com2/modempools/modempools.shtml.

VII. RESULTS AND IMPACTS

In the project, a number of new teletraffic results were obtained and related computational algorithms developed. The goals set for the project were well achieved. As mentioned in the Introduction, the methods provide basic tools needed in dimensioning of the networks and designing appropriate traffic and congestion control methods.

Besides the direct scientific results, the project had a crucial role in allowing the teletraffic group of the Networking Laboratory to reach a critical size and to get its operation on a stable basis. At the time of writing the group consists of some 15 researchers, among them three post docs.

From the educational perspective, the results were also positive. One doctoral degree was obtained during the project, and another doctoral thesis, largely based on the work done in the project, is now in the pre-examination phase. Further, a crucial part of the doctoral thesis of a visiting researcher was done in the project. In addition, two Master’s theses were completed.

A side result of the project was that a program library, Qlib, of algorithms needed in queueing theory was developed and maintained [36]. Initial work to create the
library had been done in an earlier project at VTT, but in the course of the Com^2 project several new algorithms were developed, tested and added to the library. The algorithms have been coded in Mathematica, but for many algorithms also fast implementations in C were coded. These can be called from the Mathematica by the MathLink mechanism. The library is completely open: http://keskus.hut.fi/tutkimus/com2/Qlib/index.shtml.

REFERENCES


FIT – Future Internet: Traffic Handling and Performance Analysis

Jorma Virtamo
Networking Laboratory, HUT
P.O.Box 3000, 02015 HUT, Finland
jorma.virtamo@hut.fi

Ilkka Norros
VTT Information Technology
P.O. Box 1202, 02044 VTT, Finland
ilkka.norros@vtt.fi

ABSTRACT

The FIT project has addressed a number of problems arising in controlling the traffic and providing quality of service in the Internet and in analyzing the performance of the system. An efficient recursive algorithm has been developed for calculating the flow level performance of elastic traffic under a bandwidth sharing scheme called balanced fairness (BF). The notion of BF has been extended and applied to the flow throughput estimation in ad hoc networks; also other applications of BF are being explored. Scheduling mechanisms, in general, and adaptive scheduling mechanism with delay bounds, in particular, were studied. A simple adaptive and distributed load balancing mechanism has been suggested and analyzed. In this system, traffic is gradually redistributed based on measured link loads, leading to a nearly optimal performance. Analytical results have been obtained on the performance of MAC protocols in ring networks employing optical burst switching. For traffic matrix estimation, the Gravey-Vaton method has been analyzed in detail in the case of Gaussian traffic variations. The stability problem of an overloaded network with measurement-based admission control has been analytically solved.

I. INTRODUCTION

FIT is a joint project between the teletraffic theory group of the Networking Laboratory of Helsinki University of Technology (HUT) and the VTT/UH (University of Helsinki) group. The site http://www.netlab.hut.fi/tutkimus/fit/ provides general information about the project. The project started in 2001 and continues to the end of 2004. For various reasons, the center of the gravity of the project at HUT has shifted towards the end of the project and at the time of this writing (March 2004) the project is still going on with five researchers working full time. Therefore, this report is not complete but more results will be produced.

At a general level, the objective of the project has been to study methods for traffic handling in the Internet in order to avoid or manage congestion and methods for providing Quality of Service (QoS) and QoS differentiation and, finally, to develop methods for analyzing the performance of such systems. The specific topics studied are detailed below.

Balanced Fairness (BF) discussed in Section II can, on one hand, be viewed as an ideal bandwidth sharing scheme. On the other hand, BF can be seen as a computational approximation tool for analyzing more easily implementable schemes like max-min fairness. Scheduling, which is the topic of Section III, is the key mechanism for QoS differentiation in the routers; packets belonging to different traffic classes are handled differently. Load balancing is one of the basic tasks of Traffic Engineering (TE) in a network. Section IV introduces a simple adaptive mechanism by which an almost ideal load distribution is obtained without knowing the traffic matrix. At the transport level, modern broadband networks utilize optical technology, which poses different type of traffic handling problems. In Section V, the performance of MAC protocols in an optical ring network using optical burst switching is analyzed. Many traffic handling operations at the network level require knowledge of current traffic demands between different origin-destination pairs. Estimating the traffic matrix on the basis of link load measurements, however, is a strongly under determined problem and poses a big challenge. Results of a study related to a recently introduced method are discussed in Section VI. Some traffic measurements have also been done in the project as briefly documented in the next section. Section VII deals with another traffic handling method, measurement based admission control for networks subject to overload. The analysis of a simple network model demonstrates how the lack of complete state information makes the stability of the system very sensitive to its parameters.

II. BALANCED FAIRNESS

In the Internet, the bandwidth resources of the network are shared between all concurrent traffic flows. Typically, the sharing is determined by a flow control protocol such as the TCP (Transmission Control Protocol), which is used by the majority of applications. An objective of bandwidth sharing is some kind of fairness, but what fairness means...
when there are many resources (link capacities) shared by several flows with different routes is not obvious. Many definitions have been developed, including the classical max-min fairness and proportional fairness. These are special cases of a more general class of so-called utility based fairness criteria [1].

A new notion of fairness, called balanced fairness (BF), has recently been introduced by Bonald and Proutière [2]-[4]. The most important property of BF is that it leads to completely insensitive network performance: the performance under BF depends solely on the traffic loads on different routes but not of any other traffic characteristics (e.g., flow size distribution). This is a very desirable property from the point of view of networks design and management. Another remarkable property of BF is that the performance in a real dynamic setting, where new flows arrive stochastically and depart upon completion, can be evaluated in an analytic form for many important network topologies.

The analysis of a network performance under Balanced Fairness is based on the theory of Whittle queueing networks [5], i.e. networks of Processor Sharing (PS) nodes the capacities of which, $\Phi_i(x)$, depend on the global network state $x$ in a “balanced way”. This condition means that there exists a balance function $\Phi(x)$ such that

$$\phi_i(x) = \frac{\Phi(x - c_i)}{\Phi(x)} \quad \forall i, x,$$

where $x - c_i$ denotes the network state with one class-$i$ customer less than in state $x$. Balanced Fairness refers to a capacity allocation that satisfies the balance condition (1) and at the same time utilizes the network resources maximally.

An important contribution of the FIT project was the development of a recursive method for an exact calculation of the normalization constant of a BF system [6, 7] (from the normalization constant all the performance metrics can be easily derived). Notably, the algorithm completely avoids solving the balance function explicit. The recursion is also very efficient: in a system with $n$ flow classes only $2^n$ numbers have to be recursively calculated, in contrast to some $N^n$ numbers in a direct calculation, where $N$ typically, depending on the desired accuracy, is of the order of 100.

The method has been applied to practically important topologies for access networks, like parking lot and general tree with and without access rate limits. An implementation of the recursion for a general tree network has been added to the Qlib library [8]. As an important side result it was proven in [7] that BF is Pareto efficient in all tree networks, i.e. no network resources are wasted by the balance condition. Explicit results have been worked out for several example networks. In Figure 1 a four level tree network is illustrated and Figure 2 shows the flow throughput in this network for class-10 flows going through the links 10, 7, 3 and 1.

As noted before, the distinguishing feature of BF is its insensitivity. In contrast, all utility based fairness schemes are sensitive except for some special network topologies, where they coincide with BF. Indeed, Bonald and Proutière have shown that in order for any queueing network of PS nodes to be insensitive the resource allocation has to be balanced. This result was extended to any network of symmetric queues in [9].

An extensive study of the sensitivity properties of various schemes was undertaken in [10]. The study confirms that the non-BF schemes are sensitive (their performance was evaluated by simulations). The sensitivity on the flow size distribution, however, is not very strong; more pronounced is the sensitivity with respect to so-called time scale changes. It was also found that, largely speaking, BF provides a reasonably good approximation for the performance of max-min fairness. Furthermore, it was confirmed that in a network topology called hypercycle, BF is not Pareto efficient but some capacity is wasted. This is illustrated in Figure 3. Fortunately it turns out that in a dynamic setting the impact of the capacity waste is minor. An analytical representation for the balance function of a symmetric 3-link hypercycle was obtained in [11].

The concept of BF was extended in [12] to the case where the constraints are more complicated than fixed link capacity constraints. In particular, [12] outlines an approach for the case where the flows can be split over several routes
and the constraints are then of the type encountered in multicommodity flow problems. The important insight is that in a given state \( x \), given the balance function for all states \( y < x \), the balance condition (1) fixes the capacity allocations up to a multiplicative factor \( 1/\Phi(x) \). Then, \( \Phi(x) \) can be recursively determined by making it at each point as small as allowed by whatever constraints the system is subject to.

This idea was further developed in [13], where the joint problem of scheduling and resource sharing was studied in an ad hoc network, where the simultaneous use of some links is limited by the interference. In an idealized model, the effect of scheduling can be viewed as allowing, under certain limits, a shift of capacity from one link to another one. Again, the system constraints are not just fixed link constraints. By applying the extended BF principle, it was possible to determine both the schedule and resource sharing in such a way that the resources are maximally utilized while preserving the balance property guaranteeing the insensitivity and robustness of the system. In addition, analytical tractability is retained, at least for smaller systems.

In the FIT project, several studies related BF are still going on. In [14] the possibility of insensitive adaptive routing is being studied. In insensitive routing, not only the capacity allocations are balanced but also, separately or jointly, the arrival rates determined by the probabilistic routing. In another study [15], queueing network models, such as BF, are being applied to the analysis of flow throughputs of the data traffic in a cell of a cellular network with finite user population. Application of BF to closed queueing networks was explored in [16]. A general tutorial on Balanced Fairness is also available on the web site of the project [17].

### III. SCHEDULING AND QUALITY DIFFERENTIATION IN Diffserv

During the last decade the Internet has developed into a public multiservice network that should be able to support heterogeneous applications and customers with diverse requirements. For this reason, quality of service (QoS) provisioning in the Internet has gained increasing attention. General service architectures, e.g. Differentiated Services (DiffServ) [18], have been proposed in the literature for providing QoS. In DiffServ, common resources are allocated among service classes. Packet scheduling is the mechanism primarily responsible for the allocation.

In [19] and [20], we studied issues related to quality differentiation, traffic mapping and scheduling in DiffServ. The focus was on the differentiation of two important quality parameters, capacity and delay. Two models were concerned in detail, absolute capacity differentiation and proportional delay differentiation with delay bound, and various packet schedulers were investigated by simulations. According to the results, provisioning and differentiation with static resource allocation methods is problematic but with some adaptive schedulers tunable so that consistent differentiation can be achieved. Based on these observations, we suggested that in DiffServ networks an adaptive scheduler with delay bound should be used for resource allocation. We also argued that from applications point of view it is beneficial to map different traffic types into separate service classes.

Figures 4 and 5 illustrate the behaviour of an adaptive scheduler (HPD with delay bound) in a simulation run with mixed traffic in four service classes. It can be seen that the bandwidth allocation follows quite well the development of queue lengths: if the queue of a class starts to build up, more resources are allocated to the class.

### IV. ADAPTIVE LOAD BALANCING USING MPLS

Multiprotocol Label Switching (MPLS) [21] brings up new possibilities to improve the performance of IP networks. Notably the explicit routing of MPLS facilitates balancing the load by moving traffic from a congested part of the network to some other part in a well controlled way.
In [22], we studied adaptive load balancing based on measured link loads without knowledge of the full traffic matrix. As an objective for load balancing we used minimizing the maximum link utilization and minimizing the mean delay in the network. A simple adaptive and distributed algorithm was presented to obtain these objectives. In addition, a numerical method was developed to evaluate the performance of the algorithm. The method was applied to three test networks. The results showed that the maximum link utilization and the mean delay obtained in a reasonable number of iterations are very close to the optimal values, even with random traffic fluctuations in the timescale of the measurements.

Figure 6 illustrates how our adaptive algorithm approaches the lowest possible level of the maximum link utilization (0.54) in less than 100 iterations when applied to a test network with 10 nodes, while the standard min-hop routing results in a much higher level of congestion corresponding to the maximum link utilization of 0.88. Three different adaptation granularities are applied. With a finer granularity (g = 50, 100), the adaptive algorithm converges slightly slower but much more smoothly than with a coarse granularity (g = 10).

V. OPTICAL BURST SWITCHING

Wavelength division multiplexing (WDM) is a technique, where several optical channels are used concurrently to transfer massive amounts of information, even terabits per second, in a single optical fibre. Optical networks employing WDM play already an important role in the current backbone networks and the trend is that WDM will be adopted also in MAN and finally in LAN networks.

In optical burst switching network optical bursts are used to transfer the data [25]. Each burst consists of several concatenated (IP) packets all having the same destination node and thus routed along the same path. Hence, the optical burst switching (OBS) can be seen as an intermediate step from the wavelength routed networks (i.e. circuit switching) towards the optical packet switching.

In an OBS network the necessary resources are typically reserved only for the duration of the burst. In particular, the source node first sends a control packet or frame to inform the receiver (and possibly intermediate nodes) about the coming burst. The receiver and intermediate nodes each check, one at a time, if the switches along the path can be configured to deliver the burst successfully towards its destination. Meanwhile, after a certain offset time, the actual data burst is sent along the same path without waiting for any acknowledgment from the receiver (or intermediate nodes).

Optical ring network is a suitable solution for metropolitan area networks (MAN). One cost effective solution using the OBS paradigm is described in [26, 23], where each station of the ring has a dedicated fixed “home wavelength channel” for transmitting its bursts. In addition to these data channels also a shared control channel with a certain number of circulating control frames is used to inform the other nodes about the arriving bursts resulting a time slotted system. Thus, the number of wavelength channels $W$ is equal to $N + 1$, where $N$ is the number of stations. Furthermore, each station has only one adjustable receiver. Thus, no transmissions collide in the fibre. But, as each station can listen to at most one channel at a time, burst losses may occur at the receiver in the case two or more concurrent bursts have the same destination node. The traffic pattern for a case with $N = 4$ nodes is illustrated in Figure 7.

In [24] we have studied, both analytically and numerically, the performance of the MAC protocols proposed in [26] under static traffic conditions. In particular, we have considered both random order and round-robin transmission policies. When a station operates in random order policy the transmission queue to be served next is chosen randomly.

Figure 5: Queue lengths as a function of time in a simulation run with an adaptive scheduler.

Figure 6: The maximum link utilization as a function of the number of iterations in a numerical evaluation run of an adaptive load balancing algorithm.
among the non-empty queues. In round-robin transmission policy the non-empty transmission queues are served in a fixed order so that during a one full period one burst is transmitted from each queue. In the analysis we have considered an arbitrary receiver and derived formulae for the burst blocking probability and the so called (receiver) efficiency, i.e. the proportion of the time the receiver is active. In Figure 8 the efficiency is depicted for both random order and round-robin transmission policies in a certain traffic case. It can be seen that the round-robin order, while generally being a more fair, leads to a worse efficiency, and hence, to a higher blocking probability.

