JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

1/2010

(*Contents Continued on Back Cover*)

Editorial Board

Editorial Advisory Board

ISSN 1509-4553 on-line: ISSN 1899-8852

Circulation: 300 copies

Sowa - Druk na życzenie, www.sowadruk.pl, tel. 022 431-81-40

JOURNAL OF TELECOMMUNICATIONS AND INFORMATION TECHNOLOGY

Preface

Despite the fact that a range of limitations are beginning to appear as CMOS technology is being raised to ever higher levels of perfection, it is anticipated that silicon will be the dominant material of the semiconductor industry for at least the first half of the 21st century. The forecast for microelectronics development updated in 2008 by SIA (Semiconductor Industry Association) reaches ahead to the years 2016–2022. Unfortunately, comparison with former SIA forecasts indicates that in certain aspects they become less aggressive (that is less optimistic) with time.

While the development of silicon microelectronics in the past could be attributed mostly to the reduction of the feature size (progress in lithography), today it relies more on new material (SOI, SON, SiGe or SiC) and architecture (ultra-thin body, double-gate, multiple-gate) solutions. The combination of this trend with continuous miniaturization provides the opportunity of improving IC functionality and speed of operation.

Telecommunications and information technology are arguably the most powerful drivers behind microelectronics product development nowadays. Plenty of new applications are being created for fast analog and RF circuits, as well as for information processing ones. It is clear that with the anticipated peak $f_{\max} = 425$ GHz and $f_T = 395$ GHz to be reached by RF SiGe-base bipolar transistors in 2014, according to the 2008 update of ITRS, a lot of efforts must be put into the development of appropriate material, processing, characterization and modeling. While progress in the bipolar technology is impressive, the increase of MOSFET speed is even more so. The same issue of ITRS predicts on-chip clock of $\sim$14 GHz for 2022.

High-speed isn't, however, everything. Portable wireless products push, for obvious reasons, for low-power solutions. This trend requires new architectural solutions (e.g., channel thinning), and in consequence, new materials, such as SOI (or its possible successor SON – silicon-on-nothing), where current driveability is considerably higher than in conventional MOSFETs.

In this issue the Reader will find papers devoted to the history of semiconductors, charging effects at the interfaces between high-k dielectrics and SiO_2, the influence of fluorine on the quality of the Si-SiO_2 interface, large-signal RF modeling, as well as modeling and simulation of heterogeneous systems.

Wireless applications in challenging environments require improved modulation schemes and signal processing. One example is dealing with frequency offset and phase noise in orthogonal frequency division multiplexing (OFDM) systems. In a novel scheme the frequency offset is first estimated using an autocorrelation method, and then refined by applying an iterative phase correction by means of pilot-based Wiener filtering; the method was tested in a multipath indoor environment.

The next study devoted to radio systems included in this issue found that use of chaotic spreading sequence in a multicarrier code division multiple access system (MC-CDMA) to spread spectrum and estimate the transmission channel system significantly outperforms the Walsh-Hadamard code spreading in MC-CDMA system with respect to channel identification. The proposed scheme uses a chaotic sequence generated by a logistic map as a training signal and estimate channel parameters according to dynamics of the chaotic sequence.

Global satellite-based positioning systems enable to build new telematic systems for applications and services in many branches of economy, in particular applied to mobile objects like vehicles, called mobile telematic services. A paper include in this issue presents features of such services, with a special emphasis on services foreseen in Galileo satellite positioning system, including a necessity of complementary communications between a positioned object and related surroundings.

In the age of fast internet, core and metro networks widely employ wavelength division multiplexing (WDM) in optical fiber transmission to satisfy the growing need for bandwidth. While capacity of WDM networks is adequate, failure (e.g., a cable cut) potentially leads to enormous data and revenue loss, and protection is one of the key techniques used in "survivable" WDM networks. The study compares the performance of protection schemes such as dedicated path protection (DPP), shared path protection (SPP) and shared link protection (SLP), taking into account capacity utilization, switching time and blocking probability.

The last subject covered is different: optimization of the multi-threaded interval algorithm for the Pareto-set computation, and the possibility of applying interval methods to seek the Pareto-front of a multicriterial nonlinear problem. An efficient algorithm has been proposed and tested before, the current paper presents its optimization in order to increase the speedup of multi-threaded variant, and to extend the algorithm to compute not only the Pareto-front (in the criteria space), but also the Pareto-set (in the decision space).

We hope the Readers will find this issue of the *Journal of Telecommunications and Information Technology* useful and interesting.

Andrzej Jakubowski
Lidia Łukasiak
Guest Editors

Paweł Szczepański
Editor-in-Chief

Invited paper

History of Semiconductors

Lidia Łukasiak and Andrzej Jakubowski

Abstract—The history of semiconductors is presented beginning with the first documented observation of a semiconductor effect (Faraday), through the development of the first devices (point-contact rectifiers and transistors, early field-effect transistors) and the theory of semiconductors up to the contemporary devices (SOI and multigate devices).

Keywords—band theory, laser, Moore's law, semiconductor, transistor.

1. Introduction

There is no doubt that semiconductors changed the world beyond anything that could have been imagined before them. Although people have probably always needed to communicate and process data, it is thanks to the semiconductors that these two important tasks have become easy and take up infinitely less time than, e.g., at the time of vacuum tubes.

The history of semiconductors is long and complicated. Obviously, one cannot expect it to fit one short paper. Given this limitation the authors concentrated on the facts they considered the most important and this choice is never fully impartial. Therefore, we apologize in advance to all those Readers who will find that some vital moments of the semiconductor history are missing in this paper.

The rest of this paper is organized in four sections devoted to early history of semiconductors, theory of their operation, the actual devices and a short summary.

2. Early History of Semiconductors

According to G. Busch [1] the term "semiconducting" was used for the first time by Alessandro Volta in 1782. The first documented observation of a semiconductor effect is that of Michael Faraday (1833), who noticed that the resistance of silver sulfide decreased with temperature, which was different than the dependence observed in metals [2]. An extensive quantitative analysis of the temperature dependence of the electrical conductivity of Ag_2S and Cu_2S was published in 1851 by Johann Hittorf [1].

For some years to come the history of semiconductors focused around two important properties, i.e., rectification of metal-semiconductor junction and sensitivity of semiconductors to light and is briefly described in Subsections 2.1 and 2.2.

2.1. Rectification

In 1874 Karl Ferdinand Braun observed conduction and rectification in metal sulfides probed with a metal point (whisker) [3]. Although Braun's discovery was not immediately appreciated, later it played a significant role in the development of the radio and detection of microwave radiation in WWII radar systems [4] (in 1909 Braun shared a Nobel Prize in physics with Marconi). In 1874 rectification was observed by Arthur Schuster in a circuit made of copper wires bound by screws [4]. Schuster noticed that the effect appeared only after the circuit was not used for some time. As soon as he cleaned the ends of the wires (that is removed copper oxide), the rectification was gone. In this way he discovered copper oxide as a new semiconductor [5]. In 1929 Walter Schottky experimentally confirmed the presence of a barrier in a metal-semiconductor junction [5].

2.2. Photoconductivity and Photovoltaics

In 1839 Alexander Edmund Becquerel (the father of a great scientist Henri Becquerel) discovered the photovoltaic effect at a junction between a semiconductor and an electrolyte [6]. The photoconductivity in solids was discovered by Willoughby Smith in 1873 during his work on submarine cable testing that required reliable resistors with high resistance [7]. Smith experimented with selenium resistors and observed that light caused a dramatic decrease of their resistance. Adams and Day were the first to discover the photovoltaic effect in a solid material (1876). They noticed that the presence of light could change the direction of the current flowing through the selenium connected to a battery [8]. The first working solar cell was constructed by Charles Fritts in 1883. It consisted of a metal plate and a thin layer of selenium covered with a very thin layer of gold [8]. The efficiency of this cell was below 1% [9].

3. Theory

In 1878 Edwin Herbert Hall discovered that charge carriers in solids are deflected in magnetic field (Hall effect). This phenomenon was later used to study the properties of semiconductors [10]. Shortly after the discovery of the electron by J. J. Thomson several scientists proposed theories of electron-based conduction in metals. The theory of Eduard Riecke (1899) is particularly interesting, because he assumed the presence of both negative and positive charge carriers with different concentrations and mobilities [1]. Around 1908 Karl Baedeker observed the dependence of the conductivity of copper iodide on the stoichiometry (iodine content). He also measured the Hall effect in this material, which indicated carriers with positive charge [1]. In 1914 Johan Koenigsberger divided solid-state materials into three groups with respected to their conductivity: metals,

insulators and "variable conductors" [1]. In 1928 Ferdinand Bloch developed the theory of electrons in lattices [10]. In 1930 Bernhard Gudden reported that the observed properties of semiconductors were due exclusively to the presence of impurities and that chemically pure semiconductor did not exist [1].

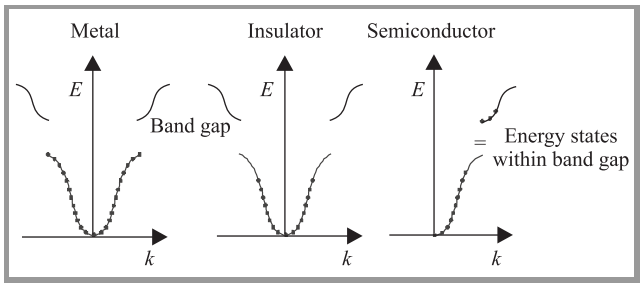


Fig. 1. Alan Wilson's theory of bands in solids.

In 1930 Rudolf Peierls presented the concept of forbidden gaps that was applied to realistic solids by Brillouin the same year. Also in 1930 Kronig and Penney developed a simple, analytical model of periodic potential. In 1931 Alan Wilson developed the band theory of solids based on the idea of empty and filled energy bands (Fig. 1). Wilson also confirmed that the conductivity of semiconductors was due to impurities [10]. In the same year Heisenberg developed the concept of hole (which was implicit in the works of Rudolf Peierls [10]). In 1938 Walter Schottky and Neville F. Mott (Nobel Prize in 1977) independently developed models of the potential barrier and current flow through a metal-semiconductor junction. A year later Schottky improved his model including the presence of space charge. In 1938 Boris Davydov presented a theory of a copper-oxide rectifier including the presence of a p-n junction in the oxide, excess carriers and recombination. He also understood the importance of surface states [11]. In 1942 Hans Bethe developed the theory of thermionic emission (Nobel Prize in 1967).

4. Devices

4.1. Point-Contact Rectifiers

In 1904 J. C. Bose obtained a patent for PbS point-contact rectifiers [12]. G. Pickard was the first to show that silicon point-contact rectifiers were useful in detection of radio waves (patent in 1906) [10]. The selenium and copper oxide rectifiers were developed, respectively, in 1925 by E. Presser and 1926 by L. O. Grondahl [10]. The selenium rectifiers were heavily used in the WWII in military communications and radar equipment [10].

4.2. The p-n Junction

During his work on the detection of radio waves Russel Ohl realized that the problems with cat's whisker detectors were caused by bad quality of the semiconductor. Therefore he melted the silicon in quartz tubes and then let it cool down. The obtained material was still polycrystalline but the electrical tests demonstrated that the properties were much more uniform. Ohl identified the impurities that created the p-n junction that he accidentally obtained during his technological experiments. He held four patents on silicon detectors and p-n junction [13].

4.3. Bipolar Transistor

In 1945 William Shockley put forward a concept of a semiconductor amplifier operating by means of the field-effect principle. The idea was that the application of a transverse electric field would change the conductance of a semiconductor layer. Unfortunately this effect was not observed experimentally. John Bardeen thought that this was due to surface states screening the bulk of the material from the field (Fig. 2). His surface-theory was published in 1947 [14].

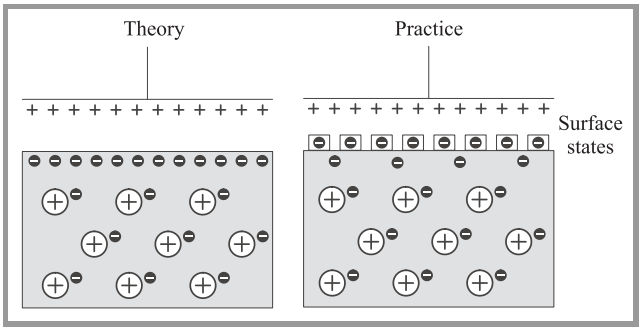


Fig. 2. The idea of surface states.

While working on the field-effect devices, in December 1947 John Bardeen and Walter Brattain built a germanium point-contact transistor (Fig. 3) and demonstrated that this device exhibited a power gain. There was, however, an uncertainty concerning the mechanism responsible for the transistor action [13]. Bardeen and Brattain were convinced that surface-related phenomena had the dominant role in the operation of the new device while Shockley favoured bulk conduction of minority carriers. About one month later he developed a theory of a p-n junction and a junction transistor [15]. Shockley, Bardeen and Brattain received the Nobel Prize in physics in 1956 (John Bardeen received another one in 1972 for his theory of superconductivity). In February 1948 John Shive demonstrated a correctly operating point-contact transistor with the emitter and collector placed on the opposite sides of a very thin slice of germanium (0.01 cm). This configuration indicated that the conduction was indeed taking place in the bulk, not along the surface (the distance between the emitter and collector along the surface would be much longer) [15]. It was only then that Shockley presented his theory of transistor operation to the coworkers [15], [16].
It is worth remembering that the crucial properties of semiconductors at the time were "structure sensitive"

(as Bardeen put it in [14]), that is they were strongly dependent on the purity of the sample. The semiconductor material with which Bardeen and Brattain worked was prepared using a technique developed by Gordon K. Teal and John B. Little based on the Czochralski method. The crystal was then purified using the zone refining method proposed by William G. Pfann [11].

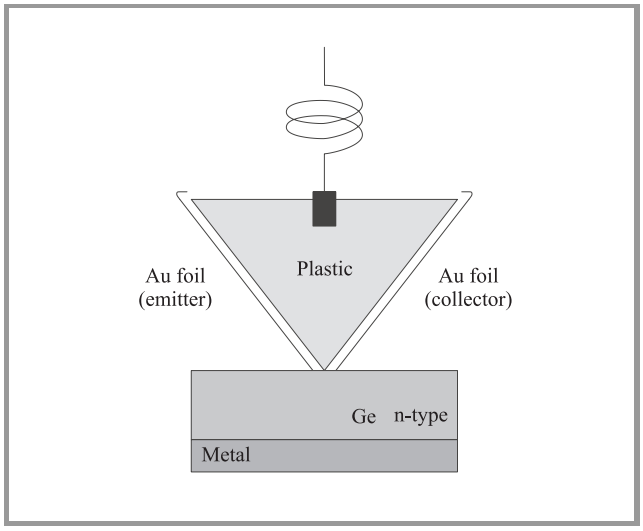


Fig. 3. The first point-contact transistor [16].

Point-contact transistors were the first to be produced, but they were extremely unstable and the electrical characteristics were hard to control. The first grown junction transistors were manufactured in 1952. They were much better when compared to their point-contact predecessor, but the production was much more difficult. As a result of a complicated doping procedure the grown crystal consisted of three regions forming an n-p-n structure. It had to be cut into individual devices and contacts had to be made. The process was difficult and could not be automated easily. Moreover, a lot of semiconductor material was wasted. In 1952 alloyed junction transistor was reported (two pellets of indium were alloyed on the opposite sides of a slice of silicon). Its production was simpler and less material-consuming and could be automated at least partially. The obtained base width was around 10 μm, which let the device operate up to a few MHz only. The first diffused Ge transistor (diffusion was used to form the base region, while the emitter was alloyed) with a characteristic "mesa" shape was reported in 1954. The base width was 1 μm and the cut-off frequency 500 MHz. It was generally understood that for most applications silicon transistors would be better than germanium ones due to lower reverse currents. The first commercially available silicon devices (grown junction) were manufactured in 1954 by Gordon Teal. The first diffused Si transistor appeared in 1955. To reduce the resistivity of the collector that limited the operation speed without lowering the breakdown voltage too much John Early thought of a collector consisting of two layers, i.e., high-resistivity one on top of a highly doped one. A transistor with epitaxial layer added was reported in 1960. In the same year Jean Hoerni proposed the planar transistor (both base and emitter regions diffused). The oxide that served as a mask was not removed and acted as a passivating layer [15].

Further improvement of speed was proposed by Herbert Kroemer. A built-in electric field could be introduced into the base by means of graded doping. Another way of introducing the electric field in the base he thought of was grading the composition of the semiconductor material itself, which resulted in graded band gap. This heterostructure concept could not be put to practice easily because of fabrication problems [17].

4.4. Integrated Circuit

The transistor was much more reliable, worked faster and generated less heat when compared to the vacuum tubes [18]. Thus it was anticipated that large systems could be built using these devices. The distance between them had, however, to be as short as possible to minimize delays caused by interconnects. In 1958 Jack Kilby demonstrated the first integrated circuit where several devices were fabricated in one silicon substrate and connected by means of wire bonding. Kilby realized that this would be a disadvantage therefore in his patent he proposed formation of interconnects by means of deposition of aluminum on a layer of SiO_2 covering the semiconductor material [15]. This has been achieved independently by Robert Noyce in 1959. In 2000 Jack Kilby received a Noble Prize in physics for his achievements.

4.5. Tunnel Diode

Leo Esaki studied heavily doped junctions to find out how high the base of a bipolar transistor could be doped before the injection at the emitter junction became inadequate. He was aware that in very narrow junctions tunneling could take place. He obtained the first Ge tunneling diode in 1957 and a silicon one in 1958. Esaki's presentation at the *International Conference of Solid State Physics in Electrons and Telecommunications* in 1958 was highly appreciated by Shockley [19]. Unfortunately, Shockley exhibited a complete lack of interest when Robert Noyce came to him to present his idea of a tunnel diode two years earlier. As a result Noyce moved to other projects [20]. The tunnel diode was extremely resistant to the environmental conditions due to the fact that conduction was not based on minority carriers or thermal effects. Moreover, its switching times were much shorter than those of the transistor. Leo Esaki received a Nobel Prize in physics in 1973 for his work on tunneling and superlattices [21], [22].

4.6. Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET)

In 1930 and 1933 Julius Lilienfeld obtained patents for devices resembling today's MESFET and MOSFET, respec-

tively. In 1934 Oskar Heil applied for a patent for his theoretical work on capacitive control in field-effect transistors [3].

The first bipolar transistors were quite unreliable because semiconductor surface was not properly passivated. A group directed by M. M. Atalla worked on this problem and found out that a layer of silicon dioxide could be the answer [23]. During the course of this work a new concept of a field-effect transistor was developed and the actual device manufactured [24]. Unfortunately, the device could not match the performance of bipolar transistors at the time and was largely forgotten [15]. Several years before Bell Laboratories demonstrated an MOS transistor Paul Weimer and Torkel Wallmark of RCA did work on such devices. Weimer made transistors of cadmium sulfide and cadmium selenide [11]. In 1963 Steven Hofstein and Fredric Heiman published a paper on a silicon MOSFET [25] (Fig. 4). In the same year the first CMOS circuit was proposed by Frank Wanlass [26]. In 1970 Willard Boyle and George Smith presented the concept of charge-coupled devices (CCD) – a semiconductor equivalent of magnetic bubbles [27]. Both scientists received a Nobel Prize in physics in 2009 for their work on CCD.

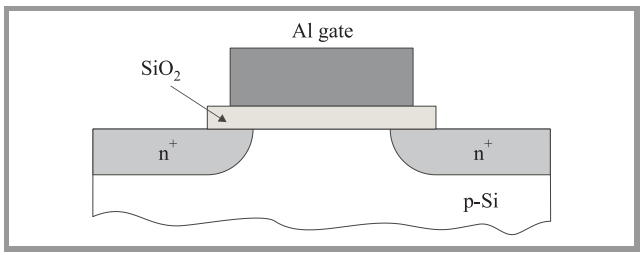


Fig. 4. A cross section of a metal-oxide-semiconductor transistor.

Early MOSFETs had aluminum gate. Development of a poly-Si gate [28] led to a self-aligned device, where the gate itself constitutes the mask for source and drain diffusion. In this way parasitic gate-to-source and gate-to-drain capacitances associated with gate overlap could be controlled. Since polysilicon had relatively high resistance, gates made of silicides of refractory metals were proposed (e.g., [29], [30]).

Reduction of the size of the device led to the so-called short-channel effects (SCE) including threshold voltage roll-off and drain-induced barrier lowering. The ways to cope with this problem include a reduction of the depth of source and drain [31] combined with efforts to avoid increased resistance (e.g., lightly doped drain [32], elevated source/drain (S/D) [33] or possibly Schottky barrier S/D [34]). Threshold voltage and punchthrough are controlled by means of the appropriate doping profile of the channel that makes it possible to maintain relatively good surface mobility (e.g., [35]). Short-channel effects are considerably reduced when gate oxide is thin. As a result of decreased thickness, gate leakage current obviously grows, increasing power consumption of the entire chip, which is an undesirable effect for battery-powered mobile systems. It is estimated that gate leakage current increases approximately 30 times every technology generation, as opposed to 3–5 times increase of channel leakage current [36]. Apart from leakage current, the reduction of gate-oxide thickness increases the susceptibility of the device to boron penetration from the poly-Si gate into the channel. A number of different high-k materials are extensively investigated.

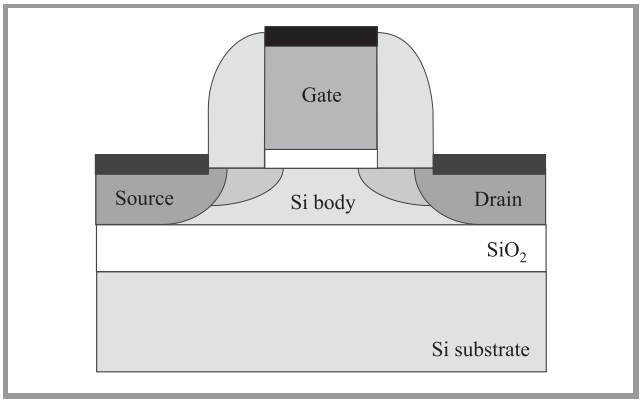


Fig. 5. A cross section of a SOI MOSFET.

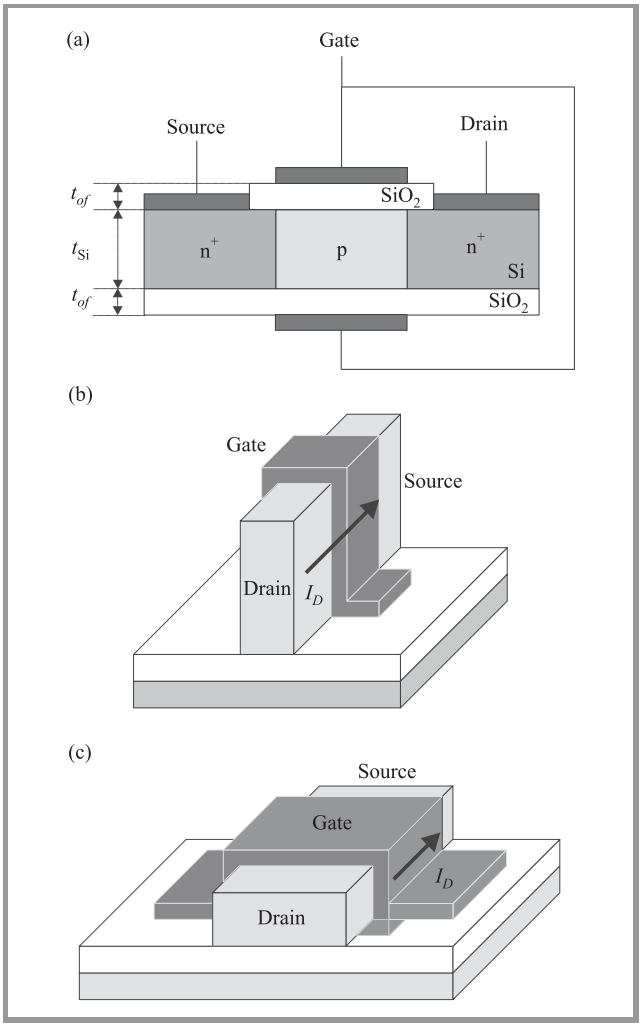


Fig. 6. Mutigate transistors: (a) double gate; (b) FinFET; (c) surrounding gate.

An interesting extension of the classical bulk MOSFET is silicon-on-insulator (SOI) – see Fig. 5 [37]. The advantage of SOI is the ease of electrical isolation of a device from the rest of the integrated circuit, which increases packing density. Moreover, the area of source and drain junctions is significantly reduced, thus decreasing parasitic capacitances. Finally, the depletion width is limited by the Si body thickness, therefore it is widely believed that SOI helps reduce short channel effects unless source-to-drain coupling through channel and BOX cannot be neglected. The properties of SOI devices are improved with the reduction of body thickness. It is believed that fully depleted ultra-thin-body SOI (FD UTB SOI) is one of the best scaling solutions. Due to excellent gate control of the channel these devices may be undoped or very lightly doped. In this way mobility is not degraded and threshold voltage is less dependent on the fluctuations of doping concentration [38]. Another advantage of SOI is that it facilitates development of new device concepts [39] (Fig. 6), but this is another story.

4.7. Semiconductor Lasers

Semiconductors are widely used for emission and detection of radiation. The first report on light emitted by a semiconductor appeared in 1907 in a note by H. J. Round. Fundamental work in this area was conducted, among other, by Losev. A very interesting description of the development of light-emitting diodes may be found in [40] while the history of photovoltaics is discussed in [8]. In this section only semiconductor lasers are mentioned briefly.

The first semiconductor lasers were developed around 1962 by four American research teams [41]. Further research in this area went in two directions, i.e., wider spectrum of materials to obtain wider wavelength range and concepts of new device structures. Herbert Kroemer and Zhores Alferov have independently come up with the idea that semiconductor lasers should be built on heterostructures. Zhores Alferov was a member of the team that created the first Soviet p-n junction transistor in 1953. He was directly involved in research aimed at development of specialized semiconductor devices for Russian nuclear submarines. The matter was of such importance for the Soviet authorities that he used to receive phone calls from very high government officials who wanted the work done faster. To fulfill those requests Alferov had to move to the lab and literally live there [42]. Later he worked on power devices and became familiar with p-i-n and p-n-n structures. When the first report on semiconductor lasers appeared, he realized that double heterostructures of the p-i-n type should be used in these devices [41]. He obtained the first practical heterostructure devices and the first heterostructure laser [42]. In 2000 Alferov and Kroemer (mentioned in Subsection 4.4) received a Nobel Prize in physics for their achievements in the area of semiconductor heterostructures used in high-speed- and optoelectronics.

Significant progress in semiconductor lasers is associated, among other, with the use of quantum wells and new materials, especially gallium nitride.

5. Summary

Silicon may be considered as the information carrier of our times. In the history of information there were two revolutions (approximately 500 years apart). The first was that of Johan Gutenberg who made information available to many, the other is the invention of the transistor. Currently the global amount of information doubles every year. Many things we are taking for granted (such as, e.g., computers, Internet and mobile phones) would not be possible without silicon microelectronics. Electronic circuits are also present in cars, home appliances, machinery, etc. Optoelectronic devices are equally important in everyday life, e.g., fiber-optic communications for data transfer, data storage (CD and DVD recorders), digital cameras, etc.

Since the beginning of semiconductor electronics the number of transistors in an integrated circuit has been increasing exponentially with time. This trend had been first noticed by Gordon Moore [43] and is called Moore's law. This law is illustrated in Fig. 7, where the number of transistors in successive Intel processors is plotted as a function of time (data after [44]).

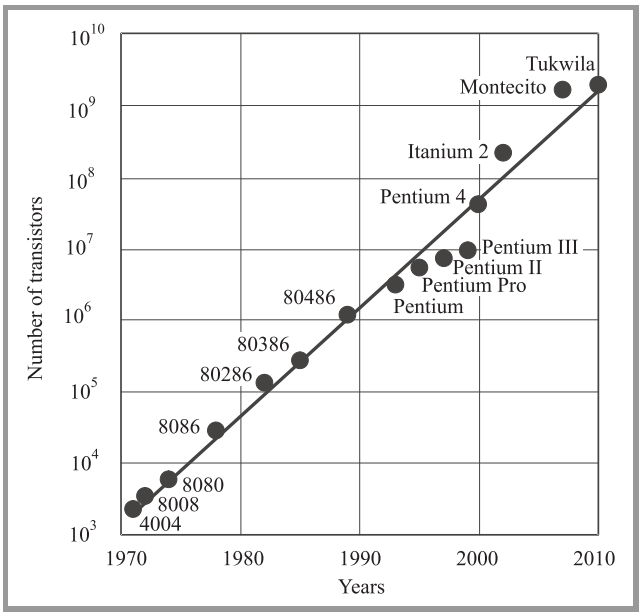


Fig. 7. Number of transistors in successive Intel processors as a function of time (data after [44]).

Even though the bipolar technology was largely replaced by CMOS (more than 90 percent of integrated circuits are manufactured in CMOS technology), Moore's law is still true in many aspects of the development trends of silicon microelectronics (obviously, with the appropriate time constant). The MOS transistor has been improved countless times but above everything else it has been miniaturized

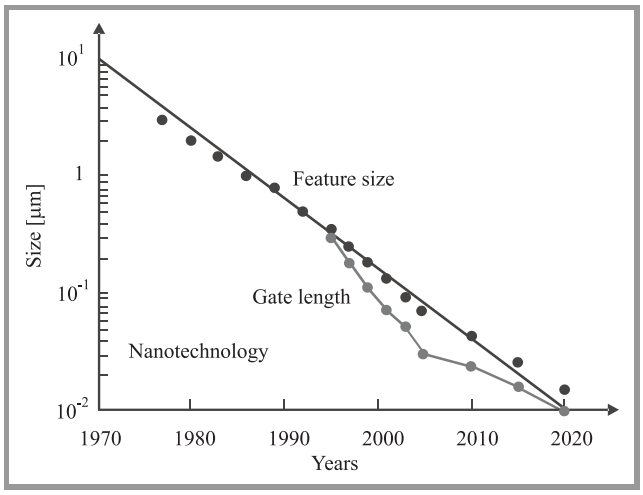


Fig. 8. Feature size as a function of time (data after [45]).

beyond imagination. The reduction of the feature size, presented in Fig. 8, is more or less exponential. The number of transistors produced per year and the average price are shown as a function of time in Fig. 9 (again the change is exponential). It is being anticipated that in 2010 approximately one billion transistors will be produced for every person living on the Earth.

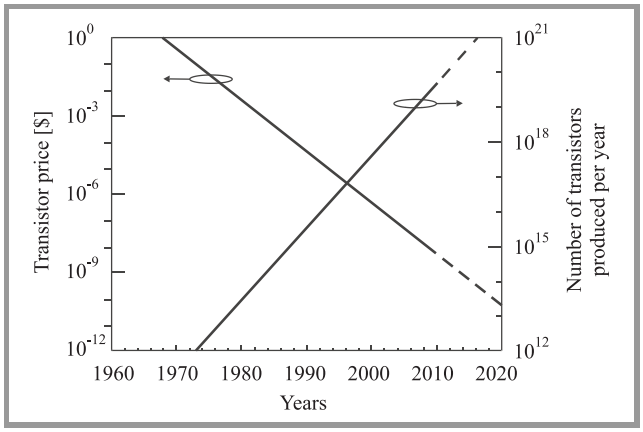


Fig. 9. Number of transistors produced per year and transistor price as a function of time (data after [46]).

We are pretty sure the future still holds a few surprises. Extensive research is being carried out on graphene, organic electronics, quantum devices, microsystems, integration of silicon with other materials and many other issues, but that is another story…

References

[1] G. Busch, "Early history of the physics and chemistry of semiconductors – from doubts to fact in a hundred years", *Eur. J. Phys.*, vol. 10, no. 4, pp. 254–263, 1989.

[2] F. Laeri, F. Schüth, U. Simon, and M. Wark, *Host-Guest-Systems Based on Nanoporous Crystals*. Weinheim: Wiley, 2003, pp. 435–436.

[3] T. K. Sarkar, R. Mailloux, A. A. Oliner, M. Salazar-Palma, and D. L. Sengupta, *The History of Wireless*. Hoboken: Wiley, 2006.

[4] J. Orton, *Semiconductors and the Information Revolution: Magic Crystals that Made IT Happen*. Amsterdam: Academic Press/Elsevier, 2009, pp. 35–36

[5] W. Mönch, *Semiconductor Surfaces and Interfaces*. Berlin-Heidelberg: Springer, 2001.

[6] Z. A. Smith and K. D. Taylor, *Renewable and Alternative Energy Sources: A Reference Handbook*. Santa Barbara: ABC-CLIO Inc., 2008, p. 157.

[7] J. Orton, *The Story of Semiconductors*. Oxford: Oxford University Press, 2004, p. 359.

[8] J. Perlin, *From Space to Earth: The Story of Solar Electricity*. Cambridge: Harvard University Press, 2002, p. 17.

[9] M. Grundman, *The Physics of Semiconductors*. Berlin-Heidelberg: Springer, 2006.

[10] L. Hoddeson, E. Braun, J. Teichmann, and S. Weart, *Out of the Crystal Maze: Chapters in the History of Solid State Physics.* New York: Oxford University Press, 1992.

[11] B. Lojek, *History of Semiconductor Engineering.* Berlin-Heidelberg: Springer, 2007.

[12] P. K. Bondyopathyay, "Sir J. C. Bose's diode detector received Marconi's first transatlantic wireless signal of December 1901 (the "Italian navy coherer" scandal revisited)", *Proc. IEEE*, vol. 86, no. 1, pp. 259–285, 1998.

[13] M. Riordan and L. Hoddeson, "The origins of the p-n junction", *IEEE Spectrum*, vol. 34, no. 6, p. 46, 1997.

[14] J. Bardeen, "Solid state physics – 1947", *Solid State Technol.*, vol. 30, no. 12, pp. 69–71, 1987.

[15] I. M. Ross, "The invention of the transistor", *Proc. IEEE*, vol. 86, no. 1, pp. 7–27, 1998.

[16] M. Riordan, L. Hoddeson, and C. Herring, "The invention of the transistor", *Rev. Mod. Phys.*, vol. 71, no. 2, pp. S336–S345, 1999.

[17] T. S. Perry, "Not just blue sky", *IEEE Spectrum*, vol. 39, no. 6, pp. 33–37, 2002.

[18] J. A. Armstrong, "Solid state technology and the computer: 40 years later small is still beautiful", *Solid State Technol.*, vol. 30, no. 12, pp. 81–83, 1987.

[19] L. Esaki, "Discovery of the tunnel diode", *IEEE Trans. Electron Dev.*, vol. 23, no. 7, pp. 644–647, 1976.

[20] L. Berlin and H. Casey, Jr., "Robert Noyce and the tunnel diode", *IEEE Spectrum*, vol. 42, no. 5, pp. 49–53, 2005.

[21] L. Esaki and R. Tsu, "Superlattice and negative conductivity in semiconductors", IBM Res. Note, RC-2418, March 1969.

[22] L. Esaki and R. Tsu, "Superlattice and negative differential conductivity in semiconductors", *IBM J. Res. Develop.*, vol. 14, pp. 61–65, Jan. 1970.

[23] M. M. Atalla, E. Tannenbaum, and E. J. Scheibner, "Stabilization of silicon surfaces by thermally grown oxides", *Bell. Syst. Tech. J.*, vol. 38, no. 3, pp. 749–783, 1959.

[24] D. Kahng and M. M. Atalla, "Silicon-silicon dioxide field induced surface devices", in *Solid State Res. Conf.*, Pittsburgh, USA, 1960.

[25] S. R. Hofstein and F. P. Heiman, "Silicon insulated-gate field-effect transistor", *Proc. IEEE*, vol. 51, no. 9, pp. 1190–1202, 1963.

[26] F. M. Wanlass and C. T. Sah, "Nanowatt logic using field-effect metal-oxide semiconductor triodes", in *Proc. Techn. Dig. IEEE 1963, Int. Solid-State Circ. Conf.*, Philadelphia, USA, 1963, pp. 32–33.

[27] W. S. Boyle and G. E. Smith, "The inception of charge-coupled devices", *IEEE Trans. Electron Dev.*, vol. 23, no. 7, pp. 661–663, 1976.

[28] R. E Kerwin, D. L. Klein, and J. C. Sarace, "Method for making MIS structures", US Patent 3 475 234, filed March 27, 1967, issued Oct. 28, 1969.

[29] T. Mochizuki, K. Shibata, T. Inoue, K. Obuchi, and M. Kashiwogi, "A new gate material for MOS devices – molybdenum silicide ($MoSi_2$)", in *Proc. ECS Conf.*, Atlanta, USA, 1977, vol. 72–2, pp. 331–332.

[30] K. C. Saraswat, F. Mohammedi, and J. D. Meindl, "WSi_2 gate MOS devices", in *Proc. IEDM Tech. Dig.*, Washington, USA, 1979, pp. 462–464.

[31] J. R. Brews, W. Fichtner, E. H. Nicollian, and S. N. Sze, "Generalized guide for MOSFET miniaturization", *IEEE Electron Dev. Lett.*, vol. 1, no. 1, pp. 2–4, 1980.

[32] W. R. Hunter, T. C. Holloway, P. K. Chatterjee, and A. F. Tasch, "New edge-defined vertical-etch approaches for submicrometer MOSFET fabrication", in *Proc. IEDM Tech. Dig.*, Washington, USA, 1980, pp. 764–767.

[33] S. S. Wong, D. R. Bradbury, D. C. Chen, and K. Y. Chiu, "Elevated source/drain MOSFET", in *Proc. IEDM Tech. Dig.*, San Francisco, USA, 1984, pp. 634–637.

[34] T. Mochizuki and K. D. Wise, "An n-channel MOSFET with Schottky source and drain", *IEEE Electron Dev. Lett.*, vol. 5, no. 4, pp. 108–110, 1984.

[35] H. Tian, R. B. Hulfachor, J. J. Ellis-Monaghan, K. W. Kim, M. A. Littlejohn, J. R. Hauser, and N. A. Masnari, "An evaluation of super-steep-retrograde channel doping for deep-submicron MOSFT applications", *IEEE Trans. Electron Dev.*, vol. 41, no. 10, pp. 1880–1882, 1994.

[36] C.-T. Chuang, K. Bernstein, R. V. Joshi, R. Puri, K. Kim, E. J. Nowak, T. Ludwig, and I. Aller, "Focusing on planar device structures and strained silicon for handling silicon scaling issues in the deep sub-100 nm regime", *IEEE Circ. Dev. Mag.*, vol. 20, no. 1, pp. 6–19, 2004.

[37] J.-P. Collinge, *Silicon-on-Insulator Technology: Materials to VLSI*. Boston/Dordrecht/London: Kluwer, 2004.

[38] T. Skotnicki, J. A. Hutchby, T.-J. King, H.-S. P. Wong, and F. Boeuf, "The end of CMOS scaling", *IEEE Circ. Dev. Mag.*, vol. 21, no. 1, pp. 16–26, 2005.

[39] J.-P. Colinge, "Multiple-gate SOI MOSFETs", *Solid-State Electron.*, vol. 48, no. 6, pp. 897–905, 2004.

[40] E. E. Loebner, "Subhistories of the light emitting diode", *IEEE Trans. Electron Dev.*, vol. 23, no. 7, pp. 675–698, 1976.

[41] N. Holonyak, "The semiconductor laser: a thirty-five year perspective", *Proc. IEEE*, vol. 85, no. 11, pp. 1678–1693, 1997.

[42] Z. I. Alferov, "Autobiography", http://nobelprize.org/nobel_prizes/physics/laureates/2000/alferov-autobio.html

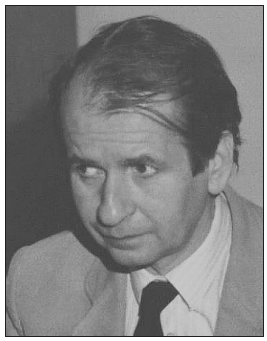
[43] G. E. Moore, "Progress in digital integrated electronics", in *Proc. IEDM Tech. Dig.*, Washington, USA, 1975, p. 103.

[44] Hardware components, Intel processor history, http://www.interfacebus.com/intel-processor-types-release-date.html

[45] International Technology Roadmap for Semiconductors, 2008, http://www.itrs.net/Links/2008ITRS/home2008.htm

[46] The singularity is near, http://singularity.com/charts

Lidia Łukasiak graduated from the Faculty of Electronics, Warsaw University of Technology, Poland, in 1988 and joined the Institute of Microelectronics and Optoelectronics the same year. She received the Ph.D. and D.Sc. degrees from the same university in 1994 and 2002, respectively. Since 2004 she has been the Vice-Director for Teaching of the Institute of Microelectronics and Optoelectronics. Her research interests include modeling and characterization of semiconductor devices and microprocessor systems.
e-mail: lukasiak@imio.pw.edu.pl
Institute of Microelectronics and Optoelectronics
Warsaw University of Technology
Koszykowa st 75
00-662 Warsaw, Poland



Andrzej Jakubowski received the M.Sc., Ph.D. and D.Sc. degrees in electrical engineering from the Warsaw University of Technology (WUT), Poland. In the years 1984–1990 and 1993–2001 the Head of the Division of Microelectronics of the Institute of Microelectronics and Optoelectronics (IMiO) and in the years 2001–2004 the Head of the Division of Microelectronics and Nanoelectronics Devices of IMiO. He was the Director of the Institute of Electron Technology (ITE) in the years 1989–1992 and Director of IMiO from 2004 till 2008. Between 1990 and 1991 he was the Chairman of the Committee of Applied Research and a member of the Prime Minister's Committee of Science and Technical Progress. He was the Vice-chairman of the Committee of Electronics and Telecommunications of Polish Academy of Sciences between 1989 and 2007. He was also the Chairman of the Microelectronics Section of this Committee from 1988 till 2003. He was the editor-in-chief of "Electron Technology" between 1990 and 1994. He is the author or co-author of more than 600 publications (journal and conference papers, monographs) and several textbooks for students, as well as popular-science papers. He delivered many invited lectures at foreign universities and international conferences in Europe, United States and Asia. He supervised 23 Ph.D. theses and more than 150 B.Sc., and M.Sc. theses. He received 5 Awards of the Minister of National Education. The main areas of his research are modeling, characterization and fabrication of semiconductor structures (e.g., MOS, SOI MOS, HBT, SiGe MOS).
e-mail: jakubowski@imio.pw.edu.pl
Institute of Microelectronics and Optoelectronics
Warsaw University of Technology
Koszykowa st 75
00-662 Warsaw, Poland

Invited paper

Charging Phenomena at the Interface Between High-*k* Dielectrics and SiO_x Interlayers

Olof Engström, Bahman Raeissi, Johan Piscator, Ivona Z. Mitrovic, Stephen Hall, Heinrich D. B. Gottlob, Mathias Schmidt, Paul K. Hurley, and Karim Cherkaoui

Abstract—The transition regions of GdSiO/SiO_x and HfO_2/SiO_x interfaces have been studied with the high-*k* layers deposited on silicon substrates. The existence of transition regions was verified by medium energy ion scattering (MEIS) data and transmission electron microscopy (TEM). From measurements of thermally stimulated current (TSC), electron states were found in the transition region of the HfO_2/SiO_x structures, exhibiting instability attributed to the flexible structural molecular network expected to surround the trap volumes. The investigations were focused especially on whether the trap states belong to an agglomeration consisting of a single charge polarity or of a dipole constellation. We found that flat-band voltage shifts of MOS structures, that reach constant values for increasing oxide thickness, cannot be taken as unique evidence for the existence of dipole layers.

Keywords—defects, dielectrics, high-k, metal oxide semiconductor.

1. Introduction

The gate function of future metal oxide semiconductor (MOS) transistors has attracted a large scientific community to an expedition into the periodic system for tracking the Dielectric Grail. Wanted is a material with acceptable energy offset values, ΔE, between the energy bands of the dielectric and the silicon crystal while, in addition, having a high enough dielectric constant, k. Yet, to fulfill the demands of low current leakage and high capacitive coupling between gate metal and transistor channel, the crucial property is the product $k \times \Delta E$ of these two quantities [1]. So far, for CMOS applications most of the efforts have been limited to metal oxides. The change from the extremely well mastered thermal SiO_2 material, to an oxide based on metals among the transition or rare earth series, has disclosed obstacles that were unnoticeable for traditional technology. Beside the problems of chemical stability between these new "high-*k*" oxides and the silicon substrate, crystallization, sensitivity to humid environment, higher concentrations of oxide traps and interface states are properties, not uncommon among these materials. Driven by technology, this has given rise to needs for understanding their microscopic properties from chemical, physical and electrical point of view.

