

# Development and Implementation of Cardiac Event Detectors in Digital CMOS

Joachim Neves Rodrigues

Lund 2005



**LUND INSTITUTE OF TECHNOLOGY**  
Lund University

Department of Electrosience  
Lund University  
P.O. Box 118  
SE-221 00 LUND  
SWEDEN

No. 56  
ISSN 1402-8662

© Joachim Neves Rodrigues 2005.  
Produced using L<sup>A</sup>T<sub>E</sub>X Documentation System.  
Printed in Sweden by *Tryckeriet i E-huset*, Lund.  
September 2005.

# Abstract

This doctoral dissertation presents the development and digital hardware realization of cardiac event detectors. Implantable medical appliances, as the cardiac pacemaker, have progressed from a life sustaining device to a device that considerably improves life quality for all ages. The number of electronic devices and household appliances in everyday life has an ongoing exponential growth. These devices contaminate their environment with electronic, magnetic or electromagnetic radiation. Pacemaker patients exposed to this environment may suffer due to malfunction of the pacemaker. Thus, the next generation of pacemakers require a low-power consuming event detector that provides reliable detection performance.

In this thesis two papers that present an artificial neural network based event detector for R-wave detection are merged to an extended manuscript. The neural network functions as a whitening filter prior to a matched filter. It is shown how the neural network responds to sudden changes in the input sequence. An algorithm that determines the initial template for matched filtering is proposed, and a continuous update of the filter impulse response is implemented in order to track long-term changes in signal morphology. Furthermore, an updated threshold function is proposed which addresses amplitude variations in the electrogram. Noise suppression and classification performance under “real-life situation” are explored by analyzing recordings from databases of electrograms and noise. Finally, the suitability for pacemaker application is discussed.

Four papers that present a low-power digital hardware implementation of a wavelet based event detector are merged and extended in the second part of this thesis. The theory of the wavelet filterbank is presented, and it is shown how the architecture was modified to achieve an area and power efficient silicon implementation. An algorithm is presented that determines automatically a threshold level during the initialization phase. A second operation mode is proposed to shut down major parts of the hardware, if the patient is at rest or in a “low-noise” environment. Power analysis on RTL-level shows that leakage power is the dominant factor in the total power figure. An estimate for leakage reduction is presented if sleep transistors are introduced between the supply rails and the logic that is shut-off in low-noise operation mode. The R-wave detector has been implemented in  $0.13\ \mu\text{m}$  low-leakage CMOS technology. The design has been routed, and, thereafter, sleep transistors are introduced in the layout. Detection performance is evaluated by means of databases containing electrograms to which five types of exogenic and endogenic interference are added. The results show that reliable detection is obtained at moderate and low SNRs.



# Contents

<b>Abstract</b>	<b>iii</b>
<b>Contents</b>	<b>v</b>
<b>Preface</b>	<b>ix</b>
<b>Acknowledgments</b>	<b>xi</b>
<b>List of Acronyms</b>	<b>xiii</b>
<b>Glossary</b>	<b>xv</b>
<b>Introduction</b>	<b>1</b>
1 Thesis Overview and Contribution . . . . .	1
1.1 Introduction . . . . .	1
1.2 Result Part . . . . .	2
2 The Heart and its Natural Control System . . . . .	5
2.1 Cardiac Charge Propagation . . . . .	5
2.2 The Malfunctioning Cardiac Control System . . . . .	9
2.3 Cardiac Fibrillation . . . . .	10
3 The Cardiac Pacemaker . . . . .	13
3.1 The Implanted Pacemaker – The Artificial Cardiac Control System . . . . .	21
3.2 Pacemaker Modes . . . . .	22
3.3 Variation in Sensing and Pacing Thresholds . . . . .	25
4 Energy and Power Dissipation in Digital CMOS . . . . .	27
4.1 Active Power Consumption . . . . .	28
4.2 Static Power Consumption . . . . .	29
4.2.1 Subthreshold Leakage . . . . .	32
4.2.2 pn Reverse Bias Leakage . . . . .	32
4.2.3 Leakage Temperature Sensitivity . . . . .	33
4.3 Active Power Minimization . . . . .	33
4.3.1 Dynamic Power Reduction . . . . .	34
4.4 Static Power Reduction Techniques . . . . .	39
4.4.1 Time-Multiplexing . . . . .	40
4.4.2 Supply Voltage Scaling . . . . .	41
4.4.3 Transistor Stacking . . . . .	42
4.4.4 Dual Threshold . . . . .	43

---

4.5	Energy vs. Power . . . . .	44
4.6	Digital Design Space . . . . .	48
4.6.1	Choice of Technology . . . . .	49
4.7	Design Considerations for Low-Leakage . . . . .	51
5	Artificial Neural Networks . . . . .	55
5.1	ANN Properties . . . . .	55
5.2	Feedforward Pass . . . . .	56
5.3	Back-Propagation Pass . . . . .	56
6	Wavelet Decomposition . . . . .	59
6.1	The Fourier Transform . . . . .	59
6.2	The Wavelet Transform . . . . .	61
6.2.1	The Mother Wavelet . . . . .	61
6.2.2	Scale . . . . .	61
6.2.3	The Continuous and Discrete Wavelet Transform . . . . .	62
6.3	DWT Realization . . . . .	63
	References . . . . .	67
<b>I</b>	<b>Implementation of an Artificial Neural Network Based Event Detector</b>	<b>77</b>
1	Introduction . . . . .	79
2	Databases . . . . .	80
2.1	EGM Database . . . . .	80
2.2	Interference Database . . . . .	82
3	Detector Structure . . . . .	84
3.1	The Whitening Filter . . . . .	84
3.2	The Time Lagged Feedforward Artificial Neural Network . . . . .	85
3.2.1	Forward Pass . . . . .	86
3.2.2	Back-Propagation Pass . . . . .	88
3.2.3	The Updated Learning Rate . . . . .	93
3.3	The Pulse-Shaping Filter . . . . .	93
3.3.1	The Initial Template for the Matched Filter . . . . .	94
3.3.2	The Update of the Impulse Response . . . . .	96
3.3.3	The Template Length of the Pulse-shaping Filter . . . . .	99
3.4	The Time-Varying Decision Rule . . . . .	99
3.4.1	Pausing of the Threshold Function . . . . .	101
4	Detection Performance . . . . .	102
4.1	Signal-to-Noise Ratio Definition . . . . .	102
4.2	Detection Performance for the Noise-Free Case . . . . .	103
4.3	Detection Performance for Interfered EGMs . . . . .	103
5	Discussion . . . . .	107

---

<b>II</b>	<b>Digital Implementation of a Wavelet Based Event Detector</b>	<b>113</b>
1	Introduction . . . . .	117
2	Materials and Methods . . . . .	119
2.1	Wavelet Filterbank and Generalized Likelihood Ratio Test . . . . .	119
2.2	The EGM Database . . . . .	122
2.3	Interference Database . . . . .	123
3	Digital Hardware Mapping and Optimization . . . . .	126
3.1	Implementation of the Wavelet Filterbank . . . . .	126
3.1.1	Dilation . . . . .	126
3.1.2	Time Alignment . . . . .	129
3.2	Implementation of the GLRT . . . . .	130
3.3	Hardware Optimization . . . . .	130
3.3.1	Optimization of the Wavelet Filterbank . . . . .	130
3.3.2	Optimization of the GLRT . . . . .	133
3.4	Sensing Threshold Initialization . . . . .	134
3.5	Dual Operation Mode . . . . .	134
3.6	Time-multiplexed Architecture . . . . .	135
3.7	Noise Detector . . . . .	136
3.8	Noise Detector Implementation . . . . .	138
4	Detection Performance . . . . .	139
4.1	Signal-to-Noise Ratio Definition . . . . .	139
4.2	Detection Performance for Noisy EGMs . . . . .	140
4.3	Detection Performance for Normal Mode . . . . .	140
5	Power Consumption . . . . .	144
5.1	Gated Supply Lines . . . . .	144
5.2	Leakage Reduction Estimation . . . . .	145
5.3	Gate Level Power Estimation . . . . .	147
5.4	Placement and Routing . . . . .	148
6	Conclusions . . . . .	148





# Preface

This thesis summarizes the results of my academic work in the Digital ASIC group at the department of Electroscience, Lund University, for the Ph.D. degree in circuit design. The main contributions of this thesis are derived from following publications:

J. Neves Rodrigues, V. Öwall, and L. Sörnmo, “QRS detection for pacemakers in a noisy environment using a time lagged artificial neural network,” *Proceedings of the 2001 IEEE International Symposium on Circuits and Systems, ISCAS 2001, Sydney, Australia*.

J. Neves Rodrigues, V. Öwall, and L. Sörnmo, “R-wave detection for pacemakers using a matched filter based on an artificial neural network,” *Proceedings of the 2002 IEEE International Conference on Neural Information Processing, ICONIP 2002, Singapore*.

J. Neves Rodrigues, V. Öwall, and L. Sörnmo, “A Wavelet Based R-wave Detector for Cardiac Pacemakers in 0.35 CMOS Technology,” *Proceedings of the 2004 IEEE International Symposium on Circuits and Systems, ISCAS 2004, Vancouver, Canada*.

J. Neves Rodrigues, V. Öwall, and L. Sörnmo, “A flexible wavelet filter structure for cardiac pacemakers: A power efficient implementation,” *Proceedings of the 2004 IEEE International Symposium on Biomedical Circuits and Systems, BIOCAS 2004, Singapore*.

J. Neves Rodrigues, T. Olsson, L. Sörnmo, and V. Öwall, “A Dual-Mode Wavelet Based R-Wave Detector using Single- $V_t$  for Leakage Reduction,” *Proceedings of the 2005 IEEE International Symposium on Circuits and Systems, ISCAS 2005, Kobe, Japan*.

J. Neves Rodrigues, T. Olsson, L. Sörnmo, and V. Öwall, “On the Digital Implementation of a Wavelet Based Event Detector for Cardiac Pacemakers,” *IEEE Transactions on Circuits and Systems. Special Issue on Biomedical Circuits and Systems: A New Wave of Technology*, accepted for publication.



# Acknowledgments

First of all I want to thank Viktor Öwall, my competent principal supervisor and travel mate during the past years. His willingness to discuss technical and other matters by any means had undoubtedly a great impact on this thesis. Our wild monkey experience in Malaysia, the rough sea in Canada, the Roykan Gaijin adventure in Japan, and “sorry, we have forgotten you, but the kitchen is closed now” in Italy, are certainly some moments I will not forget.

I was fortunate to have Professor Leif Sörnmo, as my second supervisor. Thank you for sharing your expertise and vast wisdom in biomedical signal processing and electrocardiography, and for improving my writing style.

I would like to thank Anders Berkemann for being my tutor when I took my very first steps in digital circuit design, and Thomas Olsson for his “low-level” ASIC experience, and for being my companion at IMEC.

I also extend my gratitude to the technical and administrative staff at the department, Erik for maintaining our network, Stefan for his CAD tool knowledge, Lars for fixing all the small things, and Pia, Stina, Elsbjeta, and Britta for all their assistance.

I am grateful to Fredrik and Matthias for reading parts of this thesis. I have enjoyed the company of Henrik, Hugo, Hong Tu, Martin, Peter, Zhan, Thomas, and the former Digital ASIC group members throughout the years.

Thanks to Michel for doing the pioneering work in this northern country, to Witty for being a loyal friend despite the distance, and, to Axel for sharing your thoughts.

I also would like to thank the Vinnova Competence center for circuit design (CCCD) for supporting this project and St-Jude Medical AB, Järfälla, Sweden for providing data for this study.

Finally, I would like to thank Anne for her commitment and support, and for being a wonderful mother to our son Nathan. Thanks to Nathan for being there.

Lund, October 2005

*Joachim Neves Rodrigues*



# List of Acronyms

AC	Alternating Current
ADC	Analog-to-Digital Converter
AF	Atrial Fibrillation
ANN	Artificial Neural Network
ASIC	Application-Specific Integrated Circuit
bpm	Beats Per Minute
BPEG	British Pacing and Electrophysiology Group
cc	Clock Cycle
CMOS	Complementary Metal Oxide Semiconductor
CWT	Continuous Wavelet Transform
DC	Direct Current
DSP	Digital Signal Processor <i>or</i> Digital Signal Processing
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
EDP	Energy Delay Product
EGM	Electrogram
FET	Field Effect Transistor
FFT	Fast Fourier Transform
FIR	Finite Impulse Response
FT	Fourier Transform
GLRT	Generalized Likelihood Ratio Test
HDL	Hardware Description Language
HS	High Speed
ICD	Implantable Cardioverter/ Defibrillator
LRI	Lower Rate Interval
LL	Low Leakage
LMS	Least Mean Square

LSB	Least Significant Bit
LUT	Look-Up Table
MSB	Most Significant Bit
MTCMOS	Multiple Threshold CMOS
NASPE	North American Society of Pacing and Electrophysiology
PCB	Printed Circuit Board
PDN	Pull-Down Network
PDA	Personal Digital Assistant
PDP	Power Delay Product
PUN	Pull-Up Network
RF	Radio Frequency
SoC	System on Chip
SNR	Signal-to-Noise Ratio
STFT	Short Time Fourier Transform
TLFN	Time Lagged Feedforward Network
VF	Ventricular Fibrillation
VHSIC	Very High Speed Integrated Circuit
VHDL	VHSIC Hardware Description Language
VRP	Ventricular Refractory Period
WT	Wavelet Transform

# Glossary

abdomen	Lower part of the body.
arrhythmia	Abnormal heart rhythm.
asynchronous pacing	The heart is stimulated at a constant, predefined rate.
atria	The upper chambers of the heart that receive blood from the body.
atrioventricular (AV) node	Node between atria and ventricles.
AV synchrony	Every atrial contraction is followed by a ventricular contraction.
bradycardia	A slow heart rate below 60 bpm.
cardiac output	Blood volume per minute pumped by a ventricle.
competitive pacing	Undesired pacemaker activity, competitive with the intrinsic heart rhythm.
defibrillator	An electronic device that applies a brief electrical shock to the heart, either directly or through electrodes on the chest.
depolarization	Polarity neutralization of the membrane.
ectopic rhythm	A heart rhythm that does not originate from the SA node.
ECG	Electrical activity of the heart recorded on the body surface using several electrodes.
EGM	Intracardiac activity recorded by an electrode placed inside the heart.
endocardium	Interior surface of the heart chambers.
epicardium	A conical sac of fibrous tissue that surrounds the heart.
fibrillation	Chamber quivering instead of pumping the blood effectively.

---

His bundle	Atrioventricular bundle is the extension of the atrioventricular node from the atrium across the fibrous skeleton of the heart to the ventricles.
lower rate interval (LRI)	Longest time period of two consecutive paced or sensed events in a chamber.
morphology	Structure of a sequence.
myocardium	Cardiac muscle.
Purkinje fibers	Network of conductive fibers in the ventricles.
refractory period	Period during which excitability is impossible.
sinoatrial (SA) node	A collection of cardiac muscle fibres at the junction of the superior vena cava and right atrium; origin of the cardiac cycle, known as the natural pacemaker of the heart.
synchronous pacing	Demand mode in which the pacemaker fires only when no event is sensed.
tachycardia	Rapid heart rate over 100 bpm.
thoracotomy	Surgical operation to open the chest cavity.
ventricles	The lower chambers in the heart that supply the body with blood.
ventricular refractory period (VRP)	A period after ventricular-based sensing or pacing where no sensing occurs.



# Introduction



## Chapter 1

# Thesis Overview and Contribution

This thesis deals with the development and digital hardware realization of detection algorithms for cardiac rhythm management devices, and is divided into three parts. The introduction presents briefly the mechanisms of the heart and the pacemaker, introduces low power digital circuit design, and presents the terminology of artificial neural networks and wavelet decomposition, and the second and third part present two implementations of cardiac event detectors: the first implementation is based on an artificial neural network in combination with a matched filter. The second implementation realizes a wavelet filterbank, resulting in an ultra low-power digital hardware implementation.

The chapters of this thesis cover a broad spectrum: the reader will find facts on the heart and the pacemaker, an introduction into biomedical signal processing on account of artificial neural networks and wavelet filterbanks, and digital hardware design on both high and low abstraction levels. This thesis is structured such that the reader, unfamiliar with the topic, may comprehend a digital hardware realization of a biomedical signal processing algorithm. The reader with more insight in the thematic may skip certain parts, since every section aims at one topic and may be read without the preceding sections.

### 1.1 Introduction

The first topic of the introduction is the basics of cardiology. The mechanism that keeps the heart beating is illustrated from a simplified engineering point of view with emphasis on cardiac pacing. Moreover, common cardiac diseases that require permanent pacemaker treatment are presented, together with historical highlights in artificial cardiac pacing and the functional characteristics

of cardiac pacemakers.

The second part of the introduction identifies the power consuming sources in digital circuit design. Commonly used power minimization techniques on both architectural and circuit levels are discussed. Finally, design recommendations to efficiently utilize energy sources are presented.

The signal processing part of the introduction briefly considers artificial neural networks. Moreover, the basics and advantages of wavelet decomposition are demonstrated by analyzing sequences.

## 1.2 Result Part

The result part presents two different types of R-wave detectors and is therefore split into two chapters.

**Part I** The first section proposes the application of a time-lagged artificial neural network for pacemaker application. The neural network functions as a whitening filter prior to a matched filter. It is shown how the update of the neural network weights is traced during training and how the network adapts to sudden changes. An algorithm that determines the initial template for matched filtering is proposed. A continuous update of the filter impulse response is implemented in order to track long-term changes in signal morphology. An updated threshold function is proposed which addresses amplitude variations in the electrogram. Noise suppression and classification performance under “real-life situations” are explored by analyzing recordings from databases with electrograms and noise. It is concluded that a hardware realization for pacemaker application is not feasible. The implemented features, which are required to obtain reliable detection performance, would result in a too high complexity and power dissipation, if implemented in digital hardware. Thus, a better approach that is suitable for pacemaker application is proposed.

**Part II** The second part presents a digital hardware realization of a wavelet based cardiac event detector. The theory of the wavelet filterbank is presented. Techniques such as numerical strength reduction and wordlength optimization are considered for hardware implantation. It is shown how the architecture can be modified to achieve an area and power efficient silicon implementation. An algorithm is presented that automatically determines a threshold level during the initialization phase. A second operation mode is proposed to shut down major parts of the hardware, if the patient is at rest or in a “low-noise” environment. For this purpose a low complexity noise detector has been developed that operates in supervision mode. A power analysis on RTL-level shows that leakage power is the dominant factor in the total power figure. An estimate for leakage reduction is presented if sleep transistors are introduced between the

supply rails and the logic that is shut-off in low-noise operation mode. The R-wave detector has been implemented in 0.13  $\mu\text{m}$  low-leakage CMOS technology using a standard ASIC design flow. The design has been routed, and, thereafter, sleep transistors are introduced in the layout. Detection performance is evaluated by means of databases containing electrograms to which five types of exogenic and endogenic interference are added. The results show that reliable detection is obtained at moderate and low SNRs.

The result section of this thesis is an extended version of the following publications:

### Part I

J. Neves Rodrigues, V. Öwall, and L. Sörnmo, "QRS detection for pacemakers in a noisy environment using a time lagged artificial neural network," *Proceedings of the 2001 IEEE International Symposium on Circuits and Systems, ISCAS 2001, Sydney, Australia*.

J. Neves Rodrigues, V. Öwall, and L. Sörnmo, "R-wave detection for pacemakers using a matched filter based on an artificial neural network," *Proceedings of the 2002 IEEE International Conference on Neural Information Processing, ICONIP 2002, Singapore*.

### Part II

J. Neves Rodrigues, V. Öwall, and L. Sörnmo, "A Wavelet Based R-wave Detector for Cardiac Pacemakers in 0.35 CMOS Technology," *Proceedings of the 2004 IEEE International Symposium on Circuits and Systems, ISCAS 2004, Vancouver, Canada*.

J. Neves Rodrigues, V. Öwall, and L. Sörnmo, "A flexible wavelet filter structure for cardiac pacemakers: A power efficient implementation," *Proceedings of the 2004 IEEE International Symposium on Biomedical Circuits and Systems, BIOCAS 2004, Singapore*.

J. Neves Rodrigues, T. Olsson, L. Sörnmo, and V. Öwall, "A Dual-Mode Wavelet Based R-Wave Detector using Single- $V_t$  for Leakage Reduction," *Proceedings of the 2005 IEEE International Symposium on Circuits and Systems, ISCAS 2005, Kobe, Japan*.

J. Neves Rodrigues, T. Olsson, L. Sörnmo, and V. Öwall, “On the Digital Implementation of a Wavelet Based Event Detector for Cardiac Pacemakers,” *IEEE Transactions on Circuits and Systems. Special Issue on Biomedical Circuits and Systems: A New Wave of Technology*, accepted for publication.

## Chapter 2

# The Heart and its Natural Control System

The human heart is a fist-sized muscle, usually located in the left chest. With a rate of 60–70 beats per minute it maintains the blood circulation of the body. The average human heart pumps 70 milliliters per heartbeat, and the average workload at the age of 70 has become 184,086,000 liters of blood by 2.5 billion beats. The heart consists of four chambers, an *atrium* and a *ventricle* on each side, separated by a muscular wall named *septum*. Atria and ventricles serve as blood in- and outlets, respectively. The walls of the heart are of muscular nature, referred to as the *myocardium*. The right atrium receives deoxygenated blood from the body through the *superior vena cava*. Figure 1 presents the heart and its connecting blood vessels. Blood in the atrium is pushed into the right ventricle by contraction. Thereafter, the blood is pumped into the lungs. The blood is oxygenated in the lungs before it returns to the left atrium and ventricle. Ventricular contraction pumps the blood through the *aorta* to supply the body with oxygen.

Cardiac output is due to consecutive contraction and relaxation of the myocardium. The first to contract is the atria and thereby filling the ventricles which very shortly after the atrial activity contract. Unidirectional blood flow is sustained by valves between the atria and ventricles.

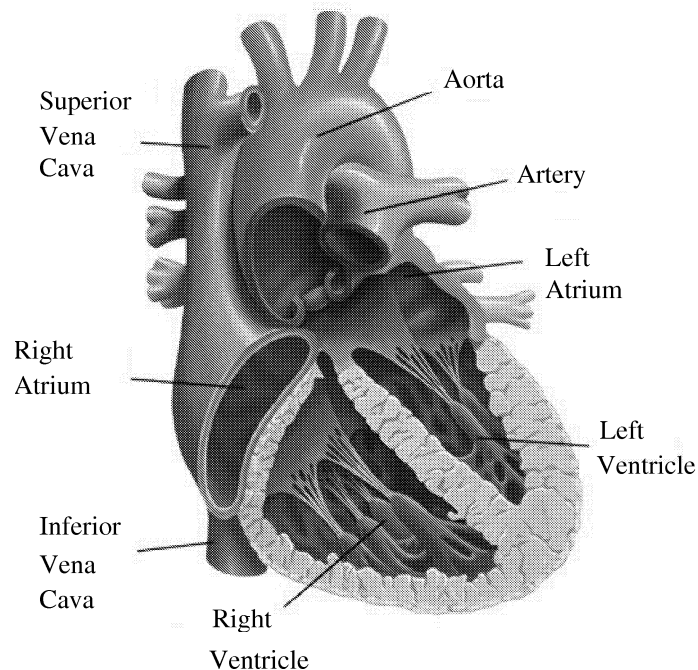
### 2.1 Cardiac Charge Propagation

The cardiac cycle is due to an electrical stimulus that originates at the natural pacemaker of the heart, called the *sinoatrial* (SA) *node*, located at the *epicardium* of the right atrium. The electrical impulse propagates to the atria, and slightly delayed through the *atrioventricular* (AV) node to the ventricles.

The short delay in impulse propagation assures that both atria and ventricles are filled with blood before contraction. The electrical impulse continues further through the conduction fiber in the septum, referred to as the *bundle of His*. Here, the impulse is split into the right and left bundle branch, and a wide network of conduction fibers, called the *Purkinje fibres*, propagate the impulse around the heart apex towards the atria. The origin of the cardiac events indicated in the electrogram and the stimuli propagation path are presented in Figure 2.

The impulse that originates at the SA node is transmitted by a change in electrical potential on the *epicardium*, along the described propagation path. The myocardial cells experience a potential difference across the cell membrane, due to unbalanced ion concentration between the exterior and interior of the cell. Thus, an electrical impulse can propagate as muscular cells are excited.

A deviation of the membrane permeability may cause a change in electrical potential as this results in an inward flow of positively charged potassium ions. Thus, a cell tends to be neutral or slightly positive, referred to as *depolarized*.

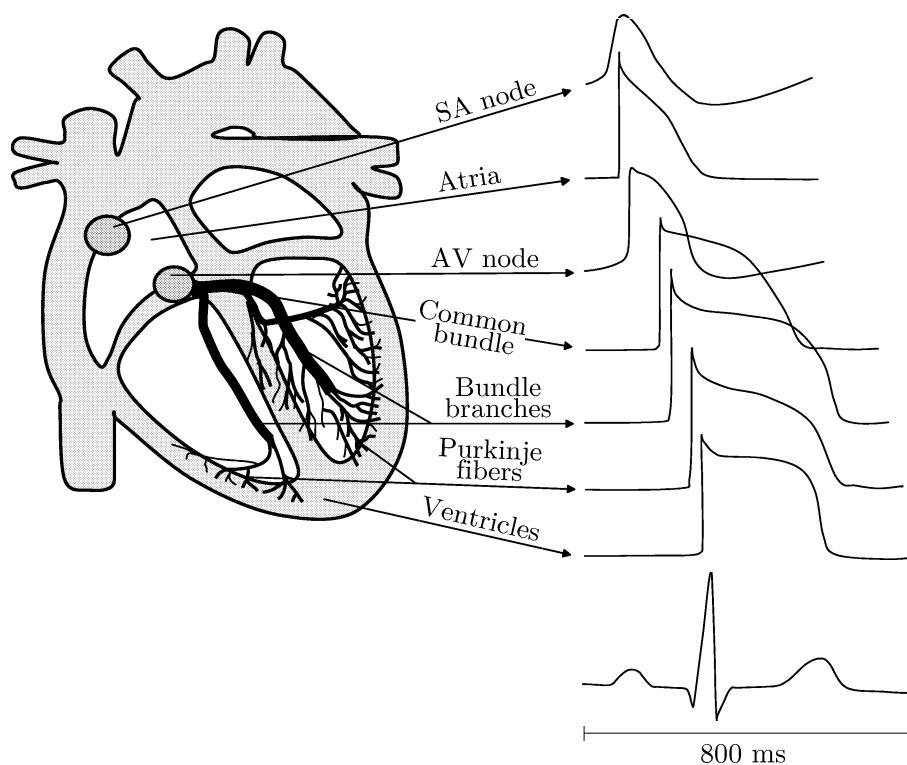


**Figure 1:** Anatomy of the heart and its connecting blood vessels. Reprinted with permission from St. Jude Medical.



Thereafter, the cell resumes to its negative potential and the cell is *repolarized*. Figure 3 illustrates the depolarization of the cardiac cells and the resultant development of the ECG. Magnitude and orientation of the dipoles during depolarization and repolarization can be represented by vectors. Summation of these vectors produces a dominant vector that indicates the charge direction on the epicardium, as presented in Figure 3. The myocardial cells cannot enter the depolarization phase immediately after repolarization. The time that needs to pass is referred to as the *refractory period*. The depolarization and repolarization correspond to the contraction and relaxation phase, respectively.

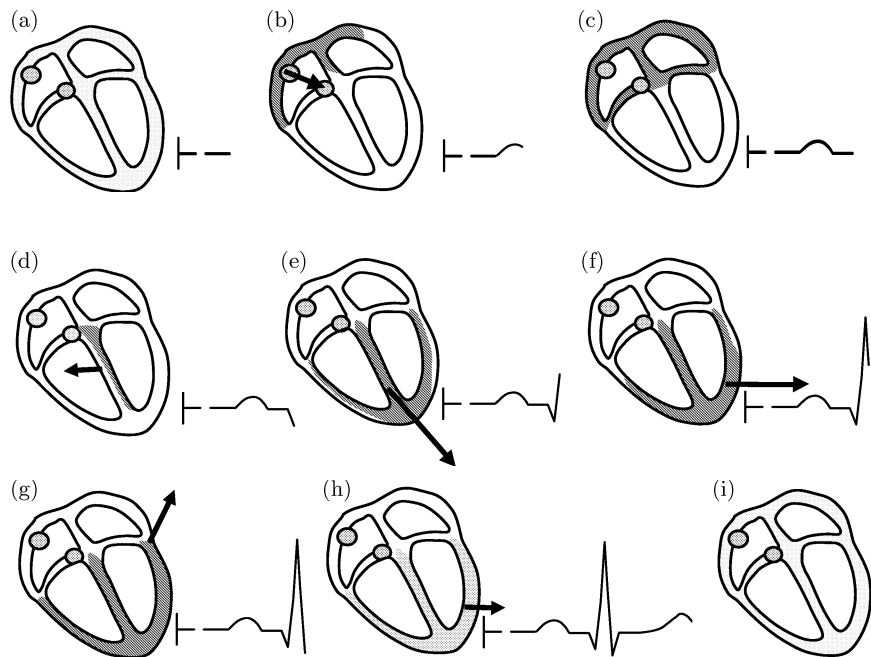
The cardiac charge propagation can be visualized by the *electrocardiogram* (ECG), which is measured on the body surface by several electrodes [1]. The first wave in the cardiac cycle, the P-wave, is the summation of the atrial cell depolarization that results in contraction and pumps blood into the ventricles,



**Figure 2:** The heart and its impulse propagation system. Morphology and timing of action potentials and their different origins in the heart and the related cardiac cycle of an ECG measured on the body surface [1].

as shown in Figure 4. The P-wave is low frequency with a spectrum well below 10 Hz, as shown in Figure 5.

The depolarization of the left and right ventricle is indicated by three waves gathered in the QRS complex that lasts about 0.1 seconds [2, 3], as illustrated in Figure 5. During depolarization of the ventricles both chambers contract and pump blood out of the ventricles, whereas the atria relax simultaneously from contraction during repolarization. The R-wave peak in the QRS complex is the sum of depolarization of individual cells in the chambers, which results in the highest signal level in the example of Figure 2. The T-wave represents repolarization of the ventricles and occurs approximately 300 ms after the QRS complex. The distance of the T-wave to the QRS complex is heart-rate dependent and becomes shorter for increasing heart rates. A typical ECG and the definition of normal time intervals are presented in Figure 4.



**Figure 3:** An ECG recorded at an electrode positioned at  $\vdash$ . (a) All cardiac cells at rest, (b) atrial depolarization, (c) electrical charge passes the AV node, (d) – (g) ventricular depolarization, (h) ventricular depolarization, and (i) all cells at rest [1].

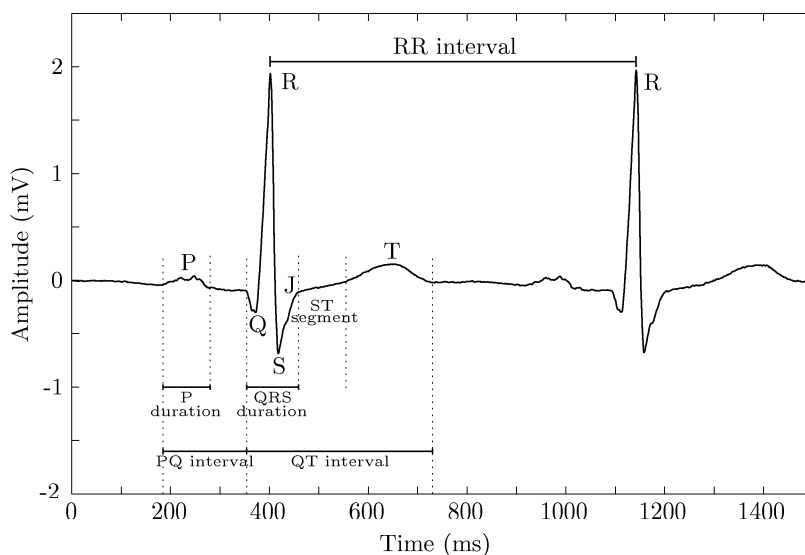
## 2.2 The Malfunctioning Cardiac Control System

Certain types of cardiac arrhythmia are, in most cases, an indication that a permanent pacemaker implant is required. If an arrhythmia becomes chronic it may cause a too slow heart rhythm. Several common types of arrhythmias exist which require pacemaker treatment, e.g., bradycardia, tachycardia, and heart block.

**Bradycardia** Bradycardia refers to a too slow heart rate to meet the demands of the body, where “too slow” depends on age and physical activity, but is usually below 60 bpm. This causes fatigue, dizziness, lightheadedness, fainting, extreme tiredness, poor exercise tolerance, and shortness of breath.

**Tachycardia** A rapid heart rate, usually above 100 bpm, during rest is referred to as tachycardia.

**Heart Block** A defect in the electrical conduction system is referred to as heart block, and prevents the electrical impulse to reach the ventricles such that cardiac activity may decrease dramatically.

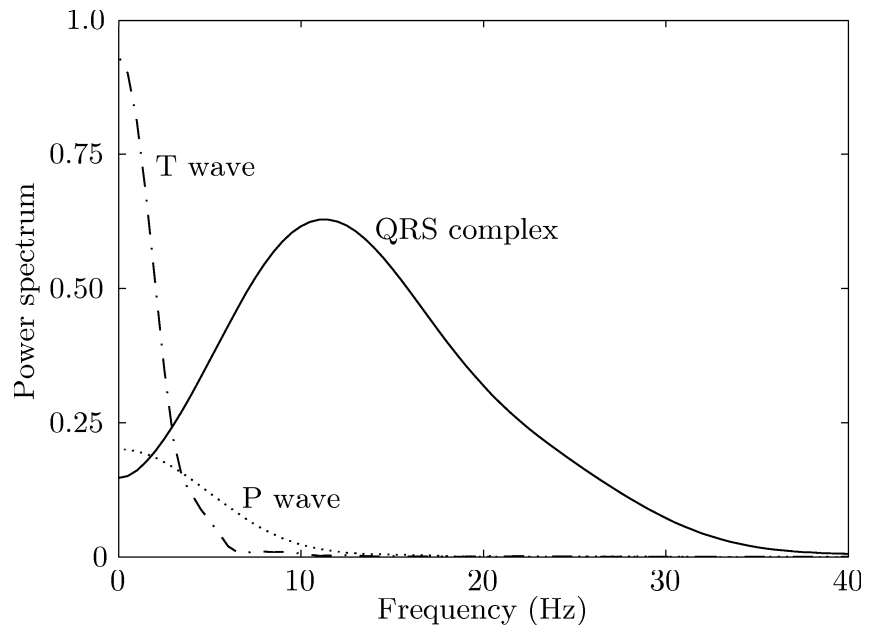


**Figure 4:** Wave definitions and durations of the cardiac cycle. The time location of J indicates where the QRS complex turns into the ST segment [1].

### 2.3 Cardiac Fibrillation

The stimuli that originates at the AV node is normally propagated as presented in Section 2.1. However, the charge propagation may be disorganized in some patients, causing *atrial* or *ventricular fibrillation*.

**Atrial Fibrillation** A cardiac cycle is usually initiated at the SA node. However, multiple regions in the atria may produce an electrical charge at different times. Such disorganized atrial stimuli generation results in *atrial fibrillation* (AF), i.e., the atria does not fill up with blood at the beginning of the cardiac cycle. These chaotic pulses, often at a rate of 300–600 bpm, are not further transmitted to the ventricles by the AV node, as such does not conduct during its recovery period. Thus, the ventricles contract at a disorganized rate and the cardiac output is reduced by approximately 10%. AF can be identified as waves having a varying amplitude, duration, and shape. The likelihood of developing atrial fibrillation increases with age and three to five percent of people over 65 suffer from atrial fibrillation.



**Figure 5:** Power spectra of the P-wave, QRS complex, and T-wave. Large variation may exist between different leads and subjects [1].

**Ventricular Fibrillation** Unsynchronized electrical activity in the ventricles result in a critical degradation of the cardiac output, i.e., the heart pumps only little or no blood, and is referred to as *ventricular fibrillation* (VF). This condition is very serious as a person collapses and sudden cardiac death will follow in minutes unless medical help is provided. Ventricular fibrillation is manifested by a particular wave morphology that is chaotic and without QRS complexes.

The ventricle may start fibrillating if an electrical impulse stimulates the ventricle while being in the *ventricle refractory period* (VRP), as the ventricle is vulnerable for fibrillation during the T-wave. Thus, an implanted pacemaker must assure that no stimulus is produced during the vulnerable time of the ventricle.



## Chapter 3

# The Cardiac Pacemaker

The cardiac pacemaker is one of the most successful implantable devices in the area of biomedical engineering. The implantable pacemaker, now in its forties, required almost nine decades of experiments and research before the first successful implantation was announced in 1958 [4]. This section gives a historical account of the predecessors of the implantable pacemaker and highlights the milestones of cardiac pacemaker development of the past decades.

**1871** The first mammal hearts to be artificially stimulated were chloroform-arrested hearts. Steiner forced cardiac arrest in three horses, one donkey, ten dogs, fourteen cats, and eight rabbits. The experiments were carried out by stabbing a 1 mm in diameter and 13 cm long, electrode needle through the chest into the ventricle. The anaesthetic level was deepened until cardiac arrest was indicated by the electrode needle stopped moving. The second electrode was connected by a moistened sponge to the epigastrium. The ventricle was stimulated by a DC stimulus that was interrupted by a metronome to successfully reanimate the animals [5].

**1872** The first human heart was paced by Greene in the United Kingdom. Sudden cardiac arrest was an occasional complication that could occur during chloroform anesthesia. It is documented that he reanimated five of seven cases of cardiac arrest by connecting hand-held electrodes, connected to a 300 V battery, to the neck and the lower left chest [6].

**1929 – 1932** Lidwill described his portable cardiac AC device used to reanimate a stillborn infant. The infant recovered completely after cardiac treatment. Lidwill's device required a wall socket and a needle inserted into the patient's ventricle [7].



**Figure 6:** A replica of the first implanted pacemaker molded in a shoe-polish can.

Hyman presented a pacing device driven by a hand-cranked spring-wound motor which provided electrical stimuli to the right atrium via an electrode needle. Unfortunately, such a pacing device was seen as an infernal machine that could interfere with the will of God, and no manufacturer dared to produce it [8].

**1948** The transistor is invented by Schockley's research group at Bell Laboratories [9].

**1952 – 1957** Paul Zoll announced the reanimation of a patient suffering from a serious cardiac disease. Zoll's external pacing device sustained the patient's heartbeat for more than 50 hours at a time [10,11]. Thereafter, the patient recovered sufficiently. The electrical stimuli was provided by a electrode needle to the ventricles. However, the external pacemaker of these days had the drawback that they were spacious, uncomfortable for the patient, and painful as the skin often was burned where the electrodes were connected to the skin.

The first wearable pacemaker was developed in 1957 by Earl Bakken [12]. Transistors were used instead of vacuum tubes. The battery-powered wearable device had the size of about a soap bar. Furman developed a long term transvenous lead that was used to maintain the heartbeat for 96 days.



**1958** Åke Senning and Rune Elmqvist implanted the first internal pacemaker in Stockholm, Sweden. The device lasted for a short time and reimplantation was required after five hours [4, 13]. The device incorporated two transistors driven by a nickel-cadmium battery. The pacemaker was molded with epoxy resin in a hockey puck sized shoe-polish can, as shown in Figure 6. The pacemaker battery was recharged by induction. An external charger unit maintained a power transfer by mutual induction. The charger coil was placed on the patient's skin above the pacemaker and power was generated by a coil within the implanted pacemaker. The charging process required several hours, usually done overnight, and the battery lasted a fortnight.<sup>1</sup>

The first *integrated circuit (IC)* was developed by Jack Kilby at Texas Instruments [14]. Almost simultaneously, Bob Noyce developed a similar device at Fairchild Semiconductor.<sup>2</sup>

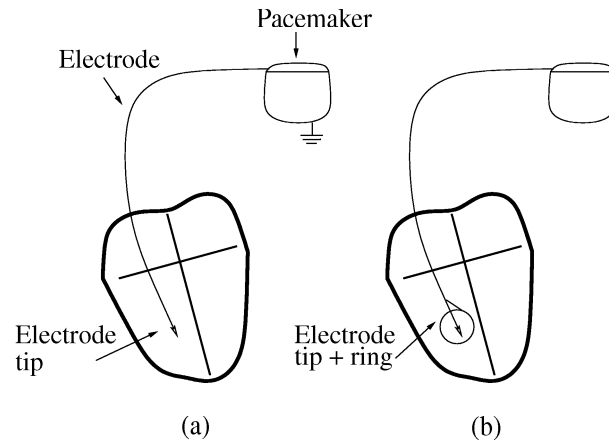
**1958 – Evolution of the Implantable Pacemaker** The major design constraints for a long-term implanted pacing device in the late 1950s were [3]:

- No interface in the skin through which infections could enter the body.
- A small battery with a high energy density that could be recharged or last for some years.
- A small pulse generator that could fit into the abdomen.
- A lead that could withstand the cardiac flexion.
- An electrode that sustained a reasonable threshold level.
- A biocompatible encapsulation.
- Circuitry shielded from both battery discharge and incursion of body fluids.
- Transistorized circuit that provides pulses of 2msec. Duration and amplitudes of 15 milliamperes at a steady rate of about 70 impulses per beat.