In [24], special attention has been given to the performance under an extremely heavy load, where each source has always bursts to be sent to all other destinations, i.e. the offered load is \( \rho = 1 \). Although this symmetric heavy traffic scenario does not hopefully exist, it gives a lower bound on blocking probability for each MAC protocol and allows comparing their worst case performances. Note that in an ideal situation the blocking probability would be zero and each receiver busy all the time, leading to an average pairwise throughput of \( C/(N - 1) \), where \( C \) is the capacity of one channel and \( N \) the number of nodes. However, in [24] we have shown that without any coordination between the nodes the actual throughput will be considerably less, i.e. about a half of that. This heavy traffic scenario is depicted in Figure 9. From the figure it can be seen that the mean burst size should be about 5 slot times or more in order to achieve a reasonably high efficiency, where the slot time is equal to the interarrival time of control frames.

VI. TRAFFIC MATRIX ESTIMATION

The traffic matrix \( x \), which gives the volume of traffic between each origin/destination (OD) pair in the network, is a required input in many network management tasks. Unfortunately, the traffic matrix cannot be directly measured. Only the link loads \( y \) and the routing matrix \( A \) are available. These satisfy the relation

\[
y = Ax.
\]

Since in any realistic network there are many more OD pairs than links, the problem of solving \( x \) from \( A \) and \( y \) is highly underdetermined and the problem is ill-posed.

To overcome the ill-posedness, some type of additional information has to be brought in to solve the problem. Typically, prior information is incorporated into the traffic matrix estimation using different traffic models. Most promising proposed methods include Bayesian inference techniques and network tomography [28, 29, 30].

The Vaton-Gravey iterative Bayesian method [31] consists of iteration and exchange of information between two “boxes”; the first calculating an estimate for the traffic matrix, given the observed link counts and parameter values of the prior, and the second updating the parameter values. Both boxes involve extensive numerical simulations or numerical algorithms.

In the FIT project, a study [27] of the above idea has been conducted with simplifying assumptions. The aim has been to gain insight into the method and, in particular, the output of the first box by examining a model simple enough to be computed analytically. Independence and normality are assumed for the OD pair distributions. This reduces the complexity of the Vaton-Gravey method. Our prior information...
consists of the mean and the covariance matrix of the Gaussian distribution. The attractive feature of this approach is that the distribution conditioned on the link counts is again Gaussian and analytical results can be obtained. When iteration is performed, it turns out that the expected value of the mean does not change in the iteration after the first conditioning on link counts has been made.

We have shown that when the mean, \( m \), and the covariance, \( C \), of the Gaussian distribution is iterated, the expectation of the result \((m, C)\) of an iteration round can be written as

\[
\begin{align*}
    m^{(i+1)} &= G \mu + H m^{(i)}, \\
    C^{(i+1)} &= \tilde{C} + GA \Sigma (GA)^T,
\end{align*}
\]

where \( G, H \) and \( \tilde{C} \) depend on the covariance of the prior distribution, while \( A \) and \( \Sigma \) are constant matrices. The index \( i \) refers to the iteration round. We have proven that the mean \( m \) does not change after the first iteration.

Figure 10 illustrates the situation when two OD pairs are to be estimated from a single link count. On the left is the prior distribution. The real distribution that we are trying to find out is shown on the right. From equation (2) we know the sum of the two variables \( x_1, x_2 \), telling that the mean of the distribution is somewhere along the line shown in the picture. The resulting distribution is shown in red, and is the estimate of the real distribution based on the prior distribution and link loads. The green dots are produced by sampling from the conditional distribution, and are seen to coincide with the explicitly calculated result, as expected.

Figure 10: Result of the first iteration.

Directions for future work may include for example studying the possibility of utilizing the covariance matrix in the estimation. This requires assuming a certain relation between the mean and the variance of the OD pairs.

**VII. MODEM POOL TRAFFIC SURVEY**

We continued collection and analysis of the HUT modem pool statistics started in the previous project (Com²) funded by the Academy of Finland. In addition to the earlier samples from years 1997 and 1998 surveyed in [35], we got new samples in 2001 and 2002. A comparative study was carried out and the results were presented in [36]. An illustration of the results for 2002 is given in Figure 11. A comparison with the results of 2001 shows that the daily usage profile has remained very much the same but the overall traffic level has decreased almost to a half, presumably due to a shift from modem to ADSL connections.

![Figure 11: The number of student users as a function of time during the weekdays in October of 2002.](image)

**VIII. STABILITY ANALYSIS OF AN ADMISSION CONTROL MECHANISM**

In DiffServ Networks, the suggested resource allocation mechanisms for bandwidth brokers are based on incomplete information about the network state. In this project, we have investigated the impact of this lack of information to the stability properties of the mechanisms. To facilitate analytical treatment, we have restricted ourselves to a simple two-server network with feedback admission control, depicted in Figure 12.

Rather surprisingly, it was found that this simple network has a non-trivial stability region. A paper on this discovery has been submitted for publication [32]. The solution of the problem is depicted in Figure 13, with unit input rate. The parameter space of the server rates is partitioned into four regions. Here \( A_1 \) is the domain where the system is never stable, no matter how strict admission control is employed. Regions \( A_2 \) and \( A_3 \) represent the situation where the overloaded network can be stabilized; \( A_2 \) requires strict enough admission control to be stable, while \( A_3 \) is stable with any
control threshold. In $A_4$ there is no overload present, and thus the network is always stable. The phase diagram illustrates how accelerating server 1 drives the system towards more stable regions, while rather paradoxically, speeding up server 2 may unstabilize the network. In [32] these results are proven and extended to more complex controllers, giving easily verifiable stability conditions for the network in terms of the system parameters.

Another related problem studied in this project was connected with traffic modelling. An interesting new limit process had been recently found (later published as [33]) that resulted in simultaneous increase of the number of sources and the timescale. A proof of an analogous result for univariate distributions with a somewhat different traffic model was found independently, but the authors of [33] extended their paper to cover this model as well. The attention was moved to another aspect of long-range dependent traffic: its behaviour under measurement-based admission control. The goal was, in rigorous mathematical terms, to understand the empirical observation of [34] that such a control almost removes long-range dependence, and to study how reliable this kind of feedback mechanisms would be. This type of question is closely connected with the question studied in [32], and there are plans to extend the type of results obtained in [32] for stability of Markovian networks to limiting distributions with heavy-tailed input processes.

IX. RESULTS AND IMPACTS

Within the FIT project so far two Master’s theses have been completed (J. Antila and V. Timonen) and one more will be completed in 2004 (S. Liu). One PhD work will be completed in 2004 (E. Hyrylä), two others are at an intermediate phase (L. Leskelä and R. Susitaival), and two more at an early phase (I. Juva and J. Leino).

The FIT project has been an important element in helping the HUT and VTT groups to strengthen their international contacts and collaboration. During the project the groups have been actively participating in the action COST279, where also work done in this project has been reported. Both the VTT and HUT groups were invited to become members in Euro-NGI, a Network of Excellence of the EU 6th Framework Programme. Altogether 58 partners are involved in this three-year activity, which was started in December 2003.

The appropriation for Senior Scientist received by Prof. Virtamo in association to this project enabled very fruitful visits to France Telecom R&D and to Cambridge University. The work on balanced fairness was initiated during these visits, and the collaboration with the researcher’s at France Telecom R&D continues. Several studies related to this topic are going in the Networking Laboratory.

Prof. Virtamo has served as a member of Scientific Committee of the forthcoming program on Queueing Theory and Teletraffic Theory at the Institute Mittag-Leffler of the Royal Swedish Academy of Sciences. Several researcher of the project are going to participate in the Program by a longer stay at the Institute during autumn 2004.

Network Laboratory of HUT hosted the 16th Nordic Teletraffic Seminar, NTS-16, which was held in Otaniemi in August 2002 [37]. The seminar was supported by a separate grant from the Academy of Finland.

The results of the FIT project have been annually presented in a half-day seminar jointly organized with the domestic COST279 project funded by Tekes. Researchers from the academia and industry have been invited to attend the seminar. Links to the programs of the last two seminars are given in [38].

The program library of teletraffic theory functions [8] has been maintained and developed. New functions and algorithms have been added into the library, notably those related to calculating the performance under balanced fairness, as well as algorithms needed for network calculations and synthetic network generation for simulation purposes.

References

BLIND SIGNAL SEPARATION IN COMMUNICATIONS RECEIVERS AND ANTENNA ARRAY SYSTEMS

Erkki Oja\textsuperscript{1} (Project Coordinator), Visa Koivunen\textsuperscript{2}, and Jyrki Joutsensalo\textsuperscript{3}

1. Helsinki University of Technology
   Neural Networks Research Centre
   P.O. Box 5400, FIN-02015 HUT, Finland
   erkki.oja@hut.fi

2. Helsinki University of Technology
   Signal Processing Laboratory
   P.O.Box 3000, FIN-02015 HUT, Finland
   visa.koivunen@hut.fi

3. University of Jyväskylä
   Lab
   Post address
   jyrki.joutsensalo@jyu.fi
ABSTRACT

In this Teletronics II project of the Academy of Finland, blind source separation (BSS) is proposed as an advanced add-on tool in DS-SS communication systems and antenna arrays. The main new results were as follows:

First, BSS was considered as a pre-processing tool for blind suppression of interfering jammer signals in DS-SS communication systems utilizing antenna arrays. The motivation for this is the ability of BSS/ICA to provide an un-jammed signal to the conventional detection, and thus improve the detection performance under different jamming scenarios. Two possible switching schemes were considered, called pre-switching and post-switching, which activate the ICA-based jammer canceler only when it is expected to improve conventional RAKE-based detection.

Second, BSS/ICA was considered to enhance the performance of conventional interference cancellation schemes. The new scheme combines the benefits of existing BSS based schemes and conventional successive interference cancellation (SIC). The motivation for this is to enhance the system capacity, since conventional interference cancellation schemes fail in highly loaded systems. However, the use of over-complete ICA somewhat circumvents this problem and hence increases the system capacity.

Third, adaptive decision feedback equalizers (DFE) and subspace methods for equalizing FIR-MIMO channels were developed. The main contribution is a novel MIMO DFE which belongs to the family of non-connected MIMO DFE's. This implies that individual DFE's are applied on the received signals. This type of DFE is used in combination with a Kalman filter in order to achieve equalization of time-varying channels.

1. INTRODUCTION

This Teletronics II project proposed three subtasks: 1. basic research into adaptive separation algorithms, 2. blind MIMO equalization and antenna array processing, and 3. blind source separation for spread spectrum receivers. On the topic of the first subtask, the participating laboratories (especially at Helsinki University of Technology) have extensive projects going on mostly outside this Teletronics programme, as parts of their Academy Center of Excellence programmes. The CIS lab also worked on separation algorithms in an EU project. It is somewhat difficult to separate the work done in the Teletronics project from those more extensive research efforts. Therefore, this report concentrates on the more concrete telecommunication approaches in subtasks 2 and 3 and refers to the theoretical work in this context.

This report first covers the basic separation algorithms, followed by subtask 2 and subtask 3. After that, some conclusions and administrative issues are reviewed.

2. BLIND SIGNAL SEPARATION BY INDEPENDENT COMPONENT ANALYSIS

Independent Component Analysis (ICA) (see the textbook recently co-authored by one of the partners in this project [1]) is an increasingly popular statistical signal processing technique. The goal of ICA is to express a set of \( m \) observed signals \( x_1(t), \ldots, x_m(t) \) at time \( t \) as linear combinations of \( n \) unknown but statistically independent components \( s_1(t), \ldots, s_n(t) \) called usually sources or source signals. The ICA problem is blind, because not only the source signals but also the mixing coefficients are unknown.

Introducing the data vector \( x(t) = [x_1(t), \ldots, x_m(t)]^T \) and the source vector \( s(t) = [s_1(t), \ldots, s_n(t)]^T \), the instantaneous noisy linear ICA mixture model is given by

\[
x(t) = As(t) + n(t)
\]

Here the \( m \times n \) unknown but constant mixing matrix \( A \) contains the mixing coefficients, and \( n(t) \) denotes the additive noise vector at time \( t \). We make the standard assumptions that \( A \) has full column rank, and that \( n \leq m \), meaning that there are at most as many source signals \( s_j(t) \) as mixtures \( x_i(t) \) [1].

The source signals \( s(t) \) are estimated using only the observations \( x(t) \) by finding an \( n \times m \) unmixing matrix \( W \) such that the \( n \)-vector \( Wx(t) \) recovers the original sources as well as possible. Because of the blindness of the problem, only the waveforms of the sources can be estimated. For estimating the unmixing (separating) matrix \( W \), many different methods have been proposed; see [1] for an extensive discussion.

In most ICA methods, the data are first pre-whitened spatially, because this makes the subsequent separation task easier [1]. In whitening, the observed mixtures \( r(t) \) are transformed linearly so that their components become uncorrelated and have unit variance:

\[
y(t) = Tr(t), \quad E[y(t)y(t)^H] = I
\]

Here \( y(t) \) is the whitened data vector, \( T \) a whitening transformation matrix, \( I \) the unit matrix, and \( H \) denotes complex conjugation and transposition. Whitening is often carried out via principal component analysis (PCA), which yields for complex-valued data the transformation matrix

\[
T = \Lambda_s^{-\frac{1}{2}} U_s^H
\]

There the matrices \( \Lambda_s \) and \( U_s \) respectively contain the eigenvalues and eigenvectors of the autocorrelation matrix \( E[r(t)r(t)^H] \) of the received data vectors \( r(t) \). When PCA is used for whitening, it is easy to reduce the dimensionality of the data vectors simultaneously if desired by using only the principal eigenvectors in \( \Lambda_s \) and \( U_s \); see [1] for details.
In our experiments, we mostly applied the complex-valued version [2, 1, 16] of the so-called FastICA algorithm to the whitened mixtures $y(t)$ for separating the sources. The core of this algorithm is updating of the $i$th column $w_i$ of the orthogonal separating matrix $W$ according to [2, 1]

$$w_i^+ = E\{y(t)|w_i^H y(t)\}^* |w_i^H y(t)|^2 - \gamma w_i,$$ (4)

where $w_i^+$ is the updated value of $w_i$, and * denotes complex conjugation. The constant $\gamma$ is 2 for complex-valued signals, and 3 for real ones. In practice, the expectation $E$ in (4) is replaced by computing the respective average over the available set of whitened data vectors $y(t)$. In addition to (4), the columns of $W$ must be orthonormalized after each step. This can be carried out for example via Gram-Schmidt orthogonalization. FastICA and its different variants are discussed thoroughly in [1]. Instead of FastICA, other ICA algorithms introduced for complex-valued mixtures could be used.

### 3. BLIND MIMO EQUALIZATION AND ANTENNA ARRAY PROCESSING

In order to improve effective data rates and consequently effective spectral efficiency, blind receiver structures have been developed. The term blindness means that the receiver has no knowledge of either the transmitted sequence or the channel impulse response. Channel identification, equalization or demodulation is then performed using only some statistical or structural properties of the transmitted and received signal. Training data can then be either completely excluded or significantly reduced, and information symbols transmitted instead. In the face of deep fades, blind methods may allow for tracking fast variations in the channel and re-acquiring operational conditions using information symbols only. The processing in blind receivers is typically nonlinear. Common design goals for blind receiver algorithms are the following: capability to identify any type of channel; fast convergence to the desired solution, capability of tracking channel time-variations, and low computational complexity.