A common attribute of high-*k* oxide films deposited on silicon is the occurrence of an SiO_x interlayer between the high-*k* material and the silicon crystal. This evokes interface electron state properties similar to those at thermal SiO_2/Si interfaces [2], [3]. Even if the interlayer lowers the effective k value of the film, it often gives better conditions for a transistor channel than those offered by a direct interface due to lower charge carrier scattering by the former. However, it must be paid for by an extra interface occurring between SiO_x and the high-*k* material [4]. As the total physical thickness of the film is in the range of 5 nm or smaller, on this length scale the transition from SiO_x to the high-*k* material can hardly be considered abrupt. It is found to include a transition region with undefined stoichiometry and thus with possible structural instabilities [5]–[11]. Recently, the occurrence of traps either singular or in dipole configurations have been noticed as the potential origin of charge sources decreasing the quality of presumptive gate insulators [11]–[16]. In the present paper, we will describe the physical and electrical properties of transition regions at GdSiO/SiO_x and HfO_2/SiO_x interfaces and demonstrate how charge carrier traffic at such positions can be interpreted in order to characterize the trap properties. Based on this reasoning, the possible existence of a dipole layer in the transition region will be addressed.

2. Properties of Transition Regions

The dominating defect causing charge carrier traps in the bulk of transition and rare-earth metal oxides is commonly considered to be the oxygen vacancy. It has similar properties as the E' center in SiO_2 and has been the object of a rich theoretical [15], [16] and experimental [10], [11] literature. In HfO_2 it has been predicted to be an amphoteric center with four [15] or five [16] charge states ranging from double donor to double acceptor behavior with the latter states positioned in the range 1–2 eV from the HfO_2 conduction band edge. Furthermore, the double acceptor level is argued to be connected with large lattice relaxation [16]. It has even been proposed that the double negatively charged energy level falls below the single negatively charged level, thus exhibiting negative-U property [17]. These trap properties will be further discussed below in relation to the results from electrical measurements.

As one example of the complex high-*k* oxide/SiO_x interface, we will discuss data from GdSiO. This dielectric was

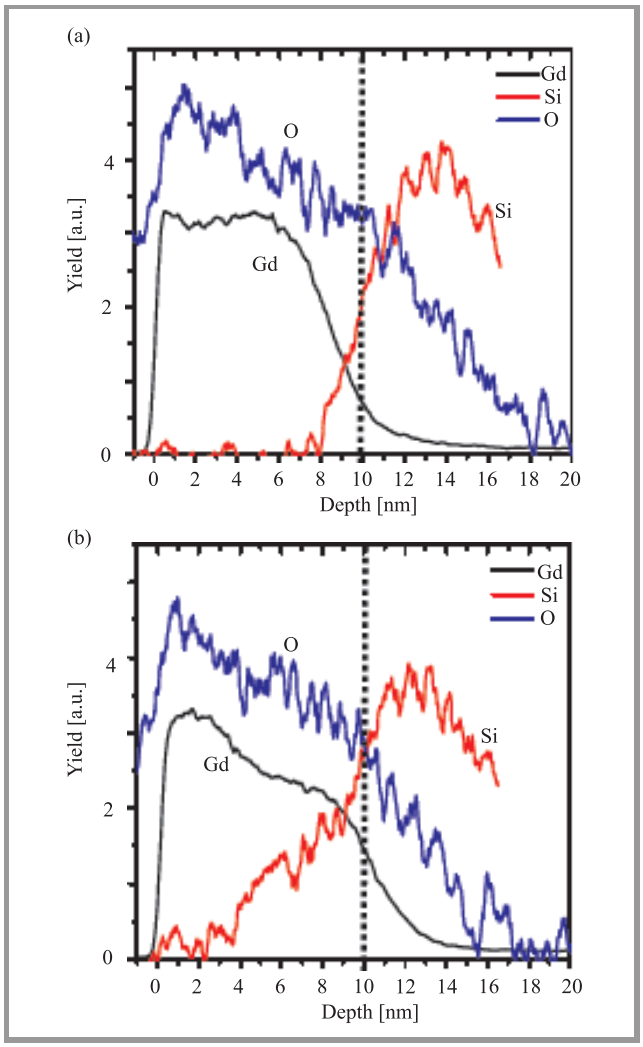


Fig. 1. MEIS data for samples with GdSiO evaporated on top of a 4 nm thermal oxide: (a) as deposited and (b) after RTA at 900°C for 1 s.

prepared by evaporating Gd_2O_3 on a 4 nm thick thermal SiO_2 layer on silicon [18]–[21]. The double layer was partly transformed into GdSiO by rapid thermal anneal (RTA) for 1 s at 900°C. The depth distribution of elements before and after the thermal anneal is shown by medium energy ion scattering (MEIS) data in Fig. 1. A steeply decreasing silicon concentration from the silicon side into the oxide directly after deposition (Fig. 1(a)), reflects a complicated diffusion process taking place already at this stage. At the 10 nm mark, the concentration of oxygen is a factor of 2 higher than the concentration of silicon. This indicates a reminiscence of SiO_2 which quickly becomes a suboxide at larger distances from this interface, where a nearly stoichiometric Gd_2O_3 takes over. The extremely high gradient of Si at the 10 nm point would be expected to stage structural instability. After the RTA at 900°C, silicon has penetrated the whole oxide layer (Fig. 1(b)) creating a GdSiO with varying concentration of Si. This structure has been demonstrated to fulfill the industrial target for low standby power, 22 nm double gate SOI transistors [18].

A second example is given by a transmission electron microscopy (TEM) picture showing the interface between HfO_2 and SiO_x in Fig. 2. This sample was prepared by reactive sputtering of Hf and O_2 followed by an anneal at 800°C for 10 min [12]. Between the white string showing the SiO_x layer and HfO_2, appearing as a black area,

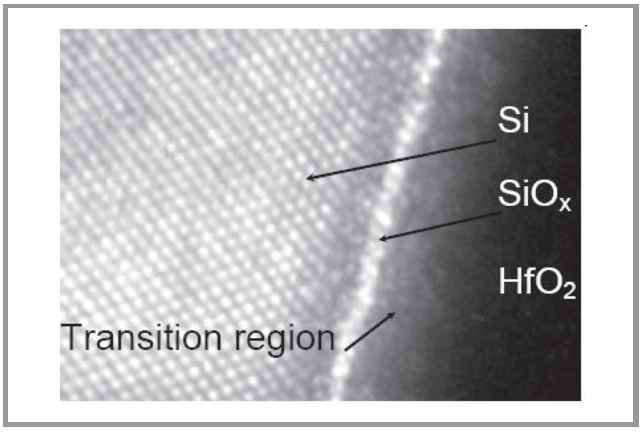


Fig. 2. TEM graph from a cross section of $Si/SiO_x/HfO_2$, where the HfO_2 layer was deposited by reactive sputtering.

a milky band is shown with a width of about 2 nm. Here, one may expect the same type of diffusion taking place as in the case of GdSiO above. As will be shown below, traps formed in this region consequently have an unusual electric behavior.

3. Charge Carrier Traffic at Transition Regions

3.1. Energy Relations and Charge Exchange

We consider an interface between a high-k oxide and an SiO_x interlayer on silicon as schematically shown in Fig. 3. The transition region is marked by a wide patterned band and includes a trap level at an energy distance ΔE below the silicon conduction band edge. A negative net charge is present in the interlayer, such that an electric field F is directed from the silicon crystal towards the high-k oxide. Assuming that the thickness of the interlayer is about 1 nm, tunneling through this layer is non-negligible and can be described as assisted by a decay of the effective density of states, N_C, from the silicon conduction band determined by $N_C \exp[-x_0/\lambda]$, where x_0 is the distance from the SiO_x/Si interface to the trap and λ is a damping factor. A captured electron can be transferred into the Si bulk by a two step process, starting with a thermally driven mechanism to the tunneling states at the energy level of the silicon conduction band edge. As long as the electric field is high enough, this process is followed by tunneling into the silicon conduction band as depicted in Fig. 3. Emitting electrons from the trap position will lower the electric field. The system

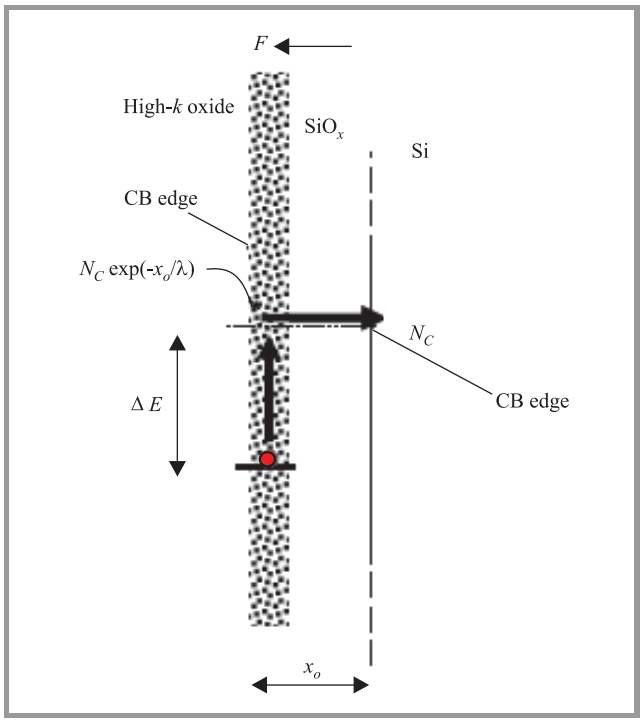


Fig. 3. Two step emission path of electrons captured in traps positioned in the transition region. The first step is thermal, to decaying states from the silicon conduction band followed by tunneling to the silicon.

may, therefore, reach a situation where the electric field becomes too low for the second transfer step to occur, which means that the emission process stops. As the tunneling probability is very sensitive to the magnitude of the electric field, this termination may be very abrupt as we will see in the experimental data following.

As long as the tunneling rate dominates, the bottleneck of this process is the thermal electron emission rate, e_n, from the trap to an effective density of states decreased by the tunneling probability

$$e_n = \sigma_n v_{th} N_C \exp\left(-\frac{x_0}{\lambda}\right) \exp\left(-\frac{\Delta E}{k_B T}\right), \tag{1}$$

where σ_n is the capture cross section representing a local transition from the tunneling state to the trap state, v_{th} is the average thermal velocity of electrons in the silicon conduction band, k_B is Boltzmann's constant and T is absolute temperature. As will be described in detail below, using thermally stimulated current (TSC) technique to determine the emission rate, the temperature is linearly increased with time during the measurement cycle. As the electric field, F, decreases due to the decrease of negative charge in the SiO_x layer, the energy distance ΔE will increase. In addition, the temperature increase will move the Fermi level in the silicon bulk to deeper energy in the band gap. For a given applied voltage across the structure, this will lower the position of the silicon conduction band edge at the SiO_x/Si interface which will increase the voltage drop across the interlayer and tend to decrease ΔE. Assuming, as a first order approximation, that both these processes are linear with temperature, we write

$$\Delta E = \Delta E_0 - \alpha T, \tag{2}$$

where ΔE_0 is the energy distance between the silicon conduction band edge and the trap level extrapolated to $T = 0$ K and α is a constant determined by the two competing processes described above. Using Eq. (2) in (1), we get

$$e_n = \sigma_e v_{th} N_C \exp\left(-\frac{\Delta E_0}{k_B T}\right), \tag{3}$$

where

$$\sigma_e = \sigma_n \exp\left(\frac{\alpha}{k_B}\right) \exp\left(-\frac{x_0}{\lambda}\right) \tag{4}$$

can be considered as an effective capture cross section, influenced by the changing energy level position and the tunneling probability. We notice that, measuring the emission rate as a function of temperature and plotting this quantity in an Arrhenius graph would give an activation energy corresponding to an extrapolation of ΔE to $T = 0$ K under the assumption of linear conditions. However, this reasoning does not take into account the properties related to local molecular dynamics of the trap volume.

For a case like the oxygen vacancy, where the transition is argued to be connected with a strong lattice relaxation [16], the emission process would be influenced also by the vibronic properties of the trap [21] and characterized by "hysteretic tunneling", adding a simultaneous trap relaxation and tunneling into the picture [22]. This would lead to two additional pre-exponential factors in the expressions for the thermal emission rate one originating from a possible thermally activated σ_n:

$$\sigma_n = \sigma_0 \exp\left(-\frac{\Delta U}{k_B T}\right), \tag{5}$$

where σ_0 depends on a combination of matrix elements including electronic and atomic wave functions involved in the process while ΔU is an activation energy originating from the vibrational properties of the trap system [21], [23]. The second effect comes from the entropy factor [21], [23], [24]:

$$X_n = \exp\left(\frac{\Delta S}{k_B}\right), \tag{6}$$

where ΔS is the change in entropy due to the change in vibrational frequency of the ionic part of the trap when the electron is released. Hence, combining the effects of the change in ΔE due to de-charging and those of a vibrating electron-ion trap system, we find from Eqs. (1)–(6):

$$\begin{aligned} e_n = {} & \sigma_0 \exp\left(-\frac{\Delta U}{k_B T}\right) \exp\left(\frac{\Delta S}{k_B}\right) \exp\left(\frac{\alpha}{k_B}\right) \\ & \times \exp\left(-\frac{x_0}{\lambda}\right) v_{th} N_C \exp\left(-\frac{\Delta H_0}{k_B T}\right), \end{aligned} \tag{7}$$

where the activation energy now is represented by an enthalpy, $\Delta H_0 = \Delta E_0 + \Delta S T$, including the heat, $\Delta S T$, stored by the local vibrational modes [21], [23], [24].

For this case the first factors in Eq. (7) make up an "effective" capture cross section, σ_e, which would be obtained from an Arrhenius plot of the thermal emission rate, e_n:

$$\sigma_e = \sigma_0 \exp\left(-\frac{\Delta U}{k_B T}\right) \exp\left(\frac{\Delta S}{k_B}\right) \exp\left(\frac{\alpha}{k_B}\right) \exp\left(-\frac{x_0}{\lambda}\right). \quad (8)$$

As will be shown below, TSC results demonstrate instabilities among the traps investigated in the HfO_2/SiO_x transition region. This is most probably originating from restructuring of the molecular arrangement around the trap volume, thus changing the matrix elements for charge carrier transition involved in σ_0 as well as quantities of the vibrational properties reflected by ΔU, ΔS, and ΔH_0 in Eqs. (7) and (8).

3.2. Thermally Stimulated Current

When measuring TSC from traps in a MOS system [12], the sample is first brought into accumulation at room temperature followed by a temperature decrease to about 50 K. During this procedure, traps at the interface and in the oxide are filled by charge carriers. At the low temperature point, the system is biased into deep depletion and a temperature increase, linear as a function of time, is applied. For an n-type semiconductor this gate bias is negative. As long as the temperature is lower than about 200 K and the total scanning time up to that temperature is shorter than about 10 min, it should be noticed that, due to the low temperature, an extremely low concentration of holes is expected in the valence band of the silicon crystal. Therefore, all de-charging processes observed as a TSC can be expected to originate from electron exchange at the insulator/silicon interface. The current created by emitted electrons is expressed by [12]

$$i(T) = \frac{C_{ox}}{C_S + C_{ox}} q N_T e_n \exp\left[-\int_{T_0}^{T_1} \beta e_n(u) du\right], \quad (9)$$

where C_{ox} and C_s are the capacitances of the oxide and the depleted semiconductor, respectively, N_T is the surface concentration of captured carriers and β is the scanning rate of the temperature, linear in time, t, such that $T = \beta t$. The function given by Eq. (9) is plotted in Fig. 4(a) for an activation energy of 0.13 eV and an effective capture cross section of 10^{-24} cm^2. Using the rather complicated expression in Eq. (8) for parameter extraction from experimental data is not practical. However, calculating the integral factor in this equation as a function of temperature, for the same input data as used above, one finds the graph shown in Fig. 4(b). It is noticed that the integral takes a value close to 1 for the initial part of the TSC curve in Fig. 4(a). Furthermore, as $C_{ox} \gg C_s$ for the deep depletion conditions used in the experiment, this part of the experimental data is proportional to the thermal emission rate e_n. Estimating the total concentration, N_T, of trap levels from the area under the TSC peak, therefore, gives a possibility to find effective capture cross sections from Arrhenius plots of e_n [12].

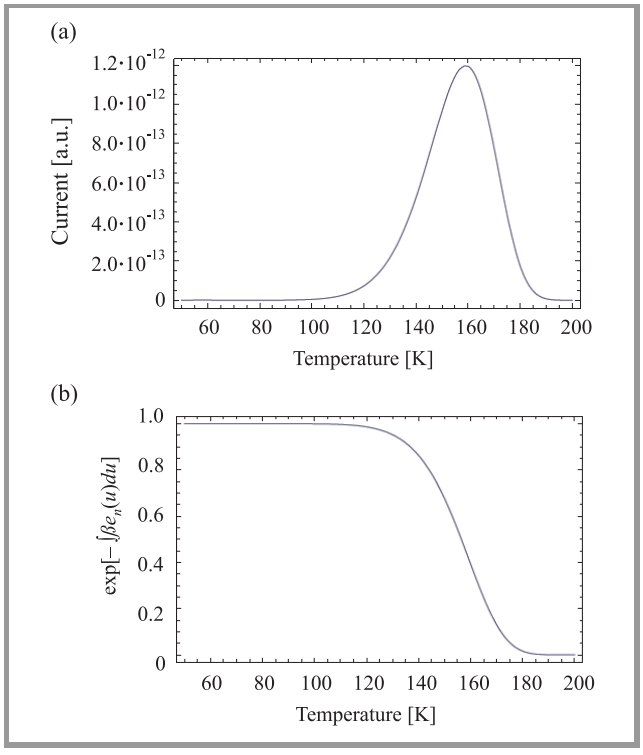


Fig. 4. (a) Theoretical plot of thermally stimulated current for an activation energy of 0.13 eV, an effective capture cross section of 10^{-24} cm^2 and a temperature scan rate of 20 K/minute. (b) The integral factor in Eq. (9) as a function of temperature. This function is close to 1 for the initial part of the curve in (a).

3.3. Experimental TSC Results on the HfO_2/SiO_x Interface

Figure 5 shows experimental TSC data from electron emission at the transition region of an MOS capacitor with an $Al/HfO_2/SiO_x/Si$ structure like the one shown by TEM in Fig. 2. Two important features can be observed in Fig. 5:

1) repeated measurement gives a different TSC peak position on the temperature scale;

2) the TSC curves are terminated before they reach the maximum point of the theoretical curves (solid curves) fitted to the experimental data.

These results are typical and occur for a large majority of the present samples with HfO_2 prepared by reactive sputtering as well as for samples prepared by atomic layer deposition (ALD) [25]. The observation (1), reveals an instability of the traps as mentioned above in relation to Eq. (9). Such instability after voltage stress under similar conditions as in the present experiment, was observed also by Bersuker *et al.* in [10]. The observation (2), can be interpreted as a result of the tunneling process involved in the electron emission. During the emission process, the negative charge in the transition region, born by the captured

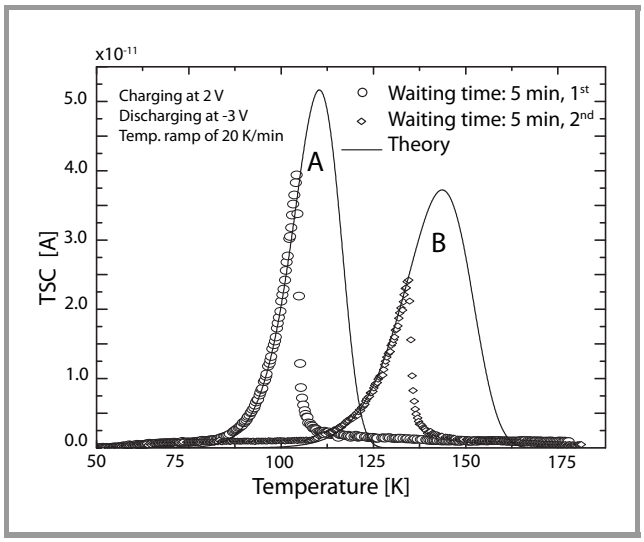


Fig. 5. Experimental TSC data (points) for MOS structures with HfO_2 prepared by reactive sputtering compared with theoretical calculations (solid curves) calculated from Eq. (9). Repeated measurement gave rise to a shift of the TSC peak along the temperature axis, reflecting structural changes of the emitting electron traps. The activation energy for the curves A and B is 0.13 eV and 0.16 eV, respectively, and the corresponding capture cross sections are $6 \cdot 10^{-22}$ cm^2 and $9 \cdot 10^{-21}$ cm^2, respectively.

electrons, will decrease. This will decrease the electric field driving the tunneling until the field strength is too weak for continued emission. As the tunneling probability is very sensitive to a change in electric field, the TSC will get an abrupt termination as shown for the two curves in Fig. 5.

4. Single Charge Versus Dipole Charge

The recent increasing interest in the properties of high-*k*/SiO_x transition layers includes novel ideas about a possible occurrence of closely separated charge planes with different polarities within this region [13]. It has been described as dipole planes occurring as a result of oxygen transfer across the interface between the high-*k* and the SiO_x material [14]. According to an idea proposed in [14], such transition would take place from the material with the highest surface density of oxygen atoms to that with lower density. For transition metal oxides, like HfO_2, this would imply that oxygen is transferred from this material into SiO_x, creating interstitials and leaving oxygen vacancies behind. As the relation between oxygen surface densities of rare-earth metal oxides, like Gd_2O_3 and SiO_x is the opposite, such transfer would instead go from the SiO_x to the high-*k* side.

The influence of a single negative charge and a dipole surface on the conduction band relations for the metal/HfO_2/SiO_x/n-type Si structure is depicted in Fig. 6. The geometries are shown in Fig. 6(a), while Fig. 6(b) and Fig. 6(c) show the charge relations and the conduction band relations for an open circuit case and a short circuit case, respectively. In order to demonstrate the specific influences of these charge planes, we assume that no other charges are present in the structure. Considering first the single charge case in the left column of Fig. 6 for an open circuit case in Fig. 6(b), where the opposite positive charge is assumed to exist at a long distance, the electron energy of the whole structure is lifted in parallel in relation to an earth plane. Short circuiting, as shown in Fig. 6(c), will cause positive charge to appear at the metal gate and in the depletion region occurring in the n-type semiconductor. Compared with an ideal structure without charge, an increased positive voltage, V_{FB}, on the gate would be needed to obtain flat-band condition for this case. Furthermore, V_{FB} would increase linearly with increasing thickness of the HfO_2 layer.

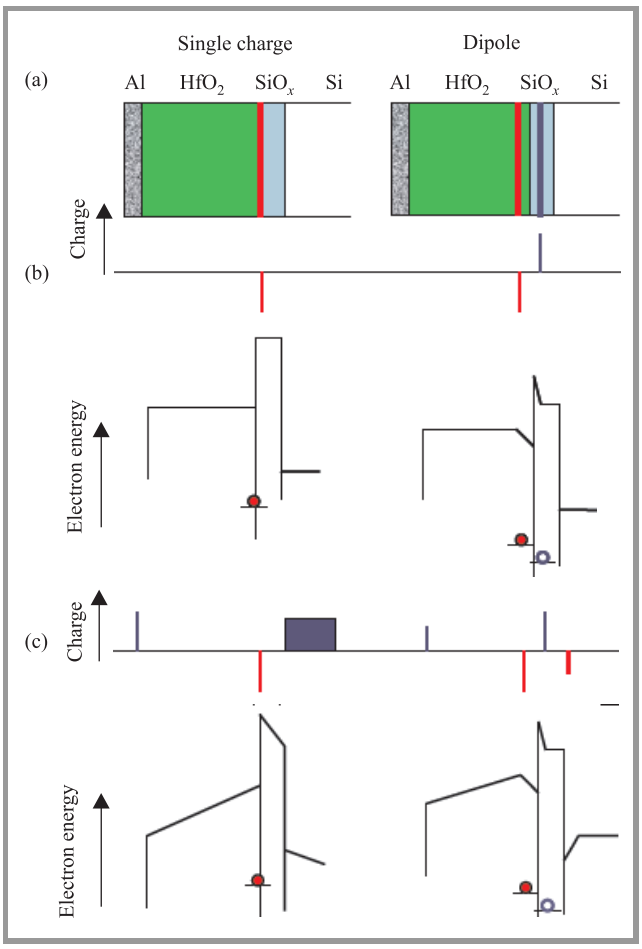


Fig. 6. Illustration of (a) charge configuration, and conduction band relations for (b) open circuit and (c) short circuit conditions of a metal/high-*k*/SiO_x/Si structure. The columns demonstrate the conditions for a single charge plane (left) and a dipole plane (right), respectively.

For a dipole layer at the HfO_2/SiO_x interface, as shown in the right column of Fig. 6, under open-circuit conditions, one would expect an electric field between the two charge planes only (Fig. 6(b)). This would create a potential drop in that domain, while constant potentials would occur outside the two planes separated by a voltage created inside the dipole. Short-circuiting this structure gives rise to a positive charge on the gate and a negative charge in

the semiconductor, forcing the latter into accumulation condition. These two charges are exactly equal and vary depending on the HfO_2 thickness in such a way that the sum of the energy drops in HfO_2 and in the semiconductor is constant and equal to qV_{FB}. In this case a negative V_{FB}, equal to the potential drop inside the dipole, is needed to obtain flat-band conditions. However, contrary to the situation with a single negative charge, the dipole combination would give a V_{FB} which is independent of the thickness of HfO_2. Such a behavior was indicated in [14] for HfO_x/Si_x interfaces.

The single charge case demonstrated by Fig. 6(a) offers a model for straightforward explanation of the anomalous TSC data as discussed in Section 3. The dipole case in Fig. 6(b) likewise gives rise to an energy relation of the conduction bands supporting electron injection into the silicon and thus a TSC with similar behavior as that from a single negative charge plane. Therefore, TSC does not give direct information on which of these two charge constellations is the source of current.

5. Discussion

In order to explain the saturating flat-band voltage and its positive sign as observed in [12] when increasing the thickness of HfO_2, the idea put forward in [13] requires injection of negatively charged traps from the high-k side into SiO_x. This would give rise to a dipole directed in the opposite way to that discussed in relation to Fig. 6. Such a model is not completely unproblematic. First, injecting an oxygen ion from the HfO_2 into the SiO_x layer might be expected to give rise to an interstitial, leaving behind a vacancy in the HfO_2. According to recent theoretical results [15], [16], vacancies are amphoteric and act as acceptors with energy levels at about 1.5 eV from the HfO_2 conduction band. They may therefore tend to be filled by electrons from the silicon conduction band and thus become negatively charged. On the other hand, the donor levels connected with oxygen vacancies are expected close to the middle of the bandgap of HfO_2 and are most probably occupied by electrons and neutral at voltages of the flat-band values at about 0.5 V in [13]. The origin of positive charge required on the HfO_2 side, therefore, is questionable.

A second problem with a dipole model may be the quantity of charge needed to obtain V_{FB} shifts in the range of 0.3 – 0.5 V as observed in [13]. As the thickness of the SiO_x layer normally is about 1 nm, the maximum value for the distance between the two charge sheets of an assumed dipole would be about that value. In order to achieve a shift of 0.3 V, this requires a charge density of more than $5 \cdot 10^{15} q$ As cm^{-2}. It can be compared with the charge contained in the TSC curves of Fig. 5, which is in the range of $10^{13} q$ As cm^{-2}. Finally, the strongest argument for a dipole model might be the saturating flat-band voltage. However, this result does not uniquely lead to a dipole configuration as shown in the following. Taking into account the continuous character of the transition region, a saturating V_{FB} may result also from single charge condition. Contrary to the assumption made in [13], the transition from HfO_2 to SiO_x cannot be considered abrupt on length scales in the range of a few nm. As noticed in Fig. 1, the oxygen concentration increases with distance from the silicon side due to diffusion into the silicon crystal. Similar behavior of oxygen concentration has been observed by elastic recoil detection analysis (ERDA) studies on HfPrO samples [26]. Therefore, it is reasonable to assume that the concentration of oxygen vacancies decreases with distance from the silicon side. Furthermore, assuming that thinner HfO_2 layers, due to the out-diffusion of oxygen to the surface, also as observed in Fig. 1, have a lower decay of vacancies, it is probable that they have a smaller decay in concentration.

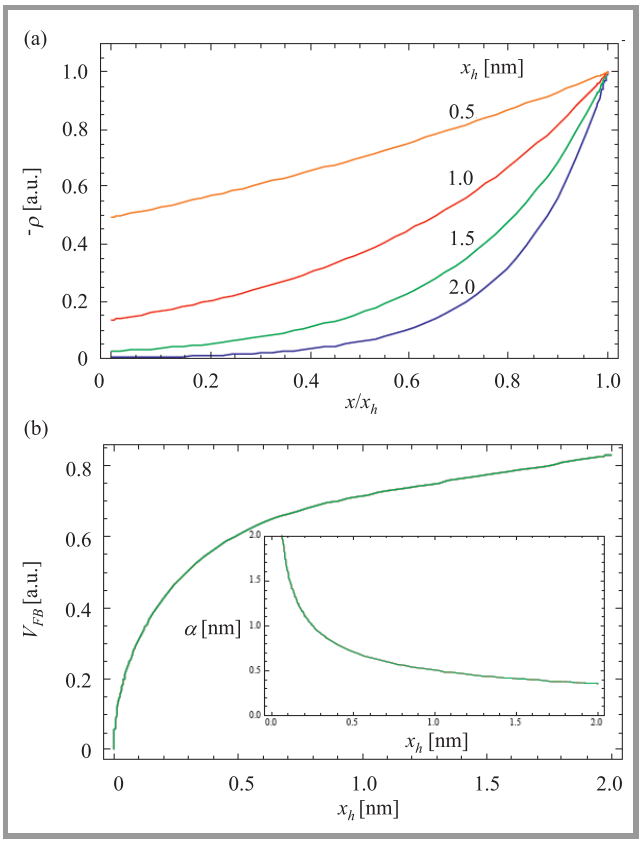


Fig. 7. (a) Assumed depth distribution of acceptor state volume concentrations in the transition region for HfO_2 with thicknesses of 0.5, 1.0, 1.5 and 2.0 nm. The point $x = 0$ is at the HfO_2 surface, while $x = x_h$ is at a reference point where SiO_2 transfers to suboxides. (b) The flat-band voltage as a function of HfO_2 thickness for acceptor concentrations as shown in (a) with the decay constant, α, varying as a function of x_h as shown in the inset of (b).

A theoretical example is shown in Fig. 7(a), where the concentration of traps is plotted for different thicknesses as a function of normalized distance from a defined reference point in the transition region. Such a point could be the one, where SiO_2 transfers to suboxides. For simplicity, the curves in Fig. 7(a) are exponential functions with decay factors, α, decreasing with the thickness x_h of HfO_2, as

shown in the inset of Fig. 7(b). Solving Poisson's equation for the functions in Fig. 7(a):

$$\frac{d^2V}{dx^2} = -\frac{\rho_0}{k\varepsilon_0}\exp\left(-\frac{x_h - x}{\alpha}\right), \tag{10}$$

where V is electrical potential, ρ_0 is the volume charge density at the reference point, x_h, ε_0 is the dielectric permittivity of vacuum and x is distance from the HfO_2 surface, we find V_{FB} from the potential at $x = 0$ as

$$\frac{k\varepsilon_0}{\rho_0}V_{FB} = -\alpha\left\{x_h + \alpha\left[1 - \exp\left(-\frac{x_h}{\alpha}\right)\right]\right\}. \tag{11}$$

The V_{FB} normalized by the pre-factor in Eq. (11) is plotted in Fig. 7(b) as a function of HfO_2 thickness, x_h. The shape of this curve, with a tendency to saturate for increasing, x_h, was interpreted in [12] as a result of a dipole layer. Figure 7(b) demonstrates that under assumptions based on MEIS data, it is possible to obtain similar behavior from a singly charged layer.

6. Conclusions

The theoretical TSC curves shown in Fig. 5 were calculated assuming an ensemble of electron states with one common discrete energy level. The agreement between the initial curvature of these graphs and the experimental data indicates that the majority of traps observed by TSC have similar properties of their charge carrier statistics and are positioned with limited spread in distance from the silicon/SiO_x interface. This excludes bulk traps in the HfO_2 as candidates and suggests a trap distribution concentrated at the HfO_2/SiO_x interface as depicted in Fig. 7(a) for delivering electrons to the TSC. Also, TSC peaks from states at the Si/SiO_x interface are ruled out by earlier experiments as they have been demonstrated to occur at temperatures above 200 K, outside the scale of Fig. 5 [12]. The extremely low values of capture cross sections in the range 10^{-26} – 10^{-19} cm^{-2}, normally found by TSC for this kind of experiments [11], [27], are explained by the tunneling process needed for carrier injection into the silicon substrate and the thermally activated process expected for carrier capture into oxygen vacancies [16]. Also, taking into account the flexible structural network anticipated in the transition region, the shift of the TSC peaks as shown in Fig. 5 may be assigned to the restructuring of oxygen vacancy defects as discussed in Section 2.

The argument in [13], for the existence of a dipole plane at the interface between HfO_2 and SiO_2 layers, was based on the observation of flat-band voltage as a function of the thickness of HfO_2. It was shown that V_{FB} saturated to a near constant value when this thickness increases as expected for a dipole charge and discussed above in connection with Fig. 6. However, such a result is not unique for a dipole. By assuming an oxygen vacancy distribution in the high-k layer as estimated from MEIS data, we have demonstrated in the present work that similar V_{FB} dependence may be the result of a single charge distribution. We conclude, therefore, that more evidence is necessary before the existence of a dipole layer occurring in high-k/SiO_x transition regions is confirmed.

Acknowledgment

The authors collaborate under the banner of the "High-k-Gang" (http://www.high-k-gang.eu/). The work has benefited from funding provided by the European Network of Excellence NANOSIL within FP7 (ICT-216171).

References

[1] O. Engström, B. Raeissi, S. Hall, O. Buiu, M. C. Lemme, H. D. B. Gottlob, P. K. Hurley, and K. Cherkaoui, "Navigation aids in the search for future high-k dielectrics: physical and electrical trends", *Solid-State Electron.*, vol. 51, iss. 4, pp. 622–626, 2007.

[2] B. Raeissi, J. Piscator, O. Engström, S. Hall, O. Buiu, M. C. Lemme, H. D. B. Gottlob, P. K. Hurley, K. Cherkaoui, and H. J. Osten, "High-k-oxide/silicon interfaces characterized by capacitance frequency spectroscopy", *Solid-State Electron.*, vol. 52, iss. 9, pp. 1274–1279, 2008.

[3] P. K. Hurley, K. Cherkaoui, E. O'Connor, M. C. Lemme, H. D. B. Gottlob, M. Schmidt, S. Hall, Y. Lu, O. Buiu, B. Raeissi, J. Piscator, O. Engström, and S. B. Newcomb, "Interface defects in HfO_2, $LaSiO_x$ and Gd_2O_3 high-k-metal-gate structures on silicon", *J. Electrochem. Soc.*, vol. 155, no. 2, pp. G13–G20, 2008.

[4] M. A. Quevedo-Lopez, P. D. Kirsch, S. Krishnan, H. N. Alshareef, J. Barnett, H. R. Harris, A. Neugroschel, F. S. Aguirre-Tostado, B. E. Gnade, M. J. Kim, R. M. Wallace, and B. H. Lee, "Systematic gate stack optimization to maximize mobility with HfSiON EOT scaling", in *Proc. ESSDERC 2006 Conf.*, Montreux, Switzerland, 2006, pp. 113–116.

[5] G. Lucovsky, Y. Wu, H. Niimi, V. Misra, and J. C. Phillips, "Bonding constraints and defect formation at interfaces between crystalline silicon and advanced single layer composite gate dielectrics", *Appl. Phys. Lett.*, vol. 74, no. 14, pp. 2005–2007, 1999.

[6] G. Lucovsky, J. P. Maria, and J. C. Phillips, "Interfacial strain induced self-organization in semiconductor dielectric gate stacks. II. Strain relief at internal dielectric interfaces between SiO_2 and alternative dielectrics", *J. Vac. Sci. Technol. B*, vol. 22, iss. 4, pp. 2097–2104, 2004.

[7] G. Lucovsky and J. C. Phillips, "Defects and defect relaxation at internal interfaces between high-k transition metal and rare earth dielectrics and interfacial native oxides in metal oxide semiconductor (MOS) structures", *Thin Solid Films*, vol. 486, iss. 1–2, pp. 200–204, 2005.

[8] F. Giustino, A. Bongiorna, and A. Pasquarello, "Equivalent thickness of thin oxide interlayer in gate insulator stacks on silicon", *Appl. Phys. Lett.*, vol. 86, iss. 19, pp. 192901-1--3, 2005.

[9] P. Broqvist and A. Pasquarello, "Band gaps and dielectric constants of amorphous hafnium silicates: a first principle calculation", *Appl. Phys. Lett.*, vol. 90, iss. 8, pp. 082907-1--3, 2007.

[10] G. Bersuker, C. S. Park, J. Barnett, P. S. Lysaght, P. D. Kirsch, C. D. Young, R. Choi, B. H. Lee, B. Foran, K. van Benthem, S. J. Pennycook, P. M. Lenahan, and J. T. Ryan, "The effect of interfacial layer properties on the performance of Hf-based gate stack devices", *J. Appl. Phys.*, vol. 100, p. 094108, 2006.

[11] J. T. Ryan, P. M. Lenahan, G. Bersuker, and P. Lysaght, "Electron spin resonance observations of oxygen deficient silicon atoms in the interfacial layer of hafnium oxide based metal-oxide-silicon structures", *Appl. Phys. Lett.*, vol. 90, p. 173513, 2007.

[12] B. Raeissi, Y. Y. Chen, J. Piscator, Z. H. Lai, and O. Engström, "Electron traps at HfO_2/SiO_x interfaces", in *Proc. ESSDERC 2008 Conf.*, Edinburgh, Scotland, 2008, pp. 130–133.

[13] K. Iwamoto, Y. Kamimuta, A. Ogawa, Y. Watanabe, S. Migita, W. Mizubayashi, Y. Morita, M. Takahashi, H. Ota, T. Nabatame, and A. Toriumi, "Experimental evidence for the flatband voltage shift of high-k metal-oxide-semiconductor devices due to the dipole formation at the high-k/SiO_2 interface", *Appl. Phys. Lett.*, vol. 92, iss. 13, pp. 132907-1--3, 2008.

[14] K. Kita and A. Toriumi, "Origin of electric dipoles formed at high-k/SiO_2 interface", *Appl. Phys. Lett.*, vol. 94, iss. 13, pp. 132902-1--3, 2009.

[15] K. Xiong, J. Robertson, M. C. Gibson, and S. J. Clark, "Defect energy levels in HfO_2 high-dielectric constant gate oxide", *Appl. Phys. Lett.*, vol. 87, iss. 18, pp. 183505-1--3, 2005.

[16] J. L. Gavartin, D. Muñoz Ramo, A. L. Shluger, G. Bersuker, and B. H. Lee, "Negative oxygen vacancies in HfO_2 as charge traps in high-k stacks", *Appl. Phys. Lett.*, vol. 89, iss. 8, pp. 082908-1--3, 2006.

[17] Y. P. Feng, A. T. Lim, and M. F. Li, "Negative-U property of oxygen vacancy in cubic HfO_2", *Appl. Phys. Lett.*, vol. 87, iss. 6, pp. 062105-1--3, 2005.

[18] H. D. B. Gottlob, M. Schmidt, A. Stefani, M. C. Lemme, H. Kurz, I. Z. Mitrovic, W. M. Davey, S. Hall, M. Werner, P. R. Chalker, K. Cherkaoui, P. K. Hurley, J. Piscator, O. Engström, and S. B. Newcomb, "Scaling potential and MOSFET integration of thermally stable Gd silicate dielectrics", *Microelectron. Eng.*, vol. 86, iss. 7–9, pp. 1642–1645, 2009.

[19] H. D. B. Gottlob, A. Stefani, M. Schmidt, M. C. Lemme, H. Kurz, I. Z. Mitrovic, M. Werner, W. M. Davey, S. Hall, P. R. Chalker, K. Cherkaoui, P. K. Hurley, J. Piscator, O. Engström, and S. B. Newcomb, "Gd silicate: a high-k dielectric compatible with high temperature annealing", *J. Vac. Sci. Technol. B*, vol. 27, iss. 1, pp. 249–252, 2009.

[20] I. Z. Mitrovic and S. Hall, "Rare earth silicate formation – a route towards high-k for the 22 nm node and beyond", *J. Telecommun. Inform. Technol.*, no. 4, pp. 51–60, 2009.

[21] O. Engström and A. Alm, "Energy concepts of insulator-semiconductor interface traps", *J. Appl. Phys.*, vol. 54, no. 9, pp. 5240–5244, 1983.

[22] W. B. Fowler, J. K. Rudra, M. E. Zvanut, and F. J. Fiegl, "Hysteresis and Franck-Condon relaxation in insulator-semiconductor tunnelling", *Phys. Rev. B*, vol. 41, no. 12, pp. 8313–8317, 1990.

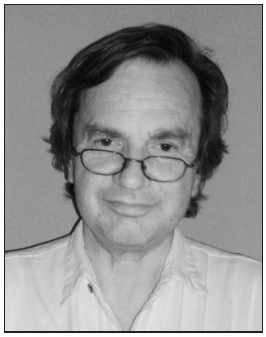
[23] O. Engström and H. G. Grimmeiss, "Vibronic states of silicon dioxide interface traps", *Semicond. Sci. Technol.*, vol. 4, no. 12, pp. 1106–1115, 1989.

[24] O. Engström, T. Gutt, and H. M. Przewłocki, "Energy concepts involved in MOS characterization", *J. Telecommun. Inform. Technol.*, no. 2, pp. 86–91, 2007.

[25] M. Johansson, "Silicon device substrate and channel characteristics influenced by interface properties", Ph.D. thesis, Chalmers University of Technology, Göteborg, Sweden, 2005.

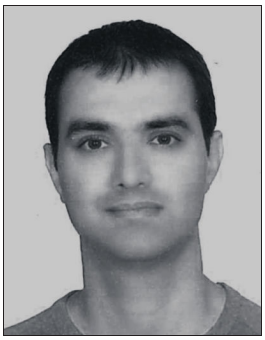
[26] B. Raeissi, "Charge carrier traffic at interfaces in nanoelectronic structures", Ph.D. thesis, Chalmers University of Technology, Göteborg, Sweden, 2010.

[27] M. Y. A. Yousif, M. Johansson, and O. Engström, "Extremely small hole capture cross sections in HfO_2/$Hf_xSi_yO_z$/p-Si structures", *Appl. Phys. Lett.*, vol. 90, iss. 20, pp. 203506-1--3, 2007.



Olof Engström received the Ph.D. degree in solid state physics from the University of Lund in 1975 and was later employed by ASEA AB for research on high power thyristors, by AB Rifa for development of MOS technology and by the Swedish Defence Research Institute for sensor research. In 1984, he came to Chalmers University of Technology as a Professor in solid state electronics. Between 1996 and 1999, he served as Dean of Chalmers School of Electrical and Computer Engineering and 1999–2002 he was the Director of the Microtechnology Center at Chalmers (MC2). From 2003 he is back in research as a Professor at the Department of Microtechnology and Nanoscience of MC2. His present research interest is in, high-k-materials and quantum dots and Schottky barriers on silicon nanowires. In 1991, he founded Samba Sensors AB, a company for production of fiberoptical pressure sensors. He is a member of the Royal Swedish Academy of Engineering Science, the Finnish Society of Science and the High-k-Gang.

e-mail: olof.engstrom@chalmers.se
Chalmers University of Technology
Department of Microtechnology and Nanoscience
SE-412 96 Göteborg, Sweden



Bahman Raeissi received the B.Sc. degree in physics from Isfahan University of Technology, Isfahan, Iran, in 2002 and M.Sc. degree in nanoscale physics and engineering, Chalmers University of Technology, Göteborg, Sweden, in 2005. He is currently a Ph.D. student at the Department of Microtechnology and Nanoscience (MC2), Chalmers University of Technology. His main research interests are fabrication and characterization of high-k dielectrics, semiconductor wafer bonding and characterization of quantum dots.

e-mail: bahman.raeissi@chalmers.se
Chalmers University of Technology
Department of Microtechnology and Nanoscience
SE-412 96 Göteborg, Sweden

Johan Piscator received the M.Sc. degree and the Ph.D. degree in electronic engineering from Chalmers University of Technology, Göteborg, Sweden, in 2003 and 2009, respectively. His main research interests are fabrications and characterization of silicon nanowires and characterization of high-k oxides for CMOS applications.

