DOI:10.15774/PPKE.ITK.2009.001

ENHANCING THE PERFORMANCE OF SENSOR SYSTEMS IN THE NANO- AND COMMUNICATION TECHNOLOGIES

Ph.D. dissertation

Barnabás Hegyi

Scientific advisers:
János Levendovszky, D.Sc.
doctor of the Hungarian Academy of Sciences
and
Árpád Csurgay, D.Sc.
full member of the Hungarian Academy of Sciences

Pázmány Péter Catholic University
Faculty of Information Technology
Multidisciplinary Technical Sciences Doctoral School

Budapest 2008

DOI:10.15774/PPKE.ITK.2009.001

DOI:10.15774/PPKE.ITK.2009.001

Acknowledgment

First and foremost, I would like to thank my scientific advisors, *Prof. János Levendovszky* and *Prof. Árpád Csurgay* for their consistent help and support and careful guidance during my doctoral studies. I am indebted to *Prof. Tamás Roska*, head of the Doctoral School for providing for the essential conditions of my work through the years.

I am also grateful to *Prof. Wolfgang Porod* for the fruitful summer I had the opportunity to spend at the Centre for Nano Science and Technology of the University of Notre Dame (ND). I would also like to extend my thanks to *Balázs Rakos*, my direct colleague in ND for sharing his experiences in nanotechnology with me. I am thankful for *Ildikó Csurgay* for her hospitality and motherly care during my first week in ND. I would further like to say thanks to *Kristóf Karacs* for his invaluable help in starting my life at the campus. Thanks are due to *Zoltán Rácz* for his support during my stay.

I thank my closest colleagues, *Gábor Matyi* and all the members of the *Wireless Sensor Networks Research Group* for the valuable discussions and their useful advices.

I am also very grateful to my fellow PhD students, especially to *Éva Bankó*, *Csaba Benedek*, *György Cserey*, *Mária Ercsey-Ravasz*, *László Füredi*, *Gaurav Gandhi*, *Gergely Gyimesi*, *Tamás Harczos*, *Kristóf Iván*, *Márton Miháltz*, *Giovanni Pazienza*, *Gergely Soós*, *Zsolt Szálka*, *Ákos Tar*, *Róbert Tibold*, *Gábor Vásárhelyi*, *József Veres*, *Béla Weiss* and *Tamás Zeffer* for all their help and the fascinating times we spent together.

I say thanks to *Ágnes Bércesné Novák* and *György Takács* for sharing their educational experiences with me during the courses we gave together.

Special thanks go out to *Istvánné Haraszti*, *Anna Csókási*, *Lívia Adorján* and *Judit Tihanyi* for their kind help in all the administrative issues I had to deal with. I extend my gratitude to all the colleagues at the various departments of the Faculty who ensured that everything ran smoothly: the *Students' Office*, the *Dean's Office*, the *Warden's Office*, the *Financial Department*, the *IT Department*, the *Library*, the *Reception* and the *cleaning personnel*.

DOI:10.15774/PPKE.ITK.2009.001

I am especially grateful to *Ági* who has encouraged me and believed in me all the time. Last but not least I thank *my family and friends* for their love and continuous support.

My research activities were supported by the following grants.

- Hungarian Scientific Research Fund (OTKA), grant T-38345 grant;
- Office of Naval Research (ONR), a Multidisciplinary University Research Initiative (MURI) grant;
- Hungarian Scientific Research Fund (OTKA), grant NI61101;
- National Office for Research and Technology (NKTH), Mobile Innovation Centre 2.1.3 project.

DOI:10.15774/PPKE.ITK.2009.001

ENHANCING THE PERFORMANCE OF SENSOR SYSTEMS IN THE NANO- AND COMMUNICATION TECHNOLOGIES

by Barnabás Hegyi

Abstract

The latest boom of information technology called the ‘revolution of sensors’ is already well underway. It started at the beginning of the present decade and is identified with the cheap mass-production of all possible sensor and actuator devices. Combining these devices with information- and communication technologies will yield novel devices and services in the forthcoming decades. This dissertation is aimed at enhancing the performance of such sensor systems mainly in two different aspects. Arrays of nanoantenna-coupled tunnel-diode detectors integrated on the top of the CNN (cellular nonlinear network) chip may serve as high-speed THz and infrared visual systems in the future. However, in order for these devices to become a reality one day, solutions are required to improve the sensitivity of the above mentioned detectors. In the first part of this study, it is shown how this can be achieved by applying double-layer insulators in between the metallic layers of the tunneling diodes used in these sensors. As opposed to the elements of the traditional wireless and ad hoc networks, the tiny wireless sensor devices applied in the various telesensing applications are usually of limited energy. As a consequence, novel types of communication techniques need to be developed in order to maximize the lifetime of the devices in these networks. In the second part of the dissertation, stochastic energy balancing packet forwarding protocols are investigated and are shown to prolong the lifetime of wireless sensor networks. Finally, in the last part, the author demonstrates how the advantages of cooperative communication techniques can be exploited to enhance the performance of medical implant communication systems.

DOI:10.15774/PPKE.ITK.2009.001

TABLE OF CONTENTS

DOI:10.15774/PPKE.ITK.2009.001

DOI:10.15774/PPKE.ITK.2009.001

DOI:10.15774/PPKE.ITK.2009.001

Chapter One

INTRODUCTION

1.1. Preface

The first boom of information technology is represented by the personal computer industry of the 80s, which is based on the cheap microprocessor technology of the 70s. The second boom is identified with the internet- and mobile communication industry, which is founded on the cheap telecommunication technologies developed in the 80s. The third boom is usually called the 'revolution of sensors', which includes the cheap mass-production of all possible sensors and actuators. Combining these devices with information- and communication technologies will yield new products and services in the forthcoming decades, which will fundamentally change our present living- and working conditions.

My worked was aimed at enhancing the performance of these sensor systems in two different aspects. In my theses, on the one hand, I have proposed a solution to improve the sensitivity of nanoantenna-coupled tunnel-diode detectors, while on the other hand, I have developed methods for the energy efficient communication between the elements of wireless sensor networks.

1.1.1. Uncooled, high-speed, CMOS-compatible, multispectral THz and infared detectors

Technological motivations

There is a world-wide, increasing demand for high-speed, multispectral THz and infrared cameras and for the real-time processing of the images obtained with these devices. The mammal-retina-like sensor-processor that is made up by the cellular wave computer (cellular nonlinear network – CNN) chip and the sensor array integrated on the top of that, perfectly suits this task. The requirements for the individual detectors to be applied are accordingly as follows: high-speed operation (1000 frames per second); room temperature operation (integrateability with existing CMOS technologies); small dimensions (integrateability into sensor arrays); spectrum selectivity.

DOI:10.15774/PPKE.ITK.2009.001

Previous works and open questions

Concerning the existing THz and infrared sensors, photon detectors [1] require cryogenic temperatures, hence cannot be integrated with CMOS technologies, while thermal detectors such as bolometers [1] are slow and not spectrum selective. Contrary to that, nanoantenna-coupled tunnel-diode detectors [2] fulfill all the four criteria described above. The only defect of these sensors is their low sensitivity.

Contribution of the theses

The results of the first group of my theses are related to the sensitivity improvement for the latter type of detectors with special focus on the construction of the tunnel-diode in use.

1.1.2. Energy efficient communications for wireless sensor networks

Technological motivations

According to the forecasts in technology, by the end of the next decade the complexity of wireless sensor networks will surpass that of the Internet. These networks – due to their easy installability – will provide new, ubiquitous services such as telemedicine and telecare [37]. Wireless sensor networks may considerably improve the accessibility, quality and – last but not least – the cost efficiency of a number of services delivered in a traditional way today.

The elements of wireless sensor networks – as opposed to the elements of the traditional ad hoc networks – are of limited communication-, processing- and storing capabilities but most of all of limited energy. (The figures for the Crossbow MICA2 mote [21] corresponding to these quantities are 76.8 kbps, kb. 7MIPS, 500 kB és 4Ah, respectively). As a consequence, traditional network protocols cannot directly be applied in wireless sensor network as – among others – they are not optimal with respect to energy consumption. Hence the lifetime of the network is a primary consideration when developing communication procedures to be used in wireless sensor networks.

DOI:10.15774/PPKE.ITK.2009.001

Previous works and open questions – "Energy efficient packet forwarding protocols"

One way to optimize network lifetime is to balance the energy the network elements spend on communication as in this manner we increase the lifetime of the node depleting the fastest (i.e. the bottleneck node). This can be achieved – among others – by properly choosing the packet forwarding protocol. Heinzelman et al. [22] divide the network into clusters: the members of a cluster forward their packets to the cluster head, which then forwards the packets to the remote base station. In order to balance the energy consumption of the network elements, cluster heads are dynamically reelected during the operation of the network. Tan et al. [23] represent the wireless sensor network as a graph, the vertices of which are the network elements, while the weight of an edge is determined based on the energy needed to establish communication between the network elements connected by the edge as well as on the remaining energy of these nodes. A minimal spanning tree is sought in the graph at regular time intervals and network elements forward sensed data along this tree towards the base station. All the existing packet forwarding protocols are based on a deterministic traffic model, in other words, they assume that network elements intend to forward sensed data to the base station periodically. However, in a number of practical applications, the need for forwarding packets is triggered by random events. Such applications are the different target tracking applications, in which networks elements forward packets to the network if and only if a target object is residing within their sensing range. Consequently, the extension of optimal packet forwarding protocols to stochastic traffic models is indispensable.

Contribution of the theses

The results of the second group of my theses are energy-optimal stochastic packet forwarding protocols. In my work, I extended the optimization of the packet forwarding protocols to the case when the traffic model is stochastic, the network elements are modeled as queuing systems and network lifetime is defined in a statistical sense.

DOI:10.15774/PPKE.ITK.2009.001

Previous works and open questions – "Cooperative communication procedures in biomedical applications"

The lifetime of the network elements can also be extended with the help of cooperative communication [38] schemes. These procedures are brought into life by the broadcast nature of wireless media, i.e. the phenomenon that the physical signals conveying the information reach not only one but also several locations in space. Thus applying multiple cooperating receivers, a more reliable information transfer can be realized and as a result, the transmission power of the transmitter can be reduced under a given reliability constraint.

The radio link between a medical implant and an off-body transceiver has only been investigated for traditional, non-cooperative links so far. Johansson [39] determined the performance of a radio link between an implant and a single base station in an indoor environment. The effect of multipath fading as well as the effect of the human body on wave propagation were taken into account applying fading margins. Thus it remains an open question whether the performance of the in-body to off-body link can be enhanced through cooperative communications.

Most of the models in the literature on radio wave propagation from medical implants are deterministic; in addition, they examine the different effects that the in-body to off-body channel is subject to (multipath fading, the effect of the body on the radiation pattern and polarization), separately. However, in order to reveal and assess the potential of cooperative communications in the present context, a random and compact model of the in-body to off-body radio channel is required.

The performance of a cooperative communication scheme is obviously affected by the choice of the relay selection procedure. Up to now the problem of distributed relay selection – as well as the mean duration and failure probability of that – has only been analyzed in the case of applying a single relay [40]. Nevertheless, one can think of scenarios in which the performance of the cooperative communication procedure in question can further be improved by selecting and applying additional relays. The dissertation among other investigates such scenarios.

DOI:10.15774/PPKE.ITK.2009.001

Contribution of the theses

Consequently, the third group of the theses is concerned with the following: (i) establishing a plausible random model of the in-body to off-body radio channel; (ii) developing a cooperative communication scheme, the performance of which is evaluated with the help of the above mentioned model. By means of the proposed cooperative communication procedure, the lifetime of the implant can considerably be extended. In addition, the thesis group includes a distributed, multiple-relay selection procedure, which can be utilized in other cooperative communication networks as well.

1.2. Methods of investigation

When investigating the current-voltage characteristics of the tunnel-diodes in the first group of the theses, I applied the fundamental results of quantum mechanics and classical electrodynamics. I calculated the different points of the characteristics using a simple quantum transport method, which is based on the scattering of the single electron wave function by a spatially varying potential. When determining the tunneling probability numerically, I solved the space-discretized version of the time-independent single electron Schrödinger equation applying the so called Quantum Transmitting Boundary Method (QTBM) boundary conditions. Based on my proposal to improve the sensitivity of the sensor, experiments were initiated at the Nanofabrication Facility of the University of Notre Dame (USA) in order to fabricate and characterize nanoantenna-coupled MIM diode detectors with double-layer insulator diodes.

In the second group of the theses, I modeled the wireless sensor network as a queuing network. According to this, I employed the fundamental results of queuing theory when evaluating the performance of the different packet forwarding procedures. I optimized the parameters of the different protocols with the help of the combinatorial optimization algorithm called simulated annealing. A common feature of these two pieces of mathematical apparatus is that both of them are based on the theory of stochastic processes, more specifically on the theory of Markov chains. I evaluated the formulas derived to assess the performance of the different procedures, numerically.

When assessing the performance of the procedures proposed in the third group of the theses, my investigations relied on the fundamental – and in certain cases special –

DOI:10.15774/PPKE.ITK.2009.001

models of wireless communications regarding electromagnetic wave propagation, noise phenomena and the operation of wireless devices. Due to the complexity of the analytical expressions applied and the large number of the random variables involved in those, I evaluated the formulas numerically using stochastic sampling (Monte Carlo simulations).

TABLE 1.I. THE PIECES OF APPARATUS APPLIED IN THE INVESTIGATIONS

	Applied apparatus
Thesis I.	Quantum mechanics, Electrostatics
Thesis II.	Queuing theory, Stochastic optimization, Markov chains
Thesis III.	Electromagnetic wave propagation models, Monte Carlo simulations

1.3. Framework of the dissertation

In Chapter 2, the DC quality factor and resistance of metal-insulator-metal (MIM) tunnel diodes with double-layer insulators are investigated. The numerical results acquired through extensive simulations are thoroughly discussed and then compared to those obtained for the traditional metal-insulator-metal tunnel diodes with single-layer insulators.

In Chapter 3, optimal statistical energy balancing packet forwarding protocols for wireless sensor networks with a stochastic traffic model are analyzed. First of all, the communication model of the network is set up and after that the performance of the different protocols is evaluated in terms of the network lifetime they can provide.

In Chapter 4, the potential of cooperative diversity to enhance the performance of medical implant communications systems is assessed. To start with, a plausible stochastic model of the in-body to off-body radio channel is established, while afterwards the various performance gains that can be achieved by means of the proposed cooperative communication scheme are quantified.

In Chapter 5, the performance of distributed, multiple-relay selection procedures – to be applied in cooperative communication networks – are studied in terms of their mean duration and failure probability.

DOI:10.15774/PPKE.ITK.2009.001

In Chapter 6, the new scientific results of the dissertation are summarized along with examples of possible applications.

1.3.1. General notes

Please remark that the notations used within a single chapter are unique, however, the same symbol may denote diverse quantities in two different chapters.

DOI:10.15774/PPKE.ITK.2009.001

Chapter Two

INVESTIGATION OF THE DC QUALITY FACTOR OF METAL-INSULATOR-METAL (MIM) TUNNEL DIODES WITH DOUBLE-LAYER INSULATORS

2.1. Introduction

As already pointed out in the preface, in order to be able to integrate the THz and infrared sensor array with the CNN chip and to realize a high-speed, multi-spectral camera in this way, the individual detectors need to fulfill the following four requirements.

- High-speed operation: the possible applications require real-time processing of images with over 1000 frames per second.
- Room-temperature operation: as CMOS devices do not operate at cryogenic temperatures, the sensors cannot considerably be cooled down.
- Small dimensions (approximately $30\mu m \times 30\mu m$): in order to have a high-resolution camera, hundreds of thousands of the detectors need to fit onto the surface of the CNN chip.
- Multi-spectral operation: the camera is required to sense and distinguish multiple 'colors' in the THz and infrared domain.

There exist several types of detectors in the THz and infrared domains. In this section, we treat three of them: photon- , thermal- and wave detectors.

The operation of photon detectors [1] is based on the interaction of the incident 'packet' of electromagnetic radiation – a photon – and the semiconductor structure making up the detector. There are two distinct types of photon detectors: photoconductors and photovoltaic detectors. In the case of the former, if the energy of the incoming photon is higher than the band gap energy of the material, it generates an excess electron-hole pair in the semiconducting material with the electron 'jumping up' to the conduction band. As a result, the conductivity of the material increases, which can detected by applying an external voltage across the material. The latter type of photon detector, the photovoltaic detector is a semiconductor diode, which generates a voltage across its terminals if the

DOI:10.15774/PPKE.ITK.2009.001

junction is irradiated. In other words, as opposed to photoconductors, which are passive and thus require biasing, photovoltaic detectors are active devices. The electron of the electron-hole pair generated by the incident photon of sufficiently low wavelength moves to the n-type, while the hole moves to the p-type material. This makes the n-type and p-type material negatively and positively charged, respectively, and as a result a voltage occurs. Though photon detectors are fast enough to meet the above speed criterion, they cannot operate at the THz frequencies without being considerably cooled down as the kT quanta of thermal noise at room temperature is much bigger than the $h\nu$ quanta of THz radiation [3].

Thermal detectors [1] exploit the phenomenon that the electromagnetic radiation impinging on a piece of material heats it up and hence changes its temperature. The simplest thermal detector is the resistive bolometer, which incorporates a membrane, a constant temperature heat sink and two leads connecting the former components. The membrane is made of a material of which resistance shows a strong dependence on temperature. The incident radiation heats the membrane up and the change in its resistance can be detected through the leads. When the incoming radiation ceases, the membrane cools down to the temperature of the sink over the leads. Though thermal detectors can be operated at room temperature, their response time is on the order of 10ms due to the relatively high heat capacity of the component absorbing the radiation. In addition, they are not spectrum selective as they take up the power of the incident wave over a wide range of frequencies. According to recent research results, antenna-coupled microbolometers (e.g. [4]), however, might be promising candidates to address these deficiencies of thermal detectors.

Wave detectors in the THz and infrared domain take advantage of the electromagnetic nature of the radiation. They are constituted by an antenna and a nonlinear rectifying element – just like the classical, crystal detector type receivers used in AM radio broadcast applications. The difference is that – due to the much shorter wavelength – these sensors in the THz and infrared domain are equipped with nanoantennas; in addition, as a rectifying component, they include metal-insulator-metal thin-film tunneling diodes, which – thanks to their electron tunneling based operation – are devices fast enough to rectify currents even at visible frequencies. Concerning the operation principle of wave detectors, the incident electromagnetic wave induces a high-frequency

DOI:10.15774/PPKE.ITK.2009.001

voltage in the antenna, which is then coupled to the rectifier generating a DC signal at the output.

Nanoantenna-coupled MIM tunnel-diode detectors are sensors of small dimensions with high-speed and room-temperature operation; furthermore – as nanoantennas can be designed to be resonant only at certain frequencies – they are spectrum selective. To sum up, this type of detector meets all the four criteria described above. One major drawback of these sensors – that may prevent their future use in high-speed, multi-spectral THz and infrared camera applications – is, however, their low sensitivity.

The present chapter of this study addresses this issue concentrating on the performance improvement possibilities related to the metal-insulator-metal diode. As it will be shown later on, the sensitivity of nanoantenna-coupled MIM tunnel-diode detectors is directly proportional to the quality factor of the diode, which – as a result – stands in the main focus of attention of this part of the dissertation. The idea to apply a double-layer insulator structure in the MIM diode – to improve the sensitivity of the system – is the main contribution of the thesis related to this part of the study.

The remainder of this chapter is organized as follows. In Section 2.2, previous work on nanoantenna-coupled MIM diode detectors is summarized along with the model of those. Section 2.3 describes a very simple quantum transport model that is used to simulate the current-voltage characteristics of MIM diodes in our investigations. In Section 2.4, the model of the MIM diode with double-layer insulator is introduced and detailed numerical results on the quality factor and resistance of this type of diode are presented along with a thorough discussion. In Section 2.5, MIM diodes with single- and double layer insulators are compared in terms of the sensitivity that can be achieved when applied in nanoantenna-coupled MIM diode detectors. Finally, conclusions are drawn in Section 2.6.

2.2. Nanoantenna-coupled MIM tunnel-diode detectors

In this section, first, previous work on nanoantenna-coupled MIM diode detectors is summarized. Afterwards the circuit model of the detectors is described in details.

DOI:10.15774/PPKE.ITK.2009.001

2.2.1. Overview of previous works

In the first antenna-coupled MIM diode detectors point contact metal-insulator-metal diodes were used [5] (late 60s). These sensors, however, were not suitable for commercial applications due to – among others – the mechanical instability of point contact diodes. The application of photolithography to integrate thin-film MIM diodes with antennas on a substrate made it possible to construct stable and reproducible antenna-coupled MIM diode detectors in the 70s (e.g. [6]). Later on, in the 90s, electron beam lithography enabled the fabrication of low-feature-sized nanoantennas – designed for the THz and infrared domains – integrated with MIM diodes with a contact area of as small as $0.012\mu m^2$ [7] (see also [8][9]). An electron-microscope image of such a detector is shown in Figure 2.1. Finally, quite recently, double-band nanoantenna-coupled MIM diode detectors were proposed in [10]. The double-band sensor consisted of two single-band detectors each of which equipped with an antenna designed to resonate at a specific frequency (Figure 2.2). A solution to alleviate the serious crosstalk between the two antennas in [10] was proposed in [11] along with the idea of applying spherical silicon microlens to increase the effective area of the antennas.

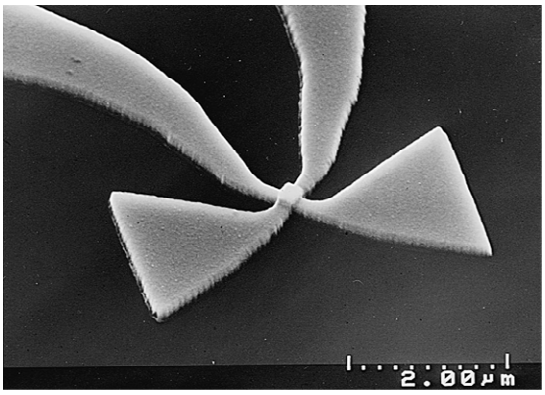

Figure 2.1. Electron-microscope image of a thin-film Ni-NiO-Ni diode with integrated bow-tie antenna [7].

The DC I-V curve of MIM diodes with single-layer insulators was treated in several theoretical papers in the literature [2][12][13][14][15]. These investigations – similarly to the present ones – are based on the simplest quantum transport model called tunneling theory. In contrary to the previously mentioned works – where analytical approximations were applied – the model is evaluated numerically in this study. The reason for that – as it will be shown in Section 2.4.2 – is that the approximations just mentioned may lead not only to quantitative but also to qualitative errors. There exist also numerous experimental studies on the DC I-V characteristics of MIM diodes with single-layer insulators

DOI:10.15774/PPKE.ITK.2009.001

[7][8][9][16]. The results of a quite an extensive experimental work can be found [16]. In that work, the resistance and quality factor of MIM diodes made of several different materials (Al-Al_2O_3-Al, Al-Al_2O_3-Ni, Al-Al_2O_3-Pt, Al-Al_2O_3-Ti and Ni-NiO-Pt) were measured and determined at different bias voltages.

Open problems

To the best of the author's knowledge, neither theoretical nor experimental results to date have been published on MIM diodes with double-layer insulators. The main goal of this chapter is to explore the properties of such devices and to evaluate their potential to improve the performance of nanoantenna-coupled MIM diode detectors.

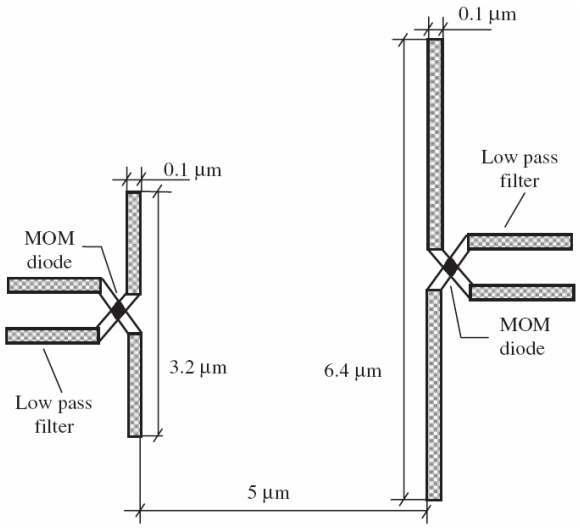

Figure 2.2. Layout of the double-band sensor designed for free-space wavelength bands of 14±2 and 24±2 µm [10].

2.2.2. Detector model

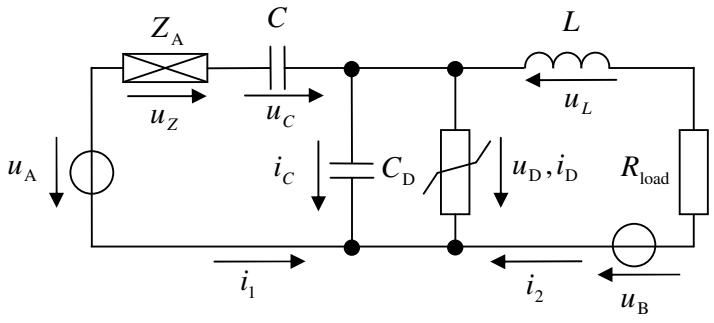

Figure 2.3. Equivalent circuit model of the detector.

DOI:10.15774/PPKE.ITK.2009.001

The equivalent circuit model of the nanoantenna-coupled MIM tunnel-diode detector [2][10][16] is shown in Figure 2.3. The antenna is modeled as an AC voltage generator (left part of the figure) with a source voltage of

$$u_{\mathrm{A}}(t) = U_{\mathrm{A}}^{1} \cos(\omega_0 t), \tag{2.1}$$

where U_{A}^{1} is the amplitude of the voltage induced by the incoming radiation, ω_0 is the angular frequency of the radiation, while t denotes time. The $Z_{\mathrm{A}} = R_{\mathrm{A}} + jX_{\mathrm{A}}$ impedance of the generator at the angular frequency of $\omega = \omega_0$ is given by $R_{\mathrm{A}}^{1} + jX_{\mathrm{A}}^{1}$, where R_{A} and X_{A} is the antenna resistance and reactance, respectively. The capacitance and the nonlinear conductance in parallel (middle of the figure) model the diode. Capacitance C_{D} represent the capacitance of the metal-insulator-metal junction, whereas the nonlinear conductance ($i_{\mathrm{D}}(u_{\mathrm{D}})$) describes the nonlinear behavior of the junction due to electron tunneling. The circuit for biasing the diode is modeled as a DC voltage generator (right part of the figure) with a source voltage of $u_{\mathrm{B}}(t) = U_{\mathrm{B}}^{0}$ and a resistance of R_{load}. Capacitance C is to separate the purely AC part of the circuit (antenna) from the DC part of that (diode and biasing circuit). This capacitance is assumed to act as an AC short-circuit and a DC open-circuit. The role and properties of inductance L is just the opposite.

The system of equations to be solved in order to determine the performance of the detector is made up of the following voltage and current rules:

$$i_C(t) + i_{\mathrm{D}}(t) + i_1(t) + i_2(t) = 0, \tag{2.2}$$

$$u_Z(t) + u_C(t) + u_{\mathrm{D}}(t) - u_{\mathrm{A}}(t) = 0, \tag{2.3}$$

$$u_L(t) + u_{\mathrm{D}}(t) - u_{\mathrm{B}}(t) - R_{\mathrm{load}} i_2(t) = 0, \tag{2.4}$$

(Note that the above system of equations – in order to be complete – needs to be supplemented with the characteristics of impedance Z_{A}, capacitance C and C_{D}, inductance L and with that of the nonlinear conductance.)

DOI:10.15774/PPKE.ITK.2009.001

The stationary (periodic) solution of the above problem can be obtained with the help of the harmonic balance analysis [17]. According to this method, all voltages and currents in the network are represented by their Fourier series expansion with a fundamental angular frequency equal to the angular frequency of the excitation. Kirchoff's rules are then formulated for each of the harmonics separately, while the resulting systems of equations are coupled through the characteristics of the nonlinear components in the network.

It is a common approximation to develop the network equations ((2.2), (2.3) and (2.4)) only for the direct (DC) and the fundamental components [17], which in the present case, take the form of

$$U_{\mathrm{D}}^{0} + R_{\mathrm{load}} I_{D}^{0} - U_{\mathrm{B}}^{0} = 0, \tag{2.5}$$

$$j\omega_0 C_{\mathrm{D}} U_{\mathrm{D}}^{1} + I_{\mathrm{D}}^{1} + I_{1}^{1} = 0, \tag{2.6}$$

$$U_{\mathrm{D}}^{1} - \left(R_{\mathrm{A}}^{1} + jX_{\mathrm{A}}^{1}\right) I_{1}^{1} - U_{\mathrm{A}}^{1} = 0, \tag{2.7}$$

where the newly introduced uppercase symbols with superscript 0 and 1 denote the direct (DC) and the fundamental components of the corresponding voltages and currents, respectively. It is easy to see (e.g. [16]) that if – furthermore – the characteristics of the nonlinear conductance ($i_{\mathrm{D}}(u_{\mathrm{D}})$) is approximated by its second-order Taylor series expansion around the DC component of the diode voltage (U_{D}^{0}), then the following additional relationships hold:

$$I_{\mathrm{D}}^{0} = i_{\mathrm{D}}\left(U_{\mathrm{D}}^{0}\right) + \frac{1}{4} i_{\mathrm{D}}^{(2)}\left(U_{\mathrm{D}}^{0}\right)\left|U_{\mathrm{D}}^{1}\right|^{2}, \tag{2.8}$$

$$I_{\mathrm{D}}^{1} = i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0}\right) U_{\mathrm{D}}^{1}. \tag{2.9}$$

Then based on (2.5), (2.6), (2.7), (2.8) and (2.9), for U_{D}^{0}, we have

DOI:10.15774/PPKE.ITK.2009.001

$$i_{\mathrm{D}}\left(U_{\mathrm{D}}^{0}\right)+\frac{1}{4}i_{\mathrm{D}}^{(2)}\left(U_{\mathrm{D}}^{0}\right)\cdot \tag{2.10}$$

$$\cdot\frac{1}{\left(1+i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0}\right)R_{\mathrm{A}}^{1}-\omega_0 C_{\mathrm{D}}X_{\mathrm{A}}^{1}\right)^2+\left(i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0}\right)X_{\mathrm{A}}^{1}+\omega_0 C_{\mathrm{D}}R_{\mathrm{A}}^{1}\right)^2}\left(U_{\mathrm{A}}^{1}\right)^2+R_{\mathrm{load}}^{-1}\left(U_{\mathrm{D}}^{0}-U_{\mathrm{B}}^{0}\right)=0.$$

Let $U_{\mathrm{D}}^{0,\mathrm{B}}$ denote the bias point DC diode voltage, i.e. the solution to (2.10) when $U_{\mathrm{A}}^{1}=0$:

$$i_{\mathrm{D}}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)+R_{\mathrm{load}}^{-1}\left(U_{\mathrm{D}}^{0,\mathrm{B}}-U_{\mathrm{B}}^{0}\right)=0\,. \tag{2.11}$$

We assume that the second term in (2.10) is so small that the solution to (2.10) (i.e. U_{D}^{0}) is found in such a small neighborhood of $U_{\mathrm{D}}^{0,\mathrm{B}}$ that

$$i_{\mathrm{D}}\left(U_{\mathrm{D}}^{0}\right)\approx i_{\mathrm{D}}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)+i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)\left(U_{\mathrm{D}}^{0}-U_{\mathrm{D}}^{0,\mathrm{B}}\right), \tag{2.12}$$

$$i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0}\right)\approx i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right), \tag{2.13}$$

$$i_{\mathrm{D}}^{(2)}\left(U_{\mathrm{D}}^{0}\right)\approx i_{\mathrm{D}}^{(2)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right). \tag{2.14}$$

i.e. in this neighborhood, the first and second derivative of $i_{\mathrm{D}}\left(u_{\mathrm{D}}\right)$ can be considered constant, while $i_{\mathrm{D}}\left(u_{\mathrm{D}}\right)$ itself can be approximated by its first-order Taylor series expansion around $U_{\mathrm{D}}^{0,\mathrm{B}}$ (bias point linearization). Then for the change in voltage across the terminals of the diode due to the incoming radiation, we have

$$\Delta U_{\mathrm{D}}^{0}=U_{\mathrm{D}}^{0}-U_{\mathrm{D}}^{0,\mathrm{B}}= \tag{2.15}$$

$$=-\frac{1}{4}\frac{i_{\mathrm{D}}^{(2)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)}{i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)}\frac{1}{\left(1+i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)R_{\mathrm{A}}^{1}-\omega_0 C_{\mathrm{D}}X_{\mathrm{A}}^{1}\right)^2+\left(i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)X_{\mathrm{A}}^{1}+\omega_0 C_{\mathrm{D}}R_{\mathrm{A}}^{1}\right)^2}\left(U_{\mathrm{A}}^{1}\right)^2\,,$$

$$\cdot\left(1+\left[i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)\right]^{-1}R_{\mathrm{load}}^{-1}\right)^{-1}.$$

where $\frac{i_{\mathrm{D}}^{(2)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)}{i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)}$ and $\left[i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)\right]^{-1}$ are the bias-point quality factor and resistance of the diode, respectively.

DOI:10.15774/PPKE.ITK.2009.001

The Johnson noise voltage of $\left[i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)\right]^{-1}$ and R_{load} in parallel [1] (bias point linearization) can be expressed as

$$U_{\mathrm{D}}^{\mathrm{n}} = \sqrt{4kTB\left(\left[i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)\right]^{-1} \times R_{\mathrm{load}}\right)}, \tag{2.16}$$

where k is Boltzmann's constant, T is the absolute temperature and B is the bandwidth. Then for the signal-to-noise ratio of the detected voltage, we have

$$\frac{\left|\Delta U_{\mathrm{D}}^{0}\right|}{U_{\mathrm{D}}^{\mathrm{n}}} = \frac{\left(U_{\mathrm{A}}^{1}\right)^{2}}{8\sqrt{kTB}}\left|\frac{i_{\mathrm{D}}^{(2)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)}{i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)}\right| \cdot \frac{\sqrt{i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)}}{\left(1+i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)R_{\mathrm{A}}^{1}-\omega_0 C_{\mathrm{D}} X_{\mathrm{A}}^{1}\right)^{2}+\left(i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)X_{\mathrm{A}}^{1}+\omega_0 C_{\mathrm{D}} R_{\mathrm{A}}^{1}\right)^{2}} \cdot \sqrt{1+\left[i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)\right]^{-1} R_{\mathrm{load}}^{-1}}. \tag{2.17}$$

It is easy to see ([2]) that – for a given value of $\frac{i_{\mathrm{D}}^{(2)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)}{i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)}$ – the signal-to-noise ratio is maximized if

$$R_{\mathrm{A}}^{1} = \frac{1}{3}\left[i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)\right]^{-1} \tag{2.18}$$

$$X_{\mathrm{A}}^{1} << \left[i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)\right]^{-1}, \tag{2.19}$$

$$\omega_0 C_{\mathrm{D}} R_{\mathrm{A}}^{1} << 1, \tag{2.20}$$

$$\omega_0 C_{\mathrm{D}} X_{\mathrm{A}}^{1} << 1, \tag{2.21}$$

and

$$\left[i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)\right]^{-1} << R_{\mathrm{load}}. \tag{2.22}$$

In that case

DOI:10.15774/PPKE.ITK.2009.001

$$\frac{\left|\Delta U_{\mathrm{D}}^{0}\right|}{U_{\mathrm{D}}^{\mathrm{n}}}=\frac{3\sqrt{3}}{128}\frac{1}{\sqrt{kTB}}\left|\frac{i_{\mathrm{D}}^{(2)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)}{i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)}\right|\left(U_{\mathrm{A}}^{1}\right)^{2}, \tag{2.23}$$

that is to say, the signal-to-noise ratio of the detected voltage is directly proportional to the bias-point quality factor of the diode, which, for this reason, stands in the main focus of attention in the remaining parts of this chapter. More specifically, we are going to investigate the quality factor and resistance of the diode at $U_{\mathrm{D}}^{0,\mathrm{B}}=0$ biasing voltage. The main reason for that is that DC sources are generally fairly noisy and thus it is preferable to avoid biasing in practice [16]. Let R_{D} and γ denote the zero-bias resistance and quality factor, respectively:

$$R_{\mathrm{D}}=\left[i_{\mathrm{D}}^{(1)}(0)\right]^{-1}, \tag{2.24}$$

$$\gamma=\frac{i_{\mathrm{D}}^{(2)}(0)}{i_{\mathrm{D}}^{(1)}(0)}. \tag{2.25}$$

Please note that also the current of the diode can be the signal to be detected just like it is done in the experiments described in [16]. It is easy to see that – in order to maximize the signal-to-noise ratio of the detected current – conditions (2.18) and (2.22) have to be replaced with

$$R_{\mathrm{A}}^{1}=\left[i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)\right]^{-1}, \tag{2.26}$$

$$\left[i_{\mathrm{D}}^{(1)}\left(U_{\mathrm{D}}^{0,\mathrm{B}}\right)\right]^{-1}>>R_{\mathrm{load}}, \tag{2.27}$$

while the others do not change. The maximum signal-to-noise ratio of detected signal is directly proportional to the bias-point quality factor of the diode in this case too.

To conclude this section, let us finally note that the reader may consider the analysis of the detector model far too detailed in view of the overall aims of this chapter. The purpose of the lengthy treatment is didactic. In several related works in the literature – such as [2], [9], [16] – it is stated that the ratio of the rectified voltage and the rectified current of the detector is simply the bias-point resistance of the diode. However, this

DOI:10.15774/PPKE.ITK.2009.001

statement is true only for the ratio of the open-circuit ($R_{\text{load}} = \infty$) rectified voltage and the short-circuit ($R_{\text{load}} = 0$) rectified current, while in general, the ratio of the formerly mention two quantities is equal to R_{load} . A proper treatment of the problem is given in [1] – in connection with photoconductor biasing.