Systems employing multiple transmitters and receivers may be modeled as Multiple-Input Multiple-Output (MIMO) systems. Such systems provide major improvement in spectral efficiency (bits/s/Hz) and link quality by exploiting the diversity (multiple independent channels between the transmitter and receiver) and array (SNR) gains. Radio spectrum is a scarce and expensive resource. The potentially high capacity of MIMO systems may be achieved only if the channels are reliably estimated. Furthermore, effective spectral efficiency in these systems is reduced by large amount of training data. Hence, there is a strong motivation to develop blind and semi-blind receivers for MIMO systems, too.

In this project, adaptive decision feedback equalizers (DFE) and subspace methods for equalizing FIR-MIMO channels were developed. These methods are briefly described in the following. Publications [29]–[37] give detailed derivations as well as performance analysis of the proposed methods.

#### 3.1. Subspace method for blind identification of MIMO systems

In blind MIMO system identification, the input-output relationship may be rewritten in the form of a low rank model if sufficient number of received data samples is available. The fact that the column space of the received data matrix and column space of the channel matrix span the same subspace may then be exploited in determining the parameters of FIR-MIMO model. A well-known noise subspace method can be used to determine the channel matrix up to an ambiguity matrix. It is a constant full rank $K \times K$ matrix, where $K$ is the number of transmitted sources.

The remaining ambiguity matrix models an I-MIMO system with non-gaussian communication signals. The system may be solved using blind source separation (BSS) if the sources can be assumed to be statistically independent. This is a reasonable assumption if signals originate from different users.

#### 3.1.1. Signal model

We assume the standard FIR-MIMO baseband signal model with $K$ transmitters and $P$ receivers, in which the received signal $x(n)$ having $P$ components arranged as a column vector is of the form

$$x(n) = \sum_{k=0}^{L} H_k s(n-k) + v(n).$$ (5)

Here $s(n) = (s_1(n), s_2(n), \ldots, s_K(n))^T$ is a $K$-dimensional signal vector $(P > K)$, $L$ is the channel order. $\{H_k\}_{k=0,\ldots,L}$ are the the unknown $P \times K$ matrix-valued impulse response coefficients associated with the transfer function $H(z) = \sum_{k=0}^{L} H_k z^{-k}$, and $v(n)$ is noise. In this work we use the notation $H = [H_0^T, H_1^T, \ldots, H_L^T]^T$.

We assume that

$$\text{rank}(H(z)) = K \quad \text{for each } z$$ \hspace{1cm} (6)

$$\text{rank}(H(L)) \quad \text{is of full column rank}.$$ \hspace{1cm} (7)

These assumptions are needed to ensure the channel identifiability by using only second order statistics of the received signal vector $x(n)$. By stacking $N+1$ observations of (5) into an $(N+1)P \times 1$ vector $X(n) = [x^T(n), x^T(n-1), \ldots, x^T(n-N)]^T$ we may write

$$X(n) = H_N(H)S(n) + V(n).$$ (8)
Here, $S(n) = [s(n)^T, s(n-1)^T, \ldots, s(n-N-L)^T]^T$, $V(n) = [v(n)^T, v(n-1)^T, \ldots, v(n-N)^T]^T$ and $\mathcal{H}_N(H)$ is the $(N+1)P \times K(L+N+1)$ channel matrix (Sylvester matrix) given by

$$
\mathcal{H}_N(H) = 
\begin{bmatrix}
H_0 & H_1 & \cdots & H_L & 0 & \cdots & 0 \\
0 & H_0 & H_1 & \cdots & H_L & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\
0 & 0 & \cdots & 0 & H_0 & H_1 & \cdots & H_L
\end{bmatrix}
$$

Assumptions (6) and (7) imply that the channel matrix is of full column rank when the channel parameter $N$ is chosen sufficiently large.

3.1.2. Channel identification

The channel may be identified using a subspace method later extended to MIMO systems. Assume that $X(n)$ given in (8) is a wide-sense stationary process and the signal $S(n)$ and noise $V(n)$ are mutually independent and of zero mean. The covariance matrix of $X(n)$ is

$$E\{X(n)X(n)^H\} = \Sigma = \mathcal{H}_N(H)\Sigma_s\mathcal{H}_N(H)^H + \Sigma_v,$$

where $\Sigma_s = E\{S(n)S(n)^H\}$ is the signal covariance matrix of full rank, and $\Sigma_v = E\{V(n)V(n)^H\}$ is the noise covariance matrix of the form $\Sigma_v = 1^2I$, where $1^2$ is the noise power. The maximum channel order $L$ is assumed to be known.

Denote the noise subspace eigenvectors by $g_i$, $i = 1, \ldots, r$. The orthogonality of the signal and noise subspaces

$$\mathcal{H}_N^H(H)g_i = 0, \ i = 1, \ldots, r,$$

allows for identification of the channel matrix $H$ up to a right multiplication of an invertible $K \times K$ ambiguity matrix.

We may partition the noise subspace eigenvectors as

$$g_i = [g_{0}^{(i)}^T, g_{1}^{(i)}^T, \ldots, g_{N}^{(i)}^T]^T, \quad (9)$$

where $g_{k}^{(i)}$, $k = 0, 1, \ldots, N$ are of size $P \times 1$. Define

$$G_i = 
\begin{bmatrix}
g_{0}^{(i)} & g_{1}^{(i)} & \cdots & g_{N}^{(i)} & 0 & \cdots & 0 \\
0 & g_{0}^{(i)} & g_{1}^{(i)} & \cdots & g_{N}^{(i)} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\
0 & 0 & \cdots & 0 & g_{0}^{(i)} & g_{1}^{(i)} & \cdots & g_{N}^{(i)}
\end{bmatrix} \quad (10)$$

It can be shown that for each column of the matrix $H$, $h_i$, $i = 1, \ldots, K$,

$$h_i^H \left( \sum_{i=1}^{r} G_i G_i^H \right) h_i = 0.$$

Moreover, the dimension of the null space of the matrix

$$C = \sum_{i=1}^{r} G_i G_i^H$$

is $K$. This implies that

$$H = BR^{-1}$$

where $B$ is a $P(L+1) \times K$ matrix of the eigenvectors of $C$ corresponding to the noise subspace eigenvalues, and $R$ is an invertible $K \times K$ matrix. We may now determine the channel matrix up to a right multiplication of an invertible $K \times K$ ambiguity matrix, i.e.,

$$B = HR^{-1} \quad (11)$$

3.1.3. Equalization using subspace method and BSS

Assume that we know the channel matrix up to the ambiguity matrix $R$, i.e. we know $\tilde{B} = HR$. Then

$$\mathcal{H}_N(B) = \mathcal{H}_N(H)\tilde{R},$$

where $\tilde{R}$ is a $K(L+N+1) \times K(L+N+1)$ block diagonal matrix

$$\tilde{R} = 
\begin{bmatrix}
R & 0 & \cdots & 0 \\
0 & R & \cdots & 0 \\
\vdots & \vdots & \ddots & 0 \\
0 & 0 & \cdots & R
\end{bmatrix}.$$ 

Consider the noise-free case of the signal model (8)

$$X(n) = \mathcal{H}_N(H)S(n).$$

Let $\mathcal{H}_N(B)^H$ be the pseudo-inverse of the matrix $\mathcal{H}_N(B)$, i.e.

$$\mathcal{H}_N(B)^H = (\mathcal{H}_N(B)B^H)\mathcal{H}_N(B)^H.$$ 

Now it is easy to see that

$$\mathcal{H}_N(B)^H X(n) = Y(n) = \tilde{R}^{-1}S(n)$$

or

$$S(n) = \tilde{R}\mathcal{H}_N(B)^H X(n).$$

Note that

$$\tilde{R}^{-1} = 
\begin{bmatrix}
R^{-1} & 0 & \cdots & 0 \\
0 & R^{-1} & \cdots & 0 \\
\vdots & \vdots & \ddots & 0 \\
0 & 0 & \cdots & R^{-1}
\end{bmatrix}$$

and consider, for example, $K$ first components of $Y(n)$. Denote these components by $y(n)$. Then

$$y(n) = R^{-1}S(n).$$
This is in a form of I-MIMO model. If the K source signals are statistically independent and non-Gaussian, we may estimate the signals s(n) up to a permutation and complex scaling, i.e.
\[ \hat{s}(n) = D P s(n), \]
where D is a complex-valued diagonal matrix and P is a real-valued permutation matrix. The separation task may be performed any BSS algorithm, for example EASI. The remaining rotation and scaling of the constellation pattern may be resolved by using differential coding and the finite alphabet property of the communication signals.

### 3.2. DFE equalizer for MIMO systems

In this subproject we derive an adaptive equalizer for MIMO system with independent data streams (spatial multiplexing). Decision Feedback Equalizer (DFE) algorithm is capable of identifying, tracking and equalizing time-varying channels (TVCs). The algorithm is derived using state-variable model. The advantages of the estimation and tracking stages can be summarized as follows: the estimator is akin to the conventional Kalman filter (KF), it is thus an exact solution to the estimation problem. The channel estimation stage does not require additional knowledge of state-space model parameters in order to estimate the state. Consequently, the noise estimation problem need to be addressed. The method presented here for estimating the noise statistics relies on covariances of measurement and observation noises. A recursive formula is derived and the quality of the noise statistics is assessed by performing a non-parametric whiteness test on the innovation sequence of the KF. Whiteness of innovations is necessary for optimal performance of KF-based channel estimators.

The main contribution of the project is a novel MIMO DFE which belongs to the family of non-connected (NC) MIMO DFE’s. This implies that individual DFE’s are applied on the received signals. This type of DFE is used in combination with the Kalman filter in order to achieve equalization of time-varying channels. The DFE structure is derived using the MMSE criterion. The MIMO DFE cost function takes into account the cross channels, hence the equalizer is able to cancel both the Inter User Interference (IUI) and the Intersymbol Interference (ISI). The main assumption is that the input signals and the noise are uncorrelated. Combining the Kalman filtering and the MIMO DFE we get a true real-time algorithm in the sense that it is recursive in time and the storage space needed to evaluate the estimates remains constant, as time progresses and the amount of received data increases. Consequently, the method can cope with a large number of parameters.

In our derivation, assuming that the input and noise processes are uncorrelated we compute pairs of feedforward-feedback filters for the received signal at each antenna based on the MMSE criterion. Let us start by defining the channel convolution matrices \( \hat{H}_{ij} \) of dimension \( N_{ch} \times N_f \), where \( N_{ch} = L + N_f - 1 \) and \( N_f \) is the FF filter length.

The DFE equalizer has \( N_f \) FF coefficients and \( N_b \) FB coefficients. This will be written as DFE(\( N_f, N_b \)). Applying the FF filter to the past \( N_f \) observed received signals and the FB filter to the past \( N_d \) estimated symbols for each output we get the soft estimate:
\[ z_j(k) = \sum_{q=1}^{N_f} f_{jq} y_j(k - q) - \sum_{q=1}^{N_d} d_{jq} \hat{x}_j(k - q). \quad (13) \]

Equalization is achieved via feedforward \( f_j = (f_{j1}, \ldots, f_{jN_f})^T \), and feedback \( d_j = (d_{j1}, \ldots, d_{jN_d})^T \) filters. These filters are obtained by minimizing the following cost functions with respect to \( f_j \) and \( d_j \):
\[ J_f = E\{(x_j - z_j)^2\}, \quad (14) \]
where \( \delta \) is the equalization delay. In equation (14) we assume that past decisions are correct. Moreover, the input sequences are assumed to be uncorrelated with each other or with the noise.

For a \( m \times n \) MIMO system we obtain:
\[ \begin{align*}
\tau_j &= \|\delta\| H_j^T H_j + \|\delta\| H_j^T P_{DFE} H_j \tau_j = \lambda I \tau_j + \mu \delta, \\
\alpha_j &= M_j \tau_j, \quad (15)
\end{align*} \]
where \( M_j \) is given by 
\[ M_j = \begin{pmatrix} I_{N_d \times N_d} & 0_{N_d \times N_{ch} - N_d} & 0 \end{pmatrix}, \]
and \( P_{DFE} = (I - M_j^T M_j) \). Furthermore \( \lambda = \frac{\sigma^2}{\sigma^2} \) and \( \delta \) is the standard basis vector, with one at the position \( \delta \), \( 0 \leq \delta \leq N_f \). The derivation of the above two equations is presented in the Appendix. We note that even that the DFE’s are non-connected, each of the feedforward filters is taking into account the cross channels of the MIMO model, hence, when performing equalization is also able to cancel the inter-user interference. The noise statistics needed by the FF filters is estimated using the noise estimation stage.

Finally, the symbol estimate \( \hat{x}_j \) at time \( k \) is obtained by:
\[ \hat{x}_i(k) = \arg \min_{\alpha \in \chi} |\alpha - z_i(k)|, \quad (16) \]
where \( \chi \) is a finite alphabet.
4. BLIND SOURCE SEPARATION FOR SPREAD SPECTRUM RECEIVERS

4.1. Interference and jammer mitigation

In this subtask, we concentrated on interference and jammer mitigation techniques [4, 10, 12]. They have been studied actively in spread spectrum (SS) communications, because they improve the overall system performance and capacity without the need for a wider spectrum. In commercial cellular SS and DS-CDMA systems, many types of interferences can appear, starting from multiuser interference inside each sector in a cell to inter-operator interference. Unintentional jamming can also be present due to co-existing systems at the same band, whereas intentional jamming arises mainly in military applications.

Interference rejection/cancellation has been considered one of the most attractive class of suboptimal solutions, and has been studied extensively in the past [12]–[23]. Parallel and successive interference cancellation (PIC and SIC, respectively) are the main categories within this class, describing the procedure by which the interference is subtracted from the original data, after regenerating the interference from the tentatively estimated data. This procedure can, naturally, be repeated many times resulting in multi-stage interference cancellation schemes. Needless to say, any interference subtractive receiver performs the better the more accurately tentative decisions are being made. This is because the interference level is then reduced the most.

Considering antenna arrays, jamming can be mitigated by utilizing spatial diversity. However, when using conventional array receivers, directions of arrivals of signals must be first estimated. This in turn requires exact prior knowledge of the positions of the receiving antenna sensors. Blind techniques [5, 8, 9, 1] relax this stringent requirement, making it possible to achieve performance gains when applied to uncalibrated arrays in which the positions of the sensors are known only roughly or not at all. Most blind techniques are based on the assumption that the original source signals are statistically independent of each other. This assumption is quite realistic here, because the jammer signal originates from a different physical source than the information bearing signal.

In this project, we extended the framework proposed in [5] in three respects. Firstly, we apply BSS techniques based on independent component analysis (ICA) [1] instead of temporal correlations. ICA takes into account also higher-order statistics by forcing statistical independence of the separated signals. This enables jammer suppression for different types of jammer signals under realistic conditions. Secondly, we propose a hybrid receiver structure in which the ICA-based pre-processing is activated only when it is expected to improve the performance of the whole receiver chain. Finally, we propose different switching criteria between ICA and RAKE branches in the receiver chain and evaluate their effects to the performance of the receiver.

Regarding single antenna reception, another application of BSS is proposed as well in the project. This is the utilization of blind source separation (BSS) techniques in interference subtractive receivers. It is shown how the parametric form of the mixing matrix can efficiently be used to refine the ICA solution, and hence avoid interference enhancement while subtracting that source from the original data. The BSS-SIC-type receiver is also developed for highly loaded systems. Recall that a major drawback for standard BSS is the case where the number of source signals (to be blindly extracted from the received data) is greater than the number of observations made. This is a commonplace situation in communications applications in which cases standard BSS model doesn’t hold anymore. The key finding in this problem was the ability of a BSS-SIC-type receiver structure to somewhat circumvent the “more sources than observations”-problem in the sense that adequate performance (in terms of bit-error probability) is still achievable even in extremely highly loaded system, whereas conventional parallel and successive interference cancellation only remain at a moderate level.