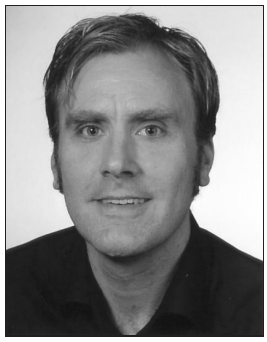
e-mail: johan.piscator@gmail.com
Chalmers University of Technology
Department of Microtechnology and Nanoscience
SE-412 96 Göteborg, Sweden

Ivona Z. Mitrovic received the Ph.D. degree in electronic engineering from the University of Liverpool, UK, in 2007, the M.Sc. degree in materials science from the University of Belgrade in 2002, and Dipl.-Ing. degree in microelectronics from the Faculty of Electronic Engineering, University of Nis, Serbia, Yugoslavia, in 1997. She took part in a research project concerning $BaTiO_3$ ceramics (1997–2001), worked as a Research Assistant (2001–2007) and a Research Associate at the University of Liverpool (2000–2009). Since June 2009, she is a Lecturer in the Solid State Electronics Research Group, Department of Electrical Engineering and Electronics, University of Liverpool. Her research interests span materials for beyond 22 nm technological node targeting energy harvesting products for medical, automotive and aerospace applications, as well as emerging technologies for energy conversion and storage.

e-mail: ivona@liverpool.ac.uk
Department of Electrical Engineering and Electronics
University of Liverpool
Brownlow Hill
Liverpool L69 3GJ, United Kingdom

Stephen Hall has been Head of the Department of Electrical Engineering and Electronics at the University of Liverpool, UK, from 2001 to the present date. He has interests spanning materials characterization, device physics and innovative device design and gate level circuits. He has over 200 conference and journal papers in the area of silicon technology, devices and circuits. These include novel measurements and contributions to the understanding of MOS related interfaces and materials quality. He has successfully designed and built novel MOS and bipolar devices in silicon for about 20 years. More recently, his work encompasses gate level circuits relating to low voltage/low power SOI, micro-power and biologically inspired concepts. He was Technical Programme Chair of ESSDERC 2008, and currently sits on the Steering Committee of ESSDERC/ESSCIRC and INFOS, for which he was vice-Chair in 2009 and is a member of the Steering Committee from 2009.

e-mail: S.Hall@liverpool.ac.uk
Department of Electrical Engineering and Electronics
University of Liverpool
Brownlow Hill
Liverpool L69 3GJ, United Kingdom



Heinrich D. B. Gottlob received his Dipl.-Ing. degree in electrical engineering and information technology from the RWTH Aachen University, Germany, in 2003 and his Dr.-Ing. degree in 2007. In 2003 he joined AMO GmbH as R&D engineer and he is currently leader of the nanoelectronics group and manager of AMO's nanotechnology lab AMICA. Research topics include novel high-*k* dielectrics and metal gate stacks as well as fully depleted ultrathin body and non-planar SOI devices. He is author and co-author of more than 30 journal papers.

e-mail: gottlob@amo.de
Advanced Microelectronic Center Aachen (AMICA)
AMO GmbH
Otto-Blumenthal st 25
D-52074 Aachen, Germany

Mathias Schmidt received his Dipl.-Ing. degree in electrical engineering from the RWTH Aachen University, Germany, in 2004. He is currently working towards his Dr.Ing. degree at the Advanced Microelectronic Center Aachen (AMICA), AMO GmbH Aachen. His research topics include ultrathin body silicon devices with planar and non-planar channels with special respect to experimental mobility extraction and integration of novel high-*k*/metal gate stacks and strained channel materials.

e-mail: schmidt@amo.de
Advanced Microelectronic Center Aachen (AMICA)
AMO GmbH
Otto-Blumenthal st 25
D-52074 Aachen, Germany

Karim Cherkaoui received the Ph.D. degree in 1998 from the Institut National des Sciences Appliquées of Lyon, France. During his Ph.D. he gained experience in the spectroscopy of point defects in semi-insulating materials. In 1999 he joined the Tyndall National Institute (formerly NMRC), University College Cork, where he has established and developed several low temperature electrical metrology techniques. His current research interests include the process development and characterization of high dielectric constant materials on Si and III-V materials for future MOS devices.

e-mail: karim.cherkaoui@tyndall.ie
Tyndall National Institute
University College Cork, Lee Maltings
Prospect Row, Cork, Ireland

Paul Hurley received his Ph.D. (1990) and B.Eng. (1985 – first class honors) in electronic engineering at the University of Liverpool, UK. He is a senior research scientist at the Tyndall National Institute, University College Cork, Ireland, where his work focuses on high dielectric constant (high-k) materials intended for use as gate level insulators in transistors for future integrated circuits. The emphasis of his current research is the formation and characterization of high-k films on silicon and III–V semiconductor substrates. He is a member of the Technical Committee of the Insulating Films on Semiconductors (INFOS) conference and the International Workshop on Dielectrics in Microelectronics (WoDiM). In addition to research activities, he is a part time lecturer in the Department of Microelectronic Engineering at University College Cork. He has published over eighty papers in the field of microelectronics.

e-mail: paul.hurley@tyndall.ie
Tyndall National Institute
University College Cork, Lee Maltings
Prospect Row, Cork, Ireland

Paper

Novel Method of Improving Electrical Properties of Thin PECVD Oxide Films by Fluorination of Silicon Surface Region by RIE in RF CF_4 Plasma

Małgorzata Kalisz, Grzegorz Głuszko, and Romuald B. Beck

Abstract—This study describes a novel technique to form good quality low temperature oxide ($< 350^\circ$C). Low temperature oxide was formed by N_2O + SiH_4:N_2 plasma in a plasma enhanced chemical vapour deposition (PECVD) system on the silicon surface reactively etched in CF_4 plasma (RIE – reactive ion etching). The fabricated oxide demonstrated excellent (for low temperature dielectric formation process) current-voltage (I–V) characteristics, such as: low leakage current, high breakdown voltage and good reliability. Experimental results indicate that the proposed method of fluorine incorporation into the SiO_2/Si inteface improves electrical parameters of MOS structures.

Keywords—capacitance-voltage characteristics, current-voltage characteristics, fluorine plasma, radio frequency reactive ion etching.

1. Introduction

As semiconductor devices are scaled down to obtain higher-performance ultra-large-scale integration (ULSI) devices, the realibility of gate-oxide films is one of the most important issues. The downscaling of gate oxide thickness improves current driving capability and reduces short-chanel effects. However, ultrathin oxide films exhibit many serious reliability problems, such as, time-dependent dielectric beakdown, interface-state generation and charge trapping (e.g., [1]). Fluorination of the gate oxide structures has been investigated as a possible candidate for solution of these problems [2]–[8]. Many aspects of the properties of fluorinated oxides have already been studied. Such oxides have been found, for example, to be more resistant to ionizing radiation [2], Fowel-Nordheim (F-N) tunneling injection stress [1]–[3] and channel hot electron stress [4]. Dramatic reduction of both hole-trapping probability and interface-trap generation under avalanche hole injection conditions of the fluorinated samples has also been reported [6]. In [7] it has been found that the degree of improvement is a function of fluorine concentration, which has created a pressure to obtain high concentrations of fluorine in silicon oxide.

A number of methods of fluorinated gate oxide fabrication has been proposed, e.g., by immersing Si wafer in HF solution with D.I. water rinse prior to oxidation [8], by ion implantation of fluorine atoms into poly-Si gate followed by a high temperature drive-in [2], or by rapid thermal processing in O_2 with diluted NF_3 [7].

An attractive method to fabricate high quality thin fluorinated gate oxides is conventional plasma enhanced chemical vapour deposition (PECVD) oxide on silicon substrate pretreated with CF_4 plasma without subsequent annealing [8]. As it has already been established in [8] application of CF_4 plasma pretreatment to the silicon surface before the PECVD gate oxide formation improved largely almost all electrical parameters, e.g., the Q_{bd} distribution, the V_{th} value.

In our recent study, it has been reported that reactive ion etching (RIE) in CF_4 plasma is a good method to incorporate high concentrations of fluorine ions into silicon surface, thus it may be considered as yet another serious candidate for fluorination process [9].

This study, in turn, investigates the structural and electrical characteristics of metal-oxide-semiconductor (MOS) structures with thin fluorinated gate oxide prepared by means of silicon dioxide RIE in CF_4 plasma prior to the deposition of gate oxide.

2. Experiments

In this work, two types of gate oxide fabrication methods were compared and investigated: PECVD deposited oxide (control – sample 4 in Table 1) and oxide deposited in a conventional PECVD tool on a surface etched in CF_4 plasma (samples 1, 2, and 3 in Table 1).

The p-type, boron-doped (100)-oriented silicon wafers with resistivity of 4–10 Ωcm were cleaned using standard procedures (SC1+SC2+HF).

Both, PECVD and RIE processes were performed in conventional RF Oxford PlasmaLab Systems. First, 13 nm thick initial PECVD oxide was deposited by at 300°C for 30 s with RF power of 10 W. The SiO_2 film was then reactively etched for 2 min in a RIE tool at room temperature by CF_4 flowing at the rate of 50 ml/min. In the experiments the RF power was set to: 80 W, 120 W and 160 W.

The complete set of experiments performed in this study is shown in Table 1.

For all samples except of the control one, fluorine distribution profiles and their concentration in the dielectric

Table 1
Matrix of experiment and process parameters

Type of sample	PECVD initial oxide reactive ion etching in RF CF_4 plasma				Final SiO_2 deposition using PECVD				
	flow of CF_4 [ml/min]	pressure [mTr]	time [min]	RF power [W]	temperature [°C]	RF power [W]	pressure [mTr]	gas flow rates [ml/min]	time [s]
1	50	200	2	80	350	10	600	$N_2O = 120$ SiH_4:$N_2 = 70$	30
2	50	200	2	120					
3	50	200	2	160					
4	–	–	–	–					

layer have been measured by ultra low energy-secondary ion mass spectroscopy (ULE-SIMS).

In order to use electrical characterization methods, MOS test structures were fabricated with the layers under investigation as gate dielectrics. Aluminium was used as gate metal and bottom electrode, allowing reliable electrical measurements.

3. Results and Discussion

3.1. SIMS Characterization of Fluorine Content

It has already been established in [9] that reactive ion etching in fluorine plasma is a good method of fabricating layers containing fluorine atoms in the silicon substrate surface region. These layers are thin (about 1.5 nm) and the concentration of fluorine incorporated into them is very high, with the maximum values exceeding 10^{19} cm^{-3}. What is important – the fluorine incorporation in the substrate may be controlled by RIE parameters, e.g., RF power or reactive gas pressure.

As it has been shown in [10], the PECVD gate oxide deposition (at 300°C) following RIE does not change the position of this profile but affects the maximum of fluorine concentration only. As a result of the deposition of silicon dioxide on the etched surface, the maximum of fluorine concentration decreases for all characterized samples fabricated within this study to about 10^{18} cm^{-3}. This suggests that due to the elevated temperature of deposition (300°C) some of the fluorine atoms can escape from the fluorinated film. Interestingly, the fluorine concentration is still significantly higher (by several orders of magnitude) than that obtained with other methods. In this way a structure consisting of silicon dioxide/oxide fluorine rich thin film/silicon (Fig. 1) is formed.

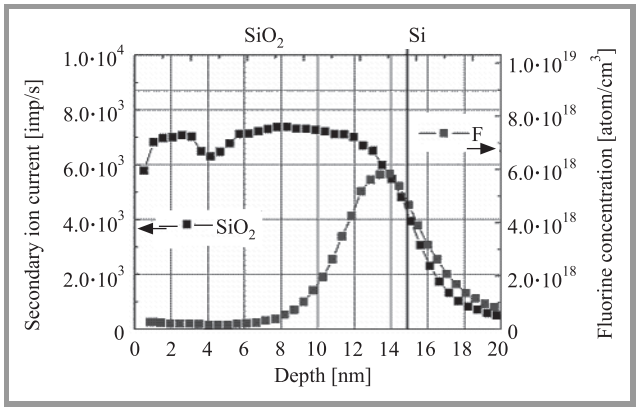


Fig. 1. SIMS profiles for structure consisting of silicon dioxide/thin fluorine-rich/silicon obtained as a result of initial oxide etching by RIE in CF_4 followed by PECVD oxide formation.

It should be stressed that although the fluorine concentration decreases after the final silicon oxide layer depositon, the RIE process still makes it possible to control (to a certain extent) the position of the fluorine profile and the value of its maximum concentration (Fig. 2). For electrical characterization discussed below we used the sample with the highest fluorine concentration at the dielectric/silicon interface of a MOS structure fabricated with RF power equal to 160 W (sample 3).

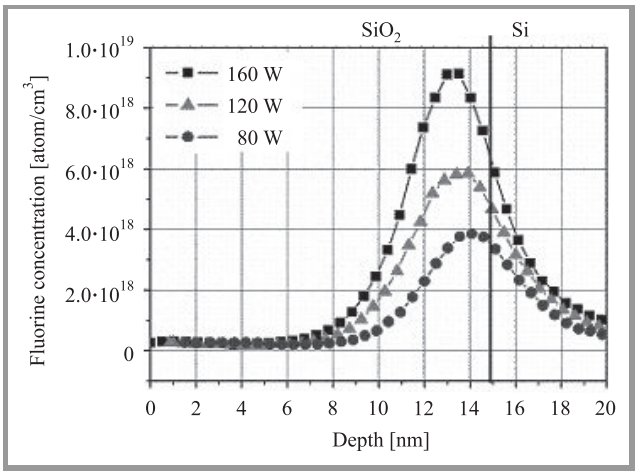


Fig. 2. Fluorine distribution profile in samples fluorinated in RIE with different RF plasma power prior to PECVD oxide deposition.

3.2. Electrical Characterization

The effect of fluorine incorporation into the gate oxide was clearly noticeable in the electrical characteristics: capacitance-voltage (*C–V*) and current-voltage (*I–V*) of MOS test capacitors.

3.2.1. Analysis of *C–V* Characteristics

As expected, the control samples with PECVD gate oxide only are characterized by very high shift of *C–V* curves towards negative voltages. *C–V* characteristics of these samples exhibit also a strong frequency dependence in the strong inversion region – see Fig. 3(a).

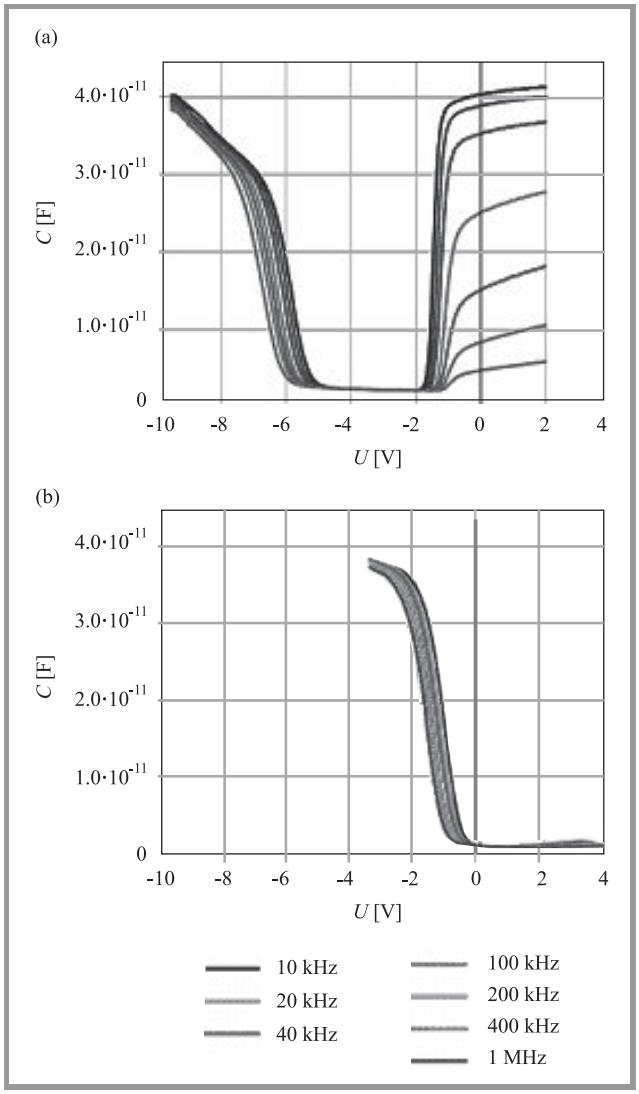


Fig. 3. The $C-V$ characteristics for MOS test structures with gate oxide prepared in different ways: (a) control sample (PECVD only) and (b) fluorinated by RIE process (RF power 160 W) + PECVD oxide.

The incorporation of fluorine by means of RIE reduces the voltage shift significantly (compare Fig. 3(a) with Fig. 3(b)). Another effect of fluorine incorporation into the interface of a MOS structure is the disappearance of frequency dispersion in *C–V* curves – see Fig. 3(b).

In Table 2 electro-physical parameters evaluated from *C–V* characteristics of the MOS test devices formed during experiment are presented.

It can be seen there, that all the measured samples have negative flat band voltage (V_{fb}) values. For the control oxide sample, the negative (V_{fb}) is very high. The possible reason is that after RIE in CF_4 plasma, we can expect poorer quality of the substrate surface due to the damage of the initial oxide layer. The very low temperature of the gate dielectric PECVD deposition (300°C) does not allow for significant improvement of the oxide quality due to thermally related effects – the temperature is simply too low. Consequently, for these samples, we can expect high effective densities of the total non-compensated charge in the oxide-silicon system (Q_{eff}) and interface traps density (e.g., D_{itmb}).

Table 2
Electrical parameters evaluated from *C–V* characteristics of the fabricated test structures

Method of fluorination	Control	RIE
C–V curves analysis at $\varepsilon = 3.9$		
N_A [cm^{-3}]	$1.7 \cdot 10^{15}$	$0.43 \cdot 10^{15}$
V_{fb} [V]	−4.15	−1.42
V_{mb} [V]	−3.4	−0.95
D_{itmb} [1/eV cm^2]	$10.3 \cdot 10^{11}$	$5.7 \cdot 10^{11}$
Q_{eff}/q [cm^{-2}]	$35.8 \cdot 10^{11}$	$7.7 \cdot 10^{11}$

In the studied samples substrate fluorination has improved the properties of both, the oxide/silicon interface and the oxide bulk (see Table 2), showing that curing mechanisms resulting form the presence of fluorine prevailed over the effects resulting from the damage created during the fluorination (RIE).

During fluorination, high energy fluorine ions break O-Si-O bonds and form two types of dangling bonds: Si and Si-O. In the mean time, fluorine ions react with dangling Si bonds and Si-O bonds and form SiF and SiOF films, which passivate the modifited surface.

The same type of dependence on fluorine concentration is observed for Q_{eff}.

As it has already been established in [10], for the sample prepared by means of RIE, the maximum of fluorine concentration is located in the oxide. Fluorine ions located in the silicon dioxide layer not only passivate SiO_2/Si interface but also react with the structural defects and damages of silicon dioxide. Therefore, the value of Q_{eff} is lower for the sample with fluorine incorporated into the interface of MOS structure than for the control sample.

3.2.2. Analysis of *I−V* Characteristics

As it can be seen from Fig. 4 fluorinated PECVD gate oxide exhibits much better properties than the control one. The very bad breakdown statistics indicate that the uniformity of the PECVD oxide is poor. On the other hand, relatively low leakage currents (until breakdown) prove, that the control PECVD oxide does not contain many structural defects that would contribute to the leakage currents.

The observed changes can be considered in terms of breakdown statistics and leakage current.

For RIE fluorination, we have observed very significant rise in the mean breakdown voltage value and very good

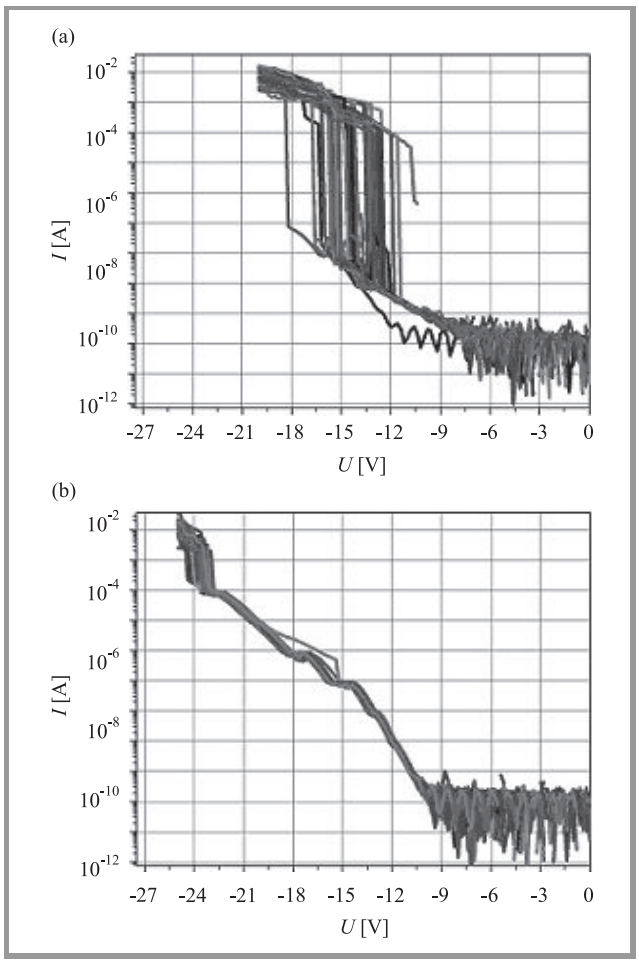


Fig. 4. The $I-V$ characteristics measured on MOS test structures with gate oxide prepared in different ways: (a) control sample (PECVD only) and (b) fluorinated by RIE process (RF power 160 W) + PECVD oxide.

statistics of breakdowns, while the leakage currents were still comparable with those of the control samples. All these parameters are very important for the feasibility of manufacturing real devices, as the requirements resulting from international technology for semiconductors (ITRS) have been challenging in this area for many years already.

4. Conclusions

The results obtained in this study show that the examined method of fluorination (RIE in CF_4 plasma of initial oxide) allows achieving very high concentrations of fluorine at the silicon surface region (of the order of $10^{19} - 10^{20}$ cm^{-3}) and the following gate oxide deposition by means of PECVD at 300°C does not reduce this concentration by more than one order of magnitude.
It is also clear from the presented results, that the presence of fluorine in such quantities at the oxide-silicon interface results in significant improvement of the electrical properties of otherwise poor-quality PECVD oxide.

The obvious results of silicon surface fluorination by means of RIE in CF_4 plasma prior to gate oxide deposition by PECVD are:

- significant reduction of both, Q_{eff} and D_{itmb};
- removal of the frequency dispersion in the inversion region of $C-V$ curves;
- increased breakdown voltage;
- significantly improved breakdown statistics.

References

[1] Y. Mitani, H. Satake, Y. Nakasaki, and A. Toriumi, "Improvement of charge-to-breakdown distribution by fluorine incorporation into thin gate oxides", *IEEE Trans. Electron Dev.*, vol. 50, no. 11, pp. 2221–2226, 2003.

[2] Y. Nishioka, K. Ohyu, Y. Ohji, N. Natuaki, K. Mukai, and T. P. Ma, "Hot-electron hardened Si-gate MOSFET utilizing F implantation", *IEEE Electron Device Lett.*, vol. 10, no. 4, pp. 141–143, 1989.

[3] P. Wrigh, N. Kasai, S. Inoue, and K. C. Saraswat, "Hot electron immunity of SiO_2 dielectrics with fluorine incorporation", *IEEE Electron Dev. Lett.*, vol. 10, no. 8, pp. 347–348, 1989.

[4] X. W. Wang, A. Balasinski, T. P. Ma, and Y. Nishioka, "Pre-oxidation fluorine implantation in Si process-related MOS characteristics", *J. Electrochem. Soc.*, vol. 139, no. 1, pp. 238–241, 1992.

[5] H. H. Tseng, P. J. Topin, F. K. Baker, J. R. Pfiester, K. Evans, and P. L. Fejes, "The effect of silicon gate microstructure and gate oxide process on threshold voltage instabilities in P+ gate chanel MOSFET's with fluorine incorporation", *IEEE Trans. Electron Dev.*, vol. 39, no. 7, pp. 1687–1693, 1992.

[6] E. F. da Silva Jr., Y. Nishioka, and T. P. Ma, "Radiation response of MOS capacitors containing fluorinated oxide", *IEEE Trans. Electron Dev.*, vol. 34, no. 6, pp. 1190–1195, 1987.

[7] J. Ahn, G. Q. Lo, W. Ting, D. L. Kwong, J. Kuehne, and C. W. Magee, "Radiation hardered metal-oxide-semiconductor devices with gate dielectrics grown by rapid thermal processing in O_2 with diluted NF_3", *Appl. Phys. Lett.*, vol. 58, no. 4, pp. 425–427, 1991.

[8] J-W. Lee, Y. Li, and S. M. Sze, "Highly reliable low temperature ultrathin oxides grown using N_2O plasma", *WSEAS Trans. Electron.*, vol. 1, no. 1, pp. 72–76, 2004.

[9] M. Kalisz, R. B. Beck, and M. Ćwil, "Reactive ion etching process (RIE) in CF_4 plasma as a method of fluorine implantation", *Vaccum*, vol. 82, no. 10, pp. 1046–1050, 2008.

[10] S. K. Lai and D. R. Young, "Effects of avalanche injection of electrons into silicon-dioxide – generation of fast and slow interface states", *J. Appl. Phys.*, vol. 52, no. 10, pp. 6231–6240, 1981.

Małgorzata Kalisz was born in Żarów, Poland, in 1977. She received the M.Sc. degree in electronics and optoelectronics material science in 2003 from the Department of Fundamental Problems of Technology, Wrocław University of Technology, Poland. She received the Ph.D. degree in microelectronics in 2009 from the Faculty of Elec-

tronics and Information Technology, Warsaw University of Technology. Since 2003, she has been with the Motor Transport Institute (ITS), Warsaw. Her research interests include the role of fluorine-containing ultrathin layer, formed on the surface during silicon and silicon dioxide reactive ion etching on the depth of boron thermal diffusion into silicon and electrical parameters of MOS structures.

e-mail: mkalisz1@mion.elka.pw.edu.pl
Institute of Microelectronics and Optoelectronics
Warsaw University of Technology
Koszykowa st 75
00-662 Warsaw, Poland

e-mail: malgorzata.kalisz@its.waw.pl
Motor Transport Institute
Jagiellońska st 80
03-301 Warsaw, Poland

Grzegorz Głuszko was born in Biłgoraj, Poland, in 1976. He received the B.Sc. and M.Sc. degrees in microelectronics from Warsaw University of Technology in 2002 and 2004, respectively. Since then he has been working on a Ph.D. thesis concerning characterization of novel MOS devices. His research interests include charge pumping studies of the quality of silicon-silicon dioxide interface, as well as modeling of MOS devices.

e-mail: ggluszko@poczta.onet.pl
Institute of Microelectronics and Optoelectronics
Warsaw University of Technology
Koszykowa 75
00-662 Warsaw, Poland

Romuald B. Beck received the M.Sc. degree in electronics from the Faculty of Electronics, Warsaw University of Technology, Poland, in 1976. From the same university, he received the Ph.D. and D.Sc. degrees in 1982 and 1996, respectively. Since 2000 he has taken the post of a Professor. Since 2005 he has been heading Microelectronic and Nanoelectronic Devices Division. His research activities have been concentrated in the area of modeling, diagnostics and technology of the metal-insulator-semiconductor devices, especially – with very thin and ultrathin oxides. In particular he is concerned with theoretical and experimental studies on different dielectric layers formation methods (including methods involving d.c. and a.c. plasmas), their kinetics and the relations between the process kinetics and the electrophysical properties of the devices, their yield and their reliability. The area of his research interest covers also dry etching methods and their implementation to modern ICs technology. In case of plasma processing he has been actively engaged in the research work on negative consequences of the use of plasma (radiation damage and contamination) and experiments aiming at curing the structural and electrophysical damage related to plasma ambient and e-beam bombardment. His research interest covers also design of novel test structure design and new diagnostics methods.

e-mail: r.beck@elka.pw.edu.pl
Institute of Microelectronics and Optoelectronics
Warsaw University of Technology
Koszykowa 75
00-662 Warsaw, Poland

Paper

The Effect of High Temperature Annealing on Fluorine Distribution Profile and Electro-Physical Properties of Thin Gate Oxide Fluorinated by Silicon Dioxide RIE in CF_4 Plasma

Małgorzata Kalisz, Grzegorz Głuszko, and Romuald B. Beck

Abstract—This study describes the effects of high temperature annealing performed on structures fluorinated during initial silicon dioxide reactive ion etching (RIE) process in CF_4 plasma prior to the plasma enhanced chemical vapour deposition (PECVD) of the final oxide. The obtained results show that fluorine incorporated at the PECVD oxide/Si interface during RIE is very stable even at high temperatures. Application of fluorination and high temperature annealing during oxide layer fabrication significantly improved the properties of the interface (D_{itmb} decreased), as well as those of the bulk of the oxide layer (Q_{eff} decreased). The integrity of the oxide (higher V_{bd}) and its uniformity (V_{bd} distribution) are also improved.

Keywords—capacitance-voltage characteristics, current-voltage characteristics, fluorine plasma, high temperature annealing process, radio frequency reactive ion etching.

1. Introduction

A key issue of ultra-large-scale-intergation technology (ULSI) is the quality of thin silicon dioxide layer used in the devices. Today, oxide thickness less than 80 Å and oxidation temperatures at about 800–900°C are required as the semiconductor industry pushes toward 1 Gbit memory chips and beyond. However, good electrical properties generally require high temperature conditions either during oxidation or as postoxidation annealing (POA) [1]. At low temperatures, ultrathin SiO_2 films prepared by the conventional plasma enhanced vapour deposition (PEVD) process do not represent satisfactory properties, e.g., they are characterized by high leakage currents and high fixed charge density. It has been demonstrated, e.g., in [1]–[6] that incorporation of small amounts of fluorine into SiO_2 can be an efficient way to improve the electrical parameters of SiO_2/Si interface.

Several ways of fluorine introduction into gate oxide, including among others: in-situ NF_3 oxidation [4] and ion implantation [5], have been tried so far.

Another useful technique to improve the quality of low temperatrure SiO_2 is high thermal annealing process.

It, thus, seems tempting to apply both of these techniques in order to get the best results possible. The problem with such an approach is that high temperature annealing of fluorinated oxides fabricated by means of the above mentioned processes seriously affects the F concentration in the oxides – as fluorine atoms escape from fluorinated gate oxides during high temperature annealing [7].

In our previous work, it has been reported that fluorination by means of reactive ion etching (RIE) in CF_4 plasma is an efficient method of manufacturing quite good quality oxides [8].

In this work, we study the effects of high temperature annealing on structures fluorinated during initial silicon dioxide reactive ion etching in CF_4 plasma prior to the plasma enhanced chemical vapour deposition (PECVD) of the final oxide.

2. Experiments

The p-type, boron-doped (100)-oriented silicon wafers with the resistivity of 4–10 Ωcm were cleaned using standard procedures prior to oxidation.

The PECVD and RIE processes were performed in conventional RF Oxford PlasmaLab systems.

The 13 nm thick initial oxide was deposited by means of PECVD at 300°C for 30 s with RF power equal to 10 W. The obtained SiO_2 film was then reactively etched for 2 min in a RIE tool at room temperature in CF_4 (50 ml/min) plasma generated with 160 W RF signal. Afterwards, the final oxide layer was deposited during PECVD at 300°C. Then, in split experiment, some of the samples were annealed in Ar at 1100°C for 30 min. The complete matrix of experiments is shown in Table 1.

For all samples, except of the control one, the fluorine distribution profiles and their concentration in the dielectric layer have been measured by ultra low energy-secondary ion mass spectroscopy (ULE-SIMS).

In order to use electrical characterization methods for evaluation of electro-physical properties, metal-oxide-semiconductor (MOS) test structures were fabricated with

Table 1
Parameters of all processes used during experiments

Type of sample	Reactive ion etching (RIE) in RF CF_4 plasma				Final SiO_2 deposition in PECVD					Thermal annealing	
	flow of CF_4 [ml/min]	pressure [mTr]	time [min]	RF power [W]	temperature [°C]	RF power [W]	pressure [mTr]	gas flow [ml/min]	time [s]	temperature [°C]	time [min]
1	50	200	2	160	300	10	600	$N_2O = 120$ $SiH_4 = 70$	30	–	–
2	50	200	2	160						1100	30
3	–	–	–	–						–	–
4	–	–	–	–						1100	30

the layers under investigation as gate dielectrics. For the purposes of comparison, the reference samples, without fluorination of the SiO_2/Si interface, were also made.

3. Results and Discussion

3.1. SIMS Characterization of Fluorine Content

It has already been established before (e.g., in [9]) that reactive ion etching in fluorine plasma can be effectively used for fabricating layers containing high concentrations of fluorine atoms. Direct application of high temperature to such a layer causes fluorine atoms to escape from the fluorinated film. Thin film of silicon oxide deposited at low temperature (300°C) on top of the fluorine-rich layer prevents them from escaping from the fluorinated layer during the subsequent high temperature annealing process.

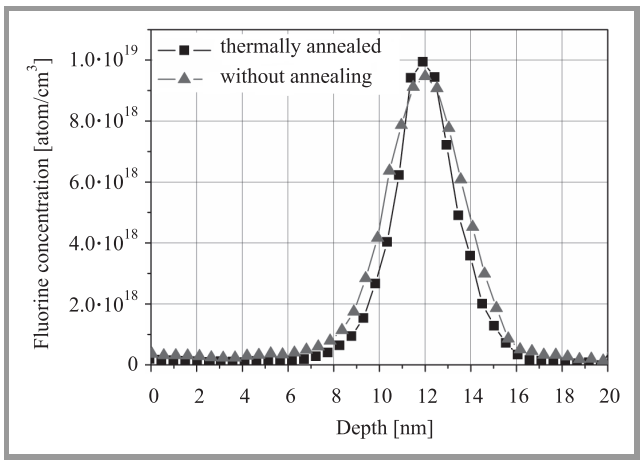


Fig. 1. The comparison of fluorine distribution profiles before and after high temperature annealing.

As presented in Fig. 1, the high temperature annealing performed in Ar at 1100°C, for 30 min had almost no effect on fluorine profile in the studied structures. The only change observed is marginal sharpening of the fluorine profile. This indicates that fluorine remains stable during high temperature annealing at its original location at the PECVD SiO_2/Si interface.

3.2. Electrical Characterization

The effect of fluorine incorporation into deposited silicon dioxide and of high temperature annealing was easily noticeable on both types of the electrical characteristics studied, i.e., capacitance-voltage ($C-V$) and current-voltage ($I-V$) characteristics of test metal-oxide-semiconductor capacitors.

3.2.1. Analysis of *C–V* Characteristics

As expected, the very high temperature annealing causes improvement of the C–V curves in both of the studied cases (Fig. 2). This improvement seems to be more spectacular for non-fluorinated sample in which apart from the decrease of the voltage shift (expressed, e.g., in terms $V_{fb} = 0$) also the frequency dispersion in the inversion region has been significantly reduced.

One may, however, notice also some consequences of high temperature annealing on the fluorinated samples.

The beneficial effects of high temperature annealing for attaining more robust oxides have been attributed to its possible ability to remove defects and damages existing in the bulk of the PECVD silicon dioxide layer. Application of high temperature annealing causes a reduction of both, D_{itmb} and Q_{eff} (see Table 2 and Fig. 3).

Similarly to the results presented in [8], within the course of this work it has been established that the fluorine atoms incorporated into the interface of SiO_2/Si improve the electrical parameters of silicon dioxide layer. However, the result of the high temperature annealing is still clearly noticeable on the measured $C-V$ curves as well as in the values of the evaluated electrical parameters, i.e., D_{itmb} and Q_{eff}. This means that the latter process is still capable of removing some defects existing in the bulk of PECVD silicon dioxide layer.

It has to be realized, however, that the degree of improvement due to high temperature annealing is by far smaller than that due to fluorination of the silicon substrate surface.

It is also worth mentioning that the degree of Q_{eff} changes is much higher than that of D_{itmb}, which means that fluorine presence in the structure is especially beneficial for curing

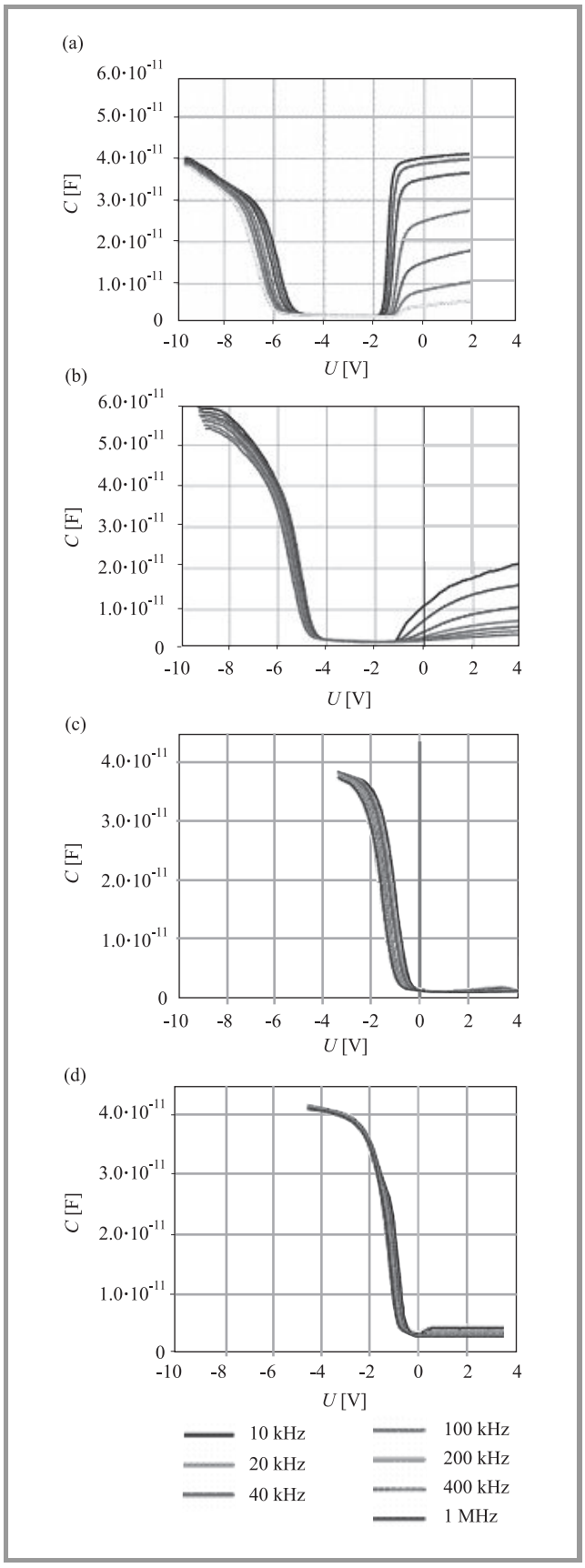


Fig. 2. The $C-V$ characteristics of samples fabricated in different ways: (a) control sample, (b) control sample annealed, (c) RIE fluorinated but not annealed, and (d) RIE fluorinated and annealed.

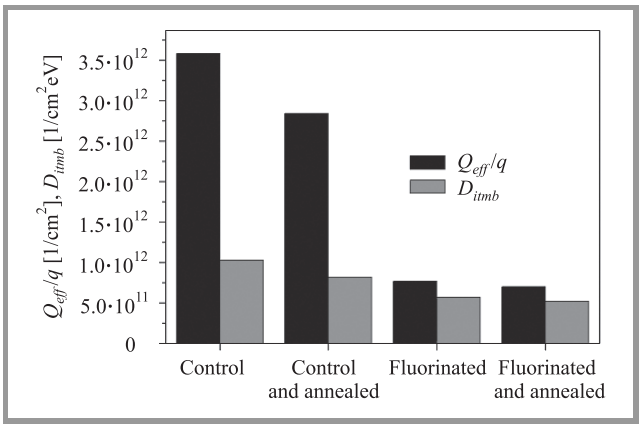


Fig. 3. Dependence of Q_{eff} and D_{itmb} on the method of fabrication the MOS test structure.

Table 2
Values of the most important electrical parameters of MOS structures evaluated from C–V characteristics of samples fabricated using different methods

Sample type	Control	Control annealed	RIE fluorinated	RIE fluorinated annealed
C–V curves analysis at $\varepsilon_{ox} = 3.9$				
N_A [cm^{-3}]	$1.73 \cdot 10^{15}$	$1.17 \cdot 10^{15}$	$0.43 \cdot 10^{15}$	$1.4 \cdot 10^{15}$
V_{fb} [V]	−4.15	−3.28	−1.42	−1.30
V_{mb} [V]	−3.4	−2.7	−0.95	−0.86
Q_{eff}/q [cm^{-2}]	$35.8 \cdot 10^{11}$	$28.4 \cdot 10^{11}$	$7.7 \cdot 10^{11}$	$7.0 \cdot 10^{11}$
D_{itmb} [1/eV cm^2]	$10.3 \cdot 10^{11}$	$8.17 \cdot 10^{11}$	$5.7 \cdot 10^{11}$	$5.18 \cdot 10^{11}$

the defects within the volume of the gate oxide, while high temperature annealing seems to be similarly effective in both areas (interface and volume of the oxide).

3.2.2. Analysis of *I–V* Characteristics

As shown in Fig. 4, application of high temperature annealing in control samples improves a little the statistics of breakdown events at the expense of higher leakage current, while having almost no effect on V_{bd} values.

Introduction of fluorine to the PECVD oxide improves both, V_{bd} values and their distribution. When high temperature annealing is additionally performed on fluorinated structures, we obtain even more narrow V_{bd} distribution, as well as higher V_{bd} values. Practically, no changes in the leakage currents values are observed.

Hence, also from the perspective of the electro-physical properties that are manifested on I–V characteristics, simultaneous application of both, fluorination and high temperature annealing is advantageous in comparison to the application of any of these methods individually.

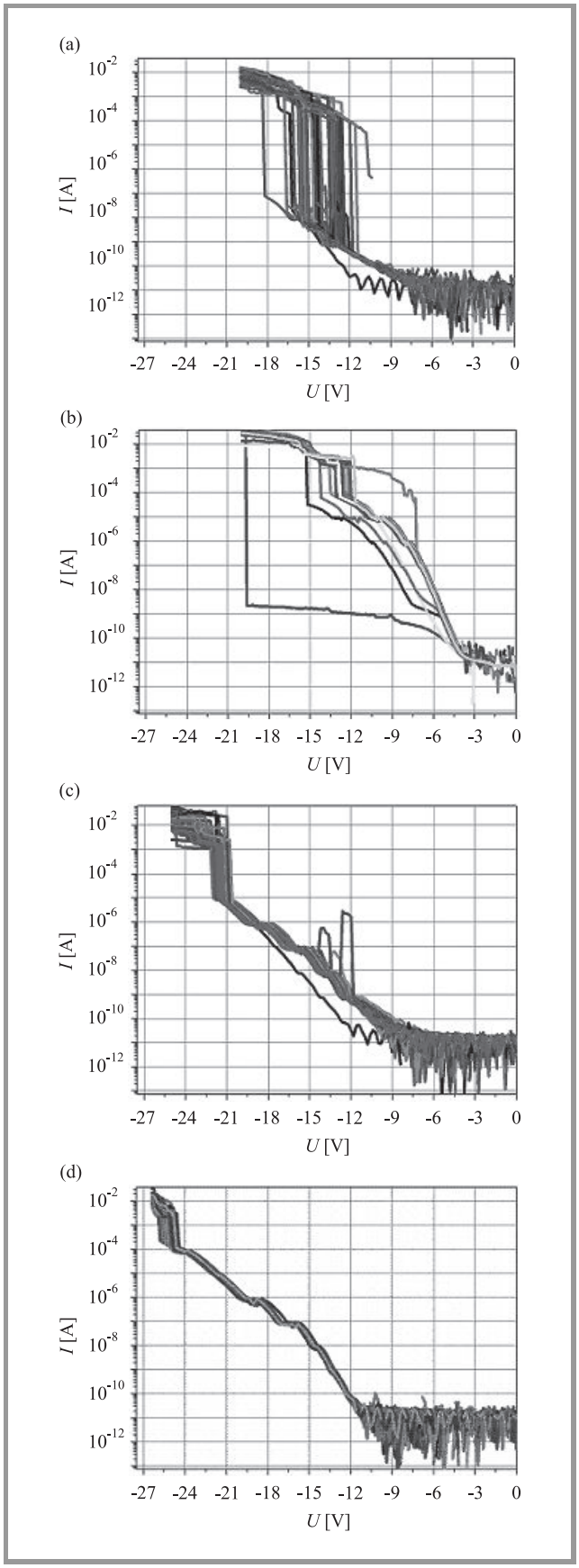


Fig. 4. The *I–V* characteristics of samples fabricated in different ways: (a) control sample, (b) control sample annealed, (c) RIE fluorinated but not annealed, and (d) RIE fluorinated and annealed.