Progress in pacemaker development has accelerated since the implantation of the first device. Several improvements have been achieved in pacing lead technology, miniaturization, battery longevity, programmability, rate adaptive pacing, telemetry, and autoprogrammability.

<sup>1</sup>The first internal pacemaker patient was Arne Larsson who passed away on December 28, 2001, at the age of 86. At the time of his death, Larsson had received 26 pacemakers over a period of 43 years.

<sup>2</sup>Kilby received the Nobel price in Physics in 2000 for his part in the invention of the integrated circuit.



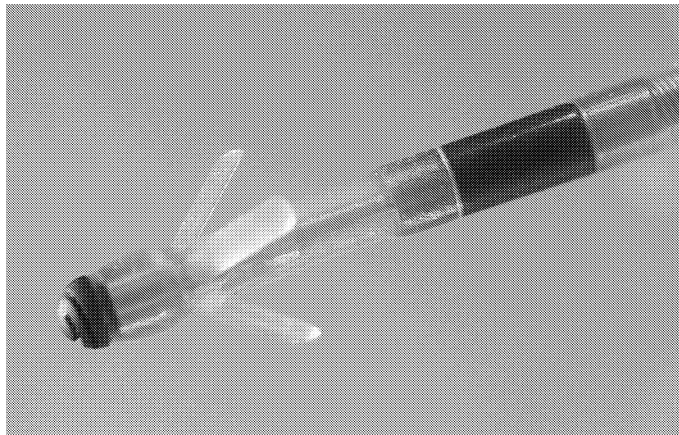
**Figure 7:** (a) The EGM is sensed through a unipolar electrode. Body tissue is the ground connection between pacemaker capsule and heart. (b) A bipolar electrode senses the EGM as potential difference between electrode tip and ring.

**Lead Technology** Breakage of the electrode wires was troublesome in the early days of the pacemaker. A durable wire that could withstand more than 36 million beats a year was a major design concern. A wire made of escapement spring of watches was developed in 1961 [15,16].

The early leads were all unipolar and gradually replaced by bipolar leads. The difference between uni- and bipolar sensing is depicted in Figure 7. A unipolar electrode uses the pacemaker capsule for GND connection, see Figure 7 (a), whereas a bipolar electrode uses a ring that surrounds the electrode tip for GND connection, see Figure 7 (b). Other improvements were the use of coaxial leads that reduced the lead diameter. A catheter electrode that could be passed into the right ventricle via a superficial chest vein made thoracotomy unnecessary [17].

Once it had been realized that the stimulation threshold may increase noticeably if the location of the electrode tip changes, research on electrode fixation techniques started. The electrode tip was anchored by barbs, hooks, loops in order to sustain a stable pacing threshold for a long time period. A typical tip of a lead used for pacemaker treatment is shown in Figure 8.

The discovery that current density at the electrode-heart interface causes cardiac stimulation, facilitated the development of smaller surface electrodes. Thereby, less current was required to achieve the same current density and, consequently, the pacemaker lifetime was prolonged.



**Figure 8:** A typical electrode tip with screw thread and barbs.

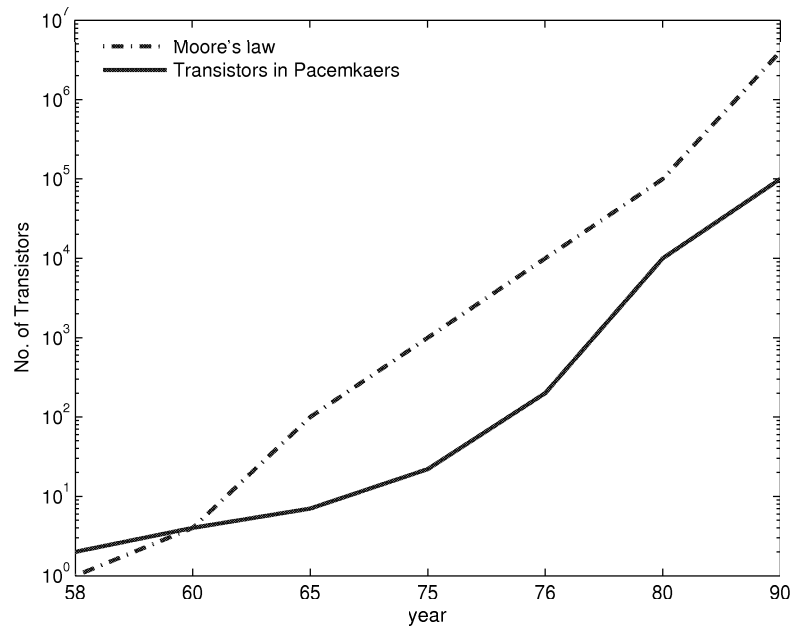
**Miniaturization** Miniaturization of cardiac pacemakers was enhanced by the application of transistors as a substitute for vacuum tubes. Thus, the dimension of a pacemaker changed from being “bread box” sized to a wearable device. Extensive functionality required more transistor which were accommodated on microprocessors. The number of transistors implemented in a pacemaker follows Moore’s law. From two transistors in 1958 the number has increased to over 200,000 in 1999, as presented in Figure 9 [18].

Hermetic sealing and biocompatible materials for encapsulation reduced the risk for moisture damage, component fracture and tissue reaction. The size of the pacemaker decreased from a device that had to be transported on a cart, to a device as small as 12.8 grams, (*Microny II SR+*, 2525, *ST-Jude Medical*), see Figure 10 [19].

**Battery Longevity** The first implantable pacemaker was powered by a rechargeable nickel-cadmium battery, which could operate for two weeks before recharging. To recharge the batteries a transmitter needed to be positioned accurately above the pacemaker. However, the majority of patients receiving permanent pacemaker treatment are elderly people who would have difficulties carrying out such a charging procedure correctly. Therefore, long lasting single-use batteries are used.

Long lasting nuclear-powered plutonium pacemakers were available for some time; however, they never became popular due to United States governmental restrictions.

The major breakthrough for pacemaker longevity occurred in 1968 when the lithium battery was invented (patented in 1971). Today’s pacemakers may

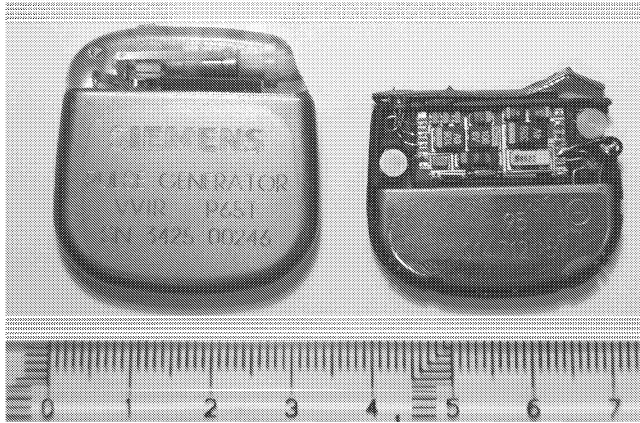


**Figure 9:** Moore's law and number of transistors implemented in a pacemaker from 1958 to 1990 [18].

last up to 20 years, e.g., the pacemaker *Regency SC+*, *St-Jude Medical* [19]. An opened pacemaker capsule is shown in Figure 10. The lithium battery has acquired approximately 50% of the total pacemaker volume. The circuitry on visible side of the PCB is basically composed of capacitors, a coil for telemetry, and an oscillator. The integrated circuits are on the other side of the PCB.

**Programmability** The pulse generator of early pacemakers maintained a fixed heart rate. However, rate adjustment was considered as early as 1957. The external pacemaker device developed by Bakken and Lillehei had a dial for rate adjustment and another to tune the electrical output [20]. In the 1960s pacemakers were available that used two insulated potentiometers for rate adjustment. The potentiometer could be adjusted with a specially shaped needle and thereby increase or decrease the pacing rate [21, 22].

Magnetic actuation was applied in 1972 to change the output pulse duration at a fixed output voltage. The pacemaker was programmed by an external device where bar magnets were manually rotated. Thereby, a rotation of the



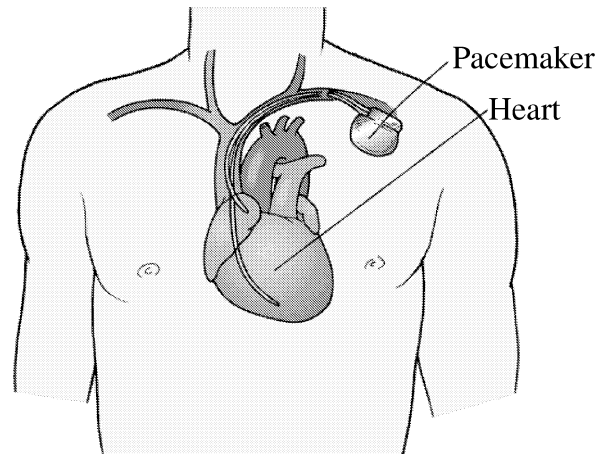
**Figure 10:** A typical pacemaker in full-scale. Half of the volume is needed for the battery.

magnets inside the pacemaker was caused and the gear train within the pulse generator adjusted.

Nowadays, radio frequency programming techniques are state-of-the-art [23]. The programmer contains a magnet that closes a magnetic switch and allows the pulse generator to be programmed.

**Rate-Adaptive Pacing** The pulse generators of the first pacemakers provided a fixed-rate stimuli. Regardless of the natural cardiac activity the ventricles were stimulated with a fixed rate, e.g., 70 or 100 bpm. However, in order to eliminate the danger of competitive pacing, i.e., an artificial stimuli is emitted during VRP and may cause fibrillation, research for rate-responsive pacing started. In 1966 clinical trials were reported with the first implanted R-wave sensitive pacemaker [24–27]. Thereafter, pacing electrodes have been used to both sense and force ventricular activity by R-wave detection, where a detected R-wave inhibited the artificial stimuli.

Rate-responsive pacing gave pacemaker patients not only a steady “natural” cardiac activity, but the rate response necessary to accomplish a normal life, e.g., exercising and working. Thus, pacemakers progressed from a life sustaining device to a device that considerably improves life quality.



**Figure 11:** Location of a pacemaker in the human body [28]. The pacemaker is sensing in both the ventricle and atrium.

**Telemetry** Bidirectional telemetry enables communication with the pacemaker. The possibility to retrieve information from the pacemaker, e.g., programming status, saved events, measured values, etc., has become very important [23]. Transfer of intracardiac data to external devices by telemetry enhances the understanding and treatment of cardiac diseases.

**Autoprogrammability** With the progress in microprocessor technology, more advanced features can be implemented in a pacemaker. Functionality, such as sensing threshold adaptation or rate-responsive pacing, makes it necessary that the pacemaker logic can adjust parameters automatically. Autoprogrammability was first used to tune parameters for rate responsive pacing, e.g., the rate needs to be increased if the patient is physically active. Other features like sensing and pacing threshold adjustments were also accomplished by autoprogrammability [29].

**Transvenous Implantation** Early pacemakers were implanted putting the patient under general anaesthesia, opening the abdomen and chest to expose the heart surface. Thus, pacemaker implantation was qualified as major surgery. Average mortality rates in the days after this surgery was about 7.5% in the 1960s [30].

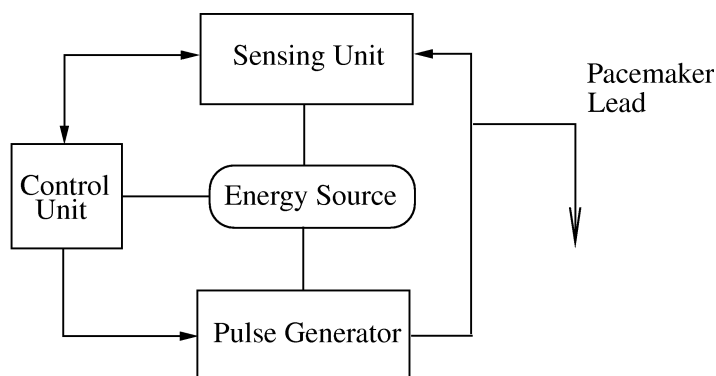
From the early 1970s, the lead was introduced in the inside of the right ventricle by making a small incision in the upper chest and advancing the

pacemaker lead down the vein. After positioning the lead in the ventricle, it was tested for functionality and electrical properties before connecting it to the pulse generator. The pacemaker was accommodated between the layers of subcutaneous tissue (underneath the skin) and muscle. Fibrous tissue grows over the electrode tip within days and binds the electrode to the inner wall of the heart. The surgery requires only local anesthesia and less than an hour for an experienced implanter. Thereafter, the patient needs some consultation for parameter adjustment. This procedure is state-of-the-art of today's pacemaker implantations.

### 3.1 The Implanted Pacemaker – The Artificial Cardiac Control System

The pacemaker is a permanent substitute or backup of the natural control system. Depending on the cardiac disease, various types of pacemakers are implemented to improve or sustain life quality. The operation mode of the pacemaker is divided into three classes: asynchronous (fixed rate), synchronous (on demand), and rate adaptive. Two different types of pacemaker exist: single chamber and dual chamber. The former type has only one lead that is either placed in the right atrium or ventricle. This type of pacemaker is often used in patients whose SA node sends out pulses too slowly (bradycardia). The latter type has two leads, one placed in the right atrium and the other in the right ventricle. Patients with a slow SA node, a blocked electrical pathway, or asynchronous atrial and ventricular contractions receive this kind of treatment.

Moreover, pacemakers are available which reduce the risk of atrial fibrillation. The pacemaker logic can be divided in three blocks: sensing unit, pacing unit, and control unit, as shown in Figure 12.



**Figure 12:** The pacemaker circuitry clustered in functional blocks [3].

**Sensing Unit** The sensing unit amplifies and filters the cardiac signal acquired at the electrode lead. Traditionally, filtering has been performed by an analog bandpass filter and a decision rule that compares the filtered signal to a reference value [3]. Thus, the logic of the sensing unit decides if an R-wave has occurred and accordingly triggers reset of the timing circuitry if the R-wave has been detected. Protection circuitry shields the sensing unit from pulses that originate from the pacing unit.

**Pacing Unit** The pacing unit generates the artificial pulse delivered to the myocardium. The voltage required to stimulate the myocardium is higher than the supply voltage of the battery, e.g., two times  $V_{Batt} = 5.6 V$ , and is generated by a pump-up capacitor. This capacitor is discharged on demand of the control unit in order to provide a stimulus. The voltage applied to the myocardium is referred to as the *stimulation threshold*.

**Control Unit** The control unit manages the timing circuitry and determines when a pulse needs to stimulate the myocardium. Today's pacemakers allow bidirectional data transfer: data is transmitted to the pacemaker to adjust parameters, e.g., sensing and stimulation threshold, timing parameters, and data that is read from the pacemaker required for diagnostics. The logic of the control unit is based on digital components accommodated on a CMOS IC [3].

**Energy Source** The energy required to drive the units for sensing, pacing, and control is provided by a lithium battery. As the battery cannot be replaced or charged, the entire pacemaker capsule needs to be replaced as soon as the battery weakens. The longevity of the battery strongly depends on the amount of pacemaker treatment needed, i.e., the number of generated pulses. The typical lifetime ranges from ten to twenty years.

## 3.2 Pacemaker Modes

The functionality of the pacemaker is deciphered by a code, see Table. 1. The table lists the code standardized by the *North American Society of Pacing and Electrophysiology (NASPE)* and the *British Pacing and Electrophysiology Group (BPEG)* [31,32]. All five positions are needed to classify available pacemakers. However, if adaptive rate pacing and multisite pacing are absent, positions I to III suffice. Pacemakers with rate modulation are identified by Position IV. Position V is used whenever the absence of adaptive rate pacing requires attention. Presence or absence of multisite pacing is denoted by all five positions [32].



**Position I: Chamber(s) Paced** Indicates the chamber that receives the electrical stimuli.

**Position II: Chamber(s) Sensed** Indicates the chambers in which spontaneous cardiac depolarizations or interference signals are sensed.

**Position III: Response to Sensing** Indicates whether sensing, as defined by Position II, inhibits pacing or triggers a pacemaker output immediately in the same chamber.

**Position IV: Rate Modulation** Indicates whether adaptive rate pacing (rate modulation) is available. Today's pulse generators are capable of comprehensive noninvasive adjustment and provide information by telemetry.

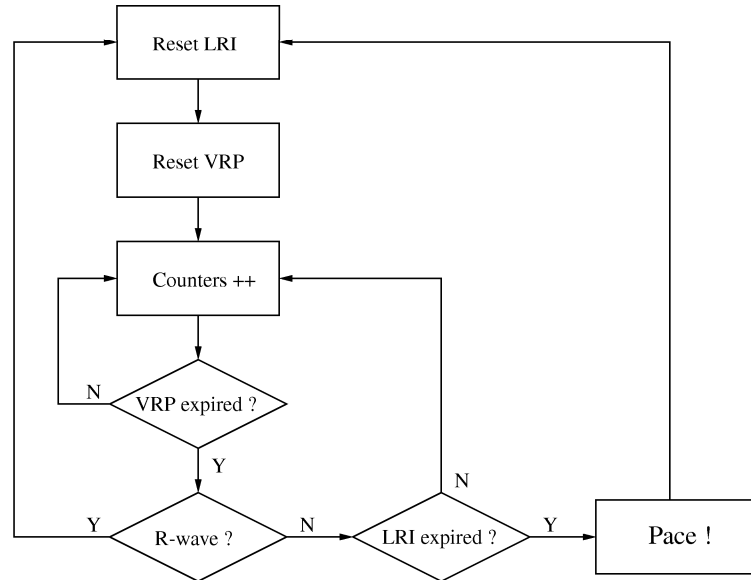
**Position V: Multisite Pacing** Indicates the presence and, to some extent, the location of multisite pacing.

The pacemaker implanted in 1958 would be coded as VOO: A ventricle is paced, no sensing, and thus no response to sensing, and, possesses only one timing interval, i.e., *lower rate interval* (LRI) [3].

A more advanced pacemaker is coded as VVI and delivers a stimulus to the ventricle whenever the heart fails to do so. This pacing mode is controlled by two timing cycles, i.e., LRI and the ventricle refractory period (VRP), as presented in the flow-chart of Figure 13. The flow diagram of a VVI mode single chamber pacemaker is a modified version of the flow diagram presented in [3]. The modification enables a sleep mode of the sensing logic during the VRP. Thus, the pacemaker logic can be switched off at least 20% of its operation

**Table 1:** Type codes for cardiac pacemakers. Atr. and Ven. denotes atrium and ventricle, respectively. Triggered and inhibited responses are indicated by T and I, respectively.

Position	I	II	III	IV	V
Category	Chamber(s) Paced	Chamber(s) Sensed	Response to Sensing	Rate Modulation	Modulation Pacing
Code	O = None A = Atr. V = Ven. D =Dual (A + V)	O = None A = Atr. V = Ven. D = Dual (A + V)	O = None A = Atr. V = Ven. D = Dual (T + I)	O = None R = Rate Modulation	O = None A = Atr. V = Ven. D = Dual (A + V)



**Figure 13:** The flow diagram of a VVI mode single chamber pacemaker is a modified version of the flow diagram presented in [3].

time and this results in reduced energy consumption and, thus, longer battery lifetime.

The timing cycle of a VVI pacemaker is always triggered either by a sensed or paced a ventricular event (R-wave). After the pacemaker has just stimulated the ventricle with a pulse, the internal counters for LRI and VRP are reset and the pacemaker increments its counters until the VRP has expired. The sensing logic can be set in sleep mode during this time period which is typically between 200 and 300 ms; when the VRP has expired, the sensing logic is activated to monitor activity in the ventricle. The sensing of a R-wave causes a restart of the pacemakers timing cycle by resetting the counter for LRI and VRP. However, if the LRI expires before a ventricle activity is sensed the pacemaker stimulates the ventricle with a pulse. To make the VVI pacemaker rate-adaptive (VVIR) the LRI must be adjustable, which can be achieved by using sensor-controlled LRI.

The most advanced pacemaker available is coded as DDD: atrial activity is sensed in both atria and ventricles, both atria and ventricles are paced taking into account six timing intervals.

### 3.3 Variation in Sensing and Pacing Thresholds

Sensing as well as pacing is carried out with the same electrode. However, due to time-dependent changes at the electrode tip, the thresholds of sensing and pacing need to be updated with time. A sensing threshold that follows changes in EGM morphology sustains reliable detection performance and, moreover, an optimized pacing threshold reduces the power consumption.

The pacing threshold does not remain static since, e.g., pain or anxiety during pacemaker surgery may cause lower thresholds whereas sleep increases the threshold level. A more significant but smooth change of the sensing threshold comes over the months. An electrode wire obtains the lowest threshold at the time of pacemaker implantation, referred to as *acute threshold*. Within four to six weeks the threshold increases to its highest level, approximately three or four times its acute level. Thereafter, the threshold falls and remains almost static at two to three times the acute level, called the *chronic threshold* [33]. These variations in pacing threshold are due to the following reasons: after initial pacemaker implantation the electrode tip has direct contact with the myocardium; inflammation of the myocardium separates the electrode from the myocardium and the threshold reaches its peak level after three weeks; after six months, fibrous tissue has grown and surrounds the electrode tip; this tissue layer usually remains unchanged and reaches the chronic threshold level.

Cardiac activity is sensed with the same electrode used for pacing. Thus, changes in conductivity, as described above, affect the sensing threshold as well. The EGM has the highest amplitude after pacemaker implantation and changes accordingly with conductivity variations at the electrode tip.



## Chapter 4

# Energy and Power Dissipation in Digital CMOS

Energy consumption is an essential design constraint for cellular devices such as mobile phones and PDAs. Consumers demand devices that have a long operation time between charging. However, for medical devices such as hearing aids and cardiac pacemakers, energy consumption is even more crucial. The batteries in a cardiac pacemaker of today can neither be replaced nor recharged. When the battery power weakens the entire device must be replaced, resulting in discomfort for the patient and high economical costs. The world's longest lasting pacemaker has an expected longevity of up to 20 years [19]. The limiting factor of lifetime in a pacemaker is usually the energy source that is progressively depleted by self-discharge and energy dissipated to drive the pacemaker components, as presented in Section 3.1. Improved features, such as the proposed wavelet-based event detection, must meet restricted energy constraints in order to be considered as a viable alternative. Therefore, it is necessary to minimize energy consumption of new circuitry.

Progress in digital CMOS technology enhances the functionality of digital hardware accommodated on very limited chip area [34–37]. However, technical issues such as heat dissipation and leakage pose obstacles that need to be overcome in current and future technologies. Thus, hardware realizations for implanted medical appliances need to be thoroughly designed in order to use the limited energy effectively.

The total power consumption of a digital circuit can be approximated as

$$P_{tot} = P_{switch} + P_{dp} + P_{stat}, \quad (1)$$

where  $P_{switch}$  is active power consumption,  $P_{dp}$  is short-circuit or direct-path consumption, and  $P_{stat}$  is static or leakage power consumption [38].

This chapter addresses power issues and is composed of three parts. The first part gives a brief introduction to the origin of power dissipation in digital circuits, as approximated in (1). The second part presents an overview of techniques on arithmetic and circuit level, used to reduce the different types of power consumption. Moreover, it is shown how coherency of power dissipating sources can be used to gain a highly effective energy reduction. The concluding part places emphasis on energy rather than on power consumption, and, it is shown that power reduction does not necessarily result in lower energy dissipation.

## 4.1 Active Power Consumption

Active power consumption, traditionally the dominating contributor to the overall power figure in digital hardware, originates from two sources: switching of the parasitic capacitors in transistors and wires, and a conducting path between  $V_{dd}$  and GND during transition. The switching power in a digital circuit is computed as

$$P_{switch} = \sum_{nodes} \alpha_i C_{L,i} f_j V_{dd,j}^2, \quad (2)$$

where  $\alpha_i$  is the switching activity in node  $i$ ,  $C_{L,i}$  the parasitic capacitance at node  $i$ ,  $f_j$  and  $V_{dd,j}$  the clock frequency and supply voltage of instance  $j$ , respectively [39–41]. The approximation in (2) assumes  $V_{dd} = V_{swing}$  and accounts for multiple supply voltages and clock frequencies. The power consumed by switching is dissipated as heat.

The parasitic capacitance of a transistor shrinks with the transistor dimensions. However, wiring capacitance gains a higher share in the total capacitance figure, with contemporary technology. The wiring capacitance of deep submicron technologies of today is an order of magnitude higher than transistor capacitance [38, 42].

The second source for active power consumption, direct-path power, is due to the conducting path between transistor power supplies, i.e.,  $V_{dd}$  and GND, and occurs when both NMOS and PMOS devices are active at the same time. This is referred to as direct-path power, approximated by

$$P_{dp} = \frac{t_r + t_f}{2} V_{DD} I_{peak} f, \quad (3)$$

where  $t_r$  and  $t_f$  are the rise and fall times, respectively, and  $I_{peak}$  is the peak current in the conducting path. Such a peak current is proportional to transistor dimensions.

The clock distribution net in a synchronous digital circuit is composed of buffers and wires, which results in a high capacitive load that need to be switched, and, thus, switching activity and clock frequency determine power

Long Channel	Short Channel	Very Short Channel	Nano Scaled
Negligible Leakage	Subthreshold Leakage	Subthreshold +Gate Leakage	Subthreshold +Gate +Reverse Bias Leakage
$L > 1 \mu\text{m}$	$L > 180 \text{ nm}$	$L > 90 \text{ nm}$	$L < 90 \text{ nm}$

**Figure 14:** Leakage origin in scaled technologies [50].

consumption. A clock-net power dissipation of nearly 50% of the total power dissipation has been observed for some designs [43, 44]. Table 2 presents power dissipation of selected synchronous clocked ASICs.

## 4.2 Static Power Consumption

Leakage is a minor contributor to the overall power figure in long channel CMOS devices above  $1 \mu\text{m}$ , as presented in Figure 14. However, according to estimates presented in [37, 51], leakage power will be the dominant power source in the future due to a lower threshold voltage in scaled technologies, as presented in Figure 15. Already in 2004 it was observed that a design in 90 nm CMOS technology dissipates more power than an identical design in 130 nm [52]. In appliances that operate at low clock frequencies leakage may already be the dominant power source in  $0.13 \mu\text{m}$  CMOS technology [53].

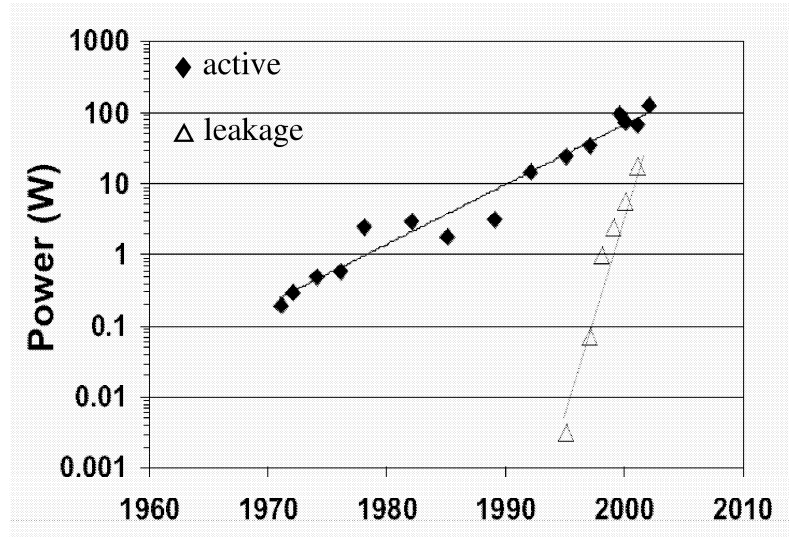
Static power consumption is due to leakage currents in the transistors [38, 41, 54], and can be approximated as

$$P_{stat} = I_{leak} V_{DD}, \quad (4)$$

where  $P_{stat}$  is linearly dependent of  $V_{dd}$  and the cumulative leakage current

**Table 2:** Percentage of power consumed in the clock-net for a selection of digital synchronous designs [45].

Design	power in clock net
high-performance CPU [43]	>45 %
MCORE microRISC [46]	36%
Alpha 21064 [47]	40%
Alpha 21164 [47]	40%
Alpha 21264 [48]	32%
TORCH MIPS R2000 [49]	36%



**Figure 15:** Processor power: active and leakage. Leakage power is going to replace the active power as dominant power source [37].

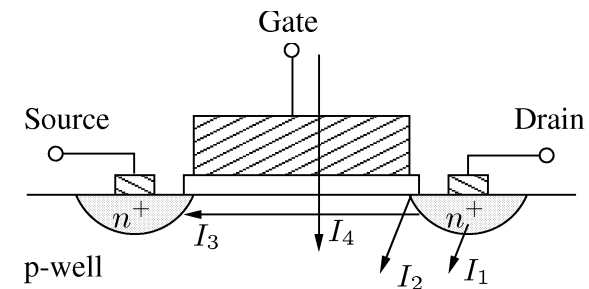
$I_{leak}$ , which consist of six different types of leakage [55]. A transistor model that illustrates these leakage currents is presented in Figure 16. The notations for the currents are:

- $I_1$ : reversed bias pn junction
- $I_2$ : gain-induced drain leakage (GIDL)
- $I_3$ : subthreshold + the drain induced barrier lowering (DIBL)
- $I_4$ : gate leakage due to oxide tunneling + gate current caused by hot-carrier injection

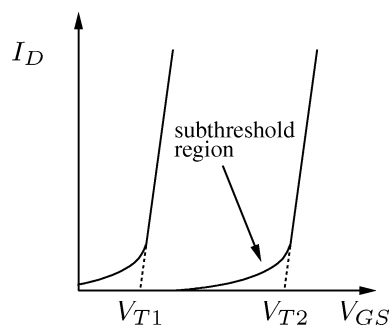
Contemporary CMOS technology suffer from subthreshold and gate leakage as shown in Figure 14. It is predicted that the application of high-k dielectrics will reduce gate leakage by a factor of 100, and is expected to be available from 2007 [50]. Technology below 90 nm will experience higher reversed bias leakage [54]. Gain-induced drain leakage as well as the pn junction leakage will gain higher significance with each future CMOS technologies [55].

The presented types of leakage are existent at certain transistor states, i.e., the leakage currents in off or on-state are denoted as  $I_{off}$  and  $I_{on}$ , respectively, as presented in Table 3. Revers-bias and oxide tunneling leakage are present in both *on* and *off* mode.





**Figure 16:** Leakage current mechanisms of a submicron transistor.



**Figure 17:** Subthreshold current for different  $V_T$ .

In submicron CMOS technology, a lower supply voltage reduces the electrical field strength and thus the power dissipation [56]. The ratio between supply voltage and threshold voltage determines the gate delay. In order to assure an improved gate delay with shrinking technology, the decrease in supply voltage

**Table 3:** Leakage presence in transistor modes.

current	transistor mode	
	ON	OFF
reverse bias pn junction	X	X
subthreshold	-	X
channel punchthrough	-	X
oxide tunneling	X	X
gate current	-	X
gain-induced drain	-	X

leads to a lower threshold voltage. Such threshold scaling results in an exponential increase of the subthreshold current. CMOS technology below 70 nm requires an oxide thickness of less than 1.5 nm, i.e., two to three layers of silicon dioxide atoms. Such thin gate oxide results in tunneling and in an exponential increase of gate leakage current [57–60], and, high-gate dielectrics have thus received much attention recently to prevent direct gate tunneling [61]. Since leakage is a limiting factor in future technologies, its reduction gains higher attraction with each new CMOS technology generation [38, 51, 62].

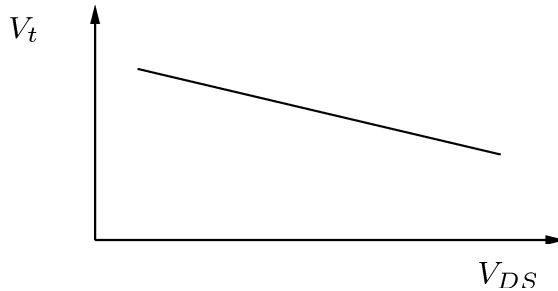
#### 4.2.1 Subthreshold Leakage

A transistor conducts as soon as  $V_{GS}$  exceeds the threshold voltage  $V_t$ . However, even when  $V_{GS}$  is below  $V_t$  (subthreshold region) a small drain-source current is present, as shown in Figure 17. The closer the threshold voltage is to zero voltage, the higher the leakage current.

**Drain-Induced Barrier Lowering** In long-channel devices the channel potential is not affected by channel length or drain bias as the depletion regions of source and drain are enough separated. However, as the source-drain distance shrinks with technology scaling, the source and drain depletion regions get closer and thereby affect the channel potential. Thus, the threshold voltage and the resulting leakage current vary with the drain bias, referred to as drain-induced barrier lowering (DIBL). Such an effect occurs if the depletion regions of source and drain are close enough to interact with each other, i.e., the source injects carriers into the channel surface. Thus, a high supply voltage to a short-channel device lowers the barrier height and thereby decreases the threshold voltage. Figure 18 illustrates how  $V_t$  changes with  $V_{dd}$  [55, 63, 64]. A lower switching threshold increases the subthreshold leakage as presented in Figure 17. Subthreshold leakage is the dominant power source in current and future CMOS technology [65].

#### 4.2.2 pn Reverse Bias Leakage

Drain and source to well junctions are reverse biased and cause a pn junction leakage current. This leakage current originates from two main sources: minority carrier diffusion/drift near the edge of the depletion region and electron-hole pair generation in the depletion region of the reverse-biased junction. pn junction reverse-bias leakage is dependent on junction area and doping concentration [66]. However, this leakage source has only a minor contribution in current and future CMOS technologies.



**Figure 18:** Drain-induced barrier lowering for short channel devices. The threshold voltage increases with  $V_{DS}$  decreasing and reduces thereby the leakage current [38].

#### 4.2.3 Leakage Temperature Sensitivity

Subthreshold leakage current is very sensitive to temperature changes. Power dissipation generates heat which in turn increases leakage power dissipation. Thus, it is necessary to sustain a low silicon temperature. This can be achieved by cooling devices, e.g., heat sinks. However, in cellular devices heat sinking is troublesome as this requires space and is resulting in higher production costs.

For implanted medical appliances the operating temperature is elevated and is equal to the body temperature. Thus, the normal operating temperature for a cardiac pacemaker lies at  $36.8^\circ\text{C}$ . An increase in operating temperature from room temperature  $25^\circ\text{C}$  to the body temperature results in an subthreshold leakage current increase of approximately 2.2 for a  $0.35\ \mu\text{m}$  CMOS technology [67]. For this technology, subthreshold leakage is the dominant component of  $I_{off}$ .

### 4.3 Active Power Minimization

Dynamic power consumption of a digital ASIC can be approximated by equation (2). Progress in technology shortens gate propagation delay with each new technology. Shorter gate delays lead to higher possible clock frequencies that results in higher switching and short-circuit power. Moreover, due to transistor size scaling it is possible to accommodate an increasing number of gates on a single die. This increases the total capacitive load and thereby the power dynamic power consumption. The predicted power consumption of a high-performance digital circuit will be more than 200 W in 65 nm technology, which puts high demands on cooling devices [68].

Dynamic power is dissipated as heat, which reduces the transistor threshold voltage, as previously presented. Since reduced threshold voltage increases the

subthreshold leakage current, it is necessary to have strong focus on dynamic power reduction as this affects leakage as well [54].

This section presents how dynamic power minimization techniques can be applied, and, how power sources are effectively addressed on architectural and circuit level. The in this section presented techniques are summarized in Table 4.

#### 4.3.1 Dynamic Power Reduction

Switching power can be reduced by a lower clock frequency, minimization of the capacitive load  $C_L$ , lower switching activity  $\alpha$ , and reduction of the supply voltage  $V_{dd}$ , according to (2). A reduction in clock frequency without computation power penalty can be achieved by parallelization. However, this leads to substantial increase in silicon area. Another option, according to (2), is the minimization of the total capacitive load. Such capacitive load reduction comes with technology scaling and is mainly due to minimization of the transistor gate and diffusion as well as wiring. Therefore, transistor dimensions should be kept at a minimum where applicable and reasonable. Unfortunately, such considerations are only possible on a layout level and most digital hardware designers are bound to a higher abstraction level, using a hardware description language (HDL). Switching activity can be minimized on an architectural level by applying certain design strategies. The parameter that achieves the highest power minimization is the supply voltage  $V_{dd}$ , as dynamic power in (2) depends quadratically on  $V_{dd}$ . This section presents design techniques used to minimize the switching activity. These design strategies can be divided in two groups: the first is optimization on architectural level; the second can additionally be applied on circuit level as presented in Table 4.

**Pipelining** The clock frequency of a digital circuit is usually determined by the critical path. The relationship between propagation delay and  $V_{dd}$  is

$$t_d \propto \frac{V_{DD}}{(V_{DD} - V_t)^\alpha}, \quad (5)$$

**Table 4:** Implementation strategies on architectural and circuit level.

Architectural	Circuit
Pipelining	Multiple $V_{dd}$
Parallelization	Multiple clocks
Wordlength Opt.	Clock gating
Arithmetical Opt.	$V_{dd}$ scaling
Strength Reduction	

where  $V_t$  is the switching threshold and  $\alpha$  the velocity saturation, i.e.,  $\alpha = 1.4$  for current CMOS technologies. [38, 69]. Thus, the maximum clock frequency, which is the reciprocal of (5), can be increased by lowering  $t_d$ . This can be achieved by introducing additional registers in the critical path, referred to as pipelining. A critical path that allows a higher clock frequency than needed permits a lower supply voltage, i.e.,  $V_{dd}$  can be scaled according to (5). Thus, pipelining is used to speed up the circuit or to reduce switching power.

**Parallelization** Parallel processing and pipelining techniques are dual aspects, i.e., if an implementation can be pipelined it can also be processed in parallel [70]. In a parallel implementation at least duplicated hardware is used to carry out computations simultaneously during one clock cycle. Hence, parallel processing results in higher computation speed. If a lower throughput is sufficient  $V_{dd}$  can be scaled. The disadvantage with a parallel implementation is the significant increase in silicon area.

**Balanced Path** A glitch develops if two or more merging signal paths in a design have different propagation delays. Implementing tree rather than chain structures and pipelining reduces the number of glitches and thereby the switching activity [39, 70]. A balanced signal path evens out the delays such that the number of ripples is minimized.

**Resource Allocation** A shared datapath is a popular technique to minimize implementation area. However, depending on the signal properties multiplexed datapaths may result in a higher switching activity [39]. If highly uncorrelated data is transmitted over the path, all data bits may flip in a worst-case situation. This charging and discharging operation increases the switching activity. The implementation of two independent datapaths may result in lower switching activity when data is correlated [41].

**Fixed-Point Wordlength Optimization** The wordlength after a mathematical operation needs to be increased in order to sustain precision and to avoid overflow, e.g., a 2's complement addition of two  $N$ -bit numbers requires a wordlength of  $N + 1$  for the sum. If several arithmetical operations are sequentially carried out in a signal path the wordlength needs to be increased accordingly, in order to avoid overflow. significantly. However, it is very unlikely that every arithmetical operation results in overflow if the wordlength is not increased. Thus, the wordlength of a result can be minimized after a completed operation by rounding or truncation. This may introduce rounding errors, and higher precision can be sustained by scaling the wordlength after several operations. Thus, wordlength optimization results in narrower

datapaths and consequently in narrower logic in the following hardware. Summarizing, the number of gates is reduced and, accordingly, capacitive load, switching activity and leakage [70].