4.2. Signal models

A standard spread spectrum system [19] with direct sequence spreading is assumed. Without loss of generality, we consider here a downlink channel (for example base-to-mobile). Thus the data describing the received block of $M$ symbols is of the form [11, 20]

$$r(t) = \sum_{m=1}^{M} \sum_{k=1}^{K} a_{km} b_{km} s_k(t - mT - d) + n(t)$$

(17)

where the symbols $b_{km}$ are sent to $K$ users via a channel characterized by a complex path gain $a_{km}$ and a path delay $d$. The delay $d$ is discrete, $d \in \{0,\ldots,(C-1)/2\}$, and remains constant for every block of $M$ data symbols. Furthermore, $s_k(\cdot)$ is $k$th user’s binary chip sequence, supported by $[0, T)$, where $T$ is the symbol duration, and $n(t)$ is Gaussian noise [20, 19].

The received signal $r(t)$ in (17) is jammed by a signal $j(t)$, which has the form

$$j(t) = \delta_p(t) \sqrt{J} e^{i(2\pi f_s (t + \phi)}$$

(18)

where $i = \sqrt{-1}$. The quantity $\delta_p(t) = 1$ with a probability $p$ during a symbol. Jamming corresponds to a continuous wave when $p = 1$ and pulsed wave at the symbol level otherwise. These two cases are used as examples of narrow-band and wide-band jamming\(^1\). The power, frequency, and

---

\(^1\) Also chip-pulsed jamming could have been considered as an example of wide-band jamming.
The DS-CDMA signal model (24) is readily a linear noisy ICA model (cf. eq. (1)) with \( m = 2C \) observations of \( n = 3K \) source components.

### 4.3. Receiver structures

#### 4.3.1. Suppression of jamming signals using ICA and RAKE

In our jamming problem, RAKE uses prior knowledge on the desired user’s code, and it can be applied also when a jammer signal is present. However, RAKE does not exploit independence of the information bearing signal from the jamming signal in any way. On the other hand, ICA relies on the strong but realistic independence assumption, but it is a blind technique which does not take into account the desired user’s code and the structure of the array steering matrix (22). In this work, we combine ICA and RAKE sequentially when it is reasonable for utilizing better the available prior information on the considered jamming problem.

The basic ICA-RAKE receiver proposed for jammer signal suppression consists of the three blocks (Preprocessing, ICA & Selection, and RAKE Detection); see reference [18]. The received vectors \( r(t) \) are first prewhitened, and then the FastICA algorithm (4) is applied to the prewhitened vectors \( y(t) \) as explained in the previous section. Due to the inherent indeterminacies of ICA [1], two additional tasks must be performed in the basic ICA-RAKE receiver. These are selection of the estimate \( \hat{r}(t) \) of the desired information bearing source \( r(t) \) among the separated two sources, and estimation of the channel. Notice that in our problem ICA is usually used for separating two sources only, which then provide estimates of the components of the vector \( z(t) \) in (21). Recall that the components of the vector \( z(t) \) are the information bearing signal \( r(t) \) and down-converted jammer signal \( j(t) \).

ICA can estimate the source signals only up to a permutation, and a complex scaling factor can be exchanged between a source and the respective column of the mixing matrix [1]. For getting rid of this ambiguity, a short preamble sequence is used, consisting of \( N_p \) training symbols \( s^d_0(t) \) \( (t = 1, 2, \ldots, N_p) \) of the desired user \( d \). The preamble is used for both conventional matched filter based channel estimation [19], and for identifying the desired information bearing source among the sources estimated by ICA. More specifically, this takes place by matching the preamble sequence \( s^d_0(t) \) \( (t = 1, 2, \ldots, N_p) \) of the desired user with the corresponding portions \( s^d_p(t) \) \( (t = 1, 2, \ldots, N_p) \) of the symbols estimated using RAKE for all the separated sources:

\[
I_d = \arg \max_i \left\{ \sum_{t=1}^{N_p} \text{dist}_H[s^d_0(t), s^d_p(t)] \right\}, \quad i = 1, \ldots, N_S
\]

Here \( I_d \) is the selected source index, \( N_S \) is the number of separated sources after whitening and ICA filtering, and

Phase of the jammer signal \( j(t) \) are denoted respectively by \( J, f_j, \) and \( \phi \). The phase is assumed to be uniformly distributed over the interval \([0, 2\pi)\).

Denoting the received antenna data \( r_1(t), \ldots, r_N(t) \) we can represent the array data more concisely in vector form as

\[
r(t) = \Theta z(t) + n(t) \tag{19}
\]

where the vector

\[
r(t) = [r_1(t) r_2(t) \cdots r_N(t)]^T \tag{20}
\]

contains the signals received at the \( N \) array elements at time \( t \),

\[
z(t) = [z(t) j(t) e^{-i2\pi f_j t}]^T \tag{21}
\]

is a two-component vector having as its elements the information signal \( r(t) \) and down-converted jammer signal \( j(t) \) at time \( t \), and the array steering matrix

\[
\Theta = \begin{bmatrix}
1 & 1 \\
e^{i\theta_1} & e^{i\theta_2} \\
\vdots & \vdots \\
e^{i(N-1)\theta_1} & e^{i(N-1)\theta_2}
\end{bmatrix} \tag{22}
\]

The \( N \)-vector \( n(t) \) is similar in form to (20), containing additive white Gaussian noise (AWGN) terms \( n_i(t) \) at each antenna element \( i = 1, 2, \ldots, N \). A comparison of the array signal model (19) with the ICA mixing model (1) shows immediately that (19) is actually a noisy mixing model with a mixing matrix \( \Theta \) and source vector \( z(t) \). Hence ICA or other BSS techniques can be applied to separation of the information signal and jammer signal. Estimates of these signals are obtained as the components of the vector \( \hat{w}(t) \), but their order and scaling is arbitrary. In case of a single antenna reception the data model can be represented as [27]

\[
r_m \overset{\text{def}}{=} \sum_{k=1}^{K} a_k (b_{k,m-1}g_k + b_{km} + b_{k,m+1}g_k) + n_m \tag{23}
\]

where a processing window size of two symbols is assumed. Here \( n_m \) denotes noise vector and the \( 2C \)-length code vectors are delayed (according to the path delay) copies of the signature sequences.

With a simple manipulation, we can get a compact representation for the data,

\[
r_m = G b_m + n_m. \tag{24}
\]

The \( 2C \times 3K \) dimensional code matrix \( G \) contains the code vectors and path strengths, while the \( 3K \)-vector \( b_m \) contains the symbols:

\[
G \overset{\text{def}}{=} \begin{bmatrix}
\cdots a_k g_k \\
\cdots a_k g_k \\
\cdots a_k g_k \\
\cdots b_{k,m-1}b_{km}b_{k,m+1} \cdots
\end{bmatrix} \tag{25}
\]

\[
b_m \overset{\text{def}}{=} \begin{bmatrix}
\cdots b_{k,m-1}b_{km}b_{k,m+1} \cdots
\end{bmatrix}
\]

The DS-CDMA signal model (24) is readily a linear noisy ICA model (cf. eq. (1)) with \( m = 2C \) observations of \( n = 3K \) source components.
\[
\sum_{t=1}^{N_p} \text{dist}_H[s_d^p(t), \tilde{s}_d^p(t)] \text{ computes the Hamming distance between the preamble sequence } s_d^p(t) (t = 1, 2, \ldots, N_p) \text{ and the corresponding estimated symbols } \tilde{s}_d^p(t) \text{ for the } i^{th} \text{ source.}
\]

Finally, conventional detection is performed for the data of the selected source \(I_d\). This leads to the basic ICA-RAKE receiver structure, where the well-known and robust RAKE method [13, 19] is used for conventional detection. We could have tried instead of RAKE other methods developed for multi-user detection [20], but in this work the primary interest was to study the capability of ICA to mitigate jammer signals prior to the actual detection.

However, the basic ICA-RAKE receiver described above as well as the BSS method proposed in [5] are still somewhat impractical as such. This is because they employ an ICA or BSS block even though it may not always be desirable. In fact, if the jammer signal is weak or even absent, the additional ICA or BSS jammer suppression block might even cause additional interference to the information bearing signal. Therefore, we improve the basic ICA-RAKE receiver by introducing two different switching strategies, pre- and post-switching schemes, which select either ICA-assisted or conventional RAKE detection in the array receiver chain.

The training symbols in the preamble can be used also to determine whether the additional jammer suppression by ICA is desirable or not. In general, this decision can be made either before or after ICA has been applied. In the pre-switching scheme, the switch decides whether there is a need to separate the jammer signal by ICA prior to conventional detection, or should conventional detection alone be performed. This decision is based on the performance of the RAKE detector on the preamble sequence \(s_d^p(t) (t = 1, 2, \ldots, N_p)\). RAKE alone is applied if sufficiently many symbols in the preamble are detected correctly:

\[
N_p = \max_{i=1, \ldots, N_r} \left\{ \sum_{t=1}^{N_p} \text{dist}_H[s_d^p(t), \tilde{s}_d^p(t)] \right\} < \delta_s N_p \quad (27)
\]

where \(\delta_s\) is the chosen threshold value for switching. In the post-switching scheme, both ICA-RAKE and RAKE branches are first active. ICA-RAKE branch provides tentative estimate \(\tilde{s}_d^p(t)\) for the \(t^{th}\) symbol \(s_d^p(t)\) of the desired user, and RAKE branch respectively tentative symbol estimate \(\tilde{s}_r^p(t)\). The final output \(\hat{s}_d(t)\) of the receiver is selected from the branch which provides better correlation with the training symbols:

\[
\max_{l, R} \left\{ \text{E}[\tilde{s}_d^{l,p}(t) s_d^p(t)], \text{E}[\tilde{s}_d^{R,p}(t) s_d^p(t)] \right\} \quad (28)
\]

That is, both branches first provide their own estimates \(\tilde{s}_d^{l,p}(t)\) and \(\tilde{s}_d^{R,p}(t)\) corresponding to the symbols \(s_d^p(t)\) belonging to the preamble, and the branch providing better symbol estimates is then selected and used for the final hard decision.

### 4.3.2. BSS/ICA Based Successive Interference Cancellation

By an interference subtractive receiver we loosely speaking mean an iterative multi-user receiver, where the estimated interference is subtracted from the received signal prior to the estimation of a particular user. The principle of this kind of receiver utilizing BSS/ICA is quite straightforward and is here only shortly revisited. Namely, the received signal is first separated by ICA. After separation, users are identified by their spreading codes. Hence, the receiver is semi-blind. Strictly speaking, for the user identification a correlation

\[
\rho(k, k') = \frac{|c_k(\hat{\tau}_k)w_{k'}^H|}{\|c_k(\hat{\tau}_k)||w_{k'}||} \quad (29)
\]

is computed for each \(k, k'\), where \(w_{k'}\) corresponds to the ICA basis vector of \(k^{th}\) source. Next, a threshold, \(\rho_1\), is set to indicate proper detection. This is needed because subtraction of erroneously detected signal would actually enhance interference. Thus, only the signal corresponding to the user \(k\) with \(\rho(k, k') > \rho_1\) for some \(k'\) is subtracted from the receiver signal. After subtraction of all such users' signals, the next ICA separation is performed for the interference-subtracted signal. The order of ICA model decreases in subtraction, which help ICA to recover the remaining users. The procedure is repeated successively until all users are detected. Furthermore, one can easily show, that phase of complex number \(\rho(k, k')\) in 29 equals to phase shift produced by ICA. This makes it possible to correct ICA's phase ambiguity.

It is of primary importance to see that what is actually subtracted from the original data is a tentative decision of the form \(a_k c_k(\hat{\tau}_k)\tilde{b}_k\) rather than \(w_k\tilde{b}_k\). Hence the ICA solution is first refined (according to the knowledge of the parametric form of the mixing matrix) before subtraction is performed.

### 4.3.3. Joint delay tracking and interference cancellation using BSS/ICA

The receiver structure described above is now developed to cope with inaccuracies in delay estimation. Recall that ICA performs purely in blind manner. The first occasion where some a prior knowledge of the users is needed is the user identification phase (29). Naturally, the timing information can also be used to generate a good initial value for ICA iterations, and hence speed up the separation [27]. Anyway, what lousy timing estimate ultimately does is that it worsen the user identification. More importantly, given that a user is nevertheless identified having an erroneous timing estimate, the subtraction of that user enhance interference the more the bigger was the timing inaccuracy. To avoid that situation the delay of each identified user could first be re-
4.4. Experiments

4.4.1. Array reception

An example of the performance evaluation is given in the following. A simulated system with $K = 8$ users was considered, spread with short Gold Codes of length $C = 31$ [13, 19]. The length of the data block was $M = 200$ QPSK symbols. The monitored user was chosen randomly, and the signal-to-noise ratio (SNR) and the signal-to-jammer ratio (SJR) were defined with respect to this desired user. The probability that a bit is jammed was chosen to be $p = 0.5$. The number of elements $N_a$ is 2. Pre-switching involved activation of the ICA-RAKE branch only if more than 10% of the training symbols were erroneously estimated by RAKE. Post-switching involved switching at the decision device, where either the soft decision outputs of the ICA-RAKE branch or the RAKE branch were decoded, based on the errors with respect to the training sequence.

In the continuous wave setting, the SJR varied against a constant SNR of 10 dB. Using a pulsed jammer which had a frequency offset resulted in the bit and block error rates shown in Figs. 1 and 2. Pre-switching is equivalent to MRC in this case, while post-switching results in a gain of 1 dB in the $-15$ to $5$ dB region, as seen from the BER curves. In fact the BLER curves show that post-switching outperforms pre-switching by at least 2 dB when the target rate is $10^{-1}$, and it provides better performance in the region of high jamming. Increasing the number of antenna elements yielded results which are comparable to the two-antenna case.

Computationally, the proposed pre-switching and post-switching schemes are more demanding than the simple RAKE receiver. Pre-switching requires three times more computation than standard RAKE, and post-switching is computationally clearly more demanding. In general, pre-switching is suitable for real-time applications, while post-switching is ideal if real-time operation is not required.

4.4.2. Single antenna reception

As an example the performance evaluation for the single antenna interference canceller, consider $K$ users transmitting data blocks of $M = 5000$ QPSK symbols. Symbols are spread using Gold codes of length $C = 31$. Two service classes are assumed, which is modelled as a power difference of 10 dB between the two user groups. Inside both groups all the users are assigned the same power. The length of the all the receivers is $2C$ and hence both the LMMSE-PIC and LMMSE detectors are truncated to that length. Recall that the optimal length for asynchronous data would be $MC$ (the whole block of symbols) which is not a sensible choice for the receiver length in practise [20].

The results demonstrate that ICA-assisted methods maintain their competence also with great numbers of users. Especially the successive ICA-receiver outperforms reference methods, SIC and PIC, clearly. This can bee seen in Fig. 3 The figure depicts BERs of the ICA-assisted methods as well as the reference methods as a function of $K$ (number of users).