4. Conclusions

The results obtained during this study show that fluorine incorporated at the PECVD oxide/Si interface by means of RIE is very stable even at high temperatures.
This effect allows combining the fluorination and high temperature annealing in order to improve the electro-physical properties of the low temperature oxide (e.g., PECVD oxide). We have also demonstrated that there are several advantages of the application of such combination. The properties of the SiO_2/Si interface (D_{itmb} decreased) and the bulk of the oxide layer (Q_{eff} decreased) are significantly improved. The integrity of the oxide (higher V_{bd}) and its uniformity (V_{bd} distribution) are also improved.

The improvement of the electro-physical properties of the SiO_2-Si systems obtained due to application of both studied steps (fluorination and high temperature annealing) is in all studied cases better than using one of these steps only.

References

[1] Y. Kuo-Lang, J. Ming-Jer, and H. Jenn-Gwo, "Fluorinated thin gate oxides prepared by room temperature deposition followed by furnace oxidation", *Solid-State Electron.*, vol. 43, no. 3, pp. 671–676, 1999.

[2] L. Vishnubhotla, T. P. Ma, H.-H. Tseng, and P. J. Tobin, "Interface trap generation and electron trapping in fluorinated SiO_2", *Appl. Phys. Lett.*, vol. 59, no. 27, pp. 3595–3597, 1991.

[3] P. J. Wright and K. C. Saraswat, "The effect of fluorine in silicon dioxide gate dielectrics", *IEEE Trans. Electron. Dev.*, vol. 36, no. 5, pp. 879–889, 1989.

[4] J. G. Huang and R. J. Jaccodine, "Fast growth of thin gate dielectrics by thermal oxidation of Si in N_2O gas ambient with low concentration of NF_3 addition", *J. Electrochem Soc.*, vol. 140, no. 2, p. L15, 1993.

[5] Y. Nishioka, K. Ohyu, Y. Ohij, N. Natuaki, K. Mukai, and T.-P. Ma, "Hot-electron hardened Si-gate MOSFET utilizing F implantation", *IEEE Electron Dev. Lett.*, vol. 10, no. 4, pp. 141–143, 1989.

[6] P. J. Wright, N. Kasai, S. Inoue, and K. C. Sarawat, "Hot-electron immunity of SiO_2 dielectrics with fluorine incorporation", *IEEE Electron Dev. Lett.*, vol. 10, no. 8, pp. 347–348, 1989.

[7] S. P. Jeng, T. P. Ma, R. Canteri, M. Anderle, and G. W. Rubloff, "Anomalous diffusion of fluorine in silicon", *Appl. Phys. Lett.*, vol. 61, no. 11, pp. 1310–1312, 1992.

[8] M. Kalisz, G. Głuszko, and R. B. Beck, "Novel method of improving electrical properties of thin PECVD oxide films by fluorination of silicon surface region by RIE in RF CF_4 plasma", *J. Telecommun. Inform. Technol.*, no. 1, pp. 20–24, 2010.

[9] M. Kalisz, R. B. Beck, and M. Ćwil, "Reactive-ion-etching (RIE) process in CF_4 plasma as a method of fluorine implantation", *Vacuum*, vol. 82, no. 10, pp. 1046–1050, 2008.

Małgorzata Kalisz – for biography, see this issue, p. 23.

Grzegorz Głuszko and **Romuald B. Beck** – for biographies, see this issue, p. 24.

Paper

Large-Signal RF Modeling with the EKV3 MOSFET Model

Maria-Anna Chalkiadaki and Matthias Bucher

Abstract—This paper presents a validation of the EKV3 MOSFET model under load-pull conditions with high input power at 5.8 GHz, as well as S-parameter measurements with low input power up to 20 GHz. The EKV3 model is able to represent coherently the large- and small-signal RF characteristics in advanced 90 nm CMOS technology. Multifinger devices with nominal drawn gate length of 70 nm are used.

Keywords—compact model, EKV3 model, large-signal, load-pull, MOSFET model, radio frequency.

1. Introduction

The boost of wireless applications in combination with the downscaling of CMOS technologies, has posed a big challenge to the RF CMOS models, which have to be consistent with the increasing demands. Especially the design of integrated power amplifiers (PAs) in RF front-ends of wireless telecommunication circuits implemented in advanced CMOS technology requires that RF models are also validated under large-signal RF conditions where the device shows a nonlinear behavior. Such validation under more realistic operating conditions, e.g., with load-pull analysis, is however quite scarce in literature [1]–[3] for recent advanced CMOS technology.

The EKV3 is a scalable compact MOSFET model which has been designed to provide ease of parameter extraction and provide the designer insight into the device behavior. The scope of the present paper is to investigate the suitability of the EKV3 model to represent both large- and small-signal RF characteristics, from weak through moderate and strong inversion under variable bias conditions, for DC analysis, Y-parameters and load-pull analysis.

2. The EKV3 MOSFET Model

The EKV3 is an analytical compact MOSFET (metal oxide semiconductor field effect transistor) model that relies on MOSFET's physics – the charge sheet theory – to describe its behavior. EKV3 is a representant of the "charge-based" MOS transistor compact models [4], [5]. It first calculates the dependence of the mobile inversion charge density Q_i on the voltages applied to the transistor. Then, it relies on Q_i, and on its particular values Q_{iS} and Q_{iD} at the source and drain ends of the channel, to calculate the drain current and to model all aspects of the device behavior, such as transconductances, transcapacitances, noise, etc.

With the downscaling of the modern advanced CMOS technologies, the complexity of the MOSFET behavior has increased. The EKV3 model has been adapted to cover these new phenomena, such as: quantum effects; polydepletion; surface roughness, phonon- and Coulomb scattering; velocity saturation and channel length modulation; charge-sharing; drain induced barrier lowering; drain induced threshold voltage shift; reverse short- and narrow channel effect; shallow trench isolation effects; edge conductance effect; gate tunneling; layout dependent stress, induced gate noise, etc. Even though the complexity of technology has increased drastically, the EKV3 model maintains a comparatively small number of parameters and the extraction of their values can be a relatively easy procedure.

Furthermore, the RF application of a scalable, bias-dependent model [6], [7], requires that the phenomena such as transmission line effects occurring in the MOS channel (referred to as non-quasi static effects) are suitably described, which is the case in EKV3 [8]. Other high-frequency specific effects are induced gate and substrate noise, as well as increased short-channel thermal noise, which are covered in the model as well.

While at low frequencies the external resistances (except source and drain) and capacitances can be ignored, at radio frequencies they play a dominant role in the behavior of the devices, so they must be carefully modeled. Special relationships for the scaling of parasitics with the number of fingers exist [5]. EKV3 provides the possibility to choose among RF macromodels containing gate resistance, and a substrate network containing up to 5 resistances. Here, a single substrate resistance shown in the schematic representation in Fig. 1 is used, which is adequate except for a very low number of fingers.

Measurements were performed on-wafer at room temperature, and the EKV301.02 model was used for the simu-

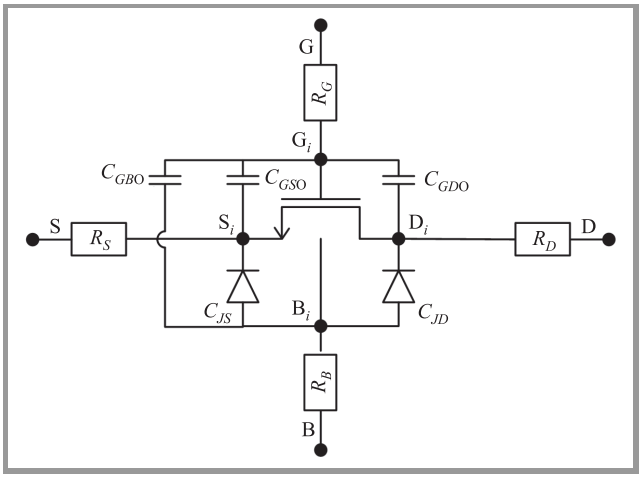


Fig. 1. EKV3 RF macromodel with a single substrate resistance.

lations. Devices with varying channel lengths, widths and number of fingers were measured to obtain a single scalable model, for NMOS and PMOS devices, similarly as shown in [7]. For the purpose of illustration, the device considered throughout the paper is an RF multifinger NMOS transistor with a minimum gate length of $L = 70$ nm, a gate width of $W = 2$ μm and number of fingers NF equal to 10. Parameter extraction was essentially performed from DC and small-signal RF measurement.

3. DC Analysis

Figure 2 shows the I_D versus V_G analysis in linear operation and saturation in linear and logarithmic scale. In addition the gate transconductance g_m versus V_G and the normalized transconductance to current ratio $g_m U_t / I_D$ versus I_D is depicted in both regions of operation. Finally, I_D versus V_D and output conductance g_{ds} versus V_D for six different values of V_G are also presented.

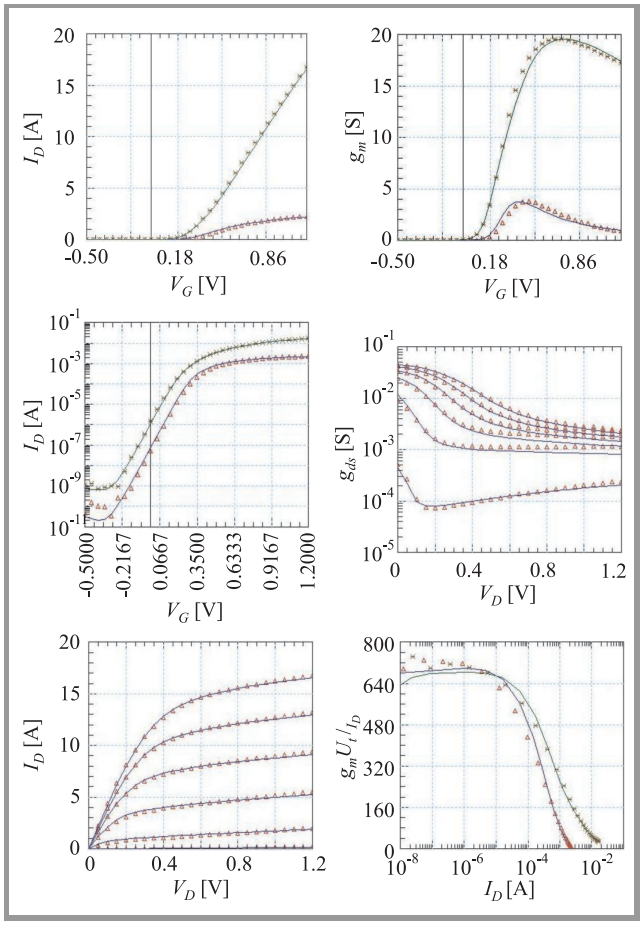


Fig. 2. Static characteristics of NMOS transistor; $L = 70$ nm; $W = 2$ μm; $NF = 10$; I_D versus V_G analysis; $V_D = 50$ mV, 1.2 V; $V_S = 0$ V; I_D versus V_D analysis; $V_G = \{0.2, 0.4, 0.6, 0.8, 1.0, 1.2\}$ V; $V_S = 0$ V. Markers: measurements, lines: EKV3 model.

The results of the DC analysis show that the EKV3 model is capable to represent with a very good accuracy the behavior of the MOSFET transistor with the incorporation of the majority of the phenomena that appear in modern CMOS ultra-deep submicron technologies.

4. Small-Signal Analysis

As the operating frequency increases to the gigahertz range, the role of the extrinsic components rivals that of the intrinsic ones. In order to have efficient circuit design and simulation, the MOS transistor models should be able to predict the behavior of the devices in a wide range of frequencies.

To evaluate the model's accuracy under small-signal conditions the transistor is considered as a two-port, where the gate is on port 1, drain on port 2, while the source is shorted to the substrate, which is the common reference terminal.

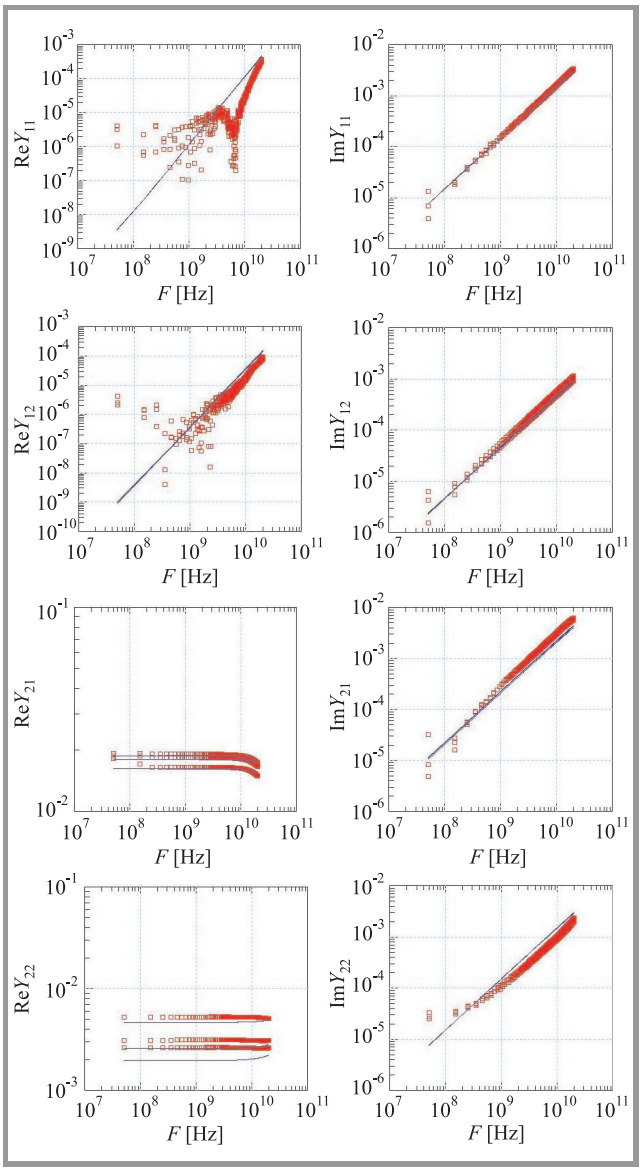


Fig. 3. Two-port small-signal Y-parameters of NMOS transistor; $L = 70$ nm; $W = 2$ μm; NF = 10; $F = 50$ MHz – 20.5 GHz; $V_G = 0.8$ V; $V_D = \{0.4, 0.6, 0.8\}$ V; $V_S = 0$ V. Markers: measurements, lines: EKV3 model.

A vector network analyzer (VNA) is then used to measure the small-signal S-parameters, which are de-embedded and then converted to Y-parameters.

In small-signal analysis, the response of the device is measured in a 50 Ω system, as a function of frequency and bias point. Then we can accurately predict the small-signal response if the device sees impedances other than 50 Ω, as the MOS transistor shows a linear behavior at low frequencies. The frequencies at which the device is tested range from 50 MHz up to 20.5 GHz.

Figure 3 shows the real and imaginary parts of the Y-parameters. These confirm that the device, under the given bias conditions, is not importantly affected by NQS (non-quasi static) effects [8], at least not below 10 GHz.

5. Large-Signal Analysis

Power amplifiers often constitute the bottleneck in using low-cost, high-density digital CMOS processes for integrating multi-gigahertz wireless application in single chips. The compact models should predict the distortion occurring in transistors operating under large-signal conditions, with acceptable accuracy. One way to check a model's capability in these conditions is to measure the load-pull characteristics.

Load-pull analysis consists of varying or "pulling" the load impedance seen by a device-under-test (DUT) while measuring the performance of the DUT. Load-pull is used to measure a DUT in actual operating conditions. This method is important since in large-signal conditions, MOS transistors show a nonlinear behavior and thus the operating point may change with power level or tuning. Load-pull analysis is usually performed at distinct frequencies, e.g., 2.4 or 5.8 GHz.

For the load-pull analysis two different measurements were carried out using two slightly different simulation setups. Firstly, output power P_{out} at operating frequency $F = 5.8$ GHz was studied when the load was varied in a specific range. Next, the gain versus input power P_{in} was examined when load (Z_L) was about 50 Ω. The difference between the two cases is that in the first one, P_{in} and terminal voltages are set but the load Z_L is varied, while in the second, Z_L is fixed at about 50 Ω and terminal voltages are set and P_{in} is varying from –20 dBm to +5 dBm. Here, for briefness, the setup that was used for the first case, using Agilent's ADS, is presented in Fig. 4.

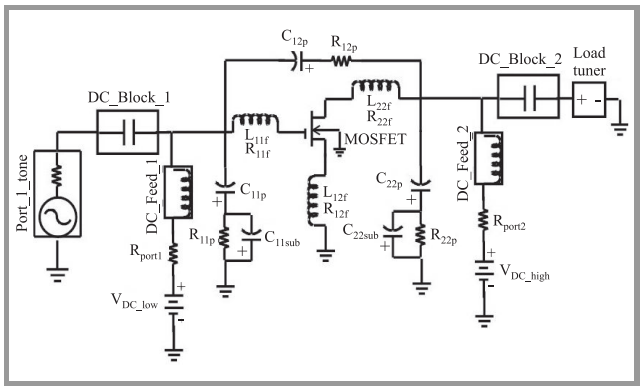


Fig. 4. Load-pull simulation setup; pulling Z_L.

The results that were obtained from the load-pull analysis when Z_L varies are displayed with the use of the Smith chart in Fig. 5. The contours represent all the different loads in the Smith chart area that result in the same transistor output power when P_{in} is constant. The node voltages that were applied to the transistor were $V_G = V_D = 0.8$ V and $V_S = 0$ V, while the input power was set at $P_{in} = 5$ dBm to ensure large-signal conditions.

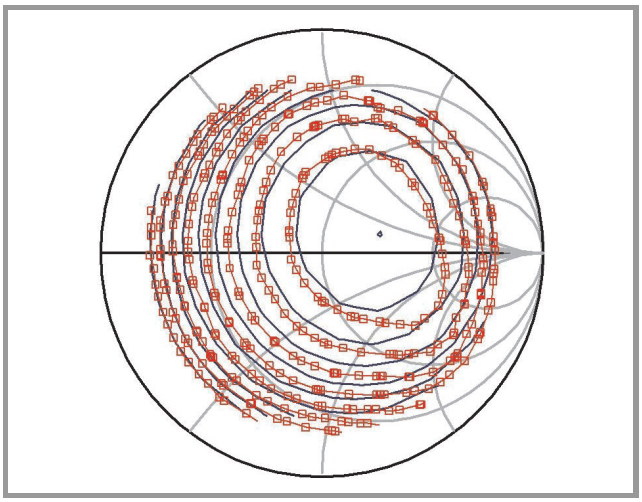


Fig. 5. P_{out} contours for an NMOS transistor pulling Z_L; $L =$ 70 nm; $W = 2$ μm; $NF = 10$; $F = 5.8$ GHz; $P_{in} = +5$ dBm; $V_G = V_D = 0.8$ V; $P_{out} =$ –4.65 – 3.5 dBm (step = 0.815 dBm). Markers: measurements, lines: EKV3 model.

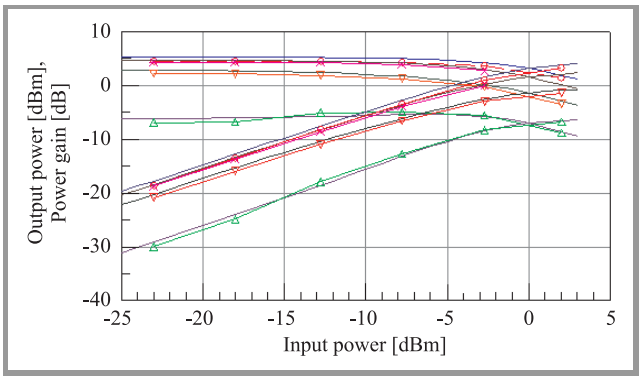


Fig. 6. Output power and power gain versus input power; NMOS transistor; $L = 70$ nm; $W = 2$ μm; $NF = 10$; $F = 5.8$ GHz; $Z_L = 50$ Ω; $P_{in} =$ –20 – +5 dBm; $V_G = 0.8$ V; $V_S = 0$ V. Lines with markers: measurements, lines: EKV3 model. $V_D = 0.2$ V (triangle), 0.4 V (reverse triangle), 0.6 V (cross), 0.8 V (circle).

The EKV3 model is also consistent when considering the output power and output power gain in relation to the input

power from small- to large-signal conditions when Z_L is set to about 50 Ω as seen in Fig. 6. The gain compression is clearly apparent when input power is increased, revealing the nonlinear behavior of the transistor under large-signal conditions.

The EKV3 model depicts the output power qualitatively well, although the model tends to slightly overestimate the measured power gain at high drain voltage. Note that these results are not dependent on the number of channel segments [4], [8] used in EKV3 indicating that NQS effects are not prominent at the given conditions of analysis. When

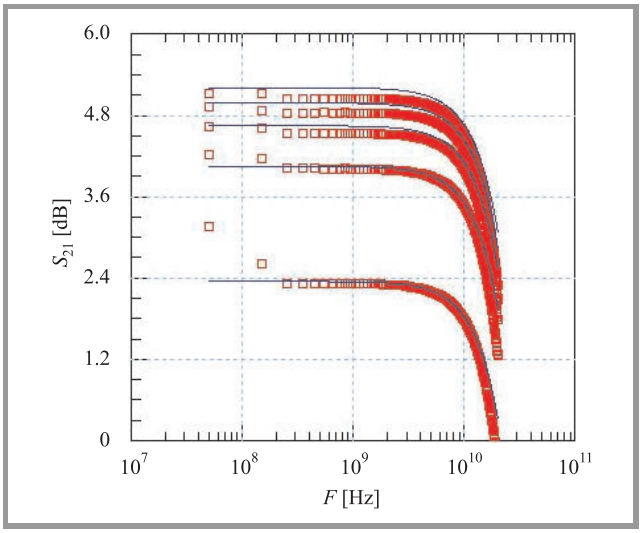


Fig. 7. Magnitude of S_{21} (small-signal) of an NMOS transistor; $L = 70$ nm; $W = 2$ μm; $NF = 10$; $V_G = 0.8$ V; $V_S = V_B = 0$ V; $V_D = \{0.4, 0.6, 0.8, 1.0, 1.2\}$ V. Markers: measurements, lines: EKV3 model.

examining the magnitude of S_{21} from small signal analysis (indicative of the RF gain), we observe that the model predicts the measurements relatively accurately, with a slight overestimation of gain by the model similarly as observed under large-signal conditions (Fig. 7).

6. Conclusion

This paper has shown the consistency of the EKV3 model in describing the MOSFET behavior covering DC analysis, small-signal S-parameter analysis up to 20 GHz, and load-pull analysis at 5.8 GHz. Extraction of RF parameters was done from the small-signal RF measurements. All results are obtained from a single RF NMOS device and the simulations are carried out with a single parameter set of the EKV3 MOSFET model, showing its capabilities in a wide range of different measurement conditions. This underlines the EKV3 model's capabilities as an RF model, including the nonlinear character of MOSFETs under high input power, which should be carefully taken into account, e.g., when designing for large-signal conditions occurring in RF power amplifiers.

Acknowledgment

We would like to thank Dr. Sadayuki Yoshitomi, Toshiba Semiconductor, for providing the measurements, and Dr. Antonios Bazigos for help with EKV3 model and parameter extraction.

References

[1] C.-W. Kuo, C.-C. Ho, and Y.-J. Chan, "Scalable large-signal model of 0.18 μm CMOS process for RF power predictions", *Solid-State Electron.*, vol. 47, no. 1, pp. 77–81, 2003.

[2] S. Yoshitomi, "Challenges of compact modeling for deep-submicron RF-CMOS devices", in *Proc. 12th Int. Conf. MIXDES 2005*, Kracow, Poland, 2005.

[3] C.-C. Wei, C.-S. Cheng, S.-W. Lin, Y.-J. Chen, H.-C. Chiu, and W.-S. Feng, "An improved BSIM4 model for 0.13-um gate-length high linearity CMOS RF transistors", *PIERS Online*, vol. 3, no. 7, pp. 1000–1004, 2007.

[4] A. Bazigos, M. Bucher, F. Krummenacher, J.-M. Sallese, A.-S. Roy, and C. Enz, "EKV3 MOSFET Compact Model Documentation, Model Version 301.02", Techn. Rep., Technical University of Crete, June 2008.

[5] C. C. Enz and E. A. Vittoz, *Charge-Based MOS Transistor Modeling: The EKV Model for Low-Power and RF IC Design*. Chichester: Wiley, 2006.

[6] S. Yoshitomi, A. Bazigos, and M. Bucher, "The EKV3 model parameter extraction and characterization of 90 nm RF-CMOS technology", in *Proc. 14th Int. Conf. MIXDES 2007*, Ciechocinek, Poland, 2007, pp. 74–79.

[7] M. Bucher, A. Bazigos, S. Yoshitomi, and N. Itoh, "A scalable advanced RF IC design-oriented MOSFET model", *Int. J. RF Microw. Comput. Aid. Eng.*, vol. 18, no. 4, pp. 314–325, 2008.

[8] M. Bucher and A. Bazigos, "An efficient channel segmentation approach for a large-signal NQS MOSFET model", *Solid-State Electron.*, vol. 52, no. 2, pp. 275–281, 2008.

Maria-Anna Chalkiadaki was born in Athens, Greece, in 1983. She received the diploma degree in electronic and computer engineering from the Technical University of Crete (TUC), Chania, Greece, in 2008. She is currently pursuing an M.Sc. degree at the same institution. Her current research is in compact modeling of advanced CMOS with emphasis on high frequency.

e-mail: machalkiadaki@isc.tuc.gr
Department of Electronics and Computer Engineering
Technical University of Crete
Chania 73100, Greece

Matthias Bucher was born in Switzerland. He received the electrical engineering and Ph.D. degrees from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1993 and 1999, respectively. In 2004, he joined the Department of Electronics and Computer Engineering, Technical University of Crete (TUC), Chania, Greece, as an Assistant Professor, where he currently is the Director of the Electronics Laboratory. His research interests are in the design of low-voltage, low-power analog/RF integrated circuits, in wide-band characterization and advanced compact modeling of single- and multi-gate nanoscale CMOS devices. He is the coordinator of the EKV3 MOSFET compact model code development. He has authored or co-authored over 55 publications in international journals, conference papers and book chapters. He is a member of the Technical Programme Committee of IEEE DDECS, was Chairman for Tutorials and Workshops at ESSDERC/ESSCIRC 2009, Athens, Greece, and an invited plenary speaker at MIXDES 2009, Łódź, Poland. He is a member of the IEEE.

e-mail: bucher@electronics.tuc.gr
Department of Electronics and Computer Engineering
Technical University of Crete
Chania 73100, Greece

Paper

Multi-Domain Modeling and Simulations of the Heterogeneous Systems

Tomasz Bieniek, Grzegorz Janczyk, Paweł Janus, Jerzy Szynka, Piotr Grabiec, Andrzej Kociubiński, Magdalena Ekwińska, Daniel Tomaszewski, and Arkadiusz Malinowski

Abstract—This paper discusses the multi-domain modeling and simulation issues of the design and analysis of heterogeneous integrated systems. Modeling and simulation methodology and tools are also discussed.

Keywords—Corona, Coventor, 3D integration, e-Cubes, heterogeneous systems, MATLAB, modeling, simulation.

1. Introduction

There is a strong demand in the market for fast and cheap development of reliable, standalone, wireless, multifunctional devices for automotive, aerospace and biomedical applications. Therefore the complexity of electronic sensors and multifunctional devices increases.

Fast development of device fabrication technology stimulates new challenging technology development and progress in wafer processing, module interconnects and packaging technology. Apart from a broad hardware experience, standalone software design simulations have to be done before hardware fabrication.

Such research and development work is done within the framework of e-Cubes [1] and SE2A [2] projects, where multimodule 3D integration of multifunctional elements is the main goal. Many various functionalities, such as power supply, power management, RF communication, digital signal processing, memory, micro electro-mechanical systems (MEMS) and many other have to be assembled into a single device. The range of applications is not limited to automotive, aerospace and biomedical domains only.

As the device integration is an important factor of the design process in improving the reliability of the device, new materials, technologies and technological concepts are being tested and verified. The reliability of vertical interconnects turns out to be one of crucial design issues. Thermal management of the device and thermo-mechanical stress simulations are prime candidates for research on reliability. From the reliability point of view standalone functional module simulations are insufficient to find out whether the modules will work if assembled into a final stacked device or not due to some thermal interactions neglected during the design verification process. For example, if one module dissipates the power, the heat flows through VIA's. It extends to other modules and affects thermal budget of other devices.

A slightly different point of view is exploited within the Corona project [3], where the main goal is to develop the best methodology and tools for fast product development. It should cover the whole flow from the idea to the final micro-nanotechnology product. The key issue is short time-to-market and efficient dataflow for distributed design and fabrication scenarios. One of the main goals of the Corona is also the support of the multi-site product development – where every module of the integrated microsystem may be fabricated in the best technology in different places/fabs and finally integrated into one multifunctional device (Fig. 1).

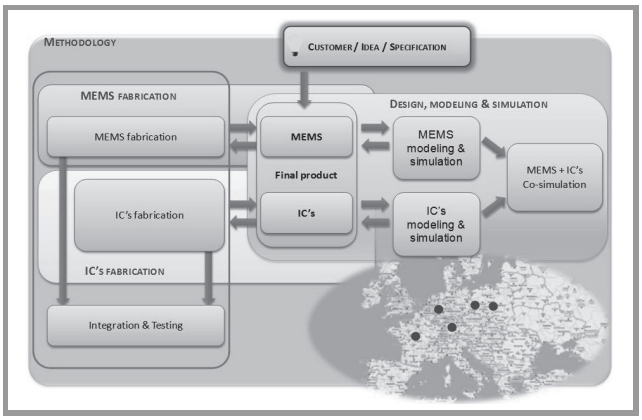


Fig. 1. Multi-site product development scenario, where every module of the integrated microsystem (MEMS+IC's) may be fabricated in the best technology in different places/fabs and finally integrated into one multifunctional device.

From this point of view different design tools are important at different design stages: from device design, through device modeling and simulation, to process development and optimization. It enables fast virtual prototyping for selected parts of the device or the whole final device.

2. Multi-Domain Modeling and Simulation Examples

As it has been mentioned before, multi-domain modeling and simulation is very important at every stage of device fabrication. Selected examples of various multi-domain modeling and simulation are presented.

For design, modeling and simulations of heterogeneous micro- and nanostructures designers may use CoventorWare [4] and ESI-CFD [5] software environments implementing multi-domain FEM (finite element method) simulation and analysis of the micro- and nanostructures. One of the key reliability issues of 3D structures is the intercon-

nection reliability. Several integration techniques have been considered within the framework of the e-Cubes project to obtain sample demonstrator structures. An example of the integrated modeling and simulation is shown in Fig. 2.

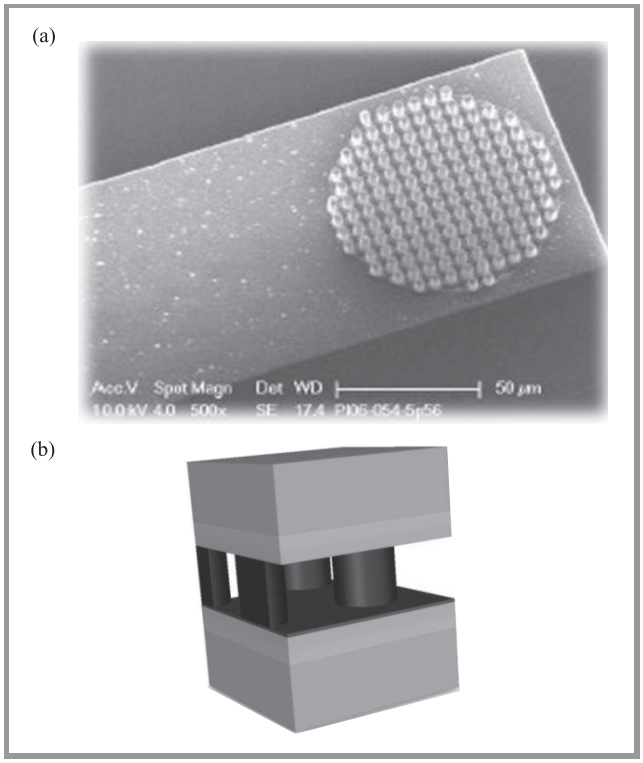


Fig. 2. (a) SEM image of a group of microinserts [6], [7] and (b) 3D model of a quarter of a group of 9 microinserts in CoventorWare.

Thermo-electro-mechanical microinsert simulations [6], [7] have been performed (Figs. 3 and 4) to find the electrical and mechanical behavior of the simulated structures and to verify possible electromigration problems and assess the electrical suitability of the examined structures. The pressure needed to be applied for thermo-compression bonding is also investigated.

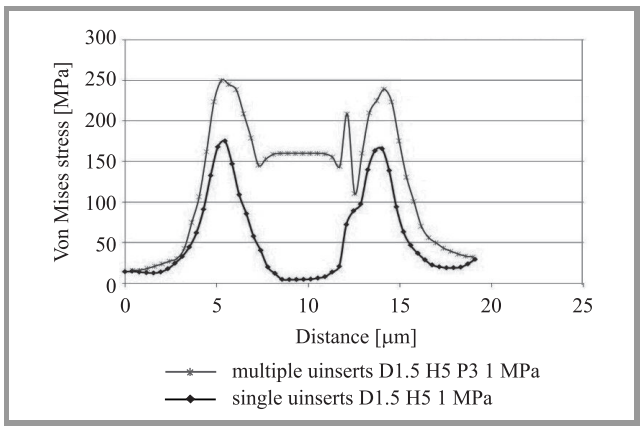


Fig. 3. Multi-domain simulation results: von Mises stress distribution across a microinsert structure under pressure of 1 MPa applied on top of the structure with single and multiple microinsert configurations (diameter of 1.5 m, height of 5 μm and pitch of 3 μm for multiple microinserts).

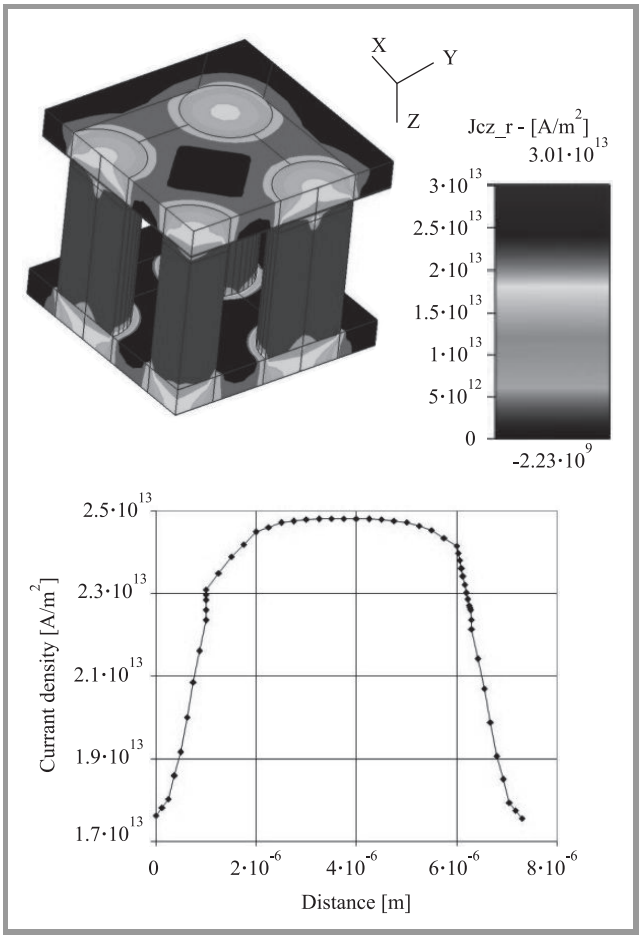


Fig. 4. Multi-domain simulation results: current density distribution through microinserts under 10 V applied between upper and bottom part.

The results presented in Figs. 3 and 4 show the von Mises stress distribution across the microinsert structure under the pressure of 1 MPa for device configurations with single and multiple microinserts. It is very important to characterize the stress distribution under different pressures according to thermo-compression bonding process usually used for 3D integration. The current density distribution shows how the current flows through the modeled 3D device. It also makes it possible to verify if one should expect such problems as, e.g., electromigration. Modeling and simulation results at this stage can help to optimize, e.g., geometry, materials and dimensions (e.g., diameter, pitch, height).

Besides modeling and simulation efforts devoted to the single structure used for 3D integration, modeling and simulation of the integrated microsystem have been also presented. Such a system is the e-Cubes project demonstrator sample. The general idea of the project is presented in Fig. 5 [8].

From the reliability point of view thermo-mechanical behavior of the whole system is one of the most important

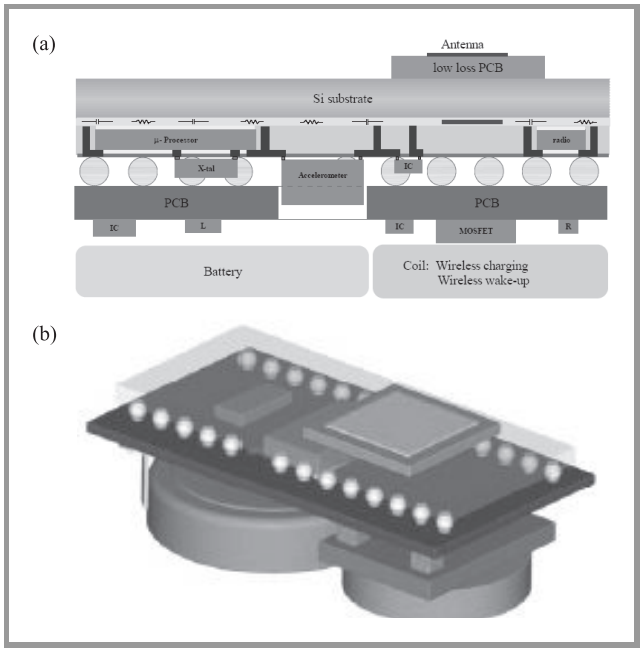


Fig. 5. (a) Physical architecture of a sensor node (whole system for health monitoring), drawing not on scale [8] and (b) mechanical design drawing (3D model) [8].

reliability parameters. Even if all standalone modules are correctly designed, simulated and one can anticipate perfect operation after assembly process, in the case of 3D integration their functionality may be affected by various side effects not taken into consideration during the design process. Each of the assembled modules of a 3D multi-module device interferes thermally and electrically with its neighborhood (see Fig. 6). The probability that a single module may affect another by, e.g., high temperature transfer is quite high. Therefore the module arrangement and complex simulation are necessary at each design stage.

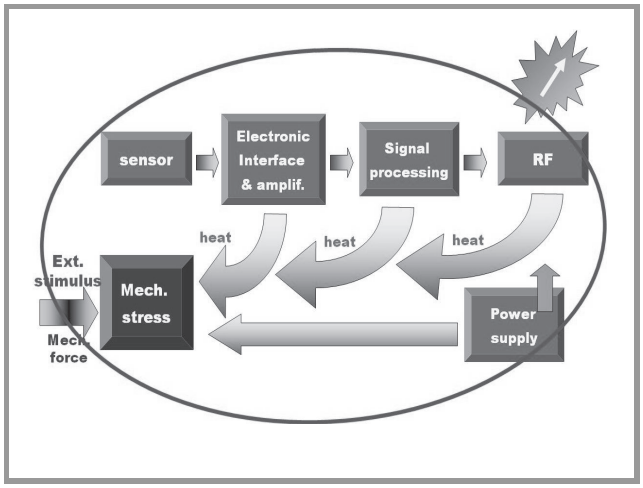


Fig. 6. Interactions between microsystem modules [9].

The 3D integration technique has been simulated along with realistic thermo-mechanical boundary conditions applied (power dissipation from different modules and parts). Also some mechanical issues regarding the encapsulation methodology were considered. Thermal simulations results show that under the assumed boundary conditions the temperature does not increase too much across the whole demonstrator device structure. Sample results of the thermal investigation are shown in Fig. 7.

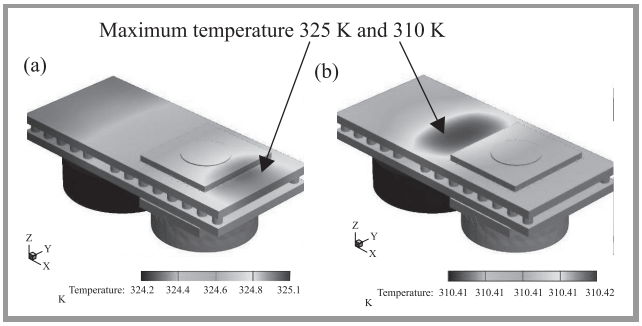


Fig. 7. Temperature distribution through a 3D model of the integrated microsystem in two modes: (a) high power mode and (b) low power mode.

Analysis of the mechanical displacement simulation leads to a major conclusion. Displacement and stress distribution is better if more fixed points are applied to the device (Fig. 8). While in the high power mode the maximum displacement reaches 24 μm on the edge of the structure, in the low power mode the displacement does not exceed 6 μm.

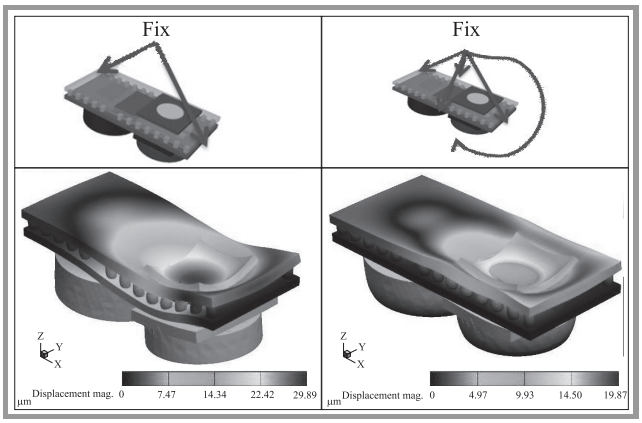


Fig. 8. Mechanical part of multi-domain simulation and modeling results – displacement magnitude (displacement and deformation factor rescaled $\times$100!!!) of structures in high power mode in two different modes of fixation to the package.

The heterogeneous device oriented (Hedoris) multi-domain simulation system [10], [11] will be developed within the framework of the projects mentioned above. The Hedoris system implements the idea of the hardware description language (HDL) of the device model based on Verilog-AMS [12] language specification used for device thermal modeling. Currently the system is mainly optimized to run thermal simulations of heterogeneous devices and is still under development. Intensive work has been undertaken including TCAD simulations oriented on the system

extension to cover thermo-mechanical internal device interactions, such as mechanical stress, which may affect the overall performance of the device.

3. Summary

The design and verification methodologies are discussed and presented. Modeling and simulation samples presented in this paper show the importance of the flexibility of the multi-site design and development framework of cooperation. The paper also shows how universal CAD tools used for design of micro- and nanodevices in distributed engineering environments should be and how important it is to correctly estimate the device reliability and keep balance between reaching the attractive time to market and fulfilling the customer requirements.