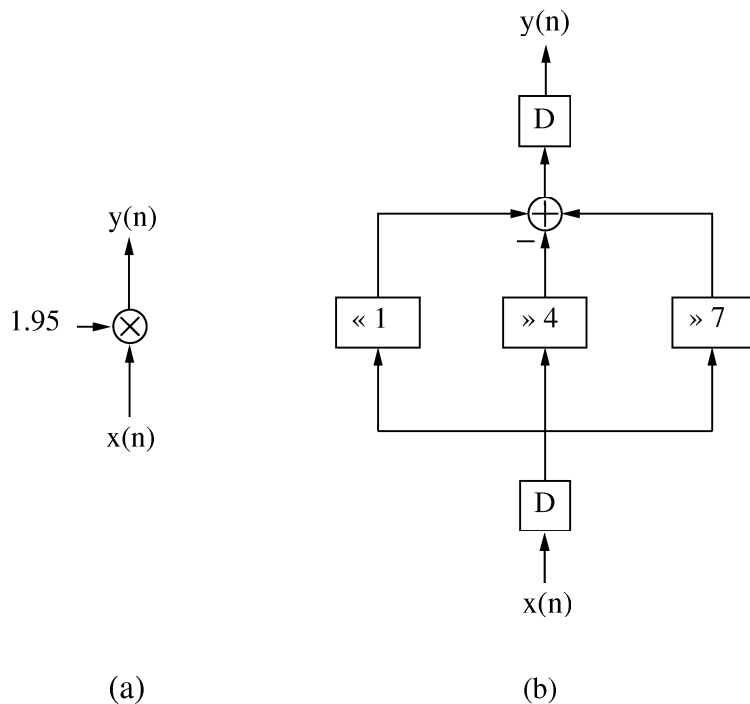
**Arithmetical Optimization** Signals in a digital circuit usually have different switching probabilities. This may range from a very high switching activity, i.e., a node switches every clock cycle, to a very low switching activity, e.g., a node which switches every  $10^{th}$  clock cycle or less. If the transition likelihood of a signal is known on architectural level, it is possible to minimize the switching activity: late introduction in the datapath of a signal with a high switching activity results in a lower switching activity internally and at the output node and can be achieved by reordering gate inputs [38].

**Numerical Strength Reduction** Numerical strength reduction can be applied to reduce arithmetical complexity in the implemented algorithm. The ranking of the basic operations in terms of required resources is:

1. Division
2. Multiplication
3. Subtraction / Addition
4. Bit-shift

The aim of numerical strength reduction is to transform a highly ranked operation into lower ranked operations, e.g., multiplication is restructured as a series of additions or bit-shift operations, or, the combination of both. Thereby, the performance in terms of area, power, and speed is improved. Figure 19 illustrates how a multiplication with a fixed-point number is accomplished by bit-shift and add operations. A multiplication with a fixed number, as shown in Figure 19 (a), is implemented by splitting the coefficient into elements, which are representable by powers of two, i.e., 1.95 is implemented as  $2 - 1/16 + 1/128$ , see Figure 19 (b). The coefficients can be hardwired by a left shift, and a truncation of 4 and 7 LSBs, respectively. The hardware that remains is an adder, which usually consumes less power than a fixed coefficient multiplier [71].

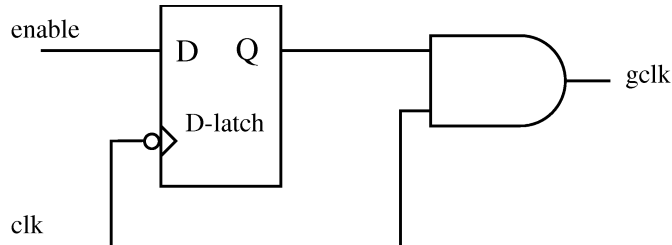
**Multiple Clock Domains** Digital designs can often be partitioned into blocks which have different demands on the throughput. These blocks may run on various clock and supply voltages and hardware that does not need to provide a high throughput can be triggered by a lower clock rate. Such local clocks can be generated by power efficient on chip generators. Thus, a partitioned ASIC that utilizes multiple clock domains is beneficial for dynamic power minimization [45].



**Figure 19:** (a) Fixed point multiplication implemented as (b) a bit-shift add instruction. The multiplicand 1.95 is implemented as  $2 - 1/16 + 1/128$ .

**Clock gating** The clock tree switching power has a significant share of the total power figure, as already previously presented. This power dissipation can be minimized by reducing the load of clock buffers or the switching activity. An efficient way to address the switching activity is the introduction of a gated clock, i.e., the clock can be enabled or disabled for some parts of the design. Thereby, it is possible to shut off parts of the design that do not operate continuously, e.g., hardware required during an initialization phase or hardware that is not required if operating in sleep mode. A logic that is used for clock gating is presented in Figure 20.

**Direct Path Power** Direct-path power can be minimized by keeping equal input and output rise and fall times [38, 41]. A possibility to eliminate the direct-path power is to keep  $V_{dd}$  below  $V_{tn} + V_{tp}$ , where  $V_{tn}$  is the threshold voltage for NMOS and  $V_{tp}$  the threshold for PMOS devices. If such a condition is met, the voltage between gate and source will always be below the threshold



**Figure 20:** Clock gating logic.

voltage of either device, and, thus, one of the transistors will always be shut off.

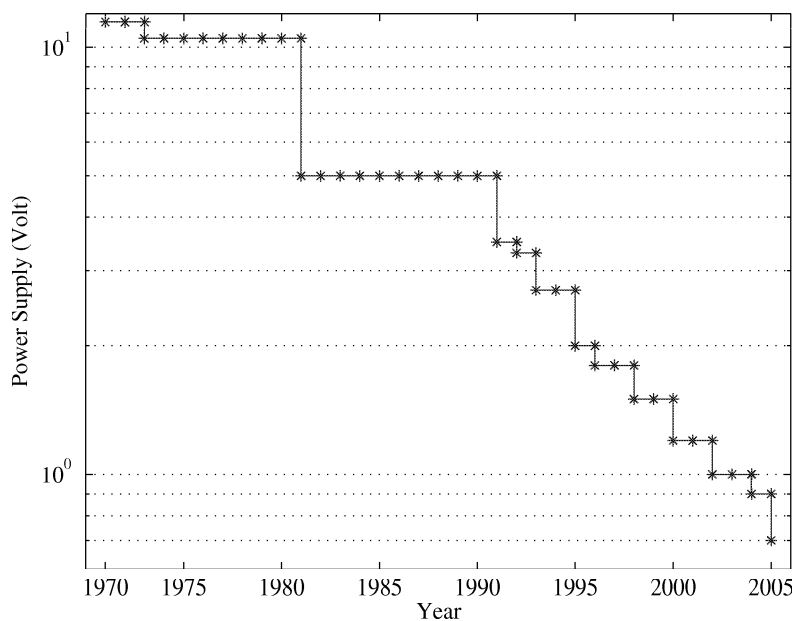
Direct-path power becomes less troublesome as the ratio of supply to threshold voltage is decreasing with shrinking technologies [38,41]. For typical  $0.13\ \mu\text{m}$  CMOS technology direct path power is not existent for supply voltages below  $0.7\ \text{V}$  and  $1.1\ \text{V}$  for high-speed and low-leakage technologies, respectively [72].

**Supply Voltage Scaling** Reduction of the supply voltage  $V_{dd}$  is the most effective power reduction technique since dynamic power consumption is quadratically dependent on  $V_{dd}$ . Various techniques, i.e., pipelining, and parallel computing are available to minimize the critical path in a design. The gate delay  $t_d \propto V_{DD}/(V_{DD} - V_t)^\alpha$  increases with decreasing power supply voltage. Thus, a shorter critical path and a reduced clock frequency facilitates a reduction in supply voltage. The application of supply voltage scaling achieves high reduction of the dynamic power consumption [41, 70].

A lower supply voltage often comes with a new generation in technology. In Figure 21 the drop in supply voltage over the last three decades is presented. It can be seen that the supply voltage has decreased significantly. From the early 90s the supply voltage decreased more rapidly compared to the 70s and 80s. From  $10\ \text{V}$  in the seventies,  $V_{dd}$  has decreased to  $0.7\ \text{V}$  using a  $65\ \text{nm}$  technology in 2005 [37, 73].

**Multiple Supply Voltages** Supply voltage scaling is an effective approach to reduce overall power consumption [38, 70]. However, for many designs this is not a suitable method as it introduces a delay increase in all gates, according to (5). Better overall performance is achieved by applying a reduced supply voltage to only some of the gates or blocks. Therefore, critical and non-critical blocks of the design are clustered, and powered by a supply voltage that meets the to the block associated delay constraints. The supply voltage of the non-critical clusters can be reduced as their speed requirement is lower with respect



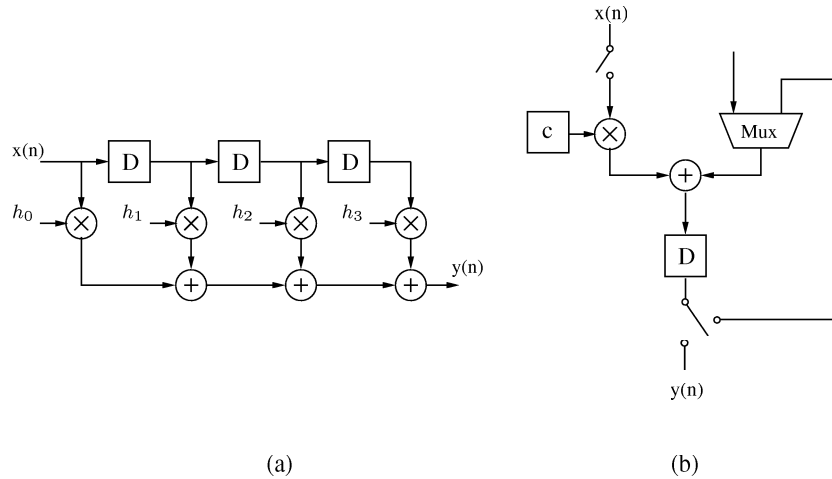


**Figure 21:** Scaling of CMOS supply voltages [37].

to the critical clusters [74,75]. Substantial local power minimization is achieved as  $V_{dd}$  has a quadratic impact on dynamic power. However, the use of multiple supply voltages requires an additional supply voltage that can either be fed through the I/Os or generated on chip by power effective DC-DC converters [76, 77]. Additionally, level converters are required whenever a low-voltage cluster drives a high voltage cluster [78].

#### 4.4 Static Power Reduction Techniques

Static power consumption contributes significantly to the overall power consumption in short-channel CMOS circuits. Leakage current is present in operation as well as in standby mode, as presented in Table 3. Therefore, leakage power needs to be addressed in both modes in order to reduce the power consumption effectively. Leakage minimization is achieved by applying architectural- and circuit-level techniques, and, if possible, by process-level techniques. However, optimization on process level is not possible in a standard design flow and, therefore, not considered in this thesis.



**Figure 22:** A FIR-filter implemented (a) direct-mapped, and (b) folded by 4.

This section discusses leakage power minimization techniques for digital CMOS circuits. On architectural level time-multiplexing is proposed for leakage reduction, and on circuit level voltage scaling, transistor stacking, and multiple threshold are presented.

#### 4.4.1 Time-Multiplexing

Leakage in a digital circuit can be reduced by minimizing the total gate width, i.e., gate count, which eliminates leakage sources. Optimization techniques that reduce gate count such as wordlength optimization or numerical strength reduction are already presented previously. Another method to minimize hardware resources is a time-multiplexed hardware realization, i.e., multiple instructions are computed by a single unit [70]. The basic concept of a time-multiplexed architecture is:

- partial computation of the operation and result storage
- reuse stored result and continue until entire operation is complete
- deliver result after completion and restart

The difference between direct mapped hardware and a folded implementation is demonstrated in Figure 22. The hardware in Figure 22 (a) and (b) is the

isomorphic-mapped and time-multiplexed realization of

$$y(n) = \sum_{k=0}^{N-1} h_k x(n-k),$$

respectively, where  $N = 4$  [79]. The hardware in Figure 22 (a) runs all operations in parallel during one clock cycle and the result  $y(n)$  is immediately valid. In Figure 22 (b) the product of  $x(n) h_0$  is computed and stored during one clock cycle, and three more cycles are required to compute a valid output, i.e., the architecture in Figure 22 (b) is folded by four. Equal throughput is possible when the clock frequency in Figure 22 (b) is four times higher than in Figure 22 (a). However, area cost of Figure 22 (b) is advantageous compared to Figure 22 (a). The time-multiplexed architecture requires control logic, memory or a register to store the coefficients, and a generalized multiplier, but the number of adders and fixed multipliers in the implementation in Figure 22 (b) is significantly reduced. A comparison of required hardware resources is presented in Table 5. The overhead for control logic and coefficient storage becomes negligible for more taps, compared to the hardware savings. Thus, leakage dissipation of the time-multiplexed architecture is approximately decreased by the folding factor.

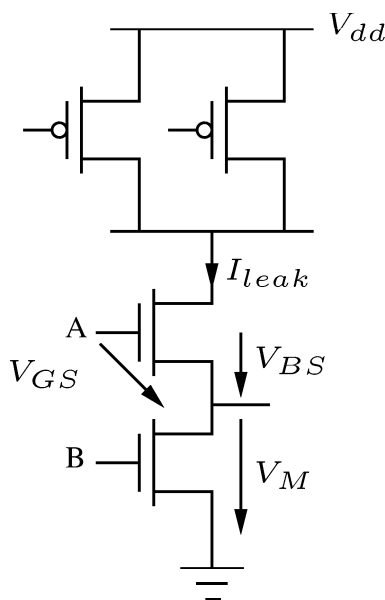
The switching power may increase in a time-multiplexed architecture as previously presented.

#### 4.4.2 Supply Voltage Scaling

Supply voltage scaling is a very effective technique to reduce the switching power due to the quadratic dependence of  $V_{dd}$ . Leakage power in (4) is linear dependent on  $V_{dd}$ . However, scaling of the supply voltage comes with a positive side effect: a lower supply voltage rises the barrier height which in turn increases the threshold voltage, referred to as DIBL, as already presented in Section 4.2.1

**Table 5:** Hardware resources for the FIR filter in Figure 22 implemented in a direct-mapped and a time-multiplexed architecture.

HW	direct mapped	folded
adder	N-1	1
fixed mult	N	-
general mult	-	1
register	N-1	1
control logic	-	1
memory	-	1
latency	1 cc	4 cc
mux	-	1

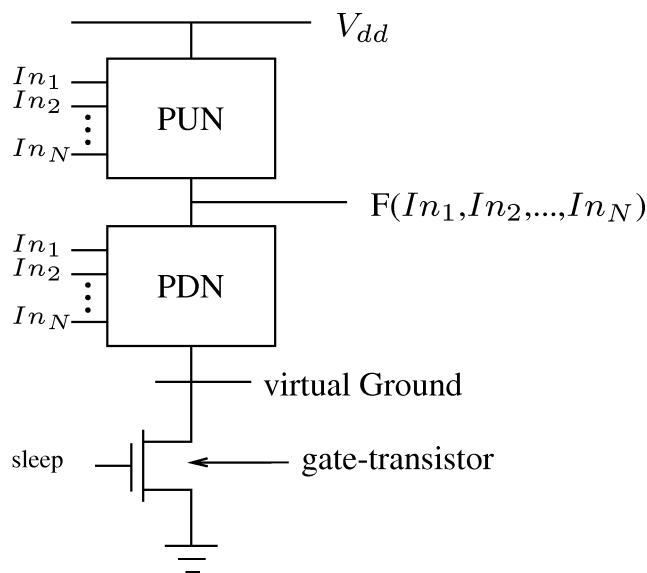


**Figure 23:** Stacked transistors in the PDN of a NAND gate.

[55]. A higher switching threshold result in less subthreshold leakage. Thus, voltage scaling enhances leakage power minimization more than linearly since  $P_{stat}$  is linearly reduced with  $V_{dd}$ , and, at the same time,  $I_{leak}$  decreases with decreasing  $V_{dd}$ . Voltage scaling for a  $0.13\ \mu\text{m}$  CMOS technology was proven to reduce leakage cubically [55, 80, 81].

#### 4.4.3 Transistor Stacking

Complementary CMOS design is comprised of a pull-up (PUN) and a pull-down network (PDN). Thus, the output of the gate is always connected through a low-resistance path to either  $V_{dd}$  or GND. Leakage current is reduced if at least two transistors in the PUN or PDN are connected in series, and the transistors are in off-mode. This is referred to as the stacking effect [82]. The stacking effect is demonstrated by two transistors in the pull-down network of a NAND gate, presented in Figure 23. If both transistors  $A$  and  $B$  are turned off, i.e.,  $A, B = 0$ , a small current between the supplies will induce a small positive voltage  $V_M$  at the intermediate node. This positive voltage at  $V_M$  increases the threshold voltage at  $A$  as the bulk-to-source voltage  $V_{BS}$  becomes negative and increases the threshold voltage. Such higher threshold voltage results in



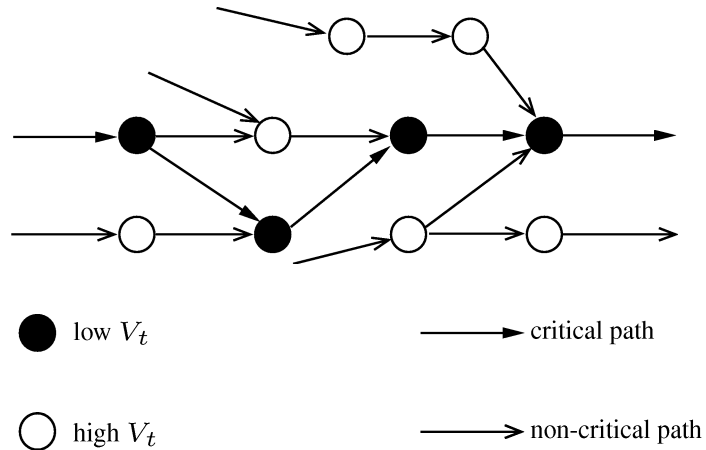
**Figure 24:** Complementary logic as a combination of a PUN and PDN driven by  $V_{dd}$  and a virtual GND. The sleep signal can originate either on- or off chip.

lower subthreshold leakage [82].

In a digital circuit sleep transistors can be introduced between the logic and the supply voltages, as presented in Figure 24. In this implementation the PUN is as usually connected to  $V_{dd}$ . The PDN is connected to a virtual GND, gated by a sleep transistor between the PDN and GND. The gate transistor is controlled on-chip by a *sleep* signal which turns the transistor off whenever the logic connected to this transistor is not needed. Thus, the logic experiences the stacking effect as soon as the *sleep* signal is low, and, thereby, leakage is reduced by at least order of tens. The application of a low-leaking gate transistor may reduce leakage by order of thousands [38, 55]. Another alternative that can be considered for leakage reduction is input-vector control [54]. However, this requires statistical information on the input vectors.

#### 4.4.4 Dual Threshold

The use of dual threshold CMOS technology provides transistors with both high- and low-threshold properties. Leakage power dissipation can be effectively minimized by dividing the design into critical and noncritical-paths. Leakage dissipation in the noncritical path is suppressed by high- $V_t$  devices, and, low- $V_t$



**Figure 25:** Dual  $V_t$  CMOS design [55].

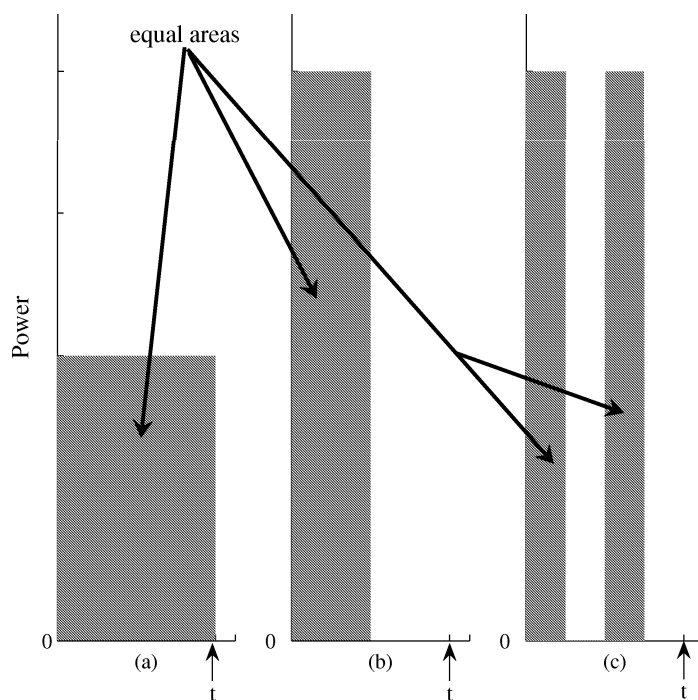
transistors in the critical path sustain speed performance, as demonstrated in Figure 25. Gate delay for a low leakage gate is about twice compared to a high speed gate in  $0.13\ \mu\text{m}$  CMOS technology [72].

The combination of both dual threshold technology and transistor stacking achieves a highly effective subthreshold leakage minimization. The implementation of a high- $V_t$  device as a sleep transistor minimizes the leakage dissipation by orders of magnitude. Optimal implemented dual- $V_t$  designs may operate at the same frequency as a design with purely low- $V_t$ , with a limited low- $V_t$  usage of 30%. It was observed that leakage power during active and standby mode is reduced by a factor 3 without any performance degradation [83].

**Reverse Body Bias** A higher  $V_t$ , and, thus, a lower leakage can be achieved by applying a negative voltage to the body bias, referred to as reversed body bias. Leakage savings of 14–55% compared to zero body bias transistors for nominal 70 nm and 50 nm transistors were observed in simulations [50]. However, this technique is not part of a traditional design flow.

#### 4.5 Energy vs. Power

Dynamic power consumption in a digital circuit is due to switching and short circuit power, as already presented in Section 4.1. Optimization techniques such as pipelining and parallelization are already presented in Section 4.3.1. Dynamic power as well as computational performance are proportional to the clock frequency  $f$ . If the clock frequency is lowered, less computational power



**Figure 26:** Process times of an arithmetical operation implemented as (a) continuous process, (b) burst mode process, and (c) time-multiplexed process. The processes run at different clock frequencies.

is available to carry out an operation. Thus, power consumption, according to (2) and (3), is reduced such that time is traded for power consumption.

However, the limiting parameter for cellular or medical implanted devices is the energy provided by the battery. Thus, energy needs to be minimized in order to prolong the device lifetime, and, unfortunately, power minimization techniques do not necessarily result in lower energy dissipation. This section illustrates what a designer should bear in mind when designing for low energy but purely focusing on power minimization techniques.

The difference between power and energy is illustrated with aid of Figure 26. The horizontal axis represents the execution time of an arithmetical operation, and the vertical axis represents the computation performance. Three different implementations perform the same logical operation and are powered by equal supply voltages. The hardware implementation in Figure 26 (a) runs

continuously during one clock cycle. No halt instruction is available and the design operates at a low clock frequency. The implementation in Figure 26 (b) operates in burst mode: the task is executed with a high clock frequency, and the hardware is halted as soon as the task has been completed. In Figure 26 (c), a time multiplexed architecture carries out the operation. The arithmetical operation is split and two clock cycles are needed to compute the result. The clock frequency is higher than in Figure 26 (b), required to provide the result at time instance  $t$ . Hardware is reused to minimize area, and, moreover, after task completion the hardware is halted as in design Figure 26 (b).

The energy dissipated by either of the implementations is represented by the area under the curves in Figure 26. In Figure 26 (a) energy is dissipated uniformly distributed during one clock cycle. The implementation in Figure 26 (b) dissipates twice the power of Figure 26 (a) in half the time, which results in equal energy dissipation compared to Figure 26 (a). The time-multiplexed architecture in Figure 26 (c) dissipates the same amount of energy as Figure 26 (a) and (b) during two clock cycles. Power dissipation of the implementation in Figure 26 (b) and (c) is higher than in Figure 26 (a). However, energy consumption of the different hardware realizations, represented by the area under the curves, is equal in all cases. The only difference is the amount of power dissipated during a certain time.

The amount of energy needed to switch the total capacitive load of an operation is approximated as

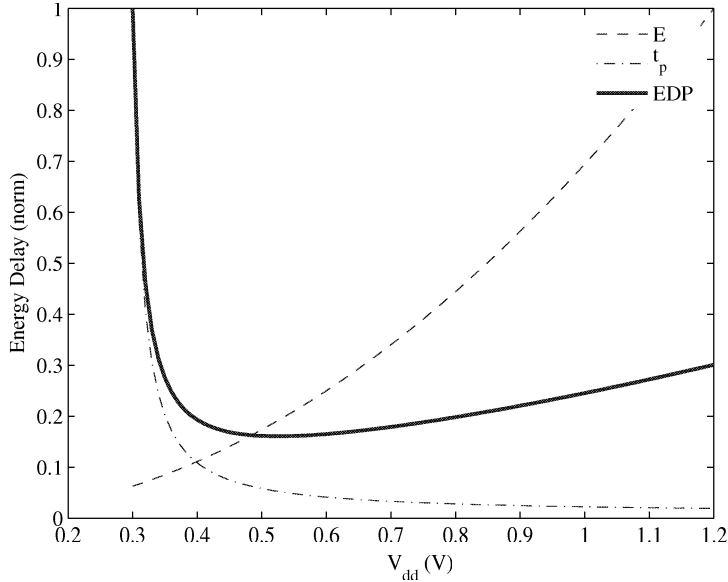
$$E_{switch} = C_L V_{dd}^2. \quad (6)$$

The energy in (6) dissipated during an operation is proportional to the capacitive load and quadratically dependent to the supply voltage [38]. A lower clock frequency results in reduced power consumption, but, at the same time, in a longer execution time. Thus, energy dissipation is not reduced by a lower clock frequency.

Due to the quadratic dependence of  $V_{dd}$  on energy dissipation in (6), scaling of the supply voltage results in significant energy savings. However, a lower  $V_{dd}$  increases propagation delay, according to (5). Thus, the maximum possible clocking frequency is reduced. A reduction in supply voltage trades energy for time [38].

The supply voltage needed to facilitate the clock frequency in Figure 26 (b) and (c) is higher than in Figure 26 (a). Thus the implementation in Figure 26 (a) permits a lower supply voltage, and, therefore, dissipates less energy than the implementation in Figure 26 (b) and (c).





**Figure 27:** Simulated delay, energy, and energy-delay plots for a typical  $0.13\ \mu\text{m}$  CMOS technology [72].

**Energy Delay Product** Scaling of the supply voltage results in a slower circuitry but benefits energy dissipation, and the opposite is true for high supply voltages. A further analysis is carried out in order to find an optimum supply voltage. The power-delay product (PDP) of a gate is defined as

$$\text{PDP} = C_L V_{dd}^2 f_{max} t_p, \quad (7)$$

where  $f_{max} = 1/(t_{pHL} + t_{pLH}) = 1/2t_p$  is the highest possible clock rate [38], which leads to

$$\text{PDP} = \frac{C_L V_{dd}^2}{2}. \quad (8)$$

The definition of PDP in (8) is used to define a measure of performance and energy as

$$\text{EDP} = \text{PDP} t_p = \frac{C_L V_{dd}^2}{2} t_p, \quad (9)$$

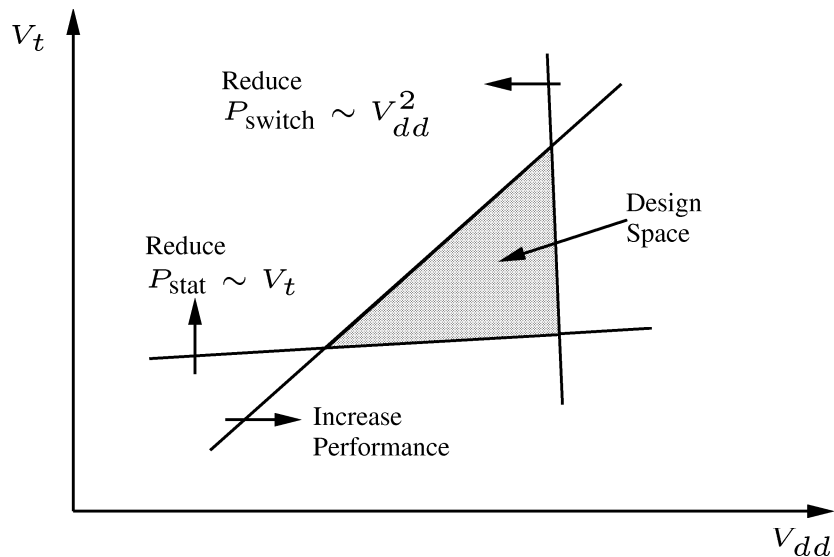
referred to as the energy-delay product. An optimum supply voltage can be found by plotting EDP as presented in Figure 27. The curves are obtained

by simulating delay, energy dissipation, and EDP for a typical  $0.13\ \mu\text{m}$  CMOS process with  $V_t = 0.28\ \text{V}$ . It can be seen that energy increases quadratically and the delay decreases exponentially with increasing supply voltage. An optimum supply voltage is found around  $0.5\ \text{V}$ . Thus, scaling of  $V_{dd}$  from  $1.2\ \text{V}$  to  $0.5\ \text{V}$  results in  $82\%$  dynamic energy savings. Lower leakage dissipation, due to a lower  $V_{dd}$ , is not considered in this analysis.

#### 4.6 Digital Design Space

The progress in technology still obeys Moore's law [34, 37]. The transistor channel length is scaled with each new technology, and, thus, transistor and gate sizes are scaled. Accordingly, more gates can be accommodated on a single die and the number of on-chip transistors is doubled every second year. The increase in performance comes at a high cost from a power perspective:

- higher chip clock frequencies,
- increase of the interconnect overall capacitance and resistance,
- exponentially growing leakage power,
- higher power density and distribution.



**Figure 28:** Design dilemma due to  $V_t$ ,  $V_{dd}$  and performance [84]. Design space is shrinking with progress in technology.

Increased clock frequency results in a higher dynamic power consumption and issues such as cross-talk needs to be addressed [38]. Moreover, the threshold voltage shrinks with technology scaling and increases the leakage power exponentially. Thus, a higher on-chip power density is the consequence of these power sources and requires advanced heat sink considerations.

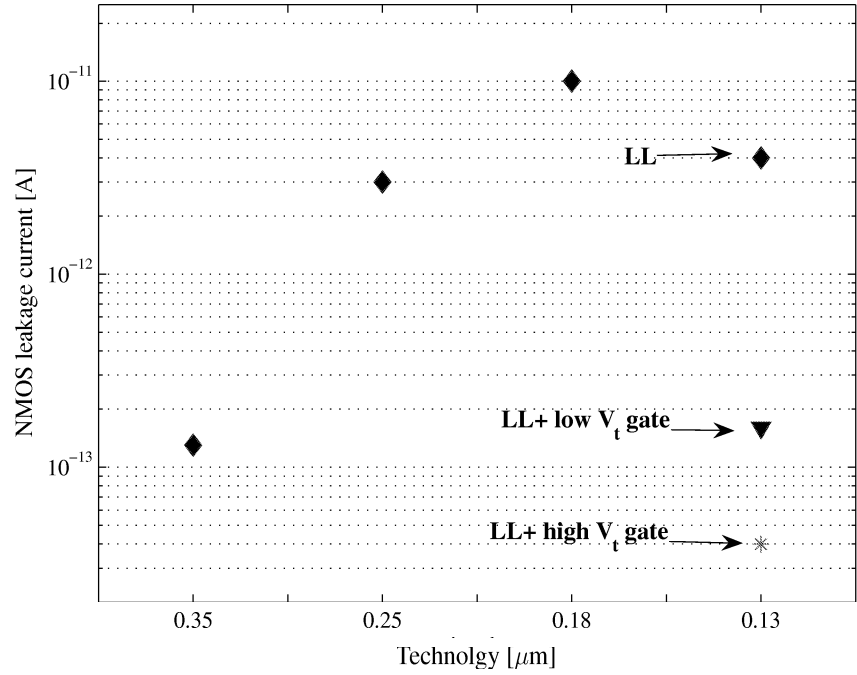
System-on-Chip (SoC) designers have to deal with this progress and it is required to continuously keep track of the roadmap since the design space is getting smaller with shrinking technology, as presented in Figure 28. Available design space, limited by performance, static power, and dynamic power dissipation, is indicated by the shaded area. Designs that have been implemented a decade ago may have to be revised in order to fit into the new design space.

#### 4.6.1 Choice of Technology

Until recently dynamic power consumption was the dominant source in digital circuits. A reduction in power consumption was achieved with each new technology due to a reduction of  $C_L$  and  $V_{dd}$ , e.g., Intel's 180 nm Pentium running at 2 GHz drew 72 W, the 130 nm Pentium running at the same speed consumed 52 W [52]. However, with device scaling leakage takes up bigger shares of the total power consumption. Intel experienced an increase in power, shifting from the 130 nm Pentium 4 (P4) to the 90 nm P4; the processors did operate at the same clock frequency, i.e., 3 GHz, but the 0.13  $\mu\text{m}$  processor consumed less power than its competitor fabricated in 90 nm. This behavior was traced back to an increase in leakage power [52].

The increase in leakage with technology progress for UMC processes is presented in Figure 29 [72]; leakage increases by a factor of 23 and 76 shifting from 0.35  $\mu\text{m}$  to 0.25  $\mu\text{m}$  and 0.18  $\mu\text{m}$ , respectively. The introduction of a low-leakage 0.13  $\mu\text{m}$  library (LL in Figure 29) that uses high- $V_t$  devices brought leakage down to a factor of 30, related to the 0.35  $\mu\text{m}$  device.

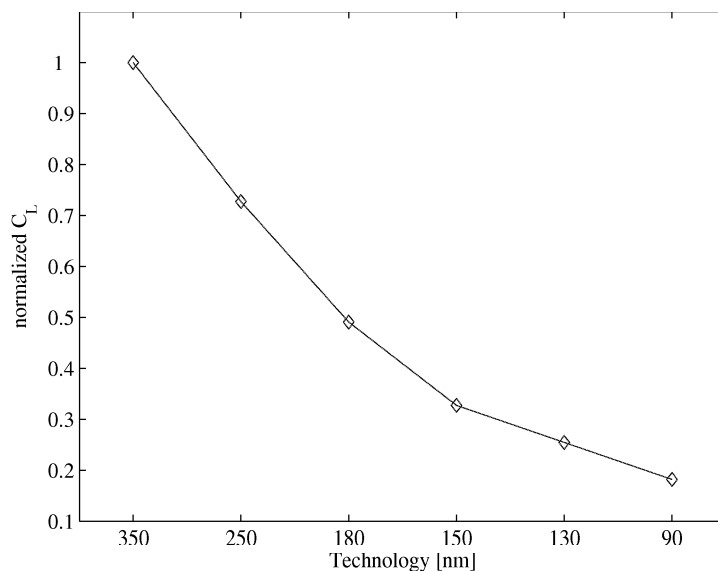
Depending on the target application, the most suitable technology needs to be determined. For devices running at very high clock frequencies a low-leakage 0.13  $\mu\text{m}$  technology is not suitable since gate delay of a low-leakage (LL) gate is approximately doubled compared to a high-speed gate [72]. However, designs that operate at lower clock frequencies can utilize low-leakage technology. Leakage power can be even further reduced by the introduction of sleep transistors in the power supply rails, as previously presented, for parts of the design that do not operate constantly. Sleep transistors reduce leakage by orders of tens and by a thousand times using high- $V_t$  and low- $V_t$  devices, respectively [38]. Leakage in Figure 29 for a 0.13  $\mu\text{m}$  devices with gated supply is approximated with a leakage reduction efficiency of 95% and 99.9% (LL+ low  $V_t$  and LL+ high  $V_t$ ) using low- and high- $V_t$  sleep transistors, respectively.



**Figure 29:** NMOS off-leakage currents in scaled technologies [72].

Moreover, scaling of the supply voltage raises  $V_t$  due to the DIBL phenomena, and, thereby reduces leakage [65]. Thus, it is expected that  $0.13 \mu\text{m}$  technology leaks less than demonstrated in Figure 29.

Gate capacitance of a typical  $0.13 \mu\text{m}$  LL CMOS technology is one fourth of the gate capacitance in  $0.35 \mu\text{m}$  CMOS technology, as shown in Figure 30 [72]. Correspondingly, dynamic energy consumption is reduced by a factor of four. A lower gate capacitance, introduction of sleep transistors, DIBL effect, and supply voltage scaling makes  $0.13 \mu\text{m}$  LL CMOS technology a reasonable choice for applications that operate at supply voltages slightly above the switching threshold. Decrease in dynamic power consumption with scaled technologies is presented in Table 6.



**Figure 30:** Normalized capacitive transistor load for scaled technologies [72].

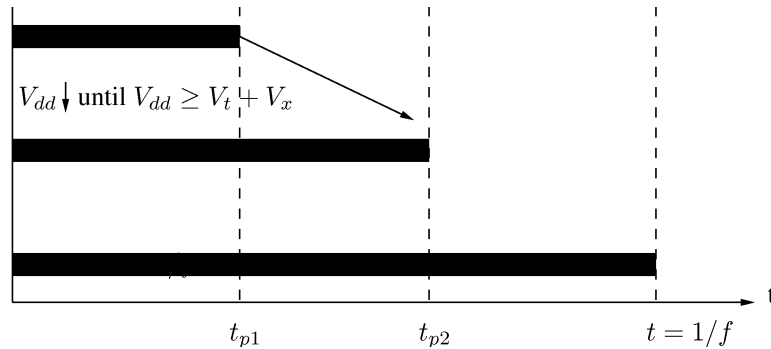
## 4.7 Design Considerations for Low-Leakage

Minimization of static power consumption is posing a new low-power design challenge for designs that cross the  $0.1\ \mu\text{m}$  barrier. Static power, mainly sub-threshold and oxide tunneling gate leakage, is dissipated by every gate regardless of its activity. This section summarizes the leakage minimization techniques which are already presented in the preceding sections [85].

The most effective approach to minimize static power is the elimination of its source. This is achievable by minimization of the total gate width, i.e., gate count. The number of gates can be reduced by keeping the design complexity

**Table 6:** Energy per device switching and gate delays for different technologies [68].

Year	2001	2004	2007	2010
Technology	130 nm	90 nm	65 nm	45 nm
Energy [fJ/device]	0.347	0.099	0.032	0.015
$t_p$ [ps]	1.6	0.99	0.68	0.30



**Figure 31:** Propagation delays in the critical path.

at the lowest possible level, and can be achieved by:

- Numerical strength reduction.
- Wordlength optimization.
- Register minimization.
- Hardware reuse (Time-Multiplexing)

Designs that operate at low frequencies may be implemented in pure LL-technology. It is beneficial to partition the design in clusters that can be powered down when not needed. Additional leakage reduction is achieved by introducing sleep transistors in the power rails.

Static and dynamic power can be reduced by lowering the supply voltage. The introduction of pipeline registers reduces the critical path, but increases the total gate width. However, a shorter critical path permits a higher maximum clock frequency. The time slack between  $f_{max}$  and the adequate clock frequency utilizes a lower supply voltage, which significantly reduces power consumption.

Figure 31 illustrates how a time slack can be utilized to reduce power consumption. The graph in Figure 31 (a) terminates at  $t_{p1}$  and represents the propagation delay of the critical path in a design. The propagation delay of the same design is increased by lowering the supply voltage. The “new” propagation delay is represented by the graph in Figure 31 (b).

The time that is available during one clock cycle is represented by Figure 31 (c). It is shown in Figure 31 (b) that  $V_{dd}$  is decreased to  $V_t + V_x$ , where  $V_x$  is a small positive voltage that prevents the circuit from malfunctioning. The original time slack  $t - t_{p1}$  has changed to  $t - t_{p2}$ . Pipelining of the critical path in Figure 31 (a) is unnecessary, since supply voltage scaling is limited by the threshold level in this example.

---

Parallel structures also permit a lower operating voltage, however, a parallel structure roughly doubles the total gate count, and, thus, increases leakage. Pipelining is therefore the favorite choice for leakage oriented low-power optimization. However, if the time slack in the critical path is very large, as in Figure 31 (a), no additional pipeline stages are necessary. The time slack can be utilized and computation performance is not affected as long as the supply voltage does not result in a propagation delay that changes the maximum clock frequency, i.e.,  $t_{p2} > t$  in Figure 31 (b).