Finally, we also notice from the experiments that the FastICA with successive interference cancellation performs clearly better than astICA alone, given that the delays are tracked, too, see Fig. 4.

5. RESULTS AND IMPACTS

We believe that scientificaly, the goals of the project have been attained. A total number of 28 articles have appeared, of which 3 are journal papers. The ideas have been positively received in many conferences which the partners have attended. Also industry representatives have shown interest.
to ideas like inference cancellation. These discussions are continuing.

In this consortium, each of the three partners used the finances essentially for the salary of one graduate student whose Ph.D. projects were done in the subtasks. One of the students, Mihai Enescu, already got his doctoral degree at the HUT / SPL in 2002 with the Thesis "Adaptive Methods for Blind Equalization and Signal Separation in MIMO Systems". Another one, Karthikesh Raju at HUT / CIS started his graduate studies in this project in 2001 and is presently finishing his doctoral thesis. His thesis defense will be in winter 2005.

On the administrative side, the leaders of the three projects have met regularly both at HUT and in appropriate conferences. The students have attended some winter / summer schools of the graduate schools like GETA.

6. REFERENCES


Concurrent Design and Fabrication of Integrated Module Boards

Jorma Kivilahti  
Professor

Lab. of Electronics Production Technology  
Helsinki University of Technology  
P.O. Box 3000, FIN-02015 HUT, Finland  
Jorma.Kivilahti@hut.fi

Ilkka Suni  
Research Manager

Microelectronics  
VTT Technical Research Centre of Finland  
P.O.Box 1208 FIN-02044 VTT Finland  
Ilkka.Suni@vtt.fi

ABSTRACT

The project had two major objectives. The first objective of the research project was to interconnect reliably bare Cu-metallised chips and integrated passive components into multilayer flexible and rigid substrates with the IMB technology. This required comprehensive design and modelling of electrical, thermal and mechanical performance as well as the interfacial compatibility - all at the chip and the module board levels. Consequently, the second objective was to develop a new concurrent design and manufacturing approach for producing reliable solderless electronic products by combining electrical, thermal and mechanical design and simulation tools and aspects into a one coherent approach. Both objectives were achieved in the project.

I. INTRODUCTION

Increasing employment of portable electronics will require the most advanced materials and manufacturing technologies; higher performance and reliability at lower cost are becoming ever more crucial. However, while striving for higher performance we will encounter also more fundamental manufacturing and reliability challenges, which require concurrent designing and simulation of electrical, thermal, and mechanical behaviour of multilayer structures used in future portable electronics. This concerns, in particular, very high-density interconnection and packaging technologies like the IMB technology being originally developed at Helsinki University of Technology [1]. It enables solderless integration of embedded IC’s and passives. Even though our approach focuses on IMB technology, but it is by no means limited to it.

Increasing (I/O) densities with finer feature sizes on chip level are unavoidably related to thinner metallisations, so there is a growing risk that the metallisations of the ULSI technology are influenced detrimentally by high current densities (electromigration) or by excessive chemical reactions between adjacent materials or by moisture and oxygen from the operational environment (corrosion). Similarly, the solder volumes are continuously decreasing on board level (PWBs), and therefore electromigration as well as thermomigration in solder joints is becoming a more serious concern [Fig. 1]. In addition to higher current densities the homologous temperatures of solder interconnections are low, and therefore relatively fast diffusion via vacancies occurs even at ambient temperatures. Besides, in small solder
joints brittle intermetallic reaction products take already a marked volume fraction of the joints in high density bare chip assemblies (e.g. area array Flip Chip, FC) diminishing their reliability. Due to this small volume effect (SVE), the electromigration and increasing stresses experienced by ever smaller solder interconnections the reliability at the component and board level is expected to become a greater concern than on a chip level.

In order to manufacture reliably and cost-effectively highly functional integrated module structures, realistic design and simulation methods will become indispensable. Electrical, thermal, mechanical and even chemical properties of such multimaterial structures must be known thoroughly - ultimately due to the fact that dissimilar materials of diminutive amounts are used in contact with each other. It is most essential that design and simulation tools are used collaboratively by developing so-called Concurrent Design Approach. Therefore, the development of such an approach was selected as one of the main targets of the project.

II. CONCURRENT DESIGN AND SIMULATION

In order to develop the above-mentioned concurrent modelling approach all the subsections were first considered separately and after that combined together as follows.

Electrical design and modelling

The IMB or an equivalent technology requires new procedures for designing and manufacturing of ultra-high density boards or modules. A test board for Power Amplifier (PA) working at 2.45 GHz was designed. Simulations were realized for specific applications, geometries and positions on the board. Using CAM board designer together with HP-HFSS (High Speed Structure Simulator) for component modelling and simulation and HP ADS (Advanced Design Systems) for electrical parameters calculations, the board design and analyses was completed. Input and output network and digital control for the PA were designed and modelled using the IMB embedded passive components. The design was done also for surface mount technology (SMT) based board. Both designs are shown in Figs.2 and 3. Then the control of characteristic impedance and input-output matching circuits were analysed to ensure good signal integrity.

Besides the PA test board electrical modelling of IMB interconnections were modelled and calculated for Ku and K band. A comparison with the flip chip assembly was then carried out. The electrical parameters of the interconnections were extracted and electrical performances analysed. Furthermore, considering the signal integrity in high-speed digital applications the electrical parameters of the embedded resistors were analysed. The concurrent electrical and thermal analysis was then performed for the best component geometry – by considering also the manufacturing capabilities.

Fig. 2. IMB board design

Fig. 3. SMT board design.

Thermal design and modelling

The thermal energy produced by different components and its effect on the components environment must be taken into account by initiating the thermal modelling already at the beginning of the electrical design of a product.

The thermal modelling for both SMT and IMB test board designs was carried out with the CFD (Computational Fluid Dynamics) technique in FLOTHERM (Fig. 4). The parameters obtained in the electrical design were used as input data for the calculation of thermal power of different components. Details of the board layout in the vicinity of the chip, including location of vias and other passive components, are also obtained from the electrical design. The natural convection of surrounding air is simulated and the contribution of radiation is included. The final results of both structures (IMB and SMT) are compared with each other in the terms of chip temperature.
In addition, the test board with an active component, thermal modelling of embedded resistors used in IMB applications are also performed by using ANSYS (Fig. 5). Different parameters of resistors including resistance, sheet resistivity and line width were examined and the optimal configuration was determined by combining the results of both the thermal and electrical modelling. In order to enable concurrent multiphysical modelling, considerable amount of work was needed [with MATLAB program] for converting the results in FLOTHERM into the format, which the finite element analysis program ANSYS could use so that more subtle thermal modelling in test board level and subsequent mechanical modelling become possible.

**Mechanical modelling**

As the electrical solder interconnections are the only mechanical connection between SMT components and the board their reliability is very important - in addition to their electrical performance. By knowing the temperature distribution of the assembly and its changes during operation the strains and stresses induced by the thermal expansion of the dissimilar materials in the assembly can be calculated with the Finite Element Method (FEM).

The temperature distributions calculated by FLOTHERM and ANSYS (Figs. 4 and 5) were input to the mechanical FE models of the studied structures and the stress and strain (Figs. 6 and 7) distributions were calculated with ANSYS.
Fig. 7. Strain distribution in IMB assembly induced by thermal power shown in Figs. 4 and 5.

The interconnections where the stresses are highest are most prone to failure, but the time to the failure depends on the whole loading history and the interconnection geometry and, naturally, on the materials used. Therefore it is very difficult task to predict the time to failure, and more work is needed to solve this very important issue. However, the calculated stress distributions provide means to analyse the critical parts in the designed layout.

Physicochemical modelling

In addition to the above-mentioned modelling approaches the physicochemical compatibility evaluation is important part of the over-all modelling of integrated component assemblies or modules. This is due to the fact that intermaterial reactions during the fabrication as well as in the use of the modules may generate reaction products, which tend to reduce the cohesion between thin material layers or even deteriorate the functions of the components. Especially the adhesion between various new interfaces between materials used in the IMB structure was found out to be crucial. Thermodynamic calculations were carried out in order to find out the stabilities of the different multimaterial interfaces found in the module structures. In addition, surface energy measurements and calculations were executed to understand the effect of different surface treatment techniques to the adhesion between different polymers and metals in IMB structures.

III. EXPERIMENTAL

The first objective of the research project - to interconnect reliably bare copper metallised chips with multilayer flexible and rigid substrates with integrated passive components utilising the IMB technology - required extensive experimental work to be carried out. The development of different component manufacturing technologies – both for Si-chips and for PWB substrates - was conducted individually as well as jointly by HUT and VTT. The investigation of diffusion barriers for Cu metallisation as carried out in the previous Telectronics-project “Integration of Microcircuits with Multilayer Substrates Using Advanced Thin-Film Processing” was also continued. In addition to the above mentioned activities the adhesion studies between Cu and polymers used in the IMB technology were incorporated into the present project. This was done owing to the observation that the adhesion between dissimilar material layers used in the modules was becoming one of the major reliability concerns.

Investigation and development of components for IMB modules was based on extensive materials research utilising scanning (SEM) and transmission electron microscopy (TEM), Rutherford backscattering (RBS), atomic force microscopy (AFM) and x-ray diffraction (XRD). Although the focus in developing components was in utilizing them in the IMB modules, the components can be adapted universally also to other thin film integration technologies.

The development of the concurrent modelling and simulation tool also required a lot of experimental work. After designing and modelling the structures were fabricated and analysed in order to verify experimentally the models used. Further, because of the compliance problems between the different modelling tools, additional programming was required to convert the data from one program to another.

IV. RESULTS AND IMPACTS

The objectives of the project were attained very well. The project gave new insight and significant fundamental results on the interactions between different material layers used in the IMB modules. Furthermore, the project produced valuable practical results concerning the fabrication of integrated passives and their utilisation in the module board technology and in other analogous technologies. Most important results of the project are considered to be the new knowledge and experience achieved about the concurrent design and modelling of the electronic circuits. In this project we were able to develop a simulation protocol that included electrical, thermal, mechanical and physicochemical tools and to combine them into one unified approach. Based solely on the results and experience achieved in this project a new design and simulation project funded by the National Technology Agency and Finnish Electronics Industry has been initiated in the beginning of this year.
The first objective of the present work was to interconnect reliably bare copper metallised chips with multilayer flexible and rigid substrates with integrated passive components utilising the IMB technology. This in turn required the investigation of various interfaces found inside the module structure. The research on diffusion barriers carried out in the Teletronic I program was continued also herein. The results verified that Ta-based barriers offer a very feasible solution to the diffusion barrier problem. The crucial effect of oxygen on the reactions in all the investigated Ta-based metalization schemes was demonstrated and the thermodynamic basis for understanding the origins of this behaviour was given.

As the research project advanced it became more and more clear that the adhesion of Cu to different polymers used in the IMB structures was one of the key issues in ensuring reliability of the module board structures. Therefore, extensive research was carried out on the ways to improve the adhesion in the module structures. Thus, the work performed in this project has supported markedly also the development of the IMB technology, which is being implemented into volume production by the Imbera Electronics established jointly by the Elcoteq Networks and Aspocomp Group. Furthermore, as it is expected that in near future photonics is becoming increasingly important field of scientific and technological activity, and because IMB-type technologies can be utilised in the fabrication of optical modules, some considerations concerning the optical properties and optical applications of the modules were taken. This in turn led to the attempts to improve some of the properties of epoxy-based polymers used in the IMB-process for optical interconnections.

Full range of components is needed in the IMB modules, and therefore new processes using copper metallisation were investigated. For example, we developed a process to fabricate high dielectric constant tantalum pentoxide thin films for integrated capacitors (Fig. 8).

The films obtained showed a stoichiometric orthorhombic \( \beta-Ta_2O_5 \) phase at an \( O_2-Ar \) sputtering gas flow ratio of 20 %. The processed thin film tantalum pentoxide MIM (Metal Insulator Metal) capacitors with copper electrodes showed good potential for further improvement. Also thin film resistors and diffusion barriers have been developed, fabricated and investigated.

The test structures shown in Figs. 2 and 3 were fabricated by utilizing both the IMB and SMT technologies. The corresponding structures already shown in Figs. 9 and 10. Some problems were encountered during the manufacturing of the IMB structure and it was not possible to finalize the assembly in time. This was partly owing to the reasons already discussed above, i.e. adhesion problems and related subjects. In addition, the fabrication process itself revealed some reliability concerns. However, the fabrication is still continuing and will be completed in April 2004.

Fig. 9. X-ray picture from the fabricated IMB structure showing the place of the chip, one integrated resistor and two integrated inductors.

Fig. 10. SMT board after assembly. The chip can be seen in the middle and is surrounded by passive components.
The second object of the research project was to develop new concurrent approach to manufacture reliable solderless electronic products by combining electrical, thermal and mechanical design and simulation tools and aspects into a one coherent approach. To achieve this goal took much more time that was initially expected mainly due to the compliance problems between different simulation programs.

Therefore, a program code had to be developed in the project to enable the transfer of simulation data from one program to another. After that the full simulation loop i.e. electrical, thermal and mechanical was finalised. The results from the concurrent simulations indicated that apparently, for the system fabricated with IMB technology, the stresses produced were much higher than those of SMT, especially in the Cu conductor layer and in the die itself. This can degrade the system performance of the circuit fabricated with the IMB process. However, for passive component on top of board, the stresses in the IMB assembly are smaller than those of SMT, which means that these passive components, such as capacitors, can be expected to show better reliability in the IMB case. The source of this behaviour can be seen from Fig. 2 a) and b) where distribution of thermal energy is shown. It is obvious that in SMT case the chip is not as hot as in the IMB case. On the contrary the surface mount passives are hotter in SMT than in IMB assembly. The results from the above simulations were also combined with physicochemical considerations to find out how different materials in the modelled structure would behave.

In order to verify the models used two sets of test structures were fabricated. The one was an IMB-module board and the other was the SMT-based assembly. The performance of the test circuits was then compared with each other with the modelling and with experimental tests. The results of the project are currently being utilized also in a new European Lead-Free Network (ELFNET) project. Without any doubt the research in this field will continue in the future and is expected to become even more intensive.

V. PUBLICATIONS


A. PEER-REVIEWS PAPERS IN THIS PROJECT


B. CONFERENCE PAPERS


Advanced Radio Channel Identification and Estimation (ARCHIE)

Timo Laakso (Project leader)
Visa Koivunen
Signal Processing Laboratory
PO Box 3000, 02015 HUT
timo.laakso@hut.fi, visa.koivunen@hut.fi
Aarne Mämmelä
VTT Electronics
P.O. Box 1100, 90571 Oulu
Aarne.Mammela@vtt.fi

ABSTRACT

This report reviews the research work of the ARCHIE project done at the Helsinki University of Technology and VTT Electronics. The goal of the project was to develop advanced channel identification and equalization methods for future mobile communication systems and wireless local area networks, and their future extensions.

The work can be divided into three parts, namely system design, signal design and receiver design. One important result was a novel classification of receiver algorithms. In addition, several original papers on iterative equalization and decoding algorithms were published in best journals and conferences.

I. INTRODUCTION

This paper includes a summary of the research work done in the ARCHIE project. The goal of the project was to develop advanced channel identification and equalization methods for future mobile communication systems and wireless local area networks, and their future extensions.

