UCLouvain
Institute for Information
and Communication Technologies,
Electronics and Applied Mathematics

Cross-Layer and Cross-Domain Power Optimization Towards Sustainable Biomedical CMOS Sensor Systems

Rémi Dekimpe

Thesis submitted in partial fulfillment
of the requirements for the degree of
Docteur en sciences de l'ingénieur et technologie

Dissertation committee:

Prof. David BOL	(advisor)	UCLouvain, Belgium
Prof. Laurent FRANCIS	(president)	UCLouvain, Belgium
Prof. Denis FLANDRE	(secretary)	UCLouvain, Belgium
Dr. Laurie MARRAULD		EHESP, France
Prof. Antoine NONCLERCQ		ULB, Belgium
Dr. Nick VAN HELLEPUTTE		IMEC, Belgium

November 2022

Abstract

Information and communication technologies (ICT) play an increasingly important role in healthcare systems, through the development of eHealth applications such as electronic health records, telemedicine, remote health monitoring, or artificial intelligence for diagnosis support. Within this context, wearable medical devices for mobile health (mHealth) emerge as a valuable tool for monitoring patients over extended time periods, in their daily environment, with the aim of enhancing disease prevention, helping with diagnosis and ensuring proper treatment. The widespread deployment of those sensor devices however introduces technical and environmental challenges due for instance to the production of battery-powered devices or the management of large data flows. In order to mitigate these challenges, some solutions can be considered for the design of the devices. Embedded processing of the acquired signals alleviates the need for wirelessly transferring all the raw data. Ultra-low-power implementations of the electronic devices and optimization of the system power consumption reduce the required battery size. Adaptable and upgradable devices finally extend the device lifetime by limiting the risk of obsolescence. This thesis explores these solutions and questions whether they can contribute to enhancing the environmental sustainability of wearable devices.

The first part of the thesis discusses the design of a baseline low-power wearable device, focusing on the use case of cardiac arrhythmia monitoring with beat-to-beat arrhythmia classification. To address the challenges of low power consumption and upgradability, the system is implemented on a custom ultra-low-power (ULP) microcontroller unit (MCU). This MCU includes a biomedical analog front-end (AFE) designed for electrocardiogram (ECG) signal acquisition with a low-noise instrumentation amplifier and a time-based ADC. In the digital back-end (DBE), the sup-

port-vector machine classification software is optimized for low computing resources and is executed on a generic ARM Cortex-M4 processor implemented with ultra-low-voltage logic, adaptive back-biasing and custom ULP SRAM memories. A prototype fabricated in 28-nm fully-depleted silicon-on-insulator CMOS technology performs the ECG signal acquisition and arrhythmia classification tasks with an average power consumption of 13.1 μW. The demonstrated inference accuracy is on par with state-of-the-art inter-patient arrhythmia classification results. The MCU-based implementation inherently allows for upgradability of the system by programming new software.

The total power consumption of the system directly depends on the power consumption of its internal blocks, which are set by their local performance level, e.g. the noise of an amplifier, the resolution of an ADC or the complexity of a processing algorithm. For the designer, the objective is to specify the performance level for each component that allows reaching the target accuracy while minimizing the power consumption. In complex mixed-signal systems, in particular when hardly predictable machine-learning algorithms are used, those local block-level specifications can be difficult to determine. In order to avoid relying on a priori assumptions or trial-and-error methods, which are inefficient and often suboptimal, we propose to search for the system global optimum using a numerical optimization approach. To this end, in the second part of this thesis, power and performance models of the different components of the arrhythmia classification system are extracted in the respective analog and digital domains, for hardware and software components. Those models allow to obtain efficiently the performance at the component and at the system levels, with a reduced simulation overhead. Several methods for extracting the models are implemented and compared. These include a knowledge-based analytical approach, a simulation-based approach relying on multi-objective optimization and an efficient hardware-aware performance simulation of digital circuits. The extracted trade-offs are verified experimentally with prototypes of the AFE designed for configurable performance level. In the third part of this thesis, the global power and performance model of the system is used in combination with a numerical optimization algorithm. Several degrees of freedom, both in the analog and digital domains, are optimized at once to find the optimal set of design parameters that minimizes the total power consumption while preserving the arrhythmia classification accuracy. The power reduction that can be obtained varies signif-

icantly depending on several factors, which are discussed and illustrated in the considered use case, e.g. choice of the degrees of freedom and their impact on the power and performance, optimality of the baseline design, operating conditions of the device and performance constraints.

The last part of this thesis proposes to take a step back and have a look at more global sustainability aspects of mHealth. Through a review of the scientific literature on the considerations towards environmental sustainability of mHealth, we show that the focus lies predominantly on the direct environmental impacts of the device life cycle and on the reduction of indirect greenhouse gas emissions coming from the patient transportation to the healthcare facilities. Given the potential of eHealth and mHealth to deeply modify the current healthcare paradigm, a systemic vision is necessary and the environmental considerations currently taken into account fail to include the specificities of mHealth devices. A few examples are the possible optimization of healthcare procedures, changes induced in the demand for medical services or rebound effects at the scale of society. Those characteristics are analyzed and classified using the LES framework originating from the ICT sustainability assessment field, which distinguishes several levels of impacts from the device life cycle to structural impacts on the socio-economic system. The advantages and difficulties of this broader approach are discussed, with a focus on the multi-disciplinary dialogue that needs to emerge in order to fully capture sustainability issues.

Acknowledgements

During these last five years, from my first steps as a young PhD student in the corridors of the Maxwell building to the completion of this research work, I have had the chance to meet, work, collaborate, discuss, or share some drinks with many inspiring people who have directly or indirectly contributed to the success of this thesis. With this note, I would like to sincerely thank them all for accompanying me and supporting me through this adventure;

For proposing me to embark on this journey, for your continued support throughout, and for the many opportunities you gave me to grow on both scientific and personal levels:

My advisor, Prof. David Bol;

For your frequent and relevant feedback, as well as your relentless attention to details:

Prof. Denis Flandre;

For contributing to improving the quality of this thesis thanks to your thorough comments and for the insightful discussion during the private defense:

Dr. Laurie Marrauld,
Prof. Antoine Nonclercq,
Dr. Nick Van Helleputte;

For agreeing to chair the jury of this thesis:

Prof. Laurent Francis;

For sharing the joys and struggles of young researchers and IC designers, from the long tape-out nights to the more relaxed lunches and coffee breaks, as well as for reviewing some parts of this dissertation:

My dear colleagues from the Electronic Circuits and Systems group, Martin Lefebvre, Maxime Schramme, Adrian Kneip, Marco Gonzalez, Pengcheng Xu, Thibault Pirson, Mathieu Xhonneux, Ludovic Moreau, Julien Verecken, Louis Golard, Pol Maistriaux, Léopold Van Brandt, Charlotte Frenkel, Roghayeh Saeidi, Cecilia Gimeno, Nicolas Chauvaux, Thomas Haine, Pierre-Antoine Haddad, Sylvain Favresse, and Augustin Wattiez;

For the collaborations that allowed me to broaden my horizon:
With the SHINE project, Prof. Sandy Tubeuf, Prof. Florence Degavre, Prof. Suzanne Kieffer, Georgiana Sandu, and Charlotte Desterbecq,
With the COPINE-IoT project, Ivan Stupia, Nafiseh Janatian, Maxime Drouguet, Prof. Christophe Craeye, and Prof. Luc Vandendorpe,
With the VENG project, Prof. Riëm El Tahry and Pascal Doguet;

For your valuable logistic and administrative support and for providing the resources that enabled me to carry out this project:
The IT support staff, Brigitte Dupont and François Hubin,
The WELCOME staff, Pascal Simon, Souley Djadjandi, Benoît Hubert, Pierre Gérard, and Valeriya Kilchytska,
The ELEN and ICTEAM administrative staff, Viviane Sauvage, Anne Gillet, Isabelle Dargent, Christel Derzelle, Nathalie Ponet, Marie-Hélène Dewael, and Jean Deschuyter;

For the many good memories shared these last few years:
The members of the Club Elec and IEEE Student Branch,
And many other colleagues in ELEN and ICTEAM;

For funding this research:
UCLouvain,
The Fonds National de la Recherche Scientifique (F.R.S.-FNRS) of Belgium;

For the invaluable time spent together, sometimes reminding me to take a necessary step back from work:
Gilles, Nicolas, and Louis,
All my friends and family;

For being there since the beginning, and for all your love and care:
Maman, Papa, Emilie, and Lucie;

For your unfailing love and support, and for our many adventures to come:
Pauline;

From the bottom of my heart, thank you all!

Contents

Acronyms

ABB	Adaptive back biasing
ADC	Analog-to-digital converter
AF	Atrial fibrillation
AFE	Analog front-end
ASIC	Application-specific integrated circuit
BB	Back bias
BBCO	Back-bias controlled oscillator
BFGS	Broyden-Fletcher-Goldfarb-Shanno (algorithm)
BioZ	Bioimpedance
CAD	Computer-aided design
CMOS	Complementary metal-oxide-semiconductor
CMRR	Common-mode rejection ratio
CP	Charge pump
CPMU	Clock and power management unit
CPU	Central processing unit
CSRO	Current-starved ring oscillator
CVD	Cardiovascular diseases
DBE	Digital back-end
DMA	Direct memory access
DS1/2	Dataset division scheme
DWT	Discrete wavelet transform
ECG	Electrocardiogram
EEG	Electroencephalogram
eHealth	Electronic health
EMG	Electromyogram
ENOB	Effective number of bits
EOG	Electrooculogram

F	Fusion beat class
FBB	Forward back biasing
FDR	False discovery rate
FD-SOI	Fully-depleted silicon on insulator
FE	Feature extraction
FNR	False negative rate
FoM	Figure of merit
FVF	Flipped voltage follower
FVP	Fixed virtual platform
GHG	Greenhouse gas
$\mathbf{G_m}$	Transconductance
GSR	Galvanic skin response
HD	High density
HDL	Hardware description language
HPF	High-pass filter
IA	Instrumentation amplifier
IC	Integrated circuit
ICT	Information and communication technologies
IDE	Integrated development environment
Io(M)T	Internet of (medical) things
IRN	Input-referred noise
IRQ	Interrupt request
ISS	Instruction set simulator
KKT	Karush-Kuhn-Tucker (conditions)
LCA	Life cycle assessment
LES	Life cycle, enabling, structural (model)
LO	Local oscillator
MCU	Microcontroller unit
mHealth	Mobile health
MLII	Modified limb II (lead)
MOEA/D	Multi-objective evolutionary algorithm based on decomposition
MUBU	Multi-objective bottom-up
N	Non ectopic beat class
NEF	Noise efficiency factor
NDR	Negative differential resistance
NSGA(-II)	Non-dominated sorting genetic algorithm (v2)
OSR	Oversampling ratio
PEF	Power efficiency factor

PG	Power gating
PLL	Phased-locked loop
PPG	Photoplethysmogram
PPV	Positive predictive value
PSO	Particle swarm optimization
PVT	Process, voltage and temperature
Q	Discarded beat class
RBB	Reverse back biasing
RBF	Radial basis function
RMS	Root mean square
RO	Ring oscillator
RoHS	Restriction of hazardous substances (directive)
RTL	Register-transfer level
S	Supraventricular ectopic beat class
SCVR	Switched-capacitor voltage regulator
Sen	Sensitivity
SFFS	Sequential forward floating search
SLSQP	Sequential least squares programming
SNR	Signal-to-noise ratio
SoC	System-on-chip
SQP	Sequential quadratic programming
SRAM	Static random access memory
(S)SCA	(Smoothed) separable case approximation
SV	Support vector
SVM	Support-vector machine
TDCD	Top-down constraint-driven
THD	Total harmonic distortion
TLM	Transaction-level modeling
TRO	Tunable ring oscillator
UFBBR	Unified frequency and back-bias regulation
ULL	Ultra-low leakage
ULP	Ultra-low power
ULV	Ultra-low voltage
V	Ventricular ectopic beat class
VC(R)O	Voltage-controlled (ring) oscillator
WEEE	Waste electrical and electronic equipment (directive)
$\mathbf{Z_m}$	Transimpedance

Author's publication list

Related journal papers

[JP1] R. Dekimpe and D. Bol, "A Configurable ULP Instrumentation Amplifier with Pareto-Optimal Power-Noise Trade-Off Achieving 1.93 NEF in 65nm CMOS," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 68, no. 7, pp. 2272–2276, July 2021.

[JP2] D. Bol, M. Schramme, L. Moreau, P. Xu, R. Dekimpe, R. Saeidi, T. Haine, C. Frenkel and D. Flandre, "SleepRunner: A 28-nm FDSOI ULP Cortex-M0 MCU With ULL SRAM and UFBR PVT Compensation for 2.6–3.6-µW/DMIPS 40–80-MHz Active Mode and 131-nW/kB Fully Retentive Deep-Sleep Mode," *IEEE Journal of Solid-State Circuits*, vol. 56, no. 7, pp. 2272–2276, July 2021.

[JP3] F. Degavre, S. Kieffer, D. Bol, R. Dekimpe, C. Desterbecq, T. Pirson, G. Sandu and S. Tubeuf, "Searching for Sustainability in Health Systems: Toward a Multidisciplinary Evaluation of Mobile Health Innovations," *Sustainability*, vol. 14, no. 9, pp. 5286, April 2022.

[JP4] R. Dekimpe and D. Bol, "ECG Arrhythmia Classification on an Ultra-Low-Power Microcontroller," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 16, no. 3, pp. 456-466, June 2022.

Related conference papers

[CP1] D. Bol, M. Schramme, L. Moreau, T. Haine, P. Xu, C. Frenkel, R. Dekimpe, F. Stas and D. Flandre, "19.6 A 40-to-80MHz Sub-4μW/MHz ULV Cortex-M0 MCU SoC in 28nm FD-SOI with Dual-Loop Adaptive Back-Bias Generator for 20μs Wake-Up from Deep Fully Retentive Sleep Mode," *2019 IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 322-324, February 2019.

[CP2] D. Bol, T. Pirson and R. Dekimpe, "Moore's Law and ICT Innovation in the Anthropocene," *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 19-24, February 2021.

[CP3] R. Dekimpe, M. Schramme, M. Lefebvre, A. Kneip, R. Saeidi, M. Xhonneux, L. Moreau, M. Gonzalez, T. Pirson and D. Bol, "SleepRider: a 5.5 μW/MHz Cortex-M4 MCU in 28nm FD-SOI with ULP SRAM, Biomedical AFE and Fully-Integrated Power, Clock and Back-Bias Management," *2021 Symposium on VLSI Circuits*, pp. 1-2, June 2021.

[CP4] F. Degavre, D. Bol, S. Kieffer, G. Sandu, C. Desterbecq, T. Pirson, R. Dekimpe and S. Tubeuf, "Accelerating Sustainability Transitions in Care and Health Systems: How Sustainable are e/m-Care and Health Innovations? ," *Transforming Care Conference*, June 2021.

Unrelated journal papers

[UJP1] R. Dekimpe, P. Xu, M. Schramme, P. Gérard, D. Flandre and D. Bol, "A Battery-Less BLE Smart Sensor for Room Occupancy Tracking Supplied by 2.45-GHz Wireless Power Transfer," *Integration*, vol. 67, pp. 8-16, July 2019.

Unrelated conference papers

[UCP1] R. Dekimpe and T. Heldt, "Reducing false asystole alarms in intensive care," *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2292-2295, July 2017.

[UCP2] R. Dekimpe, P. Xu, M. Schramme, D. Flandre and D. Bol, "A Battery-Less BLE IoT Motion Detector Supplied by 2.45-GHz Wireless Power Transfer," *2018 28th International Symposium on Power and Timing Modeling, Optimization and Simulation (PATMOS)*, pp. 68-75, July 2018.

[UCP3] L. Moreau, R. Dekimpe and D. Bol, "A 0.4V 0.5fJ/cycle TSPC Flip-Flop in 65nm LP CMOS with Retention Mode Controlled by Clock-Gating Cells," *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1-4, May 2019.

[UCP4] M. Lefebvre, L. Moreau, R. Dekimpe and D. Bol, "7.7 A 0.2-to-3.6 TOPS/W Programmable Convolutional Imager SoC with In-Sensor Current-Domain Ternary-Weighted MAC Operations for Feature Extraction and Region-of-Interest Detection," *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 118-120, February 2021.

[UCP5] R. Dekimpe and D. Bol, "Mixed-Signal Compensation of Tripolar Cuff Electrode Imbalance in a Low-Noise ENG Analog Front-End," *2022 IEEE European Solid State Circuits Conference (ESSCIRC)*, September 2022, *accepted*.

Demo presentation

[DP1] R. Dekimpe, P. Xu, M. Schramme, N. Janatian, I. Stupia, M. Drouguet, P. Gérard, C. Craeye, L. Vandendorpe, D. Flandre and D. Bol, "2.45-GHz Wireless Power Transfer for BLE-connected Smart Motion Detection Sensor," *ACM/IEEE International Symposium on Low Power Electronics and Design (ISLPED)*, August 2019.

1

Introduction

Information and communication technologies (ICT) have become a key component of our everyday life and have impacted many different sectors such as work, retail, entertainment or industry. The healthcare sector is no exception to this trend, with the ever-increasing use of, e.g., electronic health records, health informatics or remote medical consultations. The technological ecosystem that has grown around those ICT-based applications is referred to as *eHealth*, which stands for electronic health. As shown in Fig. 1.1, categories have emerged within eHealth to distinguish different types of solutions such as telehealth, telemedicine, or mobile health, even though their exact definitions vary between authors. According to the classification of Doarn *et al.* [1], telehealth is understood as the electronic means used to provide clinical and non-clinical services outside the usual medical facilities (i.e., outside clinics and medical practices). Within telehealth, telemedicine is limited to the provision of remote clinical services, typically for consultations via telephone or videoconference. Finally, mobile health, or *mHealth*, refers to the "medical and public health practice supported by mobile devices" [2]. In addition to health applications for smartphones or other portable devices, wearable medical devices are a cornerstone of mHealth.

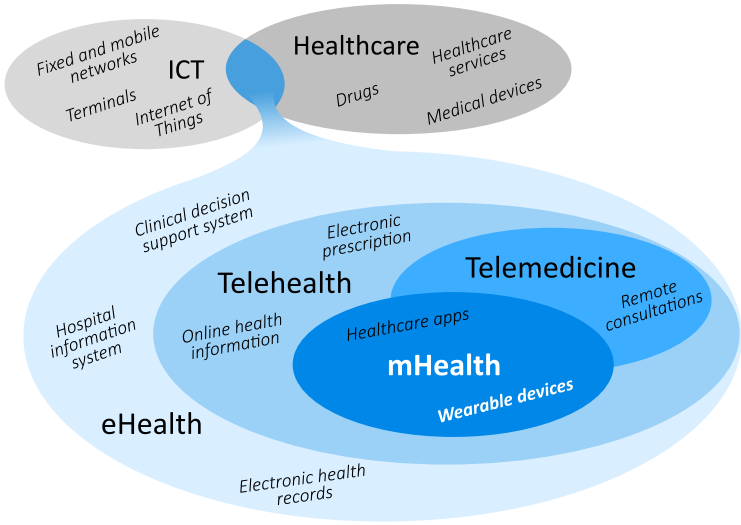

Fig. 1.1 eHealth technological ecosystem, which emerges at the intersection of ICT and healthcare. Shades of blue distinguish the subcategories of eHealth. Representative examples of each subcategory are displayed in italics. This PhD dissertation focuses on mHealth and wearable devices (in white).

1.1 Smart wearable devices

Wearable devices can be used as standalone systems or in combination with a smartphone to monitor physiological parameters and provide feedback regarding the health status of the wearer. A wide range of wearable devices already exist, with different shapes and sizes, from skin adhesive patches, to accessories such as watches or headbands, whole pieces of electronic-embedded clothing, or even composite systems [3]. Motion and fitness trackers for measuring the physical activity [4, 5] have been the main driver of commercially-available wearable devices, but other types of applications have since emerged in the industry or in research. Some examples of wearable devices are, among many others, glucose sensors for diabetic patients [6], cardiac monitors to detect arrhythmia [7] or posture monitoring garments [8].