## Chapter 5

# Artificial Neural Networks

Various types of adaptive filters represent a useful choice to achieve high performance in signal processing. However, in the area of nonlinear signal processing a better approach is required to model nonlinearities in a signal. This may be achieved by replacing the linear amplification function in a conventional adaptive filter by a nonlinear amplification function, called an *artificial neural network (ANN)*. This section illustrates the properties of an artificial neural network and presents the operation modes, i.e., feedforward and back-propagation.

### 5.1 ANN Properties

Artificial neural networks became a very popular scientific topic in the 90s with many promising achievements. The benefits of an ANN are nonlinearity, input-output mapping, adaptivity, evidential response, fault tolerance, and VLSI implementability [86]:

**Nonlinearity** An ANN consists of nonlinear *neurons*, connected in a way such that the nonlinearity is distributed throughout the entire network, as shown in Figure 32. Nonlinearity is a highly important property when modeling inherently nonlinear signals such as the ECG, EGM, or speech.

**Input-Output Mapping** The ANN is trained by subjecting classified samples to the *synaptic weights*. Each input sample results in a network response that is compared to a *desired response*. The error is used to update the weights such that it is minimized. Thus, the learning of the network is based on such an input-output mapping.

**Adaptivity** Changes in the input signal can be easily traced by a continuous update of the synaptic weights. An ANN optimized for a certain environment can be *retrained* in real-time to a modified environment, e.g., when noise occurs in a signal.

**Fault Tolerance** An ANN is fault tolerant. If one or several neurons or synaptic weights are damaged the network adapts the functioning synaptic weights to minimize the error. However, serious performance degradation is measurable if many neurons malfunction.

**VLSI Implementability** The parallel structure of an ANN is suitable for a VLSI implementation. Moreover, such parallelization results in fast computation time for complex analysis.

## 5.2 Feedforward Pass

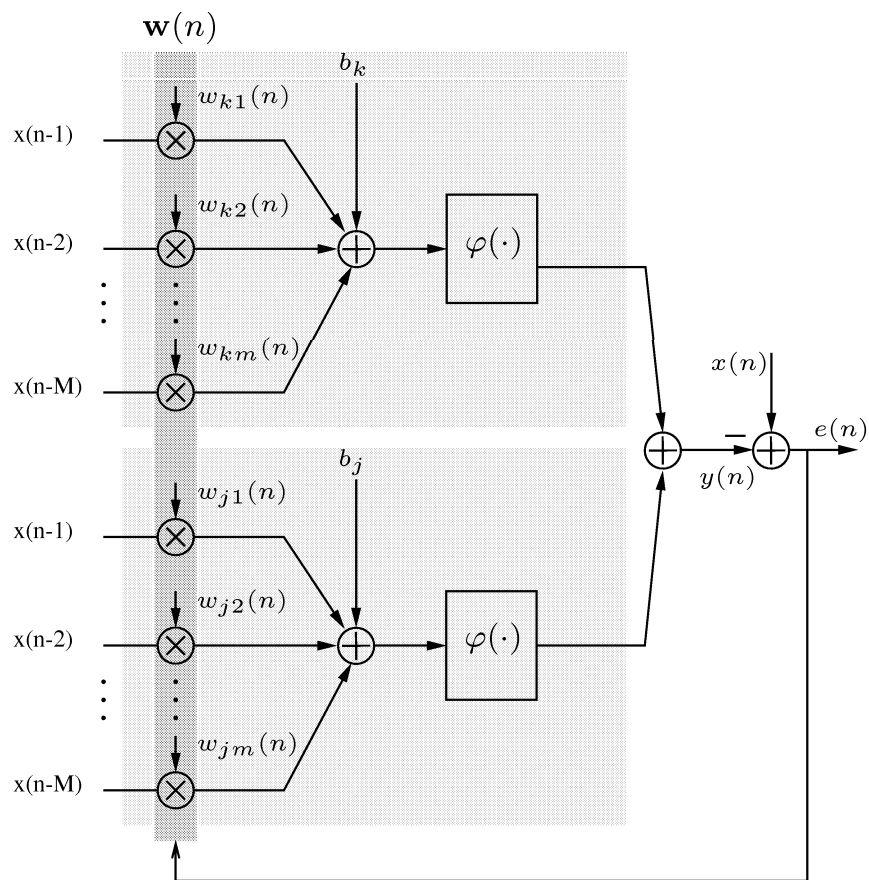
A time-lagged feed-forward neural network uses  $M$  past samples of the input vector  $x(n)$  to compute the prediction of the current sample in the feed-forward pass. The input vector  $x(n-1) \dots x(n-M)$  is fed to the neurons in the input layer. Such neurons multiply the time-varying weights with the inputs and accumulate the products. The activation function of the neuron amplifies the sum nonlinearly and the signal is ready to be transferred to the next layer. A time-lagged neural network with two input neurons is depicted in Figure 32; in this ANN the next layer is the output layer where all outputs of the preceding layer are added. The error of the prediction is computed as

$$e(n) = x(n) - y(n), \quad (10)$$

where the sample  $y(n)$  is the prediction computed by the ANN. The error  $e(n)$  is required in the back-propagation pass to minimize the prediction error by adapting the synaptic weights.

## 5.3 Back-Propagation Pass

The error  $e(n)$  in (10) is required in the back-propagation pass in order to minimize the prediction error by adapting the synaptic weights. All synaptic weights, accommodated in vector  $\mathbf{w}(n)$  are adapted using a least-mean square error criterion, and is referred to as training of the network. The next feed-forward pass makes use of the updated weights, as illustrated above, and the process continues. Thereby, the ANN keeps track of changes in signal morphology.



**Figure 32:** A typical ANN that consists of 2 neurons in the input layer, where  $x(n), x(n-1), \dots, x(n-M)$  is the input signal,  $\mathbf{w}(n)$  a vector that contains the time-varying synaptic weights,  $b_k$  and  $b_j$  the bias terms, and  $\varphi(\cdot)$  the nonlinear activation function.



## Chapter 6

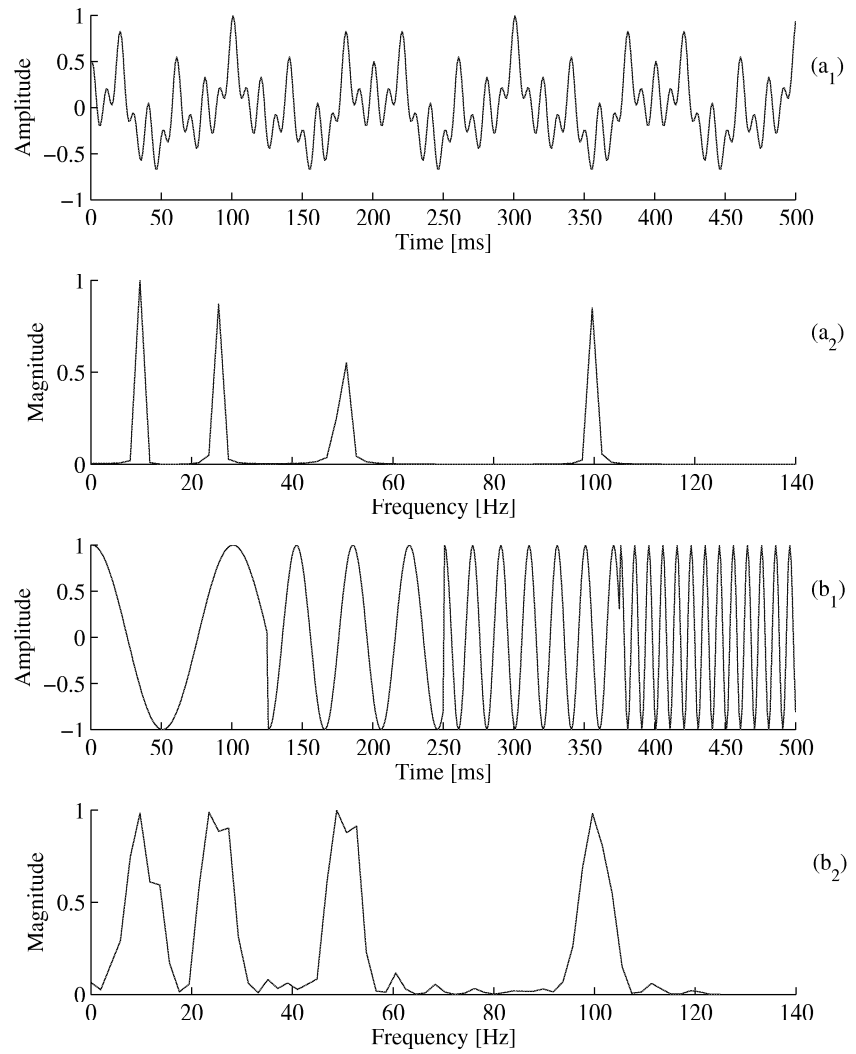
# Wavelet Decomposition

This section provides background information on wavelet decomposition. A short introduction to the Fourier transform (FT) is given to illustrate the limitations of the FT. These limitations can be overcome by making use of the wavelet transform (WT).

### 6.1 The Fourier Transform

The Fourier transform (FT) provides information on frequency content; however, the information where in time these frequency components occur is lost. A FT of a stationary and nonstationary signal is presented in Figure 33 ( $a_1$ ) and ( $b_1$ ), respectively. The signal in Figure 33 ( $a_1$ ) contains the frequencies 10, 25, 50, and 100 Hz at any time instance, whereas the signal in Figure 33 ( $b_1$ ) consists of four different frequency component at four different time intervals. The interval 0 to 125 ms contains a 10 Hz sinusoid, the interval 125 to 250 ms a 25 Hz sinusoid, the interval 250 to 375 a 50 Hz sinusoid, and the interval 375 to 500 ms a 100 Hz sinusoid. Moreover, the magnitude of the FT of both signals, displayed in Figure 33 ( $a_2$ ) and ( $b_2$ ), provides similar information on the frequency spectra.

The FT needs to be modified to obtain time localization of a frequency spectrum. Time can be localized by breaking down the total range of an FT into smaller blocks that cover a short time interval. The signal in each time interval may be assumed to be stationary; an operation which is referred to as the short time Fourier transform (STFT). Windowing a signal with the STFT thus provides the time localization. However, the use of a window of finite length reduces the frequency resolution as only a portion of the signal is covered. The resolution of the STFT is constant for time and frequencies, as presented in Figure 34 (a). The narrower the time interval the better the time resolution and the poorer the frequency resolution.



**Figure 33:** The Fourier transform of two signals, where  $(a_1)$  is a stationary and  $(b_1)$  a nonstationary signal.  $(a_2)$  and  $(b_2)$  represent the FT of the signals.

## 6.2 The Wavelet Transform

Wavelet analysis resembles Fourier analysis in the sense that the analyzed signal is decomposed into its constituent parts. The FT decomposes a signal into sine waves, whereas the WT decomposes a signal into wavelets. The sine wave is a smooth function of infinite length and distinct compared to a wavelet, which may be irregular in shape and finite. Such irregularity enhances the analysis of non-stationary signals as a discontinuity can be better captured by sharp changes.

Unlike the FT, the WT provides information on both time and frequency in the signal. Time localization is obtained by shifting and convolving a wavelet function with the signal. Frequency information is determined by applying scaled versions of the wavelet function. The application of WT solves the difficulty to chose a window function, as in the STFT, and provides flexible time and frequency resolution, see Figure 34 (b).

### 6.2.1 The Mother Wavelet

The term “wavelet” signifies an oscillatory window function of finite length and zero average as

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0. \quad (11)$$

The term “mother” implies that the functions used in the transform process are derived from a principal function, referred to as mother wavelet. Commonly used mother wavelets are named Haar, Mexican hat, and Daubechies [87,88].

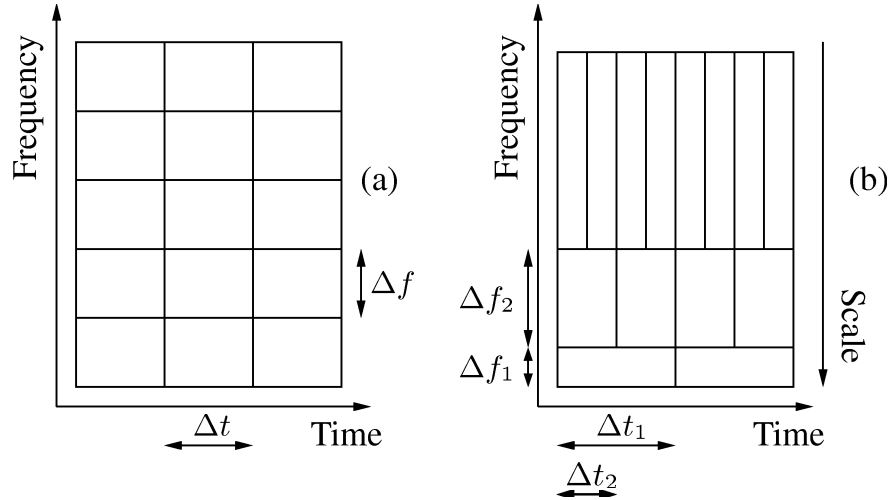
### 6.2.2 Scale

Scaling is a mathematical operation that either compresses or dilates the mother wavelet. Low frequencies are represented by coarser scales, and high frequencies by finer scales. The scale factor determines the resolution in the time-frequency plane, as presented in Figure 34 (b). The width and height of the boxes change, however, the area of a box remains constant. Each box represents an equal share of the time-frequency plane, but the time/frequency ratio changes. At low frequencies (coarse scales) the height of the box is shorter which corresponds to better frequency resolution but poorer time resolution. At higher frequencies (fine scales) the width of the boxes decrease as the height increases. This results in a better time and a poorer frequency resolution.

The area of a box in Figure 34 is equal for a distinct mother wavelet, however, different mother wavelets can result in a different area.

A scaled version of the mother wavelet is computed by

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi \left( \frac{t-u}{s} \right), \quad (12)$$



**Figure 34:** Time and frequency resolution: (a) STFT (b) WT.

where  $s$  is the scale and  $u$  the translation, and the factor  $1/\sqrt{s}$  assures equal energy for all scaled functions [1].

### 6.2.3 The Continuous and Discrete Wavelet Transform

The continuous wavelet transform (CWT) can be viewed as the convolution of the mother wavelet and the signal over the length of the data. The scaling of the mother wavelet can be defined from a minimum to a user-defined maximum, and, thus, providing very high resolution. The trade-off for the high resolution is computation time.

The continuous wavelet transform  $w(u, t)$  is defined by the correlation of a function  $f$  at a scale  $s$  and translation  $u$ ,

$$w(s, u) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right) dt. \quad (13)$$

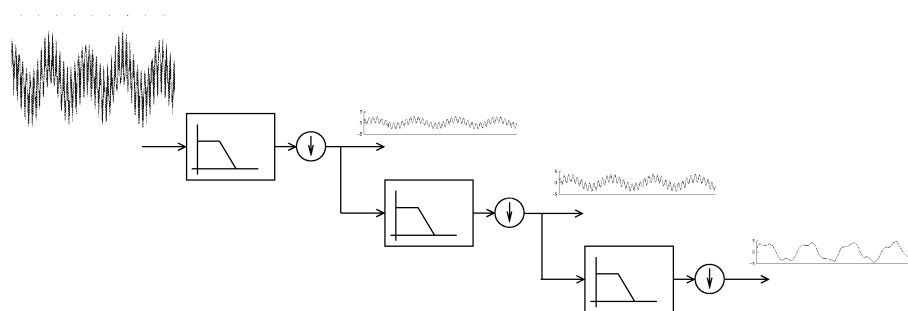
The CWT is highly redundant, and, therefore, the scaling and translation parameter  $s$  and  $u$  are discretized according to a suitable scaling grid [1]. The most popular approach to discretization is *dyadic scaling*.

$$s = 2^{-j}, \quad u = k2^{-j}, \quad (14)$$

where  $j$  and  $k$  are both integers. Thus, the discretized wavelet function is defined by

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k). \quad (15)$$





**Figure 35:** Discrete wavelet decomposition tree.

The *discrete wavelet transform* (DWT) is obtained by inserting (15) in (13)

$$w_{j,k} = \int_{-\infty}^{+\infty} f(t)\psi_{j,k}(t) dt. \quad (16)$$

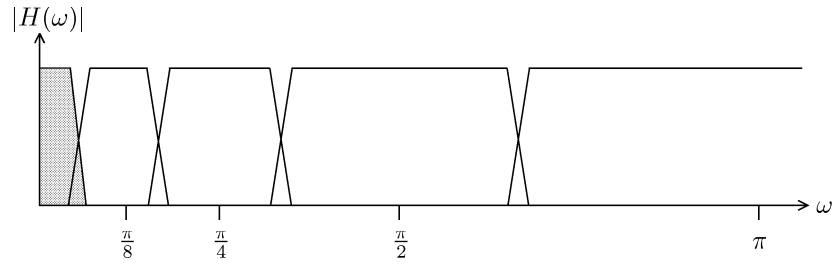
With the expression in (16) and dyadic sampling it is possible to exactly reconstruct the original function  $f(t)$ , however, the computational effort of the DWT is much lower than for CWT. The function  $f(t)$  is reconstructed by the *inverse DWT*,

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} w_{j,k}\psi_{j,k}(t), \quad (17)$$

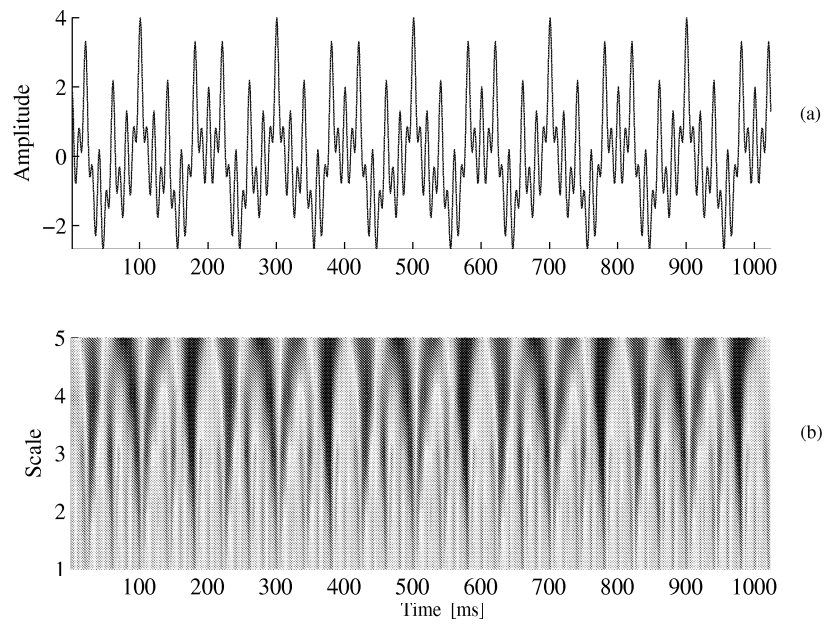
The signals previously analyzed with FT are analyzed by the DWT, as presented in Figure 37 and 38. The applied scale factors of the WT are represented by their exponent on the vertical axis. In Figure 37 it can be seen that all frequencies are present, whereas the signal in Figure 38 has non-stationary properties. The spectrogram in Figure 38 (b) indicates the localization in time for each frequency spectrum in the analyzed signal: the low frequency component, represented by a fine scale, is present the first 250 ms and at 1000 to 1250 ms, respectively. The high frequency component, represented by a coarse scale, is localized at 750 to 1000 ms.

### 6.3 DWT Realization

The discrete wavelet transform can be implemented as a tree of low and high pass filtering functions. Each branch realizes a low pass filter, as presented in Figure 35. The high frequency components are not considered for further analysis. The original signal is successively decomposed into its different scales [87].

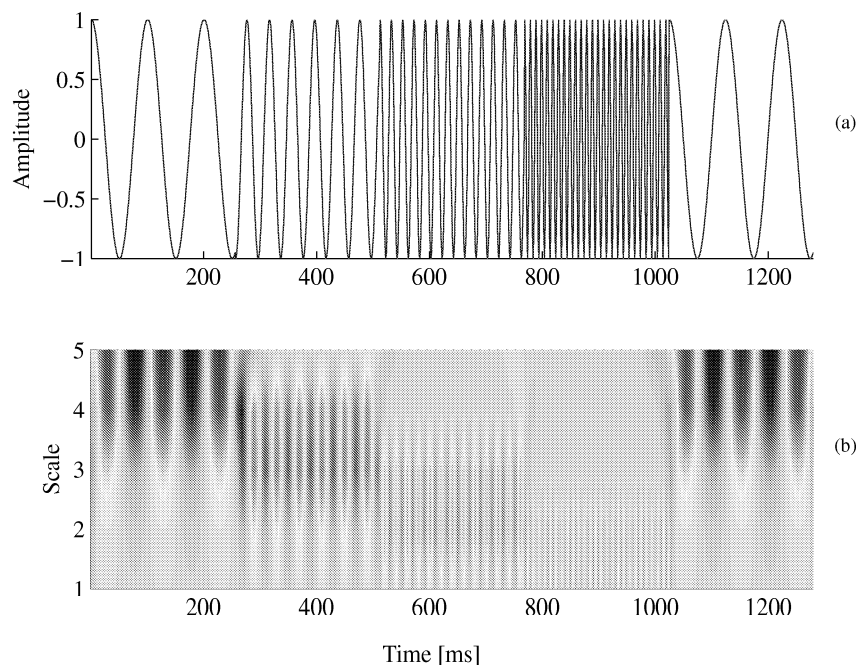


**Figure 36:** Frequency bands for the analysis tree in Figure 35.



**Figure 37:** (a) Example of a stationary signal and (b) the corresponding magnitude of the discrete wavelet transform. The number on the vertical axis represents the exponent of the scale factor.

The frequency bands of the wavelet decomposition tree in Figure 35 are presented in Figure 36. The shaded frequency band contains the lowest frequency sequence in Figure 35.



**Figure 38:** (a) Example of a nonstationary signal and (b) the corresponding magnitude of the discrete wavelet transform. The number on the vertical axis represents the exponent of the scale factor.



## References

- [1] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Amsterdam: Elsevier, 2005.
- [2] E. Berbari, *Encycl. Electric. Elec. Eng.: Electrocardiography*. New York: Wiley, 1999.
- [3] J. Webster, *Design of cardiac pacemakers*. New York, USA: IEEE Press, 1995.
- [4] R. Elmqvist, J. Landegren, S. Petersson, Å. Senning, and G. W. Ollson, "Artificial pacemaker for treatment of adams-stokes syndrome and slow heart rate," *Am Heart J*, no. 65, pp. 731–748, 1963.
- [5] F. Steiner, "Über die Elektropunktur des Herzens als Wiederbelebungsmitel in der Chlorformsyncope," *Archiv klinische Chirurgie*, no. 12, pp. 748–790, 1871.
- [6] T. Greene, "On death from chloroform; its prevention by galvanism," *Brit. Med.*, no. 1, pp. 551–553, 1872.
- [7] M. Lidwill, "Cardiac disease and anaesthesia," *Med Journal Australia*, no. 2, pp. 574–575, 1929.
- [8] A. A. Hyman, "Resuscitation of the stopped heart by intracardial therapy," *Arch. Inc. Med.*, no. 50, pp. 289–308, 1932.
- [9] W. Schockley, "The theory of pn junctions in semiconductors and pn-junction transistors," *BSTJ*, no. 28, p. 435, 1949.
- [10] P. Zoll, "Resuscitation of heart in ventricular standstill by external electric stimulation," *N. Engl. J. Med.*, no. 247, pp. 768–771, 1952.
- [11] P. Zoll, A. Linenthal, A. Norman, and M. Paul, "External stimulation of the heart in cardiac arrest," *AMA Arch. Int. Med*, no. 96, pp. 639–650, 1956.
- [12] E. Bakken, "The history of pacing," *Am J. Cardiol*, pp. 614–615, 1977.
- [13] R. Elmqvist and Å. Senning, "Implantable pacemaker for the heart," in *Medical Electronics, Proceedings of the second International Conference on Medical Electronics*, C. Smyth, Ed., 1969.
- [14] J. Kilby, "Invention of the integrated circuit," *IEEE Trans. Electron Devices*, p. 648, 1976.

- 
- [15] W. Chardack, A. Gage, and W. Greatbatch, "A transistorized, contained, implantable pacemaker for long-term correction of complete heart block," *Surgery*, no. 48, pp. 843–846, 1960.
- [16] W. Chardack, A. Gage, and A. Frederico, "Clinical experience with an implanted pacemaker," *Ann. NY Acad. Scr.*, no. 111(3), pp. 1075–1092, 1964.
- [17] L. Geddes, "Historical highlights in cardiac pacing," *IEEE Eng. Med. Biol. Mag.*, pp. 12–18, June 1990.
- [18] H. Elmqvist, "Pacemakerteknik ett kortfattat kompendium," [www.labtek.ki.se/education/kurser\\_kth\\_ki/implantat/litt/pacemaker\\_kompendium\\_elmqvist\\_vt20005.pdf](http://www.labtek.ki.se/education/kurser_kth_ki/implantat/litt/pacemaker_kompendium_elmqvist_vt20005.pdf), 2005.
- [19] "St-Jude Medical," <http://www.sjm.com>.
- [20] C. Lillehei, V. Gott, P. Hodges, D. Long, and E. Bakken, "Transistor pacemaker for treatment of complete atrioventricular dissociation," *JAMA*, no. 172, pp. 2006–2010, 1960.
- [21] A. Kantrowitz, R. Cohen, and H. Raillard, et al., "The treatment of complete heart block with an implanted controllable pacemaker," *Surg. Gynecol. Obstet.*, no. 115, pp. 415–420, 1962.
- [22] V. Parsonnet, T. E. Cuddy, and D. Escher, et al., "A pacemaker capable of external non-invasive programming," *Trans. Am. Soc. Artif. Intern. Organs*, no. 19, pp. 224–228, 1973.
- [23] A. Johansson, "Wireless communication with medical implants: Antennas and propagation," Ph.D. dissertation, Lund University, 2004.
- [24] V. Parsonnet, I. Zucker, and S. Myers, "Clinical use of an implanted standby pacemaker," *JAMA*, no. 198, pp. 769–784, 1966.
- [25] P. Samet, W. Bernstein, and D. Nathan, et al., "Atrial contribution to cardiac output in complete heart block," *Am J Cardiol*, no. 16, pp. 1–10, 1965.
- [26] A. Benchimol, A. Duenas, and M. Liggett, et al., "Contribution of atrial systole to the cardiac function at a fixed and at a variable ventricular rate," *Am J Cardiol*, no. 16, pp. 11–21, 1965.
- [27] I. Karlof, "Haemodynamic effect of atrial triggered versus fixed rate pacing at rest and during exercise in complete heart block," *Acta Med Scand*, no. 197, pp. 195–206, 1975.

- [28] “fairview.org,” [http://www.fairview.org/healthlibrary/content/ca\\_pace\\_car.htm](http://www.fairview.org/healthlibrary/content/ca_pace_car.htm).
- [29] V. Mahaux, A. Waleffe, and H. Kulbertus, “Clinical experience with a new activity sensing rate modulated pacemaker using autoprogrammability,” *Pace*, no. 13, pp. 819–820, 1990.
- [30] K. Jeffrey, *Machines in our Hearts: The cardiac pacemaker, the implantable defibrillator, and American health care*. Baltimore and London: The Johns Hopkins University Press, 2001.
- [31] V. Parsonnet, S. Furman, and N. Smyth, “Implantable cardiac pacemakers. status report and resource guideline: Pacemaker study group,” *Circulation*, no. 50A, pp. 21–35, 1974.
- [32] A. Bernstein, J. Daubert, and R. Fletcher, et al., “The revised NASPE/BPEG generic code of antibradycardia, adaptive-rate, and multisite pacing. North American Society of Pacing and Electrophysiology/British Pacing and Electrophysiology Group,” *Pace*, no. 25, pp. 260–264, 2002.
- [33] H. W. Moses, B. D. Miller, K. P. Moulton, and J. A. Schneider, *A Practical Guide to Cardiac Pacing*. Philadelphia: Lippincott Williams, 2000.
- [34] G. E. Moore, “Cramming More Components onto Integrated Circuits,” *Electronics*, vol. 38, no. 8, 1965. [Online]. Available: <ftp://download.intel.com/research/silicon/moorespaper.pdf>
- [35] I. Toumi, “The Live’s and Death of Moore’s Law,” *First Monday*, 2002. [Online]. Available: [http://firstmonday.org/issues/issue7\\_11/tuomi/](http://firstmonday.org/issues/issue7_11/tuomi/)
- [36] S. Borkar, “Obeying Moore’s Law Beyond 0.18 Micron,” in *Proceedings of the 13th Annual IEEE International ASIC/SOC Conference, Arlington, VA*, 2000, pp. 26–31.
- [37] G. E. Moore, “No Exponential is Forever: But “Forever” Can Be Delayed!” in *Proceedings of the IEEE International Solid-State Circuits Conference, ISSCC’03*, 2003, pp. 20–23.
- [38] J. M. Rabaey, A. Chandrakasan, and B. Nikolić, *Digital Integrated Cicuits*. New Jersey: Prentice Hall, 2003.
- [39] A. P. Chandrakasan, S. Sheng, and R. W. Brodersen, “Low-Power CMOS Digital Design,” *IEEE J. Solid-State Circuits*, vol. 27, pp. 473–484, 1992.
- [40] A. P. Chandrakasan and R. W. Brodersen, “Minimizing Power Consumption in Digital CMOS Circuits,” in *Proc. of the IEEE*, vol. 83, no. 4, 1995, pp. 498–523.

- 
- [41] —, *Low Power Digital CMOS Design*. Boston: Kluwer Academic Publisher, 1995.
- [42] M. Ismail and N. Tan, “Modeling Techniques for Energy-Efficient System-on-a-Chip Signaling,” *IEEE Circuits and Devices Magazine*, vol. 19, no. 1, pp. 1627–1633, 2003.
- [43] V. Tiwari, D. Singh, S. Rajgopal, G. Mehta, R. Patel, and F. Baez, “Reducing Power in High-Performance Microprocessors,” in *Proceedings of the 35th Design Automation Conference, DAC’98*. San Fransisco, CA, USA: ACM Press, June 15-19 1998, pp. 732–737.
- [44] A. J. Bhavnagarwala, B. Austin, A. Kapoor, and J. D. Meindl, “CMOS System-on-a-Chip Voltage Scaling beyond 50nm,” in *Proceedings of the 10th Great Lakes Symposium on VLSI, Chicago, Illinois, USA, 2000*, pp. 7–12.
- [45] T. Olsson, “Distributed clocking and clock generation in digital CMOS SoCASICs,” Ph.D. dissertation, Lund University, 2004.
- [46] J. Scott, L. H. Lee, J. Arends, and B. Moyer, “Designing the Low-Power MCore Architecture,” in *Workshop on Power Driven Microarchitecture*, June 1998, pp. 145–150. [Online]. Available: <http://www.cs.ccu.edu.tw/~kch91/MCORE/designwp.pdf>
- [47] P. E. Gronowski, W. J. Bowhill, R. P. Preston, M. K. Gowan, and R. L. Allmon, “High-Performance Microprocessor Design,” *IEEE J. Solid-State Circuits*, vol. 33, no. 5, pp. 676–686, May 1998.
- [48] M. K. Gowan, L. L. Biro, and D. B. Jackson, “Power Considerations in the Design of the Alpha 21264 Microprocessor,” in *Proceedings of the 35th Design Automation Conference, DAC’98*. San Fransisco, CA, USA: ACM Press, June 15-19 1998.
- [49] V. Tiwari, R. Donnelly, M. Sharad, and G. Ricarde, “Dynamic Power Management for Microprocessors: A Case Study,” in *Proceedings of the 10th International Conference on VLSI Design, 1997*, pp. 185–192.
- [50] K. Roy, “Leakage and leakage reduction techniques for nano-scale CMOS: Device & circuit perspective,” [www.research.ibm.com/aceed/2005/proceedings/roy.ppt](http://www.research.ibm.com/aceed/2005/proceedings/roy.ppt).
- [51] G. Sery, S. Borkar, and V. De, “Life is CMOS: why chase the life after?” in *Proceedings of the 39th Design Automation Conference, DAC’02*. ACM Press, June 10-14 2002, pp. 78–83.



- [52] D. Lammers, "SIA's road map affirms 3-year cycles for chips," <http://www.eetimes.com/news/latest/showArticle.jhtml?articleID=18310552>.
- [53] J. N. Rodrigues, T. Olsson, L. Sörnmo, and V. Öwall, "A dual-mode wavelet based R-Wave detector using single- $V_t$  for leakage reduction," in *Proc. 2005 IEEE Intl. Symp. on Circuits and Systems*, 2005.
- [54] A. Agarwal, H. Li, and K. Roy, "A single- $V_t$  low-leakage gated-ground cache for deep submicron," *IEEE J. Solid-State Circuits*, pp. 319–328, 2003.
- [55] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, "Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits," *Proc. of the IEEE*, vol. 91, no. 2, pp. 305–327, 2003.
- [56] S. Mukhopadhyay and K. Roy, "Modeling and Estimation of Total Leakage Current in Nano-scaled-CMOS Devices Considering the Effect of Parameter Variation," in *Proceedings of the 2003 International Symposium on Low Power Electronics and Design, ISLPED'03*, 2003, pp. 172–175.
- [57] S. H. Lo, D. Buchanan, Y. Taur, and W. Wang, "Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide MOSFETs," *IEEE Electron Device Letter*, vol. 18, pp. 206–209, May 1997.
- [58] B. Yu, H. Wang, C. Ricobene, Q. Xiang, and M. R. Lin, "Limits of gate oxide scaling in nano-transistors," *VLSI Tech. Dig.*, pp. 39–40, 2000.
- [59] H. Iwai, "Ultra thin gate oxides performance and reliability," *IEDM Tech. Dig.*, pp. 163–166, 1998.
- [60] —, "Downsizing of silicon MOSFETs beyond 0.1  $\mu\text{m}$ ," *Microelectronics Journal*, vol. 29, pp. 671–678, April 1998.
- [61] S. A. Campbell, D. C. Gilmer, X. C. Wang, H. S. Kim, and J. Yan, "MOS transistors fabricated with high permittivity  $\text{TiO}_2$  dielectrics," *IEEE Trans. Electron Devices*, vol. 44, pp. 104–109, May 1997.
- [62] N. Mohapatra, M. Desai, S. Narendra, and V. Rao, "The effect of high-k gate dielectrics on deep submicrometer CMOS device and circuit performance," *IEEE Trans. Electron Devices*, vol. 49, no. 5, pp. 826–831, May 2002.
- [63] R. Troutman, "VLSI limitations from drain-induced barrier lowering," *IEEE Trans. Electron Devices*, pp. 461–468, 1979.

- 
- [64] J. Pimbley and J. D. Meindl, "MOSFET scaling limits determined by subthreshold conduction," *IEEE Trans. Electron Devices*, pp. 1711–1721, 1989.
- [65] A. Keshavarzi, K. Roy, and C. Hawkins, "Intrinsic leakage in deep sub-micron CMOS ICs-measurement-based test solutions," *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, vol. 8, no. 6, pp. 717–723, December 2000.
- [66] R. Pierret, *Semiconductor Device Fundamentals Reading*. Boston: Addison-Wesley, 1996.
- [67] S. Borkar, "Low Power Design Challenges for the Decade," in *Proceedings of the 38th Design Automation Conference, DAC'01*. ACM Press, June 2001, pp. 78–83.
- [68] "ITRS roadmap 2003," <http://public.itrs.net/Files/2003ITRS/Home2003.htm>.
- [69] R. Gonzalez, B. Gordon, and M. Horowitz, "Supply and Threshold Voltage Scaling for Low Power CMOS," *IEEE J. Solid-State Circuits*, vol. 32, no. 8, pp. 1210–1216, Aug. 1997.
- [70] K. Parhi, *VLSI Digital Signal Processing*. New York: Wiley, 1999.
- [71] B. Parhami, *Computer Arithmetic*. New York: Oxford University Press, 2000.
- [72] "United Microelectronics Corp." <http://www.umc.com/english/process/index.asp>.
- [73] P. Bai, C. Auth, and S. Balakrishnan, et al., "A 65 nm logic technology featuring 35 nm gate lengths, enhanced channel strain, 8 Cu interconnect layers, low-k ILD and 0.57  $\mu^2$  SRAM cell," *Electron Devices Meeting, 2004. IEDM Technical Digest. IEEE International*, pp. 657–660, 2004.
- [74] M. Takahashi and el. al, "A 60-mW MPEG4 video coding using clustered voltage scaling with variable voltage scaling scheme," *IEEE J. Solid-State Circuits*, vol. 3, no. 4, pp. 1772–1780, 1998.
- [75] R. K. Krishnarnurthy, A. Alvandpour, V. De, and S. Borkar, "High-performance and Low-power Challenges for Sub-70nm Microprocessor Circuits," in *Proceedings of the IEEE 2002 Custom Integrated Circuits Conference*, May 2002, pp. 125–128.
- [76] T. Fuse, A. Kameyama, M. Ohta, and K. Ohuchi, "A 0.5 V power-supply schem for low power LSI's using multi- $V_t$  SOI CMOS technology," *Dig. Tech. Papers Symp. VLSI Circuits*, pp. 219–220, 2001.

- [77] L. Carley and A. Agarwal, "A completely on-chip voltage regulation technique for low power digital circuits," in *Proc. Int. Symp. Low Power Electronics and Design*, 1999, pp. 109–111.
- [78] Y. Kanno, H. Mizuno, K. Tanaka, and T. Watanabe, "Level converters with high immunity to power-supply bouncing for high-speed Sub-1-V LSI's," *Dig. Tech. Papers Symp. VLSI Circuits*, pp. 202–203, 2000.
- [79] L. Wanhammar, *DSP Integrated Circuits*. London: Academic Press, 1999.
- [80] S. Tyagi, et al., "A 130nm generation logic technology featuring 70nm transistors, dual  $v_t$  transistors and 6 layers of Cu interconnect," *Dig. Tech. Papers Int. Electron Devices Meeting*, pp. 567–570, 2000.
- [81] B. Chatterjee, M. Sachdev, S. Hsu, R. Krishnamurthy, and S. Borkar, "Effectiveness and Scaling Trends of Leakage Control Techniques for Sub-130 nm CMOS Technologies," in *Proceedings of the 2003 International Symposium on Low Power Electronics and Design, ISLPED'03*, 2003, pp. 122–127.
- [82] S. Mukhopadhyay, C. Neau, R. Cakici, A. Agarwal, C. Kim, and K. Roy, "Gate leakage reduction for scaled devices using transistor stacking," *IEEE Trans. on VLSI Systems*, vol. 11, pp. 716–730, 2003.
- [83] J. Tschanz, J. Kao, S. G. Narendra, R. Nair, D. A. Antoniadis, A. Chandrakasan, S. A. V. De, D. C. Gilmer, X. C. Wang, H. S. Kim, and J. Yan, "Adaptive body bias for reducing impacts of die-to-die and within-die parameter variations on microprocessor frequency and leakage," *IEEE J. Solid-State Circuits*, vol. 37, pp. 1396–1402, November 2002.
- [84] D. K. Schroder, "Some critical IC issues," <http://www.eas.asu.edu/schroder/>.
- [85] N. S. Kim and T. Austin, et al., "Leakage current: Moore's law meets static power," *Computer*, vol. 36, pp. 68–75, December 2002.
- [86] S. Haykin, *Neural Networks*, 2nd ed. New York: Prentice Hall, 1999.
- [87] S. Mallat, *A wavelet tour of signal processing*. San Diego, CA, USA: Academic Press, 1998.
- [88] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Trans. Inform. Theory*, vol. 36, no. 5, pp. 961–1005, 1990.



# Part I



# Part I

## Implementation of an Artificial Neural Network Based Event Detector

### Abstract

A matched filter composed of a Time Lagged Feedforward artificial neural Network (TLFN) and a pulse-shaping filter is used as an event detector for cardiac pacemakers. The TLFN reduces the influence of lower frequencies in the invasive electrogram (EGM) signals and conditions the EGM to optimize the performance of the dynamically updated matched filter. An algorithm that determines the initial template for matched filtering is proposed. Detector performance is studied by means of databases containing electrograms as well as different types of noise and interferences, which are added to the signals. Average detection performance in terms of detected events and false alarms for 25 dB SNR is  $P_D = 0.98$  and  $P_{FA} = 0.05$ .