A general block diagram of a modern system based on serially concatenated codes and iterative processing is shown in Fig. 1. The research covers mainly the digital baseband processing algorithms of the physical layer of the Open Systems Interconnection (OSI) architecture model.

I. ACTIVITIES

The work can be divided into three parts, namely system design, signal design and receiver design. Each of these parts will be described below.
The optimal receiver is defined to be a maximum a posteriori probability (MAP) detector where the channel is either assumed to be perfectly known or, if statistically known, be removed by averaging. The corresponding optimal receivers are the correlator and estimator-correlator. On the other hand, in channel estimation data are assumed to be known or removed by averaging. Finally, joint estimators, which are generally optimal data detectors only asymptotically, can be defined.

All the estimators are initially defined to be of the block-type, but they can be approximated with recursive versions. The estimators are based on some statistical knowledge on the channel, for example on the first- or second-order statistics of the impulse response and noise. Adaptive estimators do not necessarily rely on this a priori knowledge, but they are approximations of the optimal estimators and explicitly or implicitly estimate also the statistics. The adaptive algorithms are separately defined for unknown slowly and fast fading channels.

b) Signal design

Two papers were published on signal design. Optimal receivers cannot be defined unless the transmitted signal and the statistics of the channel are first defined.

In [3] we analyzed the effect of time and frequency domain windowing, or weighting, of the transmitted signal for least squares (LS) and minimum mean-square error (MMSE) estimators when the channel is time-variant. We considered the estimation error for different windows and we found that the windows can be selected independently. We explained the performance of the estimators with the help of the radar ambiguity function of the transmitted signal including windowing.

Pilot symbol assisted modulation and an MMSE channel predictor were used to employ feedback MMSE power control over a frequency nonselective slow Rayleigh fading channel [4]. Lag error was noticed to cause severe performance degradation, even when the channel is very slowly fading. In order to decrease the lag error, the number of estimator coefficients was found to become quite large.

c) Receiver design

Several papers were published on receiver design. The basic principle in these studies was to derive receiver algorithms, which would be nearly optimal in performance with reduced complexity compared to the optimal solutions.

In [5, 6, 7], channel estimation based on the expectation-maximisation (EM) algorithm was considered. Then, adaptive Bayesian and expectation maximisation (EM) based soft decision directed (SDD) channel estimators were derived and studied in [6, 8, 9, 10]. They approximate the maximum a posteriori probability (MAP) symbol detection in frequency-selective fading channels and they are applied to iterative turbo-processing receivers in a computationally attractive way [8, 11, 12]. Usually the second order statistics, i.e., the Doppler power spectrum and more generally the scattering function of the channel, are assumed to be known in the receiver. The EM-based estimators can be used to estimate the autoregressive moving average (ARMA) parameters of the spectrum [6].

In [11, 12], variational inference and estimation approach was used to derive low-complexity turbo receivers for single-input single-output and multiple-input multiple-output (MIMO) systems. Simulation results showed that the presented receivers achieve practically the same performance as the optimal MAP turbo receivers, while providing significant computational savings.

The application of the Berrou soft output Viterbi algorithm (SOVA) to implement an iterative turbo decoder was considered in [13]. The Berrou SOVA decoder is conceptually different from the conventional turbo decoders. The performance was found to be competitive with the traditional algorithms although the complexity is relatively low.

Joint channel and delay estimation in multiantenna CDMA receivers

In uplink (mobile-to-base station) transmission of DS-CDMA networks, the users signals are arriving to the base-station (BS) with different delays. Therefore, the system is asynchronous. The channel impulse response (CIR) of each user is different also. The asynchronous system has the advantage that the need for network signaling is reduced. On the other hand, the propagation delays of each user have to be estimated.
by the BS receiver. Hence the quality of the delay estimation has a significant impact on the overall receiver performance. This work has produced a good number of publications and doctoral thesis of Marius Sirbu.

In long-code system, the delay spread of the channel introduces inter-chip-interference (ICI) and the channels are always time varying. Multiple antennas are employed at the receiver. Consequently the performance is improved by taking advantage of the spatial diversity as well as the increased SNR.

In this work we developed novel delay and channel estimation methods [15-23]. The CIR and the propagation delays are estimated jointly. The method lends itself to an adaptive implementations since adaptive update rule may be used. The method is robust and the performance is significantly improved over existing methods. Moreover the design is simple since no user-specified threshold values are needed and the performance remains reliable even for non-minimum phase channels. The method is called the delay profile method and it determines the delay by looking for a certain pattern in the matrix containing both the CIR and delay information. Low-order AR-model is used to model the channel dynamics. In order to study symbol error rate performance, a MMSE multi-user equalizer is found based on the estimated channel matrix. The simulation results demonstrate the reliable performance of at different SNR levels, in the presence of near--far effects and in different user scenarios. The method is flexible in a sense that it may be used both in short-code and long-code DS-CDMA systems.

These methods are applicable to uplink CDMA systems such as UMTS as well as satellite navigation systems such as GPS and Galileo. Significant improvement in navigation accuracy may be achieved.

Adaptive Filtering Algorithms

Development of efficient adaptive algorithms for channel estimation and equalization is a key part of modern digital receiver design. New receiver structures based on the concept of Set Membership Filtering (update estimate only when needed) and data reuse (utilize same input several times to speed up convergence) were developed [24-33]. Part of this work was done in cooperation with prof. Paulo Diniz’ group in the Federal University of Rio de Janeiro, Brazil.

Data Selective Adaptive Filtering Algorithms

This research investigated new data selective adaptive filtering algorithms using the framework of set-membership filtering (SMF). These algorithms combine a bounded error specification on the adaptive filter with the concept of data reusing. The resulting algorithms have low average computational complexity because coefficient update is not performed each iteration. The adaptation algorithms can be adjusted to achieve a desired computational complexity by allowing a variable number of data-reuses for the filter update.

Analysis of Partial-Update Adaptive Filters

This research analyzed partial-update normalized adaptive filters. Partial-update adaptive filtering is a technique suitable for applications where the order of the adaptive filter is so high that it may impair even the implementation of low computational complexity algorithms, such as the NLMS algorithm. Partial-update adaptive filters reduce the algorithm complexity by properly decreasing the number of filter coefficients that is updated each iteration so that the filter order may be kept fixed. Order statistics are used to analyze the mean-squared error of the adaptive filter output.

IV. RESULTS AND IMPACTS

Two Dr.Tech. students (Stefan Werner and Marius Sirbu) got their degrees within the project.

Some contract research and development projects were started during the project with the industry. In the European WINNER project we are taking part in the systems engineering workpackage, which was possible due to our systematic work. WINNER includes some 40 partners, mainly from the European Union and it is led by the industrial partners, for example Siemens and Nokia. We have given several postgraduate courses at the Helsinki University of Technology and University of Oulu. A number of postgraduate students from the industry have taken part in the lectures. In addition, three one-day short courses on adaptive receiver algorithms have been given to the industry.

PUBLICATIONS


Development of the second generation ΔΣ frequency synthesis techniques
J. Kostamovaara
Department of Electrical and Information Engineering, Electronics Laboratory
Electronics Laboratory, PO.Box 4500, FI-90014 UNIVERSITY OF OULU, FINLAND
juha.kostamovaara@ee.oulu.fi

ABSTRACT

This project has produced knowledge on the improvement of the performance of ΔΣ frequency synthesizers with emphasis on the development of speed-up techniques and reduction of the in-band phase noise and output spur level resulting in the development of a ΔΣ synthesizer chip with state-of-the art performance. In addition, new synthesizer architectures have been developed and studied. Also a new sampler with embedded programmable FIR-filtering function has been developed.

I. INTRODUCTION

Frequency synthesizers are a component of wireless integrated circuits that face constantly increasing requirements for low noise and acquisition speed as new standards evolve. Recently, this has caused ΔΣ frequency synthesizers to be chosen as a solution for local oscillators in IC design. The benefits offered by ΔΣ frequency synthesizers result from the fact that they reduce the divide ratio N between the VCO and the phase detector. This reduction in N increases the comparison frequency in the synthesizer, allowing lower phase noise and higher loop bandwidths. Higher loop bandwidths, in turn, allow faster acquisition time and more suppression of noise contributed by the VCO.

Despite the fact that these synthesizers have been studied extensively, many of the detailed problems were not understood. Academic and commercial researchers have both faced apparently mysterious problems when trying to build ΔΣ frequency synthesizers. The aim of this project has been to shed light on these problems, especially on the problems of high in-band noise and spur level found typically in published ΔΣ modulators, and therefore to enable one to develop synthesizers with better performance than earlier available. This project has provided a course and several papers that address these issues. A very good ΔΣ synthesizer was developed to validate this work.

Despite the increased loop bandwidth offered by ΔΣ frequency synthesizers, it is often insufficient for some applications. One of the goals and achievements of this project has been to show how to significantly improve acquisition times for both integer-N and fractional-N synthesizers.

In addition, also totally new ΔΣ architectures were identified and suggested as described below. The other architecture relies on an RF/IF sampler, one of the components in the original proposal, which actually was seen in the course of the project to become more useful as a radio receiver component. As a result, its research was focused more on this purpose than functioning as a synthesizer component.

II. ACTIVITIES AND RESULTS

a) Speed-up techniques for frequency synthesis

Traditional PLL based synthesizers suffer from unwanted reference frequency related spurs around the synthesized carrier frequency. This is the case even with the nowadays popular fractional-N techniques. Since the loop acts as a low-pass filter for the spurs, the simple way to diminish the level of these harmful signal components is to reduce the loop bandwidth. However, as the bandwidth of the loop is reduced, the loop is made slower in the process. This is not acceptable in systems (e.g. GSM), where telecommunication standards pose limits to the time it takes for the system to switch channels. Thus, in traditional PLL based synthesizers, a trade-off must be made between channel switching speed and the purity of the output spectrum.

A new speed-up method has been proposed by us, which allows separate optimization of loop bandwidth and switching speed [1]. The so-called two-pulses method is based on using charge pulses to control the dynamics of the loop. The method can readily be applied to the ΔΣ fractional-N designs popular nowadays, as was proven in [2]. The additional circuitry is also well suited for integration since it only consists of programmable current sources.

Principle of operation of the two-pulses speed-up method

In the proposed method, the voltage at the input of the VCO is controlled using out-of-the-loop circuitry which affects the charge stored in the loop filter during the frequency hop (see Fig. 1). The response time can be very short even when the loop bandwidth is small, if the
current waveforms \(i_1(t)\) and \(i_2(t)\) have the shape shown in Fig. 1. The current sources in Fig. 1 are only connected to the loop for a short period of time after a frequency hop is initiated. Thus, they do not disturb the normal operation of the loop by adding noise or modifying the transfer function of the loop. Since the speed-up electronics is contained in a completely separate piece of circuitry, it does not necessitate changes in existing loop components such as the VCO, phase detector and dividers. Additionally, as the selected double-pulse waveform keeps the additional circuitry very simple, the method is also well suited for system level integration.

The theory presented in [1] was first derived for a mathematically simpler 2nd order PLL (i.e. \(CS_2\) and \(C_2\) not present in the loop). Fig. 2. (a) presents the phase error \(\phi(t)\) at the output of the phase detector after a 10 MHz frequency step at the output at \(t = 0\) s. The curve labelled ‘phase error without speed-up’ shows the unforced behaviour of the loop, which was intentionally made very slow to increase spurious suppression. As the height of the 1st pulse \((I_1)\) increases, the phase error response bends more and more rapidly downwards. This is exactly what is intended by the initial current pulse – a very rapid decay of phase error below zero. The increased rate of phase error decay is achieved by overdriving the output frequency of the loop as shown in Fig. 2. (a). Thus, the 1st current pulse increases the charge of \(C_1\) sufficiently to provide the VCO with an input voltage which produces a somewhat higher output frequency than the final one.

The 2nd pulse is applied at \(t = 2\) \(\mu s\), and its effect is also shown in Fig. 2. (a) The pulse has a double purpose. First, it aligns the output frequency exactly with the final value, thus zeroing the frequency error at \(t = 3\) \(\mu s\). Second, to avoid the generation of any frequency errors after this point, the phase error must also go to zero at exactly the same moment. Thus, by properly selecting the magnitudes of the 1st and 2nd pulse – \(I_1\) and \(I_2\), respectively – the phase and frequency error are both cancelled out completely at the end of the 2nd pulse, and any residual errors will be avoided as shown in Fig. 2. (a).

Figure 1 – Architecture of the loop and the selected speed-up current waveform.

Qualitatively, this means that the pulses force the areas inside the shaded boundaries in Fig. 2 (a) equal. As the sum of the areas (they have different signs) goes to zero, the phase error, being proportional to the integral of the frequency error, follows suit.

The system presented above is not directly applicable when \(C_2\) (shown in Fig. 1) is added to the loop filter, thus producing a 3rd order system typical of many practical RF synthesizers. However, by adding a second current source (\(CS_2\) in Fig. 1) to the architecture, delivering a current \(i_2(t) = i_1(t)/10\), and by selecting \(C_2 = C_1/10\), a common choice, the behaviour of the system will approximate that of a 2nd order system. This particular selection of \(i_2(t)\) ensures that the voltage across \(R_1\) will be very close to zero. Thus, \(i_1(t)\) will only flow to \(C_1\), and a proper voltage change will be produced at the input of the VCO. Fig. 2 (b) shows the resulting phase error and output frequency error responses for the theoretical 2nd order and the simulated 3rd order system, when the current levels derived from the 2nd order theory are directly used in both cases.

Figure 2 – (a) Phase error and output frequency error after a step change in input frequency and (b) phase error and output frequency error of theoretical 2nd order, simulated 3rd order and measured 3rd order loop.
Measurement results

A two-pulses speed-up system was built around an 800 MHz PLL ASIC designed in the project. Fig. 2 (b) shows the measured response of the example 3rd order loop. As expected, the measurements indicate that the presented ideas can be applied to real-life synthesizers, with an example reduction in frequency hop time to 1/7 of its original value. In practice, however, phase and frequency errors can be cancelled completely only by fine-tuning the pulse levels. As a result, the usability of the two-pulse method is limited by the sensitivity of the final phase and the frequency error at the end of the frequency hop period (in our case 10 μs) to the levels of the current pulses. This sensitivity will be a function of the loop parameters, and its maximum allowed level depends on system specifications. In our case, for example, when the 3rd order loop was used during the measurements, a frequency error of ±12 kHz was produced by a ±1% error in pulse level. The figure agrees with that predicted by 2nd order loop approximation (±17 kHz).

In the future the aim of this work is to further develop the two-pulses speed-up method so that it would automatically minimize the effects of changing loop parameters to the frequency hop speed, which was the main shortcoming of the 1st generation implementations. According to simulations, the adaptive method under development has the potential of significantly enhancing the usability of the two-pulses speed-up in real-life applications.

b) Effect of nonlinearities on the in-band noise of a ΔΣ synthesizer

In the field of ΔΣ synthesis much of the theoretical work carried out in the project was devoted to how various factors contribute to the Banerjee Figure of Merit (BFM) which is a good measure of synthesizer performance for in-band noise [3]. A theory has been developed how jitter in digital logic, charge-pump current noise, offset current noise and reference noise contribute to the in-band noise of the synthesizer. Moreover, it has been shown for the first time both analytically and experimentally how the nonlineairities of the phase detector in particular mix the high frequency quantization noise peculiar to ΔΣ frequency synthesizers into its in-band noise [4]. The analysis developed lets the designer to set specific requirements for all of the above noise sources when trying to achieve a specified in-band noise level. In addition, several circuit techniques have been developed and realized by which the above noise sources or their effects can be minimized. To serve as a base for further work, and to validate the theoretical understanding of ΔΣ synthesizers, a state of the art prototype was also developed.