Regardless of the sensor type or application, electronic systems are the backbone of wearable devices [9]. Those systems are composed of several circuit subparts, as shown in Fig. 1.2. They are capable of acquiring signals related to biophysical or biochemical processes using an analog

front-end (AFE), sometimes preceded by a transducer [10]. Typically, in the AFE, the signal is first amplified and filtered by an instrumentation amplifier (IA), which also provides a high input impedance and rejects unwanted artefacts such as electrode offset or common-mode interference. An analog-to-digital converter (ADC) then samples and quantizes the signal. The digitized signals can often be transmitted wirelessly relying on standard protocols such as Bluetooth, or using an ad hoc communication scheme. Most devices are battery-powered, but solutions based on environmental energy harvesting (e.g., thermal, motion or photovoltaic) are also emerging [11]. Some devices embed local processing in a digital backend (DBE) unit that allows to pre-process the acquired signal and possibly extract relevant information [9]. The DBE can be implemented using a central processing unit (CPU) with one or several cores executing software, or with application-specific logic circuits performing fixed tasks. All these functions of acquisition, processing, power management and communication are usually combined in one or several integrated circuits (IC), built in the majority of cases with complementary metal-oxide-semiconductor (CMOS) technology. Additional features such as display, user interfaces or actuators can be present in some cases, e.g., in smartwatches. The architecture illustrated in Fig. 1.2 is a typical example of Internet-of-Things (IoT) sensor devices, commonly used to gather, process and transmit data in a large variety of applications.

Electronic systems in wearable medical devices can record many different types of signals, with specific readout circuits in the AFE [10]. Within these signals, direct electric measurements include the measurement of biopotentials such as in the electrocardiogram (ECG) for cardiac activity, electroencephalogram (EEG) for brain waves, electromyogram (EMG) for muscle activation, or electrooculogram (EOG) for eye movements. Other electric measurements target the tissue impedance such as in bioimpedance measurement (BioZ) for respiration or galvanic skin response (GSR) to detect sweating. Some signal modalities require the use of transducers to convert the information from different physical domains such as optical signals in photoplethysmography (PPG) for blood oxygenation, or chemical signals for sensing biomolecules such as glucose or sweat electrolytes.

The monitoring of those physiological parameters continuously generates large amounts of data. For example, one day of recording typically represents from a few MB's (e.g., 3.3 MB for chest BioZ measurement with single 20-bit channel at 16 Hz [12]), to several GB's (e.g., 10.3 GB for 8-

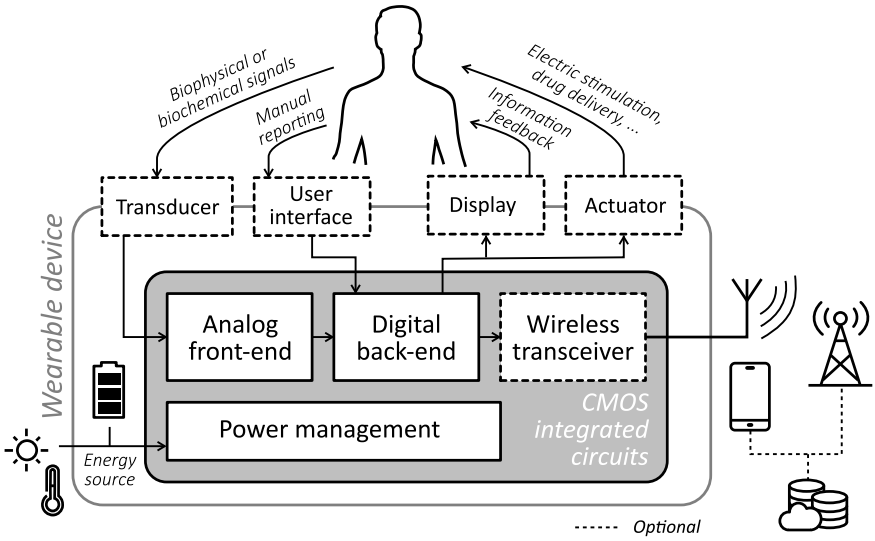

Fig. 1.2 Generic wearable device architecture. Electronic systems, and integrated circuits in particular, are the centerpiece of wearable sensor systems. Wearable devices interact with the wearer in various ways to gather information and provide feedback. They may be connected to the cloud for information storage, either directly or through a gateway smartphone. Optional components are framed with dashed lines.

channel 16-bit EMG recording at 8 kHz [13]). In the case of data acquisition systems with wireless communication, the transmission of the data requires a significant amount of power, often accounting for a large fraction of the device power budget [14]. For instance, energy intensity for short range communication such as Bluetooth Low Energy is in the order of a few nJ's per transmitted bit, as demonstrated with the latest Renesas transceiver which achieves 4 nJ/bit [15]. In the previous examples, this corresponds to an average power consumption for wireless transmission from 1.3 μW to 4.1 mW without taking payload overheads into account. When small batteries or energy harvesting are used as a power supply, this often exceeds the power budget of wearable systems which can be as low as a few μW's on average [11]. Other ad hoc communication schemes have been reported with energy efficiencies 10 to 100× higher, at the cost of shorter range, lower robustness, larger bandwidth, or the need for custom receiver gateways [16, 17, 18]. On top of that, the transmitted data needs to be transferred, stored, and analyzed downstream, either manually by a

clinician or using dedicated software on a computer or in the cloud. Assuming a large deployment of these wearable devices, along with other IoT devices, the ensuing data deluge will require a massive network and datacenter infrastructure [19]. In addition to the technical challenges of managing this increasing volume of data, the transfer, storage and processing of the data in the network and datacenters necessitates significant amounts of energy, and contributes to the environmental footprint of the ICT sector [20, 21]. To avoid these issues, an alternative is to embed the processing of the data directly on the wearable device [9, 22, 23]. This processing allows to directly infer the relevant clinical information, i.e., identify unusual events, classify normal and abnormal observations, or deduce physiological state from measured signals, so that only that meaningful information is stored or transmitted.

Local processing of the acquired data comes at the cost of embedding additional computing resources. The choice of processing algorithm depends on the type of signals and information to be retrieved. It ranges from simple signal thresholding [24], to Fourier transform for frequency analysis [25], or complex neural networks for event detection and classification [26], for example. While the use of edge processing can help reduce the power consumption of data transfers, it also has adverse effects. Firstly, new functionalities may require additional circuits and components, or increase the required silicon area. This increase in the device complexity impacts the production of the device, which requires more resources and energy to manufacture. When using the methodology and data from Pirson and Bol [27], the production carbon footprint of a basic monitoring wearable device can be estimated at approximately 1 $kgCO_2e$ of greenhouse gas (GHG) emissions[1], as shown in Fig. 1.3. This order of magnitude is also confirmed by Kokare [28] for cardiac monitoring devices. Embedding computing resources for advanced local processing can increase by up to $4\times$ the total footprint due to the required processor and memories in extreme cases [27]. Limiting the necessary hardware resources and integrating functionalities on a single chip help reduce the bill of material, and therefore mitigate the overall impact. Secondly, the device becomes dedicated to only a specific task defined by the processing algorithm. Compared to a device which transmits all the raw data, information may be

[1] While this footprint may seem reasonable, it is associated with a lot of variability, from the actual design specificities and from the environmental characterization databases. The total footprint also highly depends on the deployment scenario and the quantity of devices produced.

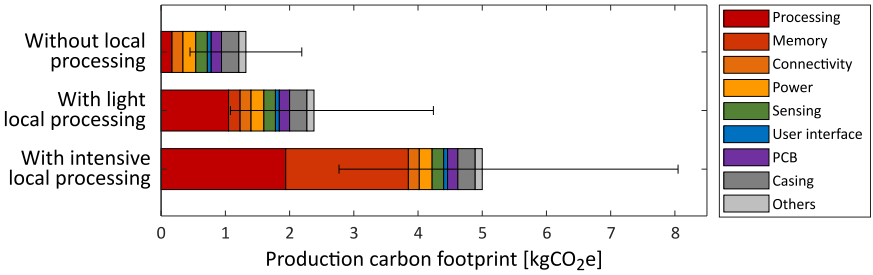

Fig. 1.3 Wearable device production carbon footprint. The carbon footprint for each component is estimated from [27], using hardware specification level (HSL) 0 for the processing, memory, PCB and others, and HSL 1 for the connectivity, power, user interface and casing. In the case of light and intensive local processing, HSL 1 and 2 are chosen respectively for the memory and processing, which corresponds to increasing levels of computing resources, e.g., up to for very large neural networks for HSL 2.

lost when only some features are transmitted. While this is not an issue as long as the extracted features correspond to the expected medical information, clinical guidelines may change and render the device obsolete, leading to its disposal. This reduces the service lifetime of the devices, increases the replacement rate, and leads to a proportional increase in the device footprint. Enabling upgrades of the processing task helps avoiding replacement of the device and contributes to preventing obsolescence. Thirdly, computing resources require energy to operate, therefore draining the batteries faster and necessitating larger energy capacity. Batteries have a detrimental impact on the environment and account for 5 to 20% of the environmental footprint of electronic devices [27, 28, 29], depending on their capacity. Optimization of the power consumption is therefore an essential aspect of wearable device design both for the patient comfort but also for environmental concerns.

1.2 Embedded processing in cardiac monitoring devices

The monitoring of cardiac activity through the acquisition of ECG signals is one of the most prevalent tasks in wearable devices, both in commercial devices and in research. Heart rate information is useful in many applications such as tracking of physical activity, stress or sleep patterns [10]. More advanced processing of the signals is also useful for the diagnosis of

cardiovascular diseases, which are among the leading causes of death in developed countries [30]. ECG signals are also relatively easy to record, as the biopotential amplitudes recorded on the skin go up to 10 mV, which is the largest amplitude compared to other non-invasive biopotential recordings [31]. Cardiac monitoring wearable devices have therefore benefited from a wide range of research, from both hardware and software/algorithm perspectives [10, 14, 32, 33], often with the goal of reducing power consumption to increase the monitoring time span, or improving the processing accuracy. Among all possible types of signal processing tasks for ECG systems, beat-to-beat arrhythmia classification consists in attributing a class to each individual beat with the objective to detect ectopic beats, i.e., beats that occur outside the normal sinus rhythm. The target classes are supraventricular (S) and ventricular (V) ectopic beats, and other "normal" non-ectopic beats (N). The frequency of occurrence of those abnormal events provides relevant information in the diagnosis of underlying heart conditions.

In this thesis, the use case of an ECG-based cardiac arrhythmia detection and beat-to-beat classification system is targeted as a representative and widespread example of wearable devices. In the system under consideration shown in Fig. 1.4, a single-lead ECG signal is measured through adhesive chest electrodes (2 differential signal electrodes and a reference electrode). The natural biopotential generated by the heart muscle is acquired by the AFE of the device using an IA and an ADC. The digitized signal is processed in the DBE to detect the heartbeat, perform feature extraction (FE) to measure relevant parameters from the signal, and classify each beat with respect to the presence of cardiac arrhythmia using a support-vector machine (SVM) classifier. At the output, the extracted information corresponds to the class of each heartbeat (i.e., N, S or V). The system can be supplied from a single supply voltage, with its internal voltages, clocks and references being generated in the clock and power management unit (CPMU). The target system is a small-size adhesive patch placed on the chest and supplied with batteries. Wireless communication can be used to transfer the beat classes to the smartphone, e.g., using Bluetooth LE. This work concentrates on the signal datapath in the AFE and DBE which need to run continuously and are the major determinants of the system inference accuracy. The challenge for this system lies in the high performance required for the acquisition and processing task. Indeed, compared to simple heart rate measurement, the acquisition of the signals requires

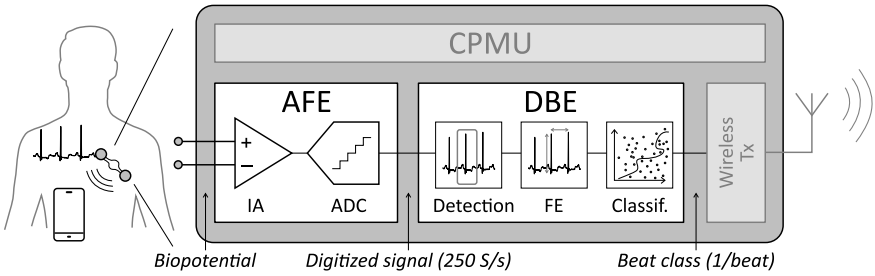

Fig. 1.4 Arrhythmia classification use case. ECG biopotentials are digitized by the AFE and processed in the DBE to infer the heartbeat class. This PhD dissertation focuses on the signal acquisition and processing.

improved signal quality to distinguish the morphologies of the different classes. The digital signal processing relies on complex machine-learning algorithms that require significant computing resources. Finally, these acquisition and processing tasks need to be performed in a limited power budget to enable long-term monitoring and limit the battery size.

Long-term cardiac monitoring and arrhythmia detection based on wearable devices can be used for multiple purposes, each necessitating different performance specifications. When used as an advanced diagnosis tool for cardiac disorders with infrequent symptoms, more focus is given to the classification performance, which requires a very high sensitivity to detect rare events. More accurate patient-specific classification models may therefore be necessary. If the device is used for the screening of populations, for enhancing prevention and identifying patients at risk, the constraints on the classification performance are relaxed, but the size and comfort of the device play an important role to ensure the patient compliance. Given the larger number of devices required to screen large populations, e.g., all people over the age of 50, the environmental impacts of those devices can become a critical factor. In this PhD dissertation, we focus on the latter use case, which emphasizes the need for low-power low-footprint devices with a generic arrhythmia classification model.

Several patches for arrhythmia detection in long-term ECG monitoring have been commercialized : Zio XT patch by iRhythms Technologies [34], SEEQ sensor by Medtronic [35], ePatch by Philips BioTel Heart [36] or Savvy ECG [37]. They can be worn from 7 to 30 days, therefore providing much longer monitoring time periods compared to traditional 24 or

72-hour ambulatory ECG monitors. However, all these commercial ECG monitoring patches rely on offline data processing in the cloud [38], which impacts the power budget, latency and environmental footprint of the device. In the literature, embedded processing of the acquired signal has been used to reduce the amount of data transferred. In particular, hardware implementations of beat-to-beat classification systems are usually based on application-specific integrated circuits (ASIC) that provide enhanced energy efficiency [39, 40]. ASIC implementations however often come at the cost of limited flexibility, which can induce obsolescence issues as upgrades are not possible when the embedded task is outdated in the current medical practice context. On the other hand, biomedical systems based on microcontroller units (MCU) provide increased versatility but have a total power consumption which is orders of magnitude higher than their ASIC counterparts [41, 42]. Additionally, such MCUs have only been demonstrated with simpler processing tasks such as detection of heartbeats and irregular rhythm [43, 44]. Ultra-low-power (ULP) implementations of MCUs may bridge the gap between the required complexity of advanced biomedical processing algorithms, high energy efficiency for limited power budget and upgradability to avoid obsolescence. The first question addressed in this thesis is therefore:

Question 1 *Can ULP MCUs enable complex on-chip real-time signal processing such as beat-to-beat arrhythmia classification, within a limited power budget of a few µW's?*

1.3 Optimization of block-level performance trade-offs for hierarchical system design

In electronic sensor systems, the top-level system performance is a function of the local specifications of its constituents, in the analog or digital domains, in hardware or software. An obvious example is the total power consumption of the system which is equal to the sum of the power consumption of each individual block. Another less trivial example is the global inference accuracy which depends non-linearly on the signal acquisition quality (e.g., intrinsic noise, distortion, resolution) and the processing complexity (e.g., data representation, number of operations, size of the model). In order to improve the system performance, the most common approach is to optimize individually each circuit constituting the electronic system, either to reduce their power consumption or improve some of their

performance metrics. For analog circuits, this improvement is typically obtained by proposing new circuit topologies or optimized sizing. For digital components, it is achieved by improving the algorithms, optimizing the software implementation or proposing new circuit architectures. However, achieving high performance at the block-level does not necessarily translate to optimal performance for the system. Indeed, improperly setting the local specifications of the system components can lead to either non-functional systems or overdesigned circuits with power and performance overheads. Therefore, another approach, complementary to the previous one, is rather to find the local specifications of the circuit blocks in the system that optimize the global performance. In power-constrained applications such as the arrhythmia classification use case, finding the minimal performance requirements for both analog and digital parts of the data path could bring additional energy savings and reduce the battery size.

When optimizing several performance metrics of a circuit simultaneously, an optimal design is one for which it is not possible to improve one metric without degrading another, as illustrated in Fig. 1.5. This design point is known as a Pareto optimum. In many cases, several optima exist, each corresponding to a different balance between performance objectives, which corresponds to the Pareto optimal front. This situation is more commonly referred to as a trade-off between several specifications and is often encountered by IC designers. For example, the noise level in an amplifier, the resolution of an ADC or the complexity of a processing algorithm are all in conflict with the power consumption of their respective blocks. When the Pareto curves of the constituting circuits are known, the optimization task at the system level becomes a question of choosing the appropriate point along the curve to balance the local performance specifications of the different blocks and reach a global optimum.

Hierarchical design methodologies have been applied to analog systems to relate local performance at the block-level with top-level system specifications [45]. The most common ones are the bottom-up and top-down design approaches [46], which have been revisited to ensure feasibility and improve optimality, such as in the multi-objective bottom-up methodology [47] or feasibility modeling with top-down constrained design [48]. To the best of our knowledge, most of these hierarchical design methodologies aim at finding a feasible and functional system design that meets a number of constraints, but not to the optimization of the overall performance. Additionally, none of these have been applied to complex mixed-signal sys-

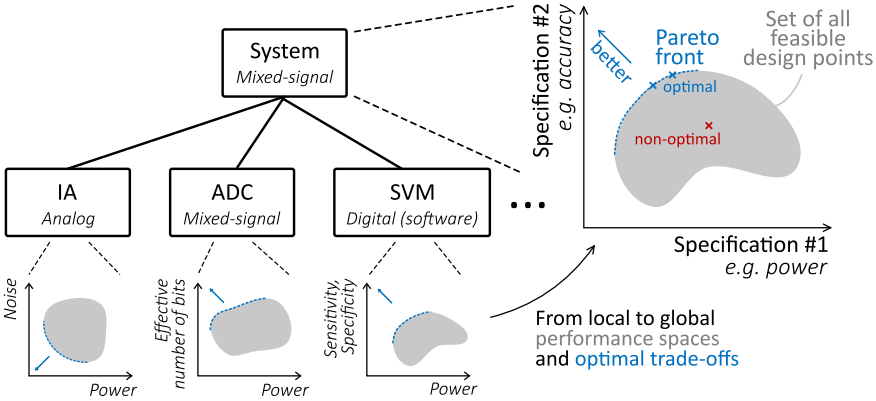

Fig. 1.5 Performance trade-offs in sensor systems. Local performance trade-offs propagate from constituting blocks to the top level and define the Pareto curve of the system. The blocks in the system are distributed in different domains: analog/digital and hardware/software.

tems, across large analog and digital domains. The increased complexity and heterogeneity of complex sensor systems introduces new challenges in the propagation of performance and constraints between the levels of hierarchy. Indeed, the relationship between signal quality in the AFE and processing performance in the DBE is highly non-linear and hardly predictable. Analog and digital blocks, including algorithm design, rely on very different design and simulation tools that deal with different signal representations. The number of degrees of freedom for the design increases with the circuit size and tends to increase the optimization complexity. In order to address this cross-domain optimization problem, the second question that we raise in this thesis is :

Question 2 *How can we tune the block-level performance requirements in order to minimize the system power consumption while maintaining the inference accuracy in complex mixed-signal sensor systems?*

1.4 Systemic approach to environmental sustainability

In electronic systems, the design of smart ultra-low-power highly-integrated sensors has often been considered as an important R&D avenue to providing environmentally sustainable technologies [49, 50]. In this work, this is exemplified by on-board processing to limit energy-hungry

data transfers, low-power operation to reduce the battery size, and reprogrammable MCUs for limiting device obsolescence. On top of that, the digitalization and virtualization enabled by the deployment of ICT infrastructure in many different sectors such as work, retail, agriculture or transport have proved useful in substituting or optimizing some aspects of their activities to reduce the environmental footprint. The combination of efficiency improvements in electronic equipments and the optimization possibilities offered by digital technologies is usually considered as having a positive outcome for the environment [51]. However, due to adverse secondary effects such as obsolescence, induced consumption, or systemic rebound effects, concerns are being expressed regarding the actual environmental footprint reduction brought by ICT in the current context [51, 52, 53]. In order to better understand the global balance between positive and negative impacts of digitalization, this calls for more comprehensive and systemic considerations of environmental impacts in the different sectors.