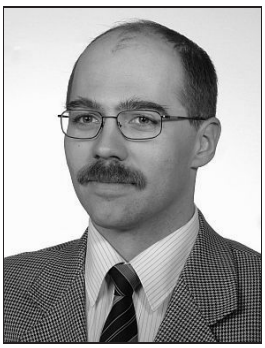
Acknowledgements

This work has been partially supported by the European Commission under support-number: IST-026461 (European project: 3D Integrated Micro/Nano Modules for Easily Adapted Application, acronym e-CUBES [1]), CP-FP 213969-2 (European project: Customer-Oriented Product Engineering of Micro and Nano Devices, acronym CORONA [2]) and Eniac SE2A (Nanoelectronics for Safe, Fuel Efficient and Environment Friendly Automotive Solutions [3]).
We would like to thank e-Cubes and CORONA project Partners for the excellent and fruitful collaboration, special thanks to Ric van Doremalen and Piet van Engen from Philips Applied Technologies.

References

[1] e-Cubes project, http://www.ecubes.org
[2] Nanoelectronics for Safe, Fuel Efficient and Environment Friendly Automotive Solutions project, http://www.eniac.eu, SE2A
[3] Corona project, http://www.corona-mnt.eu
[4] 3D MEMS Design Automation and Virtual Fabrication – Coventor, http://www.coventor.com
[5] ESI Group CFD Portal, http://www.esi-cfd.com/
[6] A. Mathewson, J. Brun, C. Puget, R. Franiatte, N. Sillon, F. Depoutot, and B. Dubois-Bonvalot, "Microstructured interconnections for high security systems", in *Proc. Electron. Syst. Integr. Technol. Conf.*, 2006, Dresden, Germany, vol. 1, pp. 126–132.
[7] A. Mathewson, J. Brun, G. Ponthenier, R. Franiatte, A. Nowodzinski, N. Sillon, G. Poupon, F. Deputot, and B. Dubois-Bonvalot, "Detailed characterisation of Ni microinsert technology for flip chip die on wafer attachment", in *Proc. Electron. Compon. Technol. Conf. ECTC'07*, Sparks, USA, 2007, pp. 616–621.
[8] P. van Engen, R. van Doremalen, W. Jochems, A. Rommers, S. Cheng, A. Rydberg, T. Fritzsch, J. Wolf, W. De Raedt, P. Müller, E. Alarcon, and M. Sanduleanu, "3D Si-level integration in wireless sensor node", in *Proc. Smart Syst. Integr. 2009 Conf.*, Brussels, Belgium, 2009, pp. 150–157.
[9] T. Bieniek, P. Janus, A. Kociubiński, P. Grabiec, G. Janczyk, and J. Szynka, "Integrated multi-domain modeling and simulation of complex 3D micro- and nanostructures", in. *Proc. Smart Syst. Integr. 2008*, Barcelona, Spain, 2008, pp. 399–402.
[10] G. Janczyk, T. Bieniek, J. Szynka, and P. Grabiec, "Reliability issues of e-Cubes heterogeneous system integration", *Microelectron. Reliab.*, vol. 48, pp. 1133–1138, 2008.
[11] G. Janczyk, T. Bieniek, J. Szynka, and P. Grabiec, "Integrated thermal modeling of heterogeneous eCubes stacked devices", in *Proc. Therminic 2008, 14th Int. Worksh. Thermal Invest. ICs Syst.*, Roma, Italy, 2008, pp. 80–84.
[12] Accelera, "Verilog-AMS Language Reference Manual", Version 2.1, 2003, http://www.accellera.com

Tomasz Bieniek received the M.Sc. and Ph.D. degrees in electronic engineering from the Warsaw University of Technology, Poland, in 2002 and 2007, respectively. The doctor thesis was devoted to plasma technologies for ultrathin dielectric layer formation. In 2006 he joined the Department of Silicon Microsystems and Nanostructure Technology of the Institute of Electron Technology, Warsaw. His research is focused on MEMS and CMOS silicon technologies and multi-domain modeling, as well as the simulation of micro- and nanostructures. He is an author and co-author of more than 40 technical papers and presentations presented in journals and at conferences.
e-mail: tbieniek@ite.waw.pl
Institute of Electron Technology
Lotników av. 32/46
02-668 Warsaw, Poland



Grzegorz Janczyk received the M.Sc. degree (with honors) and Ph.D. degree in electronic engineering from the Warsaw University of Technology (WUT), Poland, in 1999 and 2005, respectively. Since then, he has been with the Institute of Microelectronics and Optoelectronics (IMIO), VLSI Engineering and Design Automation Division, at the same university. In 2006 he joined the Institute of Electron Technology (ITE). He is an Assistant Professor in IMiO-WUT and ITE. His main Ph.D. related interests cover device modeling along with technological SOI fabrication process modeling and development of SOI-MOS transistor models. Multi-domain device modeling using Verilog language also belongs to the range of his professional interests. Statistical yield simulations and mismatch modeling are secondary field of interest. He is also C++ programmer.
e-mail: Janczyk@ite.waw.pl
Institute of Electron Technology
Lotników av. 32/46
02-668 Warsaw, Poland

Paweł Janus received the B.Sc. and M.Sc. degrees from the Technical University in Wrocław, Poland, in 1996 and 1998, respectively. In 2003 he received the Ph.D. degree in electrical engineering from the Wrocław University of Technology. Since 2003, he has been with the Institute of Electron Technology, Warsaw. His research interests include studies of materials for MEMS applications, micromachining of silicon microstructures, microsensors, and microactuators. His current research is silicon force sensors development for nanomechanical applications. He is an author and co-author of more than 30 technical papers presented in journals and at conferences.
e-mail: janus@ite.waw.pl
Institute of Electron Technology
Lotników av. 32/46
02-668 Warsaw, Poland

Jerzy Szynka graduated in 1972 and received in 1976 the Ph.D. from Technical University Dresden, Germany. The doctor thesis was devoted to cellular automata theory. In 1977 he joined to the Institute of Mathematical Machines in Warsaw, Poland, where he worked on automation project. In 1978 he joined to the Institute of Electron Technology in Warsaw. In 1992 he became Managing Director in Quantum Ltd. the software company, where he was responsible for the development and application of commercial software. From 2002 he is the Head of the Circuit and System Department of Institute of Electron Technology managing the research and development on behavioral synthesis using VHDL and Verilog description languages, silicon implementation of algorithms and design of mixed signal ASICs. Actual activities concern the design methods for heterogeneous systems.
e-mail: Szynka@ite.waw.pl
Institute of Electron Technology
Lotników av. 32/46
02-668 Warsaw, Poland

Piotr B. Grabiec graduated from the Warsaw University of Technology, Poland, in 1973, and received the Ph.D. degree in chemistry from the same university in 1985. In 1974 he joined the Institute of Electron Technology, Warsaw, where he was involved in CVD and diffusion technology research. Since 1999 he has been the Head of Silicon Microsystem and Nanostructure Department. His present activity involves fabrication of silicon ASICs, optoelectronic devices and MEMS and their integration. He has been involved in 10 EU projects and has been awarded numerous awards for development and commercialization of advanced micro-devices. He is the member of IEEE and Electrochemical Society. He is the author and co-author of more than 300 scientific papers and conference presentations. He holds 19 patents.
e-mail: grabiec@ite.waw.pl
Institute of Electron Technology
Lotników av. 32/46
02-668 Warsaw, Poland

Andrzej Kociubiński received the M.Sc. and Ph.D. degrees in electronic engineering from the Warsaw University of Technology, Poland, in 2002 and 2007, respectively. In 2001 he joined the Department of Silicon Microsystems and Nanostructure Technology of the Institute of Electron Technology, Warsaw. Since 2007, he has been with the Lublin University of Technology. His research interests include diagnostics, characterization, simulation and modeling of silicon devices. His recent works are related to SOI devices, pixel detectors on SOI wafers and integrated microsystems.
e-mail: akociub@semiconductor.pl
Lublin University of Technology
Nadbystrzycka st 38a
20-618 Lublin, Poland

Magdalena Aleksandra Ekwińska received the M.Sc. and Ph.D. degrees in micromechanics from the Warsaw University of Technology, Poland, in 2002 and 2009, respectively. The doctoral thesis was devoted to investigation of nanoscale material properties and comparison of two commonly used nanoscale tests: nanowear and nanoindentation. One of the results was the method of prediction of the results of the time consuming nanowear test based on the information from nanoindentation test only. In 2001 she was an Intern with Hysitron Company (manufacturer of nanoindentation equipment) and in 2003 with the Mechanical-Mathematical Department of the University of Belarus. From 2007 she is a member of Polish Tribology Society (PTT), which is a founder member of International Tribology Council (ITC). In 2008 she joined the Institute of Electron Technology to the Department of Silicon Microsystems and Nanostructure Technology.

Her work is focused to MEMS multi-domain modeling and simulation of micro- and nanostructures.
e-mail: ekwinska@ite.waw.pl
Institute of Electron Technology
Lotników av. 32/46
02-668 Warsaw, Poland

Daniel Tomaszewski received the M.Sc. degree from the Warsaw University of Technology, Poland, in 1980. Since then he is with the Institute of Electron Technology, Warsaw. In 1998 he received the Ph.D. degree in electrical engineering. His research interests include modeling and characterization of silicon and silicon-on-insulator devices for the purpose of IC diagnostics and design. He participated in several conferences and workshops related to these fields.
e-mail: dtomasz@ite.waw.pl
Institute of Electron Technology
Lotników av. 32/46
02-668 Warsaw, Poland

Arkadiusz Malinowski received the B.Sc. and M.Sc. degrees from the Warsaw University of Technology, Poland, in 2005 and 2007, respectively. He currently pursues the Ph.D. degree in electrical and computer engineering at the Warsaw University of Technology and the Nagoya University, Japan. In March 2004 he joined the Institute of Electron Technology, Warsaw, working in the area of TCAD semiconductor process and device simulation. His research interests include FinFET CMOS technology scaling, nanocarbon based FET and plasma nanoprocessing.
Institute of Electron Technology
Lotników av. 32/46
02-668 Warsaw, Poland
e-mail: m_arkadi@nuee.nagoya-u.ac.jp
Nagoya University
C3-1 (631), Furo-cho, Chikusa-ku
464-8603 Nagoya, Japan

Paper

Frequency Offset Compensation for OFDM Systems Using a Combined Autocorrelation and Wiener Filtering Scheme

Ali Ramadan Ali, Tariq Jamil Khanzada, and Abbas Omar

Abstract—One of the orthogonal frequency division multiplexing (OFDM) system disadvantages is its sensitivity to frequency offset and phase noise, which lead to losing the orthogonality between the subcarriers and thereby degrade the system performance. In this paper a joint scheme for frequency offset and pilot-based channel estimation is introduced in which the frequency offset is first estimated using an autocorrelation method, and then is fined further by applying an iterative phase correction by means of pilot-based Wiener filtering method. In order to verify the capability of the estimation algorithm, the scheme has been implemented and tested using a real measurement system in a multipath indoor environment. The results show the algorithm capability of compensating for the frequency offset with different transmission and channel conditions.

Keywords—channel characterization, frequency offset, OFDM measurement, Wiener filtering.

1. Introduction

Orthogonal frequency division multiplexing (OFDM) divides the base bandwidth into N orthogonal narrow subchannels transmitted in parallel. The generation of the OFDM signals at the transmitter is accomplished using inverse fast Fourier transform (IFFT), which delivers orthogonal carriers. Due to its long symbol duration together with adding a suitable guard interval (GI) to the time domain signal, OFDM system can handel the frequency selective fading resulting from the multipath effect. The source bit stream in OFDM system is first encoded, interleaved and then mapped using a common modulation scheme like quadrature amplitude modulation (QAM) or phase shift keying (PSK).

The resulting symbols are arranged in blocks with length N, and then IFFT is applied, so that, each symbol is given an orthogonal carrier. In order to cope with the multipath effect a suitable GI (longer than the channel impulse response) is inserted in the time domain signal before transmission. In OFDM system usually a so called cyclic prefix GI is used, which inserts the last portion of the OFDM symbol at the beginning. This type of GI allows for circular convolution with the channel and maintains the orthogonality of the carriers. The ability of combating the frequency selective channels has resulted in choosing the OFDM system as a standard modulation scheme for transmitting high rate services as in IEEE 802.11a/g wireless local area network (WLAN) [1], digital audio broadcasting DAB [2], digital video broadcasting (DVB) [3], and became a promising candidate for fourth generation (4G) mobile communication systems [4], [5]. On the other hand OFDM is sensitive to phase noise and frequency offsets, which cause phase rotation and loss in the orthogonality between the subcarriers.

This introduces intercarrier interference (ICI) [6]–[9] as another source of noise, which degrades the system performance. The notation to be adopted throughout this paper conforms to the following convention: bold uppercase letters $\mathbf{H}$ denote matrices, bold uppercase underlined letters $\underline{\mathbf{V}}$ denote frequency domain vectors, while time domain vectors are represented by bold lowercase underlined letters $\underline{\mathbf{v}}$, variables in time are denoted by lowercase letters $v(n)$, and in frequency by uppercase letters $V(n)$. The operator $\{.\}^H$, represents Hermitian transpose. $\mathbf{I}$ denotes the identity matrix.

2. System Model and the Measurement Setup

Making use of the FFT operation, the time domain transmitted OFDM symbol can be written as [10]

$$x(n) = \sum_{m=0}^{N-1} S(m) \mathrm{e}^{\mathrm{j}\frac{2\pi nm}{N}} \qquad 0 \le n \le N-1\,, \tag{1}$$

where $S(m)$ is the QAM symbol at the mth subcarrier. Because of the cyclic prefix, the transmitted signal is circularly convolved with the impulse response of the multipath time-variant channel. The received time domain OFDM symbol is expressed as

$$y(n) = \sum_{l=0}^{L-1} h(n,l)x(n-\tau_l) + v(n), \tag{2}$$

where $h(n,l)$ and τ_l are the complex random variable and the corresponding tap delay, respectively, at the lth path of the multipath channel, and $v(n)$ represents the additive white Gaussian noise at sample instant n. After applying FFT at the receiver side, the frequency domain OFDM symbol becomes

$$Y(k) = \frac{1}{N} \sum_{n=0}^{N-1} \Big(\sum_{l=0}^{L-1} h(n,l)x(n-\tau_l) + v(n) \Big) \mathrm{e}^{-\mathrm{j}\frac{2\pi nk}{N}}\,. \tag{3}$$

Equation (3) can be simplified [9] to

$$Y(k) = S(k)H_{k,k} + \underbrace{\sum_{m/k=} S(m)H_{k,m}}_{\text{ICI}} + V(k), \tag{4}$$

where:

$$H_{k,m} = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{l=0}^{L-1} h(n,l)\mathrm{e}^{-\mathrm{j}\frac{2\pi}{N}(k(\tau_l - n)+nk)}. \tag{5}$$

Writing (4) in vector form representation gives

$$\underline{\mathbf{Y}} = \mathbf{H}\underline{\mathbf{S}} + \underline{\mathbf{V}}, \tag{6}$$

thus the main diagonal of the matrix **H** represents the channel response at the subcarriers, while the off-diagonals represent the ICI components result from the frequency offsets and the time variation of the channel.

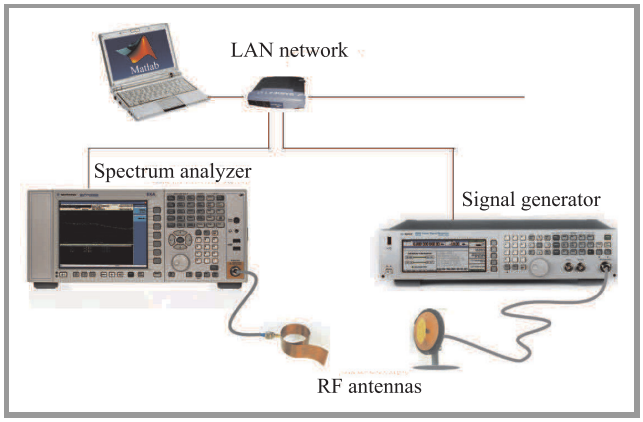


Fig. 1. The measurement system setup.

The measurement setup of the OFDM system is shown in Fig. 1. As shown from the figure, the measurement system is a combination of software and hardware components.

2.1. Software Setup

At an external PC, the baseband IQ (in-phase and quadrature-phase) OFDM signal based on IEEE 802.11g standard is generated using Matlab code. The OFDM symbols are arranged in a few frames to be transmitted. At the beginning of each frame, some training symbols (preambles) are inserted, which are used for packet detection and frequency and time offset estimation. The time domain baseband signal is sent via LAN connection to the transmitter, in order to be up-converted and transmitted through the wireless channel. The commands for controlling the power and the carrier frequency are sent to the transmitter before sending the baseband signal. On the other hand, the IQ data from the captured signal at the receiver is sent back to the external PC via the LAN for processing. Packet detection, frequency and time offset estimation and pilot-based channel estimation are all accomplished using Matlab.

2.2. Hardware Setup

2.2.1. The Transmitter

For transmitting the real time multicarrier signal, an MXG signal generator N5182A [11] has been used. This signal generator works on a frequency range between 100 kHz to 6 GHz and transmission power up to 17 dBm. The baseband IQ data is generated using Matlab and sent to the signal generator via LAN connection. The signal generator saves the IQ data, applies the IF modulation and RF up-conversion and then transmits the signal with predefined power and carrier frequency. In order to have a continuous transmission, the signal generator sends the same signal repeatedly.

2.2.2. The Antennas

The transmit antenna works between 2.4 GHz and 2.6 GHz with gain 9 dBi at the resonance frequency [12], while the receive antenna has an ultra-wideband characteristics in the frequency range from 800 MHz to 18 GHz [13]. Both antennas have been fabricated in the Institute of Electronics, Signal Processing and Communications (IESK) at the Otto von Guericke University of Magdeburg, Germany.

2.2.3. The Receiver

The EXA spectrum analyzer from Agilent [14] has been used as a receiver. This device has an absolute sensitivity of -79.4 dBm, a dynamic range of 93.1 dB, and a maximum input power of +30 dBm. The reviver down converts the RF signal and generates the IQ baseband data which is sent back to the external PC for processing.

3. Wireless Channel Characterization

Estimating the channel parameters such as the number of multipaths, the delay spread, and the Doppler spread allows for suitable design and robust behavior of the system. In this section some measurements have been made in order to estimate the channel taps and the delay spread of the channel. An OFDM signal has been sent through the channel, captured at the receiver and then processed in order to calculate the channel transfer function (CTF) and the channel impulse response (CIR). The measurements have been conducted in Lab312, which is one of the IESK laboratories at the University of Magdeburg. The lab was clustered with furniture, test and measurement equipments. The topology of the lab and the test scenario is shown in Fig. 2.
During the measurements, the transmit antenna remained fixed at the same place with different orientations, while the receive antenna was moved to different places and orientations. The signal used for measurement was a WLAN signal with 20 MHz bandwidth and carrier frequency of 2.412 GHz. The channel transfer function was calculated as the ratio between the transmitted frequency domain OFDM symbols by the received frequency domain

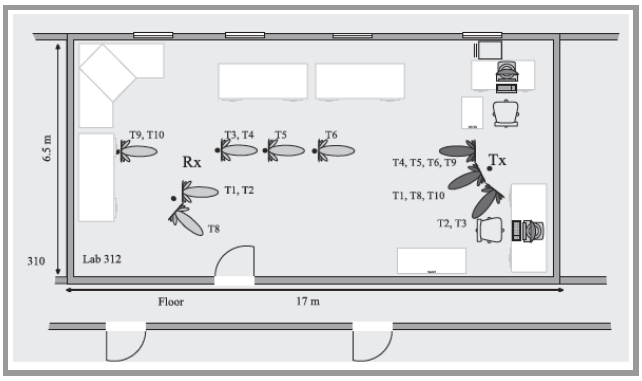


Fig. 2. The floor plan of the measurement location.

symbols. The channel impulse response h was calculated by simply applying an IFFT on the CTF.

One of the important parameters required for the characterization of any wireless channel is the root mean square delay spread τ_{rms}, which is calculated from the CIR [15] as

$$\tau_{rms} = \sqrt{\frac{\sum_{l=0}^{L-1} h_l^2 \tau_l^2}{\sum_{l=0}^{L-1} h_l^2} - \left(\frac{\sum_{l=0}^{L-1} h_l^2 \tau_l}{\sum_{l=0}^{L-1} h_l^2} \right)^2}. \tag{7}$$

In order to get a clear idea about the multipath channel, the Ricean K-factor of the channel has also been calculated by taking the ratio between the signal power in the main path to the power in the scattered paths as follows:

$$K = \frac{h_{LOS}^2}{\sum_{l=0}^{L-1} h_l^2 - h_{LOS}^2}. \tag{8}$$

Basically, the Ricean K factor lies between 0, in case of Rayleigh channel, and ∞ in the case of single path transmission. The measured channel parameters are summarized in Table 1.

Table 1

The measured channel characteristics

Test	Delay spread τ_{rms} [ns]	K factor [dB]	C_f [kHz]
T1	306.3	8.8195	653
T2	236	5.7749	847
T3	138	10.5308	1449
T4	97	10.8814	2062
T5	61	11.2222	3279
T6	58	11.3988	3448
T7	143	9.0309	1399
T8	233	10.2531	858
T9	151	10.8991	1325
T10	183	10.5308	1093

In addition, the coherence bandwidth of the channel has been approximately calculated according to the 50% of the frequency correlation function [15] as

$$C_f \approx \frac{1}{5\tau_{rms}}. \tag{9}$$

The measured parameters show that τ_{rms} lies between 306 ns and 58 ns for different locations. Figure 3 shows

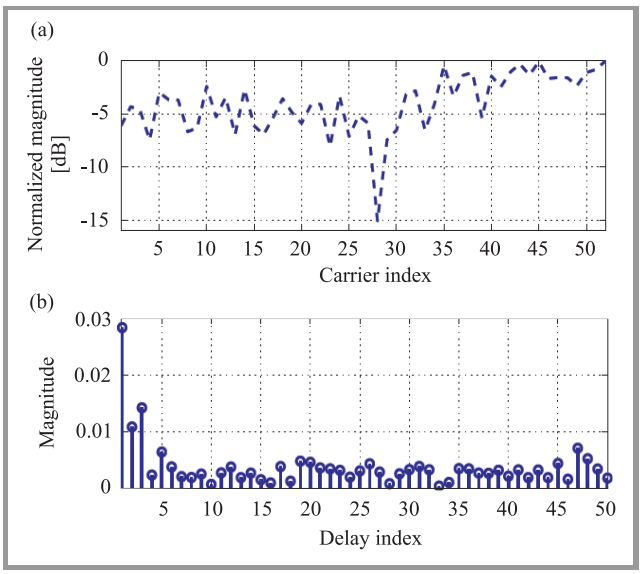


Fig. 3. The CTF (a) and the CIR (b) of a multipath channel (T1).

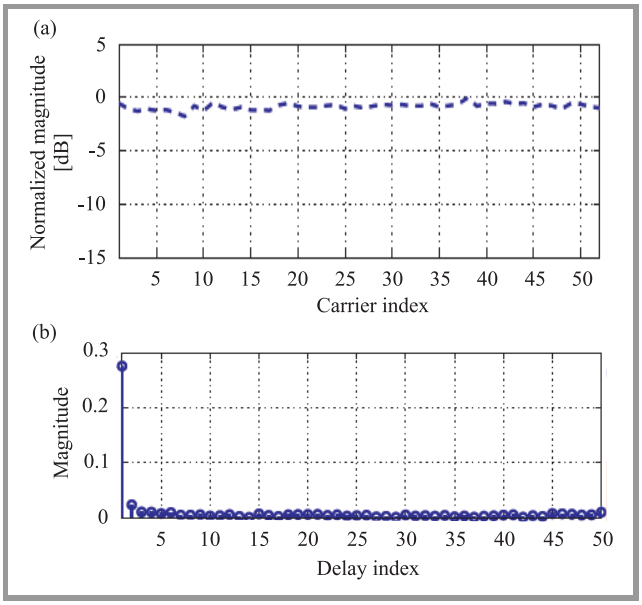


Fig. 4. The CTF (a) and the CIR (b) of a flat fading channel (T5).

the CTF and the CIR of the experiment T1, which represents a multipath channel with $\tau_{rms} = 306$ ns, while Fig. 4 represents the experiment T5, which exhibits a strong line of sight (LOS) transmission.

4. A Joint Frequency Offset and Wiener Filtering Channel Estimation Scheme

The time and frequency synchronization play an important role in designing and implementing the practical systems. In WLAN system, the signal is headed with a preamble at the beginning of each frame. This preamble is basically used for time synchronization, frequency offset estimation and for the initial channel estimation. The preamble contains two parts, the first one is called short training symbols (STS), which contains 10 short symbols occupy 8 μs. The STS are used for time synchronization and course frequency offset (CFO) estimation. The second part is called long training symbols (LTS), which contains two long symbols each with time duration of 3.2 μs as well as a GI with time duration of 1.6 μs. The LTS are used for further enhancing the time synchronization and for fine frequency offset (FFO) estimation as well as initial channel estimation.

The OFDM symbol of the WLAN system contains 64 carriers, 52 of them are active carriers, while the other 12 carriers work as frequency guard at both sides of the spectrum. For channel estimation and equalization proposes, 4 carriers out of the 52 are booked for pilots.

In order to start processing the received OFDM signal, the signal packets need to be detected first and the first sample of the OFDM symbols needs to be indicated in order to apply the FFT operation on the right symbol samples. A suitable method for detecting the packet is based on the autocorrelation of the received preamble [16], which contains short and long training symbols. The auto correlation is written in the following form

$$Z(n) = \sum_{c=0}^{N_{TS}-1} y^*(n+c)y(n+c+N_{TS}), \tag{10}$$

where $y(n)$ represents the nth incoming sample of the baseband signal, $y^*(n)$ represents the conjugate of $y(n)$ and $N_{TS} = 16$ samples in case of short training symbols. In order to avoid the expected variance of the incoming signal power, the autocorrelation needs to be normalized by a moving sum of the received signal power [17]. The moving sum of the received power can be written as

$$P(n) = \sum_{c=0}^{N_{TS}-1} |y(n+c+N_{TS})|^2. \tag{11}$$

The normalized autocorrelation used for packet detection reads

$$M(n) = \frac{|Z(n)|^2}{P(n)}. \tag{12}$$

The packet is detected at the first peak of $M(n)$ after a predefined threshold, which should be chosen to minimize the possibility of false alarm. Using only the short training symbols in packet detection gives a reasonable performance in case of good conditions of the channel, however, because of the expected time offsets at bad channel conditions, it is better to use both short and long training symbols for this purpose.

Figure 5 shows the autocorrelation function $M(n)$, whose first peak indicates the beginning of the packet (the first sample of the short training symbols), while the second peak indicates the first sample of the GI of the long training symbols. By taking the difference between the two peaks, the time offset, if found, can be estimated and then the start of the packet is corrected.

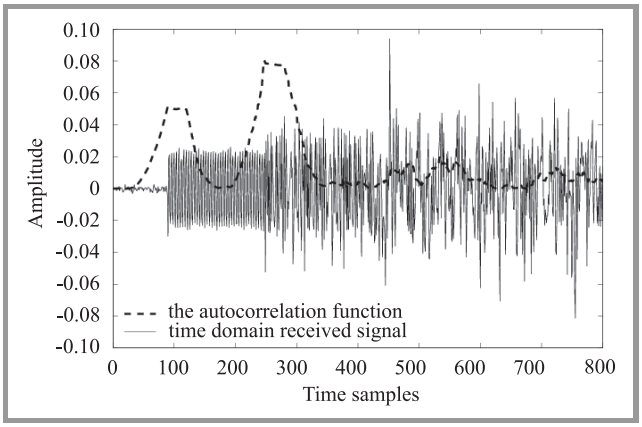


Fig. 5. The packet detection based on autocorrelation of the preamble.

Figure 6 presents a comparison between the spectrum of the transmitted and the received OFDM signals. The figure shows the effect of the widowing at the receiver side, which reduces the side loops of the spectrum.

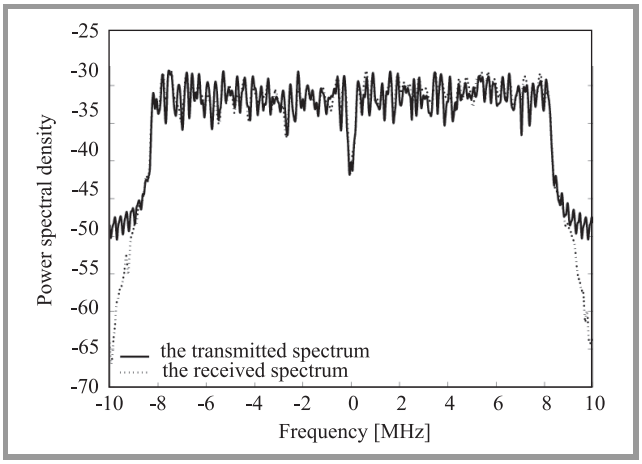


Fig. 6. A comparision between the transmitted and measured spectrum of the OFDM signal.

As a consequence of the frequency offsets between the oscillators of the transmitter and the receiver, phase errors and ICI occur on the received signal and lead to degrading the OFDM system performance. The frequency offset can be estimated from the autocorrelation of the training symbols in the preamble. If there is a frequency offset f_e,

the autocorrelation function will be modified by the phase error as

$$Z(n) = \mathrm{e}^{-\mathrm{j}2\pi f_e N T_S} \sum_{c=0}^{N_{TS}-1} |y(n+c)|^2. \tag{13}$$

The phase introduced by the frequency offset is calculated as

$$\phi = \angle Z(n).$$

The course frequency offset CFO is calculated in case of the short symbols as

$$f_e = \frac{\phi}{2\pi 16T}, \tag{14}$$

where T is the sampling time which is (50 ns) for WLAN system. For better frequency offset estimation, the estimation is accomplished in two stages CFO and then FFO estimations. After the compensation of the frequency offset, a Wiener filtering method described is applied to further correcting for the phase rotation, and at the same time equalizing the OFDM symbols. In order to enhance the performance of the frequency offset compensation, a minimum mean square error (MMSE) [18]–[20] estimator based on Wiener filtering has been used. The frequency domain MMSE estimator minimizes the following formula:

$$E\{|e|^2\} = E\{|\underline{\mathbf{H}}_p - \underline{\widehat{\mathbf{H}}}_p|^2\}.$$

The frequency domain channel matrix can be estimated by linearly filtering the received signal for each channel element. The estimated element (m,k) of the channel matrix is expressed as follows:

$$\hat{H}_{m,k} = \underline{\mathbf{W}}_{m,k}^H \underline{\mathbf{Y}}, \tag{15}$$

where $\underline{\mathbf{W}}_{m,k}$ is the Wiener filter vector that is used to estimate the channel matrix element (m,k).
The Wiener filter for each channel element is designed and optimized according to the MMSE criterion as follows:

$$\underline{\mathbf{W}}_{m,k}^{(opt)} = \arg \min_{\underline{\mathbf{W}}} E\left\{ \left| H_{m,k} - \underline{\mathbf{W}}_{m,k}^H \underline{\mathbf{Y}} \right|^2 \right\}. \tag{16}$$

The solution of Eq. (16) is given by [21]

$$\underline{\mathbf{W}}_{m,k}^{(opt)} = \mathbf{R}^{-1} \underline{\mathbf{P}}_{m,k}, \tag{17}$$

where $\mathbf{R} = E\{\underline{\mathbf{Y}}\underline{\mathbf{Y}}^H\}$ is the autocorrelation of the received frequency domain signal and $\underline{\mathbf{P}}_{m,k} = E\{H_{m,k}^* \underline{\mathbf{Y}}\}$ is the cross correlation between the received signal and the actual frequency response of the channel. The derivation of these correlation matrices is detailed in Appendix.
In order to correct the phase rotation of the captured OFDM symbols, an iterative scheme is applied to find the phase error that minimizes the following cost function:

$$MSE = E\{|\underline{\mathbf{S}}_p - \underline{\hat{\mathbf{S}}}_p|^2\}, \tag{18}$$

where $\underline{\mathbf{S}}_p$ and $\underline{\hat{\mathbf{S}}}_p$ are vectors contain the transmitted and equalized pilots.

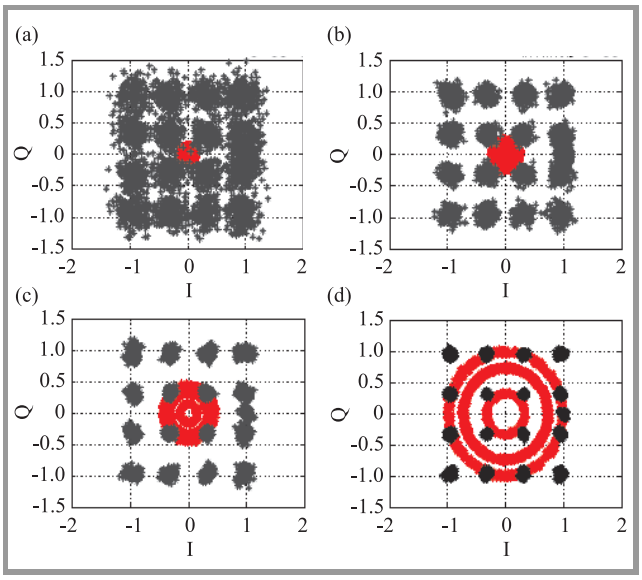


Fig. 7. The constellation of the equalized 16 QAM modulated OFDM symbols with different transmission power: (a) 0 dBm, (b) 5 dBm, (c) 10 dBm, (d) 17 dBm at distance ($d = 2$ m); gray: received symbols, black: equalized symbols.

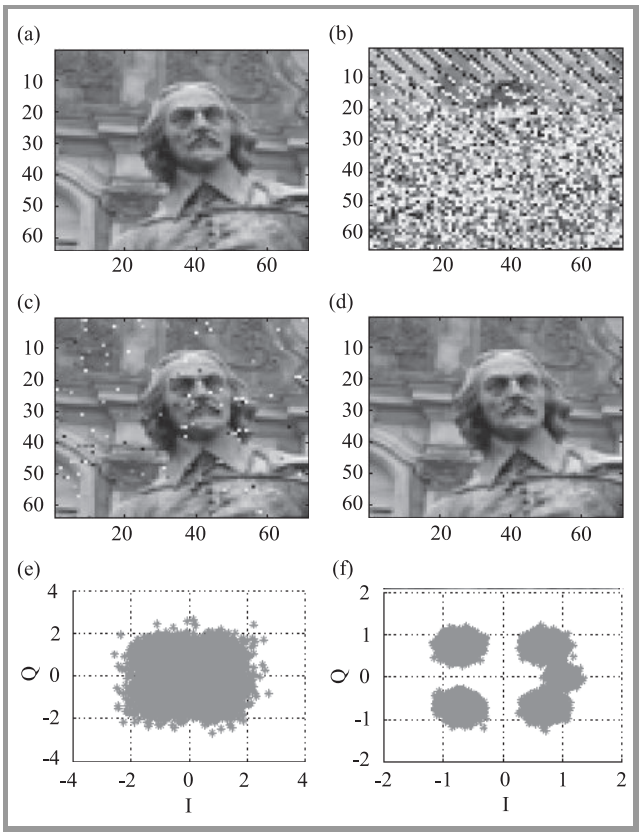


Fig. 8. An example of transmitting a gray scaled image through the wireless channel: (a) the transmitted image; (b) the received un-equalized image; (c) the equalized image with a transmission power -10 dBm; (d) the equalized image with a transmission power 0 dBm; (e) the constellation of the equalized image with -10 dBm power; (f) the constellation of the equalized image with 0 dBm power.

$\underline{\hat{\mathbf{S}}}_p$ is estimated using LS criterion as follows:

$$\underline{\hat{\mathbf{S}}}_p = \mathcal{D}(\underline{\hat{\mathbf{H}}}_p)^{-1}\underline{\mathbf{Y}}_p, \tag{19}$$

where $\underline{\mathbf{Y}}_p$ is the received pilots, $\mathcal{D}(\cdot)$ is vector to diagonal matrix conversion, and $\underline{\hat{\mathbf{H}}}_p$ is the estimated CTF, which is estimated by Wiener filtering of the received signal as follows:

$$\hat{H}_p = (\mathbf{R}_p^{-1}\underline{\mathbf{P}}_p)^H\underline{\mathbf{Y}}_p. \tag{20}$$

According to Eqs. (24) and (25), the autocorrelation $\mathbf{R}_p$ and the cross-correlation $\underline{\mathbf{P}}_p$ contain information about the delay τ_l of each channel tap. In order to correct the phase rotation, the scheme is searching for τ_0 that minimizes Eq. (18), and then applied the resulting filter for channel equalization.

Figure 7 illustrates the constellation of the equalized 16 QAM modulated OFDM symbols with different transmitted power. For simplicity, we used pilot carriers with the same magnitude.

Figure 8 shows an example of transmitting and receiving a gray scaled image through the wireless channel. The size of the image was 5.4 KB, which required 756 OFDM symbols coded by a convolutional encoder with coding rate of $R = 0.5$.

Table 2 and Fig. 9 represent the measured BER and MSE for different transmission power using 16 QAM modulation scheme for antenna separation of 2 m.

Table 2
The measured system performance with 16 QAM modulation scheme

Transmission power [dBm]	MSE [dB]	BER
–10	–5.5000	0.213
–5	–10.0000	0.0771
0	–14.5000	0.0093
5	–18.9500	$2.5431 \cdot 10^{-5}$
10	–23.5000	0

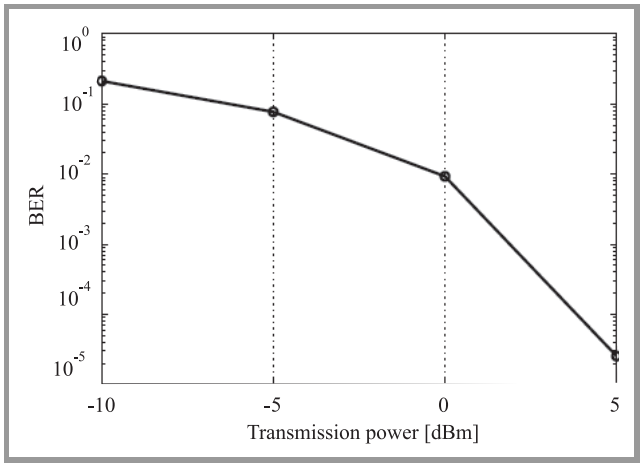


Fig. 9. The measured BER versus the transmission power with modulation scheme 16 QAM at distance ($d = 2$ m).

Table 3 and Fig. 10 represent the measured BER and MSE for different transmission power using 4 QAM modulation scheme for antenna separation of 2 m.

Table 3
The measured system performance with 4 QAM modulation scheme

Transmission power [dBm]	MSE [dB]	BER
–5	–11.5	$2.84 \cdot 10^{-4}$
0	–16	0
5	–21.2	0
7	–22.772	0
10	–26.5	0

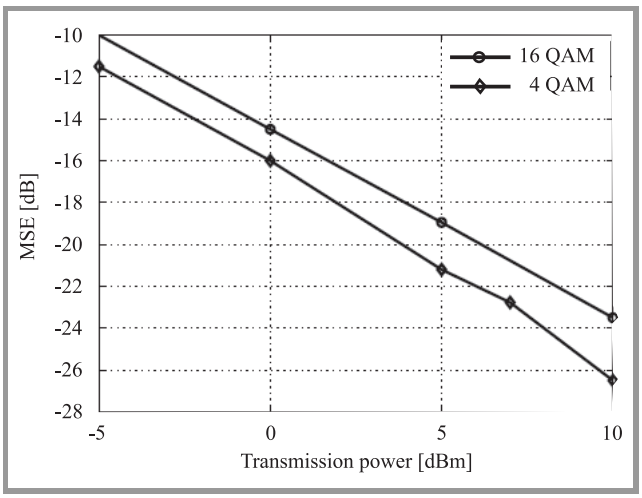


Fig. 10. The measured MSE of the OFDM symbols versus the transmission power with modulation scheme 16 QAM and 4 QAM at distance ($d = 2$ m).

Figure 11 illustrates the measured MSE of the received OFDM symbols for 16 QAM at transmission power

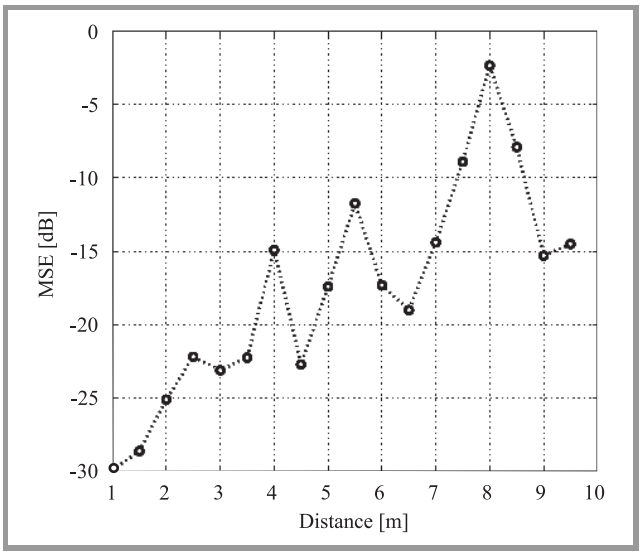


Fig. 11. The measured MSE of the OFDM symbols versus the distance between the transmitter and the receiver with modulation scheme 16 QAM and transmission power 17 dBm.

of 17 dBm and variant distance between the transmitter and the receiver. It can be seen from the figure that in addition to the path loss, the signal gets attenuated (shadowed) at certain distances because of the multipath effect.

5. Conclusion

This paper presents a method for estimating the wireless channel for OFDM system. A measurement system based on off-line measurement has been used for testing the OFDM signal. This system has been used to characterize the multipath channel and test the transmission, reception and channel estimation of the OFDM-based WLAN system. A method of joint channel and frequency offset estimation based on the signal preamble and Wiener filtering has been implemented. The measured results show that, combining the Wiener filtering and the autocorrelation can compensate for ICI and the phase rotations caused by frequency offsets and multipath effect.

Appendix

Calculating $\mathbf{R}$ and $\underline{\mathbf{P}}$

Assuming that $\mathbf{H}$, $\underline{\mathbf{S}}$, and $\underline{\mathbf{V}}$ are uncorrelated and making use of Eq. (4), the element (k,k') of $\mathbf{R}$ can be written as

$$\begin{aligned} R_{k,k'} &= E\{Y(k)Y(k')^*\} \\ &= E\Big\{\Big(\sum_{m=0}^{N-1} H_{m,k}S(m)\Big)\Big(\sum_{m'=0}^{N-1} H^*_{m',k'}S(m')^*\Big)\Big\} \\ &\quad + E\{V(k)V(k')^*\}. \end{aligned} \tag{21}$$

From Eq. (21), $\mathbf{R}$ can be separated into two parts, the pure correlation matrix and the noise

$$\mathbf{R} = \mathbf{R}' + \sigma_v^2\mathbf{I}, \tag{22}$$

where σ_v^2 is the noise variance, which is the first statistical parameter that needs to be estimated at the receiver. Also the element k' of $\underline{\mathbf{P}}_{m,k}$ is calculated as

$$\begin{aligned} P^{k'}_{m,k} &= E\{Y(k')H^*_{m,k}\} \\ &= E\Big\{\sum_{m'=0}^{N-1} H_{m',k'}H^*_{m,k}S(m')\Big\} \\ &\quad + E\{V(k')H^*_{m,k}\}. \end{aligned} \tag{23}$$

Assuming a wide sense stationary uncorrelated scattering (WSSUS) channel model, $R'_{k,k'}$ and $P^{k'}_{m,k}$ can be given by the following equations [22]:

$$\begin{aligned} R'_{k,k'} &= \sigma_s^2 \sum_{m\notin\underline{\mathbf{S}}_p}\sum_{l=0}^{L-1}\int_{-\infty}^{\infty}\Phi_{l,\mathrm{v}}(k-m,k'-m)d\mathrm{v} \\ &\quad + \sum_{m,m'\in\underline{\mathbf{S}}_p} S(m)S^*(m')\sum_{l=0}^{L-1}\mathrm{e}^{-\mathrm{j}2\pi(m-m')\Delta f\tau_l} \\ &\quad \int_{-\infty}^{\infty}\Phi_{l,\mathrm{v}}(k-m',k'-m')d\mathrm{v}; \end{aligned} \tag{24}$$

$$\begin{aligned} P^{k'}_{m,k} &= \sum_{m'\in\underline{\mathbf{S}}_p} S(m')\sum_{l=0}^{L-1}\mathrm{e}^{-\mathrm{j}2\pi(m-m')\Delta f\tau_l} \\ &\quad \int_{-\infty}^{\infty}\Phi_{l,\mathrm{v}}(k'-m',k-m)d\mathrm{v}, \end{aligned} \tag{25}$$

where

$$\Phi_{l,\mathrm{v}}(a,b) = \mathrm{sinc}(\mathrm{v}T_s - a)\,\mathrm{sinc}(\mathrm{v}T_s - b)D_l(\mathrm{v}),$$

where $\mathrm{sinc}(\cdot)$ is the sinc function and $D_l(\mathrm{v})$ is the Doppler spectrum for each tap. This term can be ignored in the case of time-invariant channels.