A phase noise plot of the tested synthesizer chip is shown in Fig. 3 [4].

<table>
<thead>
<tr>
<th>Ref</th>
<th>Tech.</th>
<th>Fout (GHz)</th>
<th>Fref (MHz)</th>
<th>Resol. (Hz)</th>
<th>In-band Noise (dBc/Hz)</th>
<th>BFM (dB)</th>
<th>Max. Power (mW)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[14]</td>
<td>0.5 μm CMOS</td>
<td>900</td>
<td>7.99</td>
<td>8</td>
<td>82</td>
<td>-202</td>
<td>43</td>
</tr>
<tr>
<td>[15]</td>
<td>0.35 μm BiCMOS</td>
<td>2.47</td>
<td>0.8</td>
<td>10</td>
<td>82</td>
<td>-201</td>
<td>16</td>
</tr>
<tr>
<td>[16]</td>
<td>0.25 μm CMOS</td>
<td>1.8</td>
<td>26</td>
<td>400</td>
<td>80</td>
<td>-171</td>
<td>70</td>
</tr>
<tr>
<td>[17]</td>
<td>0.6 μm CMOS</td>
<td>1.84</td>
<td>20</td>
<td>-75</td>
<td>-187</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[18]</td>
<td>0.35 μm BiCMOS</td>
<td>1.76</td>
<td>13</td>
<td>&lt;5000</td>
<td>-79</td>
<td>-193</td>
<td>22.6</td>
</tr>
<tr>
<td>[19]</td>
<td>0.5 μm CMOS</td>
<td>1.715</td>
<td>20</td>
<td>10</td>
<td>-90</td>
<td>-202</td>
<td>140</td>
</tr>
<tr>
<td>[20]</td>
<td>0.5 μm SiGe BiCMOS</td>
<td>2.4</td>
<td>48</td>
<td>50</td>
<td>-100</td>
<td>-211</td>
<td>135</td>
</tr>
</tbody>
</table>

The developed synthesizer chip clearly demonstrates the power of the developed circuit techniques. For example, its performance with regard to the in-band noise is almost equivalent in integer-N or fractional-N operation modes. The main source for the elevated in-band noise in the ΔΣ fractional synthesizer is the nonlinear mixing of the quantization noise into the signal band. The circuit techniques developed in the project to minimize this effect were shown to be very effective. The most important of the developed circuit techniques is the transfer of the operation point of the phase detector into its linear range by an additional offset current in the charge pump and the minimization of the active range of the ΔΣ jitter coming out of the divider by post-filtering techniques. As expected, turning off this offset in integer-N mode had no impact on in-band phase noise. However, if the offset was disabled in fractional-N mode, the in-band noise increased by as much as 20 dB, which clearly demonstrates the

![Phase noise plot of the tested synthesizer.](image-url)
detrimental effect of the non-linear mixing on the in-band noise of the synthesizer.

c) Avoiding spurs by periodical behavioural analysis

The $\Delta\Sigma$ modulator is one of the key units of the fractional-N synthesizer. The synthesized frequency is a product of a stable reference frequency and a fractional number provided by the $\Delta\Sigma$ modulator. However, it is not just a plain, static number. In fact the modulator generates a sophisticated control signal with the mean corresponding to the desired fractional number. The spectral purity of the $\Delta\Sigma$ modulator generated signal directly affects the spectral purity of the synthesized channel. Therefore, first and foremost the modulator generated signal should be properly shaped and free from spurious tones.

Short limit cycles are identified as the most severe reason for the spurious tones appearing in a modulator generated signal. A common method for randomizing modulator signal and breaking the limit cycles is known as dither. The drawbacks for using dither are extra hardware, additional noise introduced to the system and possible problems with modulator stability. Some designers use intuitive and experimental methods for randomizing modulator behaviour by inserting “seed values” into modulator registers [5]. $\Delta\Sigma$ modulator is a nonlinear system and any rigorous analysis is very difficult. Existing exact solutions are too narrow and too sophisticated to be practically applicable. Many aspects of $\Delta\Sigma$ modulator operation are still not well understood.

In this work the problem of limit cycles has been eliminated for two common modulator architectures used in fractional-N frequency synthesis. The approach is based on the original observation that the modulator sequence length can be controlled for all channels by applying predefined initial conditions and modulator scaling. This observation gave rise to a practical method for modulator design and sequence length prediction [6]. Long sequences guarantee that the quantization noise does not get concentrated on only a few dominant spurious tones. At the same time the modulator behaves in a fully predictable way. This allows collecting full information about modulator spectrum for all channels. The worst case performance for all channels can be presented in the form of spectrum envelope, see Fig. 4. The envelope show the spurious-free range in a most reliable way. This allows collecting full information about modulator spectrum for all channels. The worst case performance for all channels can be presented in the form of spectrum envelope, see Fig. 4. The envelope show the spurious-free range in a most reliable way. As shown in Fig. 4, the SFR can be controlled with respect to the modulator sequence length.

The developed design method allows a quick modulator design for the fractional-N synthesizer. Modulator SFDR is easily determined and reliable. A trade-off between the SFDR requirement and modulator resolution allows a planned and reasonable use of hardware resources.

d) New $\Delta\Sigma$ architectures

Quantization noise reduction by decreased phase step

The other research topic in the field of $\Delta\Sigma$ frequency synthesis has been the development of a new $\Delta\Sigma$ synthesizer architecture where phase quantization noise is being reduced by utilizing a fractional-N divider instead of the integer-N divider typically used in $\Delta\Sigma$ synthesizers [7]. The proposed architecture can also be considered a modification of fractional-N synthesis based on a fractional-N divider with the added feature that the control of the fractional-N divider is based on $\Delta\Sigma$ techniques.

The use of the fractional-N divider realized with a digitally controlled delay line, for example, in connection with $\Delta\Sigma$ techniques decreases the phase step and quantization noise, respectively, proportionally to the number of delay elements used in the line. With 16 elements, for example, the phase quantization step size is $2\pi/16$, giving a 24 dB reduction in quantization noise.

As a result the project introduced a new synthesizer architecture shown in Fig. 5. This architecture combines other fractional-N techniques with $\Delta\Sigma$ techniques to remove the spurs introduced by the other fractional-N techniques. This new architecture significantly reduces the quantization noise by employing $\Delta\Sigma$ techniques on a fraction of a VCO cycle rather than on a whole VCO cycle. On the other hand, if the proposed architecture is considered a modification of the fractional-N divider based synthesis, the advantages of the method are higher resolution and reduction of the level of spurs caused by delay error in the fractional divider.

In the published system the VCO cycle fraction was intended to be produced by a delay line but recent results suggest even higher performance when unit element charge pump structures are being used to achieve the same effect. The reason for the improved performance comes from the fact that in the latter realization the mismatch error between the unit element charge pumps will be noise-shaped, whereas in the delay element
realization it is only randomized. These results have, however, not been published yet.

**Figure 5** – New ΔΣ architecture.

**FD Based Synthesizer**

Yet another research topic in the field of ΔΣ frequency synthesis has been the digitization of the whole PLL loop (expect the VCO) by measuring the VCO frequency with a frequency discriminator [8]. One advantage of the digitization is the flexibility achieved in the design of the loop filter, for example. This is useful for rapid acquisition and for directly producing FSK waveforms through the synthesizer without up-conversion. Another important advantage is the possible reduction in the achievable phase noise level especially if the sampler that combines sampling filtering and mixing (see below) is used to down-mix the VCO output for the frequency-to-digital conversion. This is useful to produce synthesizers with on-chip loop filters even when charge pump currents are high in order to have low noise.

A frequency synthesizer architecture based on the above concept as shown in Fig. 6 has been studied in the project. The kernel of this architecture is the ΔΣ frequency discriminator, which converts frequency to a digitized bit stream. It enables the use of a digital loop filter, and offers great flexibility with digital signal processing.

**Figure 6** – Block Diagram of a Frequency Discriminator-based Synthesizer

A 3rd order MASH architecture frequency discriminator has been developed and investigated. An accurate linear model with phase noise analysis was also developed. Quantization noise degrading caused by the digital component delays is reported, which limits the use of this FD in RF frequency range [9]. Figure 7 shows the simulated quantization noise degradation in the 3rd order MASH frequency discriminator with a 880 MHz carrier, 10 kHz frequency modulation in the presence of 30 ps skew and 0.2 ns stage delay mismatch.

**Figure 7** – Quantization Noise Degradation in MASH Frequency Discriminator.

A specific feature of the ΔΣFD based synthesizer architecture suggested in this work is that a down-conversion mixer is introduced between the VCO and the ΔΣFD [9]. This mixer decreases the division ratio, thus decreasing the well-known 20logN in-band phase noise amplification. The mixer also decreases the input frequency of the frequency discriminator and enables the use of a higher-order frequency discriminator in the RF frequency synthesizer.

The synthesizer and discriminator are modelled in SIMULINK in time domain and frequency domain. Modelling in VHDL-AMS/Verilog-AMS is also going on.
This kind of modelling enables simultaneous high-level behavioural and more accurate circuit level simulations. A prototype synthesizer PCB with discrete components is also under construction. In the ASIC realization of the synthesizer the mixer will be realized based on the following sampling filter concept that has, however, also other potential applications and has therefore gained considerable attention in the project.

e) Sampling filters

The work in this particular field has concentrated on investigating sub-sampling (i.e. band-pass sampling) with integrated signal processing, such as partial channel selection and anti-aliasing filtering, as a potential alternative for signal conditioning and down-conversion in radio receivers. The use of a sub-sampling circuit in a receiver chain prior to baseband signal processing and A/D conversion can relieve the bandwidth constraints and minimize the power consumption of the baseband discrete-time circuit blocks by lowering their sampling frequency. However, the use of conventional simple sub-sampling circuits have usually been avoided due to their modest dynamic range, especially noise, performance resulting mainly from the lack of proper anti-aliasing filtering in front of the sampler.

Different from conventional voltage-mode sampling circuits, the sampler structure investigated in this project is based on integrative sampling of current, or charge, signal. Integrating current into a sampling capacitor within a determinate time window itself produces a continuous-time sin(x)/x-type low-pass response, which limits the noise (and signal) bandwidth of the sampler [10]. Integrating several charge samples into the sampling capacitor during the output sampling period extends the filtering properties of the sampler to discrete-time finite-impulse-response (FIR) filtering. Fig. 8 shows the principle of a general active-integrator-based charge sampling circuit with an embedded N-tap complex FIR filtering function.

In the sampler operation the front-end transconductance (Gm) cell’s output current is alternately integrated into the real (I) and imaginary (Q) channel sampling capacitors connected in negative feedback configuration. The integrated charge samples of both channels are weighted by coefficients with values +1, -1 and 0 by sets of switches controlled by I and Q channel clock signals. After the charge accumulation cycle the two quadrature outputs are sampled, or decimated, at the final output sampling rate Ts which fully determines the location of the sampler’s output signal spectrum. Thus down-conversion by sub-sampling can be combined into the operation of the sampler by decimating a band-pass input signal at a rate below its Nyquist frequency.

With an appropriate clock scheme, the sampler circuit of Fig. 8 can implement an arbitrary complex FIR filter function, the impulse response of which consists of coefficients restricted to the set {±1,0,±j}. The filter taps can be obtained e.g. by quantizing an arbitrary target impulse response by using ΔΣ modulation [11]. Due to its asymmetric response for positive and negative frequencies, a charge sampling circuit with an integrated complex FIR filter can be utilized in radio receivers for image rejection and quadrature down-conversion in addition to its partial channel selection and anti-aliasing filtering properties. Also, with a proper choice of circuit implementation, the proposed sampler can be configured to perform complex filtering and further quadrature down-conversion to baseband or to low-IF for complex (I/Q) IF signals.

Fig. 9 presents two simulated sample amplitude responses of complex charge sampling circuits with two different 3072-tap FIR filtering functions. The targeted IF input centre frequency of the designed samplers is f_c = 100 MHz, while the baseband output sampling frequency equals to f_s = 0.1042 MHz. The dashed line shows the response of the simplest possible complex band-pass sampler realization with all FIR filter coefficients having absolute value of unity. The solid line illustrates the amplitude response of a charge sampler with a more advanced FIR function utilizing ΔΣ-quantized tap coefficients. Fig. 9 shows that both filtering responses ideally provide almost similar suppression of negative frequencies (i.e. image rejection), but the wanted signal band selectivity of the sampler with the integrated ΔΣ FIR filter is superior to the simple sampler. The example sampler realization ideally provides over 40 dB of inherent anti-aliasing and out-of-band attenuation with a
modest increase in the circuit complexity. Note that due to the semi-analog-semi-digital filtering operation of the sampler, its filtering properties, such as bandwidth and centre frequency, can be modified in real time by changing the FIR filter coefficients and/or sampling frequency of the sampler. This kind of sampling filter is applicable to radio receiver structures simultaneously supporting several communication standards, and possibly eventually to fully software-defined radios.

Figure 9 – Examples of complex sampler responses.

During the project period two charge sampling circuit realizations have been designed and measured. According to the measurements, the implemented fully differential complex IF sampler with an embedded 192-tap complex FIR filter function in 0.35 \( \mu \)m CMOS has a stable dynamic range performance with 66 dB of spurious-free dynamic range (SFDR) up to 100 MHz operating frequency [10,12]. The other test circuit, fabricated in a 0.8 \( \mu \)m BiCMOS process, integrates a 192-tap complex FIR filter and a complex narrowband SC filter on the same chip providing over 44 dB of image rejection on its 26.3 kHz –3 dB bandwidth [13]. A summary of the measured results for both test circuits is given in the Table 2 below.

### Table 2. Performance of the realized sampling filters

<table>
<thead>
<tr>
<th>Technology</th>
<th>0.8 ( \mu )m BiCMOS</th>
<th>0.35 ( \mu )m CMOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 dB bandwidth</td>
<td>26.3 kHz</td>
<td>923 kHz</td>
</tr>
<tr>
<td>IIP3</td>
<td>–8 dBV (in-band)</td>
<td>+25 dBV (out-of-band)</td>
</tr>
<tr>
<td>SFDR</td>
<td>59 dB</td>
<td>66 dB</td>
</tr>
<tr>
<td>Image band rejection</td>
<td>&gt;44 dB</td>
<td>&gt;36 dB</td>
</tr>
<tr>
<td>Power consumption</td>
<td>85 mW @ 5 V</td>
<td>30 mW @ 3.3 V</td>
</tr>
</tbody>
</table>

The main technical results of the project are:

1) Development of a new speed-up method for integer- \( N \) and fractional-\( N \) type synthesizers including \( \Delta \Sigma \) synthesizers [1].

2) Analysis of the contribution of various synthesizer noise sources to the in-band noise of a \( \Delta \Sigma \) synthesizer, especially the analysis of how the nonlinearities of the phase detector in particular mix the high frequency quantization noise peculiar to \( \Delta \Sigma \) frequency synthesizers into its in-band noise. This work has resulted in the development of a state-of-the-art \( \Delta \Sigma \) synthesizer chip [4].