Some practical considerations on matching the antenna-, diode- and load resistance

The experimental results presented in [16] show that the zero-bias resistance of metal-insulator-metal diodes exhibiting pure tunneling effects falls into the $\text{M}\Omega$ domain. Since on the other hand, the antenna resistance is on the order of 100Ω [2], matching of the antenna and the diode – i.e. fulfilling (2.18) or (2.26) – is not possible in practice. (For the same reason, however, (2.19) can be readily satisfied.) In addition, due to the relatively high practical R_{D} values – just mentioned, (2.22) might also be difficult to meet as R_{load} is typically the input impedance of a low-noise voltage amplifier. However, if instead of the diode voltage it is the diode current what is to be detected, i.e. R_{load} is the input impedance of a low-noise current amplifier [16], the diode- and the load resistance can easily be matched (2.27) in practice. In that case, the signal-to-noise ratio of the detected current – assuming that (2.20) and (2.21) are valid – can be expressed as

$$\frac{\left|\Delta I_{\text{D}}^{0}\right|}{I_{\text{D}}^{\text{n}}} = \frac{\left|I_{\text{D}}^{0} - I_{\text{D}}^{0,\text{B}}\right|}{I_{\text{D}}^{\text{n}}} = \frac{\left(U_{\text{A}}^{1}\right)^{2}}{8\sqrt{kTB}} \left|\frac{i_{\text{D}}^{(2)}\left(U_{\text{D}}^{0,\text{B}}\right)}{i_{\text{D}}^{(1)}\left(U_{\text{D}}^{0,\text{B}}\right)}\right| \sqrt{i_{\text{D}}^{(1)}\left(U_{\text{D}}^{0,\text{B}}\right)}, \tag{2.28}$$

where I_{D}^{n} is the noise current of the amplifier input (R_{load}). That is to say, for a given value of the bias-point quality factor, the signal-to-noise ratio is maximized if the bias-point resistance is minimized. To sum up, if we fix the bias-point resistance at the lowest possible value, i.e. where the diode still shows pure tunneling effects, then we find again that the signal-to-noise ratio of the detected signal is directly proportional to the bias-point quality factor of the diode.

2.3. MIM diode model

We use the simplest model – called tunneling theory – to describe the metal-insulator-metal diode as a quantum transport device [18]. (Please not that, as shown in [13], tunneling current dominates the current of MIM diodes with an insulator thickness lower

DOI:10.15774/PPKE.ITK.2009.001

than 4nm at room temperature, therefore, thermionic emission is completely neglected in this study.) In this model, the metallic electrodes of the diodes are represented by electron reservoirs, each of which emits electrons with an equilibrium distribution to the so called scattering region, i.e. to the insulator layer(s) in our specific case [2][12]. The scattering of the electron wave function – in the model – is described by a single-electron time-invariant Schrödinger equation. The net flow of the electrons passing through between the reservoirs comprises the electrical current of the device.

It is a common approach to consider a one-dimensional device model [18]. In our particular case, we take the surface of the metal- and insulator layers of the diode to be parallel to the yz-plane. The dimensions of the material layers are considered infinite in this plane. The metal-insulator surfaces are assumed to be located at $x = x_l$ and $x = x_r$, i.e. the $\phi(x,V)$ potential is taken to be constant for the regions $x_l \le x$ and $x \le x_r$, respectively, with V being the external voltage. (Subscripts l and r refer to the left- and right-hand side metallic layers, respectively.)

The value of the current density for a given applied voltage is obtained from the integral as follows.

$$J(V) = \int_{\max\{\phi(x_l,V),\phi(x_r,V)\}}^{\infty} T(E_x,V) D(E_x,V) dE_x \,, \tag{2.29}$$

where $T(E_x,V)$ is the tunneling probability for an electron with E_x energy in the x direction, while $D(E_x,V)$ is the net number of the electrons with E_x energy hitting the surface of the metals per unit time and unit area.

According to [18],

$$D(E_x,V) = \frac{qmkT}{2\pi^2\hbar^3} \ln \frac{1 + e^{-\frac{1}{kT}(E_x - \phi(x_l,V) - W_{Fl})}}{1 + e^{-\frac{1}{kT}(E_x - \phi(x_r,V) - W_{Fr})}} \,, \tag{2.30}$$

DOI:10.15774/PPKE.ITK.2009.001

where q and m are the free mass of the electron, respectively, T is the absolute temperature with k being Boltzmann's constant. Variable W_{Fl} and W_{Fr} are the Fermi energies of the metals at the left- and right hand side, respectively. To a first approximation, we assume $T = 0\text{K}$ in our investigations when calculating $D(E_x,V)$. In that case,

$$D(E_x,V)=\begin{cases}\dfrac{qm}{2\pi^2\hbar^3}\left[qV-(W_{Fl}-W_{Fr})\right] \text{ if } E_x < E_{\min},\\ \dfrac{qm}{2\pi^2\hbar^3}\left[qV-(W_{Fl}-W_{Fr})\right](E_{\max}-E_x)/(E_{\max}-E_{\min}) \text{ if } E_{\min}\le E_x \le E_{\max},\\ 0 \text{ otherwise,}\end{cases} \tag{2.31}$$

where

$$E_{\min} = \min\{\phi(x_l,V)-W_{Fl},\phi(x_r,V)-W_{Fr}\}, \tag{2.32}$$

$$E_{\max} = \max\{\phi(x_l,V)-W_{Fl},\phi(x_r,V)-W_{Fr}\}. \tag{2.33}$$

The $T(E_x,V)$ tunneling probability is obtained by solving a single-electron time-invariant Schrödinger equation for the case of – as an example – a left-incident electron. As opposed to [18], instead of the effective mass of the electron, we use simply the free mass of the electron in the equation similarly to [2] and [12]:

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi(x)}{dx^2}+\phi(x,V)\Psi(x)=E_x\Psi(x). \tag{2.34}$$

Please note that, for linearly varying potentials – such as (2.50) and (2.72), the solution to the single-electron Schrödinger equation can be expressed analytically with the help of the Airy functions. Nevertheless, in order to be able to investigate possible nonlinear potential profiles in the future – such as the ones arising when taking into account effect of the possible image forces as well [13][14] – we apply the numerical method introduced in [18] to solve (2.34). (The above mentioned Airy function alternative was used to double-check the results that were obtained through the purely numerical method: the tunneling probability was expressed analytically by means of the Airy functions,

DOI:10.15774/PPKE.ITK.2009.001

while (2.29) was carried out numerically. No difference was found between the results obtained with the help of the two different approaches.)

In what follows, we briefly summarize the numerical method – described in [18] – to solve (2.34) and to calculate the tunneling probability. According to [18],

$$T(E_x,V)=\sqrt{\frac{E_x-\phi(x_r,V)}{E_x-\phi(x_l,V)}}\left|\psi_{j_r}\right|^2, \tag{2.35}$$

where ψ_{j_r} is obtained through solving the space-discretized version of the (2.34) single-electron Schrödinger equation,

$$\mathbf{H}\boldsymbol{\psi}=E_x\boldsymbol{\psi} \tag{2.36}$$

with boundary conditions

$$\frac{1}{\lambda_l-\lambda_l^{-1}}\psi_{j_l-1}-\frac{\lambda_l}{\lambda_l-\lambda_l^{-1}}\psi_{j_l}=1, \tag{2.37}$$

$$\frac{1}{\lambda_r-\lambda_r^{-1}}\psi_{j_r+1}-\frac{\lambda_r}{\lambda_r-\lambda_r^{-1}}\psi_{j_r}=0. \tag{2.38}$$

Here

$$\boldsymbol{\psi}=\left[\psi_{j_l-1}\quad\psi_{j_l}\quad\cdots\quad\psi_{j_r}\quad\psi_{j_r+1}\right]^T, \tag{2.39}$$

$$H_{i\ell}=\begin{cases}\dfrac{\hbar^2}{m}\dfrac{1}{\Delta^2}+\phi(\ell\Delta,V)\ \text{if}\ \ i=\ell,\\ -\dfrac{\hbar^2}{2m}\dfrac{1}{\Delta^2}\ \text{if}\ \ |i-\ell|=1,\\ 0\ \text{otherwise},\end{cases} \tag{2.40}$$

$i=j_l,...,j_r$,

$\ell=j_l-1,...,j_r+1$,

DOI:10.15774/PPKE.ITK.2009.001

where Δ is a suitably chosen space-discretization parameter, while j_l and j_r are integers with $x_l = j_l\Delta$ and $x_r = j_r\Delta$.

Equations (2.37) and (2.38) are the so called Quantum Transmitting Boundary Method (QTBM) [19] boundary equations for a unit-amplitude left-incident electron. Parameters λ_l and λ_r can be determined by solving the following second-order equations, respectively:

$$E_x = \frac{\hbar^2}{m} + \phi(j_l\Delta, V) - \frac{\hbar^2}{2m}\left(\lambda_l + \lambda_l^{-1}\right), \tag{2.41}$$

$$E_x = \frac{\hbar^2}{m} + \phi(j_r\Delta, V) - \frac{\hbar^2}{2m}\left(\lambda_r + \lambda_r^{-1}\right). \tag{2.42}$$

2.4. MIM diodes with double-layer insulators

The idea to apply double-layer insulator in between the metallic layers of the MIM diode originates in the following. Taking advantage of the fact that $J(0) = 0$, the zero-bias quality factor of the diode can be readily shown to have the form of

$$\gamma = \lim_{\Delta V \to 0} \frac{2}{\Delta V} \frac{J(\Delta V) + J(-\Delta V)}{J(\Delta V) - J(-\Delta V)}. \tag{2.43}$$

Based on (2.43), it is obvious that if the MIM diode is completely symmetrical to the yz-plane then the zero-bias quality factor of the diode is zero as in that case $J(-\Delta V) = -J(\Delta V)$. One alternative to break this symmetry is to fabricate the metallic electrodes from different materials [2][16], while an other is to employ two insulator layers of different electrical properties in between the metallic electrodes. (These alternatives can naturally be combined.) In this study we investigate the case when the two insulators are of different permittivity and have different work functions to the metals constituting the electrodes.

For brevity, MIM diodes with double-layer insulators will henceforth (alternatively) be referred to as MIIM (metal-insulator-insulator-metal) diodes.

DOI:10.15774/PPKE.ITK.2009.001

2.4.1. Problem formulation

Figure 2.4 shows the assumed potential profile of the MIIM diode for the case of metal electrodes of identical material and when no external voltage is applied. Please remark that the potential profile at the metal-insulator and insulator-insulator transitions may be more complicated in reality than presented in the figure. The detailed investigation of these transitions, however, is beyond the scope of the dissertation.

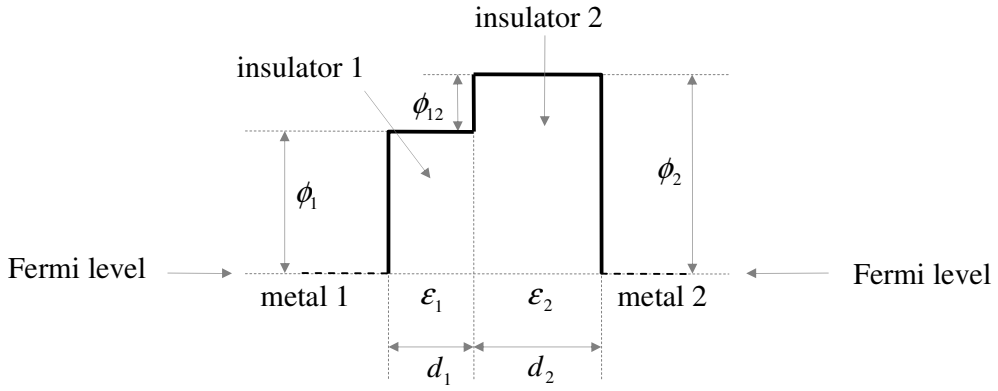

Figure 2.4. Potential profile of the MIIM diode – metal electrodes of identical material with no external voltage applied.

Parameters ε_1 and ε_2 are the relative dielectric constants of metal 1 and metal 2, respectively, while d_1 and d_2 denote the thicknesses of the insulator layers. ϕ_1 and ϕ_2 are the metal 1 to insulator 1 and the metal 2 to insulator 2 work functions, whereas ϕ_{12} is the insulator 1 to insulator 2 work function. As the metallic electrodes are of identical material, no contact potential arises and the following equation holds for the work functions:

$$\phi_1 + \phi_{12} - \phi_2 = 0. \tag{2.44}$$

For this type of diode (MIIM diode, we investigate the resistance and quality factor as a function of the following five diode parameters: the total thickness of the insulator layers (L), the ratio of the thickness of the first insulator layer to the total thickness (r_d), the ratio of the relative dielectric constants (r_ε), the average metal to insulator work function (ϕ_0) and the asymmetry factor of these work functions (α):

DOI:10.15774/PPKE.ITK.2009.001

$$L = d_1 + d_2 \tag{2.45}$$

$$r_d = \frac{d_1}{L}, \tag{2.46}$$

$$r_\varepsilon = \frac{\varepsilon_2}{\varepsilon_1}, \tag{2.47}$$

$$\phi_0 = \frac{\phi_1 + \phi_2}{2}, \tag{2.48}$$

$$\alpha = \frac{\phi_1 - \phi_2}{\phi_1 + \phi_2}. \tag{2.49}$$

When V external voltage is applied to the diode, the potential profile – which is shown in Figure 2.5 – can be expressed as

$$\phi(x,V) = \begin{cases} \dfrac{qV}{2} \text{ if } 0 > x, \\ W_\mathrm{F} + \phi_1 + \dfrac{qV}{2} - \dfrac{x}{d_1}\lambda qV \text{ if } d_1 \geq x \geq 0, \\ W_\mathrm{F} - \dfrac{qV}{2} + \phi_2 + \dfrac{d_1 + d_2 - x}{d_2}(1-\lambda)qV \text{ if } d_1 + d_2 \geq x \geq d_1, \\ -\dfrac{qV}{2} \text{ otherwise,} \end{cases} \tag{2.50}$$

where q is the charge of the electron, W_F is the Fermi energy of the electrodes and λ is the voltage drop across insulator 1 relative to the total drop (V). Based on the continuity equation for the electric displacement,

$$\lambda = \frac{\varepsilon_2 d_1}{\varepsilon_2 d_1 + \varepsilon_1 d_2} = \frac{r_d r_\varepsilon}{r_d r_\varepsilon + 1 - r_d}. \tag{2.51}$$

When deriving (2.50) and (2.51), it is assumed that the electric field inside the insulators is homogeneous and is perpendicular to the boundary surfaces of the different layers. It is also assumed that there are no surface charges at the boundary of the insulators.

DOI:10.15774/PPKE.ITK.2009.001

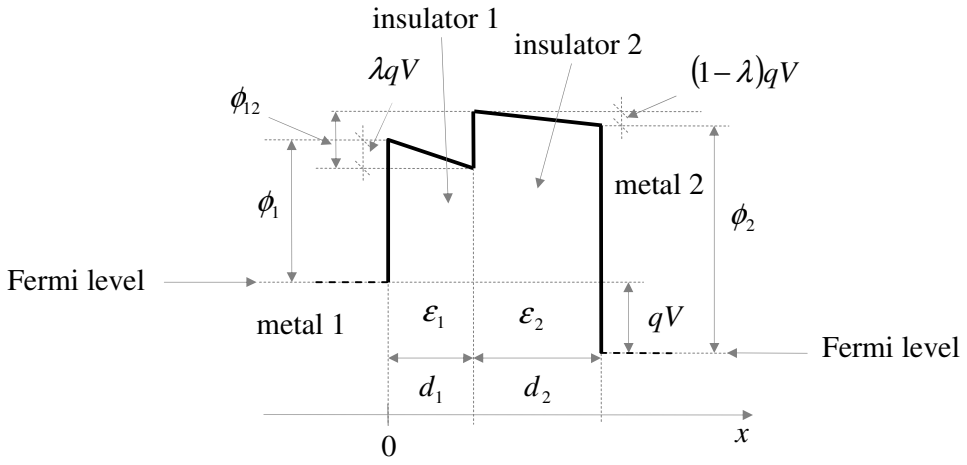

Figure 2.5. Potential profile of the MIIM diode – metal electrodes of identical material with external voltage applied.

For the sake of completeness, we present the potential profile of the MIIM diode also for the case of metal electrodes of different material (Figure 2.6.). As now the metallic electrodes are of different material, a contact potential may arise, which is given by

$$V_c = \phi_1 + \phi_{12} - \phi_2 \,. \tag{2.52}$$

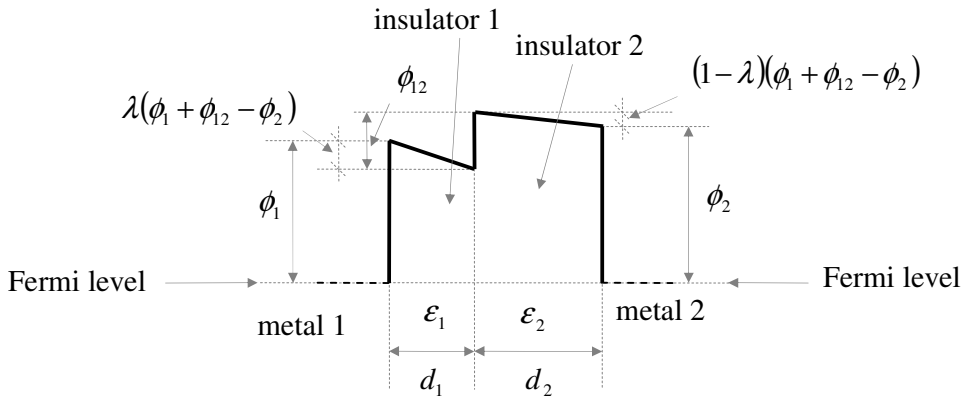

Figure 2.6. Potential profile of the MIIM diode – metal electrodes of different material with no external voltage applied.

DOI:10.15774/PPKE.ITK.2009.001

2.4.2. Results and discussion

Numerical method

The tunneling spectrum of the particular potential profile described above (2.50) is sufficiently smooth, it contains no sharp features such as narrow resonance peaks, which makes it possible to use the simple left rectangle rule for the numerical integration (2.29) without introducing significant numerical errors:

$$\int_{x_1}^{x_2} f(x)dx \approx \sum_{i=0}^{\frac{x_2-x_1}{\Delta x}-1} f(x_1 + i\cdot\Delta x)\Delta x, \tag{2.53}$$

where f is the function to be integrated, x_1 and x_2 are the limits of the integration and Δx is the step size of the numerical integration. The formula is applied to integrate $T(E_x,V)D(E_x,V)$ in (2.29).

In order to determine the different derivatives of the DC I-V curve of the diode, we use central-difference formulas of $O(h^4)$:

$$f^{(1)}(x_0) \approx \frac{f(x_0-2\Delta x)-8f(x_0-\Delta x)+8f(x_0+\Delta x)-f(x_0+2\Delta x)}{12\Delta x}, \tag{2.54}$$

$$f^{(2)}(x_0) \approx \frac{-f(x_0-2\Delta x)+16f(x_0-\Delta x)-30f(x_0)+16f(x_0+\Delta x)-f(x_0+2\Delta x)}{12\Delta x^2}, \tag{2.55}$$

where f is the function to be differentiated, x_0 is the point of differentiation and Δx is the step size of the numerical differentiation. The latter formulas are applied to differentiate function $J(V)$ given by (2.29).

The value of the parameters of the different numerical methods are set to $\Delta = 10^{-12}\,\mathrm{m}$ (step size of the spatial discretization), $\Delta E_x = 10^{-3} W_\mathrm{F}$ (step size of the numerical integration) and $\Delta V = 0.05\mathrm{V}$ (step size of the numerical differentiation). The fact that decreasing the value of these parameters does not alter the numerical results, indicate that these values provide sufficient accuracy.

DOI:10.15774/PPKE.ITK.2009.001

Diode parameter values

The cross-section of the diode is set to $a = 0.1\mu m^2$. (Note that nanoantenna-coupled tunnel-diode detectors with a cross-section area of 0.012, 0.056 and $0.0025\mu m^2$ was reported in [7], [8] and [16], respectively.) The Fermi energy of most metals lies between 1 and 15eV [20]. In the present calculations we use $W_F = 3.5eV$.

In each of the diagrams in this section, three of the five diode parameters – defined in Section 2.4.1 – were fixed, while the other two were considered as independent variables. For L and ϕ_0, the constant values $L = 1.033nm$ and $\phi_0 = 0.6448eV$ are calculated from measurement results [2] using the formulas presented in [2]. Please note that these formulas – as it will be discussed later on this section – are of limited validity. Nevertheless, the order of these values can be considered typical [2][16]. When fixed, the value of r_ε was set to 0.398, which is the ratio of the DC relative permittivity of SiO_2 and Al_2O_3 (3.9 and 9.8) – two insulators that are widely used in CMOS technology. For the invariant value of r_d, 0.6132 was used, which is – as we will subsequently see – the optimum thickness ratio corresponding to $r_\varepsilon = 0.398$. Finally, the fixed value of α was chosen to 0 as our focus is on the effects of the asymmetry in the dielectric constants.

DOI:10.15774/PPKE.ITK.2009.001

Numerical results

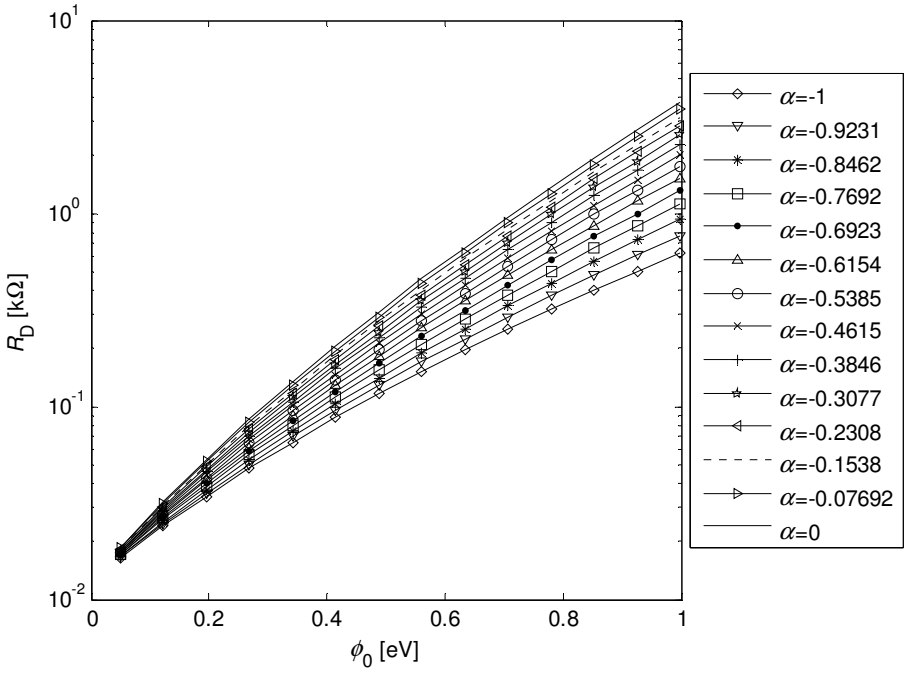

Figure 2.7. Resistance (R_D) of the MIIM diode as a function of the average work function (ϕ_0) and the asymmetry factor (α).

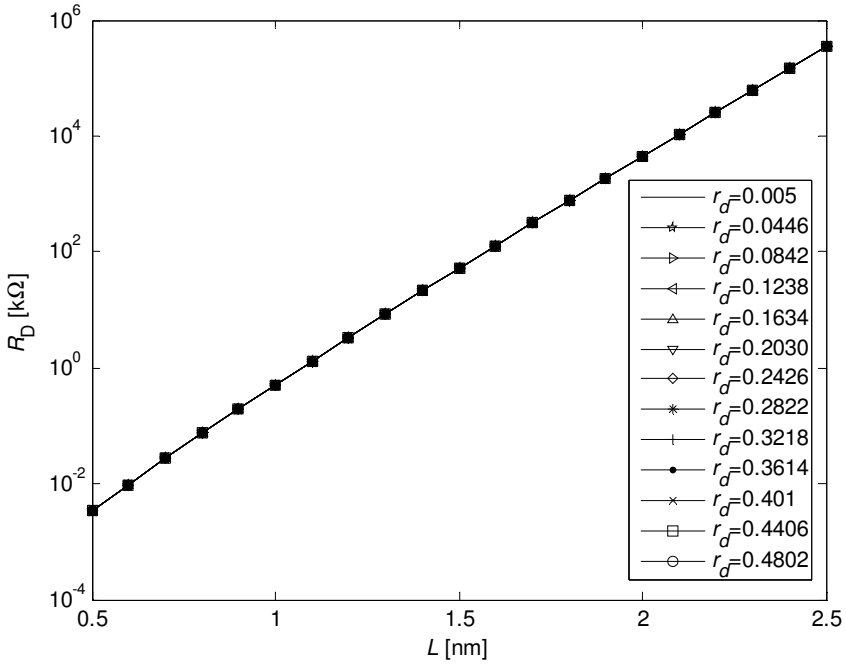

Figure 2.8. Resistance (R_D) of the MIIM diode as a function of the total thickness (L) and the thickness ratio (r_d).

DOI:10.15774/PPKE.ITK.2009.001

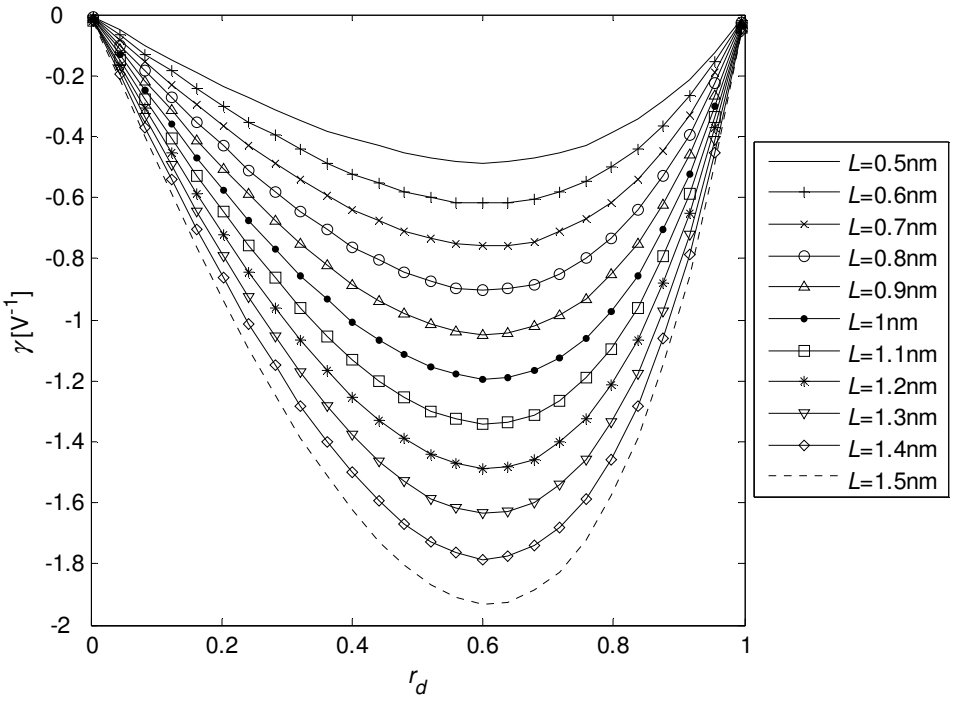

Figure 2.9. Quality factory (γ) of the MIIM diode as a function of the thickness ratio (r_d) and the total thickness (L).

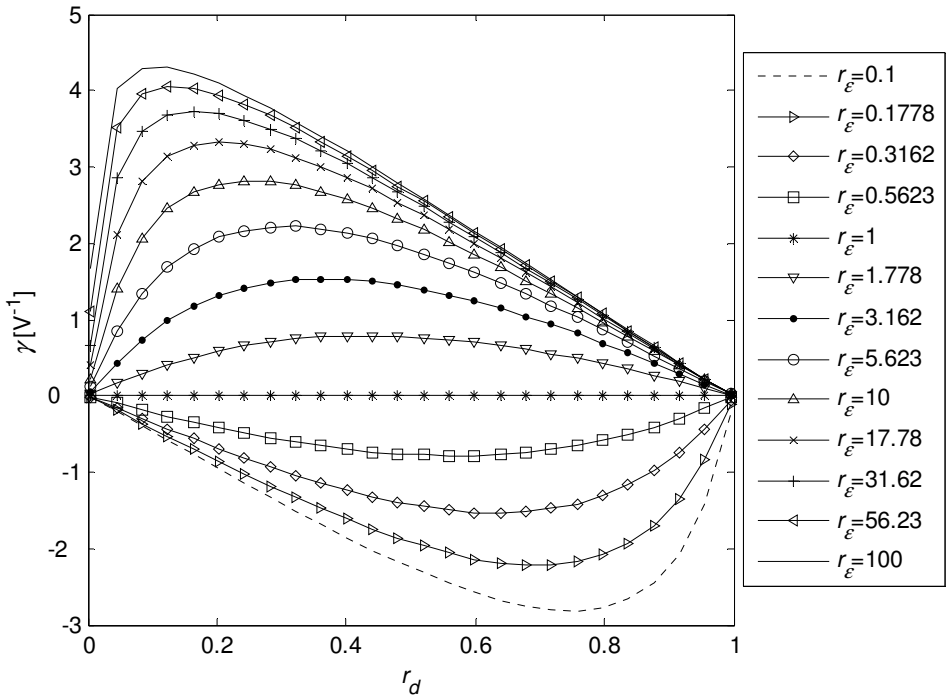

Figure 2.10. Quality factory (γ) of the MIIM diode as a function of the thickness ratio (r_d) and the ratio of the dielectric constants (r_ε).

DOI:10.15774/PPKE.ITK.2009.001

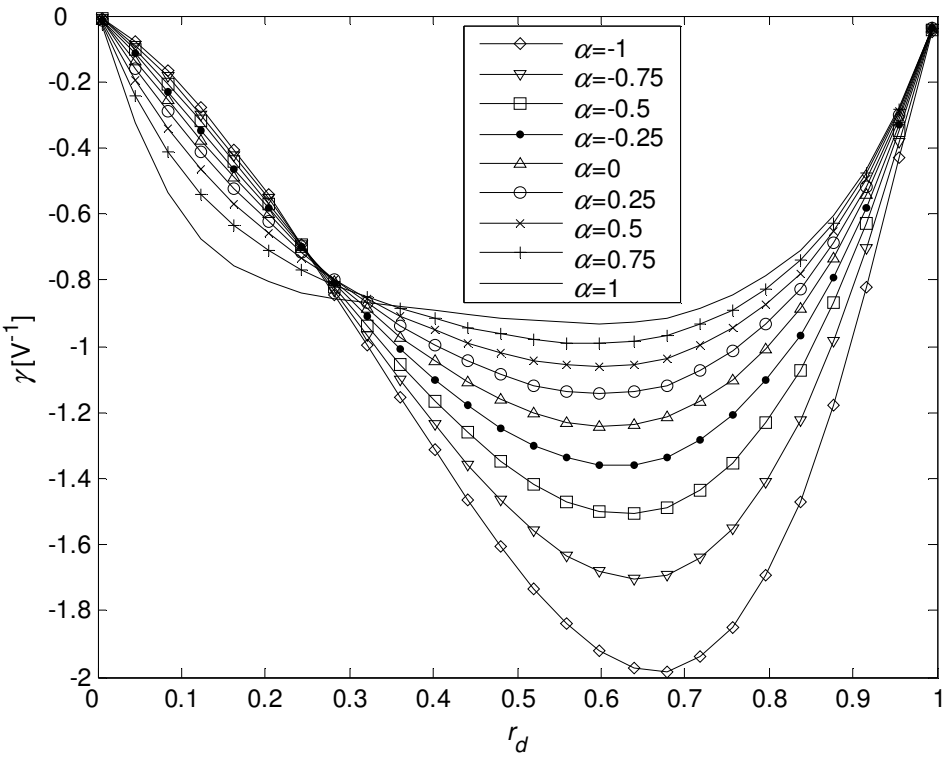

Figure 2.11. Quality factory (γ) of the MIIM diode as a function of the thickness ratio (r_d) and the asymmetry factor (α).

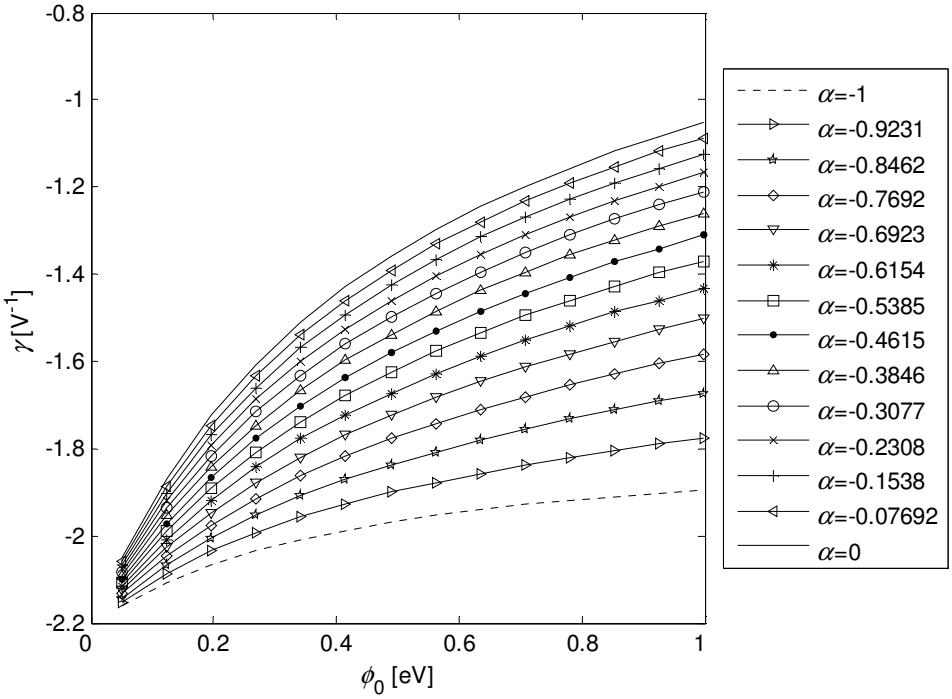

Figure 2.12. Quality factory (γ) of the MIIM diode as a function of the thickness ratio (ϕ_0) and the asymmetry factor (α).

DOI:10.15774/PPKE.ITK.2009.001

Discussion

To give a (physical) insight on why the numerical results in this section come to be, we are going to make an attempt to explain the qualitative behavior of γ as a function of the different diode parameters using the so called WKB approximation (e.g. [2][12]). According to this approximation, the tunneling probability introduced in (2.35) can be expressed as

$$T\left(E_x,V\right)=e^{-2\frac{\sqrt{2m}}{\hbar}\int_0^L\sqrt{\phi(x,V)-E_x}dx}. \tag{2.56}$$

If we assume that the tunneling current is mainly due to the electrons at the Fermi-level of the electrode of the higher potential, i.e. we take the current directly proportional to the tunneling probability of these electrons, then according to (2.43) and (2.56), γ can be approximated as

$$\gamma=\lim_{\Delta V\to 0}\frac{2}{\Delta V}\frac{T\left(W_F+\frac{q\Delta V}{2},\Delta V\right)-T\left(W_F+\frac{q\Delta V}{2},-\Delta V\right)}{T\left(W_F+\frac{q\Delta V}{2},\Delta V\right)+T\left(W_F+\frac{q\Delta V}{2},-\Delta V\right)}= \tag{2.57}$$

$$=\lim_{\Delta V\to 0}\frac{2}{\Delta V}\frac{1-\frac{T\left(W_F+\frac{q\Delta V}{2},-\Delta V\right)}{T\left(W_F+\frac{q\Delta V}{2},\Delta V\right)}}{1+\frac{T\left(W_F+\frac{q\Delta V}{2},-\Delta V\right)}{T\left(W_F+\frac{q\Delta V}{2},\Delta V\right)}}.$$

Please remark that in order for (2.57) to be in accordance with the approximation proposed above, the V-dependence of $\phi(x,V)$ has to be formulated in such a way that the potentials of the electrodes – in contrary to [2] – are shifted symmetrically as shown in (2.50). Substituting (2.56) into (2.57), we obtain

DOI:10.15774/PPKE.ITK.2009.001

$$\gamma = \lim_{\Delta V \to 0} \frac{2}{\Delta V} \frac{1 - e^{-2\frac{\sqrt{2m}}{\hbar}\int_0^L \left[\sqrt{\phi(x,-\Delta V) - W_F - \frac{q\Delta V}{2}} - \sqrt{\phi(x,\Delta V) - W_F - \frac{q\Delta V}{2}}\right] dx}}{1 + e^{-2\frac{\sqrt{2m}}{\hbar}\int_0^L \left[\sqrt{\phi(x,-\Delta V) - W_F - \frac{q\Delta V}{2}} - \sqrt{\phi(x,\Delta V) - W_F - \frac{q\Delta V}{2}}\right] dx}}. \tag{2.58}$$

Replacing the exponential expression in the numerator with its Taylor series expansion in its exponent (around 0), γ can easily be shown to have the form of:

$$\gamma = \lim_{\Delta V \to 0} \frac{1}{\Delta V} \frac{2\sqrt{2m}}{\hbar} \int_0^L \left[\sqrt{\phi(x,-\Delta V) - W_F - \frac{q\Delta V}{2}} - \sqrt{\phi(x,\Delta V) - W_F - \frac{q\Delta V}{2}} \right] dx. \tag{2.59}$$

In other words, according to (2.59), γ is directly proportional to the difference of the areas of the 'square-rooted' potential barriers that the electrons at the Fermi-level of the left and right electrode 'see' at positive and negative bias, respectively. Now again if we carry out a Taylor series expansion of the integrand in (2.59) in ΔV (around 0), we have:

$$\gamma = -\frac{2\sqrt{2m}}{\hbar} \int_0^L \frac{1}{\sqrt{\phi(x,0) - W_F}} \left(\frac{\partial}{\partial V} \phi(x,V) \right) \Bigg|_{V=0} dx. \tag{2.60}$$

(2.60) is equivalent to the formula of γ derived in [2], however, in order to obtain the same result for a specific barrier, $\phi(x,V)$ has to be formulated with symmetrical and asymmetrical electrode potential shifts for the former and latter formula, respectively. (In the symmetrical case, the potentials of metal 1 and metal 2 are shifted by $qV/2$ and $-qV/2$, respectively, while for the asymmetrical case, the shifts are 0 and $-qV$, respectively.) Please remark that while the derivation in [2] applies four approximation steps, the present one includes only two such steps. In addition, unlike [2], it contains an intermediate formula (2.59) that provides the above graphical interpretation of γ.