References

[1] C. Smith and J. Meyer, *3G Wireless with 802.16 and 802.11*. New York: McGraw-Hill, 2004.

[2] "Radio Broadcasting Systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers", ETSI EN 300 401 V1.4.1, 2006.

[3] "Digital Video Broadcasting (DVB); Interaction channel for Digital Terrestrial Television (RCT) incorporating Multiple Access OFDM", ETSI EN 301 958 V1.1.1, 2002.

[4] P. Lescuyer and T. Lucidarme, *Evolved Packet System (EPS): The LTE and SAE Evolution of 3G UMTS*. Chichester: Wiley, 2008.

[5] S. Hara and R. Prasad, *Multicarrier Techniques for 4G Mobile Communications*. Boston, London: Artech House, 2003.

[6] J. Linartz and A. Grokhov, "New equalization approach for OFDM over dispersive and rapidly time varying channel", in *Proc. PIMRC 2000 Conf.*, London, UK, 2000.

[7] Y. Liao and K. Chen, "Estimation of stationary phase noise by the autocorrelation of the ICI weighting function in OFDM systems", *IEEE Trans. Wirel. Commun.*, vol. 5, no. 12, pp. 3370–3374, 2006.

[8] M. Chang, "A novel algorithm of interference self-cancellation for OFDM systems", *IEEE Trans. Wirel. Commun.*, vol. 6, no. 8, pp. 2881–2893, 2007.

[9] W. Jeon, K. Chang, and Y. Cho, "An equalization technique for orthogonal frequency-division multiplexing systems in time-variant multipath channels", *IEEE Trans. Wiurel. Commun.*, vol. 47, no. 1, pp. 27–32, 1999.

[10] R. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. Boston, London: Artech House, 2000.

[11] "Agilent N5182A MXG Vector Signal Generator", Agilent Technol., 2007, Data sheet, ftp://ftp.testequity.com/pdf/n5182a.pdf

[12] D. Mittelstrasz, "Design Einer Antenne Bestehend aus Primaerstrahler und Dielektrischer Linse fuer das ISM-Band 2.4 GHz", Studienarbeit: Otto-von-Guericke Universitaet Magdeburg, 2007.

[13] A. Bandyopadhyay, M. Anis, A. Joestingmeier, T. Meyer, and A. Omar, "Analysis of a compact ultra-wideband antenna for radio frequency applications", in *Proc. Antenn. Propag. Soc. Int. Symp.*, Honolulu, Hawaii, USA, 2007, vol. 9, pp. 1965–1968.

[14] "Agilent EXA Signal Analyzer N9010A", Agilent Technol., 2009, Data Sheet, http://cp.literature.agilent.com/litweb/pdf/5989-6529EN.pdf

[15] T. Rappaport, *Wireless Communications: Principles and Practice*. New York: Prentice Hall, 2001.

[16] J. Heiskala and J. Terry, *OFDM Wireless LANs: A Theoretical and Practical Guide*. New York: Sams, 2001.

[17] T. Schmidl and D. Cox, "Robust frequency and timing synchronization for OFDM", *IEEE Trans. Commun.*, vol. 45, no. 12, pp. 1613–1621, 1997.

[18] H. Van Trees, *Detection, Estimation, and Modulation Theory. Part I: Detection, Estimation, and Linear Modulation Theory*. New York: Wiley, 1968.

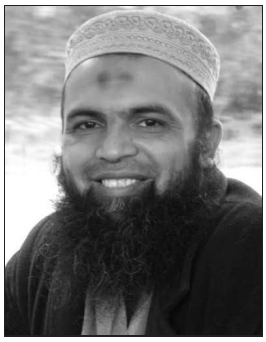
[19] H. Poor, *An Introduction to Signal Detection and Estimation*. Berlin: Springer, 1994.

[20] S. Kay, *Fundamentals of Statistical Processing*. Volume I: *Estimation Theory*. Signal Processing Series. Upper Saddle River: Prentice Hall, 1993.

[21] M. Barkat, *Signal Detection and Estimation*. Norwood: Artech House, 2005.

[22] C. Sgraja and J. Linder, "Estimation of rapid time-variant channels for OFDM using wiener filtering", *IEEE Proc. Commun.*, vol. 4, pp. 2390–2395, May 2003.

Ali Ramadan Ali received the B.Sc. degree in electrical engineering from Altahadi Universiy, Sirt, Libya, in 1997 and the M.Sc. degree in electrical engineering from Darmstadt University of Applied Science, Germany, in 2004. He works toward the Ph.D. degree at the Otto-von-Guericke University of Magdeburg, Germany. His current research field covers the areas of digital signal processing for OFDM systems and microwave filter design. He authored or co-authored over 23 technical papers in the field of wireless communications and microwave filters.
e-mail: ali.ramadan@ovgu.de
Otto-von-Guericke University of Magdeburg
Institute for Electronic, Signal Processing
and Communications (IESK)
Chair of Microwave and Communication Engineering (HF)
P.O. Box 4120
Universitaetsplatz 2
39016 Magdeburg, Germany



Tariq Jamil Khanzada received the M.Eng. degree in communication systems and networking, and B.Eng. degree in computer systems from Mehran University of Engineering and Technology (MUET), Pakistan, in 2004 and 1999, respectively. He currently works at the Institute of Electronics, Signal Processing and Communications (IESK), Otto-von-Guericke University of Magdeburg, Germany, since June 2005 in order to pursue his Ph.D. degree. Prior to start his Ph.D. studies he has been working as an Assistant Professor at the Department of Computer Systems and Software Engineering at MUET, since 2004 and as a lecturer since 2001. He is also a co-author of a practical book on computer workshop published in 2004. He is also an author of some international research publications.
e-mail: Tariq.Khanzada@ovgu.de
Otto-von-Guericke University of Magdeburg
Institute for Electronic, Signal Processing
and Communications (IESK)
Chair of Microwave and Communication Engineering (HF)
P.O. Box 4120
Universitaetsplatz 2
39016 Magdeburg, Germany

Abbas Omar received the B.Sc. and M.Sc. degrees in electrical engineering from Ain Shams University, Cairo, Egypt, in 1978 and 1982, respectively, and the Ph.D. degree in electrical engineering from the Technical University of Hamburg, Germany, in 1986. Since 1990, he has been a Professor of electrical engineering, and since 1998, a Director of the chair of microwave and communication engineering with Otto-von-Guericke University of Magdeburg, Germany. He authored or co-authored over 250 technical papers extending over a wide spectrum of research areas. His current research fields cover the areas of microwave, nuclear magnetic resonance, and ultrasonic imaging, remote sensing, microwave measurements, indoor and outdoor positioning systems, wideband wireless (terrestrial and mobile) communication, subsurface tomography and ground penetrating radar, ultra-wideband antennas, and field theoretical modeling of microwave systems and components. He is a member of the Editorial Board of some international technical periodicals.
e-mail: a.omar@ieee.org
Otto-von-Guericke University of Magdeburg
Institute for Electronic, Signal Processing
and Communications (IESK)
Chair of Microwave and Communication Engineering (HF)
P.O. Box 4120
Universitaetsplatz 2
39016 Magdeburg, Germany

Paper

Channel Identification Using Chaos for an Uplink/Downlink Multicarrier Code Division Multiple Access System

Miloud Frikel, Said Safi, Boubekeur Targui, and Mohammed M'Saad

Abstract—A scheme of chaotic spreading sequence for multicarrier code division multiple access system (MC-CDMA) is proposed to estimate the transmission channel. This system spreads spectrum and identifies the channel, simultaneously. The proposed scheme uses a chaotic sequence generated by a logistic map as a training signal and estimate channel parameters according to dynamics of the chaotic sequence. Encoding data by chaotic sequences is first built and then the orthogonal codes are used to spread the encrypted data for multiusers scheme. At the reception, first the channel parameters are identified using a training chaotic sequence in order to equalize the received data, and then the encrypted information is decoded for the desired user. The studies reveal that the proposed system (chaos plus orthogonal codes) significantly outperforms the Walsh-Hadamard code spreading in MC-CDMA system. The performance of the system is considered in the multiuser case by means of simulation. The simulation result shows that the proposed chaotic code spreading approach for channel identification achieves significant improvement in the channel identification, comparing to using others training sequence or the least square method.

Keywords—channel identification, chaos code, equalization, MC-CDMA system.

1. Introduction

The chaotic signals have some properties such as broadband, orthogonality and complexity aspects, which motivate researches in the area of communication and signal processing to investigate if chaos based communication offers advantages over classical communication systems in the last years.

Potential applications of chaos resulting directly from those aspects are spread-spectrum, multiuser communication, and cryptography [1]–[4]. In chaotic communications, a chaotic sequence is transmitted through the transmission channel [5]. If the channel is not ideal, which is often the case in practice, the transmitted signal is corrupted before it reaches the receiver. Hence, channel equalization is required to reduce the bit error rate of the receiver as small as possible.

In many practical cases, channel parameters are unknown. Hence, channel equalization must be performed from the corrupted signal alone, and this is called the blind channel equalization [6], [7]. In classical communication systems, most of the channel equalization algorithms are based on the statistical properties of the transmitted signal [8]. However, since a chaotic sequence is a deterministic signal, the statistics-based equalization techniques will not achieve optimum estimation accuracy for chaotic communication systems because they do not take into account the inherent properties specific to a chaotic signal. Various chaotic blind identification and equalization techniques based on different properties of the transmitted chaotic signal have been developed recently [9].

The multicarrier code division multiple access (MC-CDMA) system is based on the combination of code division multiple access (CDMA) and orthogonal frequency division multiplexing (OFDM) [10]–[13] which is potentially robust to channel frequency selectivity. Furthermore, it has good spectral efficiency, multiple access capability and it's easy to be implemented with fast Fourier transform (FFT). For the scheme based on a combination of CDMA and multicarrier technique spreads the original data stream over different subcarriers using a spreading code in the frequency domain [14]. Indeed, in the past decade, there is growing need for technological innovations to satisfy the increase demand for personal wireless radio communications. This technology must be able to allow users to efficiency share common resources, whether it involves the frequency spectrum and computational load. That is why the MC-CDMA [14], the one of representative of the multicarrier techniques, has been considered as a promising system for the next generation of wireless communication. One large advantage of this technology is its robustness in case of multi-path propagation, and it's capable to combat frequency selective fading, flexible to generate different data rates and provides bandwidth efficiency [12].

Multicarrier systems are very sensitive to synchronization errors such as carrier frequency offset and phase noise. Synchronization errors cause loss of orthogonality among subcarriers and considerably degrade the performance especially when large number of subcarriers presents. There have been many approaches on synchronization algorithms in literature [15]–[17]. MC-CDMA designed for multiusers using the same channel, so the problem of channel identification appears, in this paper we propose to encode chaoti-

cally data before its spreading in order to use this sequence for system identification.

2. Time and Frequency Domain MC-CDMA System Description

In order to transmit data of a large number of users, the frequency band should be, optimally, used. The objective is to transmit, in simultaneous over the same channel the maximum of information, so the use of multiplexing. Indeed, in CDMA [15], the users have access, in the same time, to the totality of the frequency band, in the receiver, to distinguish between them, we use a different codes affected for each user. That was possible thanks to the technique of spectral spreading, in condition that the emitted signals by different users have some proprieties allowing them to separate.

In opposition of the other techniques of multi-access such frequency division multiple access (FDMA) and time division multiple access (TDMA), where the capacity of the number of users is limited by the frequency and time resources, respectively, the number of users in CDMA is fixed by the proprieties of used spreading codes [18]. That is why the CDMA is an alternative to the others multiplexing techniques to increase the re-use frequency factor and eventually the spectral efficiency of communication systems. A different approach to further increase the system capacity without allocating additional frequency spectrum is the use of code multiplexing.

The MC-CDMA modulator spreads the data of each user in frequency domain (we consider a binary phase shift keying (BPSK) data), the complex symbol g_j of each user j is, firstly, multiplied by each chips $c_{j,k}$ of spreading code SC_j, and then applied to the modulator of multicarriers. Each subcarrier transmits an element of information multiplied by a code chip of that subcarrier.

We consider, for example, the case where the length L_c of spreading code is equal to the number N of subcarriers. The optimum space between two adjacent subcarriers is equal to inverse of duration T_c of chip of spreading code in order to guarantee the orthogonality between subcarriers. The MC-CDMA signal is given by

$$s(t) = \sum_{k=0}^{N-1} a_k c_{m,k} e^{2i\pi f_k t}, \tag{1}$$

where m is the user number.

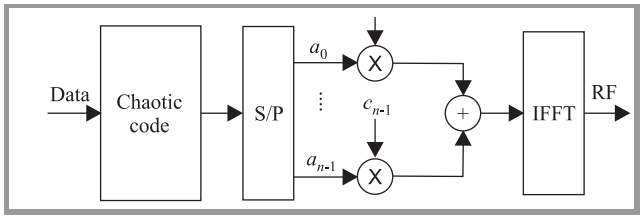


Fig. 1. Chaotic MC-CDMA transmitter. Explanation: S/P – serial/parallel.

We consider the channel invariant in time and characterized by P paths of magnitudes β_p and phase θ_p. The impulse response is given by

$$h(\tau) = \sum_{p=0}^{P-1} \beta_p e^{i\theta_p} \delta(\tau - \tau_p).$$

The relationship between the emitted signal $s(t)$ and the received signal $r(t)$ is given by

$$r(t) = h(t) * s(t) + n(t),$$

where $*$ is the convolution product and $n(t)$ is the additive white Gaussian noise (AWGN):

$$\begin{aligned} r(t) &= \int_{-\infty}^{+\infty} \sum_{p=0}^{P-1} a_k \beta_p e^{i\theta_p} \delta(\tau - \tau_p) s(t-\tau) d\tau + n(t) \\ &= \sum_{p=0}^{P-1} a_k \beta_p e^{i\theta_p} s(t-\tau) + n(t), \end{aligned} \tag{2}$$

where P is the number of paths.

The transmitter of the MC-CDMA scheme using both time domain and frequency domain is shown in Fig. 1.

The discrete chaotic sequence is: $\chi = [a_0, a_1, ..., a_{N-1}]$, where N is the chaotic sequence duration.

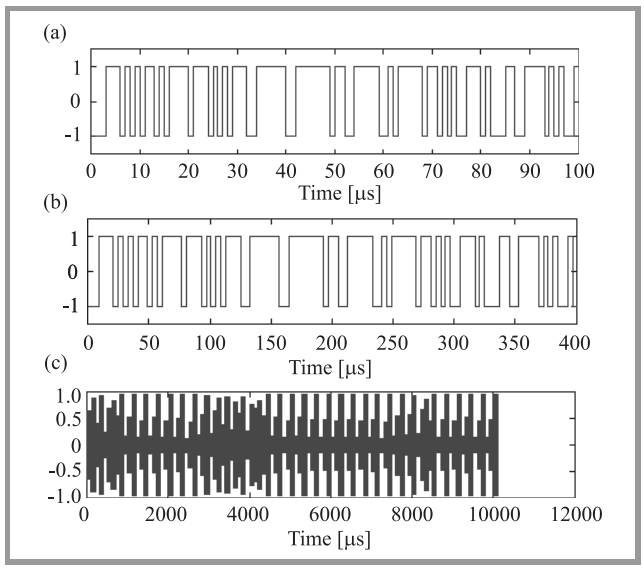


Fig. 2. Binary phase shift keying data (a), spreaded by the Walsh-Hadamard code (b), and by the chaotic code (c).

In Fig. 2, we represent an example of BPSK data spread by the Walsh-Hadamard code and by the chaotic code. In the following equation we represent an example of the Walsh-Hadamard code C (e.g., for four users):

$$C = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

In a system of M users the emitted signal through a channel is given by

$$s(t) = \sum_{m=0}^{M-1} \sum_{k=0}^{N-1} a_m[k] c_{m,k} e^{2i\pi f_k t}. \quad (3)$$

The received signal after passing through the channel is:

$$r(t) = \sum_{m=0}^{M-1} \sum_{p=0}^{P-1} \sum_{k=0}^{N-1} a_m[k] \beta_p e^{i\theta} g_{m,k} c_{m,k} e^{2i\pi(f_0 + k/T_s)} \times \delta(t-\tau_p) + n(t). \quad (4)$$

At the reception, we demodulate the signal according to N subcarriers, and then we multiply the received sequence by the code of the desired user (Fig. 3). Equalization is, then, applied to estimate the frame $g_{m,k}$.

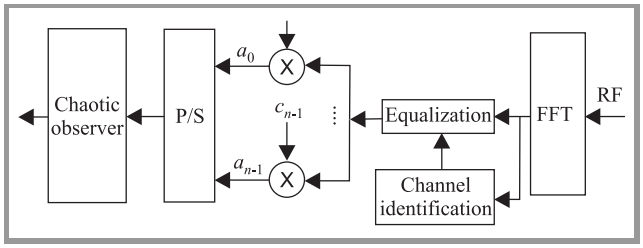


Fig. 3. Chaotic MC-CDMA receiver.

When there are M active users, the received signal is:

$$r(t) = \sum_{m=0}^{M-1} \sum_{k=0}^{N-1} h_{m,k} c_m[k] a_m[k] \cos(2\pi f_c t + 2\pi k \frac{F}{T_b} t + \theta_{m,k}) + n(t), \quad (5)$$

where the effects of the channel have been included in $h_{m,k}$ and $\theta_{m,k}$, $n(t)$ is AWGN with a one-sided power spectral density of N_0.

We assume that $m = 0$ corresponds to the desired signal. With this model, there are N matched filters with one matched filter for each subcarrier. The output of each filter contributes one component to the decision variable, ϑ_0. Each matched filter consists of an oscillator with a frequency corresponding to the frequency of the particular BPSK modulated subcarrier that is of interest and an integrator. In addition, a phase offset equal to the phase distortion introduced by the channel, $\theta_{m,k}$, is included in the oscillator to synchronize the receiver to the desired signal in time.

To extract the desired signal's component, the orthogonality of the codes is used. For the kth subcarrier of the desired signal, the corresponding chip, $c_0[k]$, from the desired user's code is multiplied with it to undo the code. If the signal is undistorted by the channel, the interference terms will cancel out in the decision variable due to the orthogonality of the codes. As the channel will distort the subcarrier components, equalization gain, $g_{0,k}$, may be included for each matched filter branch of the receiver.

Applying the receiver model to the received signal given in Eq. (4) yields the following decision variable for the kth data symbol assuming the users are synchronized in time:

$$\vartheta_0 = \sum_{m=0}^{M-1} \sum_{k=0}^{N-1} h_{m,k} c_m[k] a_m[k] \frac{2}{T_b} \int_{kT_b}^{(k+1)T_b} \cos\left(2\pi f_c t + 2\pi k \frac{F}{T_b} t + \theta_{m,k}\right) \cos\left(2\pi f_c t + 2\pi k \frac{F}{T_b} t + \widehat{\theta}_{0,k}\right) dt + \eta, \quad (6)$$

where $\widehat{\theta}_{0,k}$ denotes the receiver's estimation of the phase at the kth subcarrier of the desired signal and the corresponding AWGN term, η is given as

$$\eta = \sum_{k=0}^{N-1} \int_{kT_b}^{(k+1)T_b} n(t) \frac{2}{T_b} \cos(2\pi f_c t + 2\pi k \frac{F}{T_b} t + \widehat{\theta}_{0,k}) dt. \quad (7)$$

Assuming perfect phase correction, $\widehat{\theta}_{0,k} = \theta_{0,k}$, the decision variable reduces to

$$\vartheta_0 = a_0[k] \sum_{k=0}^{N-1} h_{0,k} + \sum_{m=0}^{M-1} \sum_{k=0}^{N-1} a_m[k] c_m[k] c_0[k] h_{m,k} \cos(\widehat{\theta}_{m,k}) + \eta, \quad (8)$$

where $\widehat{\theta}_{m,k} = \theta_{0,k} - \theta_{m,k}$. If $\theta_{0,k}$ and $\theta_{m,k}$ are i.i.d. uniform r.v.'s on the interval $[0, 2\pi]$, then $\widehat{\theta}_{m,k}$ is also uniformly distributed on the interval $[0, 2\pi]$. Note that the decision variable consists of three terms.

The first term corresponds to the desired signal's component, the second corresponds to the interference and the last corresponds to a noise term:

$$\vartheta_0 = \xi_{inf} + \beta_{int} + \eta, \quad (9)$$

where ξ_{inf} and β_{int} are the terms of information and interferences, respectively, defined by

$$\xi_{inf} = a_0[k] \sum_{i=0}^{N-1} h_{0,k}, \quad (10)$$

$$\beta_{int} = \sum_{m=0}^{M-1} \sum_{k=0}^{N-1} a_m[k] c_m[k] c_0[k] h_{m,k} \cos(\widehat{\theta}_{m,k}),$$

$$\eta = \sum_{k=0}^{N-1} \int_{kT_b}^{(k+1)T_b} n(t) \frac{2}{T_b} \cos(2\pi f_c t + 2\pi i \frac{F}{T_b} t + \widehat{\theta}_{0,k}) dt.$$

The noise can be approximated by a zero-mean Gaussian random variable with the following variance:

$$\sigma_\eta^2 = N \frac{N_0}{T_b} E[g_{0,k}^2]. \quad (11)$$

3. Chaotic Codes

Chaos based communication systems qualify as broadband systems in which the natural spectrum of the information

signal is spread over a very large bandwidth [19]. This class of systems is called spread spectrum communication systems since they make use of a much higher bandwidth than that of the data bandwidth to transmit the information. Nowadays, pseudo-noise sequences such as Gold sequences and Walsh-Hadamard sequences are so far the most popular spreading sequences and have good correlation properties, limited security and can be reconstructed by linear regression attack for their short linear complexity [20]. A chaotic sequence generator can visit an infinite number of states in a deterministic manner and therefore produce a sequence which never repeats itself [4].

There is the flexibility in choosing the spreading gain as the sequences can be truncated to any length. The search for the best set of codes contributing reduced multiple access interference (MAI) is still one of the severe requirements of future MC-CDMA systems.

This chaotic spreading code is used before the use of an orthogonal code for multiuser transmission. These chaotic and Walsh-Hadamard codes have produced good result in utility and reducing MAI [21].

A single system described by its discrete chaotic map can generate a very large number of distinct chaotic sequences, each sequence being uniquely specified by its initial value [22]. This dependency on the initial state and the nonlinear characteristic of the discrete map make the MC-CDMA system highly secure. A chaotic map is a dynamic discrete-time continuous-value equation that describes the relation between the present and next value of chaotic system. Let x_{n+1} and X_n be successive iterations of the output x and M is the forward transformation mapping function. The general form of multidimensional chaotic map is $x_{n+1} = M(x_n, x_{n-1}, ..., xn-m)$. A simple logistic map is given in the following equation:

$$x_{n+1} = \mu(1-x_n)x_n \tag{12}$$

and $1 \leq \mu \leq 4$, where μ is the bifurcation parameter and the system exhibits a great variety of dynamics depending on the value of μ ($3.6 \leq \mu \leq 6$).

In Fig. 4 we represent logistic map for different constant number μ.

Using logistic map the chaotic spreading sequences for the BPSK system is generated. After assigning different initial condition to each user, the chaotic map is started with the initial condition of the intended receiver and iterated repeatedly to generate multiple codes.

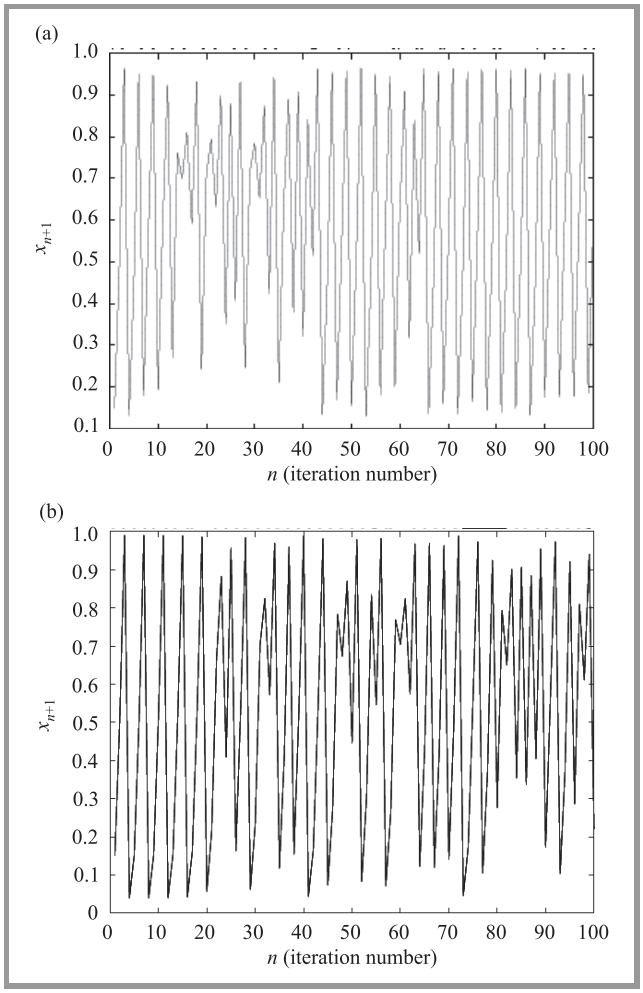


Fig. 4. The logistic map for (a) $\mu = 3.86$ and $x_0 = 0.15$, and (b) $\mu = 3.96$ and $x_0 = 0.15$.

4. Channel Identification Using Chaos

The goal of the equalization techniques is to reduce the effect of the fading and the interference while not enhancing the effect of the noise on the decision of what data symbol was transmitted. Whenever there is a diversity scheme involved, it may involve receiving multiple copies of a signal from time, frequency or antenna diversity, the field of classical diversity theory can be applied. These equalization techniques may be desirable for their simplicity as they involve simple multiplications with each copy of the signal. However, they may not be optimal in a channel with interference in the sense of minimizing the error under some criterion.

Before the equalization, the receiver estimates the channel parameters using a chaotic sequence χ of length N as a training sequence, we assume that the channel is invariant during a time T_p.

The information term is then given by

$$\mathbf{h} = \chi^{-1} r(t). \tag{13}$$

To simplify the formulation, we consider one user in this system, instead of $h_{m,k}$ we use h_k, the same with the equalizer parameters $g_{m,k}$; so $\mathbf{h} = \{h_0, h_1; h_2; \cdots; h_{N-1}\}^T$ and $r(t)$ is the received data from the chaotic sequence transmitted. We suppose that the receiver knows the transmitted chaotic sequence χ by the transmitter based on the model of logistic map Eq. (12). It is shown, from Eq. (13), that the channel parameters can be easily estimated. So the equalization term using equal gain combining (EGC) is given by

$$g_k = \frac{h_k^*}{|\, h_k \,|^2}.$$

The information is then equalized, from Eq. (10) we have:

$$\xi_{inf} = a_m \sum_{k=0}^{N-1} h_k g_k. \tag{14}$$

This technique does not attempt to equalize the effect of the channel distortion in any way. This technique may be desirable for its simplicity as the receiver does not require the estimation of the channel's transfer function. Using this scheme, the decision variable of Eq. (9) is given as

$$\vartheta_0 = \xi_{inf}^{egc} + \beta_{int}^{egc} + \eta^{egc} \tag{15}$$

with $\xi_{inf}^{egc} = a_m$ and $\beta_{int}^{egc} = \sum_{m=1}^{M-1} a_m[k]c_m[k]c_0[k]\cos(\widehat{\theta}_{m,k})$, where the noise can be approximated by a zero-mean Gaussian random with a variance of: $\sigma_\eta^2 = N\frac{N_0}{T_b}$.

5. Theoretical Analysis

5.1. Uplink Transmission

We analyze bit error rate (BER) of the proposed chaotic MC-CDMA system. We have calculated the theoretical BER for the classical equalizer EGC.
Below, some of the theoretical performance results obtained are given:

$$\sigma_{\beta int}^2 = (M-1)\overline{P}_m \text{ (variance of interferences)}$$

and

$$\sigma_\eta^2 = N\frac{N_0}{T_b} \text{ (variance of noise).}$$

We have the general form of BER [14] :

$$BER = \frac{1}{2}\text{erfc}\left(\frac{0.5(\sum_{k=0}^{N-1} h_k g_k)^2}{\sigma_{\beta int}^2 + \sigma_\eta^2}\right)^{1/2}. \tag{16}$$

In the case of EGC we have :

$$BER = \frac{1}{2}\text{erfc}\left(\frac{\frac{1}{2}(\sum_{k=0}^{N-1} h_k)^2}{(M-1)\overline{P}_m + N\frac{N_0}{T_b}}\right)^{1/2}. \tag{17}$$

The objective is to find an approximation of $h = \sum_{k=0}^{N-1} h_k$.
In the limiting case of a large number of subcarriers, $(\sum_{k=0}^{N-1} h_k)$ can be approximated by the law of large number (LLN) to be the constant $NE[h_k]$. The advantage of using the LLN is that it requires low computational complexity. Using the LLN simplifies the expression for the probability of error to [14]:

$$N \text{ is large} \Longrightarrow \gamma_0 = \sum_{k=0}^{N-1} h_k \simeq NE[h_k],$$

$$BER = \frac{1}{2}\text{erfc}\left(\frac{\frac{1}{2}N^2E^2[h_k]\,T_b}{(M-1)\overline{P}_m T_b + NN_0}\right)^{1/2},$$

$$BER = \frac{1}{2}\text{erfc}\left(\frac{\frac{\pi}{4}N\,SNR}{(M-1)\,SNR + N}\right)^{1/2}, \tag{18}$$

where N is the number of subcarriers and M is the number of users.

5.2. Downlink Transmission

Transmissions in the downlink, i.e., the transmission from the base station to the terminals through the same channel. In this section, we'll use the notation of Eq. (8), and we assume perfect phase correction for interference. The generalized decision variable given in Eq. (8), simplifies to

$$\vartheta_0 = a_0[k]\sum_{k=0}^{N-1} h_k g_k + \sum_{m=0}^{M-1}\sum_{k=0}^{N-1} a_m[k]c_m[k]c_0[k]h_k g_k + \eta\,.$$

The used codes are orthogonal and the product, $c_{k,i}c_{k,j}$ is, then, equal to 1 with the probability $1/2$ and -1 with the same probability for $i \neq j$:

$$\vartheta_0 = a_0[i]\sum_{k=0}^{N-1} h_k g_k + \sum_{m=0}^{M-1} a_m[k]\left(\sum_{k=0}^{N/2-1} h_k g_k - \sum_{k=0}^{N/2-1} h_k g_k\right) + \eta. \tag{19}$$

6. Performance Analysis

In this section, the approximations for the BER using the LLN is evaluated numerically. Using the expressions for the BER obtained for uplink transmissions in a Rayleigh fading channel, the average BER versus the number of co-channel interferers with a spreading factor $N = 128$ is shown in Fig. 5. To calculate the BER, it is assumed that the local-mean power of each interferer is equal to the local-mean power of the desired signal. The signal-to-noise ratio (SNR), which is assumed to be 10 dB, is defined as

$$SNR = \frac{\overline{p}_0 T_b}{N_0},$$

where $\overline{p}_0$ is the power of each user supposed equal for all users.

From Fig. 5, we remark that we have, approximately, the same BER results if we use both codes such as: orthogonal

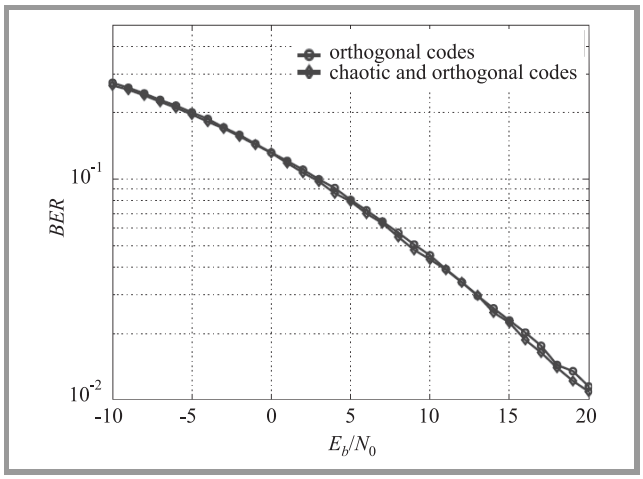


Fig. 5. Simulated BER of equalizers for chaotic MC-CDMA system.

code and chaotic code, comparing to the results obtained if we use only the orthogonal code. This result is very interesting because the use of the chaotic code allows first channel identification and this transmission is more secure than the classical MC-CDMA as combine both the Walsh-Hadamard and the chaotic codes.

7. Conclusion

Channel identification of a digital modulation technique MC-CDMA was proposed. This technique is based on the use of the chaotic sequence as a pilot frame. In fact, the chaotic code given by a nonlinear system (in this paper we have used the logistic map but other chaotic's systems could be used such Hénon map, Rössler map, Lyapunov fractal, Horseshoe map, ...) is known by the transmitter and the receiver. These codes are very sensitive to the initial conditions, so security depends on difficulty in finding these parameters.

The EGC equalization technique was used after the channel identification to correct the channel's distortion. The performance of this technique, gauged by the average bit error rate, was analytically and numerically evaluated. It is demonstrated that the performance of MC-CDMA system employing chaotic binary sequences can be superior to that for conventional codes in terms of bit error probabilities.

References

[1] S. Papadimitriou, T. Bountis, S. Mavaroudi, and A. Bezerianos, "A probabilistic symmetric encryption scheme for very fast secure communications based on chaotic systems of difference equations", *Int. J. Bifurc. Chaos*, vol. 11, pp. 3107–3115, 2001.

[2] N. K. Pareek, V. Patidar, and K. K. Sud, "Discrete chaotic cryptography using external key", *Phys. Lett. A*, vol. 309, pp. 75–82, 2003.

[3] A. T. Parker and K. M. Short, "Reconstructing the keystream from a chaotic encryption scheme", *IEEE Trans. Circ. Syst.*, vol. 48, pp. 624–630, 2001.

[4] V. Stefano and S. Gianluca, "Improving PA efficiency by chaos based spreading in MC/DS-CDMA systems", in *Proc. IEEE ISCAS Conf.*, Kos, Greece, 2006, pp. 1195–1198.

[5] A. Abel and W. Schwarz, "Chaos communication-principles schemes and system analysis", *Proc. IEEE*, vol. 90, no. 5, pp. 691–710, 2002.

[6] S. Kaiser, "Multi-carrier CDMA mobile radio system-analysis and optimization of detection, decoding, and channel estimation", Ph.D. thesis, University of Munich, Germany, 1998.

[7] S. Safi, M. Frikel, M. M'Saad, and A. Zeroual, "Blind impulse response identification of frequency radio channels: application to BRAN a channel", *Int. J. Sig. Proc.*, vol. 4, no. 1, pp. 201–206, 2007.

[8] S. Safi and A. Zeroual, "Blind parametric identification of linear stochastic non-Gaussian FIR systems higher order cumulants", *Int. J. Syst. Sci.*, vol. 35, no. 15, pp. 855–867, 2004.

[9] W. Xingang, M. Zhana, G. Xiaofeng, L. Heng, and L. Ying-Cheng, "Spread-spectrum communication using binary spatiotemporal chaotic codes", *Phys. Lett. A*, vol. 334, pp. 30–36, 2005.

[10] "Broadband Radio Access Networks (BRAN); High Performance Radio Logical Area Network (HIPERLAN) Type 2: Requirements and architectures for wireless broadband access", ETSI, TR 101 031 V2.2.1 (1999–01).

[11] "Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) layer", ETSI, TS 101 475 V1.3.1 (2001–12).

[12] F. Horlin, S. De Rore, E. Lopez-Estraviz, F. Naessens, and L. Van der Perre, "Impact of frequency offsets and IQ imbalance on MC-CDMA reception based on channel tracking", *IEEE J. Select. Areas Commun.*, vol. 24, no. 6, pp. 1179–1188, 2006.

[13] K. Fazel and L. Papke, "On the performance of convolutionnally-coded CDMA/OFDM for mobile communication system", in *Proc. IEEE Int. Symp. Pers. Ind. Mob. Radio Commun. PIMRC'93*, Berlin, Germany, 1993, pp. 468–472.

[14] N. Yee, J. M. G. Linnartz, and G. Fettweis, "Multi-carrier-CDMA in indoor wireless networks", in *Proc. IEEE Int. Symp. Pers. Ind. Mob. Radio Commun. PIMRC'93*, Yokohama, Japan, 1993, pp. 109–113.

[15] J. M. G. Linnartz, "Performance analysis of synchronous MC-CDMA in mobile Rayleigh channel with both delay and Doppler spreads", *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, 2001, pp. 1375–1387.

[16] M. Frikel, B. Targui, M. M'Saad, and F. Hamon, "Bit error rate analysis of the controlled equalization for MC-CDMA", in *IEEE ICSPC'07 Conf.*, Dubai, United Arab Emirates, 2007.

[17] S. Kaiser, "OFDM-CDMA versus DS-CDMA: performance evaluation for fading channels", in *Proc. IEEE Int. Conf. Commun.*, Seattle, USA, 1995, pp. 1722–1726.

[18] D. Mottier and D. Castelain, "SINR-based channel pre-equalization for uplink multi-carrier CDMA systems", in *Proc. IEEE Int. Symp. Pers. Ind. Mob. Radio Commun. PIMRC'02*, Lisbon, Portugal, 2002, pp. 1488–1492.

[19] S. Azou, G. Burel, and C. Pistre, "A chaotic direct-sequence spread-spectrum system for underwater communication", in *Proc. OCEANS'02 MTS/IEEE Conf.*, Biloxi, USA, 2002, vol. 4, pp. 2409–2415.

[20] C. Ling and S. Li, "Chaotic spreading sequences with multiple access performance better than random sequences", *IEEE Trans. Circ. Syst. I Fund. Theory Appl.*, vol. 47, no. 3, pp. 394–397, 2006.

[21] S. L. Georgoulis and D. G. M. Cruickshank, "Transmitter-based inverse filters for reducing MAI and ISI in CDMA-TDD downlink", *IEEE Trans. Wirel. Commun.*, vol. 3, no. 2, pp. 353–358, 2004.

[22] M. P. Kennedy, R. Rovatti, and G. Setti, "Chaotic electronics in telecommunications", in *Proc. CRC Conf.*, Boca Raton, USA, 2000.

Miloud Frikel received his Ph.D. degree from the Center of Mathematics and Scientific Computation CNRS URA 2053, France, in array processing. Currently he is with the GREYC laboratory (CNRS URA 6072) and the ENSICAEN as Assistant Professor. From 1998 to 2003, he was with the Signal Processing Lab, Institute for Systems and Robotics, Institute Superior Tecnico, Lisbon, Portugal, as a researcher in the field of wireless location and statistical array processing, after been a research engineer in a software company in Munich, Germany. He worked in the Institute for Circuit and Signal Processing of the Technical University of Munich, Germany. His research interests span several areas, including statistical signal and array processing, cellular geolocation (wireless location), space-time coding, direction finding and source localization, blind channel identification for wireless communication systems, and MC-CDMA systems.

e-mail: mfrikel@greyc.ensicaen.fr
GREYC UMR 6072 CNRS
Ecole Nationale Supérieure d'Ingénieurs
de Caen (ENSICAEN)
6, B. Maréchal Juin
14050 Caen, France

Said Safi was born in Beni Mellal, Morocco, in 1971, received the B.Sc. degree in physics (option electronics) from Cadi Ayyad University, Marrakech, Morocco, in 1995, M.Sc. and Ph.D. degrees from Chouaib Doukkali University and Cadi Ayyad University, in 1997 and 2002, respectively. He has been a Professor of information theory and telecommunication systems at the National School for Applied Sciences, Tangier, from 2003 to 2005. Since 2006, he is a Professor of applied mathematics and programming at the Faculty of Science and Technics, Beni Mellal. In 2008 he received the Ph.D. degree in telecommunication and informatics from the Cadi Ayyad University. His general interests span the areas of communications and signal processing, estimation, time-series analysis, and system identification – subjects on which he has published 10 journal papers and more than 40 conference papers. Current research topics focus on transmitter and receiver diversity techniques for single- and multiuser fading communication channels, and wide-band wireless communication systems.

e-mail: safi.said@gmail.com
Polydisciplinary Faculty
Sultan Moulay Slimane University
P.O. Box 523
Beni Mellal, Morocco

Boubekeur Targui was born in Abadla, Algeria. He received the B.Sc. degree in the electronic engineering from the University of Sidi Bel Abbess, Algeria, in 1994, and the M.Sc. and Ph.D. degrees in control engineering from the University of Claude Bernard Lyon I, Lyon, France, in 2000. Since 2002, he has held a position of Associate Professor of networking and telecommunication in the IUT of Caen, University de Caen, France. His current research interests are in equalization and synchronization in MC-CDMA systems, nonlinear observers and synchronization.

e-mail: boubekeur.targui@greyc.ensicaen.fr
GREYC UMR 6072 CNRS
Ecol Nationale Supérieure d'Ingénieurs
de Caen (ENSICAEN)
6, B. Maréchal Juin
14050 Caen, France

Mohammed M'Saad was educated at Mohammadia School of Engineering in Rabat, Morocco, where he held an Assistant Professor position in 1978. He held a research position at the Laboratoire d'Electronique et d'Etude des Systémes in Rabat, where he prepared his doctor engineering degree in process control. In November 1982, he joined the Laboratoire d'Automatique de Grenoble to work on theory and applications of adaptive control. He received his Doctorat d'Etates-Sciences Physiques from the Institut National Polytechnique de Grenoble in April 1987. In March 1988, he held a research position at the Centre National de Recherche Scientifique. In September 1996, he held a Professor position at the Ecole Nationale Supérieure d'Ingénieurs de Caen, where he is the Head of the Control Group at the GREYC. His main research areas are adaptive control theory, system identification and observation, advanced control methodologies and applications, computer aided control engineering.

e-mail: mohammed.msaad@greyc.ensicaen.fr
GREYC UMR 6072 CNRS
Ecole Nationale Supérieure d'Ingénieurs
de Caen (ENSICAEN)
6, B. Maréchal Juin
14050 Caen, France

Paper

Mobile Telematic Applications Based on Object Positioning

Kornel B. Wydro

Abstract—Global positioning systems makes possible to build a new telematic systems, enabling various applications and services in many branches of economy. A special area of such applications are various transport and transport related tasks. Among them, of a high importance are applications applied to mobile objects, for example, to all kinds of vehicles, called mobile telematic services. The paper presents features of such a services, with a special emphasis of the services being foreseen in Galileo satellite positioning system. Even conditions concerning such a service constructions are discussed, having in mind mainly necessity of complementary communications between a positioned object and related surrounding, including co-users of the movement area infrastructure as well as broader environment.

Keywords—Galileo, GPS, intelligent transport, telematics.

1. Introduction

The positioning systems, enabling object geographical positioning or localization, find – especially relatively to vehicles – more and more broad applications and use for, among others, their common and cost-free accessibility and also for progressive drop of equipment prices. But beside of popular among drivers global positioning system (GPS) navigation services, there is possibility to apply the positioning services to other needs, for example, to support some automatic control solutions, mainly related to the production processes with moving bigger physical elements in more broad spaces, e.g., mobile robots. Obviously, in such a case requirements imposed on positioning accuracy, reliability and accessibility are to be much more stronger as in nonprofessional form of GPS navigation. Yet here, for more concrete illustration of the possibilities, the problem will be discussed mainly in relation to transport applications.