3) Development of a methodology to decrease the level of spurs in a \( \Delta \Sigma \) modulator to a negligible level [6].

4) Development of a new \( \Delta \Sigma \) modulator architecture that decreases the amount of quantization phase noise by reducing the phase quantization step size [7].

5) Suggestion for a new \( \Delta \Sigma \) synthesizer architecture that is based on frequency-to-digital conversion. An essential part of the suggested architecture is a mixer that avoids the 20log\( N \) noise problem faced with standard PLL type synthesizer architectures [9].

6) Development of a new sampling filter concept that originally was intended to be part of the above new \( \Delta \Sigma \) synthesizer architecture but which in the course of the project has been found to have importance as a new discrete time, programmable FIR filtering concept that has a lot of potential applications in filtering the stages of radio receivers, especially in multi-radio environment. The concept has been verified by realized and tested circuits [10,12-13].

All the above results have been published or submitted to be published in the leading journals of the field as well as in international conferences, only a few of which are referred to below. The main results have also been published as a post-graduate course held at the University of Oulu in the autumn of 2003. The course was open also for participants from industry. So far the project has resulted in one doctoral degree (Juha Hakkinen) and one Master of Science thesis (Maciej Borkowski) and 4 other doctoral theses are currently under preparation about the research subjects of the project. Two of them will be finalized during 2004-2005 (Sami Karvonen, Tom Riley). In general the project has made it possible for the research group involved with the project to start serious work in the field of frequency synthesis. That work will continue after this particular project, also with some totally new promising ideas not presented above.
REFERENCES


On-Chip Communication Architecture for HW/SW Co-Design

Jarmo Takala¹, Timo Hämäläinen¹, and Karen Egiazarian³

¹Institute of Digital and Computer Systems
³Institute of Signal Processing
Tampere University of Technology, P.O.B. 553, 33101 Tampere, Finland
jarmo.takala@tut.fi, timo.d.hamalainen@tut.fi, karen.eguiazarian@tut.fi

ABSTRACT

Currently design of system-on-chip (SoC) products is based on the premise that the communication structure between functional blocks can be described in very imprecise fashion. This can cause severe errors in predicting the performance of the system.

In this project, future SoC designs are considered; chip interconnection architectures, formal design methodology, and spectral testing methods are developed. The project is still on-going and ends at 2004.

I. INTRODUCTION

Embedded products are often built around specialized computational nodes and interconnection architectures. The shrinking size of transistors drives the electronic systems into a single integrated circuit or system-on-chip (SoC) solutions. To cope with the increasing complexity of systems, some design objects are prepared such that they are reusable. They are usually called as intellectual property (IP) blocks. The IP blocks can be shared within the organization or obtained from external sources. However, interfacing, testing, and verification of IP blocks from a variety of sources have proven to be difficult.

From the technology point of view, an efficient interconnection architecture plays a central role in successful continuous-media applications, where large amount of data is transferred back and forth in the system. Selection of the interconnection that supports the given application domain can be of crucial importance. General-purpose interconnections might not map the utilized algorithms adequately. Furthermore, power consumption of interconnections may be too high. Wide buses can exhibit crosstalk and other implementation problems due to the pressure to route long signal lines close to each other. In addition, to be reusable, the interconnection must be well standardized and readily expandable. The design process must not be forced to start from the beginning if changes occur along the interconnection.

Formal design methods would be ideal for describing the design from specification to the final implementation. To complete the SoC design flow, testing the circuits at various levels of abstraction is essential. It consists of verification of equality of the output produced by the circuit to the designed output. The main effort in testing is to minimize the number of test patterns, yet covering all the possible faults.

This project has been considering all these problems. The objective of the project was to develop on-chip interconnection architectures and methods for future SoC designs. The project considered three research topics: a) communication architecture, b) formal design methodology, and c) spectral testing methods. The results of the topics are discussed in the following sections.

II. COMMUNICATION ARCHITECTURE

The goal for this task was to design and implement modular, flexible, and efficient communication architecture for contemporary and future system-on-chip implementations that can be seamlessly integrated to a system design flow. In order to achieve this goal, the task has been divided in five subtasks that are analysis of topologies, specification and implementation of the new communication architecture, optimization methods and integration to a system design. Each of the subtasks is elaborated in the following.

a) Analysis

The key problems in on-chip communications were first identified with the help of evaluation of existing multiprocessor systems. Table 1. summarises the main issues of this evaluation. It was concluded that the topology should be flexible and provide standardised means to attach heterogeneous functional blocks to the system. The primary target for optimizations was chosen to be communication efficiency (energy and performance) as there are many conflicting requirements.

In the next phase, detailed analyses were conducted for existing interconnection schemes. Theoretical comparisons provide rough theoretical connectivity and complexity estimates in an application independent way. Bus, crossbar and mesh topologies with variations were chosen for further analysis.
The next analysis level was conducted with real synthesised and simulated architectures that give the complexity and performance in physical area and throughput measures. These architectures are based on a set of generic communication building blocks including IP-block wrappers, distributed and centralized arbitration units, and network segments. Generic blocks allow design of comparable, parameterized architectures and very reliable results.

The communication load was tested with different transfer patterns resembling potential system-on-chip applications. As a conclusion, a bus based and mesh based architectures were found most feasible. The former gives the best cost/performance ratio, but the latter is better in performance for intense localised data traffic.

### b) Specification

Analyses led to an interconnection approach that utilizes a hierarchy of bus structures. In this type of architecture, the local interconnections are bus segments making use of their known good properties. The global on-chip connections use a more complex network structure containing buffering bridges resembling the router/switch units of dynamical topologies. This type of heterogeneous architecture can be modified according to the requirements of the application, scaling from point-to-point links to an arbitrary topology.

To name key features, data transfers in the bus segment take place in a circuit switched manner (synchronous), but transfers across bridges are packet switched (asynchronous). The arbitration is distributed which saves dedicated point-to-point signals and greatly helps modularity. In addition, all signals are identical and shared between wrappers in the network side. This makes the wrapper layout independent of the number of IP-blocks and simplifies design reuse.
widths and FIFO depths can be set only before synthesis, but other parameters affecting arbitration and power saving can be configured on the fly.

Optimization of the interconnection focused on power saving. An experimental study was conducted that compared power saving methods with the interconnection wrapper. Basic mechanisms include shutting down parts of the wrapper through clock gating, slowing down the clock speed and scaling voltage as well as total shut down of a wrapper. For run time optimizations, fine-tuning of the time slot period (TDMA arbitration) using bus monitoring is carried out.

A low bit rate video encoder was used as a test application, and a test case implementation with several processors, memories and wrappers was designed. The experiments show that power saving of around 40% can be achieved when the operation of application and the components of the implementation are optimized. It should be noted that even better power saving is possible, but not without violating the real time requirements of the application.

e) Integration

The designed interconnection architecture offers convenient interfacing for system design flow at various abstraction levels. For architecture description, the interconnection wrappers and segments are given as library components. At the lowest level, the components are physical HW descriptions that can be synthesized and implemented in real technology. At the highest level, the interconnection is given as channels and bounded FIFO interfaces.

After HW/SW partitioning, the HW will be implemented as IP-blocks, SW as processes in processor blocks and communication between all blocks using wrappers, bridges and network segments. The decision of partitioning is carried out in the co-design phase. After this, the communication optimization takes place. In this step, communication related parameters are optimized based on the data transfer pattern present in the current application.

III. FORMAL DESIGN METHODOLOGY

The work in this task was based on DisCo language and toolset [9] developed at Institute of Software Systems at Tampere University of Technology. DisCo is a formal specification and design methodology for reactive and distributed systems. It consists of an incremental design method based on superposition, specification language, and a collection of supporting tools.

The objective of this task was to develop a design methodology for DisCo, which allows a tool supported path towards hardware implementation. The actual design work is performed with refinement steps; the design focus shifts along with the stages of the methodology from high-level refinement steps to code generation for third-party tools. The developed design flow is illustrated in Fig. 1.

The main result is the DisCo-VHDL Compiler, which generates hardware descriptions from intermediate representation generated by the DisCo tools.

![Figure 1 – Principal formal design flow.](image)

a) Specification

The system is first described with DisCo specification language, which is based on the concept of temporal logic of actions (TLA) proposed in [10]. TLA is a linear-time temporal logic with an expressive power that is suitable for the transformational design style where specifications are refined using correctness-preserving steps. TLA combines two logics: logic of actions for representing relations between pairs of states and temporal logic for reasoning on infinite sequences of states.

The initial DisCo specification of the system is a collective behaviour of objects. In this model, the system state is defined by objects participating in joint actions. The superposition steps are given in a layer, which includes also the initial specification. The specification can be validated with DisCo tools.

The developed methodology towards hardware implementation consists of partitioning, allocation, scheduling and mapping phases. These are described in the following.

b) Partitioning

The system is first partitioned in specification space, i.e., the functionality of the system is specified without any apparent link to implementation. This results in a layered design style where parallel layers of superposition are used to structure the system. When the partitioning in
specification space is completed, all the parallel layers have been composed into a single, flattened DisCo layer consisting of objects and actions.

The previous layer is used as the starting point for partitioning in implementation space. We use here object-oriented partitioning, i.e., objects are partitioned first and actions are considered later. First the joint actions need to be composed into elementary actions that can be assigned to components for execution. After this phase the objects are assigned to components.

c) Allocation

When proceeding towards the implementation, we need to allocate components to target architecture. If there are several subsystems, a communication channel is needed, which implies need for interface, thus we need to create the needed interface actions.

d) Scheduling

The next task is scheduling. DisCo supports liveness in terms of strong fairness assertions; neither generalized fairness assertions can be expressed nor weak fairness assertions are supported. Therefore, these fairness assertions must be maintained interactively by the designer or theorem proving environment. The scheduling of joint actions can be performed with the aid of atomicity refinement since the responsibilities of execution can be found out easily. The active participant is the one, which modifies its state, while the other participant only allows its variables to be read.

e) Mapping

The final task is to map the DisCo specifications to a hardware description language. In this work VHDL is used. In the mapping process, the following principles are used. DisCo components are mapped to VHDL entities. If the component has interface variables, the entity has an output port. Vice versa, if the component depends on an interface variable, it has an input port. Objects are mapped onto synchronous VHDL processes. Various variables and states are naturally mapped to VHDL signals. In a similar fashion, we map the guards of instances of actions to signals. Relations, in turn, are realized with the aid of Boolean valued registers. Finally, fairness assertions are implemented with scheduler structures. These are reusable hardware structures, which are stored into a library. A scheduler obtains guard signals, evaluates them, and selects the corresponding enable action for execution. The principles of scheduler structures are discussed in detail in [11], the DisCo-VHDL Compiler is described in [12], and the entire design methodology is reported in [13].

IV. SPECTRAL TESTING METHODS

Spectral techniques are a mathematical discipline, which may be described as an area of abstract harmonic analysis devoted to the applications in engineering, primarily electrical and computer engineering. Transferring a problem from the original into the spectral domain may provide several advantages. Some numerical calculation tasks, difficult to perform in the original domain, may be simpler in the spectral domain. Additionally, many properties of the functions, realizing circuits, are much more easily detected in the spectral domain than in the Boolean domain, see, e.g. [14][15].

With the advent of VLSI and corresponding drastic increase in the density of gates on a chip, high-level functional testing is one of the most viable approaches to the testing of computer hardware. It may be remarked, that testing is the “bottleneck” of computer industry. The cost of testing is often higher, than the cost of design. Interconnection of components in the device-under-test (DUT) is too complex or not known for a user, which makes the testing a rather difficult and complex task in terms of both space and time. Traditional methods of design and testing require brute force computer search for solving optimization problems. Unlike to that, spectral methods may provide simple “analytic” solutions.

Figure 2 – Binary Decision Diagram.

a) Binary Decision Diagrams

Major problem with spectral techniques in the area of digital design is the computational complexity. Most of the spectral transform algorithms used for signal processing are not feasible here, mostly due to a much larger size of the problem instances. One possible way of dealing with this problem is to develop separate transform algorithms for various classes of functions (divide-and-conquer approach) [16]. However, such an approach is not always appropriate. An alternative way is to utilize sophisticated data structures, able to handle a wide variety of functions, where the state of the art is Binary decision diagram (BDD). BDDs have evolved over the years into a very successful data structure for Boolean function manipulation. They allow for rapid and successful computation of spectral coefficients for a wide variety of functions, realizing circuits, e.g., in [18][19][22][23][24].

However, even BDDs do not always provide a satisfactory solution. For some circuits we cannot build a BDD for a given function within a reasonable amount of space and time resources. Thus, in the case of most arithmetic circuits, BDDs express an exponential complexity in terms of input counts. For this exponential
complexity, the improvement in computer speed and memory does not considerably contribute to the solution of the problem. On the other hand, in certain tasks only a single or a few coefficients of the spectrum are of an interest. In this case, BDD based solutions are still possible even for huge circuits [15]. Otherwise, when complete spectrum must be computed, new data structures for representation of Boolean (or discrete) functions are required.

Figure 3 – Group-theoretic approach to logic design.

b) Super Decision Diagrams

Super Decision Diagrams (SrDDs) are a generalization of BDDs, based on the redistribution of information content in a complex graph with simple nodes, as a BDD, into a simple graph with complex nodes.

The increased functionality of nodes in SrDDs will provide solutions where the existing BDDs cannot be used [20][21]. This approach is supported by the existing technology which permits realizations of such nodes without increasing their cost. For example, in commercially available Look-up Table Field Programmable Logic Arrays (LUT-FPGAs), the realization of any subfunction with a given number of variables has the same cost, a complete LUT is required irrespective of the complexity and features of the subfunction realized. Due to the increased functionality of nodes, SrDDs will provide an efficient use of LUTs, unlike existing BDDs, and for the simplified structure of the graph, the total complexity of realizations will be reduced.

V. RESULTS AND IMPACTS

The main results of the project have been published in international conferences and scientific journals. The impacts of the research include citations and invited papers and presentations. Several professionals with expertise in the field have been produced, i.e., the project has produced one Dr.Tech. dissertation [13], one Dr.Tech. dissertation is under construction [1], and two graduate degrees.

The project created closer international collaboration with Univ. CA, Berkeley, USA; Dr.Tech. Kimmo Kuusilinna joined Prof. R. Brodersen’s group in Berkeley Wireless Research Center for 21 months during 2001-2002. Project had also collaboration with Prof. M. Karpovsky (Boston Univ., USA), Prof. R. Stankovic (Univ. of Nis, Serbia), and Prof. E. Dubrova (KTH, Royal Institute of Technology, Sweden).

REFERENCES


The Academy of Finland carried out the first Research Programme for Telecommunication Electronics, Telectronics, in 1998-2000. The objective was to create and support high-level basic research on scientifically essential and rapidly developing areas of telecommunications and electronics research, by combining research and expertise on different areas. The activity of Telectronics programme was continued by Telectronics II during 2001-2003. The programme was targeted to increase the knowledge in the main technologies of broadband data transfer and to produce new information to apply for general use.

This research report collection includes the scientific reports of 16 research projects included in Telectronics and Telectronics II research programmes. Together the reports cover a six-year period of targeted basic research funding for telecommunication electronics area.