---

Based on:

J. Neves Rodrigues, V. Öwall, and L. Sörnmo, "QRS detection for pacemakers in a noisy environment using a time lagged artificial neural network," *Proceedings of the 2001 IEEE International Symposium on Circuits and Systems, ISCAS 2001, Sydney, Australia, May 6-9 2001*,

J. Neves Rodrigues, V. Öwall, and L. Sörnmo, "R-wave detection for pacemakers using a matched filter based on an artificial neural network," *Proceedings of the 2002 IEEE International Conference on Neural Information Processing, ICONIP 2002, Singapore, November 18-22 2002*.





## 1 Introduction

In 1958 two breakthrough inventions in different scientific areas took place: the first implantable pacemaker was developed by Rune Elmqvist in Stockholm, Sweden, and the first integrated circuit, IC, was developed by Jack Kilby, USA [1, 2]. This was the beginning of two successful histories. In those days only visionaries would predict how much the pacemaker could benefit from the progress in IC manufacturing several decades later. Since 1958 several features, such as rate responsive sensing, programmability, miniaturizing, have been developed to improve the acceptance of pacemakers among patients suffering from heart diseases [3]. The size of the implantable pacemaker shrank from being puck-sized to 33 x 33 x 6 mm with a weight of 12.8 g (*Microny, St-Jude Medical*). The pacemaker from 1958 had rechargeable batteries, whereas the lifetime of the first pacemaker with mercury cells was 2 to 3 years. Nowadays, the longevity of the pacemaker is up to 20 years, e.g., *Regency SC+, St-Jude Medical* [4]. Despite these improvements, the event (R-wave) detection algorithm has remained largely unchanged. The event detection of today is still based on a fixed bandpass filter, followed by a programmable amplitude threshold.

During the last three decades a variety of R-wave detection algorithms for electrocardiograms, ECGs, have been proposed [5–7]. However, most of are very complex implemented in digital hardware, and, even more problematic, they do not operate in real time.

The number of electronic devices and household appliances in everyday live has an ongoing exponential growth. These devices contaminate their environment with electronic, magnetic or electromagnetic radiation. Pacemaker patients exposed to this environment may suffer due to malfunction of the pacemaker. In order to investigate how devices, such as cellular phones, electronic article surveillance systems, EAS, etc., interfere with the pacemaker, various research projects have been carried out in recent years. It turned out that several EAS systems may interfere with the pacemaker such that a degradation in performance is very likely [8–11], whereas cellular phones do not influence the pacemaker performance [12]. The need for better event detectors is therefore ever increasing and was already demanded two decades ago [13].

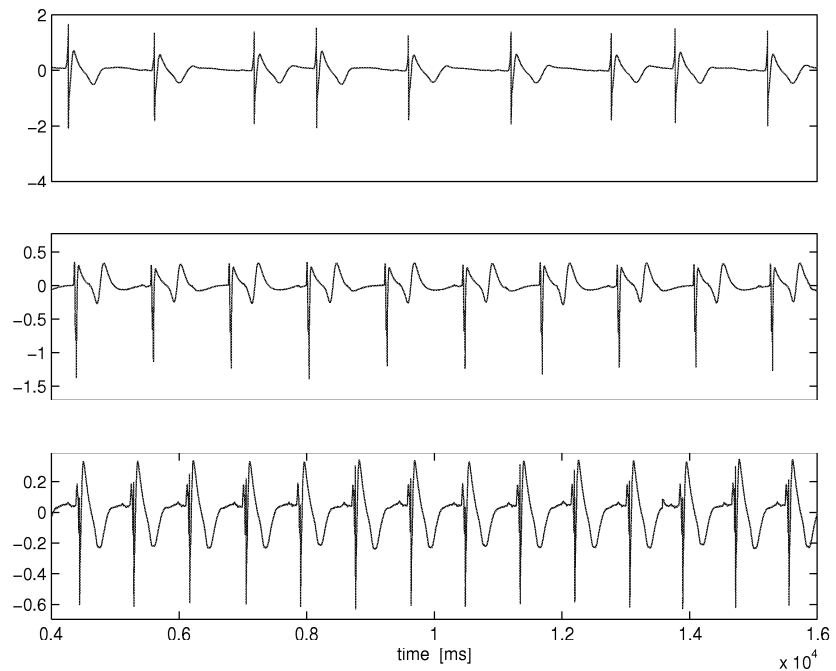
This chapter presents the databases required to demonstrate detector properties and used to carry out detector performance analysis. The detector, basically composed of a matched filter, an artificial neural network, and a decision rule is presented. Detection performance in heavily distorted environment is evaluated by subjecting recordings from an EGM and interference database to the detector. The properties and suitability of the R-wave detector for pacemaker application is discussed.

## 2 Databases

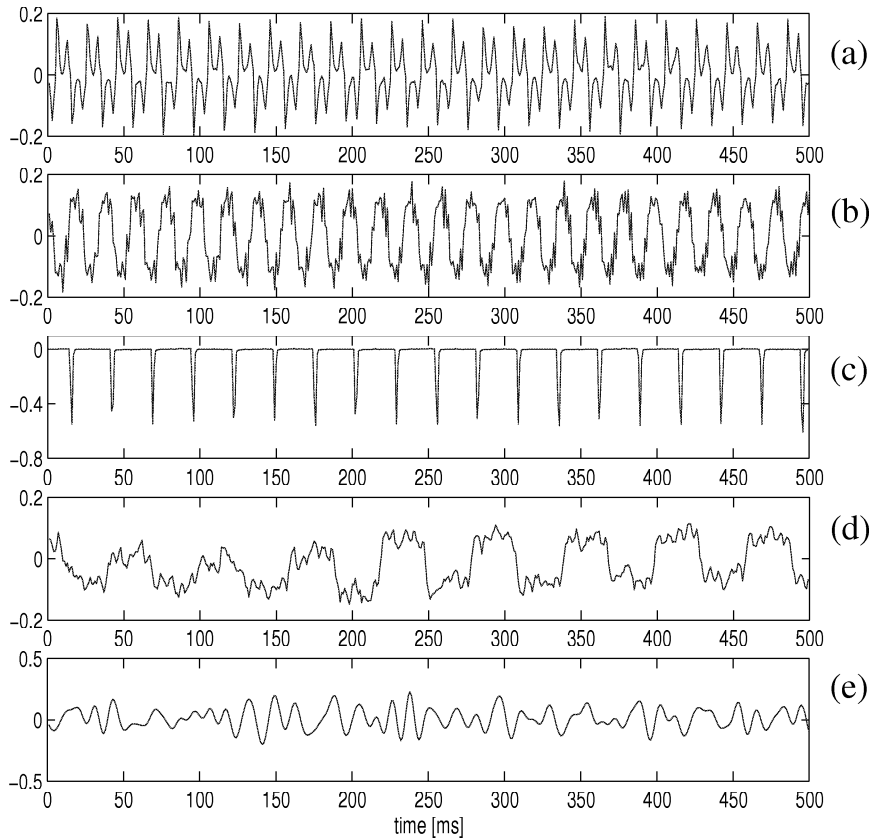
Detector properties and performance are evaluated on account of databases with EGMs and interference signals. The EGM database is a collection of individual recordings obtained during pacemaker surgery. The interference database contains different types of signals originating from household appliances or electronic devices such as a hand drill or an electronic article surveillance system (EAS).

### 2.1 EGM Database

The database contains EGMs from 50 patients, recorded from ventricular pacemaker electrodes, and is used to evaluate the performance of the detector. The EGMs were recorded either during initial implantation or pacemaker replacement, throughout hospitals in Germany (coordinated by Justus-Liebig Universität, Gießen, Germany). The recordings were obtained from patients suffering from AV block and sick sinus syndrome [14]. Most signals were recorded from a unipolar electrode, however, a few signals were recorded with a bipolar elec-

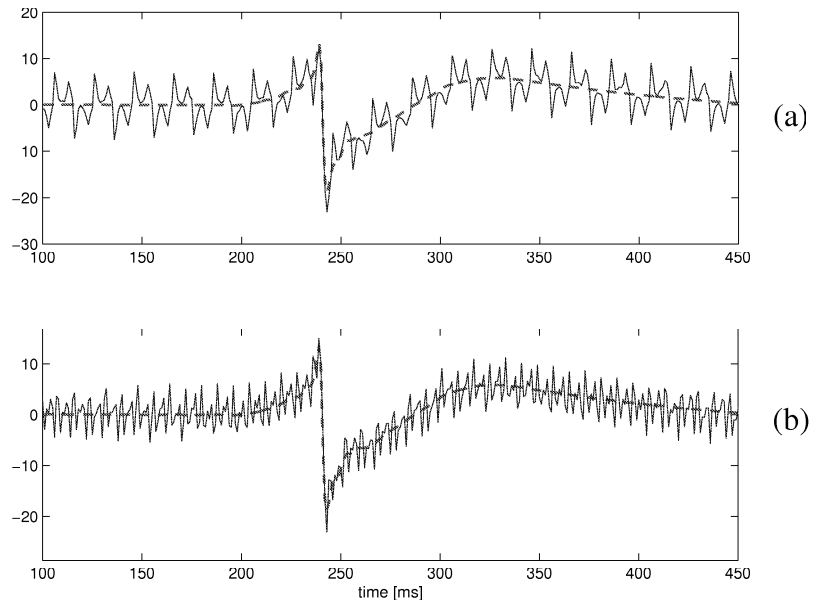


**Figure 1:** Examples of EGMs from different patients.



**Figure 2:** Exogenic interference recordings generated from (a) 500 W AC hand drill, (b) an electric hand mixer, (c) EAS 1, (d) EAS 2. Endogenic interference caused by (e) muscular activity.

trode. The sampling rate was 44.1 kHz with a resolution of 16 bits. For this particular study the signals were decimated to 1 kHz, since frequencies above 400–500 Hz were judged to be less significant to detection. In order to be compliant with the ADC in [15] a resolution of 8 bits is chosen. The recordings were annotated with respect to a time reference of each R-wave, required for performance evaluation of the detection algorithm. The annotation of an event was defined as the steepest transition phase in the cardiac cycle. Three EGM recordings from different patients are displayed in Figure 1, and illustrates the inter-patient variability in morphology.



**Figure 3:** A typical EGM interfered with (a) AC hand drill, and (b) muscular activity.

## 2.2 Interference Database

The present event detector is tested with respect to sensitivity to exogenic and endogenic interferences, originating outside and inside the body, respectively [16]. The test is done in order to simulate situations when the pacemaker patient is subjected to electronic or magnetic noise. Exogenic interference is limited to sources in everyday life, e.g., caused by electronic household appliances or electronic article surveillance (EAS) systems. Endogenic interference is represented by muscular activity [17]. Figure 2 presents examples of the different types of interference.

*Household appliances* represent a common source of interference, caused by electric and magnetic activity within the same frequency range as the R-wave. Furthermore, the magnetic field intensity is dependent on the signal transiency. In this study, recordings from an AC powered hand drill and an electric handmixer were used.

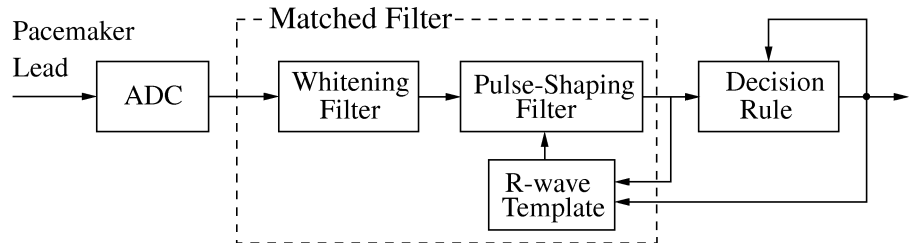
*Electronic article surveillance systems* have been identified as a common interference source [8–11]. Such systems use widely different transmission techniques which makes it difficult to generalize results to how such systems interfere with the pacemaker. In this study two systems that operate within the

---

R-wave frequency spectra have been tested. The EAS 1 system uses a 16.6 Hz triangular wave modulated with 5 or 7.5 kHz. The EAS 2 system transmits 3 ms long bursts of 58 kHz acoustomagnetic signals with a high amplitude at an interval of 27 ms. The pulse period of 30 ms (33 Hz) is considered to be the reason for possible interference with pacemakers.

*Muscular activity* is an endogenic interference source which spectrally overlaps with heart signals. In this study, signals recorded pressing the palms together have been considered. The effect of muscular noise on pacemaker performance was one of the first studies on pacemaker interference [17].

Figure 3 shows the morphology of an EGM with interferences originating from a hand drill and muscular activity, respectively.



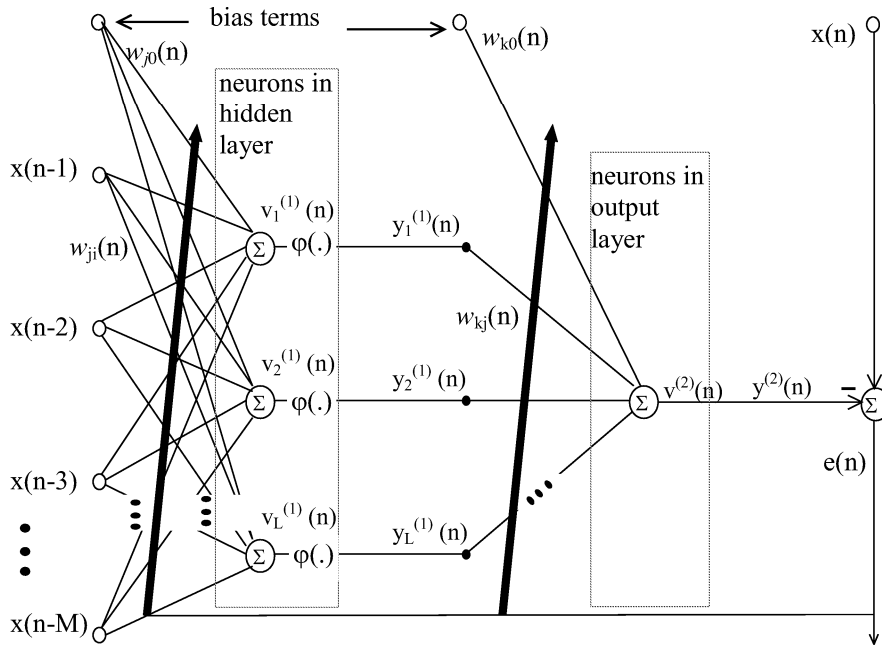
**Figure 4:** Block diagram of the detector structure.

### 3 Detector Structure

The proposed real time digital R-wave detector is basically composed of a whitening and a pulse-shaping filter, and a decision rule, as presented in Figure 4. The electrogram (EGM), obtained from the pacemaker lead, is after analog-to-digital conversion (ADC) passed to the matched filter. The matched filter consists of a whitening and a dynamically updated pulse-shaping filter. The objective of the matched filter is to maximize the SNR such that an R-wave is more easily detectable. The output of the matched filter is passed to a dynamic decision rule, which determines whether an R-wave has occurred. If an R-wave is identified the threshold of the decision rule and the template of the matched filter is updated. Such updates are performed in order to track slow changes in electrogram morphology. However, the update of the template is only carried out if the update and the template are highly correlated.

#### 3.1 The Whitening Filter

The practical application of the whitening filter is to reduce the influence of low frequency components such as the far-field P- and T-wave, as presented in Section 2. Moreover, various types of interferences that may exist in the EGM are suppressed. The application of a time lagged artificial neural network as a whitening filter for QRS detection in electrocardiograms (ECGs) has been proposed in [18]. However, the ECG is a signal recorded from the body surface with several electrodes and is significantly different compared to the EGM [14, 19]. Therefore, in this paper the proposed QRS detector in [18] is applied as an R-wave detector for EGMs.



**Figure 5:** The implemented time lagged feedforward artificial neural network.

### 3.2 The Time Lagged Feedforward Artificial Neural Network

The whitening filter in Figure 4 is an adaptive nonlinear filter, accomplished by using a fully connected time lagged feedforward neural network (TLFFN). The TLFFN can be viewed as a one-step predictor: the prediction of the current sample is computed from a number of preceding samples. The TLFFN is a supervised network consisting of a forward and a backward pass, as illustrated in Figure 5. In the forward pass, the time lagged inputs  $x(n-1) \cdots x(n-M)$  propagate through the network to produce the prediction  $y^{(2)}(n)$ . The error  $e(n)$ , is computed by taking the difference of the prediction  $y^{(2)}(n)$  to the “desired” response, which ideally would be identical to the current sample  $x(n)$ . This error is propagated backwards through the entire network, in order to update the synaptic weights and is referred to as the training of the network. In this training phase all the synaptic weights  $w_{xx}(n)$  are updated in order to minimize the prediction error  $e(n)$ . In the following sections the different parts in the TLFFN and their properties are presented.

### 3.2.1 Forward Pass

The forward pass of the implemented neural network consist of two layers. The first layer, called the input or hidden layer, has  $M$  input nodes, corresponding to the  $M$  preceding samples of the prediction, and  $L$  neurons. Each neuron in the input layer is comprised of synaptic weights  $w_{ji}(n)$ , an external applied bias  $w_{j0}(n)$ , a summing junction, and a nonlinear amplification function  $\varphi(\cdot)$ , as illustrated in Figure 5. The number of synaptic weights in the input is  $(M+1)L$ . The nonlinear amplifier is referred to as the activation function. The second layer accommodates a single neuron that exist of  $L+1$  synaptic weights, a summing junction, and an activation function.

The input samples are amplified linearly on account of the synaptic weights and summed at each node. The bias terms  $w_{j0}$  and  $w_{k0}$  either increase or decrease the input terms of the activation function depending on whether they are positive or negative [20]. The sum of the weighted inputs, including the bias terms, is then amplified using the nonlinear activation function, before propagating further to the next layer in the network. The output of the summing junctions in the hidden layer is calculated as

$$v_i^{(1)}(n) = w_{j0}^{(1)}(n) + \sum_{j=1}^M w_{ji}^{(1)}(n)x_j(n-j) \quad i = 1 \dots L, \quad (1)$$

which can be written in matrix terms as

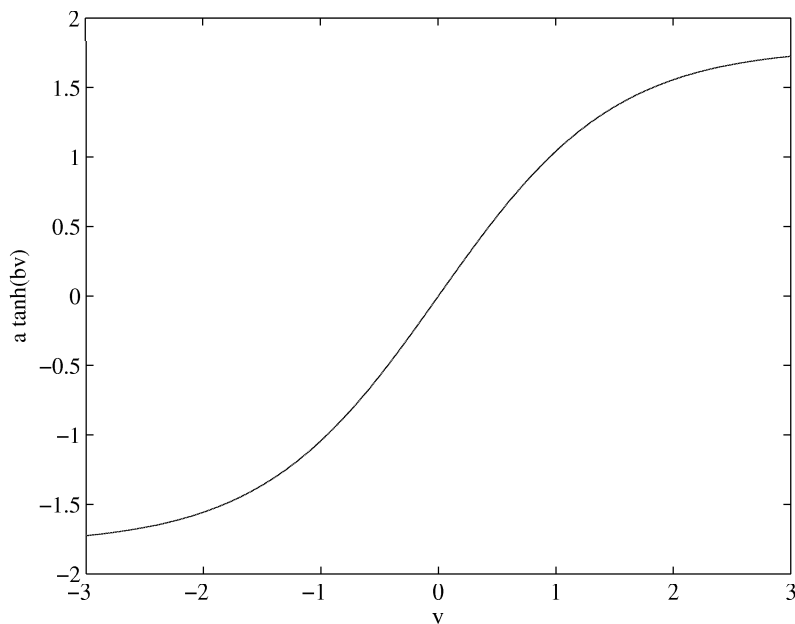
$$\begin{bmatrix} v_1^{(1)}(n) \\ v_2^{(1)}(n) \\ \vdots \\ v_L^{(1)}(n) \end{bmatrix} = \begin{bmatrix} w_{10}^{(1)}(n) & w_{11}^{(1)}(n) & \dots & w_{1M}^{(1)}(n) \\ w_{20}^{(1)}(n) & w_{21}^{(1)}(n) & \dots & w_{2M}^{(1)}(n) \\ \vdots & \vdots & \ddots & \vdots \\ w_{L0}^{(1)}(n) & w_{L1}^{(1)}(n) & \dots & w_{LM}^{(1)}(n) \end{bmatrix} \begin{bmatrix} 1 \\ x(n-1) \\ \vdots \\ x(n-M) \end{bmatrix} \quad (2)$$

$$\mathbf{v}^{(1)}(n) = \mathbf{W}^{(1)}(n)\mathbf{x}(n). \quad (3)$$

The vector computed in (3) is amplified on account of the activation function  $\varphi(\cdot)$ . The activation function in the proposed neural network is the hyperbolic tangent function as shown in Figure 6. The hyperbolic tangent function represents a graceful balance between linear and nonlinear behavior, i.e., the amplification is close to unity in the origin and saturates for larger values. Moreover, the hyperbolic tangent function is differentiable, which is a requirement in the backward pass and therefore commonly used in artificial neural networks [20]. Thus, the output of the first layer  $\mathbf{y}^{(1)}(n)$ , is calculated according to

$$\mathbf{y}^{(1)}(n) = a \tanh\left(b\mathbf{v}^{(1)}(n)\right), \quad (4)$$





**Figure 6:** Nonlinear activation function tanh with upper and lower boundary  $a=1.79$  and the temperature  $b=2/3$ .

where  $\mathbf{v}^{(1)}(n)$  and  $\mathbf{y}^{(1)}(n)$  are vectors with  $L$  elements. The steepness of the activation function is determined by  $b$ , with the upper and lower boundary defined by  $a$ . Recommended values for  $a$  and  $b$  are 1.79 and  $2/3$ , respectively [20]. These values obtain  $\varphi(1) = 1$  and  $\varphi(-1) = -1$ . The vector obtained according to (4) is the input to the next layer. In this case the next layer consists of a single neuron and corresponds to the output layer. The output of this neuron is a scalar computed according to

$$v^{(2)}(n) = w_{10}^{(2)}(n) + \sum_{j=1}^{I_1} w_{kj}^{(1)} \quad (5)$$

The expression in (5) can be written as a vector operation

$$v^{(2)}(n) = \mathbf{w}^{(2)}(n) \begin{bmatrix} 1 \\ \mathbf{y}^{(1)} \end{bmatrix}, \quad (6)$$

where

$$v^{(2)}(n) = \begin{bmatrix} w_{10}^{(2)}(n) & w_{11}^{(2)}(n) & \dots & w_{1L}^{(2)}(n) \end{bmatrix} \begin{bmatrix} 1 \\ y_1^{(1)}(n) \\ \vdots \\ y_L^{(1)}(n) \end{bmatrix}.$$

The activation function in the output layer is linear with unit gain [20]. Thus, the final output of the second layer is the linearly amplified and summed response of the first layer according to

$$y^{(2)}(n) = v^{(2)}(n), \quad (7)$$

The whitened output of the TLFN is the difference of the prediction  $y^{(2)}(n)$  to the desired value  $x(n)$  according to

$$e(n) = x(n) - y^{(2)}(n). \quad (8)$$

The output computed in (8) should ideally be a signal that only contains the whitened high frequencies of an R-wave, and is passed further to the pulse-shaping filter.

### 3.2.2 Back-Propagation Pass

In the back-propagation pass all the synaptic weights are updated with the back-propagation algorithm [20]. The objective of the back-propagation algorithm is the training of the network, which is carried out to minimize the prediction error. As defined in (8), the error  $e(n)$  is the deviation of a desired response  $x(n)$  to the current response  $y(n)$  of the network. In the general case all errors which can be calculated directly at the output of the network are squared and summed according to the following cost function

$$\mathcal{E}(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n), \quad (9)$$

where the set of  $C$  includes all the neurons in the output layer; in the presented study  $C = 1$  since only one output node exists. The error function (9) is minimized with the use of the LMS algorithm as [21]

$$\begin{aligned} \frac{\delta \mathcal{E}(n)}{\delta w_{ji}(n)} &= \frac{\delta \mathcal{E}(n)}{\delta e_j(n)} \frac{\delta e_j(n)}{\delta y_j(n)} \frac{\delta y_j(n)}{\delta v_j(n)} \frac{\delta v_j(n)}{\delta w_{ji}(n)} \\ \frac{\delta \mathcal{E}(n)}{\delta w_{ji}(n)} &= -e_j(n) \varphi_j'(v_j(n)) y_i(n). \end{aligned} \quad (10)$$

The correction  $\Delta w_{ji}(n)$  applied to the synaptic weights is performed with the

delta rule as

$$\Delta w_{ji}(n) = -\eta(n) \frac{\delta \mathcal{E}(n)}{\delta w_{ji}(n)}, \quad (11)$$

where  $\eta(n)$  is the time varying learning rate in the back-propagation algorithm. The negative sign in (11) is due to the gradient descent in weight space as shown in Figure 7. Inserting (10) in (11) obtains

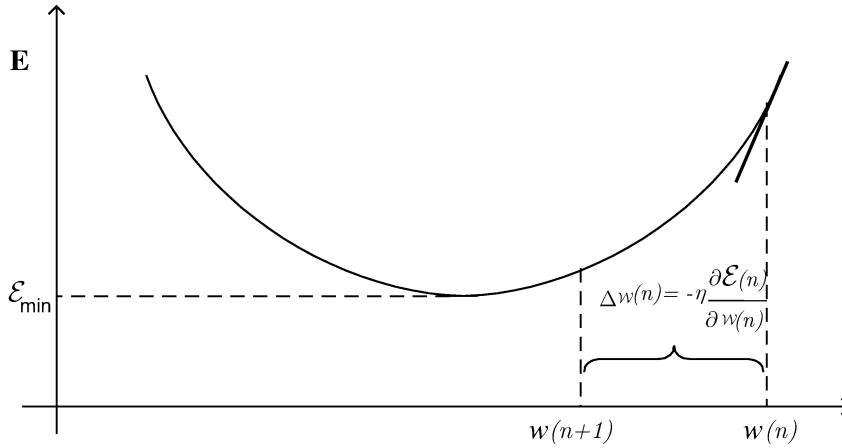
$$\Delta w_{ji}(n) = \eta \delta_j(n) y_i(n), \quad (12)$$

where the local gradient  $\delta_j(n)$  is defined as

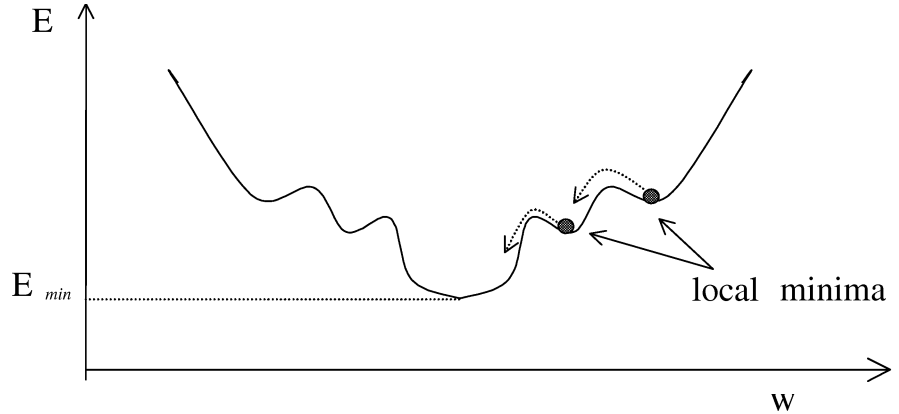
$$\begin{aligned} \delta_j(n) &= \frac{\delta \mathcal{E}(n)}{\delta v_j(n)} \\ &= \frac{\delta \mathcal{E}(n)}{\delta e_j(n)} \frac{\delta e_j(n)}{\delta y_j(n)} \frac{\delta y_j(n)}{\delta v_j(n)} \\ &= e_j(n) \varphi'_j(v_j(n)). \end{aligned} \quad (13)$$

As shown in (13), the local gradient for neuron  $j$  is the product of the corresponding error signal  $e_j(n)$  and the derivative  $\varphi'_j(v_j(n))$ .

**Momentum Term** A problem that may occur when minimizing the error  $e(n)$ , is that the procedure can get stuck in a local minima, as shown in Figure 8. A possible approach to reduce such risk is the use of a relatively high learning



**Figure 7:** Global minimum in weight space.



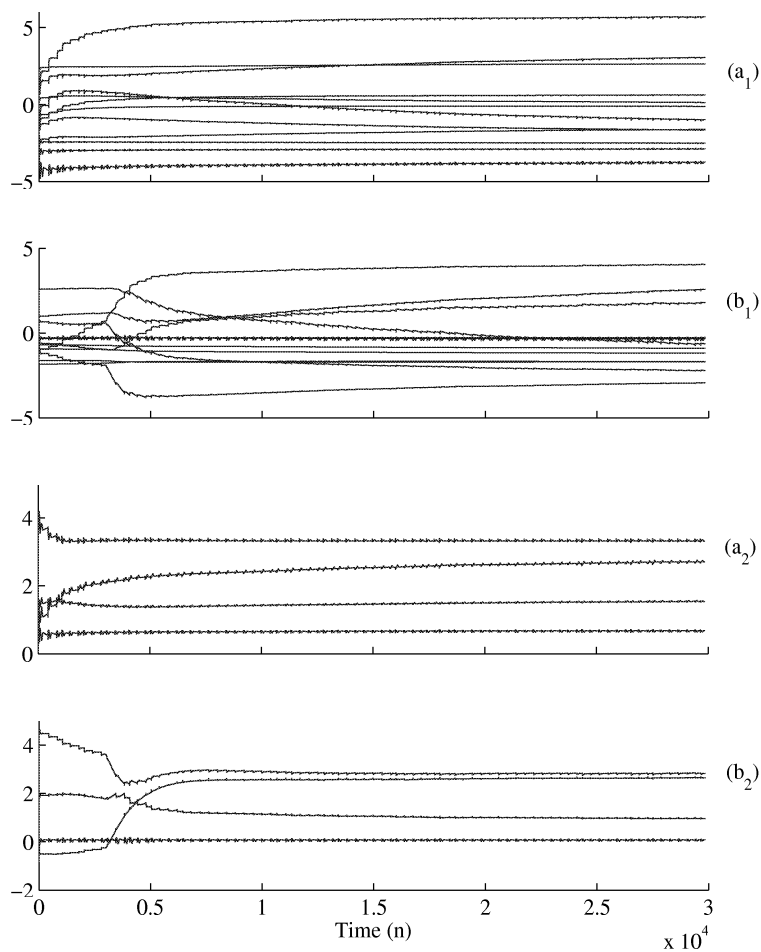
**Figure 8:** Local minima in weight space.

rate  $\eta(n)$ , which results in large changes of the synaptic weights. The drawback is that the network may become unstable. Conversely, the use of a small learning rate results in a smooth trajectory in weight space. However, this is attained at the cost of slower learning. A compromise solution is the use of a so-called *momentum term*  $\alpha$ . As shown in (14), the learning speed is accelerated with the addition of a fraction of the previous weight update to the current update as

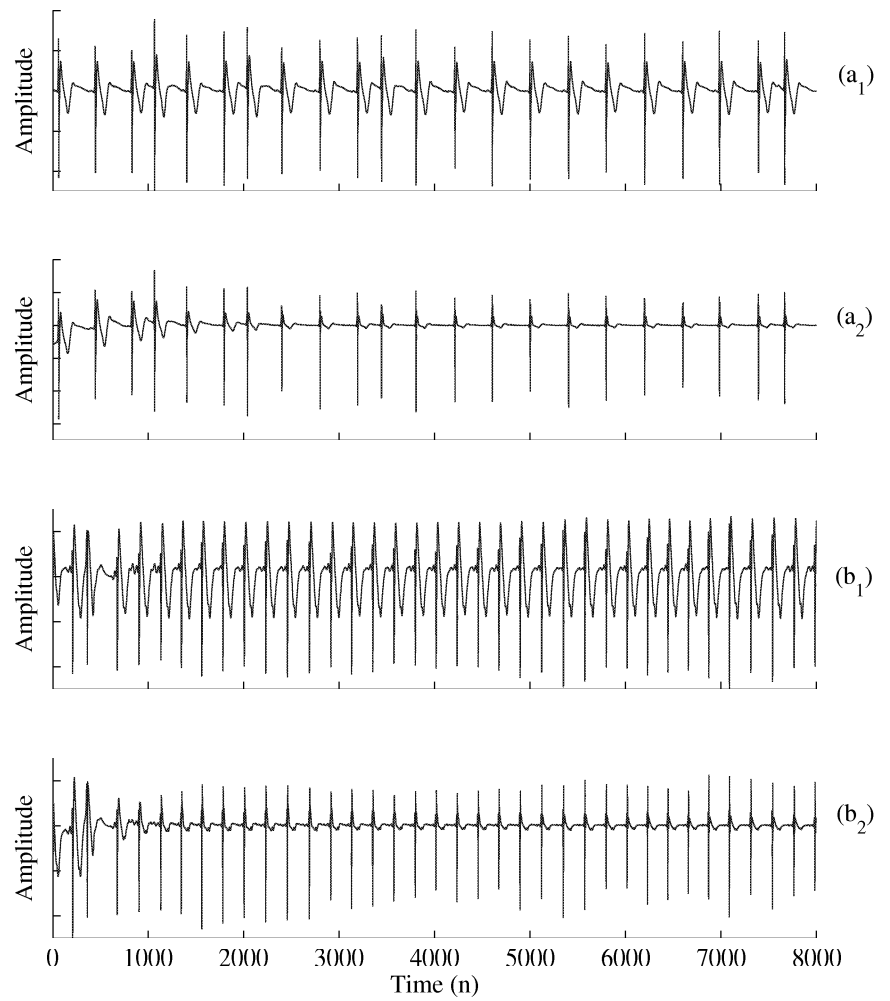
$$\Delta\omega_{ji}(n+1) = \eta(n)\delta_j(n)y_i(n) + \alpha\Delta\omega_{ji}(n). \quad (14)$$

The advantage of updating the synaptic weights according to (14) with a relatively high momentum term and a small learning rate is a fast learning and a smooth trajectory in TLFN weight space, where the risk of getting stuck in a local minima is reduced [20].

In Figure 9 the updates of the synaptic weights in the input and output layers are shown. Therefore, an undistorted and a distorted EGM are subjected to the TLFN and the weight update in the hidden and output layer is monitored. The graphs Figure 9 (a<sub>1</sub>) and (a<sub>2</sub>) represents a weight update on account of an undistorted EGM. It can be seen that the weights approximately saturate after 30000 training samples. The graphs in Figure 9 (b<sub>1</sub>) and (b<sub>2</sub>) visualize a weight update for a distorted EGM. It can be observed that several weight updates change abruptly at the time the interference is added, i.e., at  $n = 3000$ , before the weights saturate again. The small ripples, existent in all graphs, indicate a large prediction error which is due to R-wave occurrence.



**Figure 9:** The updated synaptic weights of the input and output layer. In (a<sub>1</sub>) and (a<sub>2</sub>) the training is carried out with an EGM recording without interference. (a<sub>1</sub>) represents the weights in the input layer and (a<sub>2</sub>) in the output layer. In (b<sub>1</sub>) and (b<sub>2</sub>) the EGM recording is interfered with 20 dB SNR originating from an AC hand drill. (b<sub>1</sub>) and (b<sub>2</sub>) represents the weights in the input and output layer, respectively.



**Figure 10:** EGM recordings from two patients a,b.  $(a_1)$  and  $(b_1)$  EGM recordings at the detector input.  $(a_2)$  and  $(b_2)$  EGM recordings after the whitening process.

In Figure 10 the EGM recordings of two patients before and after the whitening process are shown. It can be seen that the low frequency components are suppressed after the whitening process, such that the R-waves, indicated by large changes in amplitude, are easily detectable.

### 3.2.3 The Updated Learning Rate

The performance of a neural network is strongly dependent on the learning rate. Unfortunately, when dealing with data obtained from a nonlinear system, e.g., the human heart, it is very difficult or even impossible to predict a feasible learning rate a priori. The ideal case is that the update of the weights is paused if the prediction error is very small. Conversely, for larger errors a “speed-up” of the training is desirable. Thus, a convenient solution is to use a learning rate that has the ability to vary with time and each training set. However, the learning rate has to be limited within a certain boundary  $\hat{\eta}$  and  $\check{\eta}$  to avoid saturation in the backward pass. In this study it is proposed to let the learning rate vary according to

$$\eta(n) = \begin{cases} (1 + \gamma) \eta(n-1) & \text{if } \mathcal{E}(n) > \mathcal{E}(n-1), \eta < \hat{\eta} \\ (1 - \gamma) \eta(n-1) & \text{if } \mathcal{E}(n) < \mathcal{E}(n-1), \eta > \check{\eta} \\ \eta(n) & \text{otherwise} \end{cases} \quad (15)$$

As defined in (15), the learning rate increases if the preceding error is  $\eta$  percent smaller than the current error. Conversely, if the preceding error is  $\eta$  percent larger than the current error, the learning rate decreases. Moreover, the learning rate should be smaller in the output layer than in the front layers in order to avoid large changes at the output [20]. In this study the learning rate of the output layer is changing with the learning rate of the input layer, and is set to 70 % if the input layer rate.

### 3.3 The Pulse-Shaping Filter

The second block of the matched filter is referred to as the pulse-shaping filter [21]. In the present R-wave detector such a pulse-shaping filter is applied after the whitening process to maximize the SNR. The whitened signal contains information on the occurrence of an R-wave, but the shape of such an event has changed due to the whitening process. The impulse response of a the matched filter is a time-reversed replica of the whitened R-wave. The filter coefficients of the filter are given by

$$h(n) = \epsilon(N - 1 - n), \quad 0 \leq n \leq N - 1, \quad (16)$$

where  $e(n)$  is a vector of length  $N$  containing a whitened R-wave template [21]. The output of the pulse-shaping filter is computed as

$$y(n) = h(n) * e(n) = \sum_{k=0}^{N-1} h(n-k)e(k). \quad (17)$$

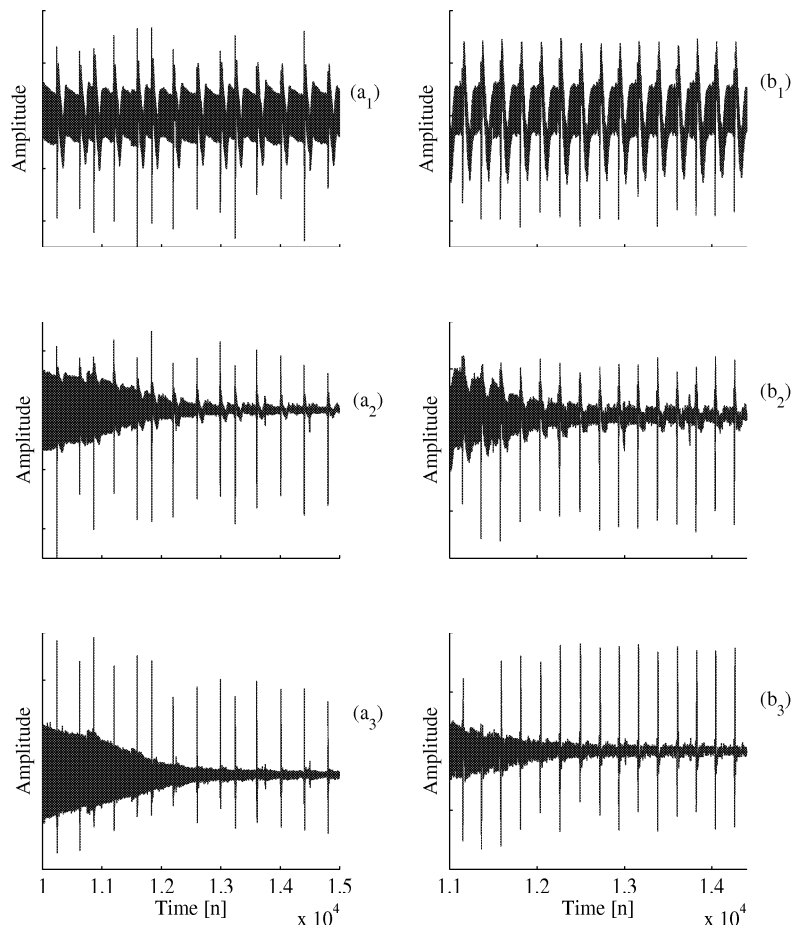
Figure 11 presents how two heavily distorted EGMs are processed at different instances of the R-wave detector. The graphs in Figure 11 ( $a_1$ ) and ( $b_1$ ) represent two EGMs which are heavily distorted by interference from an AC hand drill. In Figure 11 ( $b_2$ ) and ( $a_2$ ) the signal after the TLFN is presented. It can be seen that in both cases the SNR improves significantly after approximately 1200 input samples. This can be seen as a evidence of how the synaptic weights are adapted in order to minimize the prediction error. The graphs in Figure 11 ( $a_3$ ) and ( $b_3$ ) show the signal after matched filtering. The SNR is maximized compared to the graphs in Figure 11 ( $a_2$ ) and ( $b_2$ ), and the R-waves, indicated by spikes, are more easily detectable by the threshold function.