Positioning systems are mostly satellite supported ones, but in last time are even developed on the basis of the mobile telecommunications networks, mainly cellular ones. But is to be pointed, that satellite systems are build intentionally as positioning systems, while cellular systems offer such a possibilities thanks for their immanent manner of terminals connecting to the transmission network through the changed base stations during the move of the terminal.

A most important and known satellite systems are the global positioning system build by US mainly for military applications, but with reduced precision available for global civil usage, and Russian system called in Rusian *"globalnaja nawigacjonnaja satielitarnaja sistiema"* (GLONASS).

In the professional (precise) service option GPS allows to locate objects with accuracy of ca. 3 m, in standard one – from 100 to 50 m, dependently on local circumstances. The GLONASS in standard service option allows to locate horizontally with precision to 60 m, and vertically – to 75 m.

Portable cellular terminals in some favourable conditions, even are to be localized with accuracy of few meters, but as it is not intentional function of their applications and networks, possibility of use and reliability of such a positioning is strongly limited, therefore cannot be used as a basis for building some service systems. It comes from fact that positioning in such a case needs in principle the overlay signal fields from more than one base station, what in normal circumstances occurs not so often. Even for increasing such a possibility, should base stations emission be stronger, what is rather unfavourable for natural aims of the mobile network.

Having in mind not a satisfying actually possibilities, conditions and circumstances of GPS services use as well as expectations for more advanced services, EU during last years undertaken building an own, more modern positioning system named Galileo [1]. The system will offer to all the interested stakeholders positioning services with a high guaranteed reliability. It will create profitable circumstances for improved activities of the persons, enterprises and administration entities related to various processes with mobility factor and position detection for all the objects – if provided with proper receiver – in their interest area.

2. Basic Galileo Services

Following categories of services are foreseen in Galileo system [2]:

- **Open services** (OS), enabling unconstrained free access to signals combination delivering object geographic position and time data.
- **Safety of life services** (SoL), being improved version of open services enabling warnings to users coming nearer to some serious danger places or situations.
- **Commercial services** (CS), providing accessibility to two additional signals allowing higher transmission capacity and better positioning accuracy with guaranteed quality and also additional transmission band for information broadcasting from centres to users (500 bit/s).

- **Public regulated services** (PRS), giving positioning and time signals for users requiring continuous services with controlled access. Here two adequately coded navigation signals will be delivered.
- **Search and rescue services** (SAR), delivering broadcasted in global scale alert signals coming from systems detecting catastrophic situations and supporting systems of search and rescue COSPAS-SARSAT.

More detailed expected technical characteristics of the basic Galileo services are presented in Table 1.

Table 1
Some technical parameters of Galileo positioning services [3]

Parameters	OS	SoL	CS	PRS
Accuracy OFR	H:15 m V:35 m	– –	– –	– –
Accuracy TFR	H:4 m V:8 m	H:4 m V:8 m	0.1−10 m	H:6.5 m V:12 m
Timing accuracy	30 ns			100 ns
Integrity	No	Yes	Yes	Yes
Alert limit	–	12−20 m	2−45 m	3−15 m
Time to alert	– –	6 s	1−10 s	1−6 s
Certification	No	Yes	Possible	Yes
Serv. guarant.	No	Yes	Possible	Possible
Explanations: OFR – one frequency receiver, TFR – two frequency receiver.				

It has to be added, that some more precise services (for the positioning with accuracy better than 4 m) are to be achieved with help of local augmentation signals[1].
The accessibility of the all the Galileo signals are estimated in the level of 99.5–99.8%, accuracy – as 95% and risk of credibility (for SoL and PRS) in the range of $3.5 \cdot 10^{-7}/150$ s. Such values of parameters are to be achieved thanks for applied broader transmission bands enabling higher accuracy and stronger signals as it is in GPS or GLONASS. In result, it will make possible to use satellite navigation even in buildings and tunnels.
Defined above the "pure" Galileo services can be improved by its combination with other ground located completing technical means, enabling design and implementation of the more advanced or specially profiled applications. In such cases particularly important is proper and extended use of the local electronic communication means, ensuring information transmission between localized object(s) and surrounding elements or entities, being in the area of interest of singular user, pair communicating one to other or multilateral communication. Obviously, in such cases systems of mobile communication will play an especially important role. Altogether, it shows that Galileo positioning system can be advantageous even in constructions of, e.g., control systems for various production and building processes[2]. It seems to be particularly applicable to big, geographically wide control systems with moving elements, like, e.g., floating car data collection systems (FCD) or automatic farming operations (precision or "intelligent" farming, i.e., targeted crop and fertilizer dosing).
In this context is to be pointed, that from the service systems using localization data point of view, a very important ground element of the Galileo system is mission control system (MCS). There will be a network of 20 MCS stations whose will execute tasks of maintenance of basic services provided by Galileo, monitoring of the systems operation, analyzing signals emitted by system's satellites and spreaded transmitted data. In MCS, beside of blocks such as orbit synchronization and processing facility (OSPF), integrity processing facilities (IPF), satellite control facility (SCF) and message generation facility (MGF) in which will be created navigation telegrams, will be installed very important for building service systems blocks of precision timing facilities (PFT) and services product facility (SPF).

3. Foreseen Applications of the Galileo System

The constructors of the Galileo system expects, that on the basis of its services will be created numerous telematic service systems and applications which, from one side, generally will contribute to economic development of particular countries, to building of the new places of labour, to extend technical development and from other allow to implement solutions and applications which will enable at least partial revenue of the investments on system's building. Those expectations are more precisely specified in form of the main areas of Galileo service applications, as listed below [3].

- **Location** – positioning services with high reliability and precision for individual persons and various objects as well mobile as motionless.
- **Rescue** – systems of guidance specially adopted for firefighters, ambulances, police and other, ensuring faster rescue actions. One expects to achieve a higher efficiency of those actions, more secure transport, lessening of accidents leading to the disabilities or fatalities.
- **Guidance** – guidance assistance for the impaired, mainly blind.
- **Entrepreneurs support** – aid to all the economies branches, like agriculture, forest and water economies as, i.e., fishing, assistance in natural resources prospecting and exploitation.

[1] Generic Galileo signals will be transmitted in radio-navigation satellite service (RNSS) band, i.e., 1.164 – 1.215 GHz (signals E5a i E5b), 1.260 – 1.300 GHz (E6) and 1.559 – 1.592 GHz (E2-L1-E1). It is possible even usage of C band (5 GHz).

[2] Obviously, rather as PRS and CS services.

- **Management support** – assistance for actions in favour of environment protection, atmosphere monitoring, wild animals monitoring. Also assistance for such activities as public transport.
- **Study and research support** – research of natural environment, seismic and volcanoes activity. Providing exact time signals for telecommunications, energy transmission, banking and other.

An important part of the Galileo system will be the accompanying special user segments. Their task is defined as positioning system's exploitation as a basis for creation various forms of original Galileo services usage. Keeping in the mind that a main area of the expected services is seen transport branch, where positioning – combined with use of the proper geographic information systems (GIS) – stands for one of the most important elements which enables improving realization of the various ventures undertaken in this field, below are given some related comments, bringing closer arising problems.
To main domains having advantage of satellite positioning systems, essential in the surface transport as well as in aerial and water ones, belongs:

– monitoring of the transport entities, especially carrying passengers or hazardous materials;
– management and control of the vehicles and especially streams of those;
– supporting individual drivers of transport means;
– expanding security level in all the transport modes, especially in catastrophic or breakdowns cases.

A monitoring tasks are in this case of crucial importance, as those should allow collecting all the necessary information on systems and surrounding states and circumstances [4]. In order of providing an efficient tools in monitoring area for realization above listed tasks, it's necessary to undertake building and development of intelligent monitoring systems which will be equipped with proper algorithms for identification of – first of all – breakdown situations and will enable, for potential operators or other service teams, unambiguously detect other difficult circumstances requiring quick intervention, as well as support decisions making for, e.g., the accidents management.
In the traffic streams management and control, occurs a need of introduction of new solutions making use of the Galileo system's services. Those should be constructed with possibility of unaided defining the optimal procedures of management of traffic parameters and, by the remote control means, reducing jams and congestions. Is to be underlined, that it may effectively increase transport efficiency by reduction of losses arising from, e.g., stoppages and in this way contribute to economic advantages.
Very important possibilities are offered by new methods of drivers support with positioning data. As a driver (or vehicle) can be equipped with digital maps, data basis describing elements of the environment and algorithms enabling delivering to drivers or operators only those information, which are directly related to driving action just in given circumstances, efficiency and first of all – safety and security can be evidently improved. Such systems, completed with decision support computer aids algorithms and equipped with trajectories and routs optimization procedures for particular vehicles, considering at the same other traffic participants as well as actual circumstances and situations, will allow exchange among infrastructure's users the properly tailored information about actual traffic conditions.
What concerns more precisely the area of security, it can be pointed that Galileo constructors expects creation of enhanced systems, which thanks for precise and reliable positioning data, will enable quick realization of rescue actions, better their integration and coordination, preventing spreading of the catastrophes and reducing their results as well as enhancing of crisis management methods. Taking that into account, even important is a possibility of upgrading of existing accident's monitoring systems, aimed on stimulation or forcing of the regulations obeying and protection against various accidents and even a terrorist actions.
Other special areas of Galileo services usage are seen as follows:

- Management of energy transmission, where precise time tokens received from Galileo will enable current flows optimization and fast restoring of network after breakdowns.
- Finances, banking and insurances. In those areas the time tokens will enable integrity, authentication and safety of electronic transactions. Even continuous monitoring of the valuable or danger cargos during their transport, standard installations of suitable systems in cars enabling continuous monitoring, will be a crucial subsystems applied by insurance companies.
- Personal navigation. One estimates that it will be a domain with most broad spectrum of applications: starting from guidance support in unknown terrain and delivering actual information about it, by surveillance on disabled persons, children or workers of public services, especially during circumstances of danger, up to additional support for broadly understood recreation.
- Search and rescue. Receivers and transmitters detecting and passing on their position thanks to Galileo, will enable fast location of missing planes, vessels, vehicles and persons.
- Management in crisis situations like floods, earthquakes, forest fires. Management from command centres will be much easier thanks information received and transmitted through Galileo system.
- Crude and gas mining, environment management, agriculture and fishing – are other possible areas, where new possibilities and benefits are expected, when will be supported by Galileo services.

Listed above areas of the Galileo generic services usage suggests also a broad posibilities of applications in automatically working systems, yet under guarantees of high quality, reliability and continuos accessibility to the signals transmitted by Galileo, and consideration of the accuracy limitations.

As an example here can be called on a problem of autonomous navigation of a team of cooperating agents. The main objective in such case is the validation of localization, map building and motion planning algorithms allowing to construct a map of the surrounding environment and figure out the trajectory that each agent should follow in order to accomplish the desired tasks. To do it proper algorithms have to process the information provided by the various sensory system of each agent, e.g., range and bearing measurements coming from positioning satellite system, sonars, rangefinders, stereo cams, etc., consisting of relative absolute measurements related to static landmarks. Moreover, suitable dynamic models, accounting for the motion of each moving object, have to be considered.

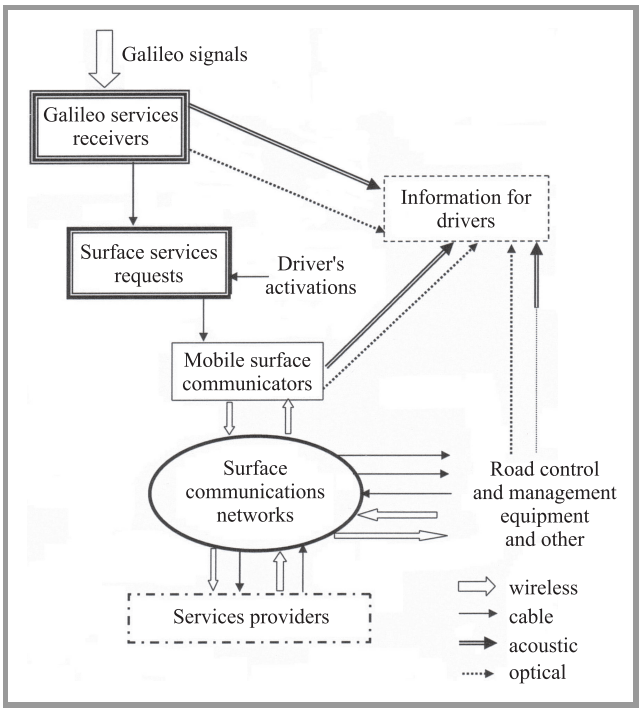


Fig. 1. Ground local communication means for positioned object.

4. Conditions for Building Systems Based on Galileo Services

As the Galileo signals have to be utilized for construction of the various services for transport and other applications for control of object movement, it is valuable to analyze conditions necessary for it, which have to be fulfilled. Those are of various kinds, mainly of technical, organizational and legal type [5]. All the kinds of conditions are of high meaning as the constructed applications can influence security of many peoples and valuable goods as it is in commonly used great transport system. A short discussion of those conditions is done below.

4.1. Technical Conditions

Making use of the geographical location data in the manner more complex than only informing driver about his position, is conditioned by accessibility to technical means enabling communication with environment being in his interest, especially situated on the co-used infrastructure and proper equipment belonging to it (e.g., signs systems). The thing is mainly concerning the auxiliary electronic communication equipment, first of all the mobile ones, making possible connections controlled by driver, as well as in automatic way. General structure of a system of communication between driver/vehicle and environment, firstly roadside equipment and mobile services providers, is shown in Fig. 1. Obviously, the proper related communication equipment is, or has to be, installed on and around the movement infrastructure.

The position and time data coming from satellite system are relayed to driver and to car-boarded automatic or/and "hand" controlled telematic call system to trigger of a proper mobile services. Those position and time parameters are to be identified – as foreseen by Galileo constructors – by range of various receivers suitable for various groups of mobile services recipients, dependently of their demands and used applications. There broad variety of receivers are planned to be build for more common use, but also a specialized receivers are foreseen for special user's segment.

The call signals will be relayed with the help of the mobile communication means to the communication network existing in surrounding. The Galileo signals receivers, as well as equipment for service's and other calls, like cellular terminals, in near future most probably will constitute standard on-board equipment.

The calls from moving object or vehicles are mostly send to the service providers and fulfilled by them with feedback information to caller in the form of a announcement or as activation signals to infrastructure's (road) information system. Obviously announcements for drivers should be preferably acoustic, but has to be even optic, as it has to reach the driver (or driving system) with highest probability. Anyway, operations on the communication means by a driver should in the minimal manner be fulfilled manually. Furthermore, all the cooperating systems have to be of high reliability ones.

In the area of dynamic development of the electronic communication systems an advice defining some specific kind of communication means are not proper (except their technical parameters, reliability and, may be prices). But of great importance is assuring of the information exchange in agreed formats, securing mutual articulated communication between all the system's elements. As a good basis for implementation of such communication manner can be

seen open communication interface for road traffic control systems (OCIT)[3] structure [6].
Other important technical factor enabling effective usage of the possibilities given by Galileo services is to have proper data basis systems and its contents. To data basis contents have to be included all the digitalized information collected and provided for users with special regard to services for moving objects. There will be digital maps, data basis with informations concerning the infrastructure and surrounding description and so on. Beside data describing some durable, constant elements, there is very important to have information on elements and occurrences of the random, not planned situations. Detecting and communicating in proper manner such an incidents makes real difficulty from the technical and formal point of view (e.g., problem of credibility).
As main elements of the information system can be recognized digital maps, geographic information systems, supporting metering and sensing equipment and information centres (all using an electronic communication network).

Digital maps. Various digital maps stands for the basic graphic description of the geographic location of the moving object and it's movement trajectory. As the map data are recorded in digital form, there is possibility to use those data not only to demonstrate area being in interest of driving entity, but also make some additional operations, like situation analysis, e.g., how is far to some defined point. There are distinguished raster maps and vector maps. The raster ones have advantage of possibility to be operated by commonly accessible software, what enables making some actualizations and corrections. The vector maps however ones enables scalability without degradation of picture quality. The raster maps use to be completed width vector layer, what may help for example to find streets. Also are used a "scanned" maps, especially when some specific elements of the area is to be demonstrated.

Geographic information systems. Beside typical cartographic data included in digital maps, an important source of basic information for driving processes are various geographic information systems, based on data bases or warehouses describing some terrain objects, which content is combined with digital maps. Is to be emphasized that usually those data are describing mainly durable objects, mostly infrastructural ones. For this reason it is important to remember actualize the in cases of reconstructions, extensions or liquidations of infrastructural elements.

Metering, sensing and controlling equipment. For the enabling collection of the information necessary for driving operations and controlling devices (e.g., road signs), the moving infrastructure have to be equipped with proper set of detecting, metering and sensing elements relaying data to the interested entities, as well as movement regulating means [4]. Taking the road traffic management as an good illustrating example can be pointed that all the traditional systems are to be used as interacting with driver, but not automatic driving means. So it is necessary to add some completing equipment to enable fulfilling contemporary needs [7].
Actually, as the moving object sensors are applied:

- magnetic loops,
- pneumatic sensors,
- piezoelectric cables,
- video cameras,
- radars,
- infrared and other passive sensors,
- active radio- and landmarks,
- FCD networks basing on GSM/GPS technologies.

It is to underline, that for traffic measurements and identifications of the vehicles and drivers are more and more applied systems with image analysis coupled with the radars measuring vehicle's speed. Also some modern thermal technologies are applied to analysis of the road surface. However, for the traffic control goals are applied so called variable message signs (VMS).
For cooperation with those elements of the infrastructure are necessary proper communication methods and means, mostly of the short range reach (ca. few hundreds meters), but of high reliability and security, as they serves also for the payments realization. As such a means com-

Table 2
Some technical parameters of communication systems used in transport telematics

Communication system	Range	Bandwidth
DSRC	0.5 – 1 km	6 – 27 Mbit/s
GSM	ca. 35 km	9.6 – 57.6 kbit/s
GSM/GPRS	ca. 35 km	53.6 – 171.2 kbit/s
GSM/EDGE	ca. 35 km	296 kbit/s
GSM/HSDPA	ca. 35 km	1.8 Mbit/s max. 7.2 Mbit/s
GSM-R	ca. 35 km	2×4 MHz
3G (UMTS)	ca. 1 km	384 – 2 000 kbit/s m
WiMAX (IEEE 802.16)	10 km	ok. 2 Mbit/s max. 75 Mbit/s
Wi-Fi (IEEE 802.11)	50 m max. few hundr. m	10 kbit/s max. 108 Mbit/s
TETRA	ca. 10 km from BS	speech 2.4 – 7.2 kbit/s data 9.6 – 28.8 kbit/s
PSTN/xDSL	few km	784 – 2 300 kbit/s
DECT	250 m	9.6 – 57.6 kbit/s
Bluetooth	up to 10 m	to 1 Mbit/s
IrDA (infrared)	few hundr. m max. few km	1 – 500 Mbit/s

[3] The standard is actually developed mostly by German-language countries: Germany, Switzerland and Austria.

monly are used systems known as (radio frequency identification – RFID, especially in dedicated short range communication – DSRC) technology or infrared links. Some approximated technical characteristics of communication systems used in transport telematics are presented in Table 2.
From other side, moving objects have to be equipped with relevant on-board communications means (in some cases even with automatically acting ones), constructed with a special consideration of a man-machine interactions rules. It is observed that temporary the on-board equipment is fitted to encompass needs arising from possibility of the realization of mainly following services:

– route selection and guidance,

– prevention against side and back-front collisions,

– signalization of the vehicle malfunctioning,

– automatic accident signalization,

– forcing the driving rules obeying,

– anti-thief protections,

– travellers comfort enhancement.

Many of such a solutions starts to be a standard or optional equipment of contemporary new, mass-produced cars.

Centers of temporary information. Obviously all the informatic structure supporting telematic systems and services, has some centres in which information is collected, analized, filtered and from which is distributed [8]. But of special attentions, as it was said earlier, are those, which operates on information regarding temporary, incidental and random situations being of particular meaning for the driving systems, especially automatic ones. By temporary situations are here understood such ones being planned for defined place, time and scope with some properly communicated prejudice (e.g., road works). As incidental or random ones have to be counted like as accidents, weather conditions changing and so on. Named two kinds of nondurable situations calls for special centres of collecting data, as information coming from high variety of sources needs careful analysis regarding credibility, integrity and value and special methods of diffusion. Obviously such a information is extremely valid for objects narrowing the places of incidents, but worthless for all others (except of those making statistical analysis)[4].

4.2. Organizational Questions

Beside technical ones, there are numerous constrains of the organizational nature concerning building methods of the considered service systems, their technical basis, operation and management, and among those, even legislative [9]. The main areas whose need extended feasibility studies and agreed implementations are:

- Financing of the building of mobile service's systems as ventures serving to broad public use at limited payments or fully free.
- Need of integration and agreement of system's building, development and management processes with the competitive infrastructure owners, operators and administrators.
- Ownership and owners rights to the systems and data collected and conitions of use of it.
- Arrangements on the cooperation between build systems and public services of rescue, security and assistance.
- Constrains coming from actual law regulations concerning the infrastuctures and traffic, as for today not sufficiently considering technical development and possibilities [10].

Financing problems are probably the crucial factors very difficult to solve. By virtue of necessity to possess highly specialized design and building potential, an administration of the infrastructure will be forced to realize discussed solutions and management of them by the outsourcing methods. This calls to take into particular consideration question of finance sources, as prvate enterpreneurs will be not intending to invest some meaninful sums on investments at which they have consciousness that the return time can be especially long and recovering of dues may be complicated.
Other difficulties may arise in the area of law, mainly related to responsibility. It is obvious, that in the transport processes may occure also serious damages and fatalities. As the advanced telematic technologies can influence on the behaviour of ifrastructure users – mainly very numerous and psychologically an phisically differentiated users – the the serious question is the possibility of recognize how far the personal responsibility is constrained by evevtual failure of telematic system as well as it concern accidents, as obligatory payments (electronic fee collection).

5. Final Remarks

Fulfilling all the expectations concerned with design, construction and exploitation of the new solutions using Galileo original system services should be a real interest of each individual or institutional user as well as of all – especially European countries – as the system is build with financing in great part from EU budget [11], [12]. In other side is to be considered, that all the phases of Galileo implementation – design, building, operation, enhancement and exploitation can bring real advantages a stimulating impacts on the economic and social processes in all the countries.

[4] In this respect interesting seems to be usage of the geotagging technique, i.e., possibility of adding on the cameras picture data describing geographical coordinates of the place in which the picture was made, as well as precise time.

So contribution of particular countries and institutions in the area of financing, science works and entrepreneurship should be an interest of their own. But it has to be executed with deep knowlege concerning all the possible constrains and threats.

References

[1] "Galileo – European satellite navigation system", Brussels, EC Directorate General Energy and Transport, 2002.

[2] "Energy and Transport – GALILEO – the programme", FRAME, Brussels, EU Directorate General Energy and Transport, 2006.

[3] "Europe's Global Satellite Navigation System", 2003, http://www.eurocontrol.fr/Newsletter

[4] I. Anderson, "Needs for network monitoring", National Roads Authority, 2003, http://www.nra.ie/Transportation/

[5] K. B. Wydro, "Analiza formalnych uwarunkowań wykorzystania systemu Galileo na potrzeby usług mobilnych", Warsaw, National Institute of Telecommunications, 2008 (in Polish).

[6] P. Haremza and D. Obcowski, "Protokół OCIT jako międzynarodowy standard umożliwiający integrację rozwiązań i podsystemów w zakresie sterowania ruchem drogowym", in *Proc. First Polish Congr. Intell. Transp. Syst.*, Warsaw, Poland, 2008 (in Polish).

[7] W. Lewiński, "Nowoczesne technologie w systemach transportowych", in *Proc. I Conf. Modern Technol. Transp.*, Szczecin, Poland, 2002 (in Polish).

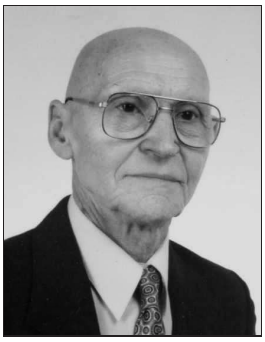
[8] K. B. Wydro, "Wymiana informacji w systemie inteligentnego transportu lądowego", *Telekomunikacja i Techniki Informacyjne*, no. 1–2, pp. 19–30, 2004 (in Polish).

[9] "Community guidelines for the development of the trans-European transport network", 2002, http://europa.eu.int/scadplus/leg

[10] I. Białowąs, "Aspekty prawne przetwarzania obrazu", in *Proc. First Polish Congr. Intell. Transp. Syst.*, Warsaw, Poland, 2008 (in Polish).

[11] "White Paper – European Transport Policy for 2010: Time to Decide", Brussels, European Commission, European Communities, 2001.

[12] "Założenia narodowej strategii rozwoju transportu na lata 2007–2013", Warsaw, Ministerstwo Infrastruktury, 2004 (in Polish).



Kornel B. Wydro was born in Poland in 1933. He earned M.Sc. and Ph.D. degrees on automatic control and telemechanics from the Electronic Faculty of Warsaw University of Technology (WUT) in 1959 and 1972, respectively. He was employed as a constructor in design offices, then as an academic teacher at WUT (1963–1998). From 1980 to 1998 he was a member of the Faculty Council and WUT Senate. Now he is a Professor and the Deputy Rector at the University College of Technology and Business in Warsaw and an Assistant Professor at the National Institute of Telecommunications. He has been the tutor of over 60 M.Sc. thesis, the author (or a co-author) of 14 books, about 200 papers or research reports, a number of presentations on scientific conferences. He translated 8 technical books from Russian and English. He is also an author of numerous reviews of papers, research works and projects. He is a current or former member of the Committees of Electronics and Telecommunications and "Poland 2000 Plus" of Polish Academy of Science, professional associations (also a member of executives), programme councils of scientific conferences, schools and organizations. He is an editor of professional journals. His scientific interests include system control, theory of information, information technology, infology, intelligent transport and information society issues.

e-mail: K.Wydro@itl.waw.pl

National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

University College of Technology and Business
Hafciarska st 11
04-704 Warsaw, Poland

Paper

Performance Comparison of Protection Strategies in WDM Mesh Networks

Hemant K. Singh, Shreya Aggarwal, Satendra Singh, Baibaswata Mohapatra, Rajendra K. Nagaria, and Sudarshan Tiwari

Abstract—Recent development in optical networking employing wavelength division multiplexing (WDM) fulfills the high bandwidth demand applications. Failure of such networks, leads to enormous data and revenue loss. Protection is one of the key techniques, which is used in designing survivable WDM networks. In this paper we compare dedicated and shared protection strategies employed in WDM mesh networks to protect optical networks failure, particularly fiber failure. Dijkstra's shortest path algorithm is considered for carrying out simulations. The paper compares the performance of protection schemes, such as, dedicated path protection (DPP), shared path protection (SPP) and shared link protection (SLP) schemes. Capacity utilization, switching time and blocking probability are the parameters considered to measure the performance of the protection schemes. Simulation results show that, SPP is more efficient in terms of capacity utilization over DPP and SLP schemes, whereas, SLP offered better switching time than both DPP and SPP schemes. The average call drop rate is minimum for shared path protection scheme and maximum for shared link protection scheme.

Keywords—blocking probability, protection, survivability, switching time, WDM networks.

1. Introduction

Wavelength division multiplexing (WDM) technology allows transmitting number of non-overlapping wavelength bands in optical networks hence, provides enormous capacities in the networks. It also provides a common infrastructure over which a range of services can be delivered. In a WDM network a lightpath is established between the source and destination for a point-to-point connection. Extremely high bandwidth (nearly 50 (Tbit/s)) offered by WDM optical network, where failure of fiber link would lead to a failure of several lightpaths traversing the link. Such failure leads to large data and revenue loss. Survivability is defined as, the ability of the network to continue to provide service in the event of failure. The lightpath traverse on different links maintains same wavelength throughout the entire path for an end-to-end communication in the absence of wavelength converter. This is known as wavelength continuity constraint. Two lightpaths sharing the same fiber link can not have same wavelength. In a network with set of demands between the node pair, determine a route and assign a wavelength for the demand is said to be routing and wavelength assignment (RWA) problem [1]–[3].

A lightpath is routed through many nodes in the network between the source and destination. There are many elements (node, link, and active components) along the path, upon failure leading to data loss. To ensure network survivability, protection schemes are being widely adopted. Protection schemes are implemented by providing some redundant capacity within the network. Upon link failure traffic is rerouted around the failed link by using this redundant capacity. On the other hand, restorations schemes involve dynamically discover the backup route and available wavelength channels to restore the traffic. Protection schemes usually implemented in a distributed fashion without using centralized control in the network, to ensure fast restoration. Most cases, protection schemes are studied under single link failure event, which means, the failed link will be repaired before another failure occurred. The maximum restoration time is 60 ms, for synchronous optical network/synchronous digital hierarchy (SONET/SDH) networks [1]. Protection schemes may be categorized as dedicated or shared based. In dedicated based protection $(1:1)$ scheme, a path disjoint wavelength channel is reserved for each lightpath which is not shared by any other backup lightpaths. On the other hand, in shared based protection $(1:n)$ scheme, a path disjoint wavelength channel is also reserved for each lightpath. However, the backup paths may be shared among different wavelength channels. As a result, backup channels are multiplexed among themselves; thereby offer better capacity efficiency over dedicated protection schemes. The details of protection strategies are explained in Section 2 by taking a simple example network.

Wavelength division multiplexing mesh networks with scheduled lightpath demand [4] is reported, where the connection setup and teardown times are known in advance also the demands between a node pair are known. The author proposed conventional integer linear programming (ILP) formulations for dedicated and shared scheduled protection without wavelength converter in the network. The study report minimum capacity utilization for fixed demand and maximize the number of demands for a fixed available capacity, while providing 100% protection for accepted connections. In [5], a RWA scheme is proposed for lightpath restoration in WDM networks, where an active multi-backup paths method is used. The author derives successive backup lightpaths up to the last node along the primary path until an available backup lightpath is found. The work has been carried out without wavelength con-

verter in the network. The algorithm is based on the iterative Dijkstra's algorithm to find a backup path for the failed primary path. Source and destination pair (s,d) as well as the cost matrix of the network were considered as inputs to the routing algorithm. Dedicated and shared resource allocation strategies [6] for survivability have been presented which reserve the resources for the primary and backup lightpaths. The author compare simulation results for different networks to evaluate the performance, which shows, shared resources strategy performs better than dedicated resources strategy in terms of blocking probability, because it utilizes the resources more efficiently.

A mixed-integer linear program (MILP) formulation for dynamic lightpath allocation for survivable WDM networks [7], is proposed by taking dedicated and shared path protection schemes. Multiple levels of services have been investigated and optimal solution is reported. Another approach called multi-commodity flow problem is widely accepted to tackle issues in WDM networks. A multi-commodity flow problem is presented [8] to address the issues of survivable network in terms of capacity allocation. The author considered different versions of node-arc and arc-path model to allocate working and spare capacity. Taking unequal arc-capacity in both the direction of the fiber links, ILP formulations is presented for capacity allocation. Our work resemble with [9], where, ILP based survivable algorithms are investigated for WDM mesh networks. This study focus on protection and restoration schemes to calculate capacity utilization and switching time for survivable schemes such as, dedicated path, shared path, and shared link protection. Restorable network design with static traffic demand is reported in [10]–[12], while, the performance of survivable algorithms with dynamic traffic is mentioned in [2], [3], [13], [14]. A review on WDM optical mesh networks is presented in [15].

Pre-configured cycle (p-cycle) is a recently proposed transport networks survivability scheme [16]–[19] combining the speed of ring networks with the capacity efficiency of mesh networks. Pre-configured cycles are ring-like pre-configured structures used to protect WDM networks against fiber failure. This protects on-cycle spans (spans that are part of the p-cycle, e.g., link B-E in Fig. 1) as well as straddling spans (spans whose end-nodes are both on the p-cycle, but that are not part of the primary paths, e.g., link B-F). Pre-configured cycle, a technique for span protection are preplanned and fully pre-connected closed loop structures of spare capacity, hence, real-time switching actions are required only at the two end nodes of the failed span to protect both on-cycle and straddling failures. Therefore switching time is quite less compared to path based and link based protection techniques. Since p-cycles are pre-planed structures consequently this scheme consumes more capacity/wavelengths compared to shared path protection [8]. Because p-cycles are formed only in the spare capacity, routing of primary paths is not affected. Also, it have been proposed to provide dual or multiple failure networks survivability [20] with re-configurable or shared p-cycles.

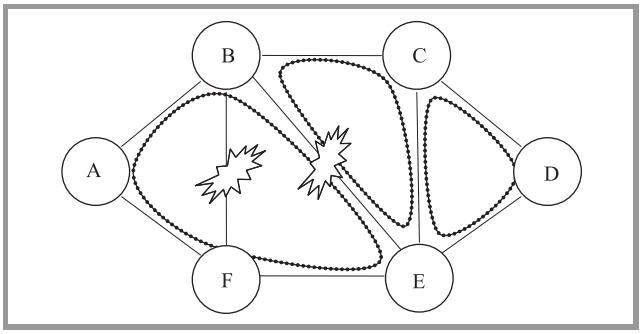


Fig. 1. Different p-cycles formed in the network.

The rest of the paper is organized as follows. Section 2 presents the simulation approach. Section 3 includes results and discussion. Conclusions of the work are presented in Section 4.

2. Simulation Approach

In this section, we will discuss the details of our proposed simulation environment developed in C language. We have taken three different protection strategies mainly, dedicated path protection, shared path protection and shared link protection schemes. Three different parameters such as, capacity utilization, switching time and blocking probability/connection drops are calculated to compare the performance of the protection strategies. We assume that, wavelength converters are not present in the networks; hence the connection between the source and the destination is established with a single wavelength link. Above protection schemes are explained by taking a simple 6-node and 8-links network arbitrarily. Taking simulation time into account, we have considered four wavelengths in each fiber to calculate capacity and seven numbers of demands/connections are considered between the node pairs randomly. Static demands are being considered and connections are exists for infinite duration of time.

2.1. Capacity Utilization

In dedicated path protection we first derive the shortest path between all node pairs using Dijkstra's shortest path algorithm. We consider these paths as the primary/working paths. In a similar fashion backup/spare paths are being derived. The wavelength assignment is done by checking the free wavelengths available on all the links in the entire path. We try to find out the minimum value of sum of working and spare capacity for all the links. For example, suppose there are two connections A→E and F→C exist. Then, the primary path would be A→B→E for connection A→E with wavelength λ_1 and the backup path is A→F→E with wavelength λ_1. For connection F→C the primary path is F→B→C with wavelength λ_1 and the backup path is F→E→C with wavelength λ_2 is reserved in this scheme.

Hence, in this scheme a total of 8 wavelength links is utilized both for primary and the backup lightpath. We derive all the possible combinations of link disjoint backup lightpaths during the simulations.

Where as, in shared path protection scheme the backup resource is not reserved. Backup resources such as, paths as well as wavelengths are shared for different connections. For example, suppose for the above two connections A→E as A→B→E with wavelength λ_1 and F→C as F→B→C with wavelength λ_2 as the primary paths may be considered. Both the backup paths A→F→E for A→E and F→E→C for B→D may use the wavelength λ_1. Even if both the backup path utilize wavelength λ_1 for the link F→E, under single fiber failure assumption, both the primary paths can not fails simultaneously. So a total of 7 wavelength links are utilized both for primary and the backup lightpaths for guaranteed service.

Link protection can be categorized as dedicated and shared. It is observed that, dedicated link protection consumes more capacity. In shared link protection we derive backup path for every links used for primary paths. From the example

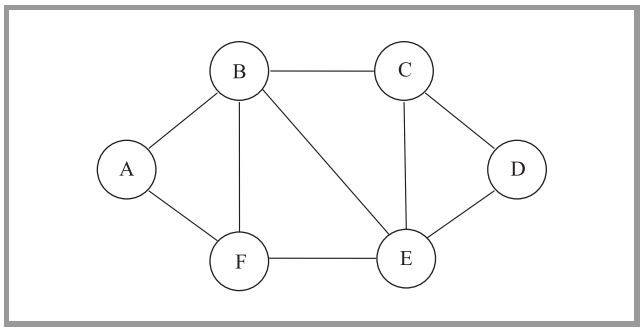


Fig. 2. Example network.

network (Fig. 2) for the connection A→E, the primary path would be A→B→E with wavelength λ_1. The backup path A→F→B→E for the failure of link A→B with wavelength λ_1 and A→B→F→E for the failure of the link B→E with wavelength λ_2 may be considered. Where as, for the connection F→C, the primary path would be F→B→C with wavelength λ_2. The backup path F→E→B→C for the failure of link F→B with wavelength λ_2 and F→B→E→C for the failure of the link B→C with wavelength λ_1 may be utilized for definite protection. Hence, a total of 13 wavelength links are utilized both for primary and the backup lightpaths. Though, the shared link protection scheme consumes more wavelength links compared to shared path and dedicated path protection, but it's switching time is less than path protection schemes.

Some of the notations we have adopted in our algorithm while performing simulations and these are given below.

- Minimize $(w_j + s_j)$, where w_j and s_j denotes working and spare capacity on link j.
  $w_j + s_j \geq W$, where W represents the total number of wavelengths available on link j.
- $\gamma_{p,w_p}^{s,d} = 1$ if the connection between the source s and the destination d used by the primary path p utilizes wavelength w_p; else zero.
- $\alpha_{p,b,w_b}^{s,d} = 1$ if the backup path b is used to protect the primary path p with wavelength w_b for the connection between the source s and the destination d; else zero.
- $\delta_{p,b,w_b}^{s,d} = 1$ if the shared backup path b is used to protect the primary path p with wavelength w_b for the connection between the source s and the destination d; else zero.
- $m_{w_b}^{l_s,l_d} = 1$ if the backup route passes through link(l_s, l_d) utilizes wavelength w_b; else zero.
- w_p, w_b are used for wavelength for primary path and backup paths, respectively.
- l_s, l_d are the link source and link destination, respectively.

Algorithm 1 describes the basic steps involves for simulating different protection schemes to calculate minimum capacity utilization.

Algorithm 1: Protection schemes

Step 1:	run Dijkstra algorithm for all node pair
Step 2:	get connection demand matrix
Step 3:	assign p, b, w_p, w_b for all connection
Step 4:	is $w_j + s_j \leq W$ for all links NO go to Step 8
Step 5:	for all W, for all links check for λ continuity NO go to Step 8
Step 6:	better than the previous record NO go to Step 8
Step 7:	over write the previous record
Step 8:	is more iteration available YES go to Step 3
Step 9:	print results

Algorithm 2 and 3 describes the wavelength assignment after routing process is completed. The fundamental assumption for the wavelength assignment is that, two lightpaths must not have same wavelengths.

Algorithm 2: λ continuity for dedicated path protection scheme

for all Links do
  for all W do
    for all node pair do
{
for all route in routeSet and Link in route
  Sum $\gamma_{p,w_p}^{s,d} = X$
}
{
for all backup route in routeSet and Link in route
  Sum $\alpha_{p,b,w_b}^{s,d} = Y$
}
$X + Y \leq 1$
end

Algorithm 3: λ continuity for shared path and shared link protection scheme

for all Links do
  for all W do
{
for all node pair, demand > 0 *do*
  for all route in routeSet and Link in route
    Sum $\gamma_{p,w_p}^{s,d} = X$
}
$X + m_{w_b}^{l_s,l_d} \leq 1$
end

2.2. Protection Switching Time

Switching time [1], [9] is the time taken from the instant a link fails to the instant the backup path of the connection is activated. In this section, we will describe the protection switching time for the different protection techniques. Assuming a link failure may be detected by the network nodes adjacent to the failed link. All network nodes participate in a distributed protocol outlined below to perform protection switching. Standard data has been taken from the literatures for calculation of switching time. We run the simulation and failed the link randomly, the average switching time calculated given below.
Let:

- t_p – the message-processing time at any node is considered as 10 μs.
- t_d – the propagation delay on each link is 400 μs.
- t_{OXC} – optical cross connect (OXC) configuration time is 10 μs.
- t_f – the time to detect a link failure is 10 μs.
- n – number of hops from the failed link source to the source node of the connection is 2.
- m – in path (link) protection, is equal to the number of hops in the backup route from the source (link-source) node to the destination (link-destination) node is 10.

2.2.1. Dedicated Path Protection

The end nodes (link-source, link-destination) of the failed link detect a failure. End nodes then send a link-fail message to the source and the destination nodes. Source node sends a connection setup message to the destination node along the backup path (which is predefined at the time of connection setup).The destination node, upon receiving the message, sends an acknowledgment (ACK) message back to the source node. This completes the protection-switching procedure and backup path is restored for the dedicated path protection. Communication continue through the backup path and the total switching time may be written as

$$t_{total} = t_f + nt_d + (n+1)t_p + 2mt_d + 2(m+1)t_p. \quad (1)$$

2.2.2. Shared Path Protection

In shared path protection backup paths are multiplexed with the available wavelengths. The end nodes of the failed link detect a failure. End nodes then send a link-fail message to the source and the destination nodes. Source node sends a connection setup message to the destination node along the backup path (which is predefined at the time of connection setup). OXCs are being configured at each intermediate node along the backup path. The destination node, upon receiving the message, sends an ACK message back to the source node. Which completes the protection-switching procedure for shared path protection and path is being restored for the shared path protection. The total switching time may be written as

$$t_{total} = t_f + nt_d + (n+1)t_p + (m+1)t_{OXC} + 2mt_d + 2(m+1)t_p. \quad (2)$$

2.2.3. Shared Link Protection

The end nodes of the failed link detect a failure. The link-source of the failed link sends a connection setup message to the link destination along the shortest backup route (which is determined in advance at the time of connection setup). OXCs are being configured at each intermediate node along the backup path around the failed link. The link destination, upon receiving the setup message, sends an ACK message back to the link source. This completes the protection-switching steps for the shared link protection and communication continue through the backup path and the total switching time may be written as

$$t_{total} = t_f + (m+1)t_{OXC} + 2(m+1)t_p + 2mt_d. \quad (3)$$

2.3. Blocking Probability

Blocking probability (BP)/connection dropped, is one parameter out of the many parameters which provides the performance of the survivable WDM networks. Request acceptance rate is the ratio of the number of connection requests accepted out of the total number of connection requests made. Where as, blocking probability is the ratio of the number of connections rejected to the total number of connection requests made. If M is the total number of connections, then blocking probability may be written as

$$BP = \frac{no.\ of\ rej.\ connection}{M} = \frac{M - no.\ of\ acc.\ connection}{M}.$$

We adopt a simple method to calculate the blocking probability in our model. Keeping the demand fixed, we calculate the number of possible connection by taking a fixed number of wavelengths on a given link. The number of call

dropped can be calculated. We increase the link capacities up to 7 number of wavelength on each fiber and present the result for the networks shown.

3. Results and Discussion

The simulation is performed with randomly taking seven numbers of connections, between the node pair. Protection schemes are implemented in random fiber failure. The simulation is carried out in Windows P-IV PC with 512 Mbit RAM. We have taken 4 wavelengths in each fiber for capacity utilization.

Table 1
Spare/protection wavelength required for the networks for different protection schemes for a demand of 7

Network	Dedicated path protection	Shared path protection	Shared link protection
NSF Net	41	36	76
India Net	42	37	57
US Net	32	29	36

From Table 1 it is observed that shared path protection consumes minimum number of wavelengths, where as, shared link protection requires maximum number of wavelengths for assured protection.
Tables 2–4 present blocking probability/call drops for different networks by taking fixed number of wavelengths and connections. Taking simulation time into account we could increase number of connections (10) and higher number of wavelengths (7) in each fiber. From the below given tables we observe that, number of connection drop is maximum for shared link protection and minimum for shared path protection techniques. As we increase the number of wavelengths the call drop should decrease in a continuous fashion, which doesn't reflect from the tables. This happened as we are taking less number of connections during our simulations.

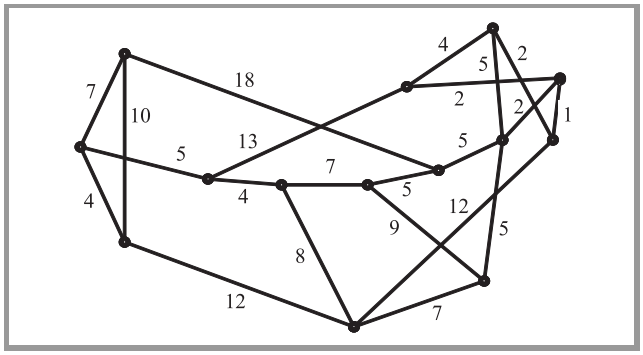


Fig. 3. NSF Net (14 nodes, 22 links).