Similarly in healthcare, the emergence of eHealth and mHealth is primarily seen as a promising solution for mitigating the environmental impacts of the medical sector [54]. Healthcare is a socio-economically important sector which has a significant environmental footprint. For example, in developed countries, the healthcare sector footprint accounts for 3.3 to 8.1% of national carbon footprints [55], and hospitals generate large amounts of liquid and solid waste, around 800 L/bed/day and 2-4 kg/bed/day respectively [56]. The digitalization of healthcare, including the emergence of mHealth technologies, has been considered as an important lever for the reduction of these impacts. Some studies have indeed estimated the savings of GHG emissions provided by eHealth applications such as telemedicine or electronic medical records [57, 58]. On the other hand, the production, use, and disposal of these new electronic devices and ICT infrastructure also comes with direct environmental impacts [28, 59, 60]. Additionally, the negative indirect effects such as rebound effects, obsolescence, or other structural changes, are often overlooked. Overall, the research efforts to better understand the positive and negative environmental impacts of emerging mHealth technologies remain limited and disparate, and they fail to evaluate mHealth as a whole system. It is therefore still uncertain whether the overall balance will tip in the direction of an increase or decrease of the sector impacts. Given the urgent need to address the environmental impacts in all sectors, including health-

care and ICT, this uncertainty is a bottleneck in the proper anticipation of the future impacts of these new technologies.

mHealth devices differ significantly from traditional medical devices. The usual approaches for environmental assessment of medical equipment may therefore miss important characteristics. For instance, wearable devices rely primarily on electronics, whose complex fabrication processes require lots of energy and resources proportionally to the device size, and induce large environmental impacts [27, 61]. Their connectivity and ubiquity enable them by design to be used outside the usual medical settings, which multiplies the possible use cases and leaves them outside traditional medical or regulatory oversight. Regarding their business model, while traditional medical devices are only bought upon request of the doctor, some wearable devices such as smartwatches, fitness trackers and other wellness-related devices are mainly acquired on individual initiative like other consumer electronics. Finally, according to some projections [62], these technologies are expected to be adopted at a very large scale, with impacts increasing proportionally. Given their specific characteristics and the potential scale of deployment and impacts, mHealth technologies have the potential to be a significant driver of healthcare environmental impacts, be it in a positive or negative direction. In order to properly anticipate the outcome of their adoption, the third question addressed in this thesis is:

Question 3 *How can we assess whether mHealth technologies contribute to the environmental sustainability of the healthcare system?*

1.5 Thesis outline

This thesis explores the optimization of CMOS integrated circuits for biomedical sensor systems, across multiple domains, analog/digital and hardware/software, and several abstraction layers from circuits to applications. The contributions and the tentative answers to the questions raised previously are divided in the following chapters, each addressing different levels of the abstraction hierarchy as shown in Fig. 1.6. Firstly, to address the challenges of managing the large quantities of data generated by wearable devices, Chapter 2 evaluates the potential of ULP MCUs to implement onboard processing with a limited energy overhead and improved upgradability. As a use case, the design of a cardiac arrhythmia classification system on a custom ULP MCU is described. Analog and digital blocks in the proposed system are optimized at the circuit-level. Secondly, we investi-

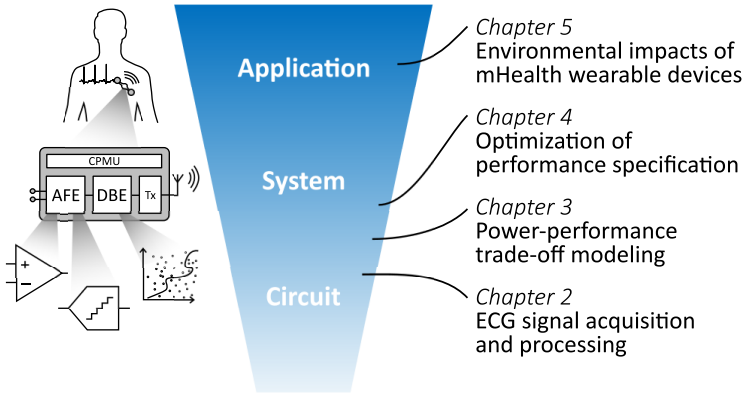

Fig. 1.6 Circuits and systems abstraction layers. Each chapter explores different parts of the abstraction hierarchy.

gate the minimization of the power consumption of the system to enable the reduction of the battery capacity. At the system-level, a cross-domain hierarchical optimization method which aims at optimizing the performance specifications of the individual circuit blocks is therefore presented, focusing on the cardiac arrhythmia classification system as a use case. The method relies on the modeling of power and performance trade-offs of the circuit, which is achieved in Chapter 3 using different approaches. Based on these these models, the block-level performance requirement are optimized numerically in Chapter 4 to find the global minimum of the system power consumption. Finally, working on the efficiency optimization of electronic devices is the usual first step to make mHealth systems sustainable, but tends to overlook secondary effects. We therefore question the current approach to sustainability of mHealth systems to include more systemic considerations in Chapter 5.

Chapter 2 As a baseline system for the cross-domain and cross-layer optimization methodology, we start by designing an electronic sensor system for wearable cardiac monitoring. While previous works focused on ASIC implementations or limited on-chip processing, we discuss the implementation of a complete beat-to-beat arrhythmia classification system on a custom ULP MCU. The architecture of the system is similar to the use case presented in Section 1.2. It includes a single-channel AFE circuit for ECG signal acquisition, and a DBE processor to execute the SVM classification software with a Cortex-M4 CPU. The low-noise instrumentation

amplifier in the AFE consumes 1.4 μW and has an input-referred noise of 0.9 μV RMS. The all-digital time-based ADC achieves 10-bit effective resolution over a 250-Hz bandwidth with an area of only 900 μm^2. The classification software reaches a sensitivity of 82.6% and 88.9% for supraventricular and ventricular arrhythmias respectively on the MIT-BIH arrhythmia database. The proposed system has been prototyped on the SleepRider SoC, a 28-nm fully-depleted silicon on insulator (FD-SOI) chip, with a core area of 0.94 mm^2. The ULP MCU achieves an energy efficiency of 5.5 μW/MHz. When performing the arrhythmia classification, the system consumes 13.1 μW on average from a 1.8-V supply. This work is the first to document the end-to-end implementation of a cardiac arrhythmia classification system on an ULP MCU, with limited energy overhead compared to ASIC counterparts and increased versatility. The results in this chapter have been presented in [CP3] for the overall MCU design and [JP4] for the specificities of the arrhythmia classification system.

Chapter 3 Starting from the baseline arrhythmia classification system implemented in the first chapter, the power and performance trade-offs of its constituting circuit blocks are modeled. After reviewing the methodologies previously proposed in the literature, we compare several approaches by applying them to the respective circuits of the target system. Analytical modeling is used to extract the Pareto curve of the ADC regarding resolution and sampling frequency. A numerical simulation-based approach is used to model the power and noise trade-off of the IA using a multi-objective genetic optimization algorithm. These analytical and numerical approaches are validated experimentally by implementing AFE prototypes with configurable Pareto-optimal performance for the IA and ADC. The configurable IA and ADC achieve a performance of respectively 0.17 to 1 μV RMS of input-referred noise, and 8 to 12 bits of resolution. Finally, the evaluation of the inference accuracy of the SVM classification software is optimized for efficient hardware-aware execution. The different modeling approaches are finally compared to highlight their respective advantages and disadvantages, e.g., in terms of accuracy, ease of implementation or execution time, and to identify in which situations they are most suitable. The IA genetic optimization, modeling, and configurability implementation have been published previously in [JP1].

Chapter 4 The local performance requirements of the individual blocks in the arrhythmia classification system are optimized in order to minimize the total power consumption while maintaining a high inference accu-

racy. Contrary to previous hierarchical design methodologies which are either suboptimal or require lots of simulations, the approach that we propose aims at providing a global performance optimum while relying on a limited number of time-consuming simulations, even for large mixed-signal systems. Firstly, performance macromodels of the individual blocks are generated and combined in a parametrized model for simulating the system power and inference performance. The parametrization is based on the performance trade-offs modeled in Chapter 3. Secondly, this system macromodel is used to find the global power minimum while limiting the degradation of the inference accuracy. This minimization problem is solved numerically using the gradient-based SLSQP optimization algorithm [63]. The results of the numerical optimization show that the proposed methodology enables the optimization of complex mixed-signal systems, across analog/digital and hardware/software domains. In the arrhythmia classification use case, the optimization results show a reduction of the power consumption between 19 and 84% depending on the starting point performance. In the nominal case, the final power consumption is 3.3 µW, while keeping the error rate degradation below 50% from its baseline. The advantages and limitations of the proposed cross-domain system optimization methodology, regarding execution time, convergence, or parameter definition, are finally discussed.

Chapter 5 As a step back from the energy efficiency optimization of wearable devices, this last chapter looks at the mHealth system as a whole and its impacts on the environmental sustainability of the healthcare sector. While eHealth and mHealth are often considered as levers for acting on the environmental impacts of healthcare, the systemic changes induced by these new technologies are often neglected. Our systematic review of the scientific literature explores the considerations on environmental sustainability of mHealth technologies. The selection of relevant articles is performed according to the PRISMA methodology and an overview of the articles addressing this subject is provided. The considerations regarding environmental impacts are analyzed through a content analysis along three dimensions: the scope of the system whose impacts are considered, the direct or indirect nature of these impacts and the indicators used to characterize them. The results of this analysis highlight the limitations of the current environmental sustainability evaluations in the literature. They show that mHealth environmental sustainability literature is still in its early stages and barely addresses the specificities of mHealth devices. More impor-

tantly, the impacts of mHealth at the system level are only considered positively, either through the reduction of transportation or the optimization of care procedures, but overlook many potential indirect impacts. To address these limitations, we use the LES model of Hilty and Aebischer [51], originally from the field of ICT, and apply it to mHealth systems in order to provide the first system-level prospective analysis of their impacts. We illustrate the systemic impacts of wearable devices through examples inspired from the literature. The review part of this chapter extends the work published in [JP3] and [CP4] regarding the aspects of ICT and environmental impacts.

2

ECG arrhythmia monitoring system

2.1 System-on-chips for cardiac arrhythmia classification

Cardiovascular diseases (CVD) are the leading cause of death and disability worldwide, especially in developed countries [30]. Among those, cardiac arrhythmias are associated with an increased risk of stroke and sudden cardiac death [64]. In order to diagnose heart diseases, one of the most effective tools is the electrocardiogram (ECG) which measures the electrical activity of the heart. However, due to the infrequent nature of arrhythmia events, long-term monitoring using a wearable device, as described in Chapter 1, is often necessary to observe them. Nevertheless, being able to use monitoring devices during the patient's daily activities imposes stringent requirements on their signal quality, size, and power consumption. To avoid the need for transmitting or storing very large amounts of data, real-time processing of the ECG signal to detect cardiac arrhythmias has emerged as a solution to extract only the relevant information when abnormal events occur [65, 66, 67].

System-on-chips (SoCs) for arrhythmia detection typically consist of an analog front-end (AFE) circuit which interfaces the electrodes and acquires the signal, and a digital back-end (DBE) processor which processes the

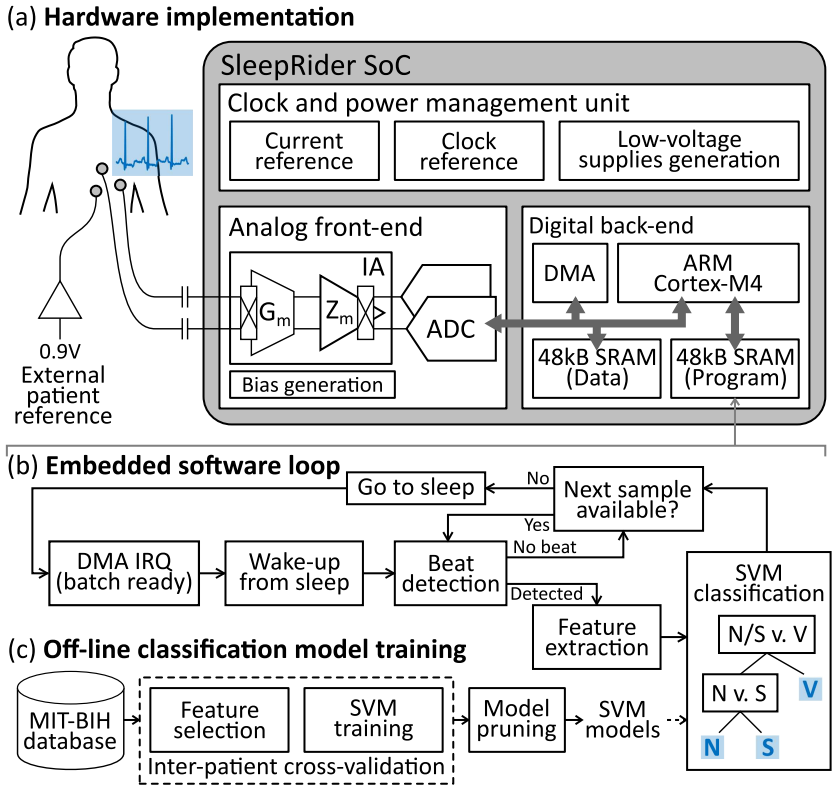

Fig. 2.1 Architecture of the arrhythmia classification system. The top schematic illustrates the hardware components of the SoC (a). The bottom block diagram describes the software execution in the DBE (b), along with the off-line training process (c). The blue color highlights the ECG signal input and the output inferred beat classes (N, S or V for normal beat, supraventricular, or ventricular arrhythmia, respectively).

data, as shown in Fig. 2.1. In the AFE, the main challenge is to improve the trade-off between noise and power consumption, while avoiding artefacts thanks to large input impedance, low distortion, and high rejection of both common-mode signals and differential electrode offset. In the DBE, the challenge resides in implementing a classification software usually based on computationally-intensive machine-learning algorithms on a platform with low hardware resources.

While most previous works have focused on optimizing only a part of the system, some have proposed full mixed-signal SoCs with both analog acquisition and digital processing capabilities [41, 42, 43, 65, 66, 68, 69, 70]. Sub-μW power consumption in the AFE has been achieved by data-dependent reconfigurable performance [68], relaxed noise and quantization constraints [43, 69], or power-efficient topologies such as current-reuse ones [65]. In the DBE, custom logic circuits with fixed hardware are often used to minimize the classification energy [65, 69, 70]. Chen *et al.* [43] and Lee *et al.* [66] have presented low-power programmable DBEs based on a CPU or a neural network accelerator, respectively. However, the demonstrated arrhythmia detection task is limited to the discrimination of abnormal rhythm with coarse temporal resolution. Others such as Schönle *et al.* [41] or Song *et al.* [42] proposed versatile SoCs including several biomedical front-ends as well as general purpose processors for signal processing. They however do not discuss the tasks performed by the DBE, and a full end-to-end performance evaluation is hence lacking.

As detailed in Chapter 1, the power consumption, upgradability and complexity of the circuits used for local data processing have an impact on the environmental sustainability of the wearable devices. While systems based on a microcontroller unit (MCU) benefit from the inherent software upgradability to reduce the risk of obsolescence, they are often outperformed by their ASIC counterparts in terms of energy efficiency and real-time execution. The total average power corresponds the sum of the power consumption in the analog block and digital blocks, which depends in turn on the energy efficiency of the hardware, and the complexity of the algorithm and its implementation efficiency in software. The global inference performance depends on the signal quality achieved in the AFE and on the detection and classification performance of the algorithm in the DBE. Any suboptimal component, in terms of power consumption or performance, would be detrimental to the global system. Bridging this performance gap for MCUs thus requires optimizing both the AFE and DBE, at the circuit-level for hardware and software aspects, with a high-level application-centred perspective.

In this chapter, we present a mixed-signal ECG SoC for arrhythmia classification based on an ultra-low-power (ULP) MCU. This approach combines a low average power consumption of 13.1 μW with high and versatile computing resources to perform advanced processing tasks such as beat-to-beat arrhythmia classification. To the best of our knowledge, this

is the first work documenting the implementation of an ECG beat-to-beat classification system on an MCU covering hardware and software aspects, including machine-learning algorithm design. Section 2.2 defines the target arrhythmia classification system. The architecture of the custom ULP MCU is described in Section 2.3. In Section 2.4, the design of the amplifier and ADC in the AFE is described, focusing on low noise and low power performance. Section 2.5 details the implementation of the classification software executed on the MCU. The measurement results of the prototype chip are provided in Section 2.6 for the AFE and DBE separately, and then for the whole system. The ULP MCU and arrhythmia classification system were presented in [CP3] and [JP4], respectively.

2.2 Proposed arrhythmia classification system

The proposed ECG arrhythmia classification system is embedded in the SleepRider MCU SoC [CP3], as shown in Fig. 2.1(a). A pair of chest gel electrodes are connected to the differential input of the instrumentation amplifier (IA) in the AFE. A third electrode is used to set the patient reference voltage through an off-chip buffer. The input signal, whose amplitude is only a few mV's for chest recordings, is amplified by the IA, featuring a current-balancing topology with chopper modulation similar to [JP1]. The differential output of the IA is digitized at 250 Hz using two single-ended time-based ADCs with 14-bit resolution. The acquired samples are temporarily stored in the SRAM data memory by the direct memory access (DMA) module while the system is in sleep mode and the DBE is inactive. When 128 samples have been stored in the memory (approximately every 0.5 s), the ARM Cortex M4 CPU inside the DBE wakes up and processes the data. A clock and power management unit (CPMU) provides a reference current, a low-frequency peripheral clock at 4 MHz and a high-frequency clock for the DBE at 64 MHz, and several low-voltage internal supplies for the different blocks

Within each batch of data from the DMA, samples are processed one at a time in the DBE as illustrated in Fig. 2.1(b), starting with the heartbeat detection algorithm from Pan and Tompkins [71]. As shown in Fig. 2.2, the beat detection consists in signal filtering and derivation to enhance the sharp edges, followed by peak detection rules with an adaptive threshold for robustness against noise. When a beat is detected, a vector of features based on the beat intervals, the downsampled signal, and wavelet

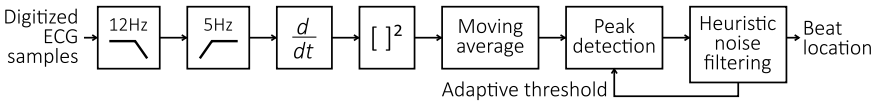

Fig. 2.2 Block diagram of the Pan and Tompkins algorithm [71], applied on the acquired samples from the AFE.

Table 2.1 Mapping of beat classes. Target classes are those considered in this work. Annotations correspond to the labels in the PhysioBank database [73].

Target class	Beat annotation
"Normal" non-ectopic beat (N)	Normal beat Left or right bundle branch block beat Atrial or nodal escape beat
Supraventricular ectopic beat (S)	Atrial premature beat Nodal escape beat Supraventricular escape beat
Ventricular ectopic beat (V)	Premature ventricular contraction Ventricular escape beat Fusion of ventricular and normal beat
Discarded (Q)	Paced beat Unclassified beat

transform coefficients is extracted. A hierarchical support-vector machine (SVM) shown in Fig. 2.1(b) with a radial basis function (RBF) kernel classifies the heartbeat as a normal beat, a supraventricular ectopic beat, or a ventricular ectopic beat (respectively N, S, or V beats), corresponding to the classes defined by the Association for the Advancement of Medical Instrumentation [72]. Table 2.1 shows the mapping between PhysioBank annotations corresponding to heartbeat abnormalities and the targeted classes [73, 74].