Substituting (2.50) into (2.60), for the zero-bias quality factor of the MIIM diode, we obtain

DOI:10.15774/PPKE.ITK.2009.001

$$\gamma = \frac{qL\sqrt{2m}}{\hbar}\left(\frac{r_\varepsilon}{\sqrt{\phi_2}} - \frac{1}{\sqrt{\phi_1}}\right)\frac{r_d(1-r_d)}{r_\varepsilon r_d + 1 - r_d}. \tag{2.61}$$

It can be easily seen that according to (2.61),

$$r_{d,\mathrm{opt}} = \arg\max_{0\le r_d\le 1}|\gamma(r_d)| = \frac{1}{1+\sqrt{r_\varepsilon}}, \tag{2.62}$$

$$\left|\gamma\right|_{r_d=r_{d,\mathrm{opt}}}\right| = \frac{qL\sqrt{2m}}{\hbar}\left|\frac{r_\varepsilon}{\sqrt{\phi_2}} - \frac{1}{\sqrt{\phi_1}}\right|\frac{1}{\left(1+\sqrt{r_\varepsilon}\right)^2}, \tag{2.63}$$

$$\gamma|_{r_d=0} = \gamma|_{r_d=1} = 0, \tag{2.64}$$

$|\gamma(r_d)|$ is bounded, continuous and concave for $0 \le r_d \le 1$, monotonic increasing for $0 \le r_d \le r_{d,\mathrm{opt}}$ and monotonic decreasing for $r_{d,\mathrm{opt}} \le r_d \le 1$.

The numerical results in Figure 2.9 show a fairly similar behavior of γ as a function of r_d. (Please remark that, for the sake of simplicity, $|\gamma|$ will be simply referred to as γ on several occasions throughout the analysis of the numerical results in this study.) In addition, it can be observed on the same diagram that the L-dependence of γ is practically homogeneous linear. As a consequence, we can conclude that (2.61) gives qualitatively accurate results in terms of the r_d- and L-dependence of the zero-biased quality factor in the case of $\alpha = 0$. Substituting $L = 1.033\mathrm{nm}$, $\phi_1 = \phi_2 = \phi_0 = 0.6448\mathrm{eV}$, $r_d = 0.6132$ and $r_\varepsilon = 0.398$ into (2.61), we obtain $\gamma = -1.492V^{-1}$, while $\gamma = -1.242V^{-1}$ is found for the same diode parameter values in Figure 2.9. In other words, the figures show that (2.61) involve some quantitative errors.

Now let us consider the results in Figure 2.10. We find on the diagram that increasing (decreasing) the value of r_ε (from 1) shifts $r_{d,\mathrm{opt}}$ towards 0 (1). This is in accordance with the behavior implied by (2.62). In addition, it can easily be verified that this formula is not only qualitatively correct with respect to $r_{d,\mathrm{opt}}(r_\varepsilon)$ but it provides fairly accurate

DOI:10.15774/PPKE.ITK.2009.001

quantitative results as well. We can conclude that the more asymmetric the relative dielectric constants are the closer $r_{d,\mathrm{opt}}$ is to 0 or 1. Considering the fact that L is only around a few nanometers, this implies serious implementation challenges if the optimal thickness ratio is to be realized. Furthermore, it can be readily seen that the C_D capacitance of the junction at $r_d = r_{d,\mathrm{opt}}$ is

$$C_\mathrm{D}\big|_{r_d=r_d,\mathrm{opt}} = \frac{a}{L}\varepsilon_0\sqrt{\varepsilon_1\varepsilon_2}\ . \tag{2.65}$$

As $\varepsilon_1, \varepsilon_2 \geq 1$, C_D is directly proportional to $\sqrt{r_\varepsilon}$ ($\sqrt{r_\varepsilon^{-1}}$) for $r_\varepsilon \geq 1$ ($r_\varepsilon \leq 1$), which means that an extreme value for r_ε results in a high value for C_D. This makes conditions (2.20) and (2.21) harder to be met and as a consequence the cut-off frequency [2] of the detector decreases (2.17). If the frequency of the radiation to be detected exceeds the former frequency, the detector performance may seriously deteriorate [2]. Another argument against pushing r_ε into the extreme domains is that the additional gain in γ due to the increase (decrease) in r_ε vanishes as we get farther and father away from 1. This saturation is clearly visualized in Figure 2.10 for both the optimal thickness ratio and the moderate values of the ratio such as $r_d = 0.5$. The latter phenomena manifests also in (2.63): it is easy to prove that – for $\phi_1 = \phi_2 = \phi_0$ – the formula is a continuous, concave function of $\log r_\varepsilon$ and has its minimum in $\log r_\varepsilon = 0$. The same holds for the absolute value of (2.61) at $r_d = 0.5$. However, as opposed to the case of $r_d = r_{d,\mathrm{opt}}$, the diode capacitance can be kept at a low value even for extreme values of r_ε if $r_d = 0.5$ is applied. Namely, if we investigate the expression of C_D,

$$C_\mathrm{D}\big|_{r_d=0.5} = \frac{2a}{L}\varepsilon_0\frac{\varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2}, \tag{2.66}$$

we find that for $r_\varepsilon >> 1$ or $r_\varepsilon << 1$ it is the smaller dielectric constant what mostly determines the capacitance. On the other hand, the loss in γ due to the r_d – r_ε mismatch at $r_d = 0.5$ – expressed as

DOI:10.15774/PPKE.ITK.2009.001

$$\frac{\gamma|_{r_d=r_{d,\mathrm{opt}}}}{\gamma|_{r_d=0.5}}=\frac{2}{1+\frac{2\sqrt{r_\varepsilon}}{1+r_\varepsilon}} \tag{2.67}$$

– is lower than a factor of 2 even for extreme values of r_ε. To sum up, if $r_{d,\mathrm{opt}}$ values lower or higher than 0.5 is hard to be realized – and a high γ and a low C_D are to be achieved – it could be more advantageous to apply a more extreme r_ε and have $r_d=0.5$.

Figure 2.11 shows that the asymmetry of the work functions can either enhance or impair the asymmetry of the diode due to the insulators of different dielectric constants. Based on the diagram, we can conclude that in order for γ to be improved at $r_d=r_{d,\mathrm{opt}}$ – relative to the $\alpha=0$ case – it is the insulator with the lower dielectric constant what needs to have a higher work function to the metal making up the electrodes. Please note that the same qualitative behavior is implied by the second factor of (2.61). However, if we consider r_d as an independent variable, this formula is no longer qualitatively correct. Whereas according to (2.61), changing α is equivalent to multiplying the $\gamma(r_d)$ curve by a constant, the diagram reveals that a change in α not only causes some shift in the optimal value of the thickness ratio but also results in a considerable change in the characteristic of the curve. Namely, while for $\alpha=0$ the curve has a definite extremum point, in the case of $\alpha=1$ its maxima gets fairly wide with a still relatively high value. The latter phenomenon could be efficiently exploited, if the device fabrication process was such that thickness ratio of the insulators could be realized only with a high uncertainty. Finally, it is interesting to remark that – in the particular case of Figure 2.11 – γ behaves just the other way around as a function of α in the left part of the r_d-axis as in the right part of that: for thickness ratios lower than approximately 0.3, it is the insulator with the higher dielectric constant what is required to have a higher work function to the metal in order for the zero-bias quality factor to be maximized.

Considering Figure 2.12, we find that – in part – (2.61) shows the same qualitative behavior as the diagram: γ decreases with ϕ_0. On the other hand, as ϕ_0 approaches zero,

DOI:10.15774/PPKE.ITK.2009.001

the zero-bias quality factor does not go to infinity as opposed to what is predicted by the formula.

To conclude this section, let us briefly investigate the zero-bias diode resistance as function of the different diode parameters. It can be observed in Figure 2.7 and Figure 2.8 that R_{D} increases nearly exponentially with L and ϕ_0, it shows some minor dependency on α and does not depend on r_d as long as $\alpha = 0$. For the sake of completeness, we present the formula of R_{D} that can be obtain using the same analytical approximations that were applied in [2] to derive the expression of the zero-bias resistance:

$$R_{\mathrm{D}} = \frac{2\sqrt{2}\pi^2\hbar^2}{q^2\sqrt{m}} \frac{L}{a\sqrt{\phi_0}} \left[r_d \frac{1}{\sqrt{1+\alpha}} + (1-r_d) \frac{1}{\sqrt{1-\alpha}} \right] e^{\frac{2\sqrt{2m}}{\hbar}\sqrt{\phi_0}L\left[r_d\sqrt{1+\alpha}+(1-r_d)\sqrt{1-\alpha}\right]}. \tag{2.68}$$

2.5. Comparison of MIM diodes with single- and double-layer insulators

In the first part of this section, after describing the potential profile assumed for MIM diodes with single-layer insulators, numerical results on the quality factor and resistance for theses type of diodes are presented, which are then discussed in brief. Finally, in the second part of the section, MIM diodes with single- and double layer insulators are compared in terms of the sensitivity that can be achieved when applied in nanoantenna-coupled MIM diode detectors.

2.5.1. MIM diodes with single-layer insulators

For brevity, MIM diodes with single-layer insulators will henceforth (alternatively) be referred to as MIM diodes.

Problem formulation

The potential profile of the MIM diode is described in Figure 2.13 [2].

DOI:10.15774/PPKE.ITK.2009.001

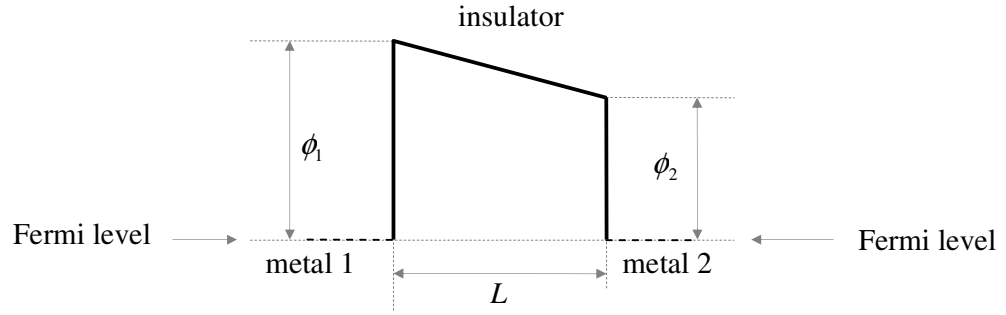

Figure 2.13. Potential profile of the MIM diode single-layer insulator – metal electrodes of different material with no external voltage applied.

Parameters ϕ_1 and ϕ_2 are the metal to insulator work functions, while L is the thickness of the insulator. As the metallic electrodes are of different material, a contact potential may arise, which can be expressed as

$$V_c = \phi_1 - \phi_2 . \tag{2.69}$$

For this type of diode (MIM diode), we investigate the resistance and quality factor as a function of the following three diode parameters: the thickness of the insulator layer (L), the average metal to insulator work function (ϕ_0) and the asymmetry factor of these work functions (α):

$$\phi_0 = \frac{\phi_1 + \phi_2}{2}, \tag{2.70}$$

$$\alpha = \frac{\phi_1 - \phi_2}{\phi_1 + \phi_2}. \tag{2.71}$$

When V external voltage is applied to the diode, the potential profile can be expressed as

$$\phi(x,V) = \begin{cases} \dfrac{qV}{2} & \text{if } 0 > x, \\ W_{F1} + \phi_1 + \dfrac{qV}{2} - \dfrac{x}{L}\left(\phi_1 - \phi_2 + qV\right) & \text{if } L \geq x \geq 0, \\ W_{F1} - W_{F2} - \dfrac{qV}{2} & \text{otherwise,} \end{cases} \tag{2.72}$$

DOI:10.15774/PPKE.ITK.2009.001

where q is the charge of the electron, whereas W_{F1} and W_{F2} are the Fermi energies of the electrodes.

Numerical results

When generating the following diagrams, the same numerical and diode parameters were used as in the case of the MIIM diode. For each of the diagrams, one of the three diode parameters – defined in this section – was fixed, while the other two were considered as independent variables. The constant value of these parameters was set to $L = 1.033\text{nm}$, $\phi_0 = 0.6448\text{eV}$ and $\alpha = 0.2913$.

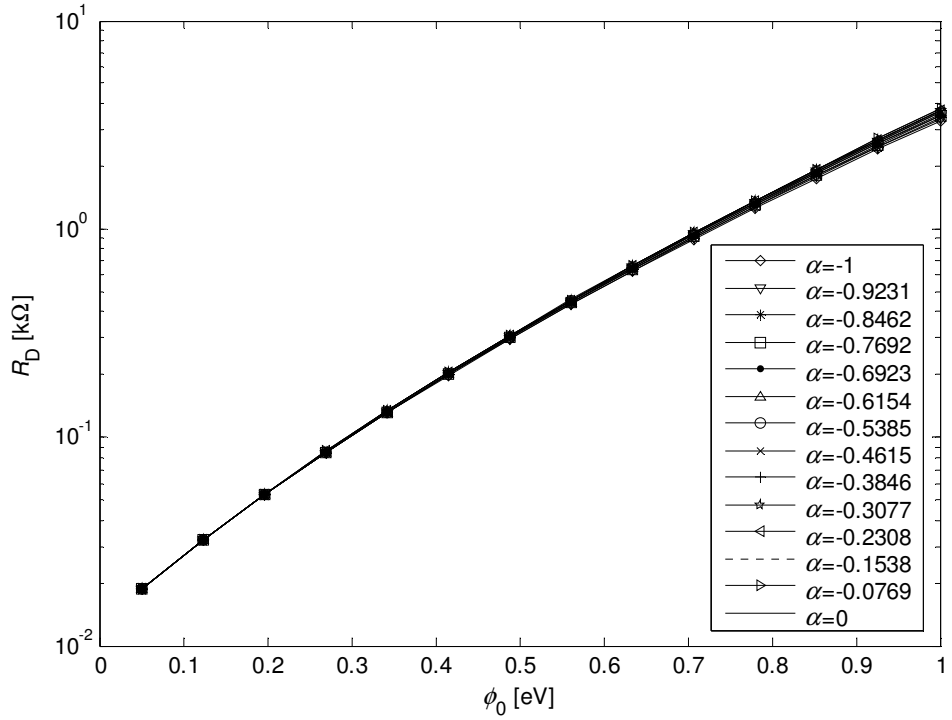

Figure 2.14. Resistance (R_D) of the MIM diode as a function of the average work function (ϕ_0) and the asymmetry factor (α).

DOI:10.15774/PPKE.ITK.2009.001

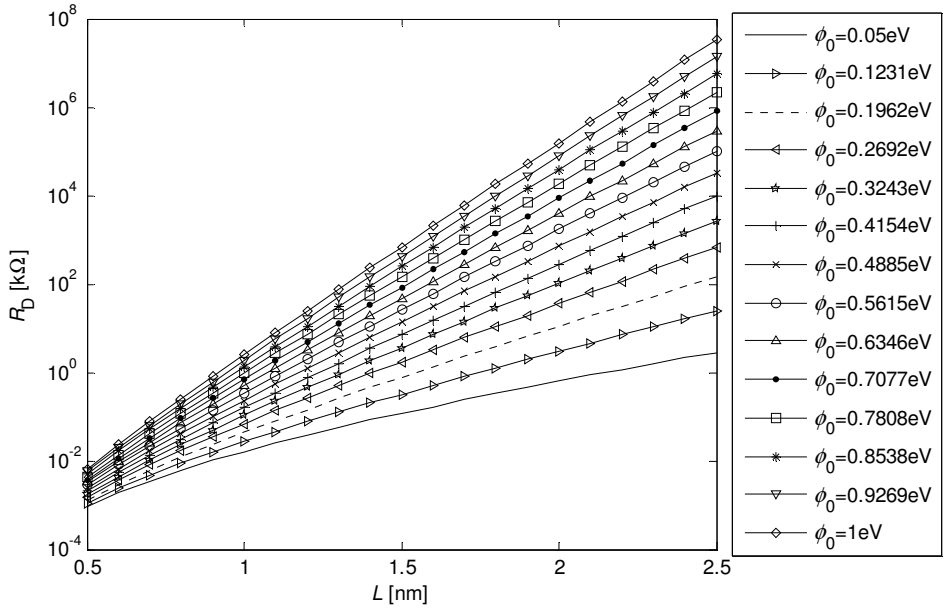

Figure 2.15. Resistance (R_D) of the MIM diode as a function of the thickness (L) and the average work function (ϕ_0).

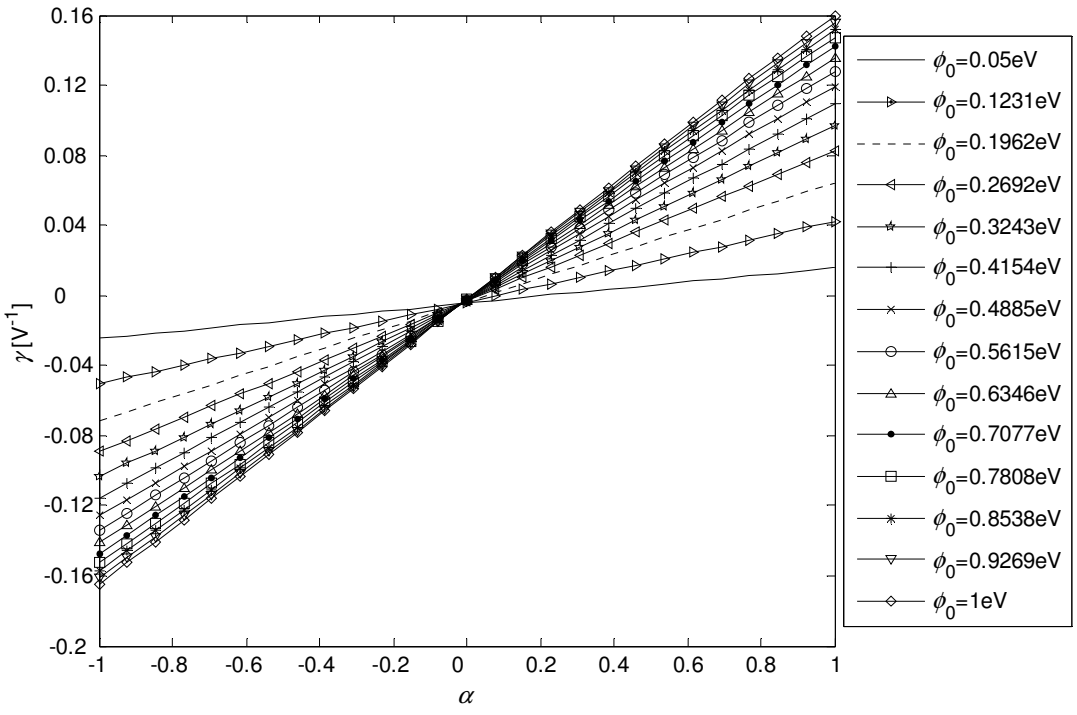

Figure 2.16. Quality factor (γ) of the MIM diode as a function of and the asymmetry factor (α) and the average work function (ϕ_0).

DOI:10.15774/PPKE.ITK.2009.001

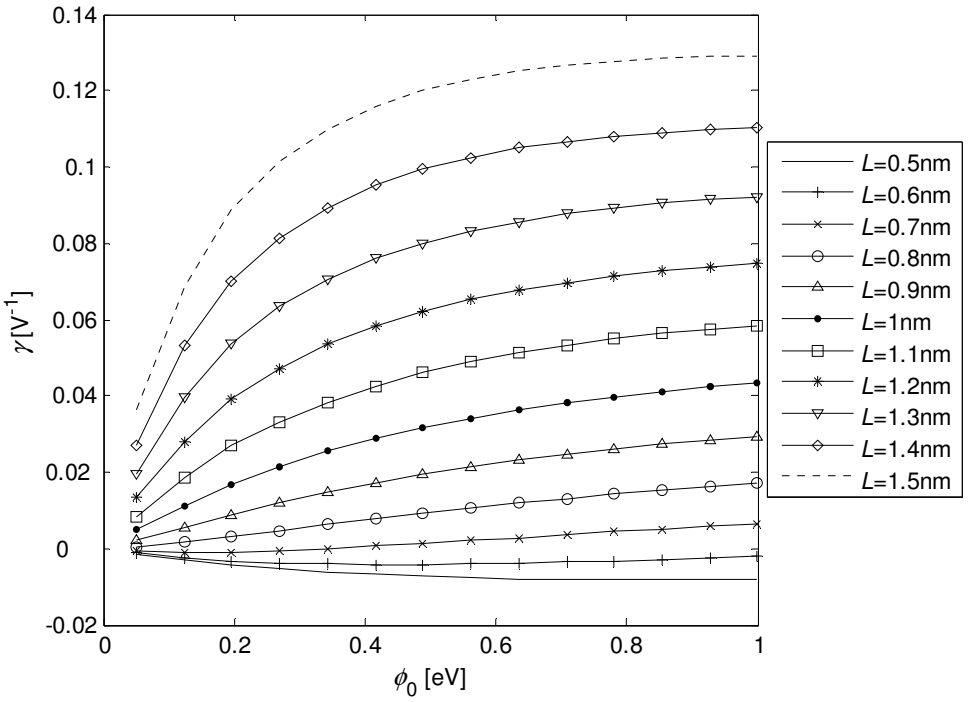

Figure 2.17. Quality factor (γ) of the MIM diode as a function of the average work function (ϕ_0) and the thickness (L).

Discussion

The resistance and quality factor of MIM diodes with single-layer insulators was investigated through analytical approximations in [2]. However – as it will be discussed in the following and was discussed above – these analytical approximations may lead not only to quantitative but also to qualitative errors. This is the main reason why the analysis in this study is carried out numerically for both types of diodes.

The expressions of the zero-bias resistance and quality factor derived in [2] are

$$R_{\mathrm{D}} = \frac{2\sqrt{2}\pi^2\hbar^2}{q^2\sqrt{m}} \frac{L}{a\sqrt{\phi_0}} e^{\frac{2\sqrt{2m}}{\hbar}\sqrt{\phi_0}L}, \tag{2.73}$$

$$\gamma = \frac{q\sqrt{2m}}{6\hbar}\alpha\frac{L}{\sqrt{\phi_0}}. \tag{2.74}$$

If R_{D}, γ, L, ϕ_0 and a are measured in Ω, V^{-1}, nm, eV and $\mu\mathrm{m}^2$, respectively, (2.73) and (2.74) take the form of

DOI:10.15774/PPKE.ITK.2009.001

$$R_{\mathrm{D}}^{[\Omega]} = 3.167 \cdot 10^{-2} \cdot \frac{L^{[\mathrm{nm}]}}{a^{[\mu\mathrm{m}^2]} \sqrt{\phi_0^{[\mathrm{eV}]}}} e^{10.25\sqrt{\phi_0^{[\mathrm{eV}]}} L^{[\mathrm{nm}]}}, \tag{2.75}$$

$$\gamma^{[\mathrm{V}^{-1}]} = 0.8538 \cdot \alpha \frac{L^{[\mathrm{nm}]}}{\sqrt{\phi_0^{[\mathrm{eV}]}}}. \tag{2.76}$$

If we substitute – as an example – $L = 1.033\mathrm{nm}$, $\phi_0 = 0.6448\mathrm{eV}$ and $a = 0.1\mathrm{nm}^2$ into (2.73) and (2.74), we obtain $R_{\mathrm{D}} = 2\mathrm{k}\Omega$ and $\gamma = 0.32\mathrm{V}^{-1}$. If we compare these figures to those found in Figure 2.14 and Figure 2.16 for the same values of L, ϕ_0 and a, i.e. $R_{\mathrm{D}} = 703.3\Omega$ and $\gamma = 0.04147\mathrm{V}^{-1}$, we can conclude that the approximations applied in [2] give rise to considerable quantitative errors.

Now let us investigate the qualitative behavior of γ as a function of L, ϕ_0 and α. Figure 2.16 shows that the α-dependence of γ is practically homogeneous linear, i.e. (2.74) gives correct results concerning this dependence. We can found on the same diagram that the dependence of γ on L is approximately linear for higher values of ϕ_0 and L as the gaps between the neighboring curves are practically equal. Consequently, (2.74) provides more or less accurate results regarding this dependence as well. However, when it comes to the ϕ_0-dependence of γ, there is an obvious contradiction between the numerical results in Figure 2.17 and (2.74). While the diagram shows that γ increases with ϕ_0, according to (2.74), γ is inversely proportional to the square root of ϕ_0, which indicate the approximations employed in [2] result also in a serious qualitative error.

As a consequence, it can be pointed out that the analytical expressions concerning R_{D} and γ, i.e. (2.73) and (2.74) are of limited validity.

2.5.2. Comparative analysis

The comparison of the MIM and MIIM diodes is carried out starting from the following design problem: given the zero-bias resistance of the diode (R_{D}) (e.g. according to (2.18)), maximize the zero-bias quality factor (γ) of that (2.23).

DOI:10.15774/PPKE.ITK.2009.001

First let us consider the design problem in the case of the MIM diode. As we saw in Section 2.5.1, R_D grows approximately exponentially with both the insulator thickness (L) and the average work function (ϕ_0), while it does not practically depend on the asymmetry factor of the work functions (α). In addition, also γ increases with ϕ_0 and L, while its dependence on α is approximately homogeneous linear. As a consequence, if we set up a practical upper bound for both L and ϕ_0, the optimum value of these parameters – that maximizes γ and fulfils the constraint concerning R_D – can be determined. The optimum value of α is obviously ± 1.

Now let us consider the design problem for the MIIM diode. If we restrict our investigations to $\alpha = 0$ – i.e. to the case when the metal making up the electrodes has the same work function value to both insulators – we can conclude based on Figure 2.7 (solid curve with no marker) and Figure 2.14 (Figure 2.8 and Figure 2.15, curve with $\alpha = 0$) that the dependence of R_D on ϕ_0 (L) is practically the same. As a result, the optimum values of ϕ_0 and L for the MIM diode, fulfill the constraint on R_D also for the MIIM diode. The value of γ corresponding to these values of ϕ_0 and L is obviously just a lower bound for the maximum value of γ (corresponding to the solution of the design problem). Since, for the MIIM diode, γ (R_D) grows as L increases (increases) and ϕ_0 decreases (increases), the value of γ can be further increased – and the constraint on R_D can be still met – by setting the value of L and ϕ_0 higher and lower relative to their optimum for the MIM diode, respectively. Nevertheless, if we compare the values of γ corresponding to the same values of ϕ_0 and L for both types of MIM diodes, we obtain a lower bound for the gain of the MIIM diode over the MIM diode. As an example, if $L \approx 1\text{nm}$ and $\phi_0 = 1\text{eV}$, the value of γ is around 0.15V^{-1} (Figure 2.16, curve with diamond marker) and -1.05V^{-1} (Figure 2.12, solid curve with no marker) for the MIM ($\alpha = 1$) and the MIIM diode, respectively. Or if – for instance – $\phi_0 \approx 0.65\text{eV}$ and $L = 1.5\text{nm}$, the value of γ is around 0.45V^{-1} (Figure 2.17, dashed curve) and -1.95V^{-1} (Figure 2.9, dashed curve) for the MIM ($\alpha = 1$) and MIIM diode, respectively. Now if – in the case of the first example – we decrease ϕ_0 to (e.g.) 0.5eV and increase L to 1.325nm, γ changes to -1.85V^{-1} (according to further calculations not presented here),

DOI:10.15774/PPKE.ITK.2009.001

while R_D is left unchanged. If – in the case of the second one – ϕ_0 is decreased to (e.g.) 0.3224eV and L is increased to 1.936nm, γ changes to -3.45V^{-1}, whereas the value of R_D is unaltered. These examples demonstrate that in the $\phi_0 \le 1\text{eV}$, $0.5\text{nm} \le L \le 1.5\text{nm}$ domain – assuming $r_\varepsilon = 0.6132$ – the gain of the MIIM diode over the other type of diode is approximately 5 or even higher.

When, in the case of the MIIM diode, $\alpha \neq 0$, the comparison gets a bit more complicated as the dependence of R_D on L and ϕ_0 (Figure 2.7) will somewhat differ from that for the MIM diode (Figure 2.14). As we saw in 2.4.2 – assuming $r_\varepsilon = 0.6132$ and $r_d = r_{d,\text{opt}}(r_\varepsilon)$ – decreasing (increasing) the value of α from 0 results in a decrease (increase) and in an increase (decrease) in R_D and γ, respectively. In other words, if $\alpha < 0$ ($\alpha > 0$), the gain of the MIIM diode over MIM one is higher (lower) relative to the $\alpha = 0$ case. In general, if the insulator with the lower dielectric constant has a higher (lower) work function to the metal making up the electrodes, the quality factor gain is higher (lower) compared to the $\alpha = 0$ case assuming $r_d = r_{d,\text{opt}}(r_\varepsilon)$.

Here and now, instead of going more into details, we stop the comparative analysis and content ourselves with the above results. The reason for that is the following. As the assumption that the different material-dependent parameters – such as ε_1, ε_2, ϕ_1 and ϕ_2 – are independent and can be selected from a continuous range is rather false, a more realistic and adequate comparison could be obtained if the analysis was carried out only over a specific set of metallic and insulating material combinations that are likely to be used in the technology. In that case, the design problem introduced at the beginning of this chapter is to be solved for both types of diodes and for each possible material combination. Then the material combination resulting in the highest zero-bias quality factor is to be picked for each diode type and then finally these values are to be compared in order to get a better idea about the possible gain of applying the double-layer insulator structure in between metallic layers. The main problem with this approach is that the metal-insulator work function values are not exactly known and can be estimated only indirectly based on other material constants or some theoretical calculations. Hence, the best way to perform the comparison described above is to fabricate diodes for each specific metal-insulator combination and to estimate the work function values based on

DOI:10.15774/PPKE.ITK.2009.001

the measurement results. As also the quantum transport model presented in Section 2.3 and the potential profiles assumed in Sections 2.4.1 and 2.5.1 may not model the diodes accurate enough, the ultimate solution is to solve the design problem experimentally for both types of diodes and for each possible material combination.

Despite all the negatives listed in the previous paragraph, the conclusion of this section is that the above numerical results demonstrate that – for typical values of the diode parameters – the detectors equipped with MIM diodes with double-layer insulators may significantly surpass the conventional ones in terms of sensitivity.

2.6. Conclusions

The investigations in this chapter have demonstrated that nanoantenna-coupled MIM diode detectors with double-layer insulator diodes have the potential to considerably outperform their conventional ‘counterparts’ in terms of sensitivity performance. As a consequence, MIM diodes with double-layer insulators – along with other performance enhancement solutions such as semi-spherical silicon microlens – may be important technology enablers for future, high-speed, multi-spectral THz and infrared camera applications. The conclusions drawn here are expected to stimulate experimental research activities on MIM diodes with double-layer insulators in the future.

The simulation curves presented in this chapter may serve as valuable guidelines when designing and fabricating MIM diodes with double-layer insulators. The investigations have – among others – shown that a moderate dielectric constant ratio results in a moderate optimum thickness ratio for the insulators, which can easily be realized even if the total thickness is only a few nanometers. When, on the other hand, the ratio of the dielectric constants takes an extreme value, the implementation of the optimum thickness ratio is no longer feasible and it is a reasonable solution to fabricate the insulators with equal thicknesses.

Finally, it has been pointed out at several occasions that the application of analytical approximations to the original tunneling theory model may introduce not only quantitative but also qualitative errors concerning the different nonlinearity properties of metal-insulator-metal devices. As a result, one should consider the employment of

DOI:10.15774/PPKE.ITK.2009.001

analytical approximations with caution and the validity of those must always be verified numerically.

The investigations concerning MIM diodes with double-layer insulators have been restricted to the case of metallic electrodes of identical material in this study. As fabricating the electrodes from different metals may result in a further performance improvement, the analysis of this more general case can be pointed out as a future research direction.

2.7. Applications

The high-sensitivity nanoantenna-coupled MIM diode detectors with double-layer insulator diodes investigated in this chapter may serve as essential components of high-speed THz and infrared visual systems in the future.These systems can primarily be applied to identify and to discriminate fast moving objects as well as to observe fast chemical and biological processes.

DOI:10.15774/PPKE.ITK.2009.001

Chapter Three

OPTIMAL SATISTICAL ENERGY BALANCING PROTOCOLS FOR WIRELESS SENSOR NETWORKS

3.1. Introduction

Due to the recent advances in electronics and wireless communication, the development of low-cost, low-power, multifunctional sensor devices have received increasing attention [24]. These nodes are compact in size and besides sensing they have limited signal processing, communication and storing capabilities. However, the limitation in size and energy makes wireless sensor networks (WSN) different from other wireless and ad hoc networks. As a result, new protocols must be developed with special focus on energy balancing in order to increase the lifetime of the network which is crucial in the case of applications where recharging of the nodes is not an option. (For such applications, please refer to Section 7.3).

Open problems

As already mentioned in the introduction, energy balance between the elements of a wireless sensor network can be achieved – among others – by properly planning the routing or packet forwarding mechanisms applied in the network. The problem of designing energy aware network layer protocols for wireless sensor networks has only been addressed for deterministic traffic models (time-driven applications) so far – e.g. in [22][23][25][26]. However, the way sensed data is generated in event-driven wireless sensor network applications – such as target tracking applications – can only be described by stochastic traffic models. In this chapter of the dissertation, energy balancing packet forwarding protocols are developed for such applications. The protocols under investigation are of stochastic nature as the target node for a packet to be forwarded is selected randomly. The main focus of attention in this study is the optimization of the underlying probability distributions – based on which the target node is assigned – in order to achieve maximum network lifetime.

The remainder of this chapter is organized as follows. In Section 3.2, the communication model of the wireless sensor network under investigation is set up. Based on this model,

DOI:10.15774/PPKE.ITK.2009.001

analytical approximations are derived for the lifetime of the network in Section 3.3. The way simulated annealing is applied to optimize the different packet forwarding protocols is described in Section 3.4. In Section 3.5, the performance of the various protocols is evaluated numerically using the formulas obtained in Section 3.3. Finally, conclusions are drawn in Section 3.6.

3.2. Communication model

In this section, the communication model of the network is described by means of the OSI seven layer model.

3.2.1. Physical layer

Radio propagation model

We model the large-scale mean path loss using the COST 231 Walfish-Ikegami model [27]. For typical parameter values and for a carrier frequency of 433MHz, the resulting expression of the large-scale mean path loss takes the form of

$$a_{\text{large-scale}}^{[\text{dB}]} = 148.3 + 22.4 d^{[\text{km}]} + 39.67 \lg d^{[\text{km}]}, \tag{3.1}$$

where d is the distance between the transmitter and the receiver.

The effects of shadow- and Rayleigh fading [28] are taken into account by means of fading margins. The distribution of the fading factor corresponding to Rayleigh fading (a_{Rayleigh}) is exponential with unit mean, while the fading factor corresponding to shadow fading (a_{shadow}) is assumed to be lognormally distributed. The probability density functions of the fading factors are given by

$$h_{a_{\text{Rayleigh}}}(x) = e^{-x} \tag{3.2}$$

and

DOI:10.15774/PPKE.ITK.2009.001

$$h_{a_{\text{shadow}}^{[\text{dB}]}}(x)=\frac{1}{\sqrt{2\pi}\sigma^{[\text{dB}]}}e^{-\frac{x^2}{2\left(\sigma^{[\text{dB}]}\right)^2}}, \tag{3.3}$$

respectively, where $\sigma^{[\text{dB}]}$ is the standard deviation of $a_{\text{shadow}}^{[\text{dB}]}$, i.e. the fading factor corresponding to shadow fading expressed in dB. A fading margin (δ) is defined as a constant that needs be added to the fading factor expressed in dB in order to have the sum lower than 0dB (1) only with a certain low probability, the outage probability (P_{out}). In other words, it needs to be added to the mean path loss in order to have the actual path loss higher than the mean path loss only with probability P_{out}. It easy to see that the equations from which the margins can be obtained, are

$$P_{\text{out}}=P\left\{a_{\text{Rayleigh}}^{[\text{dB}]}+\delta_{\text{Rayleigh}}^{[\text{dB}]}<0\text{dB}\right\}=1-e^{-10^{-\frac{\delta_{\text{Rayleigh}}^{[\text{dB}]}}{10}}} \tag{3.4}$$

and

$$P_{\text{out}}=P\left\{a_{\text{shadow}}^{[\text{dB}]}+\delta_{\text{shadow}}^{[\text{dB}]}<0\text{dB}\right\}=\frac{1}{2}\left[1+\text{erf}\left(-\frac{\delta_{\text{shadow}}^{[\text{dB}]}}{\sqrt{2}\sigma^{[\text{dB}]}}\right)\right], \tag{3.5}$$

respectively. Solving (3.4) and (3.5) for $P_{\text{out}}=16\%$ and $P_{\text{out}}=5\%$, we obtain $\delta_{\text{Rayleigh}}^{[\text{dB}]}=7.6\text{dB}$ and $\delta_{\text{shadow}}^{[\text{dB}]}=6.8\text{dB}$. (Here we assumed that $\sigma^{[\text{dB}]}=4\text{dB}$.)

Finally, if we compensate for the different fading effects, i.e. we add the margins to the mean path loss, the expression of the path loss takes the form of

$$a^{[\text{dB}]}(d)=162.7+22.4d^{[\text{km}]}+39.67\lg d^{[\text{km}]}. \tag{3.6}$$

DOI:10.15774/PPKE.ITK.2009.001

Topology model

In this study, we assume a one-dimensional (static) network topology. (After some routing protocol, e.g. the PEDAP protocol [23] has identified a path from a node to the BS, the nodes found along this path can be regarded to form a 1D topology.) In other words, the network elements are assumed to be located along a straight line with the BS at one of the ends (Figure 3.1). The sensor nodes are numbered according to their distance form the BS with node 1 and node N being the closest and farthest one, respectively. (Variable N denotes the number of sensor nodes.) The BS is numbered with index 0. The immediate neighbor of a sensor node towards and away from the BS is called its nearside and farside neighbor, respectively. Similarly, the nearside and farside neighbors of a sensor node are defined as the set of network elements that are located closer and farther away from the BS relative to the node in question. Please note that – according to the above definition – the base station is also included in the set of nearside neighbors for any of the sensor nodes.