### 3.3.1 The Initial Template for the Matched Filter

The drawback with matched filtering is that the properties of the event, such as waveshape and duration, have to be known a priori. When dealing with a well defined and specified system this task can be easily solved. However, when dealing with EGMs it is not possible to define properties for the whitened R-wave that is applicable to all possible morphologies. One solution could be to take an averaged whitened R-wave template and update it with the actual coefficients. Unfortunately, not enough data is available to compute such a template and, moreover, it is not guaranteed that such an averaged template is suitable for all the different morphologies. Thus, the most convenient solution is to initialize the matched filter with individual coefficients. This has to be done during the implantation of the pacemaker either by the surgeon, or even better, automatically. It is evident that a automatic initialization is more desirable than a supervised one for following reason: the signals are interpreted differently by different persons and hence the sensing performance is very dependent on the experience of the surgeon. The drawback is that the morphology differs from patient to patient and therefore it is not easy to find an algorithm which does the automatic initialization as good as it can be done manually.

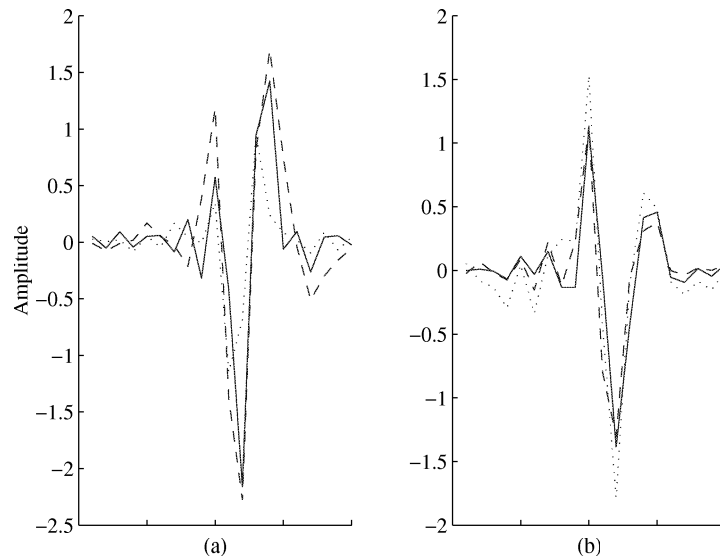
In this section an algorithm for the automatic initialization of an R-wave template is proposed. The algorithm has been tested on all the recordings in the EGM database and was able to find a typical whitened R-wave template for all the tested signals. This analysis was carried out with the assumption that the EGM was not interfered (no strong muscle contractions of the patient, which is very reasonable under the circumstance of an esthesia) during the initialization. Even though the EGM morphology differs from patient to patient





**Figure 11:** EGM recordings (a<sub>1</sub>) and (b<sub>1</sub>) are heavily by interference generated by an AC hand drill. (a<sub>2</sub>) and (b<sub>2</sub>) distorted EGMs after the TLFN. (a<sub>3</sub>) and (b<sub>3</sub>) are outputs of the pulse-shaping filter.

all the investigated QRS complexes in the signals had similar or comparable properties. The whitened R-wave had either a local maximum peak followed by a local minimum peak or a local minimum peak followed by local maximum peak, respectively, as presented in Figure 12. The signal changed from maximum  $\hat{e}$  peak to the minimum peak  $\check{e}$  within 12 ms, or vice versa, in all the



**Figure 12:** a) Typical whitened R-waves for signal #1, #10, #28. A local minimum is followed by a local maximum. b) Typical whitened R-waves for signal #13, #16, #22. A local maximum is followed by a local minimum.

investigated signals. In order to find the initial R-wave template the output of the predictor is scanned for minimum peaks the first four seconds. In respect to this minimum peaks the signal is scanned for maximum peaks prior or after the minimum peak. The center of the initial template is assigned to the larger value. If there is no sequence in the signal which meets the constraints the initialization phase can be lengthen or even it can be done manually. However, this was never necessary for the used signals.

### 3.3.2 The Update of the Impulse Response

When dealing with signals that originate from a human body it is very likely that the signal properties will change over time, e.g., the physical condition of the patient changes. With the lifetime of the today's pacemaker it is necessary for the filter parameters to adapt over time with respect to changing signal properties. Thus, the template containing the replica sequence needs to be updated. However, the vector that contains the update may be corrupt, due to presence of noise, or even worse, due to a false detection. In the long

run this will lead to a template which is not at all a replica of a whitened R-wave. This will result in an output where any event which coincides with the corrupt template will be maximized. In such a case more false detections are announced and the template would be updated with sequences, which are even more corrupt and, finally, the pacemaker would malfunction. In order to avoid such an improper update of the template, a quality constraint needs to be defined which decides whether the template should be updated. Thus, the objective is to find a quality constraint that causes an update of the template if the update  $h(n)$  is close enough to a mean  $\bar{h}_i(n)$ . In this section two algorithms are explored for their update efficiency. The first method calculates the distance  $d$  of the update to the  $\bar{h}(n)$  according to

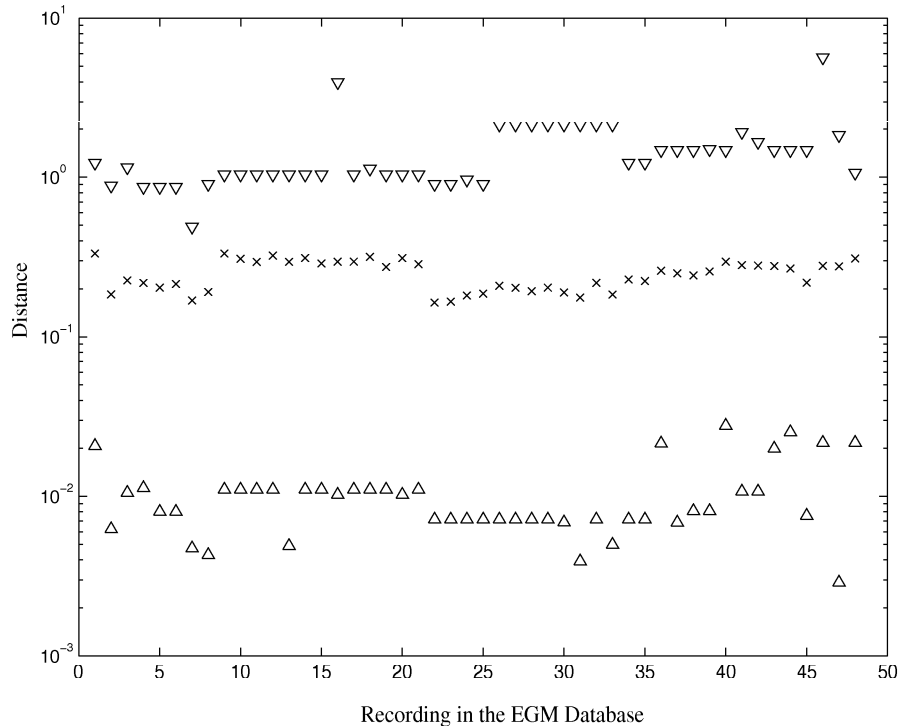
$$d_i \leq \frac{\sum_n (\bar{h}_i(n) - h_i(n))^2}{\sum_n \bar{h}_i(n)^2}. \quad (18)$$

The drawback with (18) is to define a value for  $d$ , which is applicable for any EGM morphology in the database. In Figure 13 it is shown that the average distance of each update to its correspondent mean varies between 0.005 and in the worst cases more than 2. The average distance for all means lies at 0.3. However, it is obvious that with such a distance spread of the updates it is impossible to define a value for  $d$  which guarantees a reasonable update of the template. Thus, a better approach needs to be explored.

The second proposed method is based on the computation of the cross-correlation of  $h_i(n)$  and  $\bar{h}_i(n)$ . The cross-correlation coefficient  $\rho_{h\bar{h}_i}$  of two vectors  $h_i(n)$  and  $\bar{h}_i(n)$  of the length  $N$  is defined according to [21]

$$\rho_{h\bar{h}_i} = \frac{\sum_{n=1}^N h_i(n)\bar{h}_i(n)}{\sqrt{\sum_{n=1}^N h_i^2(n) \sum_{n=1}^N \bar{h}_i^2(n)}}. \quad (19)$$

The value for  $\rho_{h\bar{h}}$  lies between ‘-1’ and ‘+1’, where ‘-1’ indicates antiphase of  $h(n)$  and  $\bar{h}(n)$  and ‘+1’ a 100% correlation, respectively. A value of zero indicates two independent vectors which might be the case due to a false detection. The normalized cross-correlation allows a comparison independent of the absolute values of the data. In Figure (14) it is shown that all cross-correlation coefficients of all the recordings in the EGM database lie between 0.2 and 1. The mean correlation coefficient for all the signals is above 0.85. In order to define a quality constraint that decides whether the template should be updated,  $\rho_{h\bar{h}}$  has to exceed 0.8 in this study. Thus,  $\bar{h}_i(n)$  will be updated with the majority of the updates  $h_i(n)$ . If  $\rho_{h\bar{h}_i}$  is below 0.8 no update of  $\bar{h}_i(n)$  will be carried out. Thus, it is likely that the template will not be updated with an

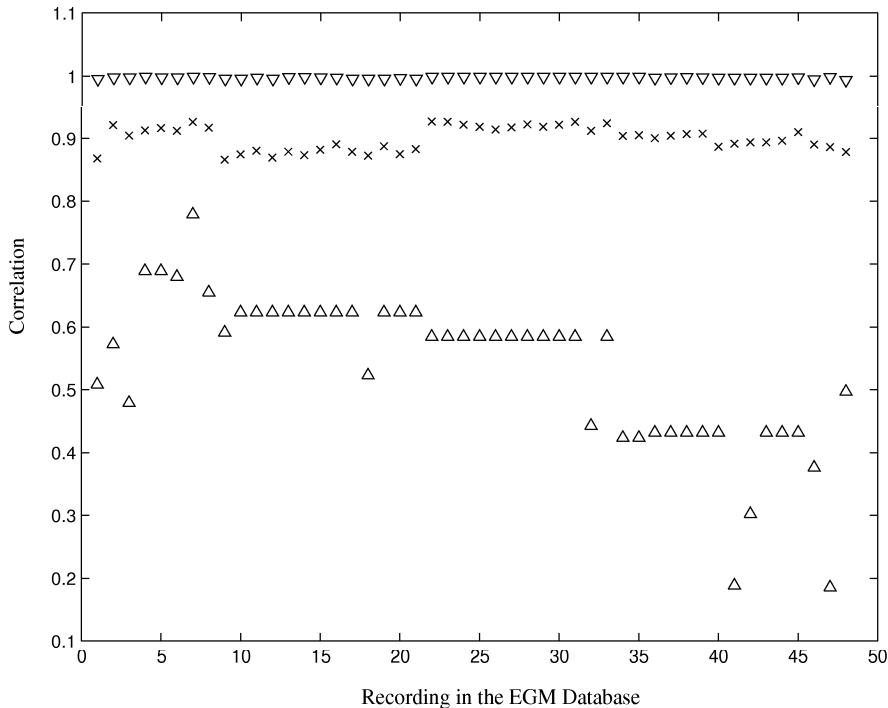


**Figure 13:** The best and worst distances of noise-free template updates to their respective mean for all the EGM recordings in the database. The minimum distance is marked with  $\Delta$ , the maximum distance with  $\nabla$  and the average distance with  $x$ , respectively.

impulse response that is heavily corrupt, e.g., generated due to a false detection. Thus, the template is unlikely to change if the EGM is heavily interfered, even over a longer time period. Moreover, it is possible to track slow changes in the EGM morphology by updating  $\bar{h}_i(n)$ . If (19) is met the template for the matched filter is updated according to

$$\bar{h}_{i+1}(n) = c \cdot \bar{h}_i(n) + (1 - c) \cdot h_i(n), \quad (20)$$

where  $c$  is an update factor. The larger  $c$  the smaller the update of the template according to (20), where  $c$  is limited as  $0 \leq c \leq 1$ . However, in order to get a mean template of the previous R-wave complexes and not to overweight a new template containing the latest impulse response, it is convenient to choose a high update factor close to 1. Another advantage is, that in the case of a false detection, although (19) is met, the influence of the update is kept rather small.



**Figure 14:** The best and worst correlations coefficients  $\rho_{h\bar{h}}$  of noise-free template updates to their respective mean. The minimum  $\rho_{h\bar{h}}$  is marked with  $\Delta$ , the maximum  $\rho_{h\bar{h}}$  with  $\nabla$  and the average  $\rho_{h\bar{h}}$  with  $x$  respectively.

### 3.3.3 The Template Length of the Pulse-shaping Filter

The maximum delay for the detector to decide whether an R-wave occurred is 40 ms [22]. The pulse-shaping filter is introducing a delay of  $\frac{N}{2} - 1$ , where  $N$  is the template length. Thus, the maximum template length is 80 ms. However, sufficient filtering performance is achievable with a shorter template. In the applied matched filter a template length is set to 48 ms, as presented in Figure 12.

## 3.4 The Time-Varying Decision Rule

After matched filtering a decision rule decides whether an R-wave has occurred. However, the sensing threshold changes significantly the first months due to time-dependent changes on the electrode tip, as presented in Section 3.3. More-

over, due the long term changing signal properties it is not sufficient to use a fixed threshold level which is programmed at pacemaker implantation. If such level is set to a too low level, many false alarms may occur, which is a threatening situation for the patient. False alarms are indicated as beats and due to this misclassification the pacemaker does not emit a pulse when required. This can lead, in the worst case, to a cardiac arrest. Conversely, if the threshold is set to a too high level, occurring beats may not be detected, and the pacemaker will emit a pulse, although not supposed to. Such additional pulses are uncomfortable for the patient and shortens the pacemaker lifetime. Thus, it is necessary that the decision rule adapts to the changing signal properties.

This section presents an algorithm the currently updates the threshold function. In the initialization phase a mean,  $\gamma_k$ , of the peak amplitude of the R-wave detections  $\hat{\gamma}_j$ , is computed according to

$$\gamma_k = \frac{1}{\sum j} \sum_{j=1}^P j \hat{\gamma}_j, \quad (21)$$

where  $P$  is the number of detections during the initialization. Multiplying  $\hat{\gamma}_j$  with  $j$  gives the early detections a lower weight. This is necessary since at the beginning of the training  $e(n)$  is usually higher due to the inadequate synaptic weight adjustments. However, after the initialization phase the threshold is smoothly updated in order to track long term signal changes according to

$$\gamma_{i+1} = \delta \gamma_i + (1 - \delta) \hat{\gamma}_i \quad (22)$$

where  $\delta$  is a updating factor and  $\gamma_k$  is used as  $\gamma_i$  immediately after the initialization phase. The value for  $\delta$  is set close to one to perform a slow update. The final threshold function is a fraction of (22) computed according to

$$\gamma_{thi} = \beta \gamma_i, \quad (23)$$

where  $\beta$  is a constant that defines the size of the fraction. Finally, the threshold defined in (23) decides whether an R-wave occurred according to

$$y^{(2)}(n) \geq \gamma_{thi}. \quad (24)$$

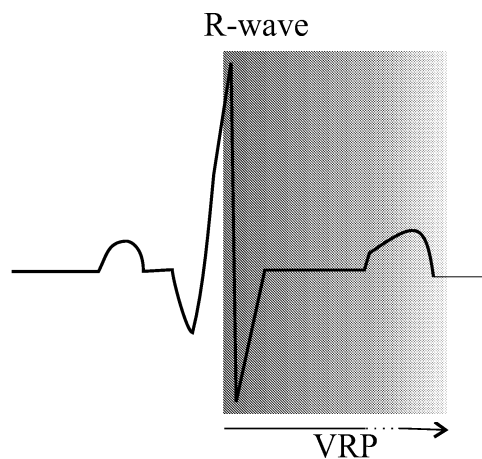
If  $y^{(2)}(n)$  exceeds  $\gamma_{thi}$  an R-wave is indicated and an update of the template for the matched filter and the threshold function according to (20) and (22) is eventually carried out. However, if

$$y^{(2)}(n) \geq 2\gamma_{thi}, \quad (25)$$

$y^{(2)}(n)$  is truncated according to

$$y^{(2)}(n) = y^{(2)}(n - 1), \quad (26)$$

to limit the influence of the update. In the case that no R-wave is detected within a certain time range the pulse generator is alerted in order to send out an artificial impulse.



**Figure 15:** The VRP is the time gray shaded time period and starts after a sensed R-wave. It lasts typically 200–350 ms.

### 3.4.1 Pausing of the Threshold Function

After an identified R-wave the detector does not expect any further event for a certain time span. This time span is called the ventricular refractory period (VRP), as presented in Section 2.1. During this time span parts of the detector can be shut off for following reasons [14]:

- An emitted pulse from the generator must not be sensed again.
- The R-wave detector should not be disturbed by an R-wave due to a paced ventricular beat
- To prevent the sensing of afterpotentials

The ventricular-based refractory period (VRP) of a single chamber pacemaker is shown in Figure 15. The VRP is a parameter that can be programmed during pacemaker implantation. In the proposed R-wave detector the VRP is set to 200 ms.

For the duration of the VRP major parts of the detector can be powered down, if implemented in digital hardware. With a VRP duration of 200–350 ms the hardware is in a sleep mode for 20–35 % of the time. This sleeping mode result in energy savings and hence a prolongation of the pacemaker lifetime. The drawback with the VRP is that the detector is absolutely inactive for a certain time period.

## 4 Detection Performance

In this section the detection performance for the R-wave detector is studied. This is carried out by adding interferences to the EGM recordings with different SNRs before subjecting them to the detector. Different SNRs are applied since the interferences were recorded in a non working condition, e.g., the AC hand drill was switched on but not drilling a hole. It is assumed that under working conditions the amplitude of the interference would increase significantly. Thus, the interferences are amplified in order to simulate a working condition of the devices. Moreover, shape and body constitution of humans vary considerably and, therefore, it is not possible to find one single estimate of how much noise can interfere with the pacemaker. However, a SNR of 20 dB corresponds to a very high interference level which should include the worst-case situation in real life. Thus, noise levels are chosen which results in SNRs of 20 and 25 dB to assure that this situation can be handled by the detector. The signals are subjected to the TLFN, the pulse-shaping filter and the decision rule, as described in the preceding sections. The detected events, according to (24), are finally analyzed by computing the *detection rate* ( $P_D$ ) and the *false alarm rate* ( $P_{FA}$ ) as

$$P_D = \frac{N_T}{N_T + N_M} \qquad P_{FA} = \frac{N_{FA}}{N_T + N_{FA}}, \quad (27)$$

where  $N_T$  is the number of true detections,  $N_M$  the number of missed detections and  $N_{FA}$  the number of false alarms, respectively. A true detection is defined as one event that occurs within 100 ms, 50 ms before and after the annotated event. Remaining events are declared as false alarms. With these quality measurements the dependency of the detection performance on changing parameters, such as SNR and threshold levels can be compared. Finally, the average performance of the R-wave detector is expressed as the mean of all  $P_D$  and  $P_{FA}$ .

### 4.1 Signal-to-Noise Ratio Definition

The analyzed signal consists of the nonstationary EGM,  $x(n)$ , to which a noise signal,  $v(n)$ , has been added. The signal-to-noise ratio (SNR) of  $y(n)$  is defined as

$$SNR = 20 \cdot \log \frac{V_x}{\sigma_V}, \quad (28)$$

where  $V_x$  is the average peak-to-peak amplitude of all the R-waves in one EGM recording and  $\sigma_V$  the standard deviation of the noise to be added. The peak-



to-peak amplitude  $V_x$  is calculated according to

$$V_x = \frac{1}{N_x} \sum_{i=1}^{N_x} \left| \max_{-50 \leq m \leq 20} \{x(R_i + m)\} \right| + \left| \min_{-50 \leq m \leq 20} \{x(R_i + m)\} \right|, \quad (29)$$

where  $x(R_i)$  is a vector containing an R-wave positioned at  $R_i$ , and  $N_x$  is the number of R-wave templates. The standard deviation  $\sigma_V$  is calculated according to

$$\sigma_V = \frac{1}{N_v} \sqrt{\left( \sum_{i=1}^{N_v} v(i) - \bar{v} \right)^2}, \quad (30)$$

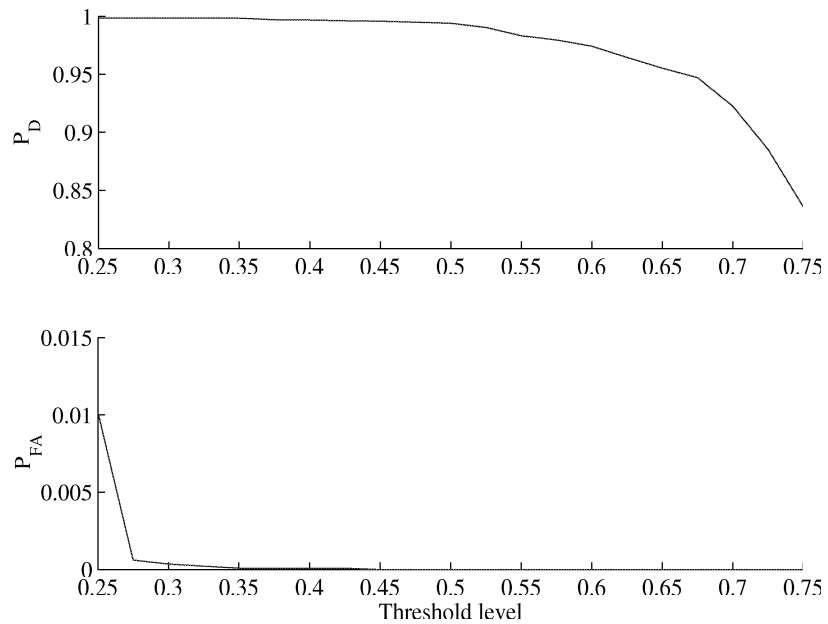
where  $\bar{v}$  is the mean of the noise signal and  $N_v$  the number of discrete samples.

## 4.2 Detection Performance for the Noise-Free Case

Despite the importance of handling heavily disturbed EGMs, the R-wave detector is mostly operating in a low-noise environment. Therefore, a performance analyzes where only “physiological” noise, such as the far-field P- and T-wave, has to be suppressed is also carried out in this study. In Figure 16 the means of  $P_D$  and  $P_{FA}$  for a threshold between 0.25 and 0.75 are shown. The maximum value for  $P_{FA}$  is 0.01 at a threshold level  $\beta = 0.25$  and a desirable  $P_{FA}$  that equals zero is achieved at a threshold level  $\beta = 0.45$ . In this threshold span the detection rate  $P_D$  drops from 0.9985 to 0.9956. From this analysis it is concluded that a threshold that compromises  $P_D$  and  $P_{FA}$  lies in the range of  $0.45 \leq \beta \leq 0.55$ . Using a threshold below the lower limit,  $\beta = 0.45$  would result in a higher false alarm and detection rate, whereas exceeding the upper limit would only reduce the detection rate.

## 4.3 Detection Performance for Interfered EGMs

Although the pacemaker patient is mostly exposed to low-noise environment, the importance of handling heavily disturbed EGMs must nonetheless be addressed. The detection performance for the circumstance when the pacemaker patient is exposed to various interferences is analyzed in this section. The EGMs are disturbed with recordings of the interference database, see Section 2.2 and 2.3. Shape and body constitution of humans vary considerably and, therefore, it is not possible to find one single estimate of how much noise can interfere with the pacemaker. However, a SNR of 20 dB corresponds to a very high interference level which should include the worst-case situation in real

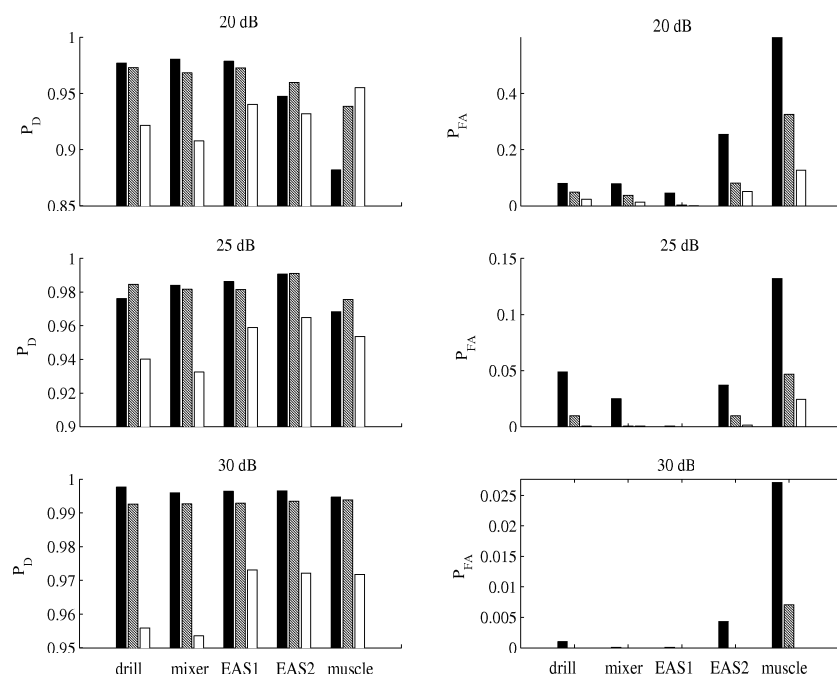


**Figure 16:** Average detection  $P_D$  and false alarm rate  $P_{FA}$  for all the EGM recordings. The threshold  $\beta$  is varied from 0.25 to 0.75.

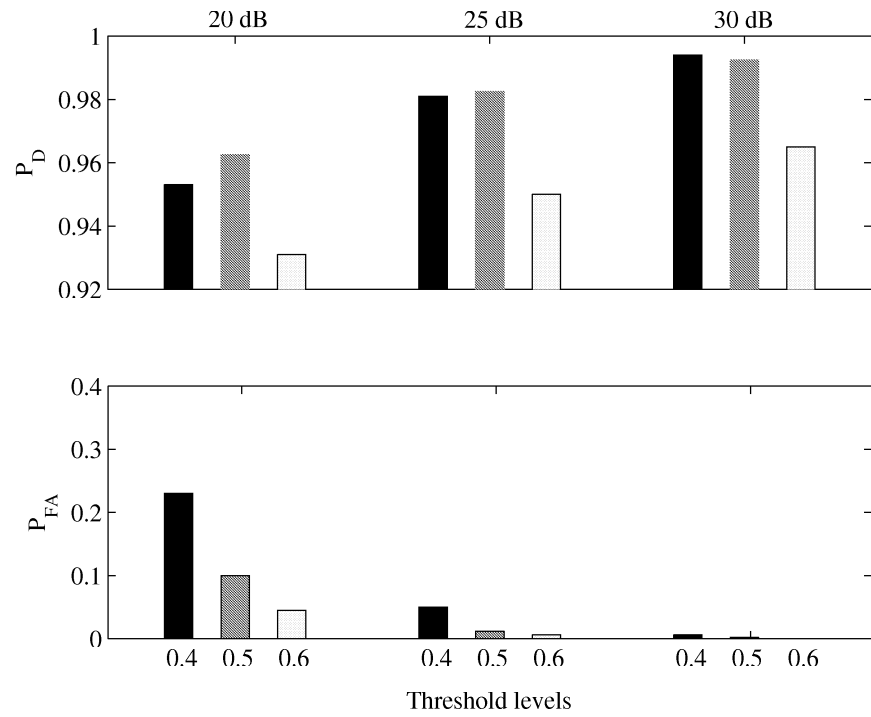
life. Thus, noise levels are chosen which results in SNRs of 20, 25, and 30 dB to assure that this situation can be handled by the detector. The threshold level  $\beta$  is varied from 0.4 to 0.6. The means of the detection rate  $P_D$  and false alarm rate  $P_{FA}$  are shown in Figure 17. The R-wave detector is most sensitive to interferences originating from EAS2 and muscle contractions at SNRs of 20 and 25 dB, as demonstrated in Figure 17. However, for a SNR of 25 dB, noise originating from a AC hand drill causes more false alarms and less detections than from EAS2.

In Figure 18 the average detection performance for disturbing all the recordings in the EGM database with the interference recordings is shown. It can be seen that the detector has reliable detection performance at 30 dB interference level for the threshold levels  $\beta = 0.4$  and  $\beta = 0.6$ . In both cases  $P_D$  is over 0.99, whereas the values for  $P_{FA}$  are below 0.006. The use of a higher threshold level results in lower values of  $P_D$ .

For higher interference levels, i.e., 20 and 25 dB, both detection and false alarm rate get worse.



**Figure 17:** Means values of detection performance in terms of detection rate  $P_D$  and false alarm rate  $P_{FA}$ . The recordings from the EGM database are disturbed with interferences originating from a hand drill, mixer, EAS1, EAS2, and muscle contraction. The applied SNRs are 20 dB, 25 dB and 30 dB. The threshold  $\beta$  is varied from 0.4 (dark bars) to 0.6 (bright bars).



**Figure 18:** Means of  $P_D$  and  $P_{FA}$  for all recordings in the EGM database interfered with all the recordings from the interference database. The threshold  $\beta$  is varied from 0.4 to 0.6.

## 5 Discussion

In this study an R-wave detector for pacemakers has been developed. The detector consists of a TLFN, a dynamically updated pulse-shaping filter and a time varying threshold level. Recordings from an EGM database are disturbed with recordings from an interference database to determine the detection performance when the pacemaker patient is exposed to electric or magnetic radiating devices. The detector properties adapt automatically to changes in the EGM morphology. The detection performance is measured in terms of detection and false alarm rates for different SNRs and threshold levels. It is observed that detection performance is sustained by the pulse shaping filter although the synaptic weights are mismatched due to sudden interference occurrence. Despite the reliable detection performance, the suitability for an implementation in a low power digital ASIC for pacemaker application is rated as low. The algorithm is very complex. The synaptic weights in the forward pass would already result in 19 generic multipliers. The nonlinear activation function can be accomplished using a lookup-table. Moreover, due to a continuously ongoing update of the synaptic weights of the TLFN, the template of the matched filter and the threshold level, the computational costs and consequently the power dissipation will be too high.

## Acknowledgment

The author is grateful to St. Jude Medical AB, Järfälla, Sweden for providing the data for this study.

## References

- [1] R. Elmqvist, J. Landegren, S. Petersson, Å. Senning, and G. W. Ollson, "Artificial pacemaker for treatment of adams-stokes syndrome and slow heart rate," *Am Heart J*, no. 65, pp. 731–748, 1963.
- [2] J. Kilby, "Invention of the integrated circuit," *IEEE Trans. Electron Devices*, p. 648, 1976.
- [3] H. Smith, N. Fernot, and W. Hillenbrand, "Concepts of rate responsive pacing," *IEEE Eng. Med. Biol. Mag.*, pp. 32–35, June 1990.
- [4] "St-Jude Medical," <http://www.sjm.com>.
- [5] G. Friesen, T. Jannett, M. Jadallah, S. Yates, S. Quint, and H. Nagle, "A comparison of the noise sensitivity of nine QRS detection algorithms," *IEEE Trans. Biomed. Eng.*, pp. 85–98, 1990.
- [6] O. Pahlm and L. Sörnmo, "Software QRS detection in ambulatory monitoring — a review," *Med. Biol. Eng. Comput.*, vol. 22, pp. 289–297, 1984.
- [7] B.-U. Köhler, C. Hennig, and R. Orglmeister, "The principles of QRS detection," *IEEE Eng. Med. Biol. Mag.*, pp. 42–57, 2002.
- [8] E. Lucas, "The effect of electronic article surveillance systems on permanent cardiac pacemakers," *PACE*, vol. 17, pp. 2021–2026, 1994.
- [9] B. Dodinot, J.-P. Godenir, and A. Costa, "Electronic article surveillance: A possible danger for pacemaker patients," *PACE*, vol. 16, pp. 46–53, 1993.
- [10] M. McIvor, J. Reddinger, E. Floden, and R. Sheppard, "Study of pacemaker and implantable cardioverter defibrillator triggering by electronic article surveillance devices," *PACE*, vol. 21, pp. 1847–1861, 1998.
- [11] J. Mugica, L. Henry, and H. Podeur, "Study of interactions between permanent pacemakers and electronic antitheft surveillance systems," *PACE*, vol. 23, pp. 333–337, 2000.
- [12] D. Hayes, P. Wang, D. Reynolds, M. Estes, J. Griffith, R. Steffens, G. Carlo, G. Findlay, and C. Johnson, "Interference with cardiac pacemakers by cellular telephones," *The New Engl. J. of Med.*, vol. 336, no. 21, pp. 1473–1479, 1997.
- [13] W. Irnich, "Interference in pacemakers," *PACE*, vol. 7, pp. 1021–1048, November/December 1984.
- [14] J. Webster, *Design of cardiac pacemakers*. New York, USA: IEEE Press, 1995.

- 
- [15] A. Gerosa and A. Neviani, "A very low-power 8-bit  $\Sigma \Delta$  converter in a  $0.8\mu\text{m}$  CMOS technology for the sensing chain of a cardiac pacemaker, operating down to 1.8V," in *Proc. 2003 IEEE Intl. Symp. Circuits Systems*, 2003.
- [16] B. Moberg and H. Strandberg, "Effects of interference on pacemakers," *Eur. J. C. P.E.*, vol. 5, pp. 146–157, 1995.
- [17] W. Irnich, "Muscle noise and interference behaviour in pacemakers: A comparative study," *PACE*, vol. 10, pp. 125–132, 1987.
- [18] Q. Xue, Y. Hu, and W. Tompkins, "Neural-network-based adaptive matched filtering for QRS detection," *IEEE Trans. on Biomedical Engineering*, pp. 317–328, 1992.
- [19] L. Sörnmo and P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*. Amsterdam: Elsevier, 2005.
- [20] S. Haykin, *Neural Networks*, 2nd ed. New York: Prentice Hall, 1999.
- [21] C. Therrien, *Discrete random signals and statistical signal processing*. Englewood Cliffs, NJ, USA: Prentice Hall, 1992.
- [22] L. Råde and B. Westergren, *Beta, Mathematics Handbook*. Studentlitteratur, 1998.





## Part II



## Part II

### Digital Implementation of a Wavelet Based Event Detector

#### Abstract

This chapter presents a digital hardware implementation of a novel wavelet based event detector suitable for the next generation of cardiac pacemakers. Significant power savings are achieved by introducing a second operation mode that shuts down 2/3 of the hardware for long time periods when the pacemaker patient is not exposed to noise, while not degrading performance. Due to a 0.13  $\mu\text{m}$  CMOS technology and the low clock frequency of 1 kHz, leakage power becomes the dominating power source. By introducing sleep-transistors in the power supply rails, leakage power of the hardware being shut off is reduced by 97%. Power estimation on RTL-level shows that the overall power consumption is reduced by 67% with a dual operation mode. Under these conditions the detector is expected to operate in the sub- $\mu\text{W}$  region. Detection performance is evaluated by means of databases containing electrograms to which five types of exogenic and endogenic interference are added. The results show that reliable detection is obtained at moderate and low SNRs. Average detection performance in terms of detected events and false alarms for 25 dB SNR is  $P_D = 0.98$  and  $P_{FA} = 0.014$ , respectively.

---

Based on:

J. Neves Rodrigues, V. Öwall, and L. Sörnmo, "A Wavelet Based R-wave Detector for Cardiac Pacemakers in 0.35 CMOS Technology," *Proceedings of the 2004 IEEE International Symposium on Circuits and Systems, ISCAS 2004, Vancouver, Canada*,

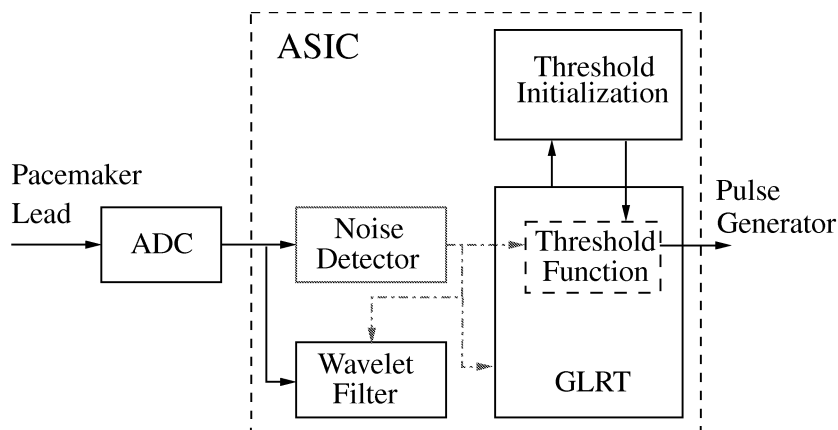
J. Neves Rodrigues, V. Öwall, and L. Sörnmo, "A flexible wavelet filter structure for cardiac pacemakers: A power efficient implementation," *Proceedings of the 2004 IEEE International Symposium on Biomedical Circuits and Systems, BIOCAS 2004, Singapore*,



J. Neves Rodrigues, T. Olsson, L. Sörnmo, and V. Öwall, "A Dual-Mode Wavelet Based R-Wave Detector using Single- $V_t$  for Leakage Reduction," *Proceedings of the 2005 IEEE International Symposium on Circuits and Systems, ISCAS 2005, Kobe, Japan*,

J. Neves Rodrigues, T. Olsson, L. Sörnmo, and V. Öwall, "On the Digital Implementation of a Wavelet Based Event Detector for Cardiac Pacemakers," *IEEE Transactions on Circuits and Systems. Special Issue on Biomedical Circuits and Systems: A New Wave of Technology*, accepted for publication.





**Figure 1:** Block diagram of the event detector and the target environment. Solid lines represent the dataflow while dashed-dotted lines the mode-select signals.

## 1 Introduction

Device longevity is a crucial design constraint in the evolving area of medical implants since replacement of implanted devices results in discomfort for the patient and high economical costs. Medical implants such as the cardiac pacemaker may last up to 20 years, e.g., *Regency SC+*, *St-Jude Medical* [1]. At the same time, reliable detection performance, closely related to longevity, is essential as the number of devices that may interfere with the pacemaker is ever increasing.

A variety of event detectors for electrocardiograms (ECGs) have been proposed during the last three decades [2–4]. However, most of them are unsuitable for pacemaker applications since they do not operate in real time. Traditionally, event detectors are composed of a bandpass filter followed by a programmable threshold level, implemented in analog circuitry [5]. The proposed implementation is optimized for digital circuitry, and the wavelet based structure offers a higher flexibility for different morphologies. Together with a low power analog-to-digital converter (ADC) a single-chip solution becomes possible.

Digital hardware is feasible for today’s pacemaker generation due to recent development in low power ADCs, e.g., the ADC in [6] operates at  $2.2\mu\text{W}$ . Moreover, shrinking technologies and advances in low power digital circuitry makes a digital solution a competitive alternative to analog solutions. In favor to analog circuitry [7], a digital implementation has the advantage of accommodating more advanced signal processing such as features for morphology classification, e.g., in implantable cardioverter defibrillators (ICDs) and data

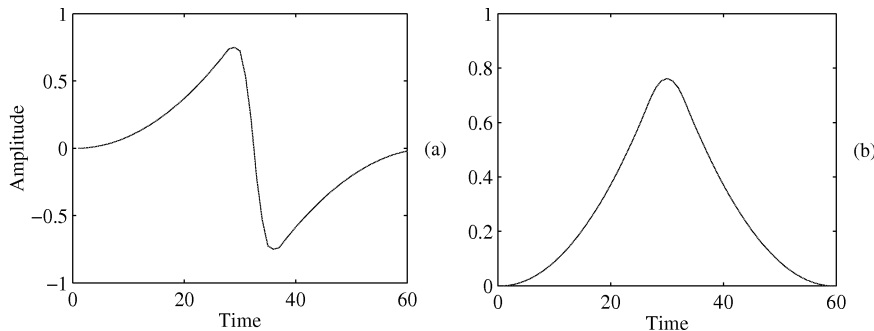
compression for postanalysis [8, 9]. Postanalysis provides better knowledge of diseases and improves pacemaker/ICD parameter tuning [10, 11].