The classification model training is performed off-line (Fig. 2.1(c)). The MIT-BIH arrhythmia database [75], which consists in 48 30-minute records with beat-specific annotations, is used for algorithm design and validation. While two channels are available in the database, only the modified limb II (MLII) lead is used in this work. The training and testing sets are made up of 22 records each following the dataset partition (DS1/DS2) and N/S/V

labelling of de Chazal *et al.* [74] (4 paced records are discarded). The training and cross-validation of the classifier follow the inter-patient scheme to avoid the optimistic bias of intra-patient training. Feature selection improves the inference performance and reduces the computational complexity for the SVM. It uses a greedy wrapper scheme with a sequential forward floating search (SFFS) algorithm [76]. To reduce the memory requirements of the SVM algorithm, a pruning step is added to reduce the number of support vectors using the separable case approximation (SCA) [77].

2.3 Ultra-low-power microcontroller design

The proposed ULP MCU was designed in collaboration with the following researchers from the Electronic Circuits and Systems group (ICTEAM - UCLouvain): Maxime Schramme, Martin Lefebvre, Adrian Kneip, Roghayeh Saeidi, Mathieu Xhonneux, Ludovic Moreau, Marco Gonzalez, and Thibault Pirson.

The architecture of the proposed MCU, codenamed SleepRider, is built around an ARM Cortex-M4 processor with several digital interfaces such as timers, DMA or off-chip communication peripherals (Fig. 2.3). The memory includes two 16-kB custom ULP SRAMs with ultra-low leakage (ULL) in sleep and two 32-kB high-density (HD) foundry SRAMs for increased memory capacity. The AFE includes an IA and a dual ADC with a differential input. The CPMU generates all clocks and ultra-low-voltage (ULV) supplies from internal references. Adaptive back-biasing (ABB) is performed for the compensation of process, voltage, and temperature (PVT) variations.

In order to reduce the need for external components and to rely on a single 1.8-V supply (*VDDIO*), switched-capacitor voltage regulators (SCVRs) generate three internal voltage supplies, respectively at 0.4 V (*VDDL*) for the logic, 0.5 V (*VDDM*) for the ULP SRAMs and 0.8 V (*VDDH*) primarily for HD SRAMs. Negative charge pumps (CPs) generate a negative supply at 0.2 V (*VSSN*) in active mode used by the ULP SRAMs. The comparators, control logic and local oscillator (LO) of the SCVRs use an auxiliary 0.9-V supply voltage (*VDDI*) from a high drop-out linear regulator. In sleep mode, the SCVR clock switches from the LO to a 500-kHz reference clock.

Operation at ULV for the digital logic allows to save power, at the cost of increased delays and sensitivity to PVT variations. In FD-SOI technologies, the back-bias (BB) voltages can be used as an additional tuning knob to change the speed and leakage trade-off of digital gates thanks to adap-

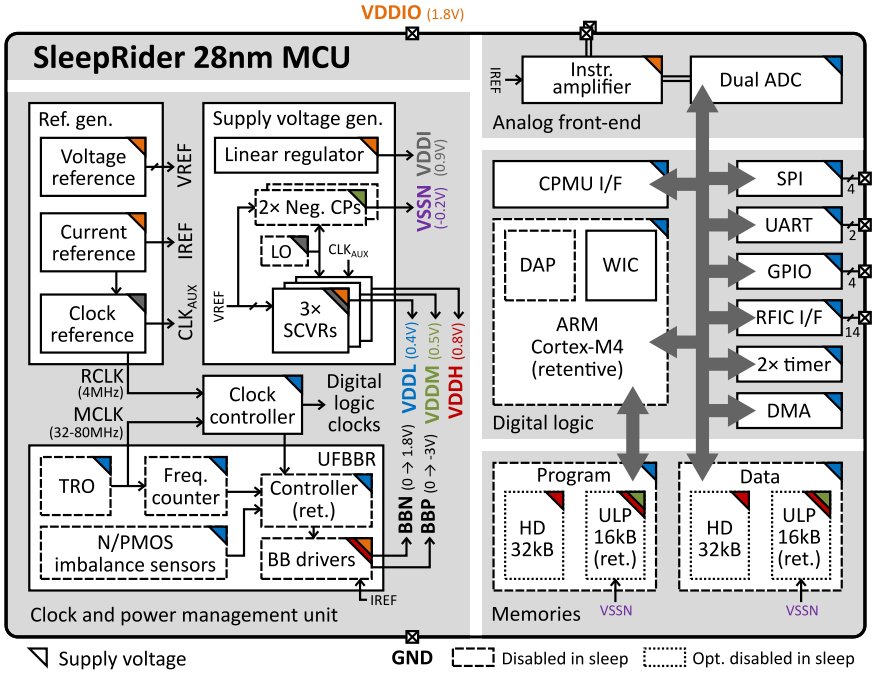

Fig. 2.3 SleepRider MCU architecture. Colored triangles identify the voltage domains. Sleep-mode status is indicated by dashed or solid lines. Digital logic and ULP SRAMs retain data in sleep mode.

tive back-biasing (ABB) methods. In particular, the unified frequency and back-bias regulation (UFBBR) technique efficiently compensates for PVT variations that affect both logic speed and N/PMOS imbalance [JP2]. However, the coupling between BBN and BBP (i.e. NMOS/PMOS BB voltages) due to the parasitic well capacitances threatens the stability of independent BBN/BBP control loops. Instead, a unified dual-output frequency-locked loop architecture shown in Fig. 2.4 is used, avoiding the need for off-chip capacitors. Logic speed and imbalance are respectively sensed by a tunable ring oscillator (TRO) – also used as the main system clock in active mode – and N/PMOS-sensitive ROs. Digitally-controlled oscillators driving switched-cap CPs perform fast proportional current actuation. The proposed UFBBR is stable in all PVT corners and locks in 16 µs. In comparison, the ABB from [78] performs only one control cycle in 14.9 µs which leads to a much longer lock time.

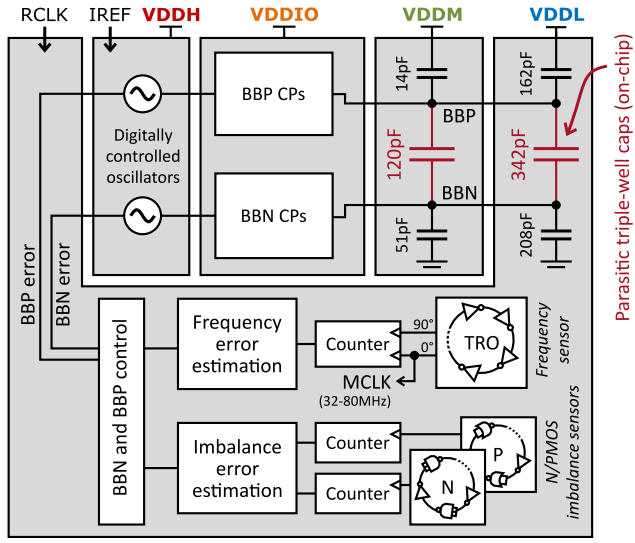

Fig. 2.4 UFBBR frequency-locked loop architecture. The UFBBR generates the main system clock MCLK and the back bias voltages in a single control loop. Despite parasitic capacitances that couple BBN and BBP voltages, the loop remains stable without the use of external capacitor.

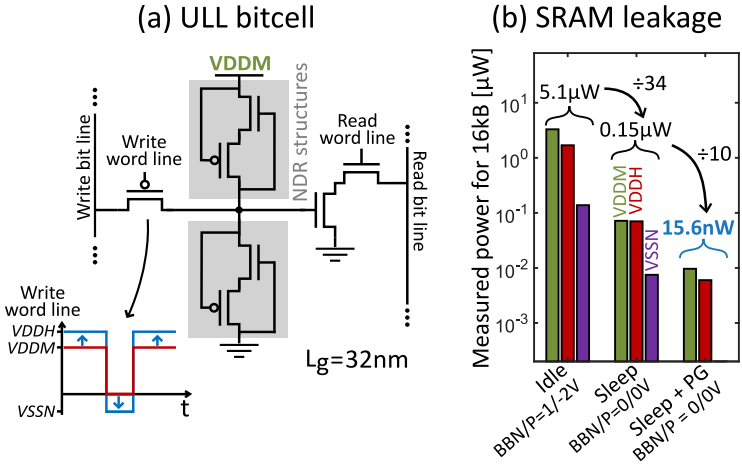

Fig. 2.5 ULP SRAM. The proposed 7T bitcell (a) with NDR latch requires boosted word line voltages to improve stability and write operations. SRAM leakage (b) is reduced thanks to zero back-bias and power gating of the periphery.

SRAM memories in microcontrollers often contribute significantly to the system overall performance and power consumption. In particular, data retention in sleep reduces the wake-up overhead, at the cost of increased leakage which is detrimental at low activity duty-cycles. Negative-differential-resistance (NDR) latches have been used in ULP SRAMs for their ABB compatibility and ULL data retention. Compared to [JP2], the ULP SRAM uses a modified 7-transistor ULL bitcell with uniform 32-nm gate length (L_g) and a PMOS-only write port (Fig. 2.5), providing 2.1× area and 15× leakage reduction in sleep. 2D sub-banking in 0.5-kB local arrays decreases the read access time by 1.5× and brings the mean read (resp. write) access energy at 64 MHz down to 37 (resp. 26) fJ/bit. Finally, power gating (PG) of the SRAM periphery brings its sleep power down to 0.98 nW/kB.

2.4 Analog front-end design

2.4.1 Low-noise instrumentation amplifier

The fully-differential current-balancing topology of the IA shown in Fig. 2.6 relies on a transconductance (G_m) and a transimpedance (Z_m) stage. In the G_m stage, the input differential voltage v_{in} is copied by the flipped voltage follower (FVF) structure with transistors P1 and P2 across resistance R_i, which induces a current difference between the left and right branches. The FVF provides a unity gain with a large input range, while readily acting as a current sensor [79]. The P3-P5 structure allows to lower the drain voltage of P1 and improves the input dynamic range of the IA. The current difference is copied to the Z_m stage thanks to the P2-P6 current mirror with current gain A and goes through resistor R_o to generate a differential voltage. The total gain A_v of the IA is defined by the product of the G_m, Z_m and A:

$$G_m = \left(\frac{g_{d,P1}}{g_{m,P1} g_{m,P2}} + R_i \right)^{-1}, \tag{2.1}$$

$$A = \frac{(W/L)_{P6}}{(W/L)_{P2}}, \tag{2.2}$$

$$Z_m = R_o, \tag{2.3}$$

$$A_v = G_m A Z_m = \frac{(W/L)_{P6}}{(W/L)_{P2}} \frac{R_o}{\frac{g_{d,P1}}{g_{m,P1} g_{m,P2}} + R_i}, \tag{2.4}$$

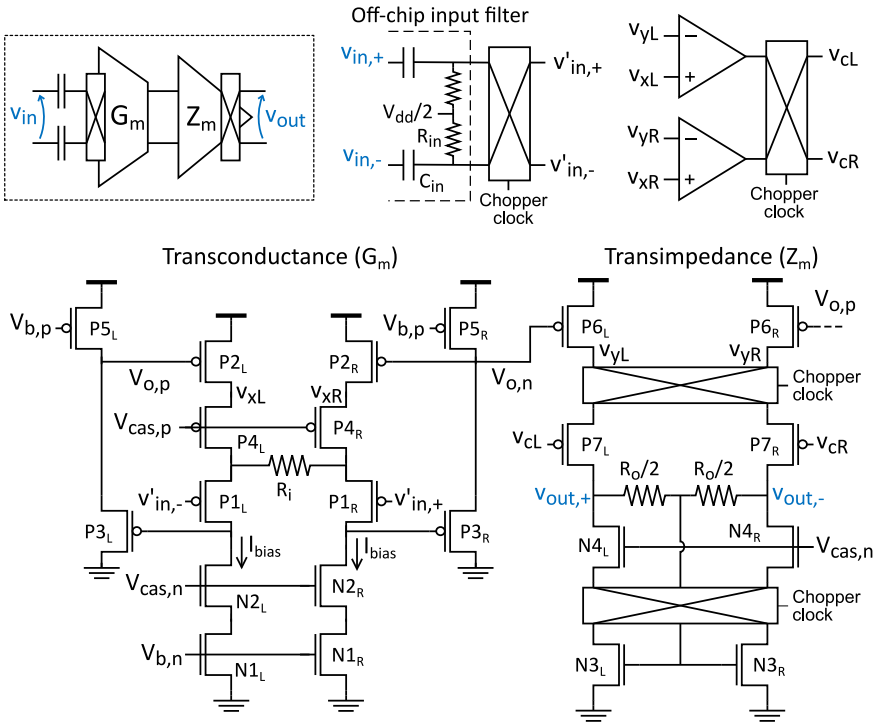

Fig. 2.6 Instrumentation amplifier circuit. In blue, the inputs v_{in} are connected to an off-chip high-pass filter and the output v_{out} are connected to the ADC. The DC bias voltages $V_{b,p/n}$ and $V_{cas,p/n}$ are generated by a bias generation circuit from a reference current originating from the power management unit.

with g_m and g_d the transistor transconductance and output conductance, respectively.

The input of the IA is AC-coupled using an off-chip RC high-pass filter (HPF) to remove the low frequencies and near-DC differential electrode offset. Chopper modulation is used to reduce the $1/f$ noise which dominates in the bandwidth of the ECG signal. High common-mode rejection is achieved by a fully differential architecture with common-centroid layout, as well as chopper modulation. Feedback operational amplifiers are used to match the V_{ds} voltages of the P6 and P2 transistors.

The power consumption and input-referred noise (IRN) are optimized using the genetic sizing algorithm described in Section 3.4. The optimal

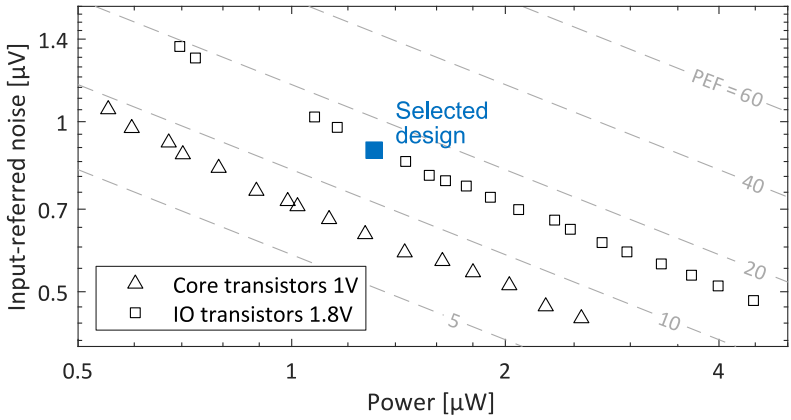

Fig. 2.7 Power-noise trade-off Pareto front in the IA. Optimal points each represent a set of sizing parameters obtained by the genetic algorithm. The dotted lines show constant-PEF contours at a 1.8-V supply.

power-noise trade-off obtained by the algorithm is shown in Fig. 2.7. A higher supply voltage benefits the input dynamic range of the amplifier, but requires the use of thick-oxide IO transistors. By sizing the P3 transistors for a larger V_{gs}, the available range increases from 0.66 V at a 1-V supply to 1.55 V at 1.8 V as shown in Fig. 2.8. This helps avoid saturation of the G_m in the presence of large common-mode external artefacts, e.g. due to powerline interference or patient motion. Compared to a design at a 1-V supply, the optimal Pareto-front in Fig. 2.7 is shifted to the right at a 1.8× higher power consumption corresponding to the change in supply. Only a limited degradation in the noise performance arises from the use of IO transistors with a +13% increase in power efficiency factor (PEF). The power loss due to the higher supply voltage is also mitigated by avoiding power conversion losses in the CPMU.

As shown in Fig. 2.7, the chosen design point corresponds to an IA power of 1.28 µW and an IRN of 0.9 µV RMS. The noise efficiency factor (NEF) and power efficiency factor (PEF) are commonly used metrics to evaluate the noise-power trade-offs in amplifiers and are defined as

$$\mathrm{PEF} = \mathrm{NEF}^2 \times V_{dd} = \frac{2v_n^2 P_{tot}}{V_T 4k_B T \pi BW} \tag{2.5}$$

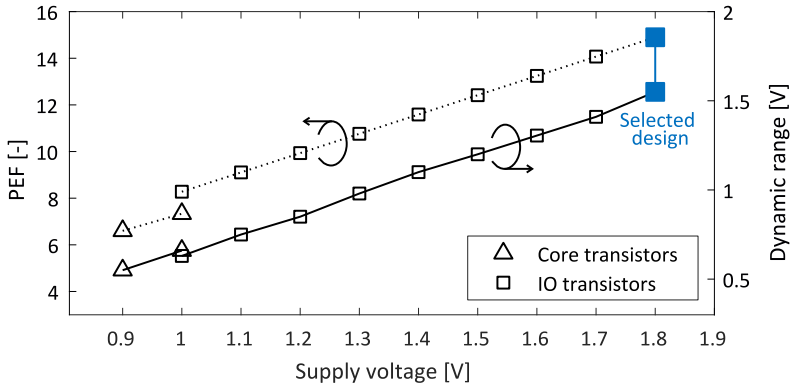

Fig. 2.8 Input dynamic range of the IA. Input dynamic range is limited by the V_{gs} of P3 and scales linearly with the supply voltage.

with V_{dd} the supply voltage, v_n the RMS IRN, k_B the Boltzmann constant, P_{tot} the total power of the IA, T the temperature and BW the noise evaluation bandwidth. The proposed amplifier achieves an NEF of 3.1 and a PEF of 15.3. The noise and power are dominated by the input G_m stage with a power consumption of 1.2 µW and 0.86 µV IRN. The noise is mostly due to the N1 (106µm/5.1µm), P1 (380µm/1.7µm), and R_i (48kΩ) devices with respectively 41%, 31%, and 12% of the total IA noise power. An additional 100 nW is consumed in the bias generation circuit for generating V_{cas} and V_b, which can be shared across several IAs in the case of a multi-channel AFE. The on-chip clock and current reference in the CPMU has a power of 520 nW and is used for the IA and several modules in the SoC. The design functionality has been verified for robustness against local mismatch and across process, voltage and temperature corners.

2.4.2 Time-based ADC

The differential signal at the output of the IA is digitized by two single-ended time-based ADCs. As shown in Fig. 2.9, a voltage-controlled ring oscillator (VCRO) performs the voltage-to-time domain conversion. The phase of the oscillator is integrated by a counter and successive samples from the counter are differentiated to obtain the frequency information. The main advantage of time-based ADCs is that they can be implemented as digital-only modules using foundry standard cells resulting in a limited design effort and low area. ULV operation at 0.4 V is used to reduce the

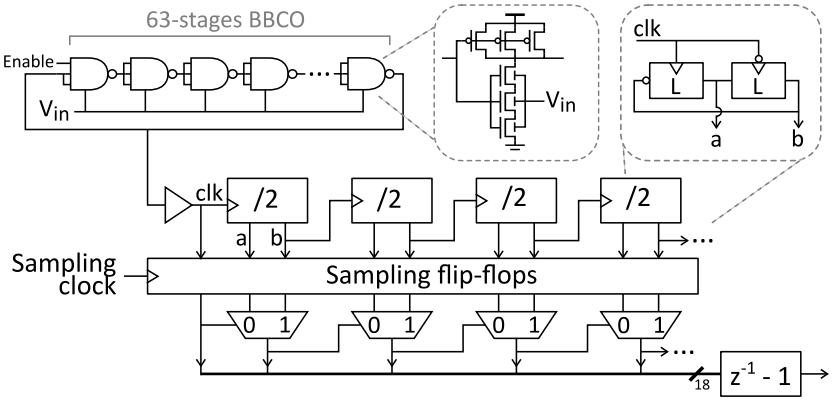

Fig. 2.9 Time-based ADC circuit. Oscillating frequency of the back-bias controlled oscillator is controlled by the input voltage V_{in}. The phase of the BBCO output clock is integrated by an asynchronous counter.

energy per cycle in the VCRO and the switching power in the counter.