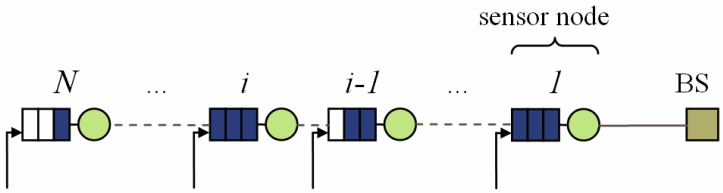

Figure 3.1. Network topology.

Finally, it is noteworthy that though the investigations in this chapter are limited to a one-dimensional topology, the analysis can be extended to higher dimensions. The way this extension can be carried out will be presented in Section 3.2.4.

Energy consumption model

We apply the power consumption model used in [29]. According to this model, the power consumed at transmission (P_{t}) and at reception (P_{r}) is expressed as

$$P_{\mathrm{t}}\left(P_{\mathrm{tr}}\right)=\eta_{\mathrm{PA}}^{-1}P_{\mathrm{tr}}+P_{\mathrm{circ,t}}, \tag{3.7}$$

and

DOI:10.15774/PPKE.ITK.2009.001

$$P_{\mathrm{r}} = P_{\mathrm{circ,r}}\,, \tag{3.8}$$

respectively. Here η_{PA} is the efficiency of the radio frequency power amplifier RF PA of the transceiver, P_{tr} is the transmission power, while $P_{\mathrm{circ,t}}$ and $P_{\mathrm{circ,r}}$ are the circuit power consumptions at transmission and at reception, respectively. (Circuit power consumption is made up of the consumption of the various additional components of the transceiver such as the mixer, the frequency synthesizer, the low noise amplifier (LNA), the intermediate frequency amplifier (IFA), etc.)

3.2.2. Link layer

The purpose of this section is to establish a relationship between the location of the nodes and the energy associated with forwarding a packet form one node to the other.

In terms of error control, we assume a threshold link model [30], i.e. packet reception is considered successful if the received power level is higher than a given value, namely, the sensitivity of the receiver (P_{sens}). Let $d_{ih_i(j)}$ denote the distance between node i and node j ($i = 1,...,N$, $j = 0,...,N$, $i \neq j$), where

$$h_i(j) = \begin{cases} j+1 \text{ if } \; j < i, \\ j \text{ if } \; j > i, \\ \text{n/a otherwise.} \end{cases} \tag{3.9}$$

(The first option in (3.9) corresponds to the case of the nearside neighbors, while the second to that of the farside neighbors.) Then the energy required to convey a packet from node i to node j ($G_{ih_i(j)}$) can be expressed as

$$G_{ih_i(j)} = \Delta t \left[\eta_{\mathrm{PA}}^{-1} \cdot 10^{\frac{a^{[\mathrm{dB}]}\left(d_{ih_i(j)}\right)}{10}} \cdot P_{\mathrm{sens}} + P_{\mathrm{circ,t}} \right], \tag{3.10}$$

where Δt is the duration of a packet. As implied in (3.10), we consider only the energy consumption associated with packet transmission in this study. Nevertheless, in Section

DOI:10.15774/PPKE.ITK.2009.001

3.3, we are going to give a hint on how the energy consumption associated with packet reception can be incorporated into the analysis.

3.2.3. Application layer

As already pointed out above, we assume an event-driven application in our model. According to this assumption, sensed data is generated at the nodes at random time instances. In other words, we apply a stochastic traffic model.

In the discrete time case, it is assumed that sensed data, i.e. packets are generated at node i according to a binomial process with parameter p_i. The binomial process as an arrival process can be described in two different ways: in terms of the number of arrivals (packet generations) (n_i) within a given time interval (K) and the time interval between consecutive arrivals (packet generations) (K_i). The former quantity is distributed according to a binomial distribution with parameters (p_i, K) (hence the name of the process), while the latter is geometrically distributed with parameter p_i. These distributions are given by

$$P\{n_i(K) = j\} = p_i^j (1-p_i)^{K-j}, \ j = 0,...,K \tag{3.11}$$

and

$$P\{K_i = j\} = p_i (1-p_i)^{j-1}, \ j = 1,2,..., \tag{3.12}$$

respectively. It is assumed that the binomial processes corresponding to the different nodes are statistically independent and as a result, y_i, the number of packets generated at node i at a given time instant is distributed according to

$$P\{\mathbf{y} = \mathbf{c}\} = \prod_{i=1}^{N} p_i^{c_i} (1-p_i)^{1-c_i}, \ c_i \in \{0,1\}, \ i = 1,...,N. \tag{3.13}$$

In other words, $y_i = n_i(1)$. (Remark that y_i is distributed according to a Bernoulli distribution with parameter p_i, which corresponds to an on/off traffic model.)

DOI:10.15774/PPKE.ITK.2009.001

In the continuous time case, it is assumed that packets are generated at node i according to a Poisson process with parameter λ_i. The Poisson process as an arrival process – just like the binomial process – can be described in two different ways: by means of the number of arrivals (n_i) within a given time interval (T) and the time interval between consecutive arrivals (T_i). The former quantity is distributed according to a Poisson distribution with parameter $\lambda_i T$ (hence the name of the process), while the latter is exponentially distributed with parameter λ_i. These distributions are given by

$$P\{n_i(K)=j\}=\frac{(\lambda_i T)^j}{j!}e^{-\lambda_i T},\ j=0,1,... \tag{3.14}$$

and

$$P\{T_i=t\}=\lambda_i e^{-\lambda_i t}, \tag{3.15}$$

respectively. Similarly to the discrete time case, it is assumed that the Poisson processes corresponding to the different nodes are statistically independent.

3.2.4. Network layer

As already mentioned above, network elements may forward their sensed data (i.e. their own packets) to the BS either directly or indirectly with the assistance of other network elements. This inherently implies that in addition to their own packets, nodes need to store and forward also the packets that were received from other nodes. The way a target node is selected at packet forwarding is dictated by the packet forwarding protocol, while the queuing mechanism determines how packets are stored in the queue and at what time instants they are forwarded.

Packet forwarding protocols

In this study, we investigate four different packet forwarding protocols.

DOI:10.15774/PPKE.ITK.2009.001

The first two protocols, namely the chain and shortcut (or single-hop) protocols serve as reference protocols in the forthcoming comparative performance analysis.

In the case of the chain protocol, each of the nodes forwards its packet to its nearside neighbor. This implies that a node may receive a packet only from its farside neighbor. The advantage of this protocol is that minimum energy is used for packet forwarding as only the closest nearside neighbor may be selected as a target node. An obvious drawback, however, is that in addition to its own packets, a node needs to forward all the packets that were generated by its farside neighbors. Let us assume an equidistant topology, uniform traffic pattern and equal initial energies. It is easy to see that, in this case, the closer a node to the BS station is, the faster it is depleted. The bottleneck node, i.e. the node that 'dies' first is the one that is closest to the BS (i.e. node 1). Obviously, in order to extend network lifetime, a solution needs to be found that makes it possible for a node to decrease the load of its nearside neighbors. (In this study, we consider the network 'alive' if and only if all of its elements are operational. A detailed description of network lifetime is given is Section 3.3.)

When the shortcut protocol is applied as the packet forwarding strategy, each node forwards its packets directly to the BS. A benefit of the protocol is that nodes do not overload their nearside neighbors, i.e. they do not use the energy of other nodes when forwarding their own packets to the BS. The energy consumption of the different nodes, on the other hand, is not balanced as nodes farther away from the BS consume more energy – per packet – than the closer ones. Assuming an equidistant topology, uniform traffic pattern and equal initial energies again, the father a node is located from the BS, the faster it is exhausted. This time the node that is the farthest away from the BS (i.e. node N) is the bottleneck node. Apparently, in order to exceed the network lifetime provided by the shortcut protocol, it needs to be upgraded in such way that nodes are enabled to forward packets also to their nearside neighbors (other than the BS).

The 'random shortcut' protocol [26] is a compromise between the chain and shortcut protocols. In the case of this protocol, a node may decide to forward its packet either to its nearside neighbor or to the BS. The decision is based on a 'loaded' coin flip. If the result of the toss is heads than the packet is forwarded to the nearside neighbor, while if is tails the packet is sent to the BS directly. Formally speaking, for node i, the nearside

DOI:10.15774/PPKE.ITK.2009.001

neighbor and the BS are selected as the target node with probability a_i and $1-a_i$, respectively.

It is clear that – in contrast to the previous two protocols – nodes operating according to the 'random shortcut' protocol have the ability to off-load both their nearside and farside neighbors. In fact, the 'random shortcut' protocol is a generalization of the chain and shortcut protocols. If $\forall i \;\; a_i = 1$ we get back the former protocol, whereas if $\forall i \;\; a_i = 0$ we obtain the latter one. As a consequence, the lifetime that can be achieved using the 'random shortcut' protocol cannot be lower than that obtained in case of the other two. Nevertheless, in order to make the most of the 'random shortcut' protocol, vector $\mathbf{a} = (a_1,...,a_N)$ needs to be optimized. Protocol optimization – including the definition of the cost function and the method of the optimization – will be treated in Section 3.3 and Section 3.4. If vector $\mathbf{a}$ is optimized, the energy consumption of the different nodes is completely balanced and as result, they go flat at the same time maximizing network lifetime. (Please remark that the previous statement is not necessarily true if either the topology is not equidistant, the traffic pattern is not uniform or if the initial energies of the different nodes are not equal. Nevertheless, complete balance is still can hold true even if we allow some variations with respect to the above parameters.)

The soundness of the above assertion can be justified with the help of the following (indirect) reasoning. The reasoning is based on the assumption that G_{ij} is a strictly monotonically increasing function of d_{ij}. Let us assume that the lifetimes of the different nodes are not equal for $\mathbf{a}_{\text{opt}}$ and that node i is the one with the lowest lifetime (indirect assumption of the proof). If ($i > 1$ and) $a_i < 1$ (i.e. if some of the packets of node i is directly forwarded to the BS), the lifetime of node i can be increased by increasing a_i (i.e. by forwarding a part of these packets is rather to node $i-1$). When doing so, the lifetime of the farside neighbors of node i is left unchanged, while that of its nearside neighbors decreases. Let us increase a_i in such a way that the lifetime of node i becomes equal to the lifetime of (at least) one of the other nodes. If this does not take place before the value of a_i reaches 1, let us set a_i to 1. It can be readily justified that network lifetime is improved in both cases, which contradicts our indirect assumption. If

DOI:10.15774/PPKE.ITK.2009.001

($i < N$ and) $a_{i+1} > 0$ (i.e. if some of the packets of node $i+1$ is forwarded to node i), the lifetime of node i can be increased by decreasing a_{i+1} (i.e. by forwarding a part of these packets rather directly to the BS). If we do so, the lifespan of the nearside neighbors of node $i+1$ increases, the lifetime of node $i+1$ decreases, whereas that of its farside neighbors remains the same. Let us decrease a_{i+1} in such a way that the lifetime of node i becomes equal to the lifetime of (at least) one of the other nodes. If this does not occur before the value of a_{i+1} reaches 0, let us set a_{i+1} to 0. The lifespan of the network is prolonged in both cases, which is again in contradiction with the indirect assumption of the proof. Finally, if ($i = 1$ or) $a_i = 1$ and ($i = N$ or) $a_{i+1} = 0$, node i does not receive any packets from other nodes and all the packets generated at node i are forwarded to its nearside neighbor. This – assuming equidistant topology, uniform traffic pattern and equal initial energies – implies that the lifetime of node i is equal to the maximum lifetime that can be reached in the network, which contradicts our indirect assumption yet again.

Figure 3.2 visualizes the operation of the 'random shortcut' protocol. The probability that the packet generated at node j is forwarded along the route shown in the figure can be expressed as

$$P = (1 - a_i) \prod_{\ell=j}^{i-1} a_\ell \,. \tag{3.16}$$

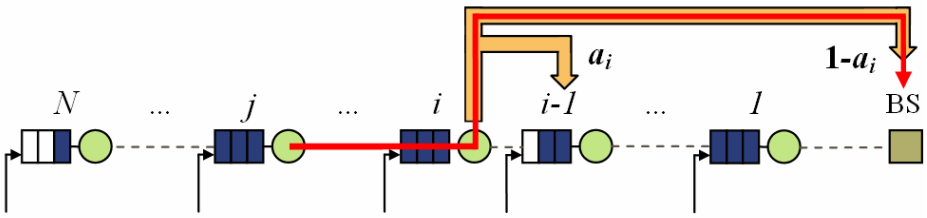

Figure 3.2. The operation of the 'random shortcut' protocol.

Generalizing the 'random shortcut' protocol, one can develop the 'sending packets to any node ahead in the chain' protocol. In the case of this protocol, a node may forward its packet to any of its nearside neighbors. (In other words, a node may receive packets from any of its farside neighbors.) This time the decision is made using a loaded dice. Please

DOI:10.15774/PPKE.ITK.2009.001

note that the dice used here is a generalized dice as it may have more or fewer than six sides. The probability that node i sends its packet to node $j-1$ ($i > j$) is denoted by A_{ij}.

As the 'sending packets to any node ahead in the chain' protocol is a generalization of the 'random shortcut' protocol, the maximum lifetime that can be achieved applying the former protocol cannot be lower than that using the latter one. (If $\forall i, j,\ j \neq 1, i\ \ A_{ii} = a_i$, $A_{i1} = 1 - a_i$ and $A_{ij} = 0$, we get back the 'random shortcut' protocol.) In order to make the best of the 'sending packets to any node ahead in the chain' protocol and achieve maximum network lifetime – similarly to vector $\mathbf{a}$ – also matrix $\mathbf{A}$ need to be optimized. (The assertion that for $\mathbf{A}_{\text{opt}}$, the energy consumption of the different nodes – assuming equidistant topology, uniform traffic pattern and equal initial energies – is completely balanced, can be proven the same way it was done for the 'random shortcut' protocol.)

Though both in the case of the 'random shortcut' and 'sending packets to any node ahead in the chain' protocol, nodes have the ability to off-load both their nearside and farside neighbors, the latter protocol may perform potentially better than the former one. The reason for that lies in the following. In both cases, the number of quantities to be balanced is N. However, when it comes to the number of independent parameters and the number of possible routes in the network, we find a major difference. While for the former protocol, the number of independent parameters is $N-1$, for the latter one, it is $N(N-1)/2$. Concerning the number of the possible (different) routes from node i to the BS, the figures are i and 2^{i-1}, respectively. As a consequence – due to the significant difference in the number of independent parameters and the number of possible routes to the BS – one can reasonably expect that the 'sending packets to any node ahead in the chain' protocol may considerably outperform the 'random shortcut' protocol in terms of network lifetime.

As already mentioned before, the above packet forwarding protocols can be extended to higher spatial dimensions. In order to do so, the concept of nearside neighbors needs to be defined also for higher-dimensional network topologies. Let us consider a spanning tree of the complete graph made up by the nodes as vertices. (The way the spanning tree

DOI:10.15774/PPKE.ITK.2009.001

is constructed is beyond the scope of this study. A possible alternative, however, is to apply the spanning tree algorithm used – as an example – in the PEDAP routing protocol [23]) Then the set of nearside neighbors of a node is defined as the set of vertices that are found along the path (open walk) that connects the node and the BS in the spanning tree. (The BS is included in this set, while the node in question is excluded from it.) The nearside neighbor of a node is defined as the node among its nearside neighbors that is at distance 1 from it. With these definitions in mind, the extension of the protocols to higher dimensions is very much straightforward. Selecting the optimum spanning tree, however, is less obvious and thus is an open research problem.

In the case of the most general packet forwarding protocol that can be thought of, nodes may forward their packets to any of the other nodes. (The probability that node i forwards its packet to node j is equal to $A_{ih_i(j)}$ in this case.) An advantage of this solution is that selecting the optimum spanning tree is no longer an issue. The optimization of matrix $\mathbf{A}$, however, gets a lot more complicated as the matrix becomes much less sparse compared to the case of the 'sending packets to any node ahead in the chain' protocol.

Finally, in order to be able to treat the above four protocols on the basis of the same principle in the forthcoming analysis, let us formulate matrix $\mathbf{A}$ also for the first two protocols. The expression of the elements of $\mathbf{A}$ is given by

$$A_{ii}=1,\ A_{ij}=0,\ i,j=1,...,N,\ i\neq j \tag{3.17}$$

and

$$A_{i1}=1,\ A_{ij}=0,\ i=1,...,N,\ \ j=2,...,N. \tag{3.18}$$

for the chain and shortcut protocols, respectively.

Queuing model

In our model, the network layer may receive packets from both the application layer and the link layer. In the former case, incoming packets contain sensed data measured by

DOI:10.15774/PPKE.ITK.2009.001

node itself, while in the latter one, packets are received from other nodes in the network. In both cases, packets are stored in a buffer or in other words, in a queue.

Packets in the queue are forwarded on a first come first served (FCFS) basis.

The operation of the queue of node i in the discrete time case is expressed by the queuing equation as follows.

$$q_i(k+1) = \min\left\{\lceil q_i(k)-1\rceil^+ + z_i(k) + y_i(k), L_i\right\}, \tag{3.19}$$

where $q_i(k)$ is the number packets in the queue of node i at time instant k, $z_i(k)$ is the number of packets that node i receives from other nodes at time instant k, $y_i(k)$ is the number packets generated at node i at time instant k, L_i is the length of the queue/buffer of node i, while

$$\lceil x\rceil^+ = \begin{cases} x \text{ if } x \geq 0, \\ 0 \text{ otherwise.} \end{cases} \tag{3.20}$$

To put it into words, at each time instant, first the number of packets in the queue is decreased by one provided that the queue is not empty (i.e. at most one packet is forwarded at a time) and afterwards the number of packets in the queue is increased by the sum of the number of packets generated at the node and the number of the packets received from other nodes. Finally, if the resulting number is higher than the size of the buffer/queue, the packets that cannot fit anymore into the buffer are simply dropped. The model of the discrete time queue is depicted in Figure 3.3.

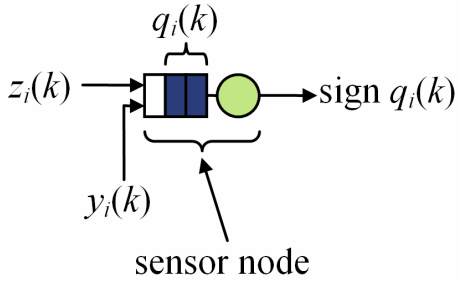

Figure 3.3. The discrete time queuing model.

DOI:10.15774/PPKE.ITK.2009.001

In the continuous time case, the inter-forwarding time intervals, i.e. the time intervals between the departure of two consecutive packets are independent, exponentially distributed random variables with parameter μ_i. (In the definition of inter-forwarding time interval, it is assumed that the queue is not empty within the interval). (As already seen above, in the discrete time case, the inter-forwarding time intervals are deterministic and are equal to 1).

3.3. Network lifetime

As both the application and routing layers are of random nature in our model, a statistical interpretation of network lifetime is required. In this section, first we give the definition of network lifetime and then derive its analytical expression. In both the discrete and continuous time model, our approach to calculate network lifetime is based on the assumption that the queuing network – made up by the interconnected nodes – is in stationary state.

3.3.1. Discrete time model

Let $\boldsymbol{\pi}^{(i)}$ denote the stationary distribution of the number packets in the queue of node i, i.e.

$$\pi_j^{(i)} = \lim_{k\to\infty} P\{q_i(k) = j\}. \tag{3.21}$$

If (3.21) exists, it can be calculated by solving the following eigenvalue problem.

$$\mathbf{P}^{(i)}\boldsymbol{\pi}^{(i)} = 1\cdot\boldsymbol{\pi}^{(i)}, \tag{3.22}$$

where

$$P_{nm}^{(i)} = \lim_{k\to\infty} P\{q_i(k+1) = n \,|\, q_i(k) = m\}.$$

Now let us determine the expression of $\mathbf{P}^{(i)}$. It is easy to see that the probability that node i receives a packet from node j $(i \neq j)$ is $A_{jh_j(i)}\left(1-\pi_0^{(j)}\right)$ in stationary state. As a result,

DOI:10.15774/PPKE.ITK.2009.001

$$P_b^{(i)} = \lim_{k\to\infty} P\{z_i(k)=b\} = \prod_{\substack{j=1,\ldots,N;\ j\neq i \\ b_j=0,1;\ \sum_{\substack{j=1\\ j\neq i}}^{N} b_j = b}} \left[A_{jh_j(i)}\left(1-\pi_0^{(j)}\right)\right]^{b_j} \left[1-A_{jh_j(i)}\left(1-\pi_0^{(j)}\right)\right]^{1-b_j} . \tag{3.23}$$

(It is noteworthy that in (3.23) we assumed that, in stationary state, the joint probability density function of (q_i) is of product form, i.e. that variables q_i are statistically independent. An interesting property of queuing networks with product form solution is that the network nodes behave as independent queuing systems, although in general they are not independent – see e.g. [33]. Sufficient conditions for a product form solution for discrete time queuing networks with batch arrivals and batch departures are given in [31] and [32].) By convolving $\lim_{k\to\infty} z_i(k)$ and $y_i(k)$ (see Section 3.2.3), finally, we obtain that

$$P_{nm}^{(i)} = \begin{cases} \sum_{\substack{b=0,1,\ldots,N-1;\ c=0,1 \\ b+c=n-m+\mathrm{sign}(m)}} P_b^{(i)} p_i^c (1-p_i)^{1-c} & \text{if } n < L_i, \\ \sum_{\substack{b=0,1,\ldots,N-1;\ c=0,1 \\ b+c\geq L_i-m+\mathrm{sign}(m)}} P_b^{(i)} p_i^c (1-p_i)^{1-c} & \text{if } n = L_i, \\ 0 \text{ otherwise.} \end{cases} \tag{3.24}$$

(The above discrete time queuing network is in fact a complex of coupled, discrete time Markov chains. The number of packets in a queue is equivalent to the state of the corresponding Markov chain. The transition matrices of the different Markov chains are given by (3.24). Coupling between the chains is established as a consequence of the packet exchange taking place between the nodes. It can be also noted that the queues of the nodes operate as $G/D/1$ queues as their service times are deterministic $(=1)$ and their arrival processes are general.) The mean of the energy consumed by node i upon forwarding a packet can be calculated as

$$\tilde{g}_i = \sum_{\substack{j=1\\ j\neq i}}^{N} A_{ih_i(j)} G_{ih_i(j)} . \tag{3.25}$$

DOI:10.15774/PPKE.ITK.2009.001

Assuming a binomial departure processes at the nodes, the probability that node i forwards n packets in time K is

$$P_{n,K}^{(i)} = \binom{K}{n} \left(1-\pi_0^{(i)}\right)^n \left(\pi_0^{(i)}\right)^{K-n}. \tag{3.26}$$

Approximating the energy consumed at packet forwarding with its mean, the probability that node i does not go flat after time K is given as

$$P_{\text{disc.}}\{\text{node } i \text{ does not go flat in time } K\} = \sum_{n=0}^{[C/\tilde{g}_i]} \binom{K}{n} \left(1-\pi_0^{(i)}\right)^n \left(\pi_0^{(i)}\right)^{K-n}, \tag{3.27}$$

where C in the initial energy of the nodes. In this study, we consider the network 'alive' if all of its elements are operational. Hence the probability that the network is no longer 'alive' after time K can be expressed as

$$P_{\text{disc.}}\{\text{the network is 'dead' after time } K\} = 1 - \prod_{i=1}^{N} P_{\text{disc}}\{\text{node } i \text{ does not go flat in time } K\}. \tag{3.28}$$

Please note that similarly to (3.23), also (3.28) is based on the assumption that the queues of the different nodes are statistically independent. The lifetime of the network (Ψ) is finally defined as the shortest time after which the probability that the network is 'dead' is equal to or higher than $e^{-\alpha}$, i.e.

$$\Psi_{\text{disc.}}(\mathbf{A}) = \min_K \left[1 - \prod_{i=1}^{N} \sum_{n=0}^{[C/\tilde{g}_i(\mathbf{A})]} \binom{K}{n} \left(1-\pi_0^{(i)}(\mathbf{A})\right)^n \left(\pi_0^{(i)}(\mathbf{A})\right)^{K-n} \geq e^{-\alpha} \right], \tag{3.29}$$

where $e^{-\alpha} \approx 1$. As both $\tilde{g}_i$ and $\pi_0^{(i)}$ depend on matrix $\mathbf{A}$, protocol optimization reduces to finding the maximum of $\Psi_{\text{disc}}(\mathbf{A})$ with respect to $\mathbf{A}$, that is to say,

$$\mathbf{A}_{\text{opt}} = \arg\max_{\mathbf{A}} \Psi_{\text{disc.}}(\mathbf{A}). \tag{3.30}$$

DOI:10.15774/PPKE.ITK.2009.001

In addition, we place the following two reliability constraints on (3.30): (i) the maximum of ratio of the number of dropped packets relative to the number of total received packets must no exceed a certain threshold ($e^{-\gamma}$); and (ii) the mean packet delay should remain lower than a predefined value (β). The constraint on the packet loss can be expressed as

$$l = \max_{i=1,\ldots,N} \frac{\sum_{n>0} n\left[\sum_{j=0,\ldots,L_i;c=0,1} \pi_j^{(i)}(\mathbf{A}) P^{(i)}_{L_i-j+\mathrm{sign}(j)-c+n}(\mathbf{A}) p_i^c (1-p_i)^{1-c}\right]}{\sum_{\substack{b=0,\ldots,N-1;c=0,1 \\ b+c>0}} (b+c) P_b^{(i)}(\mathbf{A}) p_i^c (1-p_i)^{1-c}} < e^{-\gamma}, \tag{3.31}$$

where the denominator on the left-hand side of (3.31) is the mean of the number of packets received by node i, whereas the numerator is equal to the mean of number of packets drooped by node i. (The probability that a number of n packets are dropped by node i is calculated with the help of the total probability theorem.) Finally, according to Little's theorem [33], the constraint on the mean packet delay has the form of

$$\Delta = \frac{\sum_{i=1}^{N}\sum_{j=0}^{L_i} j\pi_j^{(i)}(\mathbf{A})}{\sum_{i=1}^{N} p_i} < \beta, \tag{3.32}$$

where the denominator on the left-hand side of (3.32) is the mean of the number of packets arriving into the queuing system per unit time, while the numerator is equal to the mean of the number of packets in the queuing system. Here the expression'queuing system' refers to the queuing network made up by the interconnected nodes.

As constraints (3.31) and (3.32) are rather hard to incorporate into the solution representation of the optimization problem in (3.30), we apply penalty functions [34] in order to take (3.31) and (3.32) into account. Namely, whenever one of the constraints is not met, the value of $\Psi(\mathbf{A})$ is decreased by 25% (of its original value) for each of the constraints not fulfilled.

Finally, we present how energy consumption at packet reception can be incorporated into the above treatment of network lifetime. As, in contrast to packets generated at a node,

DOI:10.15774/PPKE.ITK.2009.001

packets to be relayed by the node need not only to be forwarded but also to be received, the energy consumed by node i when relaying a packet to node j has the form of

$$G'_{ih_i(j)} = G_{ih_i(j)} + P_{\text{circ,r}} \cdot \Delta t \,. \tag{3.33}$$

The probability that a packet in queue i is received from another node ($P_{\text{relay}}^{(\text{i})}$) can be approximated as the ratio of the mean number of packets received from other nodes to the mean number of the packets to be forwarded, i.e.

$$P_{\text{relay}}^{(\text{i})} = \frac{\sum_{\substack{j=1 \\ j\neq i}}^{N} A_{jh_j(i)}\left[1-\pi_0^{(j)}\right]}{1-\pi_0^{(i)}}. \tag{3.34}$$

As a result, in order to be able to calculate network lifetime using the same expressions as above, the formula for $\tilde{g}_i$ (3.26) has to be accordingly rewritten as

$$\tilde{g}_i = \sum_{\substack{j=1 \\ j\neq i}}^{N} \left[\left(1-P_{\text{relay}}^{(\text{i})}\right)A_{ih_i(j)}G_{ih_i(j)} + P_{\text{relay}}^{(\text{i})}A_{ih_i(j)}G'_{ih_i(j)}\right]. \tag{3.35}$$

3.3.2. Continuous time model

According to Jackson's theorem [33], the joint stationary distribution of the number of packets waiting in the queue of the nodes – constituting the continuous time queuing network described in Section 3.2.4 – is of product form. The stationary distribution of queue i will be identical to that of a $M/M/1$ queue with an arrival intensity of Λ_i, where (Λ_i) is a solution to the system of equations given by

$$\Lambda_i = \lambda_i + \sum_{j=1}^{N} A_{jh_j(i)}\Lambda_j \,. \tag{3.36}$$

(Please not that also condition $\mu_i^{-1} < \Lambda_i$ needs to be fulfilled in order for Jackson's theorem to be applicable). It is easy to see that if the queuing network is feed-forward –

DOI:10.15774/PPKE.ITK.2009.001

as in the case of the protocols introduced in Section 3.2.4 – the departure processes of the queues are in addition Poissonian (Burke's theorem [33]). As the queue of the node i is in statistical equilibrium, the intensity of its departure process is equal to that of its arrival process. According to the above, the probability that node i forwards n packets in time t is expressed as

$$P_{n,t}^{(i)} = \frac{(\Lambda_i t)^n}{n!} e^{-\Lambda_i t}. \tag{3.37}$$

Approximating the energy consumed at packet forwarding with its mean again (3.25), the probability that node i does not go flat after time t is given as

$$P_{\text{cont.}}\{\text{node } i \text{ does not go flat in time } t\} = \sum_{n=0}^{[C/\tilde{g}_i]} \frac{(\Lambda_i t)^n}{n!} e^{-\Lambda_i t}. \tag{3.38}$$

Applying the same definition of network lifetime (Ψ) as in Section 3.3.1, we have

$$P_{\text{cont.}}\{\text{the network is 'dead' after time } t\} = 1 - \prod_{i=1}^{N} P_{\text{cont}}\{\text{node } i \text{ does not go flat in time } t\}. \tag{3.39}$$

and

$$\Psi_{\text{cont.}}(\mathbf{A}) = \min_t \left[1 - \prod_{i=1}^{N} \sum_{n=0}^{[C/\tilde{g}_i(\mathbf{A})]} \frac{(\Lambda_i(\mathbf{A})t)^n}{n!} e^{-\Lambda_i(\mathbf{A})t} \geq e^{-\alpha} \right], \tag{3.40}$$

Once again, as both $\tilde{g}_i$ and Λ_i depend on matrix $\mathbf{A}$, protocol optimization can take place by maximizing $\Psi_{\text{cont.}}(\mathbf{A})$ with respect to $\mathbf{A}$, i.e.

$$\mathbf{A}_{\text{opt}} = \arg\max_{\mathbf{A}} \Psi_{\text{cont.}}(\mathbf{A}). \tag{3.41}$$

DOI:10.15774/PPKE.ITK.2009.001

Energy consumption at reception can be included in the calculations the same way it was done in the case of the discrete time model in Section 3.3.1. The only difference is that $P_{\text{relay}}^{(\text{i})}$ now takes the form of

$$P_{\text{relay}}^{(\text{i})} = \frac{\sum_{\substack{j=1 \\ j \neq i}}^{N} A_{jh_j(i)} \Lambda_j}{\Lambda_i}. \tag{3.42}$$

3.4. Optimization method

In practice, the optimal value of matrix $\mathbf{A}$ can be determined after measuring matrix $\mathbf{G}$ and vector $\mathbf{p}$ (or $\boldsymbol{\lambda}$) upon network deployment. The resulting matrix is then used throughout the entire lifetime of the network. As consequence, running time of the optimization algorithm to be used is not a critical issue and protocol optimization can be carried out off-line.

As optimization problems (3.30) (supplemented with constraints (3.31) and (3.32)) and (3.41) are complex nonlinear optimization problems, they can only be solved using some stochastic search algorithm such a simulated annealing [35][36]. This algorithm has the following two important advantages over other stochastic search algorithms [34]. Firstly, in contrast to other algorithms such as genetic algorithm, simulated annealing is very easy to implement, the implementation typically takes only a few hundred lines of computer code. Secondly, in the case of simulated annealing, theoretical performance of the algorithm can be studied analytically; namely, expressions for the upper bound of the error can be derived as a function of running time. These benefits were the main reason why the author decided to apply simulated annealing in the present investigations.

The purpose of this section is to describe how simulated annealing was applied to the optimization problems described above. The complexity of the resulting algorithms is assessed in Appendix A. A brief overview of the simulated annealing algorithm – along with a finite-time approximation of that – is provided for a better understanding in Appendix A.

DOI:10.15774/PPKE.ITK.2009.001

3.4.1. Application of simulated annealing

The application of the simulated annealing algorithm requires the specification of three distinct items: a concise problem representation, a transition mechanism and a cooling schedule.

Solution representation

A problem representation consists of a representation of the solutions and an expression of the cost function. As the cost functions have already been defined in Section 3.3, the present section focuses on solution representation.

The possible solutions of the optimization problems described above ((3.30) and (3.41)) are obviously subject to the constraints

$$\sum_{j=1}^{N} A_{ij} = 1 , \; i = 1,...,N . \tag{3.43}$$

and

$$0 \le A_{ij} \le 1 , \; i, j = 1,...,N . \tag{3.44}$$

In addition to (3.43) and (3.44), for the ‘random short-cut’ and ‘sending packets to any node ahead in the chain’ protocols, the solutions are further limited to

$$A_{ij} = 0 , \; j \ne 1, i , \tag{3.45}$$

and

$$A_{ij} = 0 , \; j > i , \tag{3.46}$$

respectively.

DOI:10.15774/PPKE.ITK.2009.001

As simulated annealing is a combinatorial optimization method, it is applicable only to discrete solution spaces. Therefore, since A_{ij} is continuous in value, it needs to be quantized when constructing the solution space S that the algorithm is applied to. One of the alternatives is to quantize A_{ij} – in the $[0,1]$ interval – for each value of i and j independently and to apply the constraints (3.43) by adding a penalty function to the original cost function (Solution 1). It is obvious that, in this case, the solution space is larger than necessary as it contains not only the feasible solutions of the optimization problem but also additional, infeasible solutions that do not meet (3.43). Consequently – as it is pointed out above that minimizing the size of the solution space S is of crucial importance – a different alternative is needed.

Instead of quantizing each of the non-zero A_{ij} variables for a given i independently, we quantize only the points of division that split up the $[0,1]$ interval into subintervals the lengths of which in fact correspond to the non-zero A_{ij} variables. (The number of points of division for a given i is 1 and $i-1$ in the case of the 'random short-cut' and 'sending packets to any node ahead in the chain' protocols, respectively.) In this way, constraint (3.43) is inherently included in the representation (Solution 2). In addition, since rearranging a given set of points of division does not change the value of the corresponding A_{ij} variables, it is only the number of points of division corresponding to a given quantization level what is in included in the representation (Solution 3). Please note that the size of the solution space is further reduced in this way.

The quantization we apply is equidistant with a number of $M+1$ quantization levels.

Let matrix $\mathbf{B}$ denote an element of the solution space S constructed in the manner described above. Then, for the 'random short-cut' and 'sending packets to any node ahead in the chain' protocols,

$$\sum_{v=1}^{M+1} B_{uv} = 1,\ B_{uv} \in \mathbf{Z}_0^+,\ u = 1,...,N-1,\ v = 1,...,M+1, \tag{3.47}$$

and

DOI:10.15774/PPKE.ITK.2009.001

$$\sum_{v=1}^{M+1} B_{uv} = u,\ B_{uv} \in \mathbf{Z}_0^+,\ u = 1,...,N-1,\ v = 1,...,M+1, \tag{3.48}$$

respectively. In other words, B_{uv} denotes the number of points of divisions that can be found at the v^{th} quantization level that corresponds to the $(u+1)^{\text{th}}$ column of matrix $\mathbf{A}$. (In (3.47) and (3.48), we took advantage of the fact that $A_{1j} = 1$ and $A_{1j} = 0$ for $j = 1$ and $j \neq 1$, respectively, i.e. there is no variable to be optimized in the first column of $\mathbf{A}$.)

It can easily be seen that the expression of $\mathbf{A}$ as a function of $\mathbf{B}$ is given by

$$A_{ij} = \begin{cases} \frac{1}{M}\arg\min_{w=1,...,M+1}\left[\sum_{v=1}^{w} B_{(i-1)v} = 1\right] & \text{if } \ j = 1, \\ 1 - \frac{1}{M}\arg\min_{w=1,...,M+1}\left[\sum_{v=1}^{w} B_{(i-1)v} = 1\right] & \text{if } \ j = i, \end{cases} \tag{3.49}$$

$i = 2,...,N$,

and

$$A_{ij} = \begin{cases} \frac{1}{M}\arg\min_{w=1,...,M+1}\left[\sum_{v=1}^{w} B_{(i-1)v} = 1\right] & \text{if } \ j = 1, \\ \frac{1}{M}\left\{\arg\min_{w=1,...,M+1}\left[\sum_{v=1}^{w} B_{(i-1)v} = j\right] - \arg\min_{w=1,...,M+1}\left[\sum_{v=1}^{w} B_{(i-1)v} = j-1\right]\right\} & \text{if } \ 1 < j < i, \\ 1 - \frac{1}{M}\arg\min_{w=1,...,M+1}\left[\sum_{v=1}^{w} B_{(i-1)v} = i-1\right] & \text{if } \ j = i, \end{cases} \tag{3.50}$$

$i = 2,...,N$,

$j = 1,...,i$,

in the case of the ‘random short-cut’ and ‘sending packets to any node ahead in the chain’ protocols, respectively.

Using the formula of combinations with repetitions, it can also readily be seen that the size of the solution space S can expressed as

DOI:10.15774/PPKE.ITK.2009.001

$$|S| = (M+1)^{N-1}, \tag{3.51}$$

and

$$|S| = \prod_{i=1}^{N-1} \binom{M+i}{i} < (M+1)^{\frac{N(N-1)}{2}} < (M+1)^{\frac{N(N+1)}{2}}, \tag{3.52}$$

for the two different protocols, respectively. The two expressions on the right-hand side of (3.52) equal to size of the solution space in the case of Solution 2 and Solution 3, respectively. Figure 3.4 shows the required computation time ($\ln|S|$) relative to Solution 1 when Solution 2 and 3 is applied. It can observed that, for larger values of N, the achievable reduction in $\ln|S|$ increases as N grows and it is mainly due to step between Solution 2 and Solution 3. We also find on the diagram that the achievable reduction decreases somewhat with M.