The proposed event detector is based on a wavelet filterbank that decomposes the input signal into subbands, followed by hypothesis testing, as presented in Figure 1 [12–16]. The threshold initialization hardware monitors the output of the hypothesis test and sets initial value. The threshold function of the hypothesis test determines whether the incoming beat is considered as cardiac activity or as noise. A dual operation mode of the detector is proposed by which major parts of the hardware can be shut down when the pacemaker patient is at rest or in a low-noise environment. Reliable detection performance is sustained by a noise detector that operates in supervision mode and reactivates the sleeping hardware when necessary, see Figure 1.

Dynamic power savings are achieved using a gated clock to shut off parts of the deactivated detector. However, as the event detector is targeted to operate at a low frequency of 1 kHz, leakage power is the main contributor to the total power figure. Therefore, leakage reduction techniques are required to efficiently address power reduction. In the present implementation gate transistors are used to effectively turn off the supply voltage and, thereby, reduce the leakage power [17–20].

In Section 2 the detector principles are presented as well as the databases used for evaluation. Section 3 describes the implementation and optimization in digital hardware. Moreover, a hardware realization for a dual operation mode is presented. The performance of the event detector is discussed in Section 4. Power optimization, an estimate for the core power consumption, and ASIC placement and routing are presented in Section 5. Finally, conclusions are presented in Section 6.





**Figure 2:** Electrogram events with (a) a typical biphasic and (b) monophasic signal [26].

## 2 Materials and Methods

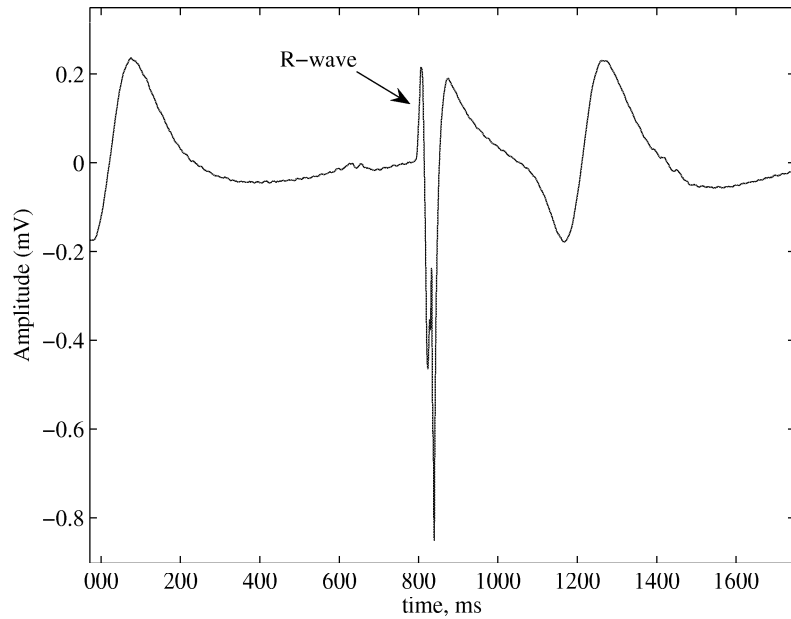
The electrical activity at the pacemaker electrode tip is reflected by the intracardiac electrogram (EGM) [21–24]. The depolarization and repolarization waves are decomposed into two perpendicular waves: one that propagates horizontally and another that propagates transversally to the myocardial wall [25]. Thus, the morphologies of these two waves differ significantly. The horizontal wave is composed of a large positive charge that rapidly changes to a negative charge resulting in a *biphasic wave*, whereas the transversal wave results in a *monophasic wave*, see Figure 2.

Ventricular depolarization usually represents the cardiac event in an EGM and is referred to as the “R-wave” in this study, see Figure 3; its duration is normally between 60 and 100 ms [5, 21].

### 2.1 Wavelet Filterbank and Generalized Likelihood Ratio Test

This section presents a brief theoretical background of the wavelet filterbank and generalized likelihood ratio test needed to comprehend the hardware implementation. A more detailed description is to be found in [16, 26].

The detector structure was developed with efficient digital hardware implementation in mind. The wavelet filterbank is a combination of a biphasic (antisymmetric) and a monophasic (symmetric) filter function that approximates biphasic and monophasic morphologies. The transfer function  $h_{q,b}(n)$  of



**Figure 3:** Electrogram of a cardiac cycle. Ventricular depolarization is reflected by the R-wave.

the biphasic wavelet filterbank is modeled as

$$\begin{aligned}
 h_{1,b}(n) &= g_b(n) \\
 h_{2,b}(n) &= f(n) * g_b(2n) \\
 h_{3,b}(n) &= f(n) * f(2n) * g_b(4n) \\
 &\vdots \\
 h_{q,b}(n) &= f(n) * \dots * f(2^{q-2}n) * g_b(2^{q-1}n),
 \end{aligned} \tag{1}$$

where  $q$  is the scale factor [16]. An analysis has shown that three scales,  $q = 2, 3, 4$ , are sufficient to cover the frequency spectrum of an R-wave [16]. The case  $q = 1$  is not considered in the design, as no prior filtering is defined. The functions  $g_b(n)$  and  $f(n)$  in (1) are defined as

$$g_b(n) = [-1 \ 1] \tag{2}$$

and

$$f(n) = [1 \ 3 \ 3 \ 1], \quad (3)$$

respectively. To achieve power-efficient hardware mapping, short filters with integer values are chosen, and, therefore,  $g_b(n)$  in (2) is chosen as a first order difference, and the impulse response  $f(n)$  in (3) was chosen as a third order binomial function [16]. The monophasic filterbank  $g_m(n)$  is modeled by reusing  $g_b(n)$  as

$$g_m(n) = g_b(n) * g_b(n) = [1 \ -2 \ 1], \quad (4)$$

such that the transfer function  $h_{q,m}(n)$  of the monophasic filterbank is modeled as

$$h_{q,m}(n) = f(n) * \dots * f(2^{q-2}n) * g_m(2^{q-1}n). \quad (5)$$

The output of the wavelet filterbank is defined as

$$\mathbf{y}(n) = \mathbf{x}^T(n)\mathbf{H}, \quad (6)$$

where

$$\mathbf{x}(n) = [x(n) \dots x(n+N-1)]^T \quad (7)$$

is the input to the wavelet filterbank;  $\mathbf{H}$  is defined as

$$\mathbf{H} = [\tilde{\mathbf{H}}_b \ \tilde{\mathbf{H}}_m], \quad (8)$$

and can be efficiently implemented by Mallat's algorithm [27]. The matrices  $\tilde{\mathbf{H}}_m$  and  $\tilde{\mathbf{H}}_b$  in (8) denote the reversals of  $\mathbf{H}_m$  and  $\mathbf{H}_b$ , respectively, where the latter is defined, defined by

$$\mathbf{H}_b = [\mathbf{h}_{2,b} \ \mathbf{h}_{3,b} \ \mathbf{h}_{4,b}]. \quad (9)$$

The matrix  $\tilde{\mathbf{H}}_m$  in (8) is computed according to (5).

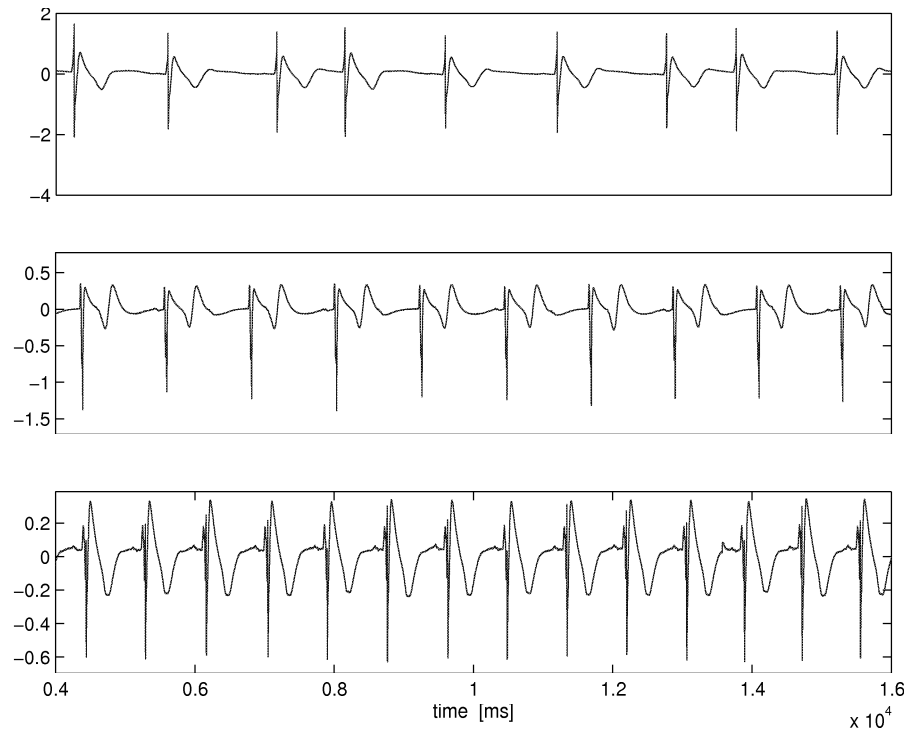
Finally, the decision signal  $T(n)$  is computed by a generalized likelihood ratio test (GLRT) as

$$T(n) = \mathbf{x}^T(n)\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{x}(n), \quad (10)$$

and compared to a threshold [12, 16]. Due to orthogonality of the mono- and biphasic functions, the matrix  $(\mathbf{H}^T\mathbf{H})^{-1}$  in (10) is symmetric and sparse with half of the elements equal to zero. Thus, half of the multiplications with the elements of the matrix in (10) do not need to be implemented. The threshold level determines the presence of an R-wave and controls the pulse generator as

$$T(n) \geq \beta T_{max}, \quad (11)$$

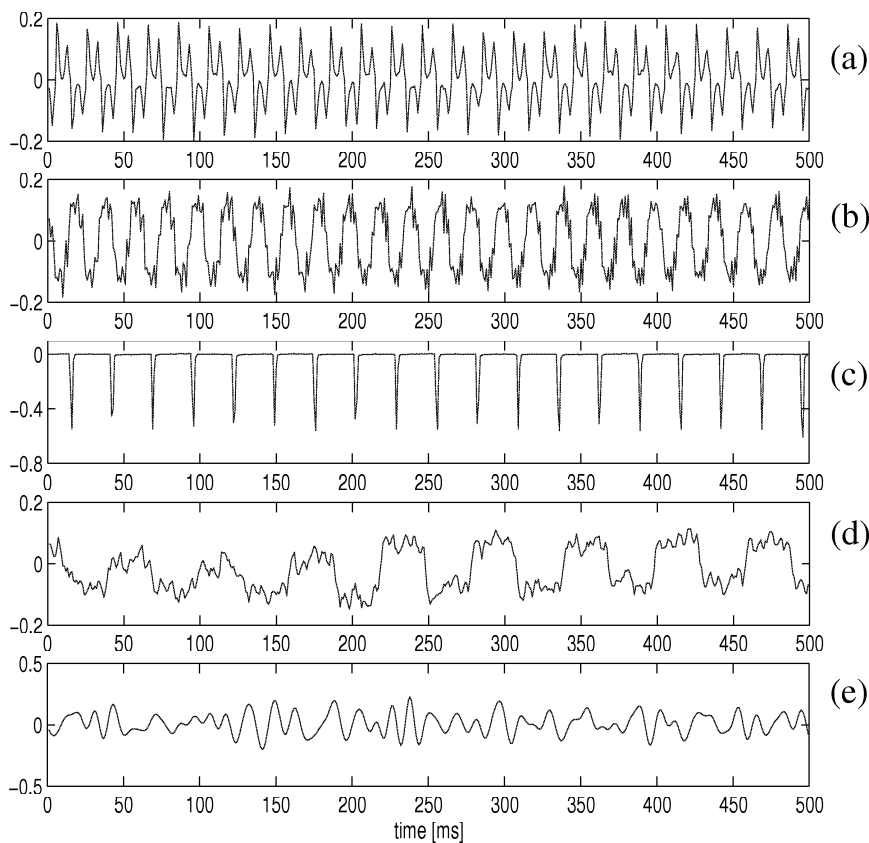
where  $\beta$  denotes a amplitude threshold fraction, and  $T_{max}$  the average value of the maximum amplitudes of the previously detected events. If the condition in (11) is met, an R-wave is detected.



**Figure 4:** Examples of EGMs from different patients.

## 2.2 The EGM Database

The database contains EGMs from 50 patients, recorded from ventricular pacemaker electrodes, and is used to evaluate the performance of the detector. The EGMs were recorded either during initial implantation or pacemaker replacement, throughout hospitals in Germany (coordinated by Justus-Liebig Universität, Gießen, Germany). The recordings were obtained from patients suffering from AV block and sick sinus syndrome [5]. Most signals were recorded from a unipolar electrode, however, a few signals were recorded with a bipolar electrode. The sampling rate was 44.1 kHz with a resolution of 16 bits. For this particular study the signals were decimated to 1 kHz, since frequencies above 400–500 Hz were judged to be less significant to detection. In order to be compliant with the ADC in [6] a resolution of 8 bits is chosen. The recordings were annotated with respect to a time reference of each R-wave, required for performance evaluation of the detection algorithm. The annotation of an event was defined as the steepest transition phase in the cardiac cycle. Three EGM

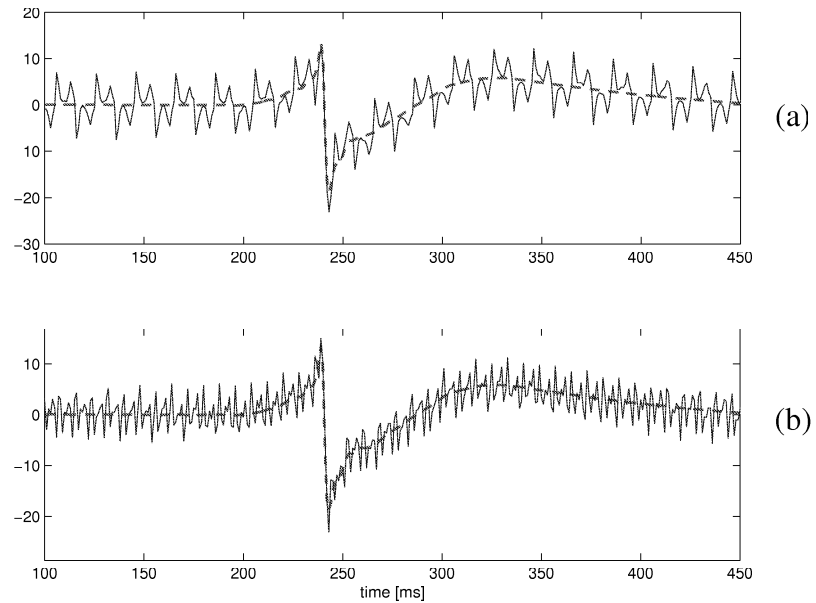


**Figure 5:** Exogenic interference recordings generated from (a) 500 W AC hand drill, (b) an electric hand mixer, (c) EAS 1, (d) EAS 2. Endogenous interference caused by (e) muscular activity.

recordings from different patients are displayed in Figure 4, and illustrates the inter-patient variability in morphology.

### 2.3 Interference Database

The present event detector is tested with respect to sensitivity to exogenic and endogenous interferences, originating outside and inside the body, respectively [28]. The test is done in order to simulate situations when the pacemaker patient is subjected to electronic or magnetic noise. Exogenic interference is limited to sources in everyday life, e.g., caused by electronic household appli-



**Figure 6:** A typical EGM interfered with 20 dB SNR from an (a) AC hand drill and (b) muscular activity.

ances or electronic article surveillance (EAS) systems. Endogenous interference is represented by muscular activity [29]. Figure 5 presents examples of the different types of interference.

*Household appliances* represent a common source of interference, caused by electric and magnetic activity within the same frequency range as the R-wave. Furthermore, the magnetic field intensity is dependent on the signal transiency. In this study, recordings from an AC powered hand drill and an electric handmixer were used.

*Electronic article surveillance systems* have been identified as a common interference source [30–33]. Such systems use widely different transmission techniques which makes it difficult to generalize results to how such systems interfere with the pacemaker. In this study two systems that operate within the R-wave frequency spectra have been tested. The EAS 1 system uses a 16.6 Hz triangular wave modulated with 5 or 7.5 kHz. The EAS 2 system transmits 3 ms long bursts of 58 kHz acoustomagnetic signals with a high amplitude at an interval of 27 ms. The pulse period of 30 ms (33 Hz) is considered to be the reason for possible interference with pacemakers.

*Muscular activity* is an endogenous interference source which spectrally over-

laps with heart signals. In this study, signals recorded pressing the palms together have been considered. The effect of muscular noise on pacemaker performance was one of the first studies on pacemaker interference [29].

Figure 6 shows the morphology of an EGM with interferences originating from a hand drill and muscular activity, respectively.

### 3 Digital Hardware Mapping and Optimization

This section describes how the wavelet filterbank and the GLRT are implemented in digital hardware. The proposed structure has been optimized with respect to wordlength and numerical strength reduction to reduce area and power consumption [34, 35].

#### 3.1 Implementation of the Wavelet Filterbank

The wavelet filterbank is comprised of scaled version of a monophasic and biphasic functions. The mother wavelets used in this implementation are scaled according to the wavelet theory described in Section 6. Low frequencies are represented by coarse scales whereas high frequencies are modeled with fine scales. The scale is increased from  $q = 2$  to  $q = 4$  in this study. This stretches the mother wavelets and is referred to as *dilation*. Consequently, propagation delay for each biphasic and monophasic output differs in every scales and needs to be adjusted before hypothesis testing.

##### 3.1.1 Dilation

In this implementation frequency components in the EGM are decomposed by (3) and (2), and the hardware realization of  $f(2n)$  and  $g_b(2n)$  is presented in this section. The narrowest filter functions  $f(n)$  and  $g_b(n)$  in the wavelet filterbank are defined as

$$f(n) = [1 \ 3 \ 3 \ 1] \quad \text{and} \quad g_b(n) = [-1 \ 1].$$

The impulse responses of these filters are dilated by inserting zeros between the coefficients which leads to

$$f(2n) = [1 \ 0 \ 3 \ 0 \ 3 \ 0 \ 1], \tag{12}$$

and

$$g_b(2n) = [-1 \ 0 \ 1]. \tag{13}$$

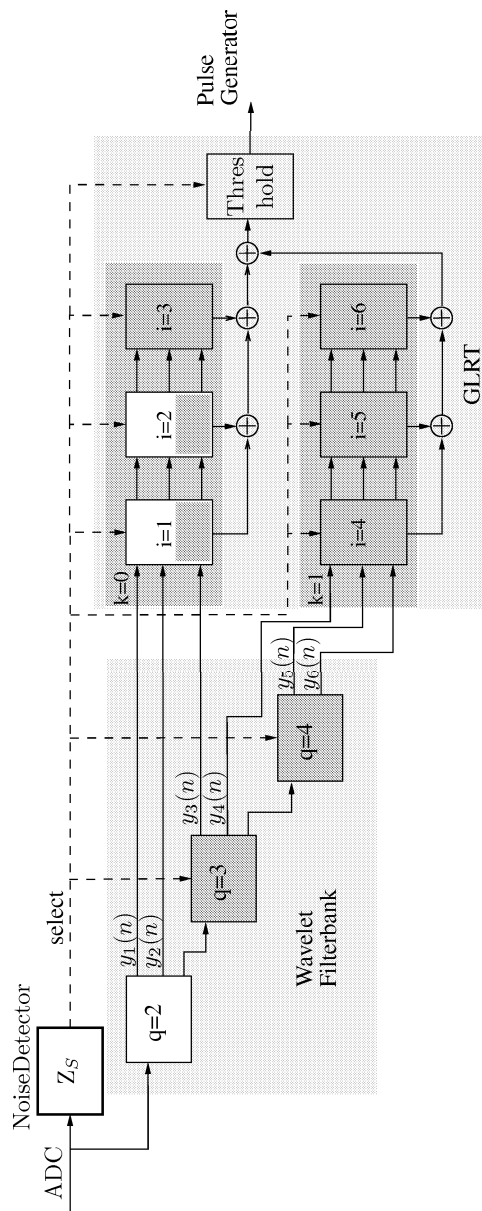
The biphasic filterbank for  $q = 2$  is realized by convoluting  $f(n)$  with  $g_b(2n)$ , as defined in (1). The discrete transfer functions of  $f(n)$  and  $g_b(2n)$  are

$$F(z) = 1 + 3z^{-1} + 3z^{-2} + z^{-3}, \tag{14}$$

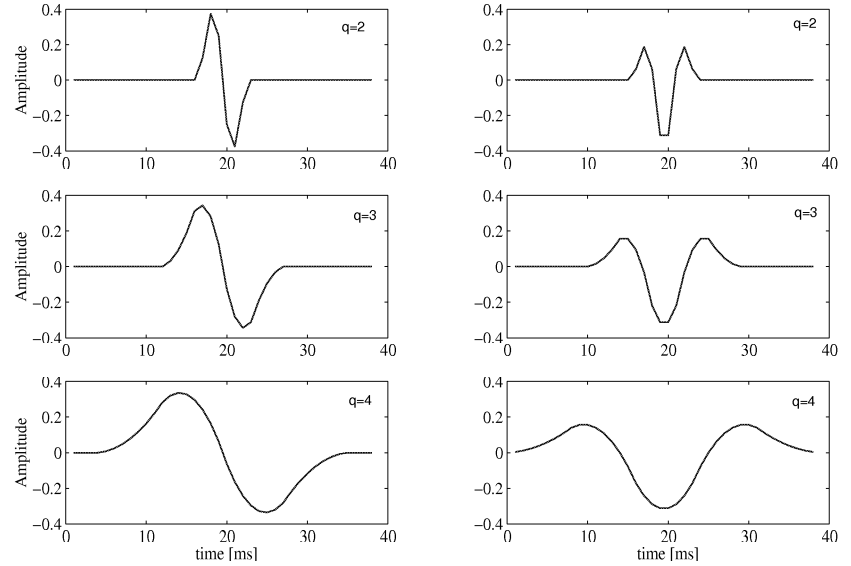
and

$$G_{2,b}(z) = -1 + z^{-2}. \tag{15}$$





**Figure 7:** Block diagram of the R-wave detector. The darker shaded blocks in the wavelet filterbank and GLRT are inactive in normal mode.



**Figure 8:** Impulse responses of the wavelet filterbank. The biphasic impulse responses  $y_{b,q}(n)$  for  $q = 2, 3, 4$  are displayed in the left panel and the monophasic impulse responses  $y_{m,k}(n)$  in the right panel.

In order to reduce hardware area the delays and multiplications in (14) are minimized by factoring out the fixed multiplication and a delay element as

$$F_1(z) = 1 + z^{-1}(3(1 + z^{-1}) + z^{-2}), \quad (16)$$

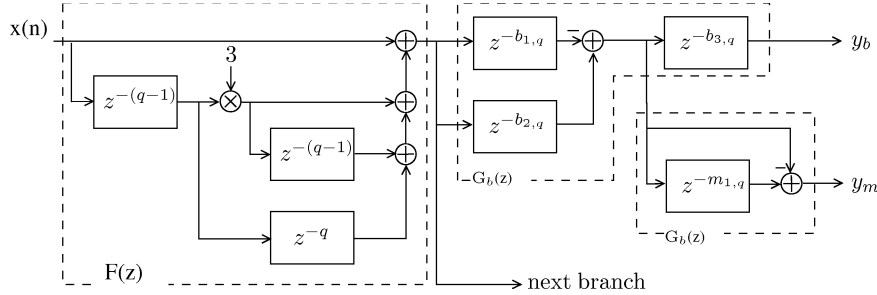
which reduces the number of delay elements and multipliers by two and one, respectively. The biphasic output  $Y_{2,b}(z)$  for  $q = 2$  is computed as

$$Y_{2,b}(z) = F_1(z) * G_{2,b}(z), \quad (17)$$

and the monophasic output is obtained by reusing  $G_{2,b}(z)$  as

$$Y_{2,b}(z) = F_1(z) * G_{2,b}(z) * G_{2,b}(z). \quad (18)$$

The branches for  $q = 3$  and  $q = 4$  are dilated and optimized in the same fashion as illustrated above. Hardware is reduced by factoring out the delays and multipliers in all branches.



**Figure 9:** Data flow diagram of the first wavelet filterbank branch using Mallat's algorithm [27]. The number of registers in  $F(z)$  is minimized, and bi- and the monophasic filter output are represented as  $y_{q,b}$  and  $y_{q,m}$ , respectively.

### 3.1.2 Time Alignment

The propagation delay of the branches differs as dilation of the impulse responses is accomplished by inserting delays. The longest impulse response is found in the branch that computes the monophasic output  $y_{4,m}$  ( $q = 4$ ). Thus, all other impulse responses need to be symmetrized in respect to  $y_{4,m}$ . This is accomplished by additional delay elements in the filterbank architecture. The biphasic and monophasic output of the first branch needs to be delayed by 16 and 15 time units, respectively. Delay elements used to postpone the biphasic impulse response are partially reused to lag the monophasic output. Thus, the number of delays is minimized, i.e., the number of delays needed to symmetrize the output of the first branch is reduced from 31 to 15. This optimization is carried out in branch two and three as well, and the total number of delays in all scales is reduced from 57 to 25. The implementation of a wavelet filterbank branch is presented in Figure 9. A summary of the additional delays needed for symmetrization is presented in Table 1, and the impulse responses of the filterbank are presented in Figure 8.

**Table 1:** Delays in  $G_b(z)$  as illustrated in Figure 9.

q	$b_{1,q}$	$ms_q$	$bs_q$
2	2	15	1
3	4	10	2
4	8	0	4

### 3.2 Implementation of the GLRT

The decision signal  $T(n)$  is computed in the GLRT as defined in (10). The input to the GLRT is computed as presented in Section 3.1. Thus, the remaining part of (10) to be implemented is the multiplication by  $(\mathbf{H}^T \mathbf{H})^{-1}$ , a matrix which is symmetric and sparse with half of its elements equal to zero,

$$(\mathbf{H}^T \mathbf{H})^{-1} = \begin{bmatrix} 4.25 & -2.81 & 0.71 & 0 & 0 & 0 \\ -2.81 & 4.47 & -1.75 & 0 & 0 & 0 \\ 0.71 & -1.75 & 1.49 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4.84 & -2.31 & 0.6 \\ 0 & 0 & 0 & -2.31 & 4.29 & -1.49 \\ 0 & 0 & 0 & 0.6 & -1.49 & 1.77 \end{bmatrix}. \quad (19)$$

The multiplication of  $\mathbf{y}(n)$  with  $(\mathbf{H}^T \mathbf{H})^{-1}$  is illustrated in Figure 10, where  $c_{i,j}$  are elements of  $(\mathbf{H}^T \mathbf{H})^{-1}$ . This representation is an aid to comprehend the hardware implementation which is depicted in Figure 11. Each elements of vector  $\mathbf{y}(n)$  is multiplied with its associated coefficient in the fixed matrix, before summing them to an element in a new vector, named  $k(n)$ . The hardware realization to the representation in Figure 10 as presented in Figure 11. The inputs  $y_1(n) \dots y_3(n)$  are fed to fixed multipliers which realize the multiplication of the first three elements in  $\mathbf{y}(n)$  with the upper part of the matrix  $(\mathbf{H}^T \mathbf{H})^{-1}$ . Each element in the input vector is multiplied with its associated elements in the matrix. This results in a hardware cost of three fixed multipliers and two adders per matrix column. Thereafter, the elements of  $\mathbf{y}(n)$  are multiplied with the column sums  $k_1(n) \dots k_6(n)$ , and the attained products are accumulated, and this sum is the decision signal  $T(n)$ . The hardware cost for this operation is six multipliers and five adders.

### 3.3 Hardware Optimization

The aim of hardware optimization is the reduction of silicon area as well as the reduction of power dissipation.

#### 3.3.1 Optimization of the Wavelet Filterbank

The wordlength of the wavelet filterbank output  $\mathbf{y}(n)$  is bit-optimized in order to reduce complexity. In a theoretically worst-case scenario mathematical operations in the filterbank leads to an extended dynamic range and therefore wordlength has to increase accordingly, in order to be sure to avoid overflow. However, this is a very pessimistic approach leading to large overhead due to excessive number of bits. In order to determine the maximum number of required bits for a more realistic scenario, all recordings in the EGM database have been analyzed using the filterbank. The internal wordlength at  $F(z)$  and

$\mathbf{y}(\mathbf{H}^T \mathbf{H})^{-1}$							$c_{11}$	$c_{12}$	$c_{13}$	0	0	0
							$c_{21}$	$c_{22}$	$c_{23}$	0	0	0
							$c_{31}$	$c_{32}$	$c_{33}$	0	0	0
							0	0	0	$c_{44}$	$c_{45}$	$c_{46}$
							0	0	0	$c_{54}$	$c_{55}$	$c_{56}$
							0	0	0	$c_{64}$	$c_{65}$	$c_{66}$
$y_1$	$y_2$	$y_3$	$y_4$	$y_5$	$y_6$		$y_1 c_{11}$	$y_1 c_{12}$	$y_1 c_{13}$	$y_4 c_{44}$	$y_4 c_{45}$	$y_4 c_{46}$
							$+y_2 c_{21}$	$+y_2 c_{22}$	$+y_2 c_{23}$	$+y_5 c_{54}$	$+y_5 c_{55}$	$+y_5 c_{56}$
							$+y_3 c_{31}$	$+y_3 c_{32}$	$+y_3 c_{33}$	$+y_6 c_{64}$	$+y_6 c_{65}$	$+y_6 c_{66}$
							$= k_1$	$= k_2$	$= k_3$	$= k_4$	$= k_5$	$= k_6$

**Figure 10:** Illustrative representation for a vector-matrix multiplication.

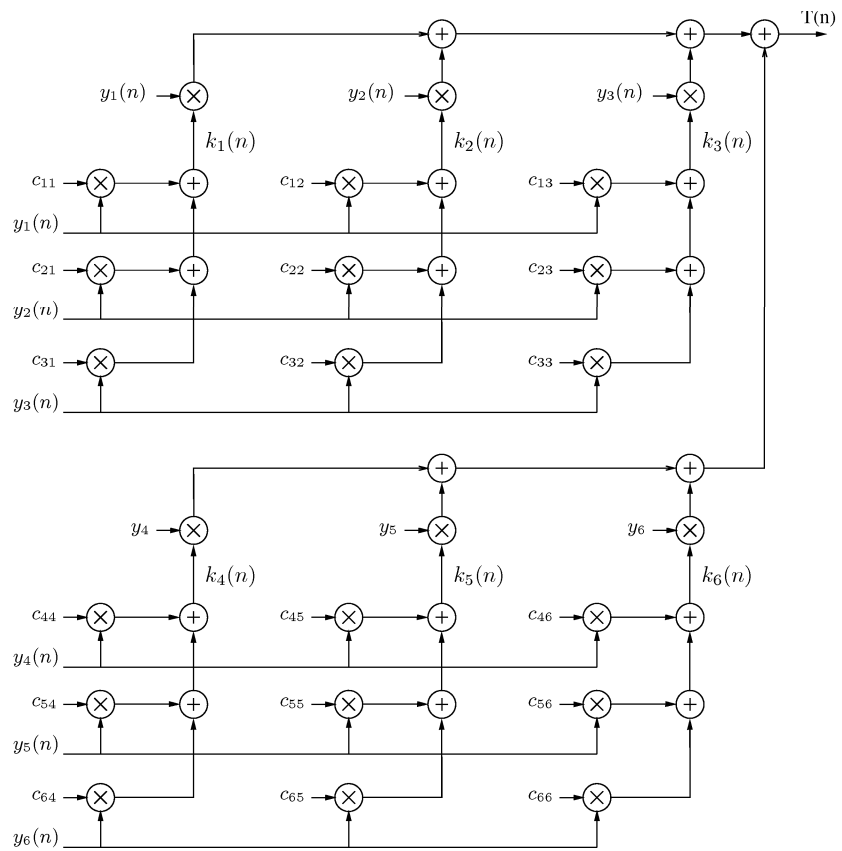
the output  $Y_1(z), \dots, Y_6(z)$  was traced in order to determine the number of bits required to represent the largest occurring number. This analysis has shown that overflow can be avoided if the dynamic range is increased by two bits compared to the input wordlength  $N$ . Thus, the wordlength,  $N_{\text{wc}}$ , for a worst-case scenario could be reduced significantly from  $N + 15$  to  $N + 2$ , see Table 2, and the target implementation has a wordlength of ten bits at  $y_1(n), \dots, y_6(n)$ . Saturation logic guarantees that the signal at  $\mathbf{y}(n)$  is upwards limited to values representable by ten bits if overflow should occur. This optimization leads to significant reductions in the filterbank and results in narrower multipliers and adders in the following GLRT.

Furthermore, the fixed multipliers in each branch of the filterbank are implemented as shift-add instructions, carried out during one clock cycle, referred to as *numerical strength reduction* [34]. This optimization results in area and power reduction.

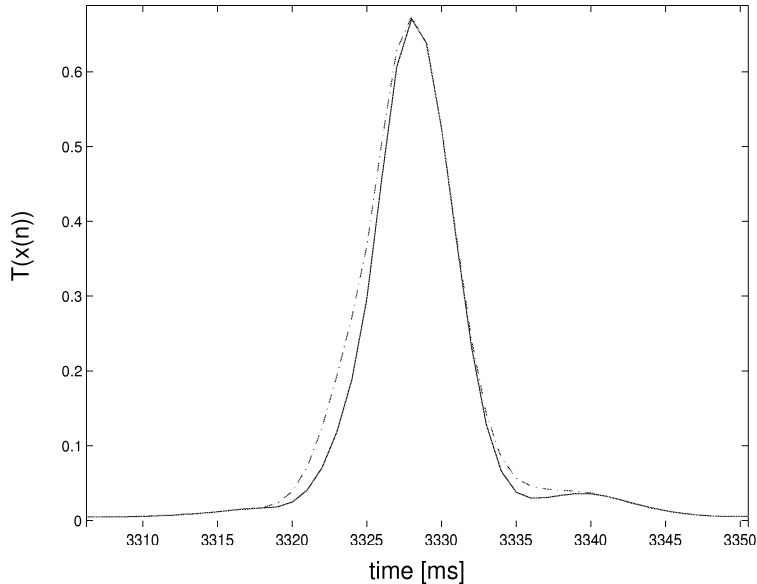
The straightforward block diagram of the filterbank in Figure 9 has an excessive number of delays. Thus, the number of registers in  $G_b(z)$  are minimized by reusing the registers needed to center the impulse responses in Figure 8. This results in a reduction of approximately 300 (1-Bit) registers.

**Table 2:** Comparison of the worst-case scenario and implemented wordlength  $N_{\text{wc}}$  and  $N_{\text{Imp}}$ , respectively, where  $N = 8$  is the width of the input signal provided from the ADC.

	$y_1(n)$	$y_2(n)$	$y_3(n)$	$y_4(n)$	$y_5(n)$	$y_6(n)$
$N_{\text{wc}}$	$N + 6$	$N + 7$	$N + 11$	$N + 12$	$N + 14$	$N + 15$
$N_{\text{Imp}}$	$N + 1$	$N + 1$	$N + 2$	$N + 1$	$N + 2$	$N + 1$



**Figure 11:** Data flow diagram of the GLRT.



**Figure 12:** Decision signal  $T(n)$  computed for floating-point values (solid line) and integer values (dashed line).

### 3.3.2 Optimization of the GLRT

Substitution of the real values in  $(\mathbf{H}^T \mathbf{H})^{-1}$  with their respective rounded integer values reduces the computational cost of the GLRT, see Table 3. As multiplication is a more complex operation than addition, complexity reduction is achieved by trading multipliers against adders [34]. Therefore, all the multiplications with the elements of  $(\mathbf{H}^T \mathbf{H})^{-1}$  are replaced with shift and add operations performing the same operation during one clock cycle. The only multiplication that remains in the GLRT is the one by  $\mathbf{H}^T \mathbf{x}(n)$ , where  $\mathbf{H}^T \mathbf{x}(n)$  is already computed in (6). The result of this operation is implemented by the multiplier inputs  $y_1(n) \dots y_6(n)$  as presented in Figure 11.

This optimization achieves a reduction of 77% in multipliers while the number of adders is increased by 41%, see Table 3. Thus, power and area is reduced in filterbank hardware.

A typical decision signal using *Matlab* floating point and integer coefficients, respectively, is presented in Figure 12. It is shown that the deviation of the GLRT output after numerical strength reduction is minor. Detection performance using the database, between optimized and original structure was compared, and remained unchanged.

**Table 3:** Summary of the strength reduction after optimization of the wavelet filterbank and the GLRT.

	not optimized	optimized	savings %
MULT	26	6	-77
ADD	32	45	+41

### 3.4 Sensing Threshold Initialization

After pacemaker implementation the acute sensing threshold is automatically computed by the proposed threshold initialization hardware [36]. This is carried out by computing a mean value of the highest amplitudes of indicated R-waves at the output of the GLRT as

$$\bar{T}_{acc} = \frac{1}{i_{max}} \sum_{i=1}^{i_{max}} T_{max,i}, \quad (20)$$

where  $i_{max}$  is the maximum number of R-waves in the initialization phase,  $T_{max,i}$  the peak amplitude of an R-wave, and  $\bar{T}_{acc}$  the mean of the R-wave amplitudes. The threshold initialization phase is triggered by resetting the device. The observation of  $T(n)$  for the highest R-wave amplitudes provides the initial amplitude  $T_{max,1}$ , used to identify the succeeding R-waves. Following R-waves are indicated by any amplitude that exceeds  $0.5T_{max,1}$ , and their highest amplitude  $T_{max,i}$  is sensed for 100 ms after crossing the threshold level. A threshold level of 50% minimizes the risk of considering noise as cardiac activity. This procedure is iterated until a predefined number of R-waves,  $i_{max}$ , is found.

The implementation of the division hardware in (20) is optimized by assigning  $i_{max}$  a number that can be represented by powers of two. Thus, the division hardware is realized by truncation of the LSBs. For this study  $i_{max}$  has been set to eight events since the signals in the EGM database are limited in time.

The hardware that computes the threshold mean is triggered by a gated clock which terminates after the initialization phase. However, in order to address the leakage power effectively, appropriate sized sleep transistors need to be placed in the supply lines. The threshold initialization hardware has an independent supply in order to carry out a power analysis.

### 3.5 Dual Operation Mode

The R-wave detector in [13] is designed to assure good detection performance when the EGM is corrupted with noise, see Figure 15. However, during long periods the pacemaker patient is not exposed to noise, e.g., during sleep (rest),



low physical activity, etc. Therefore, it is highly desirable to automatically shut off parts of the R-wave detector during such periods to save power. At the same time the R-wave detector must be able to operate with full noise suppression performance whenever necessary. Operation when the patient is not exposed to noise is referred to as normal mode as this is the case most of the time. Alert mode is when the entire filterbank and GLRT are active.

One approach is to shut off one or two branches in the wavelet filterbank during normal mode. Thus, it is also possible to shut off parts of the GLRT with respect to the inactivated branches in the filterbank. Section 4.3 presents a performance analysis of how the branches are activated or deactivated in alert and normal mode, respectively. The parts of the R-wave detector being shut off in normal mode are triggered by a gated clock [37]. This clock tree is enabled by the noise detector.

### 3.6 Time-multiplexed Architecture

An alternative technique that can be considered for leakage minimization is a time-multiplexed architecture, i.e., multiple instructions are computed by a single [34]. Such hardware transformation reduces the total gate width, equivalent to the gate count, approximately.

A possible transformation of two additions implemented in a parallel structure, carried out in the GLRT, into a time-multiplexed structure is presented in Figure 13 (a) and (b), respectively. The products  $b(b)$  and  $c(n)$  in (a) are computed simultaneously and  $k_1(n)$  is valid after one clock cycle. In (b) the first clock cycle feeds  $a(n)$  and  $b(n)$  to the adder. The addition is carried out and the sum is stored in the delay element. The next clock cycle switches  $c(n)$  and the sum stored in the delay element to the adder and the adder output is stored in the delay element. The value  $d(n)$  is valid with the next clock cycle, and, simultaneously  $a(n+1)$  and  $b(n+1)$  are subjected to the adder and the process continues. Thus, the time needed to compute  $k_1(n)$  is twice compared to the original structure. Area is traded for computation time [34]. To revive the original computation speed the clocking frequency must be doubled, and, thus, dynamic power consumption is doubled as well. However, an increase in power consumption does not necessarily increase energy dissipation, as presented in Section 4.4.1. Therefore, an analysis is carried out that determines whether an increase in clock frequency leads to higher energy dissipation.