Assuming a linear VCRO transfer function with gain K_{vcro} and free-running frequency f_0, the output digital signal of the ADC is equal to:

$$y[n] = \left\lfloor \left(\int_{(n-1)T_s}^{nT_s} f_0 + K_{vcro} V_{in}(t)\, \mathrm{d}t \right) + e_q[n-1] \right\rfloor \tag{2.6}$$

where V_{in} is the ADC input voltage, T_s is the sampling period and $e_q[n-1]$ is the quantization error from the previous sample, which leads to intrinsic first-order noise shaping as the VCRO is not reset. Additionally, due to the absence of a sample and hold, the integrating property of the VCRO also leads to low-pass filtering of the input signal, which helps for anti-aliasing [80].

The main bottleneck of VCRO-based ADCs, especially when operating at ULV in the near-threshold region, is the non-linearity of the voltage-to-frequency conversion. A typical current-starved ring oscillator (CSRO) has a strongly non-linear transfer function which strongly limits the input dynamic range as illustrated in Fig. 2.10. Previous works have dealt with this issue by adding a feedback loop to reduce the input swing of the signal [81]. In this work, we take advantage of the strong back-gate effect in FD-SOI technology to create a back-bias controlled oscillator (BBCO) that offers a much larger input dynamic range and significantly improves lin-

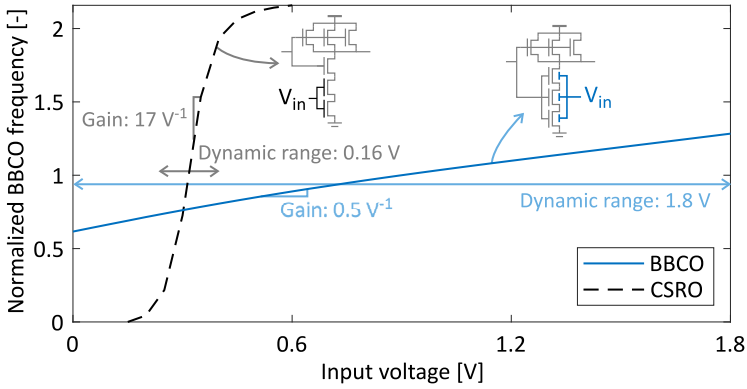

Fig. 2.10 BBCO and CSRO voltage-to-frequency transfer functions. The ROs are each built using 3-input NAND standard cells. For the BBCO, the input is connected to the NMOS back-bias. For the CSRO, the input is connected to the gates of the two bottom NMOS. The dynamic range highlights the range with a distortion below 10%.

earity [82]. By connecting the back-gate bias to the input voltage of the ADC, the threshold voltage V_{th} of the transistor can be reduced, which changes the current in each stage and the frequency of the RO. The STMicroelectronics 28-nm FD-SOI technology used in this work offers different V_{th} options that can be operated in forward back biasing (FBB, for LVT transistors) or reverse back biasing (RBB, for RVT transistors). Note that LVT PMOS and RVT NMOS devices take negative back-bias voltages while LVT NMOS and RVT PMOS take positive ones [83]. As the input voltage ranges from 0 to 1.8 V, only the LVT NMOS and RVT PMOS back-gates can be driven while the other transistors are kept at their nominal back-gate voltage.

VCRO inverting stages can be implemented with high fan-in gate types (such as NOR or NAND gates) whose inputs are short-circuited. Having more NMOS transistors in series has the effect of degreasing the total W/L and increasing the falling edge delays. The VCRO oscillation frequency is therefore made more dependent to NMOS delays compared to PMOS delays, which increases the sensitivity to the back-bias and the K_{vcro} gain. This improved gain happens however at the cost of increased non-linearity, measured here by the total harmonic distortion (THD)[1]. Fig. 2.11 shows

[1] As THD is measured with a single-tone waveform, it does not fully represent the impact of distortion on a real ECG signal with more complex time and frequency characteristics. It

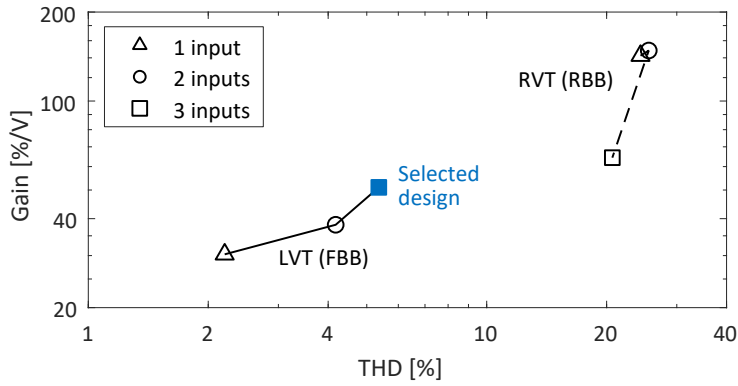

Fig. 2.11 BBCO gain-distortion trade-off for different standard cells. The 1-input gate is an inverter. The 2 and 3-input cells are NAND and NOR gates for LVT and RVT respectively. The gain is normalized to the mid-range frequency. The THD is measured assuming full-scale 1.8-V input signal.

that RBB provides a better gain but introduces a lot of distortion. In the LVT gates, the 3-input NAND is chosen as it provides a sensitivity almost twice higher than the simple inverter. The resulting THD of 5.4% (measured at full-scale input swing of 1.8 V) could be further attenuated by digital post-processing. It is however deemed acceptable in this case as it does not impact the inference performance of the DBE.

VCRO-based ADCs are also susceptible to gain variation due to PVT variations. However, ensuring an exact gain is not essential as ECG signals in wearable devices are already inherently subject to amplitude variations due to e.g. electrode placement, contact quality or patient morphology. The digital classification algorithm normalizes the signal using an adaptive gain during data pre-processing. Slow variations may be additionally compensated by measuring the 0V-input oscillation frequency and scaling the output accordingly.

The edges of the output signal of the BBCO are counted using an asynchronous ripple counter with 17 divide-by-2 stages. The asynchronous double sampling technique proposed by Baert and Dehaene [84] is used to avoid an invalid result when the sampling clock edge occurs during the transition of the counter output. This is done by latching the bits in the

however conveniently provides a first-order quantification of the distortion.

counter on the positive and negative edges of the clock to ensure that at least one of them is stable at all times, which allows to select the correct one based on the value of the preceding bit after sampling.

Finally, the resolution and quantization noise of the ADC are set by the frequency of the BBCO, which depends on the number of stages, and the sampling frequency. A smaller BBCO with higher frequency improves the resolution but increases the power consumption of the counter and has more thermal noise due to poorer averaging as illustrated in Fig. 2.12. Increasing the sampling frequency with a higher oversampling ratio (OSR) improves the input-referred noise thanks to noise shaping, but requires more processing in the DBE for low-pass filtering and decimation. Additionally, choosing a sampling frequency that is an integer divisor of the IA chopper frequency filters the output ripple thanks to the integrating effect of the counter. A too high OSR thus increases the bandwidth constraints on the IA for the signal modulated at the chopper frequency. The impact of the power consumption in the sampling flip-flops is negligible at the low frequencies considered in this work. The BBCO is implemented with 63 stages for a K_{vcro} of 12.1 MHz/V and an OSR of 4 is chosen. The resulting ADC has a resolution of 14 bits with a differential non-linearity below 0.4 LSB. The integral non-linearity goes up to 800 LSB, which is due to the uncompensated non-linearity of the BBCO.

2.5 Digital back-end software design

The data samples retrieved from the AFE are processed sequentially in the DBE. First, the heartbeats are detected by the algorithm from Pan and Tompkins which consists in a band-pass filter, a derivation, a moving average and an adaptive threshold [71]. This allows to detect the sharp peaks in the signal that likely correspond to the QRS complexes of the ECG signal shown in Fig. 2.13. Upon detection, feature extraction and classification are performed to assign a label to each individual beat. The 3 classes considered in this work are the normal heartbeats with sinus rhythm (N), supraventricular ectopic beats (S), and ventricular ectopic beats (V). Examples of these three classes are shown in Fig. 2.13. Compared to [74], the fusion beats (F) class is joined with the V class and the unknown beats (Q) class is discarded, as those two classes represent a really low percentage of the total dataset (resp. 0.8% and 0.02%) and may not be representative of the true performance of the algorithm.

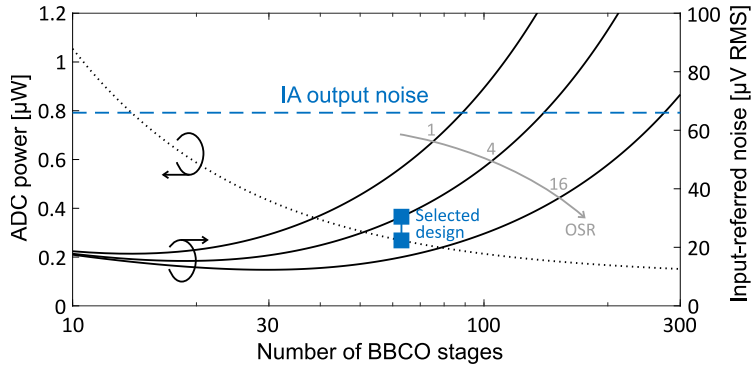

Fig. 2.12 Impact of the BBCO length on the ADC performance. The power consumption (left) includes BBCO and counter power consumption. The IRN (right) includes the quantization noise and BBCO thermal noise. The dotted line shows the noise at the input of the ADC originating from the IA.

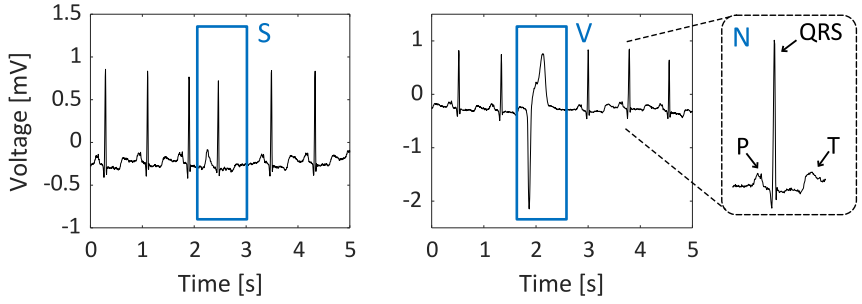

Fig. 2.13 ECG arrhythmia class examples. The supraventricular ectopic beat (S) is characterized by its prematurity and the lack of a P-wave (masked here by the T-wave of the previous beat). The ventricular ectopic beat (V) is characterized by its prematurity and a broad irregular QRS complex.

When using an MCU as a basis for the arrhythmia classification system, the detection and classification algorithm is programmed in the SRAM and executed in software. Sections 2.5.1 and 2.5.2 describe the algorithm design with the feature selection process and SVM classifier, respectively. Section 2.5.3 details the real-time execution constraints of the DBE software execution.

2.5.1 Feature selection

As observed in the examples of Fig. 2.13, while both arrhythmia types show a disruption in the normal sinus rhythm, their characteristics are quite different. The V beats differ mostly by the shape of their QRS complex. The S beats are morphologically close to the N beats with only a missing P-wave, which is often hard to observe in the baseline noise. For classification purposes, the features used to best distinguish those classes must therefore be different in each case, which is rarely done when using traditional one-against-one or one-against-all SVM classifiers. Park *et al.* [85] suggested using a hierarchical SVM (i.e. separating V beats from N/S beats with a first classifier, then a second N-versus-S classification) with different features sets to alleviate this problem. However, while they chose the features manually, we perform automated feature selection separately for each classifier stage. As summarized in Table 2.2, the initial set includes features commonly used in the literature [32], resulting in a feature vector of 121 elements for each beat before feature selection. The RR interval refers to the time duration between successive beat locations surrounding the current beat. The 20-sample average RR interval and the differences to the average are also used to highlight the heart rate variability. The morphological information of the heartbeat is extracted using downsampled raw signal samples at the location of the P and T waves and the QRS complex. Spectral information is obtained using the coefficients of a discrete wavelet transform (DWT) obtained with a Daubechies 3 (*db3*) 4-level decomposition.

The feature selection process is performed by a sequential forward floating search (SFFS) algorithm which iteratively adds or removes a feature from the set based on the improvement of the cross-validation inference performance [76]. This performance is measured by the j index (I_j) equal to the mean of the sensitivities (Sen) and positive predictive values (PPV) for the S and V classes, i.e. $I_j = (Sen_S + Sen_V + PPV_S + PPV_V)/4$ [76]. This arbitrary figure of merit, normalized here to 100%, avoids the bias due to the large proportion of N beats and focuses on the discriminative performance for the S and V arrhythmia classes. Figure 2.14 shows the performance evolution with the number of features, where it can be observed that a low number of features below 10 is already sufficient to achieve peak performance. As expected, the selected feature types are different for each classifier as shown in Table 2.2. Despite feature selection, a significant overlap between classes is still present due to inter-patient variability, as seen

Table 2.2 Number of features before and after selection. For each feature category and type, the initial number of features before selection is compared to the number of features selected for the two individual SVM in the hierarchical classifier.

Category	Name	Description	Initial	N/S v. V	N v. S
Intervals	RR_i	RR, i.e. beat-to-beat, intervals (two preceding and one following)	3	0	2
	RR_{avg}	Average of last 20 RR_i intervals	1	0	0
	RR_i - RR_{avg}	Difference between local intervals RR_i and the average RR_{avg}	3	0	2
Morphological	P_i	Regularly spaced samples from 200 to 25 ms before the QRS peak	12	0	0
	QRS_i	Regularly spaced samples from 12 ms before to 80 ms after the QRS peak	12	1	0
	T_i	Regularly spaced samples from 120 to 320 ms after the QRS peak	6	1	0
Spectral	$DWT_{L2\ det,i}$	DWT decomposition level-2 detail coefficients (approx. 31 to 63 Hz)	35	2	0
	$DWT_{L3\ det,i}$	DWT decomposition level-3 detail coefficients (approx. 16 to 31 Hz)	21	4	2
	$DWT_{L4\ det,i}$	DWT decomposition level-4 detail coefficients (approx. 8 to 16 Hz)	14	2	2
	$DWT_{L4\ approx,i}$	DWT decomposition level-4 approximation coefficients (approx. 0 to 8 Hz)	14	0	1

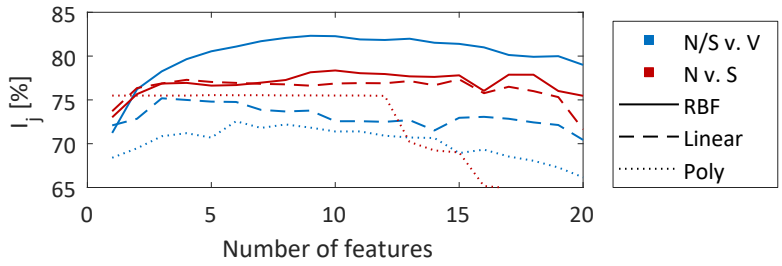

Fig. 2.14 SVM classifier cross-validation performance for different kernels. The j index (I_j) is measured separately for the "N/S v. V" and "N v. S" classifiers separately. The reported I_j is the best performance obtained for a given number of features along the SFFS process for a given classifier kernel. The corresponding set of optimal features may be different for each classifier.

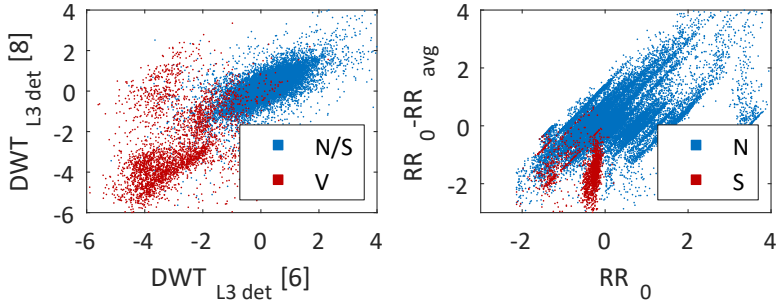

Fig. 2.15 Data distribution plots. The two best features for each classifier according to the SFFS algorithm are displayed.

in Fig. 2.15 for the two most discriminative features in each classifier.

2.5.2 Support-vector machine classifier

The choice of kernel for an SVM significantly impacts the separability of the data and the classification performance. Three different kernels are compared in Fig. 2.14. The RBF kernel performs better than the linear and polynomial kernels, especially for discriminating ventricular arrhythmias. As the classes are not fully separable, a low regularization parameter allows some misclassifications and reduces overfitting.

After training, the SVM model with RBF kernel contains 20837 support vectors (SVs) due to the low separability and low regularization parame-

ter. This corresponds to a memory storage requirement of 149.9 kB which is an issue for the SleepRider SoC with a maximum of 48 kB of program memory. This large number of SVs directly impacts the power consumption of the DBE as an exponential function needs to be calculated for each of them. Pruning the model to reduce its complexity is often possible with only a limited impact on the performance. Geebelen *et al.* [77] proposed the separable case approximation (SCA) method which consists in removing misclassified training examples from the SV set. Applied on the ECG arrhythmia classification model, the number of support vectors is strongly reduced to only 247 SVs, with 1.8 kB of memory space, while I_j decreases by only 0.5%.

2.5.3 Real-time execution

Software execution is performed on an ARM Cortex-M4 CPU. The MCU remains in sleep mode most of the time while samples are acquired by the AFE and stored in the SRAM by the DMA. A DMA cycle corresponds to 128 samples (i.e., 0.5 s) acquired and stored in memory, after which the CPU wakes up and processes these last samples sequentially. The processing execution time for the algorithm needs to remain short enough to meet the real-time constraint, as well as minimize the power consumption in active mode. The time spent on each subtask is illustrated in Fig. 2.16 with a clock running at 64 MHz. The detection algorithm is executed for each sample and takes an average of 500 cycles/sample, which includes 190 cycles for the QRS-enhancement filters. The feature extraction and classification only occurs when a beat is detected. The feature extraction takes 26 200 cycles, dominated by the DWT which lasts 17 200 cycles. The SVM classification takes 280 cycles per SV, or 69 800 in total. The number of classifications in each 128-sample DMA cycle depends on the heart rate. Assuming a worst-case scenario of 3 beats detected in a single DMA cycle, i.e. 240 bpm, the execution time is limited to 352 000 cycles. Fig. 2.16 also shows that the reduction in the number of SVs achieved with pruning also significantly decreases the execution time. Thanks the presence of ABB that switches the logic and SRAM between high-leakage fast FBB state in active mode and low-leakage zero back bias in sleep mode, the leakage overhead incurred by a higher clock frequency in active mode is limited. With a 64-MHz clock, the execution is achieved in 5.5 ms, which corresponds to a duty-cycle of only 1%. This limited time spent in active mode also provides room for higher processing workloads with more channels or higher

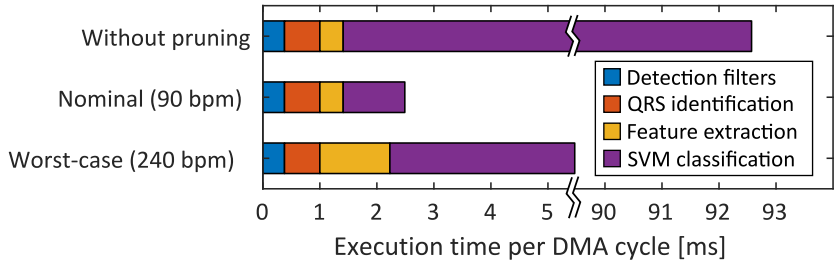

Fig. 2.16 Software execution time. The software is executed on a Cortex-M4 CPU at 64 MHz. Nominal and worst-case scenarios, respectively 90 and 240 bpm, correspond to maximum 1 and 3 detected beats per DMA cycle. One DMA cycle lasts 0.5 s.

data rates.

2.6 Results

The SleepRider SoC was prototyped in 28-nm FD-SOI on a 3.1-mm^2 chip with the core taking up 0.94 mm^2 as shown in the microphotograph of Fig. 2.17. Measurements are performed on a custom test PCB with off-chip components such as the input HPF and a patient reference circuit. The setup in Fig. 2.18 uses a Keysight N6715B power supply, a Tektronix MSO56 oscilloscope, and a laptop for JTAG programming and UART communication. In-vivo trials are performed with single-use Ag/AgCl adhesive electrodes with solid gel. Measurements were performed on one healthy male subject to test the functionality of the acquisition and detection.