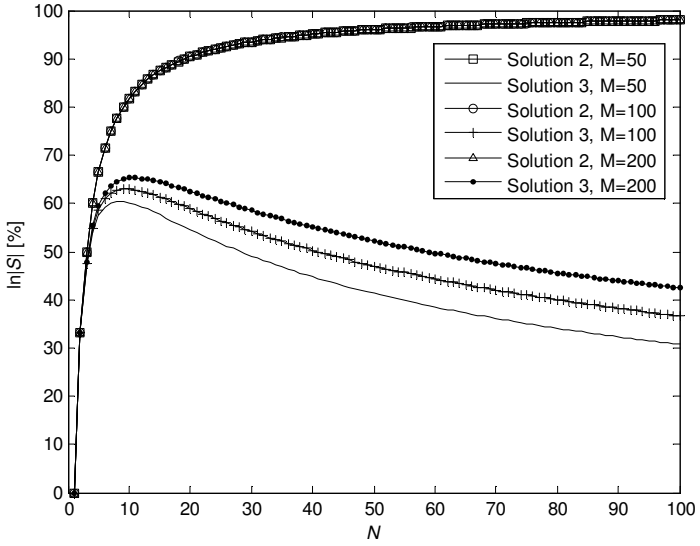

Figure 3.4. Relative computation time (ln|S|) as a function of N (number of nodes).

Finally, let us remark that the representation proposed above not only ensures minimal solution space size but it also facilitates uniformly random selection of solutions. It can readily be proven that setting the distribution of the points of division uniform over the $[0,1]$ interval is equivalent to selecting a uniformly random solution from all feasible solution that meet (3.43).

DOI:10.15774/PPKE.ITK.2009.001

Neighborhood structure and generation mechanism

A transition mechanism consists of three steps: firstly a new solution is generated from the current one according to a generation mechanism; secondly the difference in cost between the two solutions is calculated; and thirdly a decision is made on whether or not the new solution is to be accepted. Due to complicated nature of the cost functions defined in Section 3.3, the difference in cost between the two solutions cannot be determined in our case and the cost of each solution needs to be calculated separately. As, in addition, the acceptance rule has already been defined in Appendix A, this section is concerned with the neighborhood structure and generation mechanism applied.

The neighborhood structure is constructed as follows. A neighbor of a given solution is obtained if one of the points of division of the solution is moved to the previous or to the subsequent quantization level. If the point of division in question is at the first (last) quantization level and it is to be moved to the previous (subsequent) one, than it is moved to the last (first) one.

With the exception of the special cases just mentioned, neighbor generation implies a small perturbation to matrix $\mathbf{A}$ – and to the equivalent packet forwarding protocol – as only two non-zero elements are affected in one of the columns of the matrix. This in fact corresponds to a change in the packet forwarding mechanism only for a single node. (Furthermore, in the case of the 'sending packets to any node ahead in the chain' protocol, the two elements are subsequent and hence correspond to two neighboring destination nodes.) One of the two elements is increased by $\frac{1}{M}$, while the other is decreased by the same amount. However, in the special case when the point of division in question is to be moved $M+1$ quantization levels, it is the first and last non-zero element of the corresponding column what is affected. The change to the first (last) element is 0 if the point is to be moved from the first/last quantization level and the initial number of points of division at the first/last (last/first) level is at least 2 (1). Otherwise their change is limited only by ± 1.

DOI:10.15774/PPKE.ITK.2009.001

It is easy to see than the number of neighbors generated in this manner is $2(N-1)$ and $N(N-1)$ for the 'random short-cut' and 'sending packets to any node ahead in the chain' protocols, respectively. Please note that not all these neighbors are necessarily different.

Let solution $\hat{\mathbf{B}}$ denote the ℓ^{th} neighbor of solution $\mathbf{B}$. Then the expression of $\hat{\mathbf{B}}$ as a function of $\mathbf{B}$ and ℓ is the following.

$$\hat{B}_{uv} = \begin{cases} B_{uv} - 1 \text{ if } u = \hat{u} \wedge v = \hat{v}, \\ B_{uv} + 1 \text{ if } u = \hat{u} \wedge v = 1 + [\hat{v} - 2(\ell \bmod 2)] \bmod (M+1), \\ B_{uv} \text{ otherwise}, \end{cases} \tag{3.53}$$

where

$$\begin{aligned} &\hat{u} = (\ell+1) \operatorname{div} 2, \\ &\hat{v} = \arg \min_{w=1,\ldots,M+1} \left\{ \sum_{v=1}^{w} B_{\hat{u}v} = 1 \right\}, \\ &\ell = 1,\ldots,2(N-1) \end{aligned} \tag{3.54}$$

and

$$\begin{aligned} &\hat{u} = \arg \min_{i=1,\ldots,N-1} \left[\sum_{j=1}^{i} j \geq (\ell+1) \operatorname{div} 2 \right], \\ &\hat{v} = \arg \min_{w=1,\ldots,M+1} \left\{ \sum_{v=1}^{w} B_{\hat{u}v} = \arg \min_{i=1,\ldots,N-1} \left[\sum_{j=1}^{i} j \geq (\ell+1) \operatorname{div} 2 \right] - (\ell+1) \operatorname{div} 2 \right\}, \\ &\ell = 1,\ldots,N(N-1) \end{aligned} \tag{3.55}$$

for the two different protocols, respectively. In other words, $(\ell+1) \operatorname{div} 2$ is used to select a point of division, while $\ell \bmod 2$ chooses the direction in which it is moved between the quantization levels.

Generating $\hat{\mathbf{B}}$ from $\mathbf{B}$ – i.e. the generation mechanism – involves selecting the ℓ^{th} neighbor of $\mathbf{B}$, where ℓ is random variable distributed uniformly over its possible values (listed above).

DOI:10.15774/PPKE.ITK.2009.001

3.5. Performance analysis

The value of the different simulation parameters is listed in Table 3.I. Please note that the values of the parameters associated with the transceiver of the nodes, i.e. P_{sens}, η_{PA} and $P_{\text{circ,t}}$, were estimated based on the specification of the Crossbow Mica2 mote [21]. Variable D denotes the (maximum) length of the network, whereas T is the unit of time in the case of the discrete time model. Despite the fact that the capacity (C) of NiMH rechargeable batteries is around a few Ah in practice, we applied a value that is around three orders of magnitude lower in order to save computation time when calculating (3.29) or (3.40). Nevertheless, the simulation results obtained with this value provide meaningful information even for the case of much larger capacities. Finally, the value of the parameters related to the simulated annealing algorithm is listed in Table 3.II.

TABLE 3.I. SIMULATION PARAMETERS.

Parameter	Value	Unit
P_{sens}	-109 [21]	dBm
η_{PA}	17.35 [21]	%
$P_{\text{circ,t}}$	13.74 [21]	dBm
supply voltage	3	V
Δt	0.25	s
C	1	mAh
α	0.01	-
β	20	-
γ	-4	-
T	1	min
D	112.9	m

TABLE 3.II. OPTIMIZATION PARAMETERS.

Parameter	Value	Unit
χ_0	0.95	-
δ	0.1	-
ε_s	10^{-10}	-
M	50	-

Figure 3.5, Figure 3.6 and Figure 3.7 describes the lifetime, the packet loss ratio and the mean delay as a function of the buffer length in the discrete time case, respectively. The different curves belong to the different packet forwarding protocols introduced in Section

DOI:10.15774/PPKE.ITK.2009.001

3.2.4. A number of $N=10$ nodes were simulated, which were deployed equidistantly over the maximum distance D including also the BS. The value of p_i was generated randomly subject to a uniform distribution over the $[0,0.33]$ interval. The lengths of the buffer of the different nodes were set equal.

It can be observed in Figure 3.5 and Figure 3.6 that – except for the shortcut protocol – both lifetime and packet loss ratio decreases with the buffer size. (For the shortcut protocol, both quantities are obviously independent of the buffer size.) This is due to the fact the larger the buffer the less probable that the buffer gets full and as a result, fewer packets are lost and more packets need to be sent. It is also found that for $L>5$ lifetime, packet loss ratio and mean delay are independent of the buffer size. It is noteworthy the 'random shortcut' and 'sending to any node ahead in the chain' protocols provide a considerably longer lifetime than the shortcut protocol, while their performance concerning packet loss ratio and mean delay is practically identical to that of the latter protocol – the protocol that is ideal in terms of these two measures.

One could expect at first glance that – for a fixed buffer size – there is a trade-off between lifetime and packet loss ratio. However, as shown by further simulations not presented here, this is not the case. For an $\mathbf{A}$ matrix with a higher lifetime, packet loss ratio is also lower. The reason for that is that if we increase the incoming traffic of a queue then – though the packet loss ratio increases – the outgoing traffic (somewhat surprisingly) also increases. This behavior can be justified analytically e.g. for the M/M/1/K -type queuing system.

In the case of the continuous time model, a number of $N=15$ were simulated. The location of the sensor nodes were set in the following three different ways.

- The nodes were deployed equidistantly (i.e. deterministically) over the maximum distance D including also the BS;
- the locations of the nodes were generated randomly according to a uniform distribution over the maximum distance D;
- the locations of the nodes were generated randomly according to a truncated exponential distribution with parameter D^{-1} over the maximum distance D measuring the distance from the BS.

DOI:10.15774/PPKE.ITK.2009.001

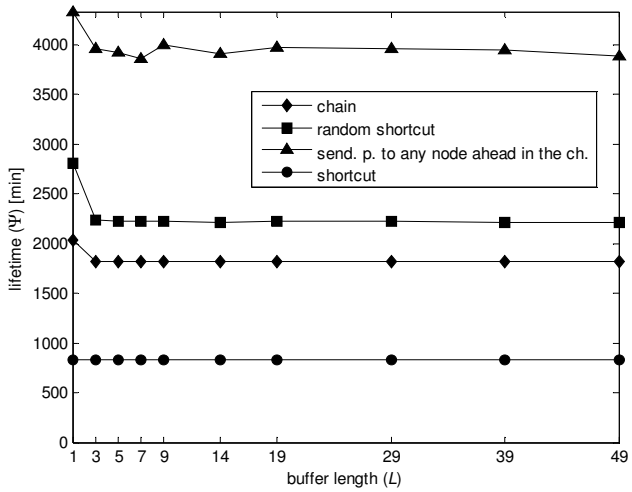

Figure 3.5. Lifetime (Ψ) as a function of the buffer length (L) (discrete time model).

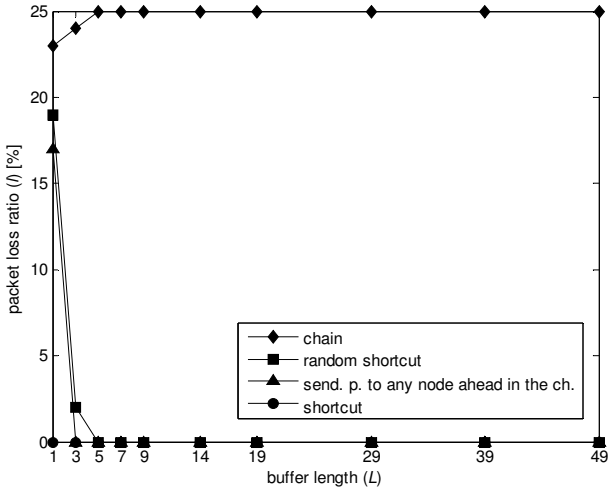

Figure 3.6. Packet loss ratio (l) as a function of the buffer length (L).

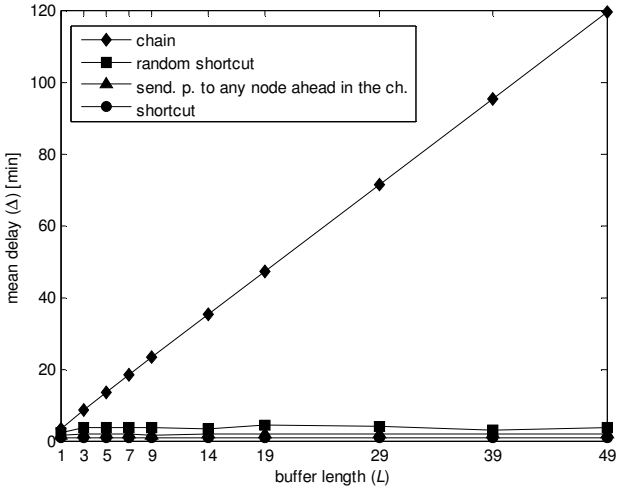

Figure 3.7. Mean delay (Δ) as a function of the buffer length (L).

DOI:10.15774/PPKE.ITK.2009.001

Figure 3.8 shows the lifetime of the network for the different protocols and node location setups. The value of λ_i was generated randomly subject to a uniform distribution over the $[0,1]$ interval. We can observe on the diagram that the 'random shortcut' protocol considerably outperforms the chain and shortcut protocols, while the 'sending packet to any node ahead in the chain protocol' shows a superior performance to all the other protocols for any of the topologies under investigation.

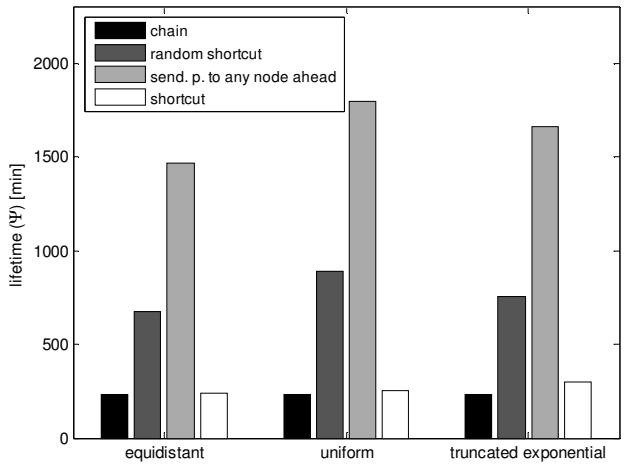

Figure 3.8. Lifetime (Ψ) (continuous time model).

In order to justify the validity of the analytical approximations introduced in Section 3.3, we carried out packet-level simulations for both models. In the discrete time model, a number of 2500 random choices of matrix $\mathbf{A}$ were tested under the assumption of $L_i = 10$. The results showed a maximum relative error less than 6% along with an average relative error of 3%. In the continuous time model, a number of 1000 random choices of matrix $\mathbf{A}$ were tested for each of the different node location setups. The results showed a maximum relative error less than 7% along with an average relative error of 3% in each case.

3.6. Conclusions

In this chapter, energy aware packet forwarding protocols for wireless sensor networks with a stochastic traffic model have been investigated. The numerical results in the previous section have clearly demonstrated that the proposed optimal stochastic packet forwarding protocols under investigation can considerably increase the lifetime of wireless sensor networks compared to the case of the traditional chain and shortcut protocols. In addition to lifespan, the performance of the former protocols has been

DOI:10.15774/PPKE.ITK.2009.001

analyzed also in terms of packet loss ratio and mean delay and has been found satisfactorily low. The investigations have also shown that the optimization of the protocols can be carried out in a reasonable computation time (few hours) for a number of $10-15$ nodes using a personal computer (PC) and MATLAB.

The analysis in this chapter has been restricted to the case of one-dimensional network topologies and negligible power consumption at reception. However, it has also been shown that the above treatment can easily be extended to two-dimensional topologies and energy consumption models in which circuit power consumption at reception is comparable to that at transmission. Hence the numerical analysis of the performance of the protocols in these more general cases can be pointed out as a possible future work. Finally, the selection of a suitable/optimal spanning tree for the 2D topologies – which is a prerequisite to the application of the protocols – can also be identified as a related open research problem.

3.7. Applications

The energy aware stochastic packet forwarding protocols investigated in this chapter can be used in a wide range of wireless sensor network applications such as intelligent home (e.g. the NETVOX Smart House System), mechanical structure monitoring (e.g. the Sensametrics Wireless Structure Monitoring), environment monitoring (e.g. the Advanced Sensor Technologies Root Zone Intelligence System), seismic activity monitoring, etc.

DOI:10.15774/PPKE.ITK.2009.001

Chapter Four

EXTENDING THE LIFETIME OF MEDICAL IMPLANTS THROUGH COOPERATIVE COMMUNICATIONS

4.1. Introduction

With the ever-growing development of information and communication technologies, novel types of services such as telemedicine and telecare have become feasible. Telemedicine is usually defined as the delivery of medical services at a distance, while telecare stands for providing health and social care services from remote locations – both offering the potential to have a favorable impact on the access, quality and cost of the above mentioned services [37]. One important technological component of these remotely delivered services is obviously wireless communications – and primarily radio communications.

Wireless (radio) devices in telemedicine and telecare systems may be located off/on and inside the human body. The latter devices – called implants – can communicate to the outside ones basically in two different ways. In the case of the inductive coupling technique, an inductive link – used many times for both data transfer and implant power supply – is established between the two devices using coils. This method provides a relatively low communication bandwidth with a fairly short communication range and requires the patient to align the external device with the implant rather accurately. On the other hand, when the carrier frequency is shifted to a higher domain – typically to the ultra high frequency (UHF) band – the propagating radio waves enable higher bit rates and longer ranges measurable in meters. This second solution offers more comfort – and e.g. continuous monitoring – to the patient but, in turn, implies greater challenges for the system and device designers as implants with this particular communication technique are generally powered by non-rechargeable batteries. The present chapter addresses these emerging challenges.

In 2002, the European Telecommunications Standards Institute (ETSI) standardized the Medical Implants Communications System (MICS) [41], which is specified to be used by active medical implants communicating to each other or to an external controller in the

DOI:10.15774/PPKE.ITK.2009.001

UHF band. The unlicensed frequency band allocated to the system is from 402 to 405 MHz. The transmitters operating in the MICS band are restricted to a maximum of $25\mu W$ equivalent transmitted power (ERP), while the maximum emission bandwidth that can be occupied by a (complete) communication session is $300kHz$. MICS shares the allocated spectrum with the Meteorological Aids Services (METAIDS), which is primarily used by weather balloons [42]. For this reason, MICS is limited to operate indoors only [37]. (Note that a similar standard was issued for the United States by the Federal Communication Commission (FCC) in 1999 [43].)

A good example of a telemedicine system operating in the MICS band is the Biotronik Home Monitoring system [44]. In this particular system, the data (trend, event, etc.) transmitted by the (battery-operated) implanted pacemaker or cardioverter-defibrillator is received by an exterior device, which then automatically forwards it to Biotronik's Service Centre via the cellular telephone system (GSM). Afterwards, the centre analyzes the data and edits a report, which can be viewed online by the attending physician or – upon request – sent to him/her by fax, e-mail or text message (SMS).

Open problems

The communication between a body-implanted device and an off-body transceiver is treated by only a relatively few papers in the literature on a link level. One such treatment is provided in [39] and [45], where the performance of a traditional (non-cooperative) radio link between a medical implant and a single base station – both located in the same hospital appointment room – is considered. The performance of the link was evaluated (e.g. in terms of channel capacity) by setting up a link budget containing various fading margins specific for the application in question. However, no effort to date in the literature has been made to unveil or to assess the potential of cooperative communications to combat fading – and hence to extend the lifetime of battery-operated implants – in medical implant communication systems. In his present attempt to do so, the author tries to adopt the existing findings on radio wave propagation from medical implants – along with the characteristics of the state-of-the-art low-power transceivers – as much as he can.

DOI:10.15774/PPKE.ITK.2009.001

Radio channels are generally subject to some kind of fading in most propagation environments. In a large number of the practical scenarios, the small dimensions of wireless network nodes prevent the use of multiple antennas on a single node for combating fading. However, by taking advantage of the broadcast nature of the wireless media, the effect of fading can be mitigated through node collaboration usually termed as cooperative diversity [38][46].

In our particular cooperative communication scenario, we consider a single room accommodating a patient with a medical implant inside his/her body (Figure 4.1). The room may be located, as an example, at the patient's home or work place or at the hospital. The implant intends to communicate its sensed data to the outside world at regular intervals. In the case of the traditional, non-cooperative communication approach, the data is received by the only receiver unit in the room, which is placed in the immediate proximity of the patient – e.g. at the bedside table, at the desk or next to the armchair depending on the activity he/she is just performing. Contrary to that, for the cooperative communication scenario, the packet transmitted by the implant is received by multiple battery-operated, wireless cooperating receiver units (CRUs) installed across the room (Figure 4.1). (In addition to the aforesaid role, these nodes may also serve as a part of a larger ambient sensor or perhaps some ad hoc network in the building incorporating the room described above [47].) After that, a couple of the CRUs are selected for cooperation. Finally, these CRUs relay their packet to the gateway CRU (G-CRU) in the room, which then accomplishes the detection of the packet sent by the implant and forwards it to the service centre through – as an example – the internet or the 3G network.

The advantage of the cooperative approach over the non-cooperative one arises from the fact that the radio channel towards the single receiver unit located in the proximity of the implant might be more adversely affected by fading than the channels towards other locations of the room. Under these circumstances, the reception of the implant packet may be more reliable in these locations even if they are farther off from the implant. Thus, by applying multiple receiver units in different locations of the room, the required transmission power of the implant can be reduced under a given reliability constraint. This reduction in transmission power can be converted into lifetime gain for the implant, which stands in the main focus of attention in this part of the dissertation. (Note that

DOI:10.15774/PPKE.ITK.2009.001

concerning the different fading effects the in-body to off-body propagation channel is subject to, multipath fading could be alleviated with a single receiver of reasonable size. The minimum required antenna separation in this case is around a quarter of a wavelength, which is approximately 20cm for the MICS band. The mitigation of shadow fading through traditional spatial diversity would, however, require an impractically large antenna separation, and therefore it is not an option when designing an implant communication system.)

Figure 4.1. The room accommodating the patient. The CRUs and the G-CRU installed across the room are highlighted by the thin circles and the thick circle, respectively. The packet transmitted by the implant is received by multiple CRUs, which then cooperatively make the decision on the implant packet.

The remainder of the chapter is organized as follows. In Section 4.2, the topology model is described along with the energy consumption model of the communication nodes. A plausible model of the radio propagation channel is established based the literature in Section 4.3. Section 4.4 details the different stages of the proposed cooperative communication procedure, while the numerical results on the achievable lifetime gain are presented in Section 4.5. Section 4.6 concludes the chapter.

4.2. Topology and energy consumption model

In this section, the topology and power consumption models applied in the forthcoming investigations are described in details.

DOI:10.15774/PPKE.ITK.2009.001

4.2.1. Topology model

In order to simplify the investigations we consider a two dimensional (2D) spatial model in terms of the location of the communicating parties in the room. In other words, the implant and the CRUs are assumed to lie in the same horizontal plane of the room.

The purpose of this study is to evaluate the potential of cooperative communications to prolong the lifespan of medical implants in general. Therefore, rather than considering a specific spatial arrangement of the CRUs in the room, the positions of the CRUs are regarded as random variables; in particular, they are assumed to be distributed according to a (two dimensional) homogeneous Poisson distribution. In accordance with this assumption, the number of CRUs in the room (N) is Poisson distributed with parameter ρA, A being the area of the room and ρ the spatial density of the CRUs. (Note that we assume that ρA is high enough for the event $N = 0$ to have a negligible probability.) The locations of the CRUs denoted by vectors $\mathbf{r}_i$ ($i = 1,...,N$) are then independent, identically distributed random variables that have a uniform distribution over the area of the room. Once again in order to investigate the problem in general, the G-CRU is chosen randomly from among the CRUs.

In order to take the mobility of the patient into account, also the location ($\mathbf{r}$) and orientation ($\mathbf{r}_0$) of the implant/patient are considered as random variables. (The orientation of the patient is identical to a designated direction of his/her body, for instance, the direction his/her chest is facing.) It is assumed that $\mathbf{r}$ follows a uniform distribution over the area of the room, while $\mathbf{r}_0$ is assumed to be uniformly distributed over the unit circle.

The distance between the implant and CRU_i and that between CRU_i and CRU_j is expressed as

$$d_{\text{I,CRU}_i} = \left|\mathbf{r}_i - \mathbf{r}\right|, \tag{4.1}$$

and

DOI:10.15774/PPKE.ITK.2009.001

$$d_{\mathrm{CRU}_i,\mathrm{CRU}_j} = \left|\mathbf{r}_i - \mathbf{r}_j\right|, \tag{4.2}$$

respectively. Finally, the orientation of CRU_i relative to the implant (ϕ_i) is given by

$$\phi_i = \angle\left(\mathbf{r}_i - \mathbf{r}, \mathbf{r}_0\right). \tag{4.3}$$

4.2.2. Energy consumption model

Numerous studies on wireless communication networks consider only transmission power as a source of power consumption. However, in many wireless sensor network – and especially in wireless medical implant – applications, circuit power consumption cannot be neglected compared to the (transmission power dependent) consumption of the radio frequency (RF) power amplifier (PA). Circuit power consumption is made up of the consumption of the various additional components of the transceiver such as the mixer, the frequency synthesizer, the low noise amplifier (LNA), the intermediate frequency amplifier (IFA), etc.

We apply the power consumption model used in [29], which takes into account not only circuit power consumption but also the losses of the RF power amplifier. The power consumed at transmission (P_{t}) and at reception (P_{r}) is expressed as

$$P_{\mathrm{t}}\left(P_{\mathrm{tr}}\right) = \eta_{\mathrm{PA}}^{-1} P_{\mathrm{tr}} + P_{\mathrm{circ,t}}, \tag{4.4}$$

and

$$P_{\mathrm{r}} = P_{\mathrm{circ,r}}, \tag{4.5}$$

respectively. Here η_{PA} is the efficiency of the RF PA of the transceiver, P_{tr} is the transmission power, while $P_{\mathrm{circ,t}}$ and $P_{\mathrm{circ,r}}$ is the circuit power consumption at transmission and at reception, respectively.

DOI:10.15774/PPKE.ITK.2009.001

Several low-power UHF transceiver designs – for biomedical applications – have recently been reported in the literature. For example, in [48], a low-consumption MICS band transceiver with a 50-μW FSK modulator/demodulator and a 1.1-mW voltage controlled oscillator (VCO) is presented. A fully-integrated frequency synthesizer for implantable medical devices was measured to have a power consumption of 580μW in [49]. A 433 MHz FM transmitter – for neural recording telemetry – dissipating 465μW was reported in [50]. The integrated UHF power amplifiers in [51] and [52] showed a power-added efficiency of 38% and 28.5%, respectively. In order to make the analysis as realistic as possible, in the forthcoming simulations, we are going to assume similar $P_{\text{circ,t}}$ and η_{PA} values for the implanted device.

4.3. Radio propagation model

In this section the radio propagation model to be used throughout the performance analysis is discussed. Based on the literature on radio wave propagation from medical implants, the author sets up a plausible statistical model of the propagation channel between the implant and the CRUs (and that between the CRUs). The establishment of such statistical model is essential when assessing the potential of cooperative diversity to extend the lifetime of medical implants.

4.3.1. Multipath fading

The radio waves emitted by the transmitter are reflected, scattered or diffracted by the objects surrounding the receiver. The components created in this way may sum up either constructively or destructively at the receiver giving rise to the phenomenon of multipath fading.

Johansson et al. measured the path loss for the MICS band in an indoor environment [45][53]. The measurements were carried out in a few-meter by few-meter furnished room at the mid-band frequency of 403MHz using two half-wavelength co-polarized dipole antennas. The observed variations of the path loss around its mean were attributed to multipath fading. The measured path loss was compared to free space loss and its large-scale mean was found to be in a good agreement with that. The excess loss that needs to be added to free space loss in order to cover a given percentage of the measured area was also calculated (Table 4.I). The investigation of the power delay profile of the

DOI:10.15774/PPKE.ITK.2009.001

radio channel was omitted in [45] and [53]. The reason for that is that the symbol time – corresponding to the maximum bandwidth of a MICS session (300kHz) – is about two orders of magnitude higher than it takes the radio waves to travel across the room. Hence, the occurence of intersymbol interference would require a wave of non-negligible power reflected off the walls around hundred times. Since this is fairly unlikely in practice, the channel can be considered as non-dispersive or, in other words, the fading is regarded to be flat.

In order to incorporate multipath fading into the present model, the author assumes Rice-fading (e.g. [54]) in the room as it is presumable that there is always a line-of-sight (LOS) multipath component (MPC) among the MPCs propagating between the implant and the CRUs. The K_r parameter of Rice-fading, called Rice-factor, is defined as the ratio of the power of the LOS MPC and that of the diffuse MPCs. Table 4.I shows the excess loss values measured in [45] along with ones that can be obtain for Rice-fading assuming different values for parameter K_r. (Note that Rice-fading with $K_r = 0$ is identical to Rayleigh-fading.) Based on the table, it can be concluded that no close match between the measured and calculated values can be found for any of the values of parameter K_r. Nevertheless, $K_r = 1$ seems to be an appropriate choice and will be used throughout this study. (Please remark that in the measurements in [45], samples were taken at a 2.5cm resolution over an area of $92\times52\text{cm}$. Considering the fact that the wavelength is around 74cm in the MICS band and that fading dips are located approx. half a wavelength apart, the size of the area can be regarded rather small to draw reliable conclusions concerning the statistics of the fading.)

TABLE 4.I. MEASURED AND CALCULATED EXCESS LOSS VALUES.

	measured [45]	**$K_R = 0\ (-\infty\text{dB})$**	**$K_R = 1\ (0\text{dB})$**	**$K_R = 4\ (6\text{dB})$**
90%	5dB	9.7dB	8.6dB	5dB
95%	8dB	12.9dB	11.7dB	6.8dB
97%	11dB	15.2dB	13.9dB	8.2dB
99%	18dB	20dB	18.7dB	11.5dB

DOI:10.15774/PPKE.ITK.2009.001

4.3.2. Polarization mismatch

A possible mismatch between the polarization of the incoming radiation and that of the receiver antenna may lead to additional losses in the radio communication link [45].

Johansson calculated the transversal radiation pattern of an implant placed in a homogeneous numerical human phantom [45]. The phantom was filled with a muscle-simulating (virtual) liquid [37]. The calculations were carried out using the finite-difference time-domain (FDTD) method for optimum (matched) polarization as well as for the polarization that is perpendicular to the transversal plane. The ratio of the calculated patterns and hence the polarization was found to show significant variations over the different directions of the plane. Scanlon et al. performed quite similar simulations and measurements for a semi-segmented (inhomogeneous) numerical phantom and a female subject, respectively, at a frequency of 418MHz [55]. Their results also confirmed large variations of the polarization with the azimuthal angle.

Nevertheless, none of the above studies has given an analytical or even a statistical description of these variations. In addition, the fact that we have very little control over the orientation of the patient makes the investigation of the polarization even more complicated. For these reasons, the author proposes the following simple probabilistic model for the (far-field) polarization of the wave transmitted towards CRU_i:

$$p^{\mathrm{t}}_{\vartheta,i} = \cos\beta_i, \tag{4.6}$$

$$p^{\mathrm{t}}_{\varphi,i} = \sin\beta_i \cdot e^{j\gamma_i}, \tag{4.7}$$

where β_i and γ_i $(i = 1,...,N)$ are independent random variables that are uniformly distributed over the interval $[0,2\pi]$, while $p^{\mathrm{t}}_{\vartheta,i}$ and $p^{\mathrm{t}}_{\varphi,i}$ are the components of the polarization in the spherical polar coordinate system centered at the implant with coordinates ϑ and φ, $\vartheta = 0$ lying in the vertical direction. Differently from the other parts of the chapter, j denotes the imaginary unit in this subsection. (Note that the polarization vector described in (4.6) and (4.7) is a random unit vector with a random phase difference between its components.) The above conditions also imply the

DOI:10.15774/PPKE.ITK.2009.001

assumption that the polarization of the waves transmitted towards the different CRUs are statistically independent.

Johansson et al. measured the path loss in the room using cross- and cross-90-polarized receiver antennas as well [45][53]. They found that the path loss with these receiver polarizations is considerably higher than that in the co-polarized case. As a consequence, it is assumed in the calculations that the channel between the implant and the CRUs does not alter the polarization.

When evaluating the effect of the mismatch between the polarization of the incident wave and that of the receiver antenna, the polarization of the receiver antenna – similarly to the polarization of the impinging wave – can be described by a unit vector. As an example, the polarization of a linearly polarized antenna – polarized in the ϑ direction, – is represented by the unit vector

$$p_{\vartheta}^{\mathrm{r,lin}} = 1, \tag{4.8}$$

$$p_{\varphi}^{\mathrm{r,lin}} = 0, \tag{4.9}$$

whereas the polarization of a right-circularly polarized antenna is described by the unit vector

$$p_{\vartheta}^{\mathrm{r,lin}} = \frac{1}{\sqrt{2}}, \tag{4.10}$$

$$p_{\varphi}^{\mathrm{r,lin}} = j\frac{1}{\sqrt{2}}. \tag{4.11}$$

We are going to apply these antenna types for the CRUs in the calculations.

Concerning the multipath fading the different polarization components are subject to, it is assumed in our model that the Rice-fading coefficients for the ϑ- and φ-components of the wave radiated towards CRU$_i$ ($a_{\vartheta,i}^{\mathrm{R}}$ and $a_{\varphi,i}^{\mathrm{R}}$) are statistically independent [54]. These are given by

DOI:10.15774/PPKE.ITK.2009.001

$$a_{\vartheta,i}^{\mathrm{R}} = \frac{1}{\sqrt{2(K_{\mathrm{r}}+1)}}\left|\sqrt{2K_{\mathrm{r}}} + x_{\vartheta,i} + jy_{\vartheta,i}\right|^2, \tag{4.12}$$

$$a_{\varphi,i}^{\mathrm{R}} = \frac{1}{\sqrt{2(K_{\mathrm{r}}+1)}}\left|\sqrt{2K_{\mathrm{r}}} + x_{\varphi,i} + jy_{\varphi,i}\right|^2, \tag{4.13}$$

where $x_{\vartheta,i}$, $y_{\vartheta,i}$, $x_{\varphi,i}$ and $y_{\varphi,i}$ $(i = 1,...,N)$ are independent, standard Gaussian random variables. The conditions above also imply that the fading the different implant-to-CRU channels are exposed to, is statistically independent. Note that this is a reasonable assumption as the fading dips investigated in [45] and [53] were measured to be approximately $40\mathrm{cm}$ apart from each other.

4.3.3. Radiation efficiency and radiation pattern

When assessing the performance of a wireless communication link, the antenna characteristics such as the radiation efficiency and the radiation pattern also need to be taken into account for both the transmitter and the receiver antenna.

For an implanted antenna, the implant itself and the body are located in the near-field of the antenna and, as a result, they have a fundamental effect on the radiation characteristics [37]. Therefore, for implanted antennas, the antenna characteristics are regarded as that of the complex made up of the implanted antenna, the implant and the body.

Johansson estimated the radiation efficiency (η) of a wireless implant in the MICS band using the FDTD method [37][45]. The radiation efficiency was found to be around $-30\mathrm{dB}$ depending on the body (i.e. homogeneous numerical phantom) shape and the depth of the implant in the phantom. Also Scanlon et al. estimated and measured the radiation efficiency of an implant placed into a numerical phantom and a female subject, respectively [55]. The values obtained in this case were close to $-20\mathrm{dB}$. According to the above results, we are going to apply similar radiation efficiency values in the coming investigations.

DOI:10.15774/PPKE.ITK.2009.001

Johansson simulated also the radiation pattern of a wireless implant under the assumption of optimum polarization matching [39][45]. The margin that has to be added to the maximum directivity to get a given percentage of coverage over all the possible angles was estimated (Table 4.II). Likewise, Scanlon et al. simulated and measured the radiation pattern of an implant during their investigations mentioned above [55]. Their results also revealed considerable variations in the radiation pattern.

Changes in body posture cause significant alterations in the radiation pattern of the implant [45][56]. This effect can be interpreted as a kind of shadow fading [37]. Johansson et al. calculated the radiation pattern of a wireless implant placed into a numerical phantom for 8 different arm positions [45][56]. The excess loss that needs to be added to the mean of the 8 directivity patterns in order to cover a given percentage of the entire spatial angle was determined (Table 4.III).

Nonetheless, no analytical or detailed statistical description of either the radiation pattern or that of the impact of body posture has been provided by these studies. Variations due to shadowing effects show a log-normal distribution in most of the practical cases [54]. Moreover, when measuring the path loss from a physical numerical phantom filled with animal organs, the variations of the path loss around its mean was also attributed to shadowing effects by Alomainy et al [57]. The measured path loss samples showed a log-normal distribution. As a consequence of these findings, the author proposes to approximate the variations in the radiation pattern and those due to the changes in body posture by the following axially symmetric 2D directivity pattern (Figure 4.2):

$$D(\phi) = \left(\int_0^{2\pi} \tilde{D}(\phi) d\phi \right)^{-1} \tilde{D}(\phi), \tag{4.14}$$

where

$$\tilde{D}^{[dB]}(\phi) = \begin{cases} F^{-1}\left(\pi^{-1}\phi\right) & if \quad 0 \le \phi \le \pi, \\ F^{-1}\left(2 - \pi^{-1}\phi\right) & \text{otherwise}, \end{cases} \tag{4.15}$$

and

DOI:10.15774/PPKE.ITK.2009.001

$$F(u)=\left(\int_{L}^{H} e^{-\frac{t^2}{2\sigma}}dt\right)^{-1}\int_{L}^{u} e^{-\frac{t^2}{2\sigma}}dt\ ,\ L\le u\le H\ . \tag{4.16}$$

This pattern has the following property: if we consider a random direction that follows a uniform distribution, then the corresponding directivity pattern value is distributed according to a truncated log-normal distribution. The distribution is truncated at both ends in order to restrict its – otherwise infinite – support to the dynamic ranges calculated in [39], [45], [56] and measured in [55].

Based on the excess loss values simulated in [45], it is plausible to model the variations due to the changes in body posture (arm movement) by means of a log-normal distribution with a standard deviation of $\sigma_{\mathrm{P}}=6.5\mathrm{dB}$ (Table 4.III). It is easy to see that the log-normal distribution that has $12\mathrm{dB}$ in between its 1st and 99th percentiles has a standard deviation of around $\sigma_{\mathrm{B}}=2.5\mathrm{dB}$. Hence – considering the former percentile as the minimum of the distribution – it is a reasonable assumption to model the variations due to the body with the help of the latter distribution (Table 4.II) The product of two log-normally distributed random variables is also log-normally distributed with a standard deviation (in dB) that is equal to the sum of the standard deviations of the two individual distributions.As a result, the untruncated distribution is estimated to have a standard deviation of around $\sigma=\sigma_{\mathrm{P}}+\sigma_{\mathrm{B}}=9\mathrm{dB}$. The distribution is truncated at its 1st and 90th percentiles, i.e. at $L\approx -21\mathrm{dB}$ and $H\approx 11.5\mathrm{dB}$, respectively. The truncation is more pronounced at the positive end as sharp dips are more typical for the radiation patterns – calculated in [39], [45], [56] and measured in [55] – than sharp peaks.