Energy dissipation of an operation in a digital ASIC is consumed as

$$E_{switch} = C_L V_{dd}^2, \quad (21)$$

where  $C_L$  is the gate capacitance that is switched during an operation.

The dynamic energy dissipation of a single operation in (21) is proportional to  $C_L$  and quadratically dependent on  $V_{dd}$ . In a time-multiplexed architecture

additional hardware, e.g., control logic and pipeline registers, is introduced, and, therefore,  $C_L$  increases by a small fraction.

The simulated critical path in the R-wave detector has a propagation delay of 16 ns which results in a large time slack using a clocking frequency of 1000 Hz. The time slack in a design is a factor that determines how much the supply voltage can be decreased without performance penalty. However, the slack in this implementation would theoretically permit a supply voltage equal to the switching threshold. Thus, the lower limit of the supply voltage is set by the switching threshold plus an additional small voltage that prevents the circuit from malfunctioning.

The highest possible clocking frequency  $f_{max}$ , according to the propagation delay in the critical path, is 62.5 MHz. The lowest possible supply voltage is approximated as

$$V_{min} = \beta_1 + \beta_2 f_{norm}, \quad (22)$$

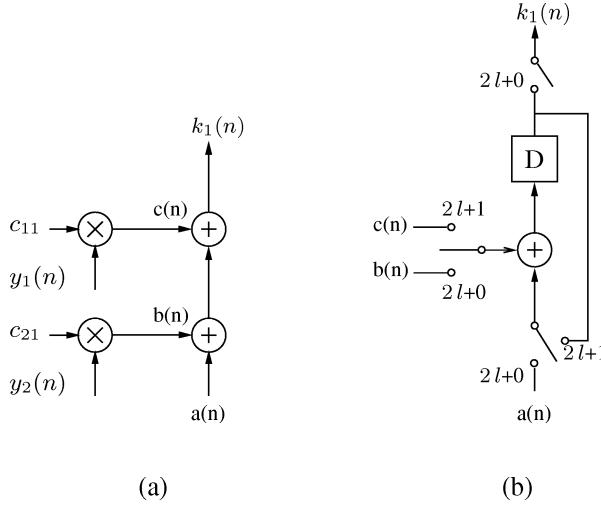
where  $\beta_1 = V_t/V_{max}$ ,  $\beta_2 = 1 - \beta_1$ , and  $f_{norm} = (f_{max}/f)^{-1}$  [38]. According to the approximation in (22), a minimized supply voltage would be equal to the threshold level using a low-leakage cell library. However, the supply voltage needs to be higher than the switching threshold. The large timing slack permits a higher clock frequency, needed to sustain computation speed in a time multiplexed architecture. A clock frequency of 2 kHz permits a propagation delay of approximately 50 ns whereas the hardware requires only 16 ns. Thus, the supply voltage remains on the same level and dynamic energy dissipation will increase by a significantly small fraction, introduced by the  $C_L$  of the extra register and control logic in a time-multiplexed architecture. However, the total gate count and thus leakage is reduced.

**Suitable Hardware in Detector** Hardware that is shut down in normal mode is suitable for a time-multiplex implementation as area reduction achieves a significant leakage minimization. In the wavelet filterbank the add operations in branch two and three can be performed by a single hardware unit. Hardware in the GLRT can be reused such that two additions and one multiplier are accommodated in a hardware unit that sequentially carries out the operations associated to  $y_3(n) \dots y_6(n)$ . Moreover, hardware in the wavelet filterbank that is shut off in normal mode is suitable for a time-multiplexed architecture as well. Remaining leakage can be further reduced by sleep transistors.

In order to get a better estimate on leakage reduction due to a time-multiplexed architecture, digital hardware needs to be fabricated.

### 3.7 Noise Detector

In order to make the R-wave detector resilient to noise, a noise detector has been supplemented, as shown in Figure 7. The noise detector operates in supervision



**Figure 13:** (a) Two additions from the GLRT in Figure 4.3.1. (b) Time multiplexed architecture where the 2 additions are carried out by a single adder and a additional pipeline stage.

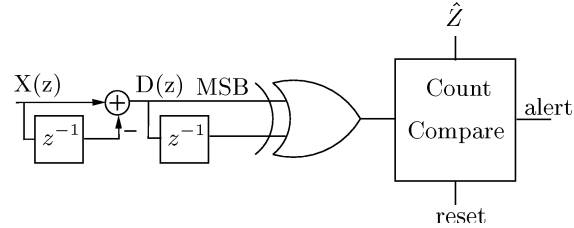
mode and guarantees full noise suppression performance by reactivating the hardware that has been shut off during normal mode. The power savings gained by deactivating parts of the R-wave detector must not be dissipated by the noise detector since the proposed modification would then lack significance. Therefore, it is necessary to design a low complexity noise detector. Noise quantification is based on a zero-crossing rate measurement  $Z_S(n)$  [39]; the number of zero crossings is the number of times a sequence changes sign. The  $Z_S(n)$  measurement on all the recordings in the database show that an typical upper bound for a patient ranges from 5 to 7 zero-crossings during 100 ms.

If the input signal has a DC component, a zero-crossing measurement cannot be carried out correctly. Therefore, any DC component of an EGM is filtered out before  $Z_S(n)$  is estimated by a differencing filter as

$$d(n) = x(n) - x(n - 1), \quad (23)$$

which has a low complexity implemented in digital hardware. The zero crossing measure in the long term, i.e., from 0 to  $\infty$ , is defined as [39]

$$Z_D = \frac{1}{2} \sum_{n=0}^{\infty} |\text{sgn}\{d(n)\} - \text{sgn}\{d(n - 1)\}|, \quad (24)$$



**Figure 14:** Zero-crossing detector logic.

where

$$\text{sgn}\{d(n)\} = \begin{cases} +1 & d(n) \geq 0 \\ -1 & d(n) < 0. \end{cases} \quad (25)$$

In order to define a short-term  $Z_S(n)$ , (24) can be used as

$$Z_S(n) = \frac{1}{2N} \sum_{m=n-N+1}^n |\text{sgn}\{d(m)\} - \text{sgn}\{d(m-1)\}|, \quad (26)$$

where  $N$  is the length of the short-term interval. Since most zero-crossings occur during an R-wave, the length of  $N$  needs to be longer than this interval, usually no longer than 100 ms [5]. However, to achieve a flexible implementation  $N$  is a programmable parameter in the target implementation.

### 3.8 Noise Detector Implementation

A zero crossing can be identified by comparing the signs of two successive samples computed in (23). Using digital hardware and two's complement representation the comparison can be carried out by analyzing the most-significant-bit (MSB) which indicates the sign of a number. A zero crossing has occurred if

$$\text{MSB}\{d(n)\} \oplus \text{MSB}\{d(n-1)\} = 1,$$

where  $\oplus$  is the XOR function. The number of zero-crossings, indicated by a low to high transition at the XOR gate, is accumulated for the time  $N$ , see Figure 14. Noise is detected when  $Z_S(n)$  exceeds an upper bound  $\hat{Z}_S$ , causing the R-wave detector to switch to alert mode.  $\hat{Z}_S$  differs for every patient and needs to be programmed during pacemaker surgery or check-up. The  $Z_S$ -accumulator is reset after  $N$  samples are processed to start a new  $Z_S$  determination for the next input sequence. Thus, a counter that provides a reset signal after  $N$  samples is needed in addition to the schematic in Figure 14. The  $Z_S$ -accumulator is implemented by a register and an adder. The simple noise detector structure results in very little area and power overhead when implemented in digital hardware. More sophisticated alternatives can be considered but would result in higher complexity.

## 4 Detection Performance

The performance of the implemented detector is analyzed by adding various interferences to the EGM recordings such that different SNRs are obtained. Detection performance is measured by computing the probability of missed detection,  $P_D$  and false alarms,  $P_{FA}$  as

$$P_D = \frac{N_T}{N_T + N_M} \quad \text{and} \quad P_{FA} = \frac{N_{FA}}{N_T + N_{FA}}, \quad (27)$$

where  $N_T$  is the number of true detections,  $N_M$  the number of missed detections, and  $N_{FA}$  the number of false alarms. A true detection is defined as an event that occurs within 50 ms of the annotation, whereas events outside this interval are declared as false alarms.

### 4.1 Signal-to-Noise Ratio Definition

The analyzed signal consists of the nonstationary EGM,  $x(n)$ , to which a noise signal,  $v(n)$ , has been added. The signal-to-noise ratio (SNR) of  $y(n)$  is defined as

$$SNR = 20 \cdot \log \frac{V_x}{\sigma_V}, \quad (28)$$

where  $V_x$  is the average peak-to-peak amplitude of all the R-waves in one EGM recording and  $\sigma_V$  the standard deviation of the noise to be added. The peak-to-peak amplitude  $V_x$  is calculated according to

$$V_x = \frac{1}{N_x} \sum_{i=1}^{N_x} \left| \max_{-50 \leq m \leq 20} \{x(R_i + m)\} \right| + \left| \min_{-50 \leq m \leq 20} \{x(R_i + m)\} \right|, \quad (29)$$

where  $x(R_i)$  is a vector containing an R-wave positioned at  $R_i$ , and  $N_x$  is the number of R-wave templates. The standard deviation  $\sigma_V$  is calculated according to

$$\sigma_V = \frac{1}{N_v} \sqrt{\left( \sum_{i=1}^{N_v} v(i) - \bar{v} \right)^2}, \quad (30)$$

where  $\bar{v}$  is the mean of the noise signal and  $N_v$  the number of discrete samples.

## 4.2 Detection Performance for Noisy EGMs

Although the pacemaker patient is mostly exposed to low-noise environment, the importance of handling heavily disturbed EGMs must nonetheless be addressed. The detection performance for the circumstance when the pacemaker patient is exposed to various interferences is analyzed in this section. The EGMs are disturbed with recordings of the interference database, see Section 2.2 and 2.3.

Shape and body constitution of humans vary considerably and, therefore, it is not possible to find one single estimate of how much noise can interfere with the pacemaker. However, a SNR of 20 dB corresponds to a very high interference level which should include the worst-case situation in real life. Thus, noise levels are chosen which results in SNRs of 20 and 25 dB to assure that this situation can be handled by the detector. A typical EGM that is disturbed by interference and the corresponding decision signal  $T(n)$  is shown in Figure 15.

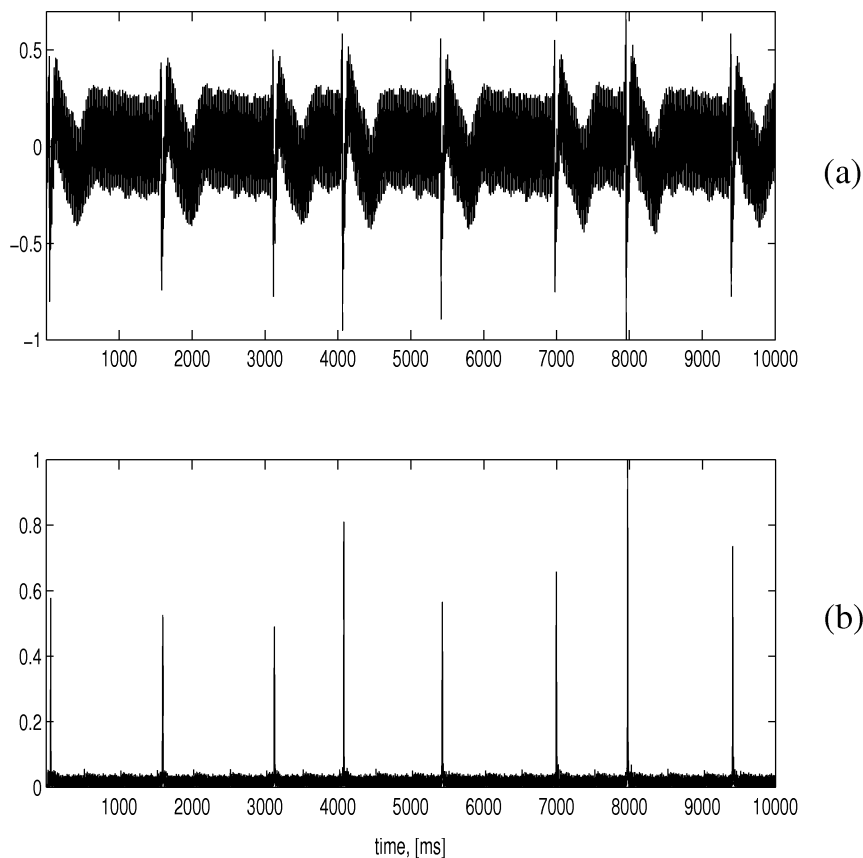
The threshold level  $\beta$  is varied from 0.3 to 0.5. A low value for  $\beta$  produces high rates of  $P_D$ , however,  $P_{FA}$  will also increase as more false events will exceed the threshold. A high  $\beta$  leads to a lower value for  $P_D$  and  $P_{FA}$ .

It can be observed that the detector is more sensitive to interference that originates from the EAS2 system and muscular activity than from other sources, see Figure 16. This is the case for  $P_D$  and  $P_{FA}$  for both noise levels. Interference that originates from muscle contractions is the most difficult noise to suppress in all the tested EGMs [29]. The results show that the detector attains reliable detection performance at moderate to low SNRs. For 20 dB and 25 dB SNR the average performance for all noise sources is  $P_D = 0.88$  and  $P_{FA} = 0.13$  and  $P_D = 0.98$  and  $P_{FA} = 0.014$ , respectively. Moreover, a threshold level that strikes a good balance between  $P_D$  and  $P_{FA}$  is  $\beta = 0.4$ .

## 4.3 Detection Performance for Normal Mode

The pacemaker is mostly operating in a low-noise environment and, therefore, a performance analysis is carried out where no additional noise is added to the EGM. The threshold level  $\beta$  is 0.4, see Table 4. The total number of 3200 events for all the recordings in the EGM database has been analyzed.

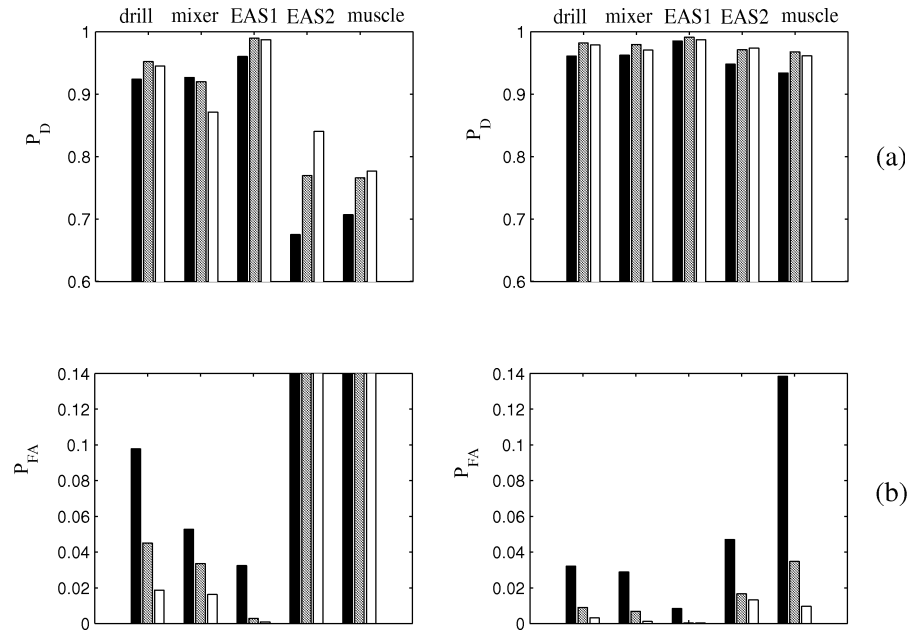
In order to find out which branches of the filterbank can be shut off without performance degradation for all recordings in the EGM database, the branches have been activated in different combinations. The mode of a single branch, active or inactive, binary coded. If a branch is active it is coded as 1 in a 3 digit word, e.g., 110 indicates that branches one and two ( $q = 2, 3$ ) are active while three ( $q = 4$ ) is inactive. However, as the branches are connected in series it is not possible the entirely shut off the first branch if the second or third branches is active, see Figure 7. For such cases, only the filtering part



**Figure 15:** (a) An EGM recording distorted by an AC hand drill (20 dB SNR). (b) The output  $T(n)$  of the GLRT.

$G_b(z)$  of the preceding branches are shut off, see Figure 9.

The average detection and false alarm rate,  $P_D$  and  $P_{FA}$ , respectively, is presented in Table 4. In none of the simulated cases, the detection performance drops below 0.97 whereas the highest  $P_{FA}$  rate is less than 0.021, for a threshold level  $\beta = 0.4$ . The highest detection and lowest false alarm rates,  $P_D$  and  $P_{FA}$ , are obtained if all branches, i.e.  $111$ , in the filterbank are operating,  $> 0.99$  and  $< 0.001$ , respectively. However, the difference between  $111$  and  $100$  is negligible as only a minor difference for  $P_{FA}$  is measurable. Furthermore, it is possible to shut off block three to six in the GLRT, whereas block one and two operate partially, see Figure 7. As the first branch has to operate partially in all the combinations it is of further advantage to inactivate filterbank two and



**Figure 16:** Means of detection performance in terms of  $P_D$  and  $P_{FA}$ . The recordings from the EGM database are disturbed with interferences originating from a hand drill, mixer, EAS1, EAS2 and muscle contraction. The applied noise results in SNRs of (a) 20 dB, and (b) 25 dB. The threshold level  $\beta$  is set to following values: 0.3 (black bars), 0.4 (grey bars), 0.5 (white bars).

**Table 4:** Detection performance for the noiseless case. The mode column indicates which branch operates,  $\beta = 0.4$ .

Mode	$P_D$	$P_{FA}$
001	0.973	0.021
010	0.997	0.001
011	0.972	0.020
100	0.997	0.001
101	0.994	0.021
110	0.997	0.001
111	0.997	< 0.001



**Table 5:** Comparison of detection performance for noisy EGMs, 25 dB SNR. The R-wave detector operates in alert mode and forced normal mode,  $\beta = 0.4$ .

mode	normal		alert	
	$P_D$	$P_{FA}$	$P_D$	$P_{FA}$
-	0.997	0.001	0.997	< 0.001
drill	0.926	0.083	0.982	0.009
mixer	0.983	0.019	0.980	0.007
EAS1	0.975	0.024	0.991	0.003
EAS2	0.513	0.082	0.970	0.016
muscle	0.960	0.083	0.970	0.034

three since the amount of hardware that can be shut off is higher compared to other combinations.

In the target implementation not all presented modes will be implemented. Only two modes will be considered, *100* for the normal mode and *111* for the alert mode.

**Noisy Signal in Normal Mode** The noise detector reactivates branch two and three if the EGM is corrupted by noise, and, thereby, sustains filtering performance. Nevertheless, it is of interest to analyze detection performance for the circumstance that noise is present but not indicated by the noise detector. To evaluate detection performance for such a circumstance, noise is added (25 dB SNR) to the recordings in the EGM database. The situation that no noise is detected is simulated by a forced normal operation mode, i.e., branch two and three are permanently off. Detection performance is analyzed by computing  $P_D$  and  $P_{FA}$  for  $\beta = 0.4$  for all the recordings in the database, see Tab. 5. It can be seen that performance degrades if the EGM is disturbed and the R-wave detector continues the normal mode operation. The  $P_D$  rate for EAS2 drops to approximately 0.52 which is unacceptable. Thus, it is necessary to switch to alert mode and to reactivate branch two and three to sustain reliable performance.

## 5 Power Consumption

The power consumption of a digital ASIC is defined as

$$P = P_{dyn} + P_{dp} + P_{leak}, \quad (31)$$

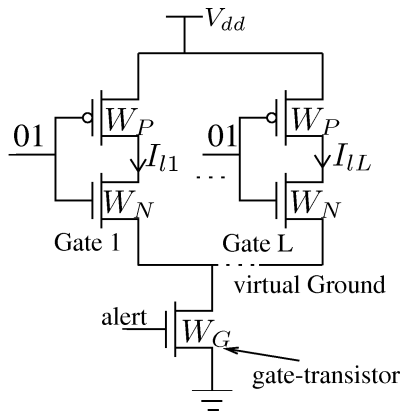
where  $P_{dyn}$  is the switching power,  $P_{dp}$  the direct-path power, and  $P_{leak}$  the leakage power [40]. For a long time, dynamic power consumption has been the dominant source whereas leakage power has been ignored for most applications. However, with shrinking technology and decreasing threshold voltage  $V_t$ ,  $P_{leak}$  represents a substantial or dominant share of the total power. Leakage power is consumed as long as the supply voltage is switched on, regardless of the switching activity. For the presented design leakage power is the main contributor due to the low clock frequency and correspondingly low switching activity.

### 5.1 Gated Supply Lines

The R-wave detector is operating at a low clock frequency of 1 kHz, implying that leakage power will have a large share of the total power consumption. To reduce the leakage current effectively three approaches can be applied: a multithreshold CMOS (MTCMOS) process, transistor stacking or the combination of transistor stacking and MTCMOS [19], [17,41]. Using MTCMOS the design is implemented using high  $V_t$  devices for the noncritical and low- $V_t$  devices for the critical path. Transistor stacking, on the other hand, cuts off one of the supply rails and thereby reduces the leakage power. This can be implemented by using a standard CMOS process without dual  $V_t$ . Combining the two techniques by using high- $V_t$  transistors for transistor stacking will lead to a higher leakage reduction but requires extra process steps with corresponding costs [19].

The presented design has been implemented in a UMC low-leakage process that provides high- $V_t$  devices. Furthermore, transistor stacking is applied to achieve substantial leakage power reduction. An extra gate transistor is introduced in the leakage path and can either be placed between the power supply,  $V_{dd}$ , and the cells or between ground, GND, and the cells, see Figure 17 [40]. The gate-transistor is turned on and off in the alert and normal mode, respectively, and thus the cell supply voltage is gated which achieves significant leakage reduction. An extra sleep transistor reduces leakage current by orders of ten [40] and the gate-transistor can be shared among multiple cells which amortizes the area overhead of an extra transistor.

**Transistor Sizing** The gate-transistor must be large enough to sink the current flowing through the cells during alert mode. However, a too large transistor degrades the stacking effect and introduces an area overhead. The dimensions of a transistor that match the needs to gate a single cell can be



**Figure 17:** Schematic of the gate-transistor circuit simulation;  
 $L_P/W_P = 0.13/100$ ,  $L_N/W_N = 0.13/33$ ,  $L_{Gate}/W_{gate} = 0.13/33$ .

determined analytically. Unfortunately, this is not the case if several cells with different properties are driven by one gate-transistor, which is the case in the proposed design. Therefore, *Spice* level simulations are carried out to find out how leakage current is influenced by the chosen size of the gate-transistor, see Figure 17. The static resistance of a transistor can be computed as

$$R_{\{P,N\}} = \frac{L_G}{W k'_p |V_{GS} - V_t - V_{DS}|}, \quad (32)$$

where  $L_G$  is the channel length,  $W$  is the channel width and  $k'_p$  the process transconductance, which for the provided cell library results in  $R_{\{P,N\}} < 1\Omega$  for a  $33\mu m$  wide NMOS and a  $69\mu m$  wide PMOS transistor [40]. Thus  $R_{\{P,N\}}$  is negligible compared to the wire resistance for power routing. The width of the gate-transistor in Figure 17 is set to  $33\mu m$  and loaded with an increasing number of balanced inverters with the transistor dimension of  $L_P/W_P = 0.13/100$  and  $L_N/W_N = 0.13/33$ . The dimensions of the inverter are restricted to the maximum dimensions allowed by the simulation tool.

## 5.2 Leakage Reduction Estimation

The number of equivalent inverter cells with respect to the R-wave detector hardware that is shut off needs to be determined. Therefore, the number of cells to shut off in normal mode and their corresponding leakage power are identified. A brief overview of these cells is presented in Table 6. This overview represents (88%) of the hardware being shut off in normal mode, and is considered to

**Table 6:** Brief summary of the cells that are shut off in normal mode. The summary represents approximately 88 % of the implemented hardware.

	Number	$I_{leak}$	INV equivalent
D-latch	395	46	1211
NAND	365	11	277
NOR	329	11	241
FA	633	43	1815
XOR	61	11	45
AND2	91	18	109
INV	791	15	791
total	2665	-	4489

be sufficient to determine the sleeping hardware leakage in this simulation. In order to be consistent with the schematic in Figure 17, the inverter leakage power of the cells in Table 6 is interpolated to the leakage power using only inverters, e.g., an AND2 cell has 20 % higher leakage than an INV cell, thus 91 AND2 cells are approximated by 109 INV cells. Thus, an equivalent leakage power to the shut down part of the detector would result in  $\sim 4500$  minimum sized inverter cells. The maximum dimension for a transistor is restricted by the simulation tool, and is 625 times the minimum size, which is indicated as load factor, (L). With the dimensions of the inverter cell in the schematic in Figure 17 a load of at least 7 can be assumed.

In order to find an estimate for leakage reduction, the schematic in Figure 17 is used, since gate-transistors cannot be included in a gate level power-simulation. The width of the PMOS transistor is thrice that of the NMOS width to balance the inverter cell. The leakage current  $I_l$  is measured for having either NMOS/PMOS of the inverter gate in on or off mode, with equal likelihood. The average of these measurements is presented in Table 7, where  $I_{gl1}$  and  $I_{l1}$  is the gated leakage current and the non-gated leakage current, respectively. Simulation on *Spice* level results show that the leakage reduction rate increases with the number of inverters driven by the gate-transistor. As presented in Table 7, gated-ground is more suitable if a large number of cells will be shut-off, and the area overhead for a NMOS gate-transistor is 52 % smaller compared to the PMOS solution. Thus gated ground is chosen for this silicon implementation. The expected leakage reduction is approximately 96 % according to the simulation results presented in Table 7.

**Table 7:** Leakage current reduction (red.) using gated  $V_{dd}$  and gated GND. The load is the number of cellblocks that are supplied by the gate-transistor.  $W_P = 100 \mu m$ ,  $W_N = 33 \mu m$ ,  $L_P = L_N = 0.12 \mu m$ ,  $W_G = 33 \mu m$ , L: # inverters in Figure 17.

L	PMOS			NMOS		
	$I_{gl1}$ [nA]	$I_{l1}$ [nA]	Red.	$I_{gl1}$ [nA]	$I_{l1}$ [nA]	Red.
1	46.4	317	85 %	50.1	317	84 %
2	27.6		91 %	32		89 %
5	18.4		94 %	15		95 %
10	12.5		96 %	7.9		97 %

### 5.3 Gate Level Power Estimation

In order to confirm the expected power savings, the power dissipation of the two modes is estimated on gate level [42]. Since sleep-transistors cannot be included in a gate level simulation, results of the leakage reduction estimation in Table 7 are used. The power estimation in normal mode is carried out by removing the hardware that is shut off from the netlist, i.e., the hardware neither leaks nor switches. However, leakage is included in the estimate by adding a fraction, according to Table 7, of the power difference between normal and alert mode and a reduction of 97 % is considered. The power analysis confirms that leakage power is the dominant power source, using the provided technology and a clock frequency of 1 kHz, see Table 8. Using solely a gated clock would not result in equal power savings since this only effects the dynamic power consumption and leakage current still exists as long as the cells are connected to the supplies. Thus, the power saving would only be 29 % if no gated-supply is used. The presented estimates only serve for identification of the dissipation sources and to provide a relative measurement of the power savings. A more accurate estimate will be available on the fabricated ASIC level. Due to a large time slack in the critical path it is possible to gain further power savings by lowering  $V_{dd}$ , as a low supply voltage lowers the barrier height and thereby decreases the threshold voltage [43–45].

However, such estimates cannot be obtained in the current gate-level simulation. The ASIC core is expected to operate in the 100 nW region in the target implementation.

**Table 8:** Core power estimation at gate level using a  $0.13\ \mu\text{m}$  low-leakage library.  $f=1\ \text{kHz}$ ,  $V_{dd} = 1.2$ .

	Alert [nW]	Normal [nW]	Reduction
dynamic	24.6	6.8	72 %
short-circuit	22.1	6.1	72 %
leakage	67.8	< 25	> 63 %
total	114	< 38	> 67 %

## 5.4 Placement and Routing

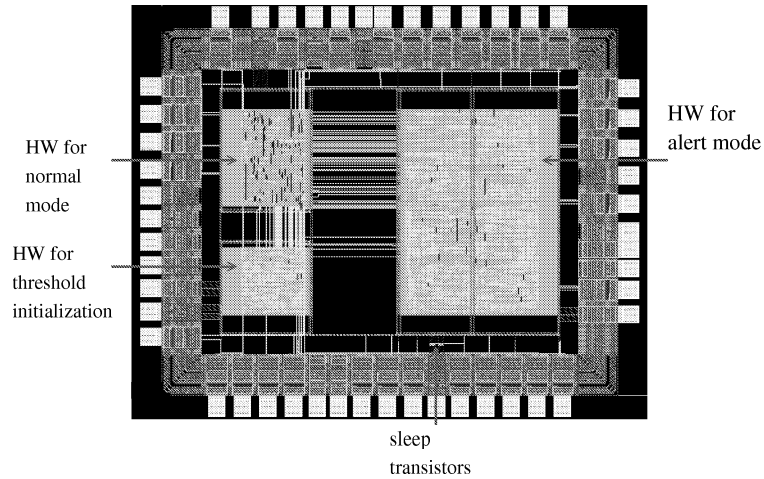
The ASIC core consists of three blocks: the hardware used in normal and alert mode, respectively, and the hardware of the threshold initialization are accommodated in separate blocks. All blocks share a common  $V_{dd}$ , however, GND is connected individually, as presented in Figure 18. Individual GND supply is used in the verification process to measure the current of each hardware block. Sleep transistor that gate GND are introduced between the GND pad and the core ring for the logic that is powered down in normal mode. The transistor gates are manually connected to the alert signal that originates in the noise detector, and, GND supply is automatically gate on account of the sleep transistors.

The total chip area, inclusive pads, is  $2.8\ \text{mm}^2$  using a  $0.13\ \mu\text{m}$  LL UMC process, see Figure 18. The ASIC has been sent for fabrication and further power analysis will be carried on the fabricated ASIC.

The design is pad-limited as various control signals are fed to the IOs for verification purpose. These control signals are not needed in the target implementation and therefore it is possible to reduce the chip size significantly. Moreover, the sleep-transistors can be accommodated in the supply pads which results in further area reduction.

## 6 Conclusions

The implementation of a wavelet based R-wave detector in  $0.13\ \mu\text{m}$  low-leakage UMC technology has been presented in this paper. The design has been power optimized by applying strength reduction, as well as wordlength and register minimization. The inclusion of a noise detector facilitates a dual operation mode. Thus, 2/3 of the hardware can be shut down if the pacemaker patient is at rest or not exposed to interferences (normal mode). Reliable detection performance is sustained by reactivating the sleeping hardware whenever necessary.



**Figure 18:** Layout of the routed ASIC.

Gate transistors in the GND supply lines are used to address the dominating leakage power and to effectively reduce the power consumption when operating in normal mode. Gate-level power estimation predicts 67% power savings if operating in normal mode; no performance degradation is measurable for such cases. The total chip area is  $2.8 \text{ mm}^2$  and is expected to operate in the sub- $\mu\text{W}$  range.

The ASIC has been fabricated in  $0.13 \mu\text{m}$  UMC low-leakage technology and further performance measures as well as power analysis will be carried out.

## Acknowledgment

The authors are grateful to Dr. Magnus Åström for contributing his expertise and to St. Jude Medical AB, Järfälla, Sweden for providing the data for this study.





## References

- [1] “St-Jude Medical,” <http://www.sjm.com>.
- [2] G. Friesen, T. Jannett, M. Jadallah, S. Yates, S. Quint, and H. Nagle, “A comparison of the noise sensitivity of nine QRS detection algorithms,” *IEEE Trans. Biomed. Eng.*, pp. 85–98, 1990.
- [3] O. Pahlm and L. Sörnmo, “Software QRS detection in ambulatory monitoring — a review,” *Med. Biol. Eng. Comput.*, vol. 22, pp. 289–297, 1984.
- [4] B.-U. Köhler, C. Hennig, and R. Orglmeister, “The principles of QRS detection,” *IEEE Eng. Med. Biol. Mag.*, pp. 42–57, 2002.
- [5] J. Webster, *Design of cardiac pacemakers*. New York, USA: IEEE Press, 1995.
- [6] A. Gerosa and A. Neviani, “A very low-power 8-bit  $\Sigma\Delta$  converter in a  $0.8\mu\text{m}$  CMOS technology for the sensing chain of a cardiac pacemaker, operating down to 1.8V,” in *Proc. 2003 IEEE Intl. Symp. Circuits Systems*, 2003.
- [7] S. A. P. Haddad, N. Verwaal, R. Houben, and W. A. Serdijn, “Optimized dynamic translinear implementation of the gaussian wavelet transform,” in *Proc. 2004 IEEE Intl. Symp. on Circuits and Systems*, 2004.
- [8] J. Jenkins and S. Caswell, “Detection algorithms in implantable cardioverter defibrillators,” *Proc. IEEE*, vol. 84, no. 3, pp. 428–445, March 1996.
- [9] R. Coggins and M. Jabri, “A low-complexity intracardiac electrogram compression algorithm,” *IEEE Trans. Biomed. Eng.*, vol. 46, no. 1, pp. 82–91, January 1999.
- [10] A. Auricchio, W. Hartung, C. Geller, and H. Klein, “Clinical relevance of stored electrograms for implantable cardioverter-defibrillator (ICD) troubleshooting and understanding of mechanisms for ventricular tachyarrhythmias,” *Am. J. Cardiol.*, vol. 78, pp. 33–41, September 1996.
- [11] B. Nowak, “Taking advantage of sophisticated pacemaker diagnosis,” *Am. J. Cardiol.*, vol. 83, pp. 172–179, March 1999.
- [12] S. Kay, *Fundamentals of statistical signal processing : detection theory*, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall, 1998, vol. II.
- [13] J. N. Rodrigues, V. Öwall, and L. Sörnmo, “A wavelet based R-wave detector for cardiac pacemakers in  $0.35\mu\text{m}$  CMOS technology,” in *Proc. 2004 IEEE Intl. Symp. on Circuits and Systems*, 2004.

- 
- [14] —, “A flexible wavelet filter structure for cardiac pacemakers: a power efficient implementation,” in *Proc. 2004 IEEE Intl. Symp. on Biomed. Circuits and Systems*, 2004.
- [15] J. N. Rodrigues, T. Olsson, L. Sörnmo, and V. Öwall, “A dual-mode wavelet based R-Wave detector using single- $V_t$  for leakage reduction,” in *Proc. 2005 IEEE Intl. Symp. on Circuits and Systems*, 2005.
- [16] M. Åström, S. Olmos, and L. Sörnmo, “Wavelet-based event detection in implantable cardiac rhythm management devices,” *IEEE Trans. Biomed. Eng.*, p. accepted for publication, 2005.
- [17] A. Agarwal, H. Li, and K. Roy, “A single- $V_t$  low-leakage gated-ground cache for deep submicron,” *IEEE J. Solid-State Circuits*, pp. 319–328, 2003.
- [18] S. Mukhopadhyay, C. Neau, R. Cakici, A. Agarwal, C. Kim, and K. Roy, “Gate leakage reduction for scaled devices using transistor stacking,” *IEEE Trans. on VLSI Systems*, vol. 11, pp. 716–730, 2003.
- [19] S. Narendra, V. De, S. Borkar, D. Antoniadis, and A. Chandrakasan, “Full-chip subthreshold leakage power prediction and reduction techniques for sub-0.18 $\mu\text{m}$  cmos,” *IEEE J. Solid-State Circuits*, vol. 39, pp. 501–510, 2004.
- [20] C. Long and L. He, “Distributed sleep transistor network for power reduction,” *IEEE Trans. on VLSI Systems*, vol. 12, pp. 937–946, 2004.
- [21] E. Berbari, *Encycl. Electric. Elec. Eng.: Electrocardiography*. New York: Wiley, 1999.
- [22] E. Berbari, J. Dyer, P. Lander, and D. Geselowitz, “Simulation of intracardiac electrograms with a moving dipole source. role of electrode geometry and high-pass filtering,” *Journ. of Electrocard.*, vol. 27, pp. 146–150, 1994.
- [23] W. Irnich, “Intracardiac electrograms and sensing test signals: Electrophysiological, physical and technical considerations,” *PACE*, vol. 8, pp. 870–888, November/December 1985.
- [24] G. Myers, Y. Kresh, and V. Parsonnet, “Characteristics of intercardiac electrograms,” *PACE*, vol. 1, pp. 90–103, January-April 1978.
- [25] W. Irnich, “Interference in pacemakers,” *PACE*, vol. 7, pp. 1021–1048, November/December 1984.
- [26] M. Åström, “Detection and classification in electrocardiac signals,” Ph.D. dissertation, Lund University, May 2003.

- 
- [27] S. Mallat, *A wavelet tour of signal processing*. San Diego, CA, USA: Academic Press, 1998.
- [28] B. Moberg and H. Strandberg, "Effects of interference on pacemakers," *Eur. J. C. P.E*, vol. 5, pp. 146–157, 1995.
- [29] W. Irrnich, "Muscle noise and interference behaviour in pacemakers: A comparative study," *PACE*, vol. 10, pp. 125–132, 1987.
- [30] B. Dodinot, J.-P. Godenir, and A. Costa, "Electronic article surveillance: A possible danger for pacemaker patients," *PACE*, vol. 16, pp. 46–53, 1993.
- [31] E. Lucas, "The effect of electronic article surveillance systems on permanent cardiac pacemakers," *PACE*, vol. 17, pp. 2021–2026, 1994.
- [32] M. McIvor, J. Reddinger, E. Floden, and R. Sheppard, "Study of pacemaker and implantable cardioverter defibrillator triggering by electronic article surveillance devices," *PACE*, vol. 21, pp. 1847–1861, 1998.
- [33] J. Mugica, L. Henry, and H. Podeur, "Study of interactions between permanent pacemakers and electronic antitheft surveillance systems," *PACE*, vol. 23, pp. 333–337, 2000.
- [34] K. Parhi, *VLSI Digital Signal Processing*. New York: Wiley, 1999.
- [35] B. Parhami, *Computer Arithmetic*. New York: Oxford University Press, 2000.
- [36] H. W. Moses, B. D. Miller, K. P. Moulton, and J. A. Schneider, *A Practical Guide to Cardiac Pacing*. Philadelphia: Lippincott Williams, 2000.
- [37] K. Nazifi and G. Hansson, "Industry's first RTL power optimization feature significantly improves power compiler's quality of results," [www.synopsys.com/news/pubs/rsvp/spr98/rsvp\\_spr98\\_6.html](http://www.synopsys.com/news/pubs/rsvp/spr98/rsvp_spr98_6.html), 1998.
- [38] N. S. Kim and T. Austin, et al., "Leakage current: Moore's law meets static power," *Computer*, vol. 36, pp. 68–75, December 2002.
- [39] J. Deller, J. Hansen, and J. Proakis, *Discrete-time Processing of Speech Signals*. New York: Wiley, 2000.
- [40] J. M. Rabaey, A. Chandrakasan, and B. Nikolić, *Digital Integrated Circuits*. New Jersey: Prentice Hall, 2003.
- [41] Y. Ye, S. Borkar, and V. DE, "A new technique for standby leakage reduction in high-performance circuits," in *Symp. on VLSI Circuits Digest of Tech. Papers 1998*, 1998.

- 
- [42] W. Qifa, T. Yujing, and X. Wei, “Oki techno centre design team achieves lowest power consumption using power compiler,” [www.synopsys.com/news/pubs/compiler/art1lead\\_oki-dec02.html](http://www.synopsys.com/news/pubs/compiler/art1lead_oki-dec02.html).
  - [43] R. Troutman, “VLSI limitations from drain-induced barrier lowering,” *IEEE Trans. Electron Devices*, pp. 461–468, 1979.
  - [44] J. Pimbley and J. D. Meindl, “MOSFET scaling limits determined by subthreshold conduction,” *IEEE Trans. Electron Devices*, pp. 1711–1721, 1989.
  - [45] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand, “Leakage Current Mechanisms and Leakage Reduction Techniques in Deep-Submicrometer CMOS Circuits,” *Proc. of the IEEE*, vol. 91, no. 2, pp. 305–327, 2003.