Table 2
Number of connection blocked taking different protection strategies for demand of 10 for NSF Net (Fig. 3)

Wavelengths	Dedicated path protection	Shared path protection	Shared link protection
1	9	8	9
2	7	5	8
3	5	3	5
4	3	0	4
5	0	0	2
6	0	0	0
7	0	0	0

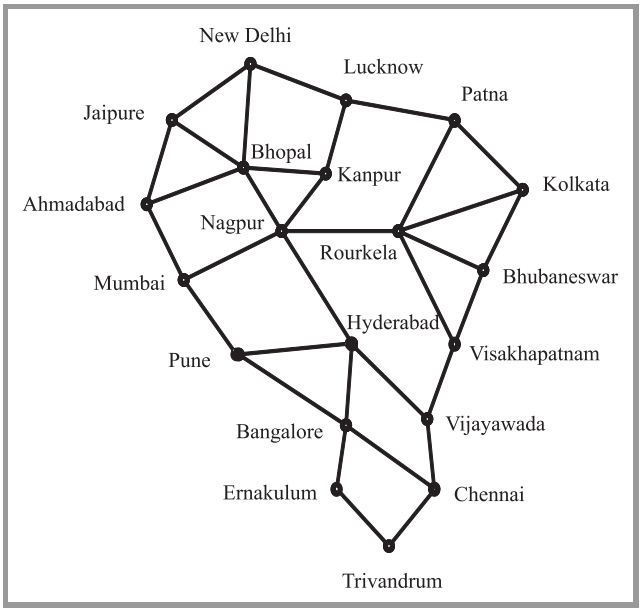


Fig. 4. India Net (20 nodes, 33 links).

Table 3
Number of connection blocked taking different protection strategies for demand of 10 for India Net (Fig. 4)

Wavelengths	Dedicated path protection	Shared path protection	Shared link protection
1	9	8	8
2	6	3	8
3	5	0	7
4	1	0	5
5	0	0	4
6	0	0	1
7	0	0	0

The average switching time is calculated and shown in Table 5. From the above results, as a network designer, we may provide different protection techniques as per the customer's requirement. For example, to generate maximum revenue we may employ shared path protection scheme by utilizing minimum number of wavelengths for a fixed demands. Where as, from quality of service point of view, where protection switching time is the parameter as

a priority, we may consider shared link protection scheme, while providing a definite protection.

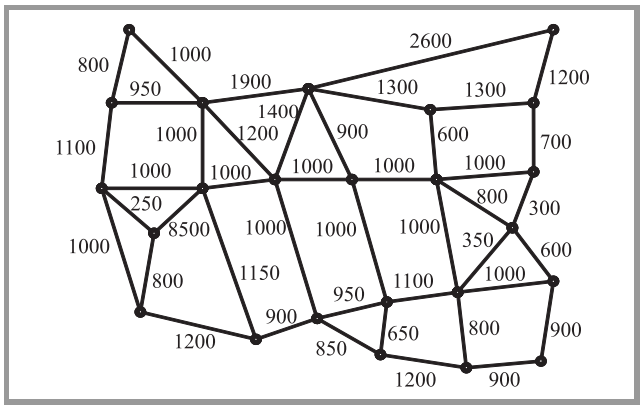


Fig. 5. US Net (24 nodes, 43 links).

Table 4
Number of connection blocked taking different protection strategies for demand of 10 for US Net (Fig. 5)

Wave-lengths	Dedicated path protection	Shared path protection	Shared link protection
1	9	9	9
2	8	8	8
3	7	0	3
4	0	0	2
5	0	0	2
6	0	0	0
7	0	0	0

Table 5
Average protection switching time for different protection strategies in milliseconds for the networks

Network	Dedicated path protection	Shared path protection	Shared link protection
NSF Net	4.96	5.02	3.36
India Net	9.06	9.17	3.36
US Net	3.73	3.78	1.7

4. Conclusions

Survivability is an essential and challenging issue in high speed networks. In this paper we examined different protection schemes such as dedicated path, shared path and shared link protection to manage WDM optical network failure. In particular, we examine capacity utilization, blocking probability and protection switching time for dedicated path, shared path and shared link protection strategies for WDM mesh networks. The chosen performance parameters such as wavelength consumption, blocking probability and switching are the measures of the quality of service. This shows the network performances under failure scenario without modifying the existing architecture.

The results show out of these three protection schemes, spare capacity utilization is minimum for shared path protection, where as it is maximum for shared link protection compared to dedicated path protection scheme. On the other hand, protection switching time is minimum for shared link protection scheme and it is maximum for shared path protection scheme. The average blocking probability or the call drop rate is minimum for shared path protection scheme and it is maximum for shared link protection scheme. Though *p*-cycle offers better protection switching time compared to shared path protection scheme but its wavelength consumption is worse. Moreover, further investigations are in progress to develop efficient algorithms to minimize network capacity utilization as well as switching time and blocking probability, for survivable WDM mesh networks.

References

[1] R. Ramaswami and K. N. Sivarajan, *Optical Networks: A Practical Perspective*. New Delhi: Morgan Kaufmann, 2002.

[2] G. Mohan and A. K. Somani, "Routing dependable connections with specified failure restoration guarantees in WDM networks", in *Proc. IEEE INFOCOM Conf.*, Tel Aviv, Israel, 2000, pp. 1761–1770.

[3] G. Mohan, C. S. R. Murthy, and A. K. Somani, "Efficient algorithms for routing dependable connections in WDM optical networks", *IEEE/ACM Trans. Netw.*, vol. 9, no. 5, pp. 553–566, 2001.

[4] C. V. Saradhi, L. K. Wei, and M. Gurusami, "Provisioning fault-tolerant scheduled lightpath demands in WDM mesh networks", in *Proc. Broadband Netw., BROADNETS Conf.*, San Jose, USA, 2004.

[5] M. A. Azim, X. Jiang, P.-H. Ho, M. M. R. Khandker, and S. Horiguchi, "A new scheme for lightpath restoration in WDM networks", in *Proc. Parall. Proces. Worksh. ICPPW'04 Conf.*, Montreal, Canada, 2004.

[6] S. Rani, A. K. Sharma, and P. Singh, "Resource allocation strategies for survivability in WDM optical networks", *Opt. Fiber Technol.*, vol. 13, no. 3, pp. 202–208, 2007.

[7] S. Zhong and A. Jaekel, "Optimal priority-based lightpath allocation for survivable WDM networks", in *Proc. Comput. Commun. Netw. ICCCN 2004*, Rosemont, USA, 2004, pp. 17–22.

[8] J. L. Kennington, E. V. Olinick, and G. Spiride, "Basic mathematical programming models for capacity allocation in mesh-based survivable networks", *J. Manage. Sci.*, vol. 35, no. 6, pp. 629–644, 2007.

[9] S. Ramamurthy, L. Sahasrabuddhe, and B. Mukherjee, "Survivable WDM mesh networks", *J. Lightw. Technol.*, vol. 21, no. 4, pp. 870–883, 2003.

[10] B. T. Doshi, S. Dravida, P. Harshavardhana, O. Hauser, and Y. Wang, "Optical network design and restoration", *Bell Labs Techn. J.*, vol. 4, no. 1, pp. 58–84, 1999.

[11] M. Sridharan, M. V. Salapaka, and A. K. Somani, "Operating mesh survivable WDM transport networks", in *Proc. SPIE Int. Symp. Terabit Opt. Netw.*, Boston, USA, 2000, pp. 113–123.

[12] M. Sridharan, A. K. Somani, and M. V. Salapaka, "Approaches for capacity and revenue optimization in survivable WDM networks", *J. High Speed Netw.*, vol. 10, no. 2, pp. 109–125, 2001.

[13] C. Sivakumar and G. Mohan, "Protocol for rapid restoration in WDM optical networks", *J. Comput. Commun.*, vol. 28, no. 1, pp. 86–96, 2005.

[14] I. Cerutti, A. Fumagalli, and S. Sheth, "Performance versus cost analysis of WDM networks with dynamic traffic grooming capabilities", in *Proc. Comput. Commun. Netw. ICCCN 2004*, Rosemont, USA, 2004, pp. 425–430.

[15] B. Mohapatra, R. K. Nagaria, and S. Tiwari, "A review of survivable WDM optical networks for unicast and multicast traffic", *Int. J. Comput. Sci. Eng. Syst.*, vol. 3, no. 3, pp. 257–264, 2009.

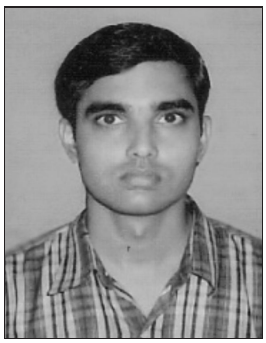
[16] W. D. Grover and D. Stamatelakis, "Cycle-oriented distributed preconfiguration: ring-like speed with meshlike capacity for self-planning network restoration", in *Proc. IEEE Int. Conf. Commun. ICC 1998*, Atlanta, USA, 1998, pp. 537–543.

[17] D. Stamatelakis and W. D. Grover, "IP layer restoration and network planning based on virtual protection cycles", *IEEE J. Select. Areas Commun.*, vol. 18, no. 10, pp. 1938–1949, 2000.

[18] D. Stamatelakis and W. D. Grover, "Theoretical underpinnings for the efficiency of restorable networks using preconfigured cycles ("*p*-cycles")", *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1262–1265, 2000.

[19] W. D. Grover and D Stamatelakis, "Bridging the ring-mesh dichotomy with *p*-cycles", in *Proc. IEEE/VDE DRCN 2000 Conf.*, Munich, Germany, 2000, pp. 92–104.

[20] H. Wang and H. T. Mouftah, "*p*-cycles in multi-failure network survivability", in *Proc. ICTON 2005 Conf.*, Barcelona, Spain, 2005, pp. 381–384.



Hemant Kumar Singh received the B.Tech. degree in electronics and communication engineering from the Motilal Nehru National Institute of Technology, Allahabad, India, in 2009.

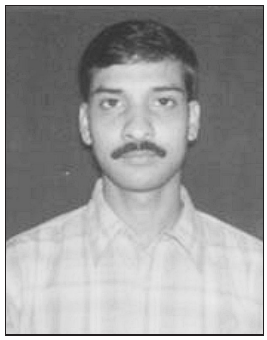
e-mail: hksingh1987@gmail.com
Department of Electronics
and Communication Engineering
Motilal Nehru National Institute of Technology
Teliergani, Allahabad-211004, India

Shreya Aggarwal received the B.Tech. degree in electronics and communication engineering from the Motilal Nehru National Institute of Technology, Allahabad, India, in 2009.

e-mail: shreya5055@gmail.com
Department of Electronics
and Communication Engineering
Motilal Nehru National Institute of Technology
Teliergani, Allahabad-211004, India

Satendra Singh received the B.Tech. degree in electronics and communication engineering from the Motilal Nehru National Institute of Technology, Allahabad, India, in 2009.

e-mail: silversurfer.mnnit@gmail.com
Department of Electronics
and Communication Engineering
Motilal Nehru National Institute of Technology
Teliergani, Allahabad-211004, India



Baibaswata Mohapatra received the B.Eng. degree in electronics and telecommunication engineering from the Utkal University, Bhubaneswar, India, in 1999, the M.Tech. degree in electronics and communication engineering from the U.P. Technical University, Lucknow, India, in 2006. He is currently working towards the Ph.D. degree with the M.N. National Institute of Technology, Allahabad, India, since 2007. He is a life member of Indian Society of Technical Education (India) and has more than 7 years of teaching and research experience. His research interest include, routing and resource optimization in WDM optical networks.
e-mail: rel0707@mnnit.ac.in
Department of Electronics
and Communication Engineering
Motilal Nehru National Institute of Technology
Teliergani, Allahabad-211004, India

Rajendra Kumar Nagaria is an Associate Professor of electronics and communication engineering at the Motilal Nehru National Institute of Technology (MNNIT), Allahabad, India. He received a B.Tech. and M.Tech. in electronics engineering from the Kamla Nehru Institute of Technology (KNIT) Sultanpur, India, and Ph.D.(Eng.) from the Jadavpur University, Kolkata, India. He has been over 22 years of teaching and research experience. He has published more than forty research papers of national and international reputes. His name is enlisted in an exclusive directory Marquis Who's Who in the world. He is

also nominated for the award as International Educator of the year 2005, by International Biographical Centre, Cambridge England. He is fellow of professional bodies like Institute of Engineers (India) and Indian Society of Technical Education. He has guided the thesis of many PG students and presently four research scholars are working under his supervision. His area of interest is mixed-mode signal processing, high speed networks/VLSI design.
e-mail: rkn@mnnit.ac.in
Department of Electronics
and Communication Engineering
Motilal Nehru National Institute of Technology
Telierganj, Allahabad-211004, India

Sudarshan Tiwari received the B.Tech. degree in electronics engineering from the I.T.BHU, Varanasi, India, in 1976, the M.Tech. degree in communication engineering from the same institution in 1978 and Ph.D. degree in electronics and computer engineering from the IIT Roorkee, India, in 1993. Presently, he is a Professor and Head of Department of Electronics and Communication Engineering, Motilal Nehru National Institute of Technology (MNNIT), Allahabad, India. He has also worked as Dean Research and Consultancy of the institute from June 2006 till June 2008. He has more than 28 years of teaching and research experience in the area of communication engineering and networking. He has supervised a number of M.Tech. and Ph.D. thesis. He has served on the program committee of several seminars, workshops and conferences. He has worked as a reviewer for several conferences and journals both nationally and internationally. He has published over 78 research papers in different journals and conferences. He has served as a visiting Professor at Liverpool John Moore's University, Liverpool, UK. He has completed several research projects sponsored by government of India. He is a life member of Institution of Engineers (India) and Indian Society of Technical Education (India), he is a member of Institution of Electrical and Electronics Engineers (USA). His current research interest include, in the area of WDM optical networks, wireless ad hoc and sensor networks and next generation networks.
e-mail: stiwari@mnnit.ac.in
Department of Electronics
and Communication Engineering
Motilal Nehru National Institute of Technology
Telierganj, Allahabad-211004, India

Paper

Optimization of the Multi-Threaded Interval Algorithm for the Pareto-Set Computation

Bartłomiej J. Kubica and Adam Woźniak

Abstract—Previous investigations of the authors surveyed the possibility of applying interval methods to seek the Pareto-front of a multicriterial nonlinear problem. An efficient algorithm has been proposed and its implementation in a multi-core environment has been done and tested. This paper has two goals. First one is to tune the developed algorithm to increase the speedup of the multi-threaded variant. The second one is to extend the algorithm to compute not only the Pareto-front (in the criteria space), but also the Pareto-set (in the decision space). Numerical results for suitable test problems are presented.

Keywords—interval computations, multicriterial analysis, multi-threaded programming, Pareto set, POSIX threads, shared-memory parallelization.

1. Introduction

It is well known that interval methods can be used as a precise and robust tool to solve nonlinear problems of various types (see, e.g., [1]), in particular multicriterial optimization problems (see, e.g., [2]–[4]). A multicriterial optimization problem is commonly encountered in practical applications (e.g., [5]–[7]). It is a problem of the following form:

$$\begin{aligned} &\min_x q_k(x) \qquad k = 1,\ldots,N \ , \\ &\text{s.t.} \\ &g_j(x) \le 0 \qquad j = 1,\ldots,m \ , \\ &x_i \in [\underline{x}_i, \overline{x}_i] \qquad i = 1,\ldots,n \ , \end{aligned} \tag{1}$$

where decision variable $x = (x_1,\ldots,x_n)^T \in \mathbb{R}^n$. In the sequel we shall denote the set of points satisfying the above conditions as X (the set of feasible points). Precisely, we seek the Pareto-set and Pareto-front of the above problem, i.e., the set of all non-dominated points $x \in X$ and the image of such set.

In this paper we recall a previously developed algorithm [4] and its parallelization using the *Pthreads* library [8]. Than, we try to optimize the parallel version to obtain high performance.

2. Basics of Interval Computations

Now, we shall define some basic notions of intervals and their arithmetic. We follow a widely acknowledged standards (cf., e.g., [1], [9], [10]).

We define the (closed) interval $[\underline{x}, \overline{x}]$ as a set $\{x \in \mathbb{R} \mid \underline{x} \le x \le \overline{x}\}$. We denote all intervals by brackets; open ones will be denoted as $]\underline{x}, \overline{x}[$ and partially open as: $[\underline{x}, \overline{x}[$, $]\underline{x}, \overline{x}]$. (We prefer this notation to using the parenthesis that are used also to denote sequences, vectors, etc.)

Following [11], we use boldface lowercase letters to denote interval variables, e.g., $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, and $\mathbb{IR}$ denotes the set of all real intervals.

We design arithmetic operations on intervals so that the following condition was fulfilled: if we have $\odot \in \{+,-,\cdot,/\}$, $a \in \mathbf{a}$, $b \in \mathbf{b}$, then $a \odot b \in \mathbf{a} \odot \mathbf{b}$. The actual formulae for arithmetic operations (see, e.g., [1], [9], [10]) are as follows:

$$\begin{aligned} [\underline{a}, \overline{a}] + [\underline{b}, \overline{b}] &= [\underline{a} + \underline{b}, \overline{a} + \overline{b}] \ , \\ [\underline{a}, \overline{a}] - [\underline{b}, \overline{b}] &= [\underline{a} - \overline{b}, \overline{a} - \underline{b}] \ , \\ [\underline{a}, \overline{a}] \cdot [\underline{b}, \overline{b}] &= [\min(\underline{ab}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{ab}), \max(\underline{ab}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{ab})] \ , \\ [\underline{a}, \overline{a}] \ / \ [\underline{b}, \overline{b}] &= [\underline{a}, \overline{a}] \cdot \left[1/\overline{b}, 1/\underline{b}\right] \ , \qquad 0 \notin [\underline{b}, \overline{b}] \ . \end{aligned}$$

Links between real and interval functions are set by the notion of an *inclusion function* (see, e.g., [1]); also called an *interval extension* (e.g., [10]).

Definition 1: A function $\mathsf{f}: \mathbb{IR} \to \mathbb{IR}$ is an *inclusion function* of $f: \mathbb{R} \to \mathbb{R}$, if for every interval $\mathbf{x}$ within the domain of f the following condition is satisfied:

$$\{f(x) \mid x \in \mathbf{x}\} \subseteq \mathsf{f}(\mathbf{x}) \ . \tag{2}$$

The definition is analogous for functions $f: \mathbb{R}^n \to \mathbb{R}^m$.

When computing interval operations, we can round the lower bound downward and the upper bound upward. This will result in an interval that will be a bit overestimated, but will be *guaranteed to contain the true result of the real-number operation*.

The quality of an interval approximation is often measured by the width of an interval, $\operatorname{wid}\mathbf{x} = \overline{x} - \underline{x}$.

3. The Algorithm to Approximate the Pareto-Set

In [4] an algorithm to seek the Pareto-front has been proposed. It subdivides the criteria space in a branch-and-bound (b&b) manner and inverts each of the obtained sets using a variant of the SIVIA procedure (i.e., set inversion via interval analysis, see [12]). Some additional tools

(like the componentwise Newton operator) are applied to speedup the computations.
The algorithm is expressed by the following pseudocode.

Algorithm 1: Pseudocode of the algorithm

```
compute_Pareto-front (q(·), x^(0), ε_y, ε_x)
// q(·) is the interval extension of the function
// q(·) = (q_1, ..., q_N)(·)
// L is the list of quadruples (y, L_in, L_bound, L_unchecked)
y^(0) = q(x^(0));
L = { (y^(0), {}, {}, {x^(0)}) };
while (there is a quadruple in L, for which wid y ≥ ε_y)
    take this quadruple (y, L_in, L_bound, L_unchecked)
        from L;
    bisect y to y^(1) and y^(2);
    for i = 1,2
        apply SIVIA with accuracy ε_x
            to quadruple (y^(i), L_in, L_bound, L_unchecked);
        if (the resulting quadruple has a nonempty
                interior, i.e., L_in ≠ ∅)
            delete quadruples that are dominated by ȳ^(i);
        end if
        insert the quadruple to the end of L;
    end for
end while
end compute_Pareto-front
```

Please note that it is sufficient to *break* the SIVIA procedure after finding an interior subbox. This leads to two variants of our algorithm, as described in [4]: "breaking SIVIA" and "non-breaking SIVIA".
Also, It should be noted that, while the "non-breaking SIVIA" variant computes both the Pareto-front and Pareto-set, the "breaking" one leaves several boxes (from the decision space) unchecked. To compute the Pareto-set we have to add a "finishing" procedure, described later in the paper.

4. A Multi-Threaded Variant

Threads are a most commonly used tool to parallelize computations in a shared-memory environment. In opposite to "heavy" processes threads run in a common address space – they can share some of the variables and data structures (and obviously have private ones, too).
In our implementation the list L from the algorithm is shared and each thread has an instance of the main `while` loop.
Obviously, operations of fetching a quadruple from L, inserting a quadruple to L and deleting dominated quadruples have to be synchronized. A single mutex (mutual exclusion lock) associated with the list is proper here.
A bit more complicated issues are related to checking if all boxes have already been investigated or not – each thread has to check not only if the list is empty, but also if other threads have finished computations or not. A *conditional variable* is used there.
We define a table `finish_thread[]` of booleans – each thread has a corresponding element, but the array is shared by all threads. Obviously there is a mutex (as always with the conditional variable) to synchronize operations on the array. Initially each element of the array is set to zero (i.e., "do not finish").
When a thread realizes that the queue of quadruples is empty, it sets its flag to true and checks if other threads did. If so, it resumes all the threads, using `pthread_cond_broadcast()` (so that they could terminate) and finishes the work. Otherwise it suspends the execution, using `pthread_cond_wait()`.
On the other hand when a thread adds a new quadruple to the queue, its `pthread_cond_signal()`signals it to one of the waiting threads.
And when a thread wakes up, it checks all flags in `finish_thread[]` once more and either terminates or resets its own flag and continues work.

5. Changes to the Algorithm

5.1. How to Increase the Efficiency of the Parallel Algorithm?

There are two major problems that decrease performance of the parallel algorithm:

- threads have to wait for each other when manipulating shared resources (in our case – the list of quadruples);
- the parallelism causes that quadruples that in the serial variant would be deleted in initial iterations are unnecessarily processed by other threads.

To minimize the influence of the first problem we have to make all operations that have to be synchronized as quick as possible. In our case the list of quadruples is implemented as a unidirectional linked list with a shortcut to the last element. Consequently, insertion of an element at the end is quick. On the other hand retrieving the box for which $\operatorname{wid}\mathbf{y} \geq \varepsilon_y$ requires a linear search of the list.
A simple improvement (similar to the one used in interval unicriterial global optimization algorithms; see, e.g., [10]) is to use two separate lists: the list L of boxes that are still processed and a new list S of small boxes that are not going to be bisected anymore.
Now the operation of obtaining the first box from L is as efficient as inserting a box to its end. The only costly operation that remains is the procedure of deleting dominated boxes, but as the list is now divided in two parts, this procedure improves, too.
Obviously, the list S must have its own mutex to synchronize operations on it ("insert" and "delete dominated").

Still the second of previously mentioned problems remains an important drawback of the method. Boxes that are currently processed by one of the threads are not removed by the "delete dominated" procedure and are going to be uselessly processed for several iterations.
To deal with this problem we add a third shared resource – a queue (implemented by a table of length roughly equal to the number of threads) of criteria vectors used lately to delete boxes that they dominate. New quadruples to be processed are compared with the values in the list – dominated ones are rejected.
Obviously, operations on this queue have to be synchronized, but – as these operations often require reading only – a readers-writer lock is more sufficient than a mutex there. As each thread reads and writes from it alternatingly (precisely: two reads than one write), no starvation is possible.

5.2. What to Do with the Lists of Unchecked Boxes?

The procedure to finish the computation for remaining quadruples is simple. The ordinary SIVIA procedure can be used on them; only with the "breaking SIVIA" flag unset.
What is more interesting is the parallelization of this part of the program. Two models were used to create threads for this computation:

- "many finishing threads" – the main thread iterates through the list L and creates a specific thread to finish the computations for each of the elements;
- "N finishing threads" – a given number of threads are created; they iterate through the list simultaneously and finish computations for different elements.

Obviously, both variants require proper extensions to the structure of elements, stored in L:

- in the first case we have to add a field to store tid of the finishing thread;
- in the second case we add to each element a flag finished and a mutex to protect it.

6. Numerical Experiments

Results for two test problems are going to be presented. The first one is constrained, but seems to be simple:

$$\begin{aligned}
&\min_{x_1,x_2}\Big(q_1(x_1,x_2) = -(5x_1+12x_2-x_1^2-x_2^2)\ ,\\
&q_2(x_1,x_2) = -(x_1+x_2)\Big)\ ,\\
&\text{s.t.}\\
&-2x_1-x_2+12\le 0\\
&-x_1+x_2-2\le 0\\
&4x_1-2x_2-47\le 0\\
&x_1,x_2\in[0,50]\ .
\end{aligned} \tag{3}$$

The second one is taken form [13]. It is a good benchmark for multicriterial optimization problems, because minimized functions are complicated and its Pareto-front and Pareto-set are both nonconnected (suitable figures are presented in [4]):

$$\begin{aligned}
\min_{x_1,x_2}\Big(q_1(x_1,x_2) = &-\big(3(1-x_1)^2\exp(-x_1^2-(x_2+1)^2)\\
&-10\big(\frac{x_1}{5}-x_1^3-x_2^5\big)\exp(-x_1^2-x_2^2)\\
&-3\exp(-(x_1+2)^2-x_2^2)\\
&+0.5(2x_1+x_2)\big),\\
q_2(x_1,x_2) = &-\big(3(1+x_2)^2\exp(-x_2^2-(1-x_1)^2)\\
&-10\big(-\frac{x_2}{5}+x_2^3+x_1^5\big)\exp(-x_2^2-x_1^2)\\
&-3\exp(-(2-x_2)^2-x_1^2)\big)\Big),\\
x_1,x_2\in[-3,3]\ .&
\end{aligned} \tag{4}$$

Due to the nondeterministic nature of parallel computations, results for four runs are presented for each of the multi-threaded variants.

Table 1
Problem (3), $N=1$, $\varepsilon_y=10^{-3}$, $\varepsilon_x=10^{-4}$

Alg. variant	Non-breaking	Breaking
old $T(1)$	182	23
new $T(1)$	181	24
f	18319573	3274377
∇f	15517892	921836
g	405435	209820
∇g	559212	414915
bis y	159127	47625
bis x	279605	155730
L	91309	27434
L_{in}	1	486
L_{bound}	433627	202645

Table 2
Problem (3), $N=2$, old, breaking SIVIA, $\varepsilon_y=10^{-3}$, $\varepsilon_x=10^{-4}$

No.	1	2	3	4
$T(2)$	44	48	56	33
f	3274377	3313486	3274502	3274397
∇f	921838	932688	921960	921842
g	209820	211848	209833	209820
∇g	414915	419485	414972	414915
bis y	47625	48588	47627	47626
bis x	155730	157169	155739	155730
L	27440	27986	27443	27443
L_{in}	487	488	488	489
L_{bound}	202689	205345	202718	202714

The program was implemented in C++, using C-XSC 2.2.1 library [14] for interval computations. The parallelization was done using POSIX threads [15].
Computations were performed on a machine with Intel S775 Core 2 Quad Q6600 2.4 GHz processor and 2 GB RAM, under control of Linux Slackware 12.0 operating system (with the 2.6.21.5-smp kernel). The GCC compiler was used in version 4.1.2.

Table 3
Problem (3), $N = 4$, old, breaking SIVIA, $\varepsilon_y = 10^{-3}$, $\varepsilon_x = 10^{-4}$

No.	1	2	3	4
$T(4)$	46	68	86	96
f	3706197	3706826	3274545	3275720
∇f	1037810	1038116	921864	922546
g	231104	231214	209820	209960
∇g	463436	463596	414915	415246
bis y	58882	58880	47629	47636
bis x	170578	170665	155730	155836
L	33728	33722	27447	27444
L_{in}	491	488	494	487
L_{bound}	231764	231732	202752	202716

Table 4
Problem (3), $N = 2$, new, breaking SIVIA, $\varepsilon_y = 10^{-3}$, $\varepsilon_x = 10^{-4}$

No.	1	2	3	4
$T(2)$	20	21	16	18
$T(1)/T(2)$	1.20	1.14	1.5	1.33
f	3644854	3624613	3246557	3222754
∇f	1022666	1021852	917838	912122
g	229376	229735	209855	209144
∇g	456835	456669	412969	410852
bis y	58306	58026	47919	47119
bis x	169214	169555	155630	155270
L	32958	32741	27173	26777
L_{in}	483	489	477	486
L_{bound}	225999	223533	199139	197038

Tables 1–11 present a few variants of the algorithm:

- a single-threaded program, using breaking or non-breaking SIVIA algorithm variants;
- "old" multi-threaded implementations of breaking or non-breaking SIVIA algorithm variants; they do not use modifications presented in this paper;
- "new" multi-threaded implementations of breaking or non-breaking SIVIA algorithm variants; they use the modifications presented in this paper.

The "new" implementations have "finishing threads" as described above; their number may be equal to the number of threads that execute the b&b method (2 or 4) or there might be an "indefinite number of finishing threads", which is explicitly marked then.

Table 5
Problem (3), $N = 4$, new, breaking SIVIA, $\varepsilon_y = 10^{-3}$, $\varepsilon_x = 10^{-4}$

No.	1	2	3	4
$T(4)$	12	14	12	15
$T(1)/T(4)$	2.00	1.71	2.00	1.60
f	3222390	3637486	3247590	3616950
∇f	912900	1022034	919978	1022494
g	209513	229731	210574	230025
∇g	411207	456761	414036	457160
bis y	47053	58121	47830	57734
bis x	155442	169501	156188	169853
L	26544	32640	26943	32428
L_{in}	490	475	488	484
L_{bound}	195903	224062	197702	222481

Table 6
Problem (3), $N = 4$, new, non-breaking SIVIA, $\varepsilon_y = 10^{-3}$, $\varepsilon_x = 10^{-4}$

No.	1	2	3	4
$T(4)$	273	200	213	202
f	18539981	18445743	18408647	18451755
∇f	15924704	15814380	15763692	15848002
g	405435	405435	405435	405435
∇g	559212	559212	559212	559212
bis y	157672	157538	157668	157481
bis x	279605	279605	279605	279605
l	89164	89018	89102	88866
L_{in}	1	1	1	1
L_{bound}	420030	420318	420451	417555

Please note that differences between the "old" and "new" variant of the algorithm rely on synchronization primitives and data structures management only, so they do not affect the number of criterion functions evaluations, gradients evaluations, etc. Only times of computation differ as it can be seen in Tables 1 and 7.
Notation for Tables 1–11 is as follows:

- $T(N)$ – computation time in seconds (for N threads),
- f – number of criterion evaluations,
- ∇f – number of criterion gradient evaluations,
- g – number of constraints evaluations,
- ∇g – number of constraints gradients evaluations,
- bis y – number of bisections in the criteria space,
- bis x – number of bisections in the decision space,
- L – number of resulting quadruples,
- L_{in} – number of resulting interior boxes,
- L_{bound} – number of resulting boundary boxes.

In captions of the tables we write:

- test problem number;
- "old" – for the algorithm described in [8] or "new" – for the modified method, presented here;
- "$N = \ldots$" for the number of threads, adding "indefinite" if the number of finishing threads was such;
- "breaking SIVIA" or "non-breaking SIVIA";
- accuracies: ε_y and ε_x.

For Tables 2, 3 and 6 a slowdown was obtained instead of a speedup, so we do not compute any speedup $T(1)/T(N)$ for them.

Table 7
Problem (4), $N = 1$, old, $\varepsilon_y = 0.2$, $\varepsilon_x = 10^{-3}$

Alg. variant	Non-breaking	Breaking
old $T(1)$	941	254
new $T(1)$	942	251
f	18591128	5786259
∇f	15089688	3358486
bis y	441	440
bis x	1689116	819561
L	174	173
L_{in}	284989	301741
L_{bound}	424703	398416

Table 8
Problem (4), $N = 4$, indefinite, new, breaking SIVIA, $\varepsilon_y = 0.2$, $\varepsilon_x = 10^{-3}$

No.	1	2	3	4
$T(4)$	129	125	126	146
$T(1)/T(4)$	1.95	2.01	1.99	1.92
f	6070586	5914523	5912838	6834256
∇f	3527204	3434490	3431324	3957086
bis y	451	462	448	480
bis x	860940	837719	837227	966762
L	181	178	178	203
L_{in}	315906	308057	307955	352962
L_{bound}	418477	406901	407869	476708

Table 9
Problem (4), $N = 2$, new, breaking SIVIA, $\varepsilon_y = 0.2$, $\varepsilon_x = 10^{-3}$

No.	1	2	3	4
$T(2)$	130	129	130	131
$T(1)/T(2)$	1.93	1.95	1.93	1.92
f	5896384	5786532	5834466	5929859
∇f	3422286	3359012	3386998	3441636
bis y	441	442	441	441
bis x	835183	819592	826520	839983
L	177	173	175	178
L_{in}	307555	301741	304292	309399
L_{bound}	406781	398416	402308	409390

Table 10
Problem (4), $N = 4$, new, non-breaking SIVIA, $\varepsilon_y = 0.2$, $\varepsilon_x = 10^{-3}$

No.	1	2	3	4
$T(4)$	299	293	275	291
$T(1)/T(4)$	3.15	3.21	3.42	3.23
f	20239650	20132048	18648308	19925301
∇f	16244024	16147668	15089772	15997082
bis y	488	494	451	484
bis x	1819498	1807109	1701210	1793874
L	200	201	182	200
L_{in}	337125	339912	298077	334196
L_{bound}	500867	507761	446478	498749

Table 11
Problem (4), $N = 4$, new, breaking SIVIA, $\varepsilon_y = 0.2$, $\varepsilon_x = 10^{-3}$

No.	1	2	3	4
$T(4)$	68	67	67	66
$T(1)/T(4)$	3.69	3.75	3.75	3.80
f	6169467	5967139	6055588	5851089
∇f	3579196	3463062	3510842	3396582
bis y	454	441	452	445
bis x	874123	845219	856911	828741
L	185	180	183	175
L_{in}	319963	311211	314298	304949
L_{bound}	427422	411917	418824	403354

7. Results Analysis

As it was stated in [4], our algorithm can compute the approximation of the whole Pareto-front in nonconnected case, compared to pointwise approximation with only 15 points, obtained by classical methods (see [13], [16]). It means that potentialities of the proposed algorithm are interesting and we hope that it can be used to solve practical problems.

As we can see in Tables 1–3, the old variant of the parallel algorithm, presented in [4] achieved no speedup for test problem (3). It is surprising, but apparently, the penalty for synchronization is too large. Fortunately, due to changes made to the program a speedup is obtained (Tables 4 and 5).

The "indefinite number of finishing threads" variant seems inefficient – on 4 processors its performance was comparable to "N finishing threads" on 2 processors for problem (4). Although the operating system managed to schedule the large number of threads properly, it clearly consumed too much resources. Clearly, as thread creation and joining is relatively expensive on today architectures, it seems optimal to have the number of threads (approximately) equal to the number of processors/cores.

For problem (3) it did not work at all – it required too much memory for a single process. Please note, also, that as this variant was easy to implement in *Pthreads*, it would be very difficult to implement, e.g., in classical *OpenMP* (older than version 3.0) that does not use tasks.

Changes made to previously created algorithm resulted in speedup of the parallel implementation. While the older version achieved no speedup for problem (3), the modified did. Anyway, the speedup was not as great as for problem (4).
It is also worth noting that the "breaking SIVIA" variant of the algorithm occurred to parallelize better than traditional, "non-breaking SIVIA" one. If SIVIA is broken after finding an interior box, much work is moved from the branch-and-bound method (which requires relatively much synchronization between concurrent threads) to the "finishing" part of the algorithm which requires no synchronization and can even be classified as "embarrassingly parallel".

8. Conclusions

Interval methods seem to be well suited to approximate the Pareto-set of a multicriterial optimization problem. Efficient parallelization, based on POSIX threads, targeted for a mutli-core environment has been proposed by authors.
Thanks to proper use of several synchronization primitives and suitable algorithm tuning, the program parallelizes well on 4 cores, allowing speedup over 3.82 for test problem (4). Testing the algorithm on higher number of cores will be subject to our future research.

Acknowledgment

The research has been supported by the Polish Ministry of Science and Higher Education under grant N N514 416934.

References

[1] L. Jaulin, M. Kieffer, O. Didrit, and E. Walter, *Applied Interval Analysis*. London: Springer, 2001.
[2] V. Barichard and J. K. Hao, "Population and Interval Constraint Propagation Algorithm", in *Second International Conference on Evolutionary Multi-Criterion Optimization (EMO 2003), Faro, Portugal, April 8–11, 2003*, LNCS, vol. 2632. Berlin-Heidelberg: Springer, 2003, pp. 81–101.
[3] G. R. Ruetsch, "An interval algorithm for multi-objective optimization", *Struct. Multidiscip. Opt.*, vol. 30, no. 1, pp. 27–37, 2005.
[4] B. J. Kubica and A. Woźniak, "Interval methods for computing the Pareto-front of a multicriterial problem", in *The Seventh International Conference on Parallel Processing and Applied Mathematics PPAM 2007, Gdańsk, Poland, September 2007*, LNCS, vol. 4967. Berlin-Heidelberg: Springer, 2008, pp. 1382–1391.
[5] M. Marks and E. Niewiadomska-Szynkiewicz, "Multiobjective approach to localization in wireless sensor networks", *J. Telecommun. Inform. Technol.*, no. 3, pp. 59–67, 2009.
[6] W. Ogryczak, "Reference point method with importance weighted partial achievements", *J. Telecommun. Inform. Technol.*, no. 4, pp. 17–25, 2008.
[7] C. Gomes da Silva and J. C. N. Clímaco, "A note on the computation of ordered supported non-dominated solutions in the bi-criteria minimum spanning tree problems", *J. Telecommun. Inform. Technol.*, no. 4, pp. 11–15, 2007.
[8] B. J. Kubica and A. Woźniak, "A multi-threaded interval algorithm for the Pareto-front computation in a multi-core environment", in *PARA 2008 Conf.*, Trondheim, Norway, 2008.
[9] E. Hansen, *Global Optimization Using Interval Analysis*. New York: Marcel Dekker, 1992.
[10] R. B. Kearfott, *Rigorous Global Search: Continuous Problems*. Dordrecht: Kluwer, 1996.
[11] R. B. Kearfott, M. T. Nakao, A. Neumaier, S. M. Rump, S. P. Shary, and P. van Hentenryck, "Standardized notation in interval analysis", http://www.mat.univie.ac.at/ neum/software/int/notation.ps.gz
[12] L. Jaulin and E. Walter, "Set inversion via interval analysis for nonlinear bounded-error estimation", *Automatica*, vol. 29, no. 4, pp. 1053–1064, 1993.
[13] I. Y. Kim and O. L. de Weck, "Adaptive weighted-sum method for bi-objective optimization: Pareto front generation", *Struct. Multidiscip. Opt.*, vol. 29, no. 2, pp. 149–158, 2005.
[14] "C-XSC library", http://www.xsc.de
[15] "POSIX threads programming", https://computing.llnl.gov/tutorials/pthreads
[16] E. Zitzler, M. Laumanns, and M. Thiele, "SPEA2: improving the strength Pareto evolutionary algorithm for multiobjective optimization", in *Evolutionary Methods for Design Optimization and Control*, K. Giannakoglou, D. Tsahalis, J. Periaux, K. Papailiou, and T. Fogarty, Eds. Barcelona: CIMNE, 2002.

Bartłomiej Jacek Kubica received his Ph.D. in computer science in 2006 from the Warsaw University of Technology (WUT), Poland. Since 2005 he is employed at WUT. Currently he works as an Assistant Professor in Complex Systems Group. He coorganizes interval sessions at PPAM conferences and organizes at PARA. He co-authored a book on parallel programming and wrote several papers and presentations. His research interests focus on interval methods, parallel computations and optimization algorithms.
e-mail: bkubica@elka.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Adam Woźniak received Ph.D. in control science in 1975 from the Warsaw University of Technology (WUT), Poland. He is employed at the Institute of Control and Computation Engineering of WUT since 1970. Now he is a reader in Systems Control Division, Complex Systems Group. His research interests include control of complex systems, robot control, decision support systems, multicriteria optimization, game theory, multiagent systems including mechanism design and auctions, interval methods applications.
e-mail: A.Wozniak@ia.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland

Information for Authors

Journal of Telecommunications and Information Technology **(JTIT)** is published quarterly. It comprises original contributions, dealing with a wide range of topics related to telecommunications and information technology. **All papers are subject to peer review**. Topics presented in the JTIT report primary and/or experimental research results, which advance the base of scientific and technological knowledge about telecommunications and information technology.

JTIT is dedicated to publishing research results which advance the level of current research or add to the understanding of problems related to modulation and signal design, wireless communications, optical communications and photonic systems, voice communications devices, image and signal processing, transmission systems, network architecture, coding and communication theory, as well as information technology.

Suitable research-related papers should hold the potential to advance the technological base of telecommunications and information technology. Tutorial and review papers are published only by invitation.

Manuscript. TEX and LATEX are preferable, standard Microsoft Word format (.doc) is acceptable. The author's JTIT LATEX style file is available:
http://www.itl.waw.pl/publ/jtit/doc/jtitaut.zip

Papers published should contain up to 10 printed pages in LATEX author's style (Word processor one printed page corresponds approximately to 6000 characters).

The manuscript should include an abstract about 150–200 words long and the relevant keywords. The abstract should contain statement of the problem, assumptions and methodology, results and conclusion or discussion on the importance of the results. Abstracts must not include mathematical expressions or bibliographic references.

Keywords should not repeat the title of the manuscript. About four keywords or phrases in alphabetical order should be used, separated by commas.

The original files accompanied with pdf file should be submitted by e-mail: redakcja@itl.waw.pl

Figures, tables and photographs. Original figures should be submitted. Drawings in Corel Draw and PostScript formats are preferred. Figure captions should be placed below the figures and can not be included as a part of the figure. Each figure should be submitted as a separated graphic file, in .cdr, .eps, .ps, .png or .tif format. Tables and figures should be numbered consecutively with Arabic numerals.

Each photograph with minimum 300 dpi resolution should be delivered in electronic formats (TIFF, JPG or PNG) as a separated file.

References. All references should be marked in the text by Arabic numerals in square brackets and listed at the end of the paper in order of their appearance in the text, including exclusively publications cited inside. Samples of correct formats for various types of references are presented below:

[1] Y. Namihira, "Relationship between nonlinear effective area and mode field diameter for dispersion shifted fibres", *Electron. Lett*., vol. 30, no. 3, pp. 262–264, 1994.

[2] C. Kittel, *Introduction to Solid State Physics*. New York: Wiley, 1986.

[3] S. Demri and E. Orłowska, "Informational representability: Abstract models versus concrete models", in *Fuzzy Sets, Logics and Knowledge-Based Reasoning*, D. Dubois and H. Prade, Eds. Dordrecht: Kluwer, 1999, pp. 301–314.

Biographies and photographs of authors. A brief professional author's biography of up to 200 words and a photo of each author should be included with the manuscript.

Galley proofs. Authors should return proofs as a list of corrections as soon as possible. In other cases, the article will be proof-read against manuscript by the editor and printed without the author's corrections. Remarks to the errata should be provided within one week after receiving the offprint.

A copy of the JTIT is provided to each author of paper published.

Journal of Telecommunications and Information Technology has entered into an electronic licencing relationship with EBSCO Publishing, the world's most prolific aggregator of full text journals, magazines and other sources. The text of *Journal of Telecommunications and Information Technology* can be found on EBSCO Publishing's databases. For more information on EBSCO Publishing, please visit www.epnet.com.

(*Contents Continued from Front Cover*)

Editorial Office

National Institute
of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

tel. +48 22 512 81 83
fax: +48 22 512 84 00
e-mail: redakcja@itl.waw.pl
http://www.itl.waw.pl/jtit

INSTYTUT ŁĄCZNOŚCI
PAŃSTWOWY INSTYTUT BADAWCZY