The normalization in the expression of $D(\phi)$ (4.14) is necessary in order for the directivity pattern ($D(\phi)$) to have a unit mean. This normalization also implies the (optimistic) assumption that all of the emitted power of the implant is concentrated in the horizontal plane in question.

DOI:10.15774/PPKE.ITK.2009.001

TABLE 4.II. SIMULATED AND CALCULATED MARGIN VALUES.

	body (transversal plane) [45]	**log-normal distribution** $\sigma = 2.5$dB
1%	-	0dB
99%	12dB	11.6dB

TABLE 4.III. SIMULATED AND CALCULATED EXCESS LOSS VALUES.

	arm movement (saggital plane) [45]	**log-normal distribution** $\sigma = 6.5$dB
90%	8.7dB	8.3dB
95%	11.4dB	10.7dB
97%	13.2dB	12.2dB
99%	14.9dB	15.1dB

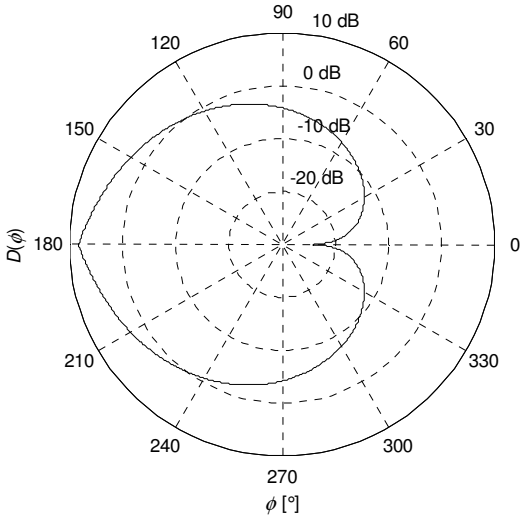

Figure 4.2. The assumed directivity pattern ($D(\phi)$) of the implant. The pattern has the property that if we consider a random direction that follows a uniform distribution, then the corresponding directivity pattern value is distributed according to a (truncated) log-normal distribution.

4.3.4. The resulting propagation model

Finally – based on the previous sections – the author proposes to piece together the propagation model in the following two equations, which concern the implant-to-CRU$_i$ and the CRU$_i$-to-CRU$_j$ channels gains ($h_{\mathrm{I,CRU}_i}$ and $h_{\mathrm{CRU}_i,\mathrm{CRU}_j}$), respectively ($i, j = 1,...,N$).

DOI:10.15774/PPKE.ITK.2009.001

$$\left|h_{\mathrm{I,CRU}_i}\right|^2 = \left(\frac{4\pi\lambda}{d_{\mathrm{I,CRU}_i}}\right)^2 \left|a^{\mathrm{R}}_{\vartheta,i}\, p^{\mathrm{t}}_{\vartheta,i}\, p^{\mathrm{r}}_{\vartheta,i} + a^{\mathrm{R}}_{\varphi,i}\, p^{\mathrm{t}}_{\varphi,i}\, p^{\mathrm{r}}_{\varphi,i}\right|^2 \eta D(\phi_i), \tag{4.17}$$

$$\left|h_{\mathrm{CRU}_i,\mathrm{CRU}_j}\right|^2 = \left(\frac{4\pi\lambda}{d_{\mathrm{CRU}_i,\mathrm{CRU}_j}}\right)^2 \left|a^{\mathrm{R}}_{\vartheta,ij}\, p^{\mathrm{r}}_{\vartheta,i}\, p^{\mathrm{r}}_{\vartheta,j} + a^{\mathrm{R}}_{\vartheta,ij}\, p^{\mathrm{r}}_{\varphi,i}\, p^{\mathrm{r}}_{\varphi,j}\right|^2, \tag{4.18}$$

where the first factor in the equations is the expression of free space loss with λ being the wavelength in air. Variables $p^{\mathrm{r}}_{\vartheta,i}$ and $p^{\mathrm{r}}_{\varphi,i}$ are the ϑ- and φ-components of the unit vector describing the polarization of the antenna of CRU_i, respectively. It is implicitly assumed in (4.17) and (4.18) that the CRUs are equipped with isotropic and lossless antennas and that the CRU_i-to-CRU_j channels are subject to the same type of multipath fading as the implant-to-CRU_i channels. The fading factors for the former channel are denoted by $a^{\mathrm{R}}_{\vartheta,ij}$ and $a^{\mathrm{R}}_{\varphi,ij}$.

The propagation model proposed in this section is obviously just an approximation. This – among other things – is due to the fact that it has been put together based on the results of some studies that treated the different aspects of the propagation channel separately. The author had to rely on this solution since – to the best of his knowledge – no paper to date investigates the properties of the in-body to off-body propagation channel in an indoor environment directly and as a whole. As an example, it is assumed in the model that all of the power received at a given CRU originates from a single wave coming from the direction of the implant [45] and it is scattered only in the close vicinity of the CRU. This is not necessarily true in practice as waves of non-negligible power that are reflected from relatively distant walls may also arrive at the CRU. Another limitation of the model is the assumption that even the closest CRUs are in the far field of the implant-body complex. However, the establishment of a more accurate statistical channel model – either by FDTD simulations or by measurements – requires a considerable amount of time and effort, and therefore it is out of the scope of this paper.

4.4. Cooperative communication scheme

In this section the different stages of the proposed cooperative communication scheme as well as the definition of the lifetime gain are presented in details.

DOI:10.15774/PPKE.ITK.2009.001

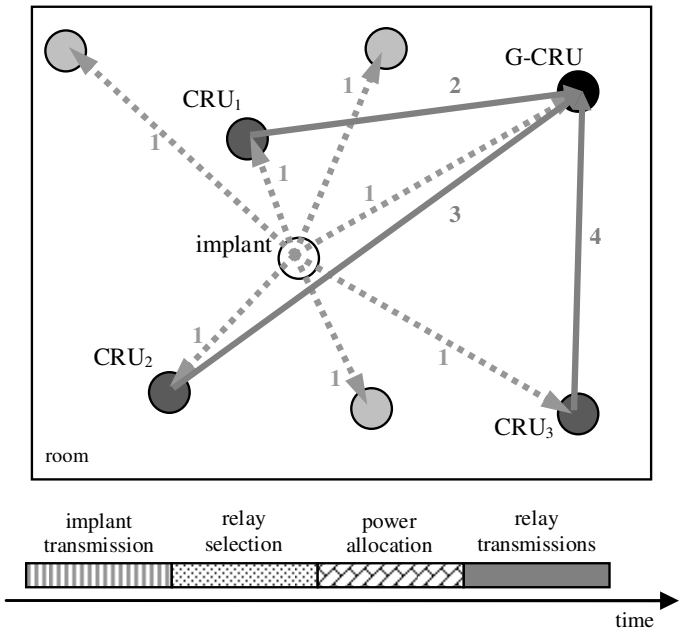

Figure 4.3. The outline of the proposed cooperative communication scheme in space (top) and time (bottom). The relaying and non-relaying CRUs are depicted by light grey and dark grey circles, respectively. The communication taking place during the implant transmission and relaying stages is represented by the dashed (1) and solid (2, 3, 4) arrows, respectively. For the sake of simplicity, the messages exchanged during relay selection and power allocation are not depicted.

The outline of the cooperative communication procedure is depicted in Figure 4.3. At the beginning of the procedure the packet transmitted by the implant – containing the sensed data – is received by all the CRUs in the room. After that the relay selection stage takes place, during which the CRUs that are the most suitable for forwarding the implant packet to the G-CRUs are selected. In the course of the next stage (i.e. the power allocation stage), the transmission power level of the relays is determined. Finally, the selected CRUs relay their packet to the G-CRU, which then optimally combines the different versions of the implant packet including the one that is directly received.

As the goal of this chapter is mainly to point out to the possible advantages of cooperative diversity in medical implant communications, we consider only the simple case when the implant uses the same transmission power level at each and every transmission. The more complex problem of implant transmission power control is omitted in this study.

DOI:10.15774/PPKE.ITK.2009.001

4.4.1. Relaying stage

There are a number of relaying methods treated in the literature of cooperative communications. The most widely investigated of these are decode-and-forward and amplify-and-forward (e.g. [58]). In the case of the former one, the packet received by the relay is decoded and re-encoded before retransmission. For the latter – which is also the one to be used in the present investigations – the relay simply amplifies and retransmits the analog signal received from the source.

In terms of medium access, relays may amplify and forward their received signal *one after the other* (Figure 3) or *at the same time*, both alternatives having their pros and contras. For the latter solution called simultaneous transmission based relaying [59], the noise of the destination (in our present case, that of the G-CRU) is added only to the combination of the relayed packets. On the other hand, this alternative necessitates a complex radio hardware (HW) as the relayed signals have to be phase adjusted in a distributed fashion to sum up constructively [60]. For the former solution usually termed as repetition based relaying (e.g. [58]), the situation is just the opposite. Though the noise of the destination is added to each of the relayed packets, this alternative requires a low-complexity radio. For this reason, the former solution is chosen to be employed in the forthcoming analysis.

Since the CRUs – with the exception of the G-CRU – are also assumed to be battery-operated, the power consumption of these nodes might also be of importance. As a consequence, a limit is introduced on the total transmission power of the relaying CRUs denoted by P_{CRU}. Furthermore, in order to enhance the efficiency of transmission power utilization, the problem of optimal power allocation (among the relays) is also treated in the forthcoming investigations.

The different versions of the implant packet are optimally combined at the G-CRU according to maximum ratio combining [54]. It can be readily shown that the resultant signal-to-noise ratio (SNR) of the combined relay packet is given by

$$SNR = \sum_{i=1}^{n} \frac{B_i x_i}{1 + C_i x_i}, \tag{4.19}$$

DOI:10.15774/PPKE.ITK.2009.001

with

$$B_i = C_i \cdot \frac{P_{\mathrm{I}}}{\sigma_{\mathrm{n}}^2}\left|h_{\mathrm{I,CRU}_i}\right|^2, \tag{4.20}$$

$$C_i = \frac{P_{\mathrm{CRU}}\left|h_{\mathrm{CRU}_i,\mathrm{G\text{-}CRU}}\right|^2}{P_{\mathrm{I}}\left|h_{\mathrm{I,CRU}_i}\right|^2 + \sigma_{\mathrm{n}}^2}, \tag{4.21}$$

where n is the number of selected relays, x_i is the proportion of the total relay power (P_{CRU}) allocated to CRU_i, P_{I} is the transmission power of the implant and σ_{n}^2 is the variance of the receiver noise. (Note that in this subsection, CRU_i refers to the i^{th} one among the CRUs selected for relaying, $i = 1,...,n$, $n < N$.)

The problem of optimal power allocation among the relaying CRUs can be formulated as

$$\mathbf{x}_{\mathrm{opt}} = \arg\max_{\mathbf{x}} SNR, \tag{4.22}$$

subject to the constraint that

$$\sum_{i=1}^{n} x_i = 1,\ 0 \le x_i \le 1,\ i = 1,...,n. \tag{4.23}$$

The optimization problem of the identical form, though in a somewhat different context, was treated in [61]. The solution has the form of

$$x_{i,\mathrm{opt}} = C_i^{-1}\left(\sqrt{B_i}\, g(n_1) - 1\right)^+, \tag{4.24}$$

where

$$g(k) = \left(\sum_{j=1}^{k} C_{\delta_j}^{-1}\sqrt{B_{\delta_j}}\right)^{-1}\left(1 + \sum_{j=1}^{k} C_{\delta_j}^{-1}\right),\ 1 \le k \le n, \tag{4.25}$$

DOI:10.15774/PPKE.ITK.2009.001

n_1 is the largest integer that

$$\sqrt{B_{\delta_{n_1}}}\, g(n_1) > 1, \;,\; 1 \le n_1 \le n\,, \tag{4.26}$$

$$\{\delta_1, ..., \delta_n\} \equiv \{1, ..., n\}, \tag{4.27}$$

such that

$$B_{\delta_1} \ge B_{\delta_2} \ge ... \ge B_{\delta_n}\,, \tag{4.28}$$

and

$$(x)^+ = \begin{cases} x \;\; if \;\; x > 0, \\ 0 \;\; \text{otherwise.} \end{cases} \tag{4.29}$$

Remark that CRUs with $x_{i,\text{opt}}$ equal to 0 do not actually relay the implant packet.

The expression of the resultant SNR of the combined relay packet under the assumption of optimal power allocation is given by

$$SNR_{\max} = \sum_{i=1}^{n_1} \frac{B_{\delta_i}}{C_{\delta_i}} - \left(\sum_{i=1}^{n_1} \frac{\sqrt{B_{\delta_i}}}{C_{\delta_i}} \right)^2 \left(1 + \sum_{i=1}^{n_1} C_{\delta_i}^{-1} \right)^{-1}. \tag{4.30}$$

It is easy to see that, for the hypothetic case when P_{CRU} converges to infinity (CRUs with unlimited transmission power), the resultant SNR of the combined relay packet is given by

$$\lim_{P_{\text{CRU}} \to \infty} SNR_{\max} = \sum_{i=1}^{n} \frac{B_i}{C_i} = \sum_{i=1}^{n} \frac{P_1}{\sigma_{\text{n}}^2} \left| h_{\text{I,CRU}_i} \right|^2. \tag{4.31}$$

DOI:10.15774/PPKE.ITK.2009.001

The expression behind the summation sign in (4.31) is the SNR of the implant-to-CRU$_i$ channel ($SNR_{\mathrm{I,CRU}_i}$) and will henceforth be referred to as the received SNR of CRU$_i$.

The combined relay packet and the packet that is received directly are optimally combined by the G-CRU according to MRC. Thus the resultant SNR of the packet to be decoded by the G-CRU is expressed as

$$SNR_{\mathrm{res}} = \frac{P_{\mathrm{I}}}{\sigma_{\mathrm{n}}^2}\left|h_{\mathrm{I,G\text{-}CRU}}\right|^2 + SNR_{\max}. \tag{4.32}$$

4.4.2. Relay selection

The problem of selecting a single (opportunistic) relay in a fully distributed manner was treated in [40] and [62]. The proposed solution is based on the method of distributed timers and could easily be extended to solve our problem (multiple relay selection) as well (please refer to Chapter 5). However, due to the relatively low number of CRUs, i.e. potential relays in the room, the following simple centralized relay selection procedure is applied.

The procedure is initiated after the transmission of the implant packet. At this time the CRUs (except for the G-CRU) send their received SNR to the G-CRU one after the other. The G-CRU then selects a number of n CRUs based on a selection criterion that takes into account the channel gain of both the implant-to-CRU$_i$ and the CRU$_i$-to-G-CRU channels. At the end, the G-CRU broadcasts the outcome of the selection procedure to the CRUs. (Note that in both the present and the subsequent subsection, $i = 2,...,N$ with the G-CRU being CRU$_1$.)

Concerning the relay selection criterion, two cases are investigated. For the criterion called *complex*, the G-CRU selects the CRUs with the highest value for $SNR_{\mathrm{I}\to\mathrm{CRU}_i\to\mathrm{G-CRU}}$. This quantity is the received SNR for the implant-to-CRU$_i$-to-G-CRU communication channel for the hypothetic case when all of the total relay power is allocated to CRU$_i$ and is expressed as

DOI:10.15774/PPKE.ITK.2009.001

$$SNR_{\mathrm{I}\to\mathrm{CRU}_i\to\mathrm{G-CRU}} = \frac{B_i}{1+C_i}. \tag{4.33}$$

Note that the same criterion is used to select a single opportunistic relay in [60]. For the criterion named *simple*, the G-CRU simply selects the CRUs with the highest received SNR.

Based on (4.31), one can immediately find that selecting the nodes with the highest received SNR is the optimal relay selection criterion when $P_{\mathrm{CRU}} \to \infty$. On the other hand, since the expression of $SNR_{\max}$ for the finite P_{CRU} case (4.30) is fairly complicated and most probably cannot be 'decoupled' as (4.31), the optimal relay selection criterion for this case remains an open research problem.

4.4.3. Power allocation

At the beginning of the power allocation procedure (taking place after the relay selection stage), the G-CRU calculates $g(n_1)$, which is then broadcasted to the relaying CRUs [61]. Finally, the relays determine their transmission power level based on the received value according to (4.24) [61].

Note that, in order to carry out the required calculations in the relay selection and power allocation procedures, the G-CRU can estimate the CRU_i-to-G-CRU channel gains based on the transmissions of the CRUs at the beginning of the former procedure provided that those are carried out with a predefined transmission power value. Similarly, assuming reciprocity [40] for the CRU_i-to-G-CRU channels, CRU_i can estimate the CRU_i-to-G-CRU channel gain based on – as an example – the transmission of the G-CRU at the end of the same procedure.

4.4.4. Lifetime gain

In this part, the lifetime gain to be evaluated in the performance analysis is defined.

As already mentioned in the introduction, in the case of the traditional, non-cooperative communication approach, only a single receiver unit is used in the immediate vicinity of

DOI:10.15774/PPKE.ITK.2009.001

the patient. We are going to model the location of this unit ($\mathbf{r}_{\mathrm{SRU}}$) as if it were identical to that of the CRU that is found the closest to the implant in the statistical topology model introduced in Section 4.2, i.e.

$$\mathbf{r}_{\mathrm{SRU}} = \mathbf{r}_{\arg\min_{i=1,\ldots,N}|\mathbf{r}_i-\mathbf{r}|}. \tag{4.34}$$

In this way, both the relative proximity and the uncertain placement of the single receiver unit (SRU) are taken into account in our model. The resultant SNR of the implant packet when only a single receiver unit is employed, is given by

$$SNR_{\mathrm{res}} = \frac{P_{\mathrm{I}}}{\sigma_{\mathrm{n}}^2}\left|h_{\mathrm{I,SRU}}\right|^2, \tag{4.35}$$

where $h_{\mathrm{I,SRU}}$ is the gain of the implant-to-SRU channel.

The reliability constraint we place on the communication link from the implant to the outside world is formulated as

$$P\left(SNR_{\mathrm{res}}\left(P_{\mathrm{I}}\right) < SNR_{\mathrm{req}}\right) \le P_{\mathrm{out}}, \tag{4.36}$$

where SNR_{req} and P_{out} are the required SNR and the maximum outage probability, respectively. In other words, the transmission power of the implant shall be high enough in order for the outage event of an unacceptably low resultant SNR to have a sufficiently low probability.

Finally, the lifetime gain is defined as the relative difference in lifetime between the cooperative and non-cooperative approaches. It is a reasonable assumption that the energy consumption of the implant is dominated by the transmission activity and that the energy consumption associated with the other activities such as reception or sensing can be neglected due to either their relatively low duty-cycle or power requirement. As a direct consequence of this, lifetime is assumed to be directly proportional to the

DOI:10.15774/PPKE.ITK.2009.001

reciprocal of the power consumed at transmission (P_{t}) and, hence, the lifetime gain takes the form of

$$G_{\mathrm{lifetime}} = \frac{P_{\mathrm{t}}\left(P_{\mathrm{I}}^{\text{non-coop}}\right)}{P_{\mathrm{t}}\left(P_{\mathrm{I}}^{\text{coop}}\right)} - 1. \tag{4.37}$$

Here $P_{\mathrm{I}}^{\text{non-coop}}$ and $P_{\mathrm{I}}^{\text{coop}}$ denotes the minimum transmission power level of the implant that satisfies (4.36) for the non-cooperative and cooperative communication scheme, respectively.

4.5. Performance analysis

The model elaborated in the preceding sections is obviously analytically intractable generating the need for approximations. However, in order to obtain controllably accurate results, the performance analysis is carried out numerically using Monte Carlo simulations. For details on how Monte Carlo simulations are performed in this study, please refer to Appendix B.

TABLE 4.IV. SIMULATION PARAMETERS.

Parameter	**Value**	**Unit**
η	-31.1 [37]	dB
η_{PA}	33.33	%
$P_{\mathrm{circ,t}}$	$-3/0.5$	dBm / mW
SNR_{req}	0	dB
P_{out}	10^{-2}	-
bandwidth	300kHz	kHz
CRU noise figure	8	dB
total CRU noise power over thermal noise power	20 [42]	dB
λ	74.35	cm

The performance of the non-cooperative and cooperative approaches are compared under the assumption of different room sizes. We consider a $3\times3\mathrm{m}$, a $5\times5\mathrm{m}$ and a $7\times7\mathrm{m}$ room. (The corresponding room area (A) values are $9\mathrm{m}^2$, $25\mathrm{m}^2$ and $49\mathrm{m}^2$,

DOI:10.15774/PPKE.ITK.2009.001

respectively.) In order for the distance of the patient and the closest CRU to have a mean of approx. 1m, the spatial density of the CRUs (ρ) was set to 0.36m^{-2}, 0.34m^{-2} and 0.31m^{-2}, respectively. (These values were determined by trial and error.) In terms of CRU antenna polarization, two cases are investigated. For the first one, CRUs are equipped with circularly polarized antennas, while for the second one, CRU antennas are linearly polarized with a uniformly random orientation.

Figure 4.4 describes the minimum implant transmission power ($P_{\text{I}}^{\text{coop}}$) as a function of the number of relaying CRUs (n) under the assumption of infinite total relay power ($P_{\text{CRU}} = \infty$). (The additional parameter values – not yet defined above – are listed in Table 4.IV.) The minimum implant transmission power for the non-cooperative approach ($P_{\text{I}}^{\text{non-coop}}$) is also plotted – at the tick labeled 'SRU' – in the same diagram. (In case the value of variable n is higher than the number of non-gateway CRUs in the room, i.e. $N-1$, the actual number of relaying CRUs is limited to $N-1$.) The diagram shows that the transmission power of the implant can considerably be reduced by employing the proposed cooperative strategy. We can observe, in addition, that a slightly better performance can be obtained if the CRUs are supplied with circularly polarized antennas. Figure 4 also shows that the minimum implant transmission power decreases as the room size increases. Furthermore – while, for the two larger rooms, the performance improves as the number of relaying CRUs grows – in the case of the smallest room, applying more than one relaying CRUs does not result in an additional decrease in transmission power. Both of the latter phenomena are certainly due to the fact that the larger the room, the more CRUs are installed providing a higher degree of diversity.

Figure 4.5 describes the lifetime gain (G_{lifetime}) as a function of the number of relaying CRUs (n) for $P_{\text{CRU}} = \infty$ and for circularly polarized CRU antennas. The lifetime gain for the non-cooperative approach – which is by definition equal to 0 – is also plotted as a reference in the same diagram. It should be noted that the lifetime of the implant can significantly be prolonged by applying the cooperative communication approach described above. We also find that – though the minimum transmission power of the implant is approx. 3dB lower for the $7\times7\text{m}$ room than for the $5\times5\text{m}$ room (Figure 4.4) – the difference in lifetime gain between the two larger rooms is not that pronounced.

DOI:10.15774/PPKE.ITK.2009.001

The reason for that is when the transmission power of the implant is around -15dBm or lower, the power consumption of the implant (at transmission) is already dominated by the circuit power consumption ($P_{\text{circ,t}}$) rather than the consumption of the RF PA ($\eta_{\text{PA}}^{-1}P_{\text{I}}$).

Figure 4.6 shows the minimum implant transmission power ($P_{\text{I}}^{\text{coop}}$) as a function of the number of relaying CRUs (n) for $P_{\text{CRU}} < \infty$, for circularly polarized CRU antennas and for a room size of $5\times5\text{m}$. The total transmission power of the relaying CRUs (P_{CRU}) power was set to -30dB and -40dB relative to $P_{\text{circ,t}}$ (circuit power consumption) multiplied by the actual number of relaying CRUs. The simulations were performed under the assumption of different power allocation strategies (optimal or uniform) and relay selection criteria (complex or simple). The curve corresponding to $P_{\text{CRU}} = \infty$ is also plotted as a reference. Our first observation is that when the transmission power of the relaying CRUs is set to approx. -30dB (1000 times) lower than $P_{\text{circ,t}}$, the performance is practically identical to that of the system with $P_{\text{CRU}} = \infty$, no matter what power allocation strategy and relay selection criterion we choose. (This obvious and remarkable difference in the performance of the implant-to-CRU$_i$ and CRU$_i$-to-CRU$_j$ communication channels is mainly due to the fact that the latter channels are not subject to the losses caused by the body tissues.) As a consequence, the power allocation stage can be omitted from the proposed cooperative communication scheme, while in terms of relay selection, it is sufficient to use the simple criterion. On the other hand, when the transmission power of the relaying CRUs is set to approx. -40dB relative to $P_{\text{circ,t}}$, an apparent difference in performance is found for the different power allocation strategies, while the different relay selection criteria – except for $n = 1,2$ – show practically the same performance. Nevertheless, in order for the choice of the power allocation strategy to have a considerable impact on the power consumption of the CRUs, the propagation conditions between the CRUs would need to be much more adverse than assumed in this study. (The same holds for the choice of the relay selection criterion concerning system performance.) Such conditions could perhaps occur if the room (or the part of the building) under investigation were – as an example – extremely densely furnished.

DOI:10.15774/PPKE.ITK.2009.001

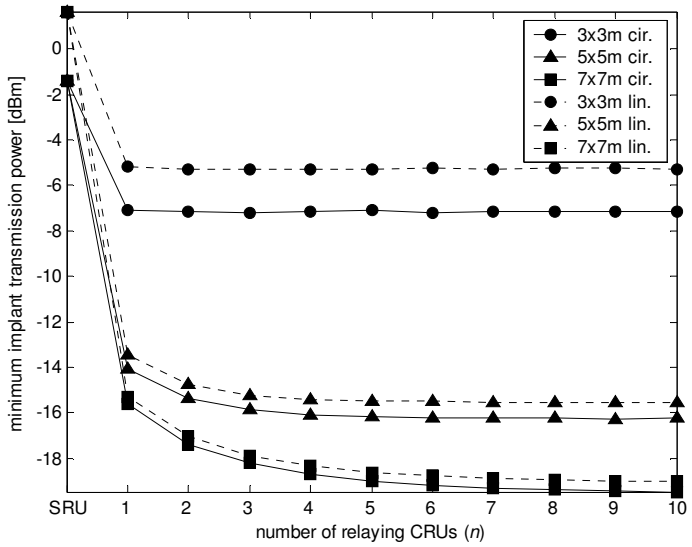

Figure 4.4. Minimum implant transmission power vs. number of relaying CRUs (n) ($P_{CRU}=\infty$). The case of the traditional, non-cooperative approach is also plotted and referred to as 'SRU'. The different curves belong to different room sizes and CRU antenna polarizations (circular or linear).

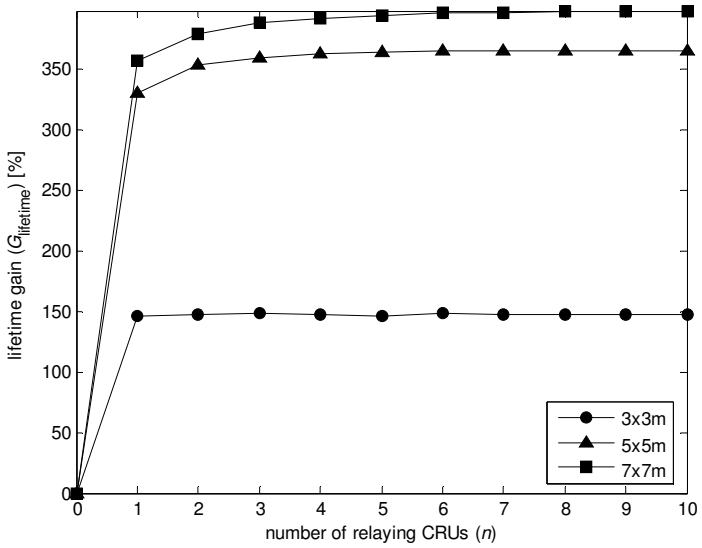

Figure 4.5. Lifetime gain ($G_{lifetime}$) vs. number of relaying CRUs (n) ($P_{CRU}=\infty$, circular CRU antenna polarization). The case of the traditional, non-cooperative approach is also plotted and referred to as 'SRU'. The different curves belong to different room sizes.

DOI:10.15774/PPKE.ITK.2009.001

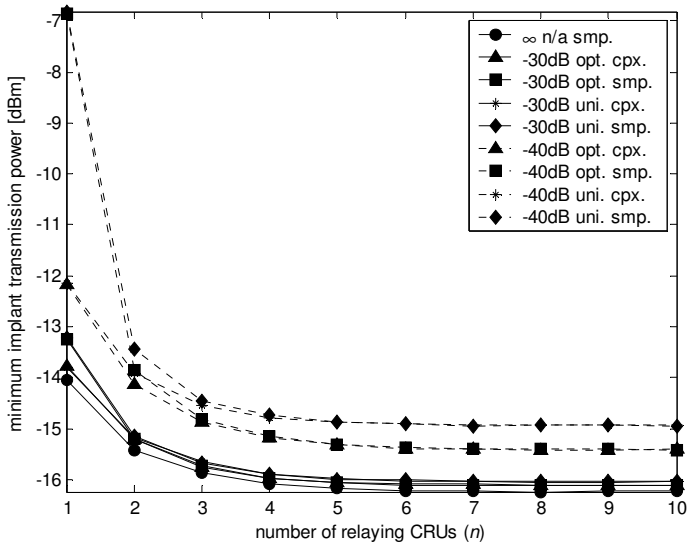

Figure 4.6. Minimum implant transmission power vs. number of relaying CRUs (n) ($P_{CRU}<\infty$, circular CRU antenna polarization, 5x5m room). The different curves belong to different PCRU values, power allocation strategies (optimal or uniform) and relay selection criterias (complex or simple). The curve corresponding to $P_{CRU}=\infty$ is also plotted as a reference.

4.6. Conclusions

The investigations above have demonstrated that the power consumption of the implant – stemming primarily from the data communication towards the outside world – can significantly be reduced by means of the proposed cooperative communication scheme. Since it is a plausible assumption that the presence of the patient is limited to only a couple of rooms during the day, this reduction in power consumption may also result in an actual and considerable lifetime gain for the implant in practice. In addition to extending the lifetime of the implant, the suggested collaborative procedure offers the alternative benefit of a more frequent and more intensive data communication between the implant and the outside world. Finally, the proposed method provides more comfort to the patient as the SRU does not need to be worn on his/her belt all the time or carried with him/her upon changing his/her location in the room. (Remark that, when worn on the belt or placed off-body close to the patient, the SRU is apparently in the near-field of the implant. For this reason, it is difficult to tell – without proper calculations or measurements – whether, at all, it is more advantageous to attach the SRU to the patient or it is more beneficial to install it off-body.)

DOI:10.15774/PPKE.ITK.2009.001

The numerical results presented in the previous section were produced based on an approximate radio propagation model and, as a consequence, the exact figures should be considered with some caution. Nonetheless, as the order of magnitude of the transmission power levels calculated for the implant are believed to be correct, this study will hopefully stimulate further and more detailed investigations of the subject in the future.

The verification of the numerical results could be accomplished either by computer simulations or by experiments. In light of the former studies in the field, the FDTD method is an appropriate choice for the simulations. With the continuous increase in the available computational power, these calculations could be run even for areas larger than a single room.

The experimental verification of the results is less obvious. The implantation requires an invasive intervention, and, for this reason, it should be preceded by a more extensive and accurate simulation work even if the surgery is carried out – for instance – only on a (healthy) animal. One option is to employ patients with an implant already in their body. The problem with this alternative is that it demands the development of customized external devices that conform to both the proposed cooperative communication scheme and the specific implanted device in question.

A kind of qualitative verification of the results, however, could easily be accomplished without an actual implantation. Assuming the availability of a number of ordinary, programmable radio transceivers such as wireless sensor nodes, one could be installed on-body – simulating the implant – while the others would be placed off-body in the experimental room. (In order to imitate the massive losses due to the body tissues for the in-body to off-body communication link, the transmission power of the on-body transmitter could be set a few tens of dBs lower than that an actual implant would have.) As also the radiation pattern of on-body transmitters was measured to show large variations [37], the experimental arrangement – just outlined above – may also be suitable for demonstrating the advantages of cooperative diversity in medical implant communications.

DOI:10.15774/PPKE.ITK.2009.001

4.7. Applications

The cooperative communication procedures described in this chapter can primarily be utilized in those telemedicine and telecare applications (e.g. BIOTRONIK Home Monitoring [44]) in which the implanted wireless sensors are of limited energy and non-rechargeable. Such implants are – for instance – the pacemakers and cardioverter defibrillators (pl. Biotronik Philos II DR-T pacemaker and Lumax DR-T ICD). In addition, the procedures can be well employed in such possible applications in which the implanted sensors can be recharged through – as an example – inductive coupling. In these cases, the frequency of the necessary recharges can significantly be decreased, which obviously improves the quality of life of the person carrying the implant.

DOI:10.15774/PPKE.ITK.2009.001

Chapter Five

EFFICIENT, DISTRIBUTED, MULTIPLE-RELAY SELECTION PROCEDURES FOR COOPERATIVE COMMUNICATIONS

5.1. Introduction

Cooperative communication schemes [38] for wireless networks without central control require a distributed procedure in order to select the relays that assist the communication between the source and destination. The problem of distributed relay selection was addressed for the first and – in terms of failure probability and duration – the only time so far in [40] and [62]. The algorithm introduced in these papers to select a single opportunistic relay forms the base of the forthcoming investigations. It is based on the method of distributed timers and assumes only channel reciprocity and the availability of instantaneous channel measurements. It does not require either a priori information on the network geometry or explicit communication between the potential relays.

Open problems

As already mentioned above, the problem of distributed relay selection has only been investigated for a single relay so far ([40] and [62]). In this study, the work presented in [40] and [62] is extended for the case of multiple-relay selection and a numerical analysis of the problem is provided for various relay selection criteria and timer settings (distortion functions) for the first time. (Remark that a quite similar distributed method is used to select three wireless sensor nodes to participate in a target localization task in [63].)

In the forthcoming investigations, the performance of the different relay selection procedures is analyzed in connection with a somewhat different cooperative communication scenario than the one presented in [40] and [62]. While, in the latter works, a unicast communication scenario with deterministic node locations and equal power for the source and relay was investigated, here an anycast cooperative communication scheme with random node distribution and transmission power balancing between the source and relay nodes is being studied. In this scheme, the packet sent by the energy constrained bottleneck (source) node (BN) is simultaneously received by a

DOI:10.15774/PPKE.ITK.2009.001

couple of less energy constrained nodes in the proximity of the BN. (The latter nodes will henceforth be referred to as 'energy abundant' nodes, EANs.) After that, a processing node (PN) is selected among these nodes in a distributed fashion along with a number of $n-1$ additional cooperating (relay) nodes (EAN_2, ..., EAN_n). Afterwards, the power allocation procedure takes place, for which refer to [59][61]. Finally, the selected relay nodes amplify and forward their received analogue signal to the PN, which then optimally combines the received waveforms. This scenario is termed as anycast since the BN packet is not intended for a specific node in the network, i.e. any of the cooperating nodes can be considered as destination.

5.2. The model

It this section, the network topology, radio propagation model as well as the different relaying schemes applied in the forthcoming investigations are discussed in brief.

5.2.1. Network topology

The EANs are distributed around the BN according to a 2D homogeneous Poisson distribution with parameter λ. Note that, for a large area, this approximates a uniform distribution.

5.2.2. Radio propagation environment

It is assumed that the radio channels between the BN and the EANs and those between the EANs are subject to independent, flat Rayleigh fading, i.e. the corresponding $r_{\mathrm{BN,EAN}_i}$ and $r_{\mathrm{EAN}_i,\mathrm{EAN}_j}$ fading factors are independent, exponentially distributed random variables with unit mean. It is also assumed that the fading is slow enough for the transmitted symbols within a single period of the communication cycle (including the BN transmission, the node selection procedure, the power allocation procedure and the relaying phase) to experience the same channel gains. Finally, the large-scale mean path loss is considered to be proportional to a specific negative α power of the distance between the transmitter and the receiver. (For specific values of α, please refer to Section 5.4.)

To summarize, for the BN-to-EAN_i and the EAN_i-to-EAN_j channel gains ($h_{\mathrm{BN,EAN}_i}$ and $h_{\mathrm{EAN}_i,\mathrm{EAN}_j}$), the following equations hold, respectively:

DOI:10.15774/PPKE.ITK.2009.001

$$\left|h_{\mathrm{BN,EAN}_i}\right|^2 = A \cdot d_{\mathrm{BN,EAN}_i}^{-\alpha} \cdot r_{\mathrm{BN,EAN}_i}, \tag{5.1}$$

$$\left|h_{\mathrm{EAN}_i,\mathrm{EAN}_j}\right|^2 = A \cdot d_{\mathrm{EAN}_i,\mathrm{EAN}_j}^{-\alpha} \cdot r_{\mathrm{EAN}_i,\mathrm{EAN}_j}, \tag{5.2}$$

where $d_{\mathrm{BN,EAN}_i}$ and $d_{\mathrm{EAN}_i,\mathrm{EAN}_j}$ is the distance between the BN and EAN_i and that between EAN_i and EAN_j, respectively, while A is a propagation model dependent constant.

5.2.3. Relaying schemes

In the case of the simultaneous transmission based scheme the selected relay nodes (EAN_2, ..., EAN_n) send their packet to the PN (EAN_1) simultaneously. Assuming phase adjustment between the relays, the SNR of the relayed BN packet – under the assumption of optimal power allocation in the relaying stage [59] – has the form of

$$SNR_{\max} = \sum_{i=2}^{n} \frac{P_{\mathrm{BN}}}{\sigma_{\mathrm{n}}^2} \frac{\left|h_{\mathrm{BN,EAN}_i}\right|^2 \left|h_{\mathrm{EAN}_i,\mathrm{PN}}\right|^2}{\left(\left|h_{\mathrm{BN,EAN}_i}\right|^2 + \sigma_{\mathrm{n}}^2 P_{\mathrm{BN}}^{-1}\right)\eta^{-1} + \left|h_{\mathrm{EAN}_i,\mathrm{PN}}\right|^2}, \tag{5.3}$$

σ_{n}^2 being the variance of receiver noise, P_{BN} the transmission power of the BN and η the total transmission power of the relays relative to P_{BN}.