2.6.1 Microcontroller power consumption

When tested on standard benchmarks to assess plain computing resources, the measurements and characteristics in Table 2.3 show the highest level of computing performance and integration compared to other ULP MCUs, with low power in sleep and active modes. The logic and SRAM minimum energy point is 4.8 µW/MHz at 56 MHz (CoreMark). This results in a peak energy efficiency similar to our previous ULP MCU [JP2] while providing 1.7× higher computing performance. In Table 2.3, the proposed ULP MCU has among the highest computing performance, as measured on the CoreMark testbench, thanks to its higher clock frequency. While

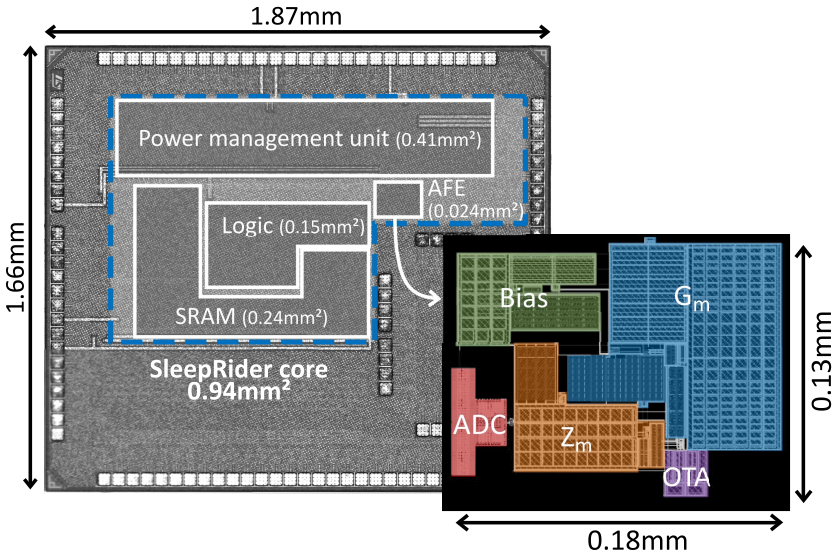

Fig. 2.17 SleepRider chip microphotograph and AFE layout. Unlabeled parts of the chip are unrelated to this work.

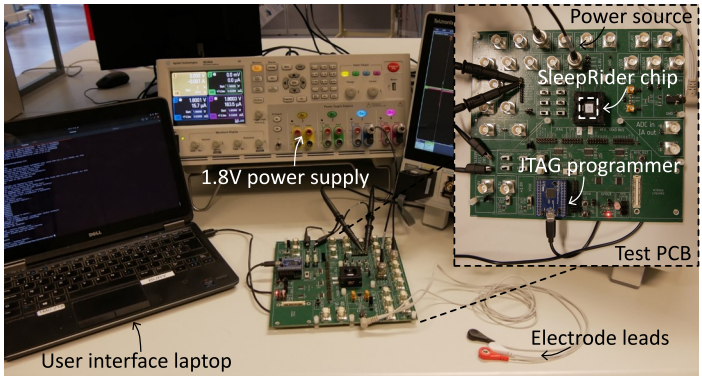

Fig. 2.18 System measurement setup with custom test PCB.

this performance allows to operate with a low duty-cycle as shown in Section 2.5.3 thanks to the low-resource algorithm, the MCU could have been designed for a lower target frequency, which would have likely improved the energy efficiency. It also provides the possibility to execute other tasks on the same hardware.

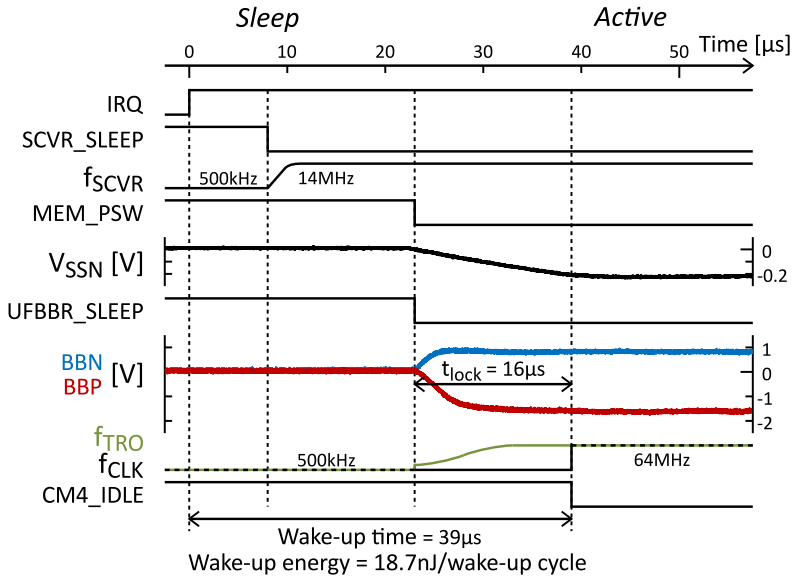

Fig. 2.19 PMU wake-up sequence measurement. The blocks of the system wake up successively after the interrupt request (IRQ). The SCVR start ($SCVR\_SLEEP$) and its frequency f_{SCVR} increases to boost its power output. The memory power switch is closed (MEM_PSW) and the $VSSN$ negative voltage required for writing is generated. The UFBBR starts to generate the back bias voltages and the high-frequency clock (f_{TRO}). Finally, the Cortex-M4 CPU exits the sleep mode ($CM4\_IDLE$).

Sleep power is 5.8 µW with 32-kB memory retention, excluding SCVR conversion losses. Each wake-up cycle from sleep to active mode takes 39 µs and has an energy overhead of 18.7 nJ as illustrated in Fig. 2.19.

2.6.2 AFE signal acquisition

The measured IA shows a differential gain of 37 dB and a maximum -3-dB bandwidth of 1 kHz. The transfer function in Fig. 2.20 shows a flat response in the ECG signal bandwidth. The common-mode attenuation measured with a 300-mV_{pp} input signal stays below 40 dB over the whole bandwidth and increases down to 90 dB at lower frequency. This corresponds to a common-mode rejection ration (CMRR) above 80 dB. The

Table 2.3 Ultra-low-power MCU comparison.

	This work	Ambiq Micro [86] Apollo3 Blue	Myers [87] VLSIC 2017	Prabhat [88] ISSCC 2020	Lallement [89] SSCL 2018	Lee [90] JSSC 2020	Song [42] TBioCAS 2019	Bol [JP2] JSSC 2021
Technology	28nm FD-SOI	40nm ULP eFlash	65nm LP	65nm LP	22nm FD-SOI	55nm DDC	55nm CMOS	28nm FD-SOI
CPU	CM4	CM4F	CM0+	CM33	CM0+	CM0	CM4F	CM0
Memory	96kB SRAM	384kB RAM 1MB Flash	**2kB ROM 12kB SRAM**	128kB ROM 20kB RAM	**12kB SRAM**	**8kB SRAM**	192kB SRAM	64kB SRAM
Embedded PCM	**All clocks Int. supplies Back bias**	Main clock Int. supplies	**All clocks Int. supplies**	**All clocks Int. supplies**	Back bias	**All clocks Int. supplies Back bias**	Main clock Int. supplies	Main clock Int. supplies Back bias
External components	**1.8V supply 3C (SCVR) 2C+2R (AFE)‡**	1.8V supply Crystal osc. 1L+2C (Buck)	**1.2V supply**	**1-1.5V supply 1L+1C (Buck)**	0.4-0.8V supply Main clock	1.2/2.4/-1.2V supplies	1.4V supply Crystal osc. 6C (LDO/VREF) 2C (AFE)	1.8V supply Ref. clock 3C (SCVR) **1C (UFBBR)**
PVT compensation	UFBBR	N/A	AFS	UFVR	**Limited ABB**	AVS + ABB	✗	UFBBR
AFE	**IA + ADC**	ADC	✗	✗	✗	✗	**(IA+ADC)×8ch.**	✗
Max freq. at MEP supply [MHz]	80	96	**0.2**	**2.35**	20	**0.5**	80	80
Active power at MEP [µW/MHz]	**5.5 at 56MHz**	32.8* at 48MHz	7.6 at 0.2MHz	20 at 2.35MHz	**1.13 at 20MHz**	**4.7 at 0.5MHz**	23† at 80MHz	**3.5 at 48MHz**
Wake-up time [µs]	39	25	N/A	**180**	0.3	N/A	N/A	20
Lowest power state with retention* [µW]	7.7 (32kB) 9 (96kB)	1.8 (8kB) 6.7 (384kB)	0.14 (4kB)	0.01 (4kB)	N/A	0.7 (8kB)	N/A	12.7 (64kB)
Peak efficiency [CM/mW]	**622 at 192 CM**	104 at 164 CM	324 at 0.5 CM	201 at 9.4 CM	**✗(not enough RAM)**	495 at 1.2 CM	149 at 273 CM	**666 at 112 CM**

* Including DC/DC conversion losses. † Digital only. ‡ Optional.
Red and blue fonts highlight MCU limitations and top performance respectively.
AFS: adaptive frequency scaling - AVS: adaptive voltage scaling - UFVR: unified frequency and voltage regulation - CM: CoreMark

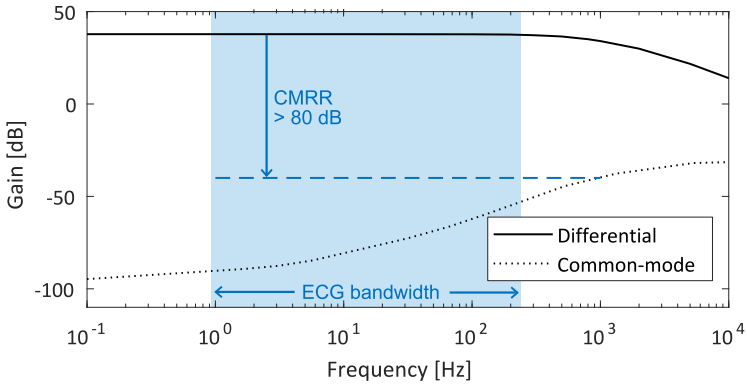

Fig. 2.20 IA transfer response. The differential and common-mode gains are measured with 5 mV_{pp} and 300mV_{pp} amplitudes respectively. The measurements were performed without the off-chip input high-pass filter.

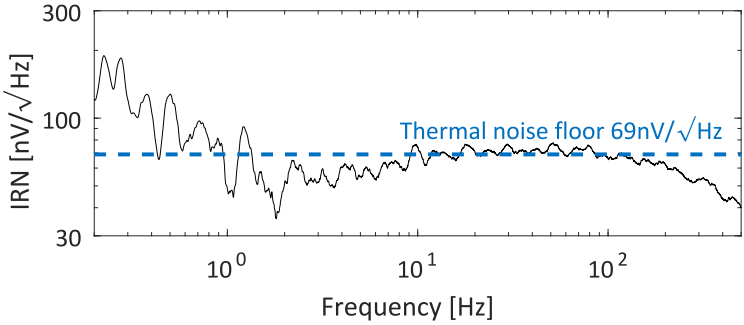

Fig. 2.21 IA noise spectral density. The flicker noise corner is reduced to 1 Hz thanks to chopper modulation and the total noise is minimized using numerical optimization.

measured input-referred noise spectral density of the IA[2] in Fig. 2.21 is at 90 $nV/\sqrt{Hz}$ with a corner frequency below 1 Hz thanks to chopper modulation. The resulting IRN is 0.75 µV RMS integrated over the 1-to-100-Hz frequency range. A sample signal measured in-vivo through chest electrodes is shown in Fig. 2.22. An electrode distance of 8 cm is chosen as it is the limit before QRS amplitude starts to decrease.

[2]The noise from the electrodes is not taken into account. In the case of wet electrodes with low impedance, the intrinsic noise of the IA dominates. However, for dry electrodes, the higher impedance of the skin-electrode interface can lead to noise levels above 1 µV [91].

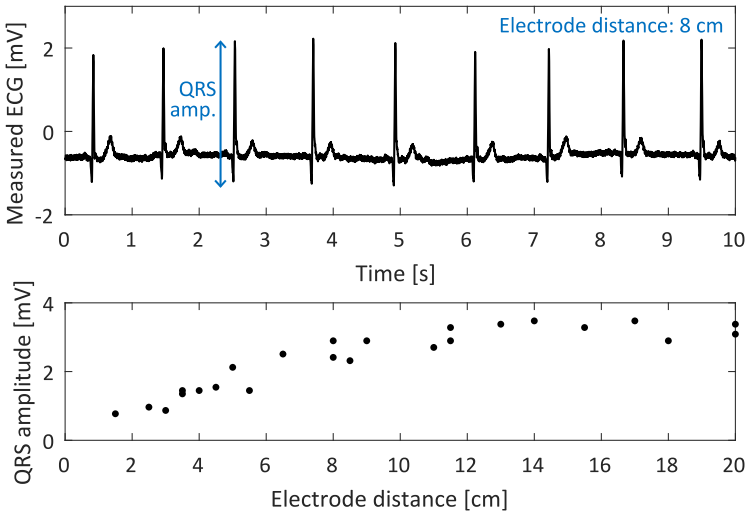

Fig. 2.22 Acquired input-referred ECG signal sample from on-chest electrodes. The bottom graph shows QRS amplitude for measurements on the same person with multiple placements on the chest using the MLII orientation.

Each single-ended ADC consumes 0.36 µW for a resolution of 14 bits with an OSR of 4. Unfortunately, the noise in the ADC was higher than expected because of an underestimation of the jitter due to flicker noise in the BBCO, which was confirmed in simulation. This caused an additional 490 µV of noise referred to the ADC input in the ECG bandwidth which limits the SNR to 62 dB. This issue could be easily avoided by removing the chopper demodulation switches at the output of the IA and performing the demodulation in the digital domain.

The main performance metrics of the AFE are summarized in Table 2.4 in comparison with examples from the state of the art. The IA achieves among the best power-noise trade-offs with an NEF of 2.3. The silicon area is also much lower than other works, especially thanks to the all-digital time-based ADC in a 28-nm technology taking up only 900 μm^2.

2.6.3 DBE inference

Given the long duration of the MIT-BIH arrhythmia database [75], inference performance on the full dataset can hardly be evaluated on the SoC and is thus emulated off-line. In order to take into account the impact of

Table 2.4 Low-noise AFE comparison

	This work	Mondal [68] TBioCAS 2021	Yang [92] TBioCAS 2021	Hsu [93] JSSC 2020	Xu [94] TBioCAS 2018	Zeng [95] JSSC 2019	Pochet [81] TBioCAS 2021	Park [96] JSSC 2022
Technology	28nm FD-SOI	65nm CMOS	0.13µm CMOS	0.18µm CMOS	0.18µm CMOS	0.13µm CMOS	65nm CMOS	0.11µm CMOS
Supply [V]	0.4 / 1.8	0.8	1.2	1.8	1.2	1	0.8 / 1.2	1 / 1.5
Area [mm^2]	**0.024**	0.8	0.53	0.72	1.1	**0.04**	**0.075**	0.75
Power [µW]	**2.1**	**0.77**	**4**	5.3	33	7.5	5.8	**3.8**
IA gain [dB]	37	22 - 43	20 - 46	42 - 50	38	29	-	61 - 74
IA bandwidth [Hz]	1k	250	125	1k	200	300	1k	380
THD [%]	0.3 @10mV$_{pp}$	1.3 @11mV$_{pp}$	-	0.01 @12mV$_{pp}$	-	0.2 @4mV$_{pp}$	-	0.22 @2.5mV$_{pp}$
IA IRN [µV$_{RMS}$]	**0.75** (1-100)	1.9 (1-250)	**0.4** (0.1-100)	1.2 (1-100)	**0.8** (1-100)	1.4 (0.5-150)	3.5 (1-1k)	**0.36** (0.5-300)
NEF / PEF	**2.3** / 9.8	4.4 / 15.5	**1.1** / 1.5	8 / 115	16.1 / 311	12.4 / 154	6.8 / 55	**1.6** / 2.5
CMRR [dB]	>80	82	84	100	100	112	100	92
Input impedance [MΩ]	>50	105	-	-	-	20	50	2000
Max. sample rate [Hz]	32k	1k	250	2.5k	5M	-	200k	-
ENOB [bit]	10 (**14**[a])	8.8	**16**	9.2	10.4	**16**	**15**	10
Conversion energy [fJ/bit]	**22**	81	208	1172	**6.2**	-	**0.5**	-

Blue color highlights designs with an area below 0.1 mm^2, power below 5 µW, IA IRN below 1 µV, NEF below 4, ADC resolution above 12 bits and conversion energy below 50 fJ/bit.

[a] Without the BBCO flicker noise, e.g. if removed by chopper modulation.

Table 2.5 Beat-to-beat arrhythmia classification algorithms. All the works perform beat-to-beat arrhythmia classification using similar classes (with optional F and Q classes) and use the MIT-BIH database for training and testing.

	This work	Villa [97] EMBC 2019	Chen [98] EMBC 2020	Mondéjar [99] BSPC 2019	Guo [100] BBE 2019	Liu [39] ISSCC 2021	Wu [40] IEEE Access 2019	Tang[69] TBioCAS 2021
Training	**Inter-patient**	**Inter-patient**	**Inter-patient**	**Inter-patient**	**Inter-patient**	Patient-specific	Patient-specific	Patient-specific
Hardware	MCU	✗	✗	✗	✗	ASIC	ASIC	FPGA
Classes	3	5	4	4	5	5	5	3
Classifier	SVM	SVM	NN	SVM	NN	NN	NN	NN
Features	RR, DWT, Samples	RR, VMD	RR, Samples	RR, DWT HOS, Morph	Samples	Morph, Samples	Samples	RR, Morph
Performance	Sen_S: **82.6** PPV_S: 34.4 Sen_V: 88.9 PPV_V: 80.8	Sen_S: **83** Spe_S: 29 Sen_V: 88 Spe_V: 84	Sen_S: 64.4 PPV_S: 66.4 Sen_V: 90.7 PPV_V: 85.5	Sen_S: 78.1 PPV_S: 49.7 Sen_V: **94.7** PPV_V: **93.9**	Sen_S: 62.7 PPV_S: 61.2 Sen_V: 91.3 PPV_V: 88.3	Acc: 99.2	Sen_S: 78.4 PPV_S: **84.4** Sen_V: 92.1 PPV_V: **95.5**	Sen_S: 88.8 PPV_S: **76.1** Sen_V: **95.1** PPV_V: **95.2**

Blue color highlights highest performance results for each indicator.
NN: neural network - RR: heartbeat intervals - Samples: raw signal samples - Morph: morphological indicators - VMD: value mode decomposition - HOS: higher order statistics.

Table 2.6 Classification confusion matrix, on testing set DS2 from the MIT-BIH database.

		Label					
		N	S	V	F	-	PPV [%]
Result	N	40597	244	10	263	4	98.7
	S	2748	1509	118	7	19	34.4
	V	684	74	3074	117	99	80.8
	-	13	4	2	1		
Sen [%]		92.2	82.6	88.9			

analog non-idealities on the inference performance, a Python model simulates the noise and transfer response characteristics of the IA and ADC before performing the classification. The characteristics of the AFE model were matched to the measured hardware performance. Execution directly on the SoC is validated using shorter signal samples.

As recommended in [74], 22 30-minute records (DS1) are used for training and validating the SVM while 22 other records (DS2) are used for testing. Leave-one-out cross-validation on the DS1 records was used for algorithm training and validation. The confusion matrix resulting of the DS2 set classification is shown in Table 2.6. This respectively corresponds to a sensitivity and positive predictive value of 82.6% and 34.4% for the S class and 88.9% and 80.8% for the V class ($I_j = 71.7\%$).

The largest source of misclassifications are normal beats classified as supraventricular beats which degrades PPV_S. Most of these (2081 beats) arise from atrial fibrillation (AF) episodes, which are not labeled as supraventricular arrhythmias despite presenting a disruption of the regular sinus rhythm. AF segments can be detected with coarse resolution using heart rate variability [101], but it is not included in this work.

This work is the first implementing an inter-patient arrhythmia classification algorithm on an MCU. The performance obtained is comparable to other works using the inter-patient training scheme, as shown in Table 2.5.

2.6.4 System performance

When performing arrhythmia classification, the SleepRider SoC consumes 13.1 μW including DC/DC conversion losses from the 1.8-V supply. This

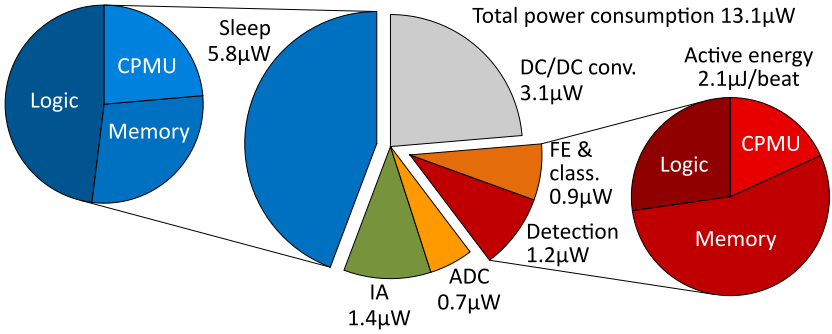

Fig. 2.23 Power breakdown of the SleepRider SoC during arrhythmia classification. The center pie chart shows the distribution of the average power among different tasks. The left and right pie show the distribution among different modules in sleep and active modes. Wake-up energy is negligible.

is lower than the value of 19.8 µW we reported in [CP3], thanks to the reduced wake-up overhead enabled by the use of the DMA and reduced power in the AFE with a different configuration. The average power distribution among different tasks and modules is illustrated in Fig. 2.23. Sleep power dominates the average power consumption due to the very low duty-cycle ($< 1\%$), despite the zero back biasing for the logic. Other methods such as power gating could be used to reduce the leakage, at the cost of a longer wake-up latency. The active power of the DBE accounts for only 21% of the total energy thanks to the high energy efficiency of the SleepRider MCU which achieves 4.8 nJ/sample for the detection and 0.9 µJ/beat for the classification.

Compared to an implementation of the cardiac arrhythmia classification software on a commercial ULP MCU (Apollo3 Blue from Ambiq Micro [86]), the proposed system achieves a 4.2× reduction of the average power consumption. When comparing only the power consumption in the AFE and DBE related to signal acquisition and processing, the reduction factor reaches 9.8×. While the SleepRider SoC displays a much higher energy efficiency in active mode, the sleep mode power consumption is lower in the Apollo3 Blue MCU, thanks to power gating of the idle blocks which is not available in the proposed system.

Table 2.7 compares several mixed-signal ECG SoCs from the state of the art. The SleepRider SoC integrates all necessary components for ECG sig-

Table 2.7 ECG mixed-signal SoC comparison

	This work[e]	Mondal [68] TBioCAS 2021	Bose [65] JSSC 2020	Lee [66] SSCL 2020	Deepu [70] TCAS-II 2018	Chen [43] JSSC 2015	Tang [69] TBioCAS 2021	Schönle [41] JSSC 2018	Song [42] TBioCAS 2019
Technology	28nm FD-SOI	65nm CMOS	0.18µm CMOS	28nm FD-SOI	0.35µm CMOS	65nm CMOS	0.18µm CMOS	0.13µm CMOS	55nm CMOS
Supply [V]	1.8	0.8	0.02	1	1 / 1.5	0.6 / 0.4	1	1.5 / 3.5	2
AFE	1 × ECG	1 × ECG	1 × ECG	1 × ECG	1 × ECG	1 × ECG	1 × ECG	9 × ExG BioZ PPG	2 × ECG BioZ PPG
DBE	Cortex-M4	Off-chip FPGA	Comparators (mixed-signal)	DNN core	Detection accelerator	Cortex-M0+ + FFT & FIR	Off-chip FPGA	4-core PULP	Cortex-M4f
PMU	✓	✗	✓	✗	✗	✗	✗	✓	✓
Task	Arrhythmia classification[c]	Heartbeat detection	Heartbeat detection	Arrhythmia detection[d]	Heartbeat detection	Arrhythmia detection[d]	Arrhythmia classification[c]	-	-
AFE power	2.1µW	0.77µW	0.5µW	92nW	0.75µW	19nW	0.15µW	285µW	40µW
DBE energy	1.2µJ/detect. 0.9µJ/class.	-	55nJ/detect.[a]	30µJ/class. (1×/30s)	0.9µJ/detect.[a]	0.45µJ/class. (1×/10s)	-	33µW/MHz[b]	23µW/MHz[b]
System power	13.1µW	-	4.6µW	1µW	2.3µW	64nW	-	-	-

[a] a Assuming 100 beats/minute.
[b] No documented task.
[c] Beat-to-beat classification.
[d] Time segment classification.
[e] Results do not include the system optimization described in Chapter 4.

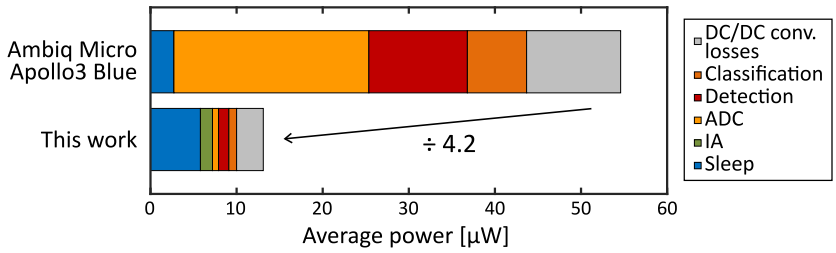

Fig. 2.24 Comparison with commercial ULP MCU [86]. DC/DC conversion losses for the Ambiq Apollo3 MCU assume an 80% power conversion efficiency across all blocks in the total measured power.

nal acquisition and processing, as well as the power and clock management. In addition to being more versatile, arrhythmia classification performed on such an ULP MCU achieves an energy efficiency close to other works while performing a much more complex task.

2.7 Conclusion

To our knowledge, this work is the first tackling end-to-end ECG beat-to-beat arrhythmia classification implemented on a ULP MCU. It aims at meeting the constraints of power consumption, signal quality and inference performance inherent to wearable monitoring devices while providing enough versatility to perform complex tasks. A prototype chip has been fabricated in 28-nm FD-SOI technology and tested with the classification software. The low-power and low-noise AFE acquires the ECG signal with 2.1 µW. The software-based DBE performs beat-to-beat arrhythmia classification with a high energy efficiency of 2.1 µJ/beat (including beat detection and classification) and an accuracy of 91.4%. The system has an average power consumption of 13.1 µW and has been tested in-vivo to validate the functionality.

The main remaining performance bottleneck of the proposed implementation is the power consumption in sleep. While the custom SleepRider MCU has the advantage of having a low wake-up time and energy overhead and full digital retention in sleep, it is not a critical factor in the arrhythmia classification use case when the DMA is used, as the wake-up frequency is only 2 Hz. The wake-up overhead would become significant in an application that requires a wake-up frequency above approximately

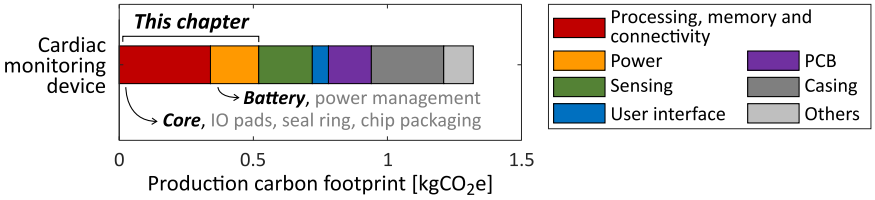

Fig. 2.25 Carbon footprint distribution for a cardiac monitoring wearable device. The respective shares of the components is estimated using the methodology from [102], similarly to Fig. 1.3. In this chapter, the design of the SleepRider SoC and optimization of the power consumption enable a reduction in the footprint of the IC for the processing and memory, and of the battery for the power supply of the device.

50 Hz, as the average power would be increased by almost 1 µW. In the presented application, a reduction of the power consumption could have been obtained with additional power gating techniques which only keep the required AFE and CPMU components active in sleep, at the cost of retention loss in the DBE logic.

The low power consumption of the system enables long-term cardiac monitoring with limited battery capacity, i.e., 9.4 mAh at 1.8 V for one month operation. By embedding the electronic sensor in a small device worn on the chest as an adhesive patch, this allows the patients to comfortably continue their daily activities. To validate the proposed system in this application, some limitations still need to be addressed. First, such a long-term monitoring device would likely require dry electrodes, compared to the wet Ag/AgCl electrodes used for system testing. The impact of the higher impedance and noise of these electrodes on the system performance therefore needs to be evaluated. Second, the cardiac arrhythmia classification performance demonstrated with our system is on par with the state of the art. However, inter-patient classification is intrinsically limited by patient-to-patient variability, which bounds the accuracy achieved by these systems. While the resulting error rates lead to uncertainty for diagnosis purposes, these devices can be used conveniently for mass screening of cardiac disorders, with a more precise secondary evaluation for identified higher-risk patients. A more accurate classification can be obtained with a patient-specific model, which induces practical constraints for model training. Acceptable levels of sensitivity and specificity for each use case need to be defined from a clinical perspective.

Table 2.8 System improvement with local signal processing. The system designed in this chapter is compared to the wireless transfer of the raw ECG signal, as discussed in Chapter 1.

	Transmission of raw signal	**Local data processing**
Technology	28nm FD-SOI	28nm FD-SOI
Embedded DBE	✗	✓
TX data rate [bps]	3000	< 4
Power[b] [μW]		
AFE	2.1	2.1
DBE	-	2.1
TX[a]	12	0.016
Total	*14.1*	*4.2*
Silicon area [mm^2]		
AFE	0.02	0.02
DBE	-	0.39
TX[a]	0.61	0.61
Total	*0.63*	*1.02*
Battery lifetime[bc] [year]	3	10
Production carbon footprint [gCO_2e]		
Integrated circuit[d]	16	25
Battery[c]	35	35
Total	*51*	*60*

[a] Assuming a Bluetooth Low-Energy TX front-end with 4 nJ/bit and 0.61 mm^2, similar to [15] in 22-nm CMOS.
[b] Sleep power and DC/DC conversion losses not considered.
[c] Continuous operation, assuming a 3-V 120-mAh coin cell battery. While a typical monitoring period is limited to one month, longer lifetime allows to reuse it across several patients.
[d] Based on [102] for bulk CMOS. Only the front-end is included, back-end adds 25 to 37% of energy consumption [103]. Silicon area does not include power and clock management, nor the overhead for pads and seal ring.

The proposed arrhythmia classification system presented in this chapter demonstrates that ULP MCUs can perform complex real-time signal processing with a very limited power budget of only a few μW's. This performance can however only be achieved thanks to careful optimization of all system domains, with application-specific low-power and low-noise AFE, ultra-low-power digital operation thanks to robust ultra-low-voltage operation with adaptive back-biasing, and resource-optimized software im-

plementation. From the overall carbon footprint for the production of a cardiac monitoring wearable system illustrated in Fig. 2.25, these circuit optimizations address the aspects related to integrated circuits for the processing, memory and connectivity, as well as the batteries. As detailed in Table 2.8, the reduction of transmitted data through local data processing optimize the battery lifetime, with a limited silicon footprint overhead. The very low footprint reported for ICs and batteries are made possible by the low power consumption and low-complexity circuits. Embedded processing requires additional computing resources, consisting here in the logic and memories, which take up an additional 0.39 mm^2, compared to the 0.024 mm^2 of the AFE. In addition, ULP implementation of MCUs usually benefit from more advanced fabrication nodes, which may impact the environmental footprint of production [102]. MCU implementations however provide a versatile and upgradable alternative to ASIC implementations.

In this chapter, the total power consumption of the system was minimized by optimizing the respective power consumption of the different circuit blocks, for a given set of local performance specifications. However, a trade-off often exists between the target performance of a circuit and the minimum power consumption that can be achieved. Chapter 3 explores these power-performance trade-offs for the different circuits designed in this chapter. In Chapter 4, the local performance specifications are considered as design parameters at the system level to minimize the total power consumption while maintaining a sufficient accuracy for the global system.

3

Block-level power/ performance optimization and Pareto front modeling

3.1 Performance trade-offs in circuit design

Modern mixed-signal system-on-chips (SoCs) are made up of several heterogeneous circuit blocks, analog or digital, involved in various functions such as control, signal acquisition, data processing, clock generation, voltage regulation, or communication. The cardiac monitoring system with embedded arrhythmia classification presented in Chapter 2 is an example of such mixed-signal SoC with both an analog front-end (AFE) for signal acquisition and a digital back-end (DBE) for data processing. In each circuit block, the performance evaluation usually relies on multiple metrics, such as power consumption, speed, accuracy or precision, linearity, etc. At design time, several of those metrics can be considered simultaneously as optimization objectives, resulting in a multi-objective optimization problem. Each of these performance objectives is defined as a function $f_i(\mathbf{x})$ of the design parameters $\mathbf{x} = (x_1, x_2, ...)$. The parameters $\mathbf{x}$ correspond to the sizing variables of a given circuit topology, such as device dimensions, bias voltages or clock frequencies [45]. Without loss of generality, only the

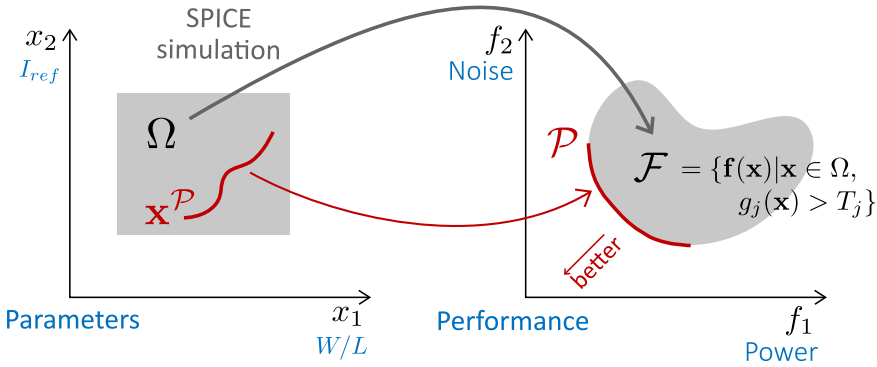

Fig. 3.1 Pareto optimum in multi-objective optimization problems. Simulation maps the circuit parameters $\mathbf{x}$ to the circuit performance metrics $\mathbf{f}(\mathbf{x})$. The Pareto curve $\mathcal{P}$ is the boundary of the achievable performance space $\mathcal{F}$ in the optimum direction. The blue text provides an illustration of the optimization task for an instrumentation amplifier use case. While the illustration provided here is limited to 2-D parameter and objective spaces, higher dimensions are often encountered.

minimization of these objective functions is addressed in this work. In the case of a function to be maximized, its inverse or opposite function must be considered.

Mathematically, the set of feasible design parameters Ω, which contains all circuit configurations $\mathbf{x}$ that are physically achievable, maps to a set of feasible performance $\mathcal{F}$, as illustrated in Fig. 3.1 for a 2-D objective space. Performance constraints can also be defined on the set of achievable performance to account for additional circuit requirements, in which performance metrics are assigned a minimum or maximum value, $g_j(\mathbf{x}) > T_j^{min}$ or $g_j(\mathbf{x}) < T_j^{max}$. Similarly to physical constraints, performance constraints shrink the feasible objective space, with the difference that a function mapping the circuit parameters to the performance metric needs to be evaluated while physical constraints can be verified immediately. As the mapping is usually non-linear for performance objectives and constraints, an accurate evaluation of the circuit performance is typically obtained using a SPICE circuit simulator.

In the set of achievable performance, the subset that optimizes performance is of interest as it can be inversely mapped to optimal design parameters $\mathbf{x}^{\mathcal{P}}$. When several performance objectives are considered, there is

rarely a single combination of parameters that optimizes all metrics simultaneously, as these are often competing against each other, which leads to a performance trade-off. The set of optimal points is known as the Pareto front $\mathcal{P}$ (Fig. 3.1), i.e., all points for which it is not possible to improve one objective without degrading another [46]. For instance, in the case of an instrumentation amplifier (IA), there is a trade-off between, e.g., its power consumption and its intrinsic noise. The Pareto curve can typically be used to help the designer select the optimal performance trade-off for his application. In hierarchical system design, Pareto-optimal models of circuit blocks can also be used at the system level to take into account low-dimensional low-level design information [45].

In this chapter, the local power and performance trade-offs of the constituting blocks of the electrocardiogram (ECG) monitoring and arrhythmia classification system of Chapter 2 are explored and modeled. Section 3.2 starts by describing methods in the literature for performance space modeling, which have been used to obtain information about circuit trade-offs. To model the components of the target system, several modeling methods are then applied in order to meet the different requirements of each block in the system. Firstly, the link between resolution, sampling frequency and power of an analog-to-digital converter (ADC) are explored using an analytical approach in Section 3.3. Secondly, in Section 3.4, a numerical approach based on genetic optimization algorithms is used to model the power and noise trade-off of an IA. To reduce the execution time of this simulation-intensive method, two solutions are introduced to mitigate the number of simulation runs. In order to validate these numerical and analytical approaches experimentally, two prototypes with configurable block-level performance, respectively an IA in TSMC 65-nm LP CMOS and an ADC in GF 22-nm FD-SOI, are implemented using the Pareto fronts obtained with the respective methods. The optimization and design of a configurable IA were presented in [JP1]. Finally, the particular case of performance modeling for DBE is discussed in Section 3.5. Due to the dependence of the inference performance on the input signal quality, a more explicit simulation method is presented. This chapter concludes with the comparison of the different approaches and their respective advantages and disadvantages depending on the type of circuit evaluated.

3.2 State of the art of performance trade-off modeling

In the literature, the exploration and modeling of performance trade-offs have mainly been studied from the perspective of analog circuit design. While digital design has been mostly automated thanks to the development of powerful computer-aided design (CAD) tools, analog design still mostly relies on the designer know-how and trial-and-error. This lack of efficient tools makes the manual design of analog blocks difficult due to the large number of possible topologies and sizing parameters. In this context, extracting the Pareto-optimal trade-off curve of a given circuit, which relates its sizing parameters to the best achievable performance metrics, allows to select the desired trade-off, obtain the related circuit sizing and possibly compare different topologies.

In practice, the main approach used for evaluating trade-offs in analog circuits is an analytical one, though often implicitly [104, 105]. Using their knowledge of the circuit, designers derive the main equations defining the circuit performance. These equations are then used to make decisions and size the circuit devices appropriately. While this approach is only as effective as the knowledge of the designer, simple equations can be obtained quickly and provide initial circuit insights. However, it is usually difficult to obtain an exact model that captures all second-order effects. Additional selected simulation data can be incorporated in the design to calibrate the model and provide more accurate results [106].

When analytical approaches are limited by the complexity of the circuit or their lack of accuracy, simulation-based approaches can provide an alternative solution. They rely on a large number of SPICE simulations to explore the design space, usually in addition to numerical tools to guide the exploration. By tuning many design parameters, these methods can reach optimal circuit sizing with accurate performance evaluation. However, they suffer from the large number of time-consuming circuit simulations that are necessary to map the performance space. In the literature, different methods originating from the optimization field have been applied to analog circuit sizing to extract their Pareto front. Numerical optimization methods applied to circuit sizing can be classified in several categories: equation-based, simulation-based or surrogate model-based [107, 108].

- In equation-based methods, the objective performance is estimated by expert-derived equations, which can be either extracted similarly to the analytical approach presented above [105] or generated auto-

matically [109]. By combining these equations with numerical optimization, the optimal sizing points can be found. The advantage of these methods is that the calculation of these equations is often much faster than SPICE simulations of the circuit, but they can suffer from the accuracy and simplicity bottlenecks inherent to the analytical models [107].

- In simulation-based methods, full-SPICE simulations evaluate the circuit performance. Many examples of numerical optimization for analog circuit sizing have been studied in the literature, most often on simple blocks such as operational amplifiers [110]. A variety of optimization techniques have been used to deal with multiple performance objectives and to obtain a set of Pareto-optimal design points. The optimization of the circuit parameters can be performed by solving a series of single-objective optimization problems, for example by using weights to scalarize the multi-dimensional performance objectives [111, 112]. Alternatively, the multi-objective optimization problem, i.e., finding the optimal Pareto front in a single optimization run, can be solved by dedicated algorithms. Multi-objective optimization algorithms mostly consist in evolutionary algorithms, such as particle swarm optimization (PSO) [113], multi-objective evolutionary algorithm based on decomposition (MOEA/D) [114], or non-dominated sorting genetic algorithm (NSGA-II) [115]. Those algorithms rely on bio-inspired iterative heuristics to make a set of candidate solutions converge towards the Pareto curve. The main bottleneck of these methods is the very long simulation time required to evaluate up to hundreds or thousands of design points [108].

- Finally, surrogate model-based methods have recently emerged to benefit from the accuracy of simulation-based methods while avoiding the execution time drawback. They consist in iteratively building surrogate models of the performance functions using machine learning regression methods such as Gaussian processes [116] or Bayesian neural networks [108], and using those models to solve the multi-objective optimization problem. With this approach, only the best candidates according to the multi-objective optimization problem are evaluated using SPICE simulations and used to refine the surrogate model.

When many performance metrics are considered, the selection of valid de-

sign parameter sets can be accelerated using hierarchical simulation, as performed in the case of exhaustive space search in [117].

On the digital side, hardware synthesis and physical implementation predominantly rely on computer-aided design tools. Thanks to standardized cells and memory macros, those tools are able to translate user requirements through the multiple abstraction layers into designs that optimize given metrics such as power consumption, speed or area. The power and performance trade-off of microcontroller-based digital circuits can be swept at the architecture level, e.g., by choosing a different type of core or optional accelerators, or at the implementation level by changing the target design specifications such as the operating frequency. For a given logic design and target frequency, methods to optimize the physical parameters such as supply voltage or threshold voltage flavour can be integrated in the synthesis flow [118]. Different implementations can be compared one-to-one to evaluate their respective speed, efficiency and power trade-offs. A fixed digital hardware, regarding both the architecture and physical implementation, will be considered in this work and the design trade-offs at this level will not be explored.

Regarding the design of the software algorithms, many studies have explored the trade-off between accuracy and cost, which can be expressed in terms of execution time, energy or computing resources. A lot of research has been carried out on computer-based implementations, but significantly less tackled this issue in smart embedded systems which suffer from additional energy and resource constraints. In sensor systems with on-chip classification, the cost-accuracy trade-off can be modified either by modifying the input data, e.g., by reducing the sampling rate or resolution [119], or by changing the parameters of the classification algorithm, such as the training hyperparameters of a machine-learning model [120, 121]. The evaluation of the software accuracy in the case of machine learning models usually consists in the exhaustive simulation with a grid search in the design space, due to the presence of discrete parameters [119, 120, 121]. The estimation of the power consumption is not a straightforward task in the case of software execution on a microcontroller as it strongly depends on the software code being executed. In the literature, methodologies for extracting power models can be classified as either a hardware-oriented bottom-up approach, or a system-oriented top-down approach. Bottom-up approaches start from the description of the hardware components and their respective energy consumption and model the total power consump-

Table 3.1 Power-performance trade-off modeling methods.

Section	Approach	Method	Use-case circuit
3.3	Knowledge-based	Equation derivation	ADC
3.4	Simulation-based	Genetic optimization	IA
3.5	Direct emulation	Software simulation	DBE

tion based on their activity factors (number of memory accesses, number of instructions of each type, etc.) [122, 123, 124]. This approach typically provides accurate results but requires a lot of data and simulation results down to the hardware level. System-oriented approach operate in a top-down fashion by modeling the total power consumption based on software parameters, such as function calls or variable values [124, 125]. This approach provides coarser results but requires less data to calibrate the model. It works well for software tasks with regular instruction patterns, such as loops or vector computations, which are often present in machine-learning algorithms.

In the following sections, different approaches are used to model the power and performance trade-offs of the circuit blocks in the system of Chapter 2. The respective approaches and methods used, as well as the circuits modeled in each case, are summarized in Table 3.1.

3.3 Knowledge-based analytical approach

When the trade-offs between different performance metrics can be easily grasped, deriving manually the main equations of the circuit can already provide an acceptable approximation. While those may not capture all second-order non-idealities in the circuit, they have the advantage of providing a model which can be understood by the designer. In the system presented in Chapter 2, the relationship between resolution, sampling frequency and power consumption in the analog-to-digital converter (ADC) can be extracted in this manner with a few design parameters and equations. This section presents the analytical extraction of performance and power trade-offs for the ADC. The model is then verified experimentally by prototyping an ADC with configurable resolution and power in GlobalFoundries 22-nm FD-SOI technology.

3.3.1 ADC performance trade-offs

The ADC presented in Section 2.4.2 is a time-based ADC with a back-bias controlled oscillator (BBCO) as input stage to perform the voltage-to-frequency conversion, followed by a counter to integrate its phase, and a sampling stage. The key performance metrics of the ADC in the classification system are its dynamic power consumption, effective resolution and sampling frequency. The effective quantization resolution is measured with the effective number of bits (ENOB), which is calculated on the basis of the signal-to-noise ratio (SNR) in this work, taking into account its quantization noise and thermal noise in the BBCO. Let us mention that the distortion is discarded as the subsequent signal processing algorithm is resilient to limited non-linearity (Section 2.4.2). Moreover, the flicker noise in the BBCO is negligible compared to thermal noise thanks to chopper modulation in the ADC. Finally, errors in the counter due to metastability are mitigated by the asynchronous double sampling technique. Finally, the leakage power consumption is negligible compared to dynamic power when the ADC is active.

The performance metrics considered for the ADC are mainly limited by the BBCO characteristics in this topology. The performance of BBCOs can generally be tuned with the following parameters: supply voltage, number of stages, load capacitor, transistor length and parallel device multiplier [82]. When working with standard cells only, not all those parameters can be accessed directly and the remaining factors are the supply voltage, number of stages in the RO, transistor length and threshold voltage (depending on the choice of cell library), gate fan-in (i.e., pure inverter or NAND/NOR gate with shorted inputs), and cell driving strength. In this work, the supply voltage is considered fixed by the SoC architecture, at 0.65 V for this analysis. Given the necessity to slow down the oscillator, as the target ECG frequency range is at only 250 Hz, the transistor length and threshold voltage are maximized in the available libraries, which corresponds to ultra-high-V_t (UHVT) transistors with 36-nm gate length in 22-nm FD-SOI technology from Global Foundries. Compared to the work presented in Chapter 2, the UHVT transistors operate in reverse back biasing (RBB) because the PMOS transistors are laid out over an N-Well and the NMOS transistors over a P-Well, which means that positive back-bias voltages can be applied to the PMOS devices. The type of cell is chosen as 3-input NOR gates to balance gain and linearity, similarly to the design of Section 2.4.2. The remaining BBCO design parameters taken into account

for the ADC performance modeling are the number of stages and cell driving strength. In this analysis, the cell driving strength will be modeled as cells connected in parallel. This approximates the choices of cells found in the standard cells libraries, which display less consistent width sizing. Finally, the sampling frequency can be directly controlled in the circuit by the clock controller and is included as a design parameter.

Before modeling mathematically the performance trade-offs, the impact of these selected design parameters can be understood intuitively separately.

- **Number of stages** When increasing the number of stages in the BBCO, its oscillation frequency at the output decreases. The input-referred noise due to thermal noise jitter remains constant, contrary to flicker noise which decreases due to its correlation property and averaging. The dynamic power remains unchanged as the propagation speed of the transition in the ring oscillator does not change. In the counter, the reduced BBCO frequency induces a lower number of possible transitions per sampling period and therefore a lower resolution, while also reducing the dynamic power. Overall, the number of stages allows to trade quantization resolution for power.

- **Cell driving strength** Increasing the driving strength of each stage does not change the oscillation frequency, due to the equal increase in capacitance and driving current [82]. However, as the voltage-to-frequency transfer function does not change and the parasitic capacitance increases, the input-referred noise of the BBCO is reduced. The cell driving strength therefore tunes jitter noise and power.

- **Sampling frequency** Increasing the sampling frequency obviously provides a better temporal resolution. However, if other parameters remain unchanged, as the integration period of the counter decreases, the quantization resolution decreases. As the sampling frequency is low and the power consumption of the sampling flip-flops is negligible, a limited change in the sampling frequency has a negligible impact on the power consumption. The sampling frequency exchanges temporal for quantization resolution.

3.3.2 Modeling via the derivation of circuit equations

Based on the intuitive insights from the previous section, mathematical models of the performance trade-offs are now derived. In the considered

time-based ADC circuit, the resolution is defined by the difference between the numbers of periods of the BBCO in a period of the sampling clock when minimum and maximum input voltages are applied. The quantization resolution, without taking intrinsic noise and distortion into account, is therefore given by

$$N_Q(N) = \log_2\left(\frac{2\left(f_{osc}(N, v_{min}) - f_{osc}(N, v_{max})\right)}{F_s}\right), \tag{3.1}$$

where $f_{osc}(N, v)$ is the transfer function of a BBCO with N stages, v_{min} and v_{max} are respectively the lower and upper bounds of the input voltage range, and F_s is the sampling clock frequency. The transfer function $f_{osc}(N, v)$ is inversely proportional to the length of the BBCO :

$$f_{osc}(N, v) = f_{osc}(N_0, v)\frac{N_0}{N}, \tag{3.2}$$

with N_0 a reference BBCO length.

The model of BBCO jitter depends on the type of noise considered. As thermal noise is uncorrelated, the associated input-referred jitter σ_{th} varies only with the cell driving strength. However, the jitter due to flicker noise σ_{fl} depends both on the cell driving strength and on the number of stages :

$$\sigma_{th}(M, T_s) = \sigma(M_0, T_s)\sqrt{\frac{M_0}{M}}, \tag{3.3}$$

$$\sigma_{fl}(N, M, T_s) = \sigma(N_0, M_0, T_s)\sqrt{\frac{N_0 M_0}{NM}}, \tag{3.4}$$

where $\sigma(N, M, T_s)$ is the standard deviation of input-referred jitter noise sampled at a frequency $F_s = 1/T_s$, and M is the number of devices in parallel to model the driving strength. Fig. 3.2 shows the validation of the models of BBCO performance (Eqs. 3.2, 3.3 and 3.3) against simulation data. The model is calibrated using the simulation of a reference BBCO with 7 stages and 1 cell per stage ($N_0 = 7$ and $M_0 = 1$).

Using the number of quantization bits and the BBCO jitter from Eqs. 3.1 and 3.4, we can infer the effective number of bits (ENOB) for the ADC, excluding distortion:

$$\text{ENOB} = \frac{\text{SNR}_{\text{dB}} - 1.76}{6.02}, \tag{3.5}$$

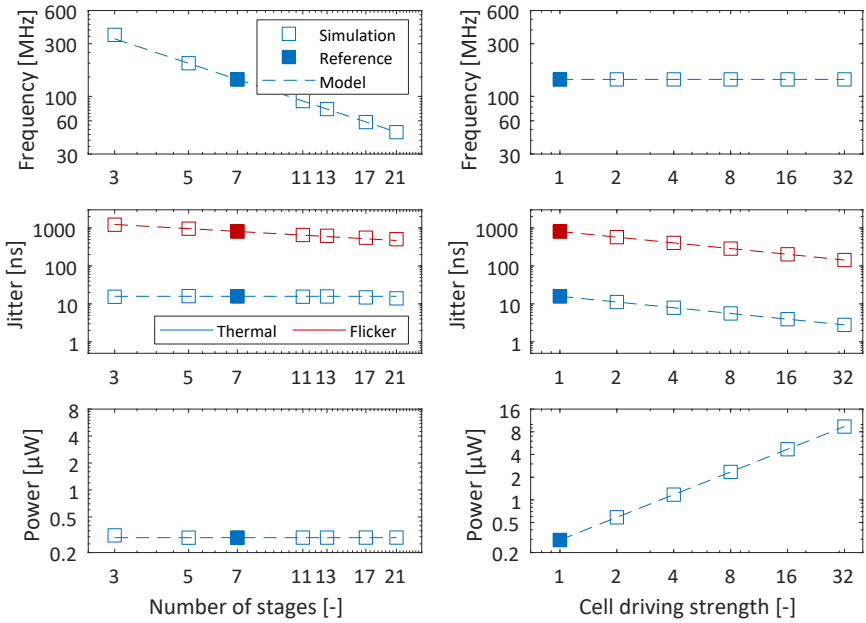

Fig. 3.2 ADC model verification. Eqs. 3.2, 3.3, 3.4, and 3.9 are compared to simulation results, by sweeping independently the number of stages N and the cell driving strength M. Results are obtained from steady-state simulation of BBCOs with UHVT NOR3 gates, at 0.65 V and 25 °C.

$$\mathrm{SNR_{dB}} = 10\log_{10}\left(\frac{(v_{max}-v_{min})^2}{8(n_q^2+n_j^2)}\right), \tag{3.6}$$

$$n_q = \frac{v_{max}-v_{min}}{2^{N_Q}\sqrt{12}}, \tag{3.7}$$

$$n_j \approx \sigma(N, M, T_s), \tag{3.8}$$

where $\mathrm{SNR_{dB}}$ is the signal-to-noise ratio expressed in dB, and n_q^2 and n_j^2 are the input-referred quantization and jitter noise power.

The power consumption of the ADC is equal to the sum of the power consumption in the BBCO and in the counter. As the sampling frequency is low, the power consumption of the sampling flip-flops and decoder is negligible. For a given BBCO stage and average input voltage, the power

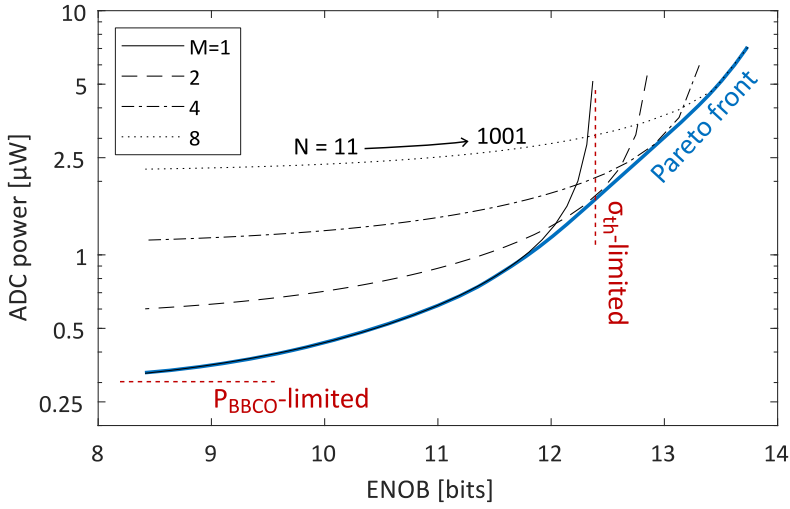

Fig. 3.3 ADC resolution and power trade-off model. For a given driving strength, the power consumption is bounded by the BBCO power and the effective resolution by the jitter (in red). The envelope of the different curves shows the power and performance Pareto front for the ADC design (in blue).

consumption of the BBCO is proportional to the cell driving strength:

$$P_{BBCO}(M) = P_{BBCO}(M_0)\frac{M}{M_0}. \tag{3.9}$$

The BBCO power is however independent of the number of stages in the ring oscillator, as validated in Fig. 3.2. The dynamic power consumption of the counter depends linearly on the output frequency of the BBCO. This results in a total ADC power of

$$P_{ADC} = P_{BBCO,M} + P_{counter} = P_{BBCO}(M) + 2f_{osc}(N, v_{avg})E_{dyn}, \tag{3.10}$$

where v_{avg} is the mean input voltage and E_{dyn} is the energy per cycle of the counter. It follows from Eq. 3.5 and 3.10 that a trade-off between resolution and power consumption in the ADC occurs when changing the number of stages N or cell driving strength M. This trade-off can be modified by design as illustrated in Fig. 3.3.

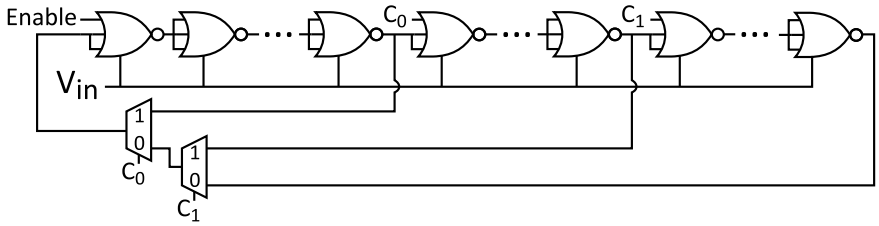

Fig. 3.4 Performance configurability implementation in the ADC. Digital control signals C_i allow configuring the length of the BBCO.

3.3.3 Experimental validation with configurable ADC

As a digital-only circuit, performance configurability can be easily implemented in the proposed time-based ADC. Based on the analysis in the previous section, the power and effective resolution trade-off can be tuned by changing the length of the BBCO or the cell driving strength. Changing the driving strength, or equivalently connecting several cells in parallel, is not trivial due to the added parasitic capacitance on the nets between successive stages. It is therefore not used as a tuning parameter for the performance configurability in this work. This has a limited impact on the Pareto curve when only the designs with low power and low resolution are targeted, due to the limit of minimally-sized stage driving strength as shown in Fig. 3.3. The BBCO length can be reduced more conveniently by closing the RO at different points in the loop depending on the value of control signals, as shown in Fig. 3.4. Multiplexers are used to connect the first stage of the BBCO with the output of intermediate cells in the loop. After each intersection, one input of the following NOR3 gate is connected to the control signal to stop the propagation of the transition in the unused part of the RO. The introduction of this length selection slightly impacts the RO frequency, which can be compensated by removing a few stages if necessary. Additionally, the sampling frequency can also be modified dynamically. As the frequency is generated as a divided clock from the low-frequency reference clock on the SoC, a different division factor can be selected.

In the AFE implemented in the SleepRider prototype of Chapter 2, the flicker noise dominates the BBCO jitter at low frequency. As two single-ended ADCs are used as a differential ADC, chopper modulation can be used to modulate the intrinsic noise at higher frequency and filter it in the digital domain as shown in Fig. 3.5. The flicker noise is thereby strongly

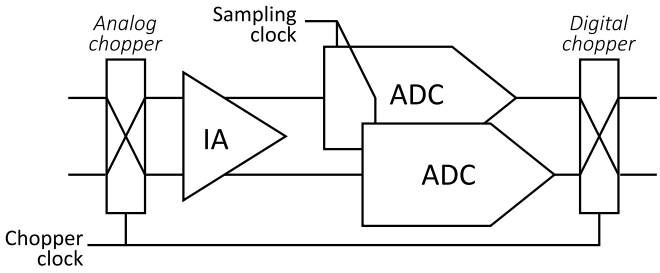

Fig. 3.5 ADC chopper modulation. Chopper demodulation is moved after the ADC to reduce BBCO flicker noise.

reduced and the thermal noise dominates the jitter noise. As the preceding block in the AFE is the IA, which also includes chopper modulation, no demodulation is performed between the blocks. This technique constraints the sampling frequency to be an integer multiple of the chopper frequency. The power consumption of chopper modulation in the digital domain is negligible thanks to the low sampling frequency.

This configurable ADC was prototyped and included in the ICare SoC fabricated in 22-nm FD-SOI from Global Foundries. The ADC is measured with a sampling frequency of 4 kHz and a chopper frequency of 2 kHz, for a signal bandwidth of 500 Hz (oversampling ratio of 4). Thanks to the use of standard cells, the layout of the ADC is very compact as shown in Fig. 3.6, with an area of 600 μm^2 for a single-ended ADC or 1200 μm^2 for a differential ADC. Fig. 3.7 shows the power and effective resolution trade-off for several configurations, with a range of 0.3 to 1 μW for 8 to 12 bits. The Walden figure of merit (FoM), which is defined as

$$\mathrm{FoM}_W = \frac{P}{2^{ENOB} f_{nyq}} \tag{3.11}$$

with P the power consumption and f_{nyq} the Nyquist sampling frequency, ranges from 0.23 to 1.3 pJ per conversion step. The configurability measurements show good agreement with the model results calibrated with simulation data. Slight discrepancies can be explained by the impact of parasitic capacitances in the BBCO.