In the case of the so called repetition based scheme, the selected relay nodes send their packet to the PN, one after the other. As the expression of $SNR_{\max}$ for this scheme – corresponding to optimal power allocation – is rather complex, it is not presented here. For details, though in a somewhat different context, please refer to [61].

5.3. Node selection procedure

The distributed node selection procedure is constituted by two consecutive stages.

5.3.1. Stage 2: distributed PN selection procedure

As soon as the BN terminates its transmission, stage 2 is started and each of the EANs that received the BN packet starts a countdown timer with an initial value inversely

DOI:10.15774/PPKE.ITK.2009.001

proportional to its received SNR from the BN. Then these nodes listen to the channel and transmit a 'buzz' (a very short signal) as soon as their timer expires, provided that they did not hear a buzz up to that moment. The EAN transmitting the one and only buzz in this stage – i.e. the one with the best received signal strength – is elected as the PN. In this chapter the focus is on the performance of the relay selection procedure, hence – though they can be obtained in a similar fashion – the results for the PN selection procedure are not presented here.

5.3.2. Stage 2: distributed, multiple-relay selection procedure

Stage 2 is initiated at the end of the buzz of the recently selected PN. The EANs involved in the relay selection procedure start their countdown timer again. However, this time the initial value of the timer is set inversely proportional to a value given by a more complex selection criterion that takes into account the gain of both the BN-to-EAN$_i$ and the EAN$_i$-to-PN channels. The specific relay selection criteria investigated in this study as well as the way the initial value of the timers are exactly set are going to be discussed in details in the subsequent paragraphs. (Here, differently from the previous parts, EANs other than the PN are indexed according to their distance to the BN.) After that they listen to the channel again and transmit a buzz as soon as their timer expires, provided that they heard less than $n-1$ buzzes up to that moment (in stage 2). Stage 2 terminates when the last (i.e. the $(n-1)^{\text{th}}$) node transmits its buzz. In this manner, the EANs with the lowest initial timer value, i.e. the ones that are the most suitable for relaying according to a given selection criteria are selected.

Concerning the relay selection criterion, three cases are investigated. For the criterion called *opportunistic*, nodes set the initial value of their timer inversely proportional to the expression behind the summation sign in (5.3), $SNR_{\text{BN}\to\text{EAN}_i\to\text{PN}}$, which is the received SNR of the BN-to-EAN$_i$-to-PN communication channel for the hypothetic case when all of the total relay power is allocated to EAN$_i$. (EAN$_i$ can estimate the EAN$_i$-to-PN channel gain based on the fixed-power buzz transmitted by the PN assuming that the channel is reciprocal.) The same criterion is used to select a single opportunistic relay (hence its name) in [60]. In the case of the criterion named *minimum*, the timers are set to a value inversely proportional to

DOI:10.15774/PPKE.ITK.2009.001

$$\min\left\{\left|h_{\mathrm{BN,EAN}_i}\right|^2,\left|h_{\mathrm{EAN}_i,\mathrm{PN}}\right|^2\right\}, \tag{5.4}$$

which is the minimum of the (power) gains of the source-to-relay and relay-to-destination channels [40][62]. Finally, for the criterion termed *harmonic*, nodes set their timer inversely proportional to

$$2\left(\left|h_{\mathrm{BN,EAN}_i}\right|^{-2}+\left|h_{\mathrm{EAN}_i,\mathrm{PN}}\right|^{-2}\right)^{-1}, \tag{5.5}$$

which is the harmonic mean of the (power) gains of the source-to-relay and relay-to-destination channels [40][62]. (Note that the latter two criteria have the property that – under the assumption of deterministic node locations and Rayleigh-faded radio channels – the probability density function of the initial value of the timers can be expressed in a closed form [40][62] when no distortion function is applied.)

Based on (5.3), one can immediately find that selecting the nodes with the highest $SNR_{\mathrm{BN}\to\mathrm{EAN}_i\to\mathrm{PN}}$ is the optimal node selection criterion for the simultaneous transmission based scheme. Considering the expression of the minimum and harmonic criteria, we see that these criteria do not take into account – among others – the possibility that the transmission power of the relays (EANs) may be considerably higher than that of the source (BN). In that case, as a result, EAN_i with $\left|h_{\mathrm{BN,EAN}_i}\right|^2 << \left|h_{\mathrm{EAN}_i,\mathrm{PN}}\right|^2$ may get ranked higher than EAN_j with $\left|h_{\mathrm{EAN}_j,\mathrm{PN}}\right|^2 << \left|h_{\mathrm{BN,EAN}_i}\right|^2 < \left|h_{\mathrm{BN,EAN}_j}\right|^2$, when in fact the ranking of these relays should be performed just the other way around as – due to the relatively high transmission power of the relays – $SNR_{\mathrm{BN}\to\mathrm{EAN}_i\to\mathrm{PN}}$ is determined mostly by the source-to-relay channel gain in both cases. This in the end leads to a performance degradation in terms of the gain of the cooperation.

Since the expression of $SNR_{\max}$ for the repetition based scheme is fairly complicated and most probably cannot be 'decoupled' as (5.3), the optimal relay selection criterion for this scheme remains an open research problem.

DOI:10.15774/PPKE.ITK.2009.001

As the RX-to-TX switch time of real radios is non-zero [40][62], there is a finite probability that while a node with a lower initial timer setting is switching from RX to TX mode (in order to transmit its buzz), the timer of a node with a higher initial timer value expires. In this case each of the two nodes will assume the same role for itself and also their buzzes may collide, which means that the relay selection procedure fails. Accordingly, the event of failure can be expressed as

$$\bigcup_{i=1}^{n-1}\left(T_{i+1}-T_i<c\right), \tag{5.6}$$

where T_i is the initial value of the timer of the node with the i^{th} lowest initial timer value in stage 2 and c is the RX-to-TX switch time of the radios. Here it was assumed that all the potential relays can hear each other (they use sufficiently high transmission power for their buzzes) and that the distances between the nodes in the network are low enough for the propagation delay to be neglected compared to c. If some of these conditions are not met, the very same treatment (5.6) can still be used, only some additional terms – such as the maximum propagation delay between the relays – need to be incorporated into c as described in [40] and [62]. (Under the assumption that the duration of the buzzes – which can be as short as a bit [40] – is not higher than c, the complement of (5.6) is not only a necessary but also a sufficient condition for the buzzes not to collide.)

According to the above, the mean duration of the relay selection procedure to be evaluated in the simulations is defined as

$$D=\mathrm{E}\{T_{n-1}\}. \tag{5.7}$$

The initial value of the timers in [40] and [62] is set directly proportional to the reciprocal of the selection criteria (5.4) and (5.5). However, it is a reasonable assumption that we might obtain a better performance if we use a suitable monotonic increasing function of the reciprocal of the selection criteria to set the timers. (This way the order of selection is left unchanged.) In the present investigations two such distortion functions are used:

DOI:10.15774/PPKE.ITK.2009.001

$$f_1(x) = x^{\frac{2}{\alpha}}, \tag{5.8}$$

$$f_2(x) = \left(\sqrt{1 + 4x \left(\sqrt{\eta} + \frac{1}{\sqrt{\eta}} \right)^{-2}} - 1 \right)^{\frac{2}{\alpha}}. \tag{5.9}$$

(For specific values of α and η, please refer to Section 5.4.) The expression of f_2 in (5.9) is derived by developing a quite a rough approximation of the mean of the reciprocal of $SNR_{\mathrm{BN}\to\mathrm{EAN}_i\to\mathrm{PN}}$ as a function of i and then taking the inverse of that. The intuition behind this idea is that if we distribute the initial value of the timers evenly the probability of failure will be minimized. For the detailed derivation of (5.9), please refer to Appendix C. The expression of f_1 in (5.8) is obtained similarly but for (5.4) and (5.5). (It can be easily seen that, for large values of η, (5.9) is directly proportional to (5.8)).

5.4. Simulation results and discussion

As the model set up in the preceding sections is most probably analytically intractable, the performance analysis is carried out numerically using Monte Carlo simulations. For details on how Monte Carlo simulations are performed in this study, please refer to Appendix B.

Figure 5.1 and Figure 5.2 shows the failure probability of the relay selection procedure as a function of the mean duration of the procedure for the different selection criteria and distortion functions. The value of parameters n, α, η, P'_{BN} were set to 10, 4, $10\mathrm{dB}$, $5\mathrm{dB}$ and 2, 4, $30\mathrm{dB}$, $10\mathrm{dB}$, respectively. (For the definition of P'_{BN} please refer to the Appendix.) It can be seen on the diagrams that all criteria perform significantly better if some (suitable) distortion function is used. The most remarkable performance enhancement – which is around an order of magnitude – is observed in the case of the opportunistic criterion and $n = 10$ (Figure 5.1). Comparing the performance achieved with the different distortion functions for this criterion, we find that for $\eta = 10\mathrm{dB}$ (Figure 5.1), applying distortion function f_2 does result in a lower failure probability than applying f_1, while for $\eta = 30\mathrm{dB}$ (Figure 5.2), no difference can be noted. The latter is

DOI:10.15774/PPKE.ITK.2009.001

certainly due to the fact that, for high values of η, the two functions are virtually equivalent. As expected, for a given value of the mean duration, the probability of failure is much lower for $n = 2$ (a single relay – Figure 5.1) than for $n = 10$ (a number of 9 relays – Figure 5.2). The explanation for that is the higher n is, the more random variables need to be squeezed into the same time interval and the higher the probability is that two of these get closer to each other than a predefined value (c). We can also observe on both diagrams that the minimum and the harmonic criteria perform practically in the same way both when no distortion function and when distortion function f_1 is used. Note that this is not in contradiction with the results presented in [40] and [62], where the former criterion was found to perform better than the latter one. For this, the reason is that in these papers – instead of the mean duration of the selection procedure – it is the factor applied to the reciprocal of (5.4) and (5.5) what was kept constant during the comparison.

It is found in Figure 5.1 that all three criteria require no more than 10000 time units (measured in c) to achieve a failure probability lower than 1%. For sophisticated radios $c \approx 1\mu s$ [40], which results in a mean duration of 10ms even when the number of cooperating nodes is $n = 10$. The latter time interval is much shorter than the coherence time of the channel, which, for a mobility of $1 ms^{-1}$ (corresponding to walking people) and a carrier frequency of $433MHz$, is approximately $100ms$.

DOI:10.15774/PPKE.ITK.2009.001

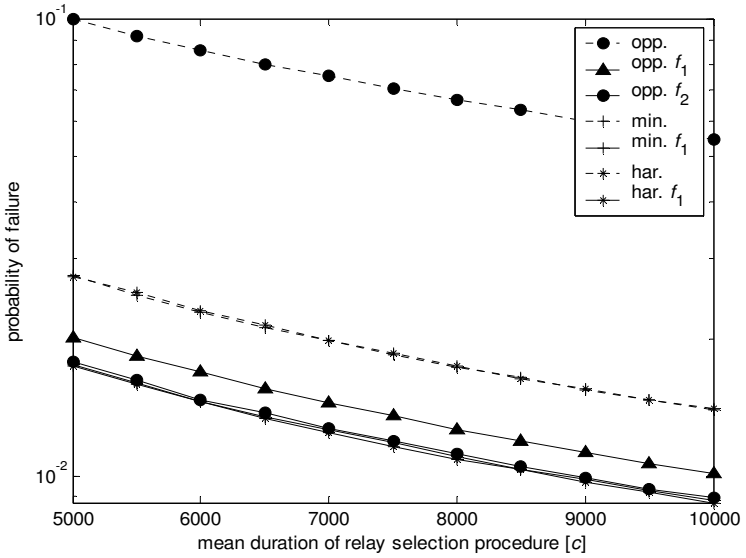

Figure 5.1. Probability of failure ($P\left\{\bigcup_{i=1}^{n-1}(T_{i+1}-T_i<c)\right\}$) as a function of the mean duration of the relay selection procedure (n=10 – a number of 9 relays, α=4, η=10dB, P'_{BN} =5dB).

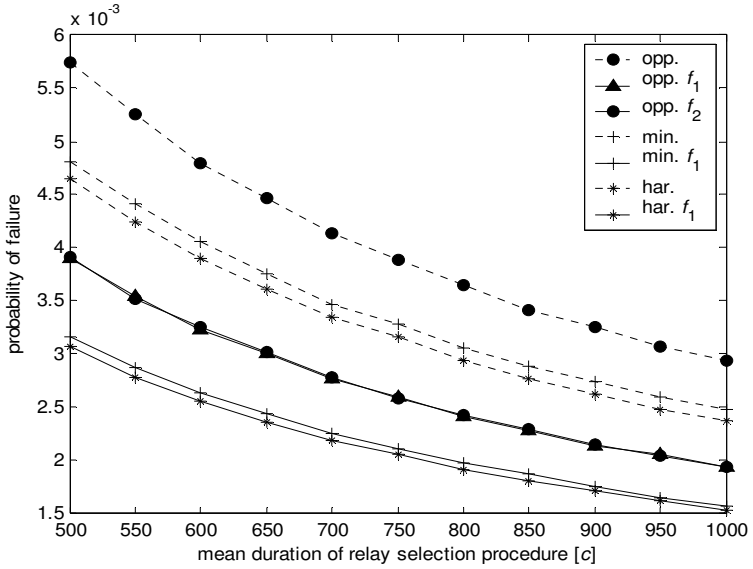

Figure 5.2. Probability of failure ($P\left\{\bigcup_{i=1}^{n-1}(T_{i+1}-T_i<c)\right\}$) as a function of the mean duration of the relay selection procedure (n=10 – a single relay, α=4, η=30dB, P'_{BN} =10dB).

The gain of the different cooperative communication schemes was evaluated in terms of the achievable lifetime gain for the BN. This quantity is defined as the relative difference in BN lifetime between the cases with ($n>1$) and without relays ($n=1$). It is assumed that the power consumption of the BN is dominated by the consumption of its radio

DOI:10.15774/PPKE.ITK.2009.001

frequency power amplifier, and as a consequence, the lifetime of the BN is assumed to be directly proportional to the reciprocal of its transmission power. The value of the latter quantity was determined under the reliability constraint that the overall (uncoded) mean bit error probability of the BN-to-EANs communication link is not higher than a predefine parameter, $\overline{BEP}_{\text{req}}$. Figure 5.3 describes the lifetime gain as a function of the number of relay nodes for the different relay selection criteria and relaying schemes. The value of parameters α, η and $\overline{BEP}_{\text{req}}$ was set to 4, 30dB and 10^{-2}, respectively, while for the modulation, BPSK was assumed. Now, if we fix the mean duration of the selection procedure and consider the case of $n = 2$, we can conclude based on Figure 5.2 and Figure 5.3 that there is a trade-off between the gain of the cooperative communication scheme and the failure probability of the involved distributed relay selection procedure. That is to say, the opportunistic criterion performs better than the others in terms of lifetime gain, while the situation is just the opposite when it comes to failure probability. The reason for the former finding is already given in the previous section, while the latter is probably due to the fact that the variance of the initial value of the timers relative to the difference of their mean is lower for the minimum and harmonic criteria than for the opportunistic one.

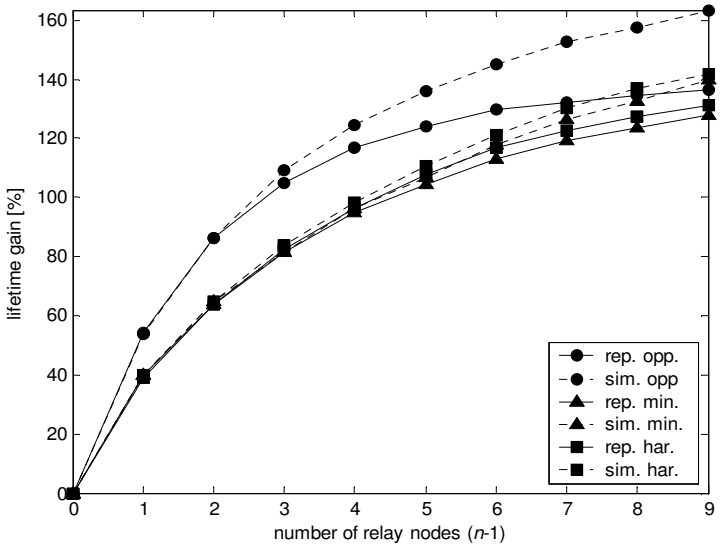

Figure 5.3. Lifetime gain as a function of the number of relays (α=4, η=30dB, $\overline{BEP}_{\text{req}}=10^{-2}$, BPSK modulation).

DOI:10.15774/PPKE.ITK.2009.001

5.5. Conclusions

The investigations above have demonstrated that the failure probability of the above described distributed relay selection procedure can considerably be decreased by applying a suitable distortion function that transforms the reciprocal of the relay selection criterion into initial timer value. The results have also shown that – for slow-fading environments – the mean duration of the selection procedure can be much shorter than the typical coherence time of the channel even if multiple relays are to be selected. Finally, we have provided an example that the choice of the selection criterion may imply a trade-off between the cooperation gain and the failure probability of the involved distributed relay selection procedure.

The distortion functions proposed here – though they do improve the performance of the selection procedure – are clearly suboptimal. In order to assess how much additional performance enhancement can be achieved, an attempt was made to further optimize the distortion function with simulated annealing using a piecewise linear approximation. The preliminary results showed that – for practically interesting failure probability values ($<10^{-2}$) – no improvement could be obtained in a reasonable computation time. This was due to the fact that – for the lack of a closed form expression for the failure probability – the value of the cost function (i.e. failure probability) had to be determined by stochastic sampling (Monte Carlo simulations). (If the precision of stochastic sampling was lowered the optimization algorithm executed in a plausible running time, however, no improvement relative to the suboptimal solutions could be obtained in this case as the estimate of the cost function was rather rough during the optimization). For this reason, it is pointed out as a future plan to set up a cost function which adequately estimates system performance and whose possible components – such as the mean and variance of the initial timer values – can faster be computed.

5.6. Applications

The distributed multiple-relay selection procedures analyzed in this chapter can be applied in such specific cooperative communication scenarios in which employing multiple relays yields a substantial cooperation gain and the coherence time of the channel is high enough to accommodate a relatively longer relay selection procedure and a possibly longer relaying stage (repetition based relaying). The latter condition is

DOI:10.15774/PPKE.ITK.2009.001

fulfilled if the movement of the communicating parties as well as that of the interacting objects in the radio propagation environment are relatively slow. (Such conditions can occur – as an example – in indoor environments, where the fastest objects are typically walking people.) Please note, however, that the results in this study do not necessarily apply to an arbitrary cooperative communication scenario as they were determined assuming a specific topology-, radio propagation-, power consumption- and communication model (anycast). As a consequence, the performance of the distributed, multiple-relay selection procedures should be determined and optimized for each individual case.

DOI:10.15774/PPKE.ITK.2009.001

Chapter Six

CONCLUSIONS AND APPLICATIONS

The aim of this dissertation has been to enhance the performance of sensor systems made up by emerging sensor devices and existing communication and information technologies. Accordingly, in this dissertation, several different solutions have been proposed to improve the performance of such sensor systems. In the first part of the study, it has been demonstrated that the sensitivity of nanoantenna-coupled tunnel-diode detectors – used in high-speed, multicolor, THz and infrared cameras – can considerably be improved by applying metal-insulator-metal diodes with double-layer insulators in these devices. In the second part of the dissertation, stochastic energy balancing packet forwarding protocols have been investigated and shown to significantly prolong the lifetime of wireless sensor networks. Finally, in the last part, the author has demonstrated that the performance of medical implant communication systems can greatly be enhanced using cooperative communication techniques.

The high-speed, multicolor, THz and infrared cameras can primarily be applied to identify and to discriminate fast moving objects as well as to observe fast chemical and biological processes.

The energy aware packet forwarding protocols can be used in a wide range of wireless sensor network applications such as intelligent home (e.g. the NETVOX Smart House System), mechanical structure monitoring (e.g. the Sensametrics Wireless Structure Monitoring), environment monitoring (e.g. the Advanced Sensor Technologies Root Zone Intelligence System) and seismic activity monitoring.

The cooperative communication procedures described in the third group of the theses can primarily be utilized in those telemedicine and telecare applications (e.g. BIOTRONIK Home Monitoring [44]) in which the implanted wireless sensors are of limited energy and non-rechargeable. Such implants are – for instance – the pacemakers and cardioverter defibrillators (pl. Biotronik Philos II DR-T pacemaker and Lumax DR-T ICD). In addition, the procedures can be well employed in applications in which the

DOI:10.15774/PPKE.ITK.2009.001

implanted sensors can be recharged through – as an example – inductive coupling. In these cases, the frequency of the necessary recharges can significantly be decreased, which obviously improves the quality of life of the person carrying the implant.

DOI:10.15774/PPKE.ITK.2009.001

Chapter Seven

SUMMARY

7.1. Main findings and results

In this dissertation, various solutions have been proposed to enhance the performance of present and future sensor systems. Firstly, it has been demonstrated that the sensitivity of nanoantenna-coupled tunnel-diode detectors can considerably be improved by applying metal-insulator-metal diodes with the double-layer insulators in these devices. Secondly, it has been shown that the lifetime of wireless sensor networks can significantly be prolonged by means of the proposed optimal statistical energy balancing packet forwarding protocols. Thirdly and finally, the author has demonstrated how the performance of medical implant communication systems can be enhanced using cooperative communication techniques.

7.2. New scientific results

Thesis I: Investigation of the quality factor of metal-insulator-metal diodes with double-layer insulators.

(The author's related publications: [3][6].)

The sensitivity of detector type radio receivers is directly proportional to the quality factor of the diode used as the detector. The quantity just mentioned is defined as the ratio of the second and the first derivatives of the current-voltage characteristics. Metal-insulator-metal diodes coupled to nanoantennas form miniaturized detector type radio receivers, which can be used to detect electromagnetic radiation in the THz and infrared domains. The results of the present thesis are related to the sensitivity of the above mentioned detector.

DOI:10.15774/PPKE.ITK.2009.001

I.1. Using a single electron Shrödinger equation and the corresponding QTBM boundary conditions, I have determined the quality factor of metal-insulator-metal diodes with double-layer insulators as a function of the insulator thicknesses, metal-insulator work functions and the dielectric constants of the insulators.

I calculated the quality factor as a function of the following five diode parameters: the total thickness of the insulators (L), the ratio of the insulator thicknesses (r_d), the ratio of the dielectric constants of the insulators, the average metal-insulator work function and the asymmetry of the work functions. The results have among others shown that the ratio of the insulator thicknesses has a fundamental impact on the magnitude of the quality factor (Figure 2.9). In the case of aluminum oxide and silicon dioxide – two insulators widely used in CMOS technology – the optimum thickness ratio has proven to be approximately 0.6 , the implementation of which can be considered feasible even for a total thickness of only a few nanometers.

I.2. I have compared the quality factor of metal-insulator-metal diodes with single- and double-layer insulators and found that the latter may exceed the former by around an order of magnitude.

When carrying out the comparison, I assumed that the resistance of the diode is given in advance as a design parameter. I determined the maximum quality factor for both types of diodes assuming parameter values that are typical and feasible in CMOS technology. My investigations have shown that – thanks to the asymmetry in the dielectric constants – nanoantenna-coupled MIM diode detectors with double-layer insulator diodes may outperform the conventional detectors – with single-layer insulator diodes – by around an order of magnitude in terms of sensitivity.

DOI:10.15774/PPKE.ITK.2009.001

Thesis II: Optimization of energy aware protocols for one-dimensional network topology.

(The author's related publications: [2].)

The elements of wireless sensor networks are typically of limited energy. This energy is completely depleted with time and the sensor node in question ceases to operate. In wireless sensor networks, sensed data is usually collected at a base station (BS). The network elements may forward their data to the BS either directly or indirectly with the assistance of other network elements. The choice of the packet forwarding strategy has a fundamental impact on the energy consumption of network elements and as a result, on network lifetime. The results of the present thesis are related to packet forwarding protocols achieving maximum network lifetime.

My investigation involved the following packet forwarding protocols. For the chain protocol, node i forwards its packets to node $i-1$ (i.e. the nearside neighbor), while in the case of the shortcut protocol, packets are sent directly to the BS (i.e. node 0). When the 'random shortcut' protocol is applied as the packet forwarding strategy (Figure 3.2), each node tosses a 'loaded' coin and – based on the result of the toss – decides to forward its packet either to its nearside neighbor or to the BS. (Variable a_i denotes the probability that the packet is sent to the nearside neighbor.) Finally, in the case of 'sending packets to any node ahead in the chain' protocol, a node may choose to forward its packet to any of its nearside neighbors including the BS. (Variable a_{ij} denotes the probability that node i forwards its packets to the node $j-1$.)

The lifetime of the network can be defined in various ways. In my investigations I considered the network 'alive' if all of its elements were operational. As packet generation and – for the 'random shortcut' and 'sending packets to any node ahead in the chain' protocols – packet forwarding are random, also the remaining energies of the nodes are random variables. As a result, a statistical interpretation of network lifetime is required. In my work, the latter quantity – denoted by Ψ – is defined as the shortest

DOI:10.15774/PPKE.ITK.2009.001

(discrete) time, after which the probability that the network is 'dead' is higher than $e^{-\alpha}$, i.e.

$$\Psi = \min\left\{K \middle| P\left(\min_i c_i(K) \le 0\right) > e^{-\alpha}\right\}, \tag{7.1}$$

were $c_i(K)$ is the remaining energy of node i t time instant K.

II.1. I have extended the performance analysis of the chain- and shortcut protocols from the case without buffers to the case with buffers. I have modeled the packet forwarding mechanism by means of Markov chains, derived the stationary distribution and based on these, I have given a quantitative description of network lifetime.

Based on the above definition, the probability that the network is no longer operational after (discrete) time K is given by,

$$P(K) = 1 - \prod_{i=1}^{N} \sum_{k=0}^{[C/\tilde{g}_i]} \binom{K}{k} \left(1-\pi_0^{(i)}\right)^k \left(\pi_0^{(i)}\right)^{K-k}, \tag{7.2}$$

where N is the number of network elements, C is their initial energy, $\tilde{g}_i$ is the mean of the energy consumed by node i when forwarding a packet, while $1-\pi_0^{(i)}$ is the probability that node i sends a packet at given time instant. For the case investigated in Figure 7.1, a network lifetime of 102% was obtained for shortcut protocol relative to the chain protocol.

II.2. I have generalized the above protocols to the 'random shortcut' protocol and derived the optimum $(\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_N)$ vector that results in minimum energy consumption and maximum network lifetime.

The operation of the 'random shortcut' protocol is shown is Figure 3.2. The probability that the packet generated at node *j* is forwarded along the route shown in the figure is

DOI:10.15774/PPKE.ITK.2009.001

$$P = (1 - a_i) \prod_{\ell=j}^{i-1} a_\ell \,. \tag{7.3}$$

The optimization problem concerning vector $\mathbf{a}$ takes the form of

$$\mathbf{a}_{\text{opt}} = \arg\max_{\mathbf{a}} \Psi(\mathbf{a}). \tag{7.4}$$

[In the discrete time case the optimization was subject to the following two constraints: (i) the maximum of the ratio of the number of dropped packets relative to the number of total received packets must no exceed a certain threshold; (ii) the mean packet delay should remain lower than a predefined value.] For the 'random shortcut' protocol, a network lifetime of 272% was obtained relative to the chain protocol, which considerably exceeds the lifetime provided by the protocols described in Theses II.1.

II.3. Based on the above, I have generalized the 'random shortcut' protocol, i.e. introdced the 'sending packets to any node ahead in the chain' protocol and derived the optimum $\mathbf{A}$ matrix that provides minimum energy consumption and maximum network lifetime.

This time, the optimization problem can be expressed as

$$\mathbf{A}_{\text{opt}} = \arg\max_{\mathbf{A}} \Psi(\mathbf{A}). \tag{7.5}$$

(The element in the i^{th} row and i^{th} column of matrix $\mathbf{A}$, i.e. a_{ij} expresses the probability that node i forwards its packets to node $j-1$.) In the case of this protocol, a network lifetime of 626% was obtained relative to the chain protocol, which significantly surpasses the lifetime achieved by any of the protocols introduced in Thesis II so far. The increase in network life, for both the 'random shortcut' and 'sending packets to any node ahead in the chain protocol, is due to following two reasons. Firstly, nodes that go flat slower can relieve the ones that get depleted faster by forwarding them fewer packets. Secondly, nodes that are exhausted faster can send

DOI:10.15774/PPKE.ITK.2009.001

their packets also to the ones that are closer to them. It this manner, the consumption of the network elements gets balanced and they are depleted practically at same time.

Comparative performance analysis

I have compared the different protocols in terms of energy consumption and obtained the following ranking (Table 7.I, Figure 7.1). The results in Figure 7.1 were obtained for an equidistant topology.

TABLE 7.I. RANKING OF THE DIFFERENT PACKET FORWARDING PROTOCOLS.

Protocol	**Ranking**
'sending packets to any node ahead in the chain'	1
'random shortcut'	2
chain	3-4
shortcut	3-4

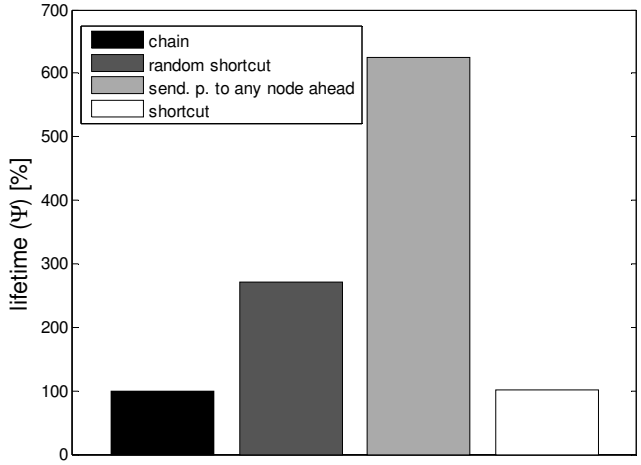

Figure 7.1. Performance of the different packet forwarding protocols.

Thesis III: Cooperative communication procedures in biomedical applications.

(The author's related publications: [1][4][5].)

It is a characteristic feature of wireless media that the physical signals conveying the information reach not only the receiver but also several other locations in space. Besides its harmful effects, e.g. interference, this phenomenon has also some benefits. Namely, one can take advantage of this effect to economize on the transmission power of the often

DOI:10.15774/PPKE.ITK.2009.001

energy-constrained transmitter. Since the signals reach not only the receiver but also several other nodes in the network, information transfer can be made more reliable if these network elements share their received signals cooperating in this manner. The results of the thesis group are related to this idea.

In our particular cooperative communication scenario, we consider a single room accommodating a patient with a medical implant inside his/her body. The packet transmitted by the implant is received by multiple battery-operated, wireless cooperating receiver units (CRUs) installed across the room. After that, a couple of the CRUs are selected for cooperation in a centralized or distributed fashion. Finally, these CRUs relay their packet to the gateway CRU (G-CRU) in the room, which then detects the packet sent by the implant and forwards it to the service centre through some traditional network (e.g. GSM, 3G, internet, etc.) (Figure 4.3).

At the beginning of the relay selection procedure known as the method of distributed timers, the potential relays set the initial value of their countdown timers according to a given selection criterion. After that they start their timers and listen to a common radio channel. As soon as the timer of one of the potential relays expires, the node transmits a buzz (a very short signal) informing the other participants that the relay was selected and the procedure terminated.

III.1. I have constructed a compact and stochastic model of the in-body to off-body communication channel based on the existing results on radio wave propagation from medical implants. Applying this model, I have shown that the lifetime of the implant can considerably be extended by means of the proposed cooperative communication scheme.

Based on the results in the literature, I applied a Rician multipath fading model, while concerning the effect of the body on the radiation characteristics I employed an axially symmetric pattern with a truncated log-normal distribution. I estimated the value of the different parameters in these models on the basis of the published results. In my investigations, I considered the location of the CRUs as well as the position and orientation of the patient (i.e. the implant) as random variables. With respect to the relaying method, I applied the solution in which the relays amplify and retransmit the

DOI:10.15774/PPKE.ITK.2009.001

received analogue signal one after the other. After the G-CRU combines the relayed packets with the packet that is received directly, the resultant signal-to-noise ratio of the combined packet – assuming optimal transmission power among the relaying CRUs – takes the form of

$$SNR_{\text{res}} = \frac{P_{\text{I}}}{\sigma_{\text{n}}^2}\left|h_{\text{I,G-CRU}}\right|^2 + \sum_{i=1}^{n_1}\frac{B_{\delta_i}}{C_{\delta_i}} - \left(\sum_{i=1}^{n_1}\frac{\sqrt{B_{\delta_i}}}{C_{\delta_i}}\right)^2\left(1+\sum_{i=1}^{n_1}C_{\delta_i}^{-1}\right)^{-1}, \tag{7.6}$$

where

$$B_i = C_i \cdot \frac{P_{\text{I}}}{\sigma_{\text{n}}^2}\left|h_{\text{I,CRU}_i}\right|^2, \tag{7.7}$$

$$C_i = \frac{P_{\text{CRU}}\left|h_{\text{CRU}_i\text{,G-CRU}}\right|^2}{P_{\text{I}}\left|h_{\text{I,CRU}_i}\right|^2 + \sigma_{\text{n}}^2}, \tag{7.8}$$

$$n_1 = \arg\max_{1\le k\le n}\sqrt{B_{\delta_{n_1}}}\,g(k), \tag{7.9}$$

$$g(k) = \left(\sum_{j=1}^{k}C_{\delta_j}^{-1}\sqrt{B_{\delta_j}}\right)^{-1}\left(1+\sum_{j=1}^{k}C_{\delta_j}^{-1}\right), \tag{7.10}$$

and

$$B_{\delta_1} \ge B_{\delta_2} \ge \ldots \ge B_{\delta_n},\ \{\delta_1,\ldots,\delta_n\} \equiv \{1,\ldots,n\}. \tag{7.11}$$

Here P_{I} is the transmission power of the implant, n and P_{CRU} are the number and total transmission power of the relaying CRUs, respectively, σ_{n}^2 is the variance of receiver noise, while $h_{\text{X,Y}}$ is the gain of the channel between node X and Y. I also incorporated a realistic power consumption model into the analysis along with the typical figures for the state-of-the-art low-power transceivers. The reliability constraint I imposed on the communication link from the implant to the outside world is formulated as

DOI:10.15774/PPKE.ITK.2009.001

$$P\left(SNR_{\text{res}}\left(P_{\text{I}}\right) < SNR_{\text{req}}\right) \leq P_{\text{out}}\,, \tag{7.12}$$

where SNR_{req} is the required signal-to-noise ratio, whereas P_{out} is the outage probability. For the traditional, non-cooperative link, I took the position of the off-body receiver (single receiver unit – SRU) identical to the location of the closest CRU in the cooperative scenario.

My investigations have shown that the transmission power of the implant can considerably be reduced by means of the proposed cooperative communication procedure. Taking also the power consumption model into account, the application of the procedure results in a significant lifetime improvement for the implant (Figure 4.5). I examined two relay selection criteria in the analysis. In the case of the complex criterion, the channel gain of both the implant-to-CRU and the CRU-to-G-CRU channel is taken into account, while for the simple criterion the gain of only the implant-to-CRU channel is considered. The result have shown that – for the specific values of the parameters used in the model – uniform power allocation combined with the simple criterion performs practically the same way as optimal power allocation combined with the more complex criterion.

III.2. Based on the existing single-relay selection procedure, I have derived a distributed, multiple-relay selection procedure. I have shown that choice of the distortion function has a fundamental impact on the performance of the procedure and that for a low number of relays, the mean duration of the procedure is much lower than the coherence time of the radio channel experienced in practice.

I extended the existent method into a multiple-relay selection algorithm in such a way that the potential relays keep listening and the procedure does not terminate until a given number of buzzes are transmitted on the channel. This number is equal to the number of relays to be selected (Figure 7.2). As the RX-to-TX switch time of real radios is non-zero, there is a finite probability that while a potential relay node with a lower initial timer setting is switching from RX to TX mode (in order to transmit its buzz), the timer of another node with a higher initial timer value expires. In this case,

DOI:10.15774/PPKE.ITK.2009.001

each of the two nodes will assume the same role for itself and as result, the relay selection procedure fails. The event of failure can be expressed as

$$\bigcup_{i=1}^{n}(T_{i+1}-T_i<c), \tag{7.13}$$

where T_i is the i^{th} lowest initial timer value, n is the number of relays to be selected and c is the RX-to-TX switch time of the radios. The distortion function is used to transform the reciprocal of the selection criterion into initial timer value.

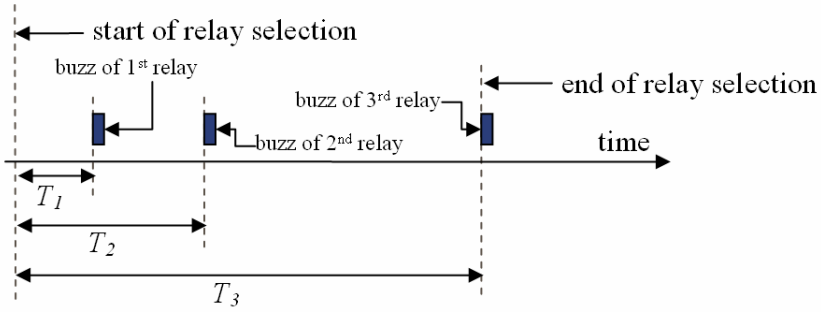

Figure 7.2. The relay selection procedure (n=3).

Using Monte Carlo simulations, I determined the probability of failure as a function of the mean duration of the procedure assuming different selection criteria and distortion functions. For $n=9$, this is shown in Figure 5.1. For sophisticated radios, the value of parameter c is around $1\mu s$. The investigations have shown that – under the assumption of a low number of relays ($n<10$) and a suitably chosen distortion function – the mean duration of the procedures is much lower than the typical coherence time of the channel (100ms), while at the same time the probability of failure remains lower than a practically acceptable limit (1%).

7.3. Examples of application

The high-speed, multicolor, THz and infrared cameras can primarily be applied to identify and to discriminate fast moving objects as well as to observe fast chemical and biological processes.

The energy aware packet forwarding protocols can be used in a wide range of wireless sensor network applications such as intelligent home (e.g. the NETVOX Smart House

DOI:10.15774/PPKE.ITK.2009.001

System), mechanical structure monitoring (e.g. the Sensametrics Wireless Structure Monitoring), environment monitoring (e.g. the Advanced Sensor Technologies Root Zone Intelligence System), seismic activity monitoring.

The cooperative communication procedures described in the third group of the theses can primarily be utilized in those telemedicine and telecare applications (e.g. BIOTRONIK Home Monitoring [44]) in which the implanted wireless sensors are of limited energy and non-rechargeable. Such implants are – for instance – the pacemakers and cardioverter defibrillators (pl. Biotronik Philos II DR-T pacemaker and Lumax DR-T ICD). In addition, the procedures can be well employed in applications in which the implanted sensors can be recharged through – as an example – inductive coupling. In these cases, the frequency of the necessary recharges can significantly be decreased, which obviously improves the quality of life of the person carrying the implant.

DOI:10.15774/PPKE.ITK.2009.001

Chapter Eight

APPENDICES

8.1. Appendix A

In this part of the Appendices, a brief overview of the simulated annealing algorithm – along with a finite-time approximation of that – is provided for a better understanding based on [34]. In addition, in the last part of this section, the complexity of the particular simulated annealing algorithms described in Section 3.4.1 is assessed.

8.1.1. Simulated annealing

In condensed matter physics, annealing in known as a thermal process for obtaining low energy states of a solid. It consists of the following two stages. First, the temperature of the solid is increased to a maximum value where the solid melts, and then it is carefully decreased until the particles of the solid arrange themselves in the ground state of that. In the liquid phase, the particles organize themselves randomly, while in the ground state they are arranged in a highly structural lattice and the energy is minimal. The ground state of the solid is reached only if the maximum temperature is sufficiently high and the cooling is carried out sufficiently slowly.

The physical annealing process – and the evolution of the state of a solid to thermal equilibrium – can be simulated with the help of the Metropolis algorithm. In this algorithm, a sequence of states of the solid is generated as follows. Given the current state i with energy E_i, the subsequent state j is obtained by applying a perturbation mechanism – involving a small distortion such as a particle displacement – to the current state. The energy of the subsequent state is E_j. The probability that the transition from state i to state j is accepted, i.e. state j becomes the current state, is expressed as

$$P_T\{i \to j\} = \begin{cases} 1 \text{ if } E_j \le E_i, \\ e^{\frac{E_i - E_j}{k_B T}} \text{ if } E_j > E_i, \end{cases}, \tag{8.1}$$

DOI:10.15774/PPKE.ITK.2009.001

where k_B and T denotes the Boltzmann constant and the temperature, respectively. The acceptance rule described above is the known as the Metropolis criterion. If the temperature is lowered sufficiently slowly, the solid reaches thermal equilibrium at each temperature, which is characterized by the Boltzmann distribution.

The simulated annealing algorithm is obtained by applying the Metropolis algorithm to generate a sequence of solutions of a combinatorial optimization problem. The analogy between a physical system and a combinatorial optimization problem is based on the following equivalences. The solutions in a combinatorial optimization problem are equivalent to the state of the physical system, while the cost (f) of a solution is equivalent to the energy (E) of a state. In the simulated annealing algorithm, the role of temperature is played by the so-called control parameter (c), whereas the analogue of the Metropolis criterion is termed as acceptance criterion, which is formally identical to (8.1). The equivalent of the perturbation mechanism is called generation mechanism, which determines how a subsequent solution is selected from the neighborhood of the current one. Finally, the neighborhood structure defines for each solution a set of solutions that are close to it in some sense. The set and its elements are called the neighborhood and the neighbors of the given solution, respectively.

The behavior of the simulated annealing algorithm can be studied using the theory of finite Markov chains. It can be shown that if the generation mechanism fulfills certain conditions than the homogeneous Markov chain ($X(k)$) associated with the algorithm has a stationary distribution at each value of c, whose components are given by

$$\lim_{k\to\infty} P_c\{X(k)=i\} = \frac{e^{-\frac{f(i)}{c}}}{\sum_{j\in S} e^{-\frac{f(j)}{c}}}, \tag{8.2}$$

where S denotes the space of the solutions. (Please observe the similarity between (8.2) and the Boltzmann distribution). It is easy to see that

DOI:10.15774/PPKE.ITK.2009.001

$$\lim_{c\downarrow 0}\lim_{k\to\infty} P_c\{X(k)=i\}=\frac{1}{|S_{\text{opt}}|}\chi_{(S_{\text{opt}})}(i) \tag{8.3}$$

or

$$\lim_{c\downarrow 0}\lim_{k\to\infty} P_c\{X(k)\in S_{\text{opt}}\}=1, \tag{8.4}$$

where S_{opt} is the set of optimal solutions, while $\chi_{(A)}(a)$ equals to 1 if $a\in A$ and it is 0 otherwise. Equations (8.3) and (8.4) reflect the basic property of the simulated annealing algorithm, i.e. the guarantee that it asymptotically finds an optimal solution.

The same equations, on the other hand, indicate that the implementation of the algorithm along these lines would require the generation of a sequence of infinitely long homogeneous Markov chains at descending values of the control parameter. This is obviously impractical. One way to help this problem is to replace the infinitely long chains with finite ones in the sequence. In this manner, the sequence of infinitely long homogeneous Markov chains is reduced to a single inhomogeneous Markov chain of infinite length. It can be proven that if the generation mechanism and the sequence of the control parameter values satisfy certain conditions than the inhomogeneous Markov chain associated with the algorithm converges in distribution to

$$\lim_{k\to\infty} P\{X(k)=i\}=\frac{1}{|S_{\text{opt}}|}\chi_{(S_{\text{opt}})}(i) \tag{8.5}$$

or, in other words,

$$\lim_{k\to\infty} P\{X(k)\in S_{\text{opt}}\}=1. \tag{8.6}$$

Nevertheless, the application of the inhomogeneous Markov chain introduced above is by no means the ultimate solution. Namely, it can be shown that in order for (8.5) to be approximated arbitrarily closely, a number of transitions that – for most problem

DOI:10.15774/PPKE.ITK.2009.001

instances – is larger than the solution space, is required leading to a exponential-time execution of the algorithm. Hence, the simulated annealing algorithm needs to be approximated to run in finite (polynomial) time.

8.1.2. Finite-time approximation

A finite-time approximation of the simulated annealing algorithm can be implemented by generating homogeneous Markov chains of finite length with a finite sequence of decreasing values of the control parameter. In order to realize this, a set of parameters making up the so-called cooling schedule needs to be specified. The parameters included in a cooling schedule are the *initial value* of the control parameter, the *decrement function* for decreasing the value of the control parameter, the *stop criterion* determining the final value of the control parameter and the finite *length* of each homogeneous Markov chain. The first three parameters obviously define a finite sequence of values of the control parameter.

In this study, we apply the polynomial-time cooling schedule described in [34]. The initial value of the control parameter, i.e. c_0 is set in such a way that the probability that a proposed transition is accepted is equal to the χ_0 *initial acceptance ratio* at $c = c_0$ and statistical equilibrium. The decrement function is expressed as

$$c_{k+1} = \frac{c_k}{1 + \dfrac{c_k \ln(1+\delta)}{3\sigma_{c_k}}}, \; k = 0,1,\ldots, \tag{8.7}$$

where δ is the *distance parameter*, while σ_c is the standard deviation of the cost in the case of the distribution given by (8.3). The stop criterion is given by

$$\frac{c_k}{\langle f \rangle_\infty} \left. \frac{\partial \langle f \rangle_c}{\partial c} \right|_{c=c_k} < \varepsilon_s, \tag{8.8}$$

where ε_s is the *stop parameter*, whereas $\langle f \rangle_c$ is the mean value of the cost in the case of the distribution given by (8.3). Finally, the length of the Markov chains is taken equal to the Θ size of the neighborhoods, i.e.

DOI:10.15774/PPKE.ITK.2009.001

$$L_k = L = \Theta\,,\ k = 0,1,\ldots\,. \tag{8.9}$$

The way c_0, σ_c, $\langle f \rangle_c$ as well as the left-hand side of (8.8) can be approximated, is described in [34]. Similarly, for further details and for the ideas behind the cooling schedule, please refer to [34].

According to [34], the finite-time approximation of the simulated annealing algorithm presented above requires a computation time T that satisfies the following relation

$$T = O\left(\tau \cdot \Theta \cdot \ln|S|\right), \tag{8.10}$$

where τ denotes the computation time to carry out a single transition. For most combinatorial optimization problems, τ and Θ can be chosen polynomially in the problem size. As a consequence, if also $\ln|S|$ is polynomial in the problem size, the approximate algorithm runs in polynomial time.

Based on (8.10), it is obvious that a concise representation of the possible solutions of the original problem, i.e. mapping of those onto S is essential in order to reduce computation time.

8.1.3. Complexity analysis

In this part, the complexity of the optimization algorithms described in Section 3.4.1 is assessed.

Let us evaluate how the T computation time of the optimal solution depends on N and M assuming the solution representation and neighborhood structure described in Section 3.4.1. (For the definition of N and M, please refer to Section 3.2.1 and Section 3.4.1, respectively.) Recall that, in the case of the ‘random short-cut’ and ‘sending packets to any node ahead in the chain’ protocols,

DOI:10.15774/PPKE.ITK.2009.001

$$\Theta = 2(N-1), \qquad \ln|S| = (N-1)\ln(M+1) \tag{8.11}$$

and

$$\Theta = N(N-1), \qquad \ln|S| = \sum_{i=1}^{N+1} \ln\binom{M+i}{i} \approx \frac{N(N-1)}{2}\ln(M+1), \tag{8.12}$$

respectively. (The approximate expression on the right-hand side of the (8.12) belongs to Solution 2 in Section 3.4.1.) Hence

$$T = O\left(\tau \cdot N^2 \cdot \ln M\right) \tag{8.13}$$

and

$$T = O\left(\tau \cdot N^4 \cdot \ln M\right), \tag{8.14}$$

respectively. As τ might also increase with N, the complexity of finding the optimal solution for the 'random short-cut' and 'sending packets to any node ahead in the chain' protocols is at least $O(N^2)$ and $O(N^4)$, respectively. Concerning M, the complexity of the problem is much more moderate, it is only $O(\ln M)$.

8.2. Appendix B

In this section, the Monte Carlo simulations performed in this study are discussed in detail. First, the theory of sampling is described in brief for a better understanding based on [65] and [66] and after that the way it is applied in the simulations is presented.

Let us consider a sample of a number of n independent observations (x_i, $i = 1,...,n$) of a random variable (x) with an unknown (population) mean μ. Then variable

DOI:10.15774/PPKE.ITK.2009.001

$$t = \frac{\bar{x}_n - \mu}{s_n / \sqrt{n}}, \tag{8.15}$$

is distributed according to Student's (Gosset's) t-distribution with $n-1$ degrees of freedom. Here $\bar{x}_n$ and s_n are the sample mean and the estimator for the population standard deviation (i.e. the sample variance), respectively,

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i , \tag{8.16}$$

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 . \tag{8.17}$$

The probability density function of Student's t-distribution with $n-1$ degrees of freedom can be expressed as

$$f_{n-1}(t) = \frac{\Gamma\left[\frac{1}{2}n\right]}{\sqrt{(n-1)\pi}\,\Gamma\left[\frac{1}{2}(n-1)\right]\left(1+\frac{t^2}{n-1}\right)^{\frac{n}{2}}}, \tag{8.18}$$

where Γ is the gamma function. Then the probability that the population mean μ falls into the $\left(\bar{x}_n - \frac{s_n}{\sqrt{n}}a, \bar{x}_n + \frac{s_n}{\sqrt{n}}a\right]$ confidence interval is given by

$$P\left\{\bar{x}_n - \frac{s_n}{\sqrt{n}}a < \mu \le \bar{x}_n + \frac{s_n}{\sqrt{n}}a\right\} = P\{-a < t \le a\} = F_{n-1}(a) - F_{n-1}(-a) = 1 - 2\cdot F_{n-1}(-a) = \tag{8.19}$$
$$= 2\cdot F_{n-1}(a) - 1 = 1 - \beta,$$

where $F_{n-1}(t)$ is the cumulative density function of Student's t-distribution with $n-1$ degrees of freedom. The probability expressed in (8.19) is called the level of confidence. Putting (8.19) into other words, a is the upper $1-\frac{\beta}{2}$ percentile of Student's t-

DOI:10.15774/PPKE.ITK.2009.001

distribution with $n-1$ degrees of freedom. The values of a are listed in Table 8.I as a function of n and β.

TABLE 8.I. THE VALUE OF a AS FUNCTION OF n AND β.

n	$\beta=10\%$	$\beta=5\%$	$\beta=1\%$
1	6.314	12.71	63.66
2	2.920	4.303	9.925
5	2.015	2.571	4.032
10	1.812	2.228	3.169
20	1.725	2.086	2.845
50	1.676	2.009	2.678
100	1.660	1.984	2.626
∞	1.645	1.960	2.576

Please note that when n becomes large, Student's t-distribution converges to the standard normal distribution.

Now let us see how the above theory is applied in the Monte Carlo simulations performed in this study. The confidence level and the minmum number of observations is set to 99% and 1000, respectively. As the latter value is fairly high, the $n \approx \infty$ approximation is used, and as a result, a is set to 2.576 (Table 8.I). When estimating the population mean μ with sample mean $\bar{x}_n$, the relative error of the estimation becomes

$$\varepsilon = \left| \frac{\bar{x}_n - \mu}{\mu} \right|. \tag{8.20}$$

As the number of observations, i.e. n is relatively high, μ is approximated with $\bar{x}_n$ in the denominator of (8.20) in this study:

$$\varepsilon \approx \left| \frac{\bar{x}_n - \mu}{\bar{x}_n} \right|. \tag{8.21}$$

As a consequence, the number of observation used in this study in order to estimate μ with a confidence level of 1% and a maximum relative precison γ is given by

DOI:10.15774/PPKE.ITK.2009.001

$$n' = \arg\min_{n \geq 1000} \left(2.576 \cdot \frac{s'_n}{\bar{x}'_n} \frac{1}{\sqrt{n}} < \gamma \right), \tag{8.22}$$

where if $n = 1000$ then $\bar{x}'_n$ and s'_n is a realization of $\bar{x}_n$ and s_n for the same realization of x_i ($i = 1,...,1000$), respectively, while if $n > 1000$ then

$$\bar{x}'_n = \left(1 - \frac{1}{n}\right)\bar{x}'_{n-1} + \frac{1}{n}x', \tag{8.23}$$

$$s'^2_n = \left(1 - \frac{1}{n-1}\right)s'^2_{n-1} + \bar{x}'^2_{n-1} + \frac{1}{n-1}x'^2 - \left(1 + \frac{1}{n-1}\right)\bar{x}'^2_n, \tag{8.24}$$

where x' is a given realization of x. Finally, the estimated value of μ is expressed as

$$\mu = \bar{x}'_{n'}. \tag{8.25}$$

The value of the relative precision γ is set between 0.625% and 2.5%.

In this study, three quantities are estimated using the method described above: the outage probability in Section 4.5 and the failure probability and mean bit error probability in Section 5.4. The latter quantity is estimated directly, while the former two can be estimated in the following indirect way. Let us consider an auxiliary random variable that is equal to 1 if an outage event occurs and is equal to 0 if no outage event takes place. It is easy to see to that mean of this variable is equal to the outage probability. As a consequence, the outage probability can be estimated by estimating the mean of this auxiliary random variable applying the method described above. The failure probability of the relay selection procedure in Section 5.4 can be estimated in the same way.

8.3. Appendix C

In this section, the formula of f_2 (5.9) is derived.

Based on Section 5.2,

DOI:10.15774/PPKE.ITK.2009.001

$$SNR_{\mathrm{BN}\to\mathrm{EAN}_i\to\mathrm{PN}}^{-1} = P_{\mathrm{BN}}'^{-1} d_{\mathrm{BN,EAN}_i}'^{\alpha} r_{\mathrm{BN,EAN}_i}^{-1} + \left(1 + P_{\mathrm{BN}}'^{-1} d_{\mathrm{BN,EAN}_i}'^{\alpha} r_{\mathrm{BN,EAN}_i}^{-1}\right)\eta^{-1} P_{\mathrm{BN}}'^{-1} d_{\mathrm{EAN}_i,\mathrm{PN}}'^{\alpha} r_{\mathrm{EAN}_i,\mathrm{PN}}^{-1}, \quad (8.26)$$

where $d'_{\mathrm{BN,EAN}_i}$ ($d'_{\mathrm{EAN}_i,\mathrm{PN}}$) is $d_{\mathrm{BN,EAN}_i}$ ($d_{\mathrm{EAN}_i,\mathrm{PN}}$) normalized with $\sqrt{\lambda^{-1}}$, i.e. twice the mean distance of a node and its closest neighbor [64], while

$$P'_{\mathrm{BN}} = \frac{P_{\mathrm{BN}}}{\sigma_{\mathrm{n}}^2} A \lambda^{\frac{\alpha}{2}}, \quad (8.27)$$

i.e. it is the mean received SNR from the BN in a distance that is equal to twice the mean distance of the BN and the closest EAN (i.e. $\sqrt{\lambda^{-1}}$). (Here EANs are indexed according to their distance to the BN, EAN_1 being the closest EAN to the BN.)

Now let us determine the mean of (8.26) as a function of i:

$$\begin{aligned}
&\mathrm{E}\left\{SNR_{\mathrm{BN}\to\mathrm{EAN}_i\to\mathrm{PN}}^{-1}\right\} \approx P_{\mathrm{BN}}'^{-1}\mathrm{E}\left\{d_{\mathrm{BN,EAN}_i}'^{\alpha}\right\}\mathrm{E}\left\{r_{\mathrm{BN,EAN}_i}^{-1}\right\} + \\
&+\left(1 + P_{\mathrm{BN}}'^{-1}\mathrm{E}\left\{d_{\mathrm{BN,EAN}_i}'^{\alpha}\right\}\mathrm{E}\left\{r_{\mathrm{BN,EAN}_i}^{-1}\right\}\right)\eta^{-1} P_{\mathrm{BN}}'^{-1}\mathrm{E}\left\{d_{\mathrm{EAN}_i,\mathrm{PN}}'^{\alpha}\right\}\mathrm{E}\left\{r_{\mathrm{EAN}_i,\mathrm{PN}}^{-1}\right\} \approx \\
&\approx P_{\mathrm{BN}}'^{-1}\mathrm{E}\left\{d_{\mathrm{BN,EAN}_i}'^{\alpha}\right\} I + \left(1 + P_{\mathrm{BN}}'^{-1}\mathrm{E}\left\{d_{\mathrm{BN,EAN}_i}'^{\alpha}\right\} I\right)\eta^{-1} P_{\mathrm{BN}}'^{-1}\mathrm{E}\left\{d_{\mathrm{EAN}_i,\mathrm{EAN}_1}'^{\alpha}\right\} I \approx \\
&\approx I \cdot P_{\mathrm{BN}}'^{-1}\left(\mathrm{E}\left\{d_{\mathrm{BN,EAN}_i}'^{2}\right\}\right)^{\frac{\alpha}{2}} + \left[1 + I \cdot P_{\mathrm{BN}}'^{-1}\left(\mathrm{E}\left\{d_{\mathrm{BN,EAN}_i}'^{2}\right\}\right)^{\frac{\alpha}{2}}\right]\eta^{-1} I \cdot P_{\mathrm{BN}}'^{-1}\left(\mathrm{E}\left\{d_{\mathrm{EAN}_i,\mathrm{EAN}_1}'^{2}\right\}\right)^{\frac{\alpha}{2}} = \\
&= I \cdot P_{\mathrm{BN}}'^{-1} i^{\frac{\alpha}{2}} + \left[1 + I \cdot P_{\mathrm{BN}}'^{-1} i^{\frac{\alpha}{2}}\right]\eta^{-1} I \cdot P_{\mathrm{BN}}'^{-1}(i+1)^{\frac{\alpha}{2}} \approx \\
&\approx I \cdot P_{\mathrm{BN}}'^{-1}\left[\left(1+\eta^{-1}\right) i^{\frac{\alpha}{2}} + I \cdot P_{\mathrm{BN}}'^{-1}\eta^{-1} i^{\alpha}\right].
\end{aligned} \quad (8.28)$$

The first approximation in (8.28) implies the assumption that variables $d'_{\mathrm{BN,EAN}_i}$, $r_{\mathrm{BN,EAN}_i}$, $d'_{\mathrm{EAN}_i,\mathrm{PN}}$ and $r_{\mathrm{EAN}_i,\mathrm{PN}}$ are uncorrelated, while the second one involve that EAN_1 is elected as the PN all the time (i.e. with unit probability). (Constant I – introduced at this stage – denotes the mean of the reciprocal of an exponentially distributed variable with unit mean. The latter quantity is not finite, however, as we will see later on, this will not prevent us from deriving the formula of the distortion function.) In the next step, the $\alpha/2^{\mathrm{th}}$ moment of variables $d'^{2}_{\mathrm{BN,EAN}_i}$ and $d'^{2}_{\mathrm{EAN}_i,\mathrm{EAN}_1}$ is approximated by the $\alpha/2^{\mathrm{th}}$ power

DOI:10.15774/PPKE.ITK.2009.001

of their mean. (Though $\mathrm{E}\{d'^{\alpha}_{\mathrm{BN,EAN}_i}\}$ can be expressed in a closed form [64], the same is not true for $\mathrm{E}\{d'^{\alpha}_{\mathrm{EAN}_i,\mathrm{EAN}_1}\}$, and therefore both quantities are approximated throughout the derivation). Then – according to [64] – $\mathrm{E}\{d'^{2}_{\mathrm{BN,EAN}_i}\}$ is replaced with i, while $\mathrm{E}\{d'^{2}_{\mathrm{EAN}_i,\mathrm{EAN}_1}\}$ $(i>1)$ can easily be shown to be equal to $i+1$. Finally, $i+1$ is approximated by i, which is adequate for large values of i.

We can observe in the approximate expression of $\mathrm{E}\{SNR^{-1}_{\mathrm{BN}\to\mathrm{EAN}_i\to\mathrm{PN}}\}$ (i.e. the mean of the initial value of the timer of EAN_i when no distortion function is applied) that the difference in its value for two consecutive values of i is not constant. In order to distribute the mean of the initial value of the different timers more evenly – which is expected to minimize the probability of failure – we should apply a suitable monotonic increasing function to transform $SNR^{-1}_{\mathrm{BN}\to\mathrm{EAN}_i\to\mathrm{PN}}$ (i.e. the reciprocal of the selection criterion). To that end, let us express i as a function of $\mathrm{E}\{SNR^{-1}_{\mathrm{BN}\to\mathrm{EAN}_i\to\mathrm{PN}}\}$ based on (8.28):

$$i\left(\mathrm{E}\{SNR^{-1}_{\mathrm{BN}\to\mathrm{EAN}_i\to\mathrm{PN}}\}\right)=\left\{\frac{1}{2}P'_{\mathrm{BN}}I^{-1}(\eta+1)\cdot\right.$$
$$\left.\cdot\left[\sqrt{1+4\mathrm{E}\{SNR^{-1}_{\mathrm{BN}\to\mathrm{EAN}_i\to\mathrm{PN}}\}\left(\sqrt{\eta}+\frac{1}{\sqrt{\eta}}\right)^{-2}}-1\right]\right\}^{\frac{2}{\alpha}}. \tag{8.29}$$

Finally assuming that the above function and $\mathrm{E}\{.\}$ commute, (5.9) seems an appropriate choice for transforming $SNR^{-1}_{\mathrm{BN}\to\mathrm{EAN}_i\to\mathrm{PN}}$ into a value that is directly proportional to the initial value of the timers.

DOI:10.15774/PPKE.ITK.2009.001

BIBLIOGRAPHY

Uncooled, high-speed, multispectral, CMOS-compatible THz and infrared detectors

[1] J. D. Vincent, "Fundamentals of infrared detector operation and testing," *John Wiley & Sons*, New York – Chichester – Brisbane – Toronto – Singapore, 1989.

[2] A. Sanchez, C. F. Davis, Jr., K. C. Liu, and A. Javan, "The MOM tunneling diode: theoretical estimate of its performance at microwave and infrared frequencies," *Journal of Applied Physics*, 1978, vol. 49, pp. 5270–5277.

[3] A. Csurgay, W. Porod, "Surface plasmon waves in nanoelectronic circuits", *International Journal of Circuit Theory and Applications*, 2004, vol. 32, 339–361.

[4] A. S. Weling, P. F. Henning, D. P. Neikirk, S. Han, " Antenna-coupled microbolometers for multispectral infrared imaging", *Proc. Infrared Technology and Applications XXXII, SPIE*, 2006, vol. 62061F.

[5] L. O. Hocker, D.R. Sokoloff, V. Daneu, A. Szoke, A. Javan, "Frequency mixing in the infrared and far-infrared using a metal-to-metal point contact diode," *Applied Physics Letters*, 1968, vol. 12, pp. 401–402.

[6] M. Heiblum, S. Y. Wang, J. R. Whinnery, T. K. Gustafson, "Characteristics of integrated MOM junctions at dc and at optical frequencies," *IEEE Journal of Quantum Electronics*, 1978, vol. 14, pp. 159-169.

[7] C. Fumeaux, W. Herrmann, F. K. Kneubühl, H. Rothuizen, "Nanometer thin-film Ni–NiO–Ni diodes for detection and mixing of 30 THz radiation," *Infrared Physics & Technology*, 1998, vol. 39, pp. 123–183.

[8] I. Wilke, Y. Oppliger, W. Herrmann, F. K. Kneubühl, "Nanometer thin-film Ni-NiO-Ni diodes for 30 THz radiation," *Applied Physics A-Materials Science & Processing*, 1994, vol. 58, pp. 329–341.

[9] I. Codreanu, F. J. Gonzalez, G. D. Boreman, "Detection mechanisms in microstrip dipole antenna-coupled infrared detectors," *Infrared Physics & Technology*, 2003, vol. 44, pp. 155–163.

[10] G. Matyi, "Nanoantennas for uncooled, double-band, CMOS compatible, high-speed infrared sensors", *International Journal of Circuit Theory and Applications*, 2004, vol. 32, pp. 425–430.

[11] G. Matyi, "The electrodynamics of nanoantenna-coupled MOM diode sensors," Ph. D. dissertation, Faculty of Information Technology, Pázmány Péter Catholic University. Budapest, Hungary, 2007.

[12] R. Stratton, "Volt-current characteristics for tunneling through insulating films," *Journal of Physics and Chemistry of Solids*, 1962, vol. 23, pp. 1177–1190.

[13] J. G. Simmons, "Potential barriers and emission-limited current flow between closely spaced parallel metal electrodes," *Journal of Applied Physics*, 1964, vol. 35, pp. 2472–2481.

[14] J. G. Simmons, "Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film," *Journal of Applied Physics*, 1963, vol. 34, pp. 1793–1803.

[15] T. E. Hartman, "Tunneling Through Asymmetric Barriers", *Journal of Applied Physics*, 1964, vol. 35, pp. 3283–3294.

DOI:10.15774/PPKE.ITK.2009.001

[16] B. Rakos, "Investigation of metal-oxide-metal structures for optical sensor applications," Ph.D. dissertation, Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN, USA, 2006.

[17] Gy. Fodor, "Hálózatok és rendszerek analízise 2. rész," *Műegyetemi Kiadó*, Budapest, 2000.

[18] W. R. Frensley, "Heterostructures and quantum devices," *Academic Press*, San Diego, 1994, ch. 9 (Quantum Transport).

[19] C. S. Lent, D. J. Kirkner, "The quantum transmitting boundary method," *Journal of Applied Physics*, 1990, vol. 67, pp. 6353–6359.

[20] N. W. Ashcroft, N. D. Mermin, "Solid state physics," *Saunders College*, Philadelphia, 1976.

Energy-aware packet forwarding protocols

[21] MPR/MIB User's Manual, Rev. A, September 2005, Document 7430-0021-07, www.xbow.com

[22] W. R. Heinzelman, A. Chandrakasan, H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," *Proc. 33rd Hawaii International Conference on System Sciences*, Island of Maui, HI, USA, 2000, p. 8020.

[23] H. O. Tan, I. Korpeoglu, "Power efficient data gathering and aggregation in wireless sensor networks," *Sigmoid Record*, 2003, vol. 32, pp. 66–71.

[24] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, C. Erdal, "Wireless sensor networks: a survey," *Computer Networks*, 2002, vol. 38, pp. 393–422.

[25] S. Lindsey, C. Raghavendra, K.M. Sivalingam, "Data gathering algorithms in sensor networks using energy metrics," *IEEE Transactions on Parallel and Distributed Systems*, 2002, vol. 13, pp. 924–935.

[26] M. Haenggi, "Energy-balancing strategies for wireless sensor networks," *Proc. 2003 International Symposium on Circuits and Systems,* Bangkok, Thailand, 2003, Vol.4, pp. 828–831.

[27] 3GPP TR 43.030 V7.0.0, "Radio network planning aspects," www.3gpp.org.

[28] A. F. Molisch, "Wireless Communications," *John Wiley & Sons*, Chichester, 2005, ch. 5.

[29] L. Song, D. Hatzinakos, "Cooperative transmission in poisson distributed wireless sensor networks: protocol and outage probability," *IEEE Transactions on Wireless Communications*, 2006, vol. 5, pp. 2834–2843.

[30] S. Savazzi, U. Spagnolini, "Energy aware power allocation strategies for multihop-cooperative transmission schemes", *IEEE Journal on Selected Areas in Communications*, 2007, vol. 25, pp. 318–327.

[31] W. Henderson, P. G. Taylor, "Product form in networks of queues with batch arrivals and batch services," *Queueing Systems*, 1990, vol. 6, pp. 71–87.

[32] W. Henderson, P. G. Taylor, "Some new results on queueing networks with batch movement", *Journal of Applied Probability*, 1991, vol. 28, pp. 409–421.

[33] L. Kleinrock, "Queueing Systems. Vol I: Theory," *John Wiley & Sons*, New York – London – Sydney – Toronto, 1975.

DOI:10.15774/PPKE.ITK.2009.001

[34] E. Aarts, J. Korst, "Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing," *John Wiley and Sons*, Chichester – New York – Weinheim – Brisbane – Toronto – Singapore, 1989.

[35] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, "Optimization by simulated annealing," *Science*, 1983, vol. 220, 1983, pp. 671–680.

[36] V. Cerny, "Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm," *Journal of Optimization Theory and Applications*, 1985, vol. 45, pp. 41–45.

Cooperative communication procedures in biomedical applications

[37] P. S. Hall, Y. Hao, "Antennas and Propagation for Body-Centric Wireless Communications," *Artech House*, Boston – London, 2006, ch. 8–9.

[38] J. N. Laneman; G. W. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Transactions on Information Theory*, 2003, vol. 49, pp. 2415–2425.

[39] A. J. Johansson, "Performance of a radio link between a base station and a medical implant utilising the MICS standard," *Proc. 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, San Francisco, CA, USA, 2004, vol. 3, pp. 2113–2116.

[40] A. Bletsas, A. Khisti, D. P. Reed, A.Lippman, "A simple cooperative diversity method based on network path selection," *IEEE Journal on Selected Areas in Communications*, 2006, vol. 24, pp. 659–672.

[41] Electromagnetic compatibility and radio spectrum matters (ERM); Short range devices (SRD); Ultra low power active medical implants (ULP-AMI) and peripherals (ULP-AMI-P) operating in the frequency range 402 MHz to 405 MHz; Part 1: Technical characteristics and test methods, ETSI Standard EN 301 839-1 V1.2.1, 2007.

[42] Sharing between the meteorological aids service and medical implant communication systems (MICS) operating in the mobile service in the frequency band 401-406 MHz, ITU-R SA.1346 Recommendation, 1998.

[43] FCC Standard 47 CFR 95.601-95.673 Subpart E, 1999.

[44] Biotronik Home Monitoring, http://www.biotronik.de.

[45] A. J. Johansson, "Wireless communication with medical implants: antennas and propagation," Ph.D. dissertation, Department of Electroscience, Faculty of Engineering, Lund University, Lund, Sweden, 2004.

[46] E. Sendonaris, E. Erkip, B. Aazhang, "User cooperation diversity – part I.: system description," *IEEE Transactions on Communications*, 2003, vol. 51, pp. 1927–1938.

[47] G-Z. Yang, "Body Sensor Networks," *Springer-Verlag*, London, 2006, ch 5.

[48] A. Tekin, M. R. Yuce, J. Shabani, L. Wentai, "A low-power FSK modulator/demodulator for an MICS band transceiver", *Proc. 2006 IEEE Radio and Wireless Symposium*, San Diego, CA, USA, 2006, pp. 159–162.

[49] L. F. Tanguay, M. Sawan, "A fully-integrated 580 μW ISM-band frequency synthesizer for implantable medical devices," *Proc. International Symposium on Signals, Circuits and Systems 2007*, Iasi, Romania, 2007, vol. 1, 4 pp.

[50] N. M. Neihart, R. R. Harrison, "A low-power FM transmitter for use in neural recording applications," Proc. *26th Annual International Conference of the IEEE*

DOI:10.15774/PPKE.ITK.2009.001

Engineering in Medicine and Biology Society, San Fransisco, CA, USA, 2004, vol. 1, pp. 2117–2120.

[51] M. .M. El-Desouki, M. J. Deen, Y. M. Haddara, "A low-power CMOS class-E power amplifier for biotelemetry applications," *Proc. 35th European Microwave Conference*, Paris, France, 2005, vol. 1, 4 pp.

[52] S. M. Abdelsayed, M. J. Deen, N. K. Nikolova, "A fully integrated low-power CMOS power amplifier for biomedical applications," *Proc. 8th European Conference on Wireless Technology*, Paris, France, 2005, pp. 261–263.

[53] L. B. Picasso, P. Jansson, "Simulation and measurement of radio wave propagation in the 400 MHz MICS band," M.Sc. thesis, Department of Electroscience, Faculty of Engineering, Lund University, Lund, Sweden, 2004.

[54] A. F. Molisch, "Wireless Communications," *John Wiley & Sons*, Chichester, 2005, ch. 5, 7, 13.

[55] W. G. Scanlon, J. B. Burns, N. E. Evans, "Radiowave propagation from a tissue-implanted source at 418 MHz and 916.5 MHz," *IEEE Transactions on Biomedical Engineering*, 2000, vol. 47, pp. 527–534.

[56] A. J. Johansson, A. Karlsson, "Wave-propagation from medical implants – influence of arm movements on the radiation pattern", *Proc.18th Nordic Conference on Radio Science and Communications*, Stockholm, Sweden, 2002.

[57] A. Alomainy, Y. Hao, Y. Yuan, Y. Liu, "Modelling and characterisation of radio propagation from wireless implants at different frequencies," *Proc. 9th European Conference on Wireless Technology*, Manchester, UK, 2006, pp. 119-122.

[58] J. N. Laneman, D. N. C. Tse, G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, 2004, vol. 50, pp. 3062–3080.

[59] P. Larsson, H. Rong, "Large-scale cooperative relaying network with optimal relaying coherent combining under aggregate relay power constraints", *Proceedings of Working Group 4, World Wireless Research Forum WWRF8 meeting*, Beijing, China, February 26-27, 2004.

[60] A. Bletsas, H. Shin, M. Z. Win, "Outage optimality of opportunistic amplify-and-forward relaying," *IEEE Communications Letters*, 2007, vol. 11, pp. 261–263.

[61] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Transactions on Signal Processing*, 2007, vol. 55, pp. 4683–4695.

[62] A. Bletsas, A. Lippman, D. P. Reed, "A simple distributed method for relay selection in cooperative diversity wireless networks, based on reciprocity and channel measurements,", *Proc. IEEE 61st Semiannual Vehicular Technology Conference*, Stockholm, Sweden, 2005, vol. 3, pp. 1484–1488.

[63] J. Lee, K. Cho, S. Lee, T. Kwon, Y. Choi, "Distributed and energy-efficient target localization and tracking in wireless sensor networks," *Computer Communications*, 2006, vol. 29, pp. 2494–2505.

[64] M. Haenggi, "On distances in uniformly random networks," *IEEE Transactions on Information Theory*, 2005, vol. 51, pp. 3584–3586.

[65] V. B. Iversen, "Teletraffic Engineering and Network Planning," ch. 15, http://oldwww.com.dtu.dk/education/34340/telenook.pdf.

[66] E. W. Weisstein, "Student's t-Distribution," from MathWorld – A Wolfram Web Resource, http://mathworld.wolfram.com/Studentst-Distribution.html.

DOI:10.15774/PPKE.ITK.2009.001

THE AUTHOR'S PUBLICATIONS

Journal papers

[1] **B. Hegyi**, J. Levendovszky, "Energy balancing cooperative diversity for wireless sensor networks," *International Journal of Communication Networks and Distributed Systems*, vol. 1, no. 4–6, 2008, pp. 524–543.

[2] **B. Hegyi**, J. Levendovszky, "Optimal statistical energy balancing protocols for wireless sensor networks," *WSEAS Transactions on Communications*, vol. 6, no. 5, May 2007, pp. 689–694.

[3] **B. Hegyi**, Á. Csurgay, W. Porod, "Investigation of the nonlinearity properties of the DC I-V characteristics of metal-insulator-metal (MIM) tunnel diodes with double-layer insulators," *Journal of Computational Electronics*, vol. 6, no. 1-3, Sep. 2007, pp. 159–162.

[4] **B. Hegyi**, J. Levendovszky, "Enhancing the performance of medical implant communication systems through cooperative diversity," *IEEE Journal on Selected Areas in Communications Special Issue on Wireless and Pervasive Communications for Healthcare*, under review.

Conference papers

[5] **B. Hegyi**, J. Levendovszky, "Efficient, distributed, multiple-relay selection procedures for cooperative communications," in *Proc. International Symposium on Wireless Pervasive Computing 2008*, Santorini, Greece, May 7-9 2008, pp. 170–174.

[6] **B. Hegyi**, Á. Csurgay, W. Porod, "Investigation of the nonlinearity properties of the DC I-V characteristics of metal-insulator-metal (MIM) tunnel diodes with double-layer insulators," *11th International Workshop on Computational Electronics*, Vienna, Austria, May 25-27 2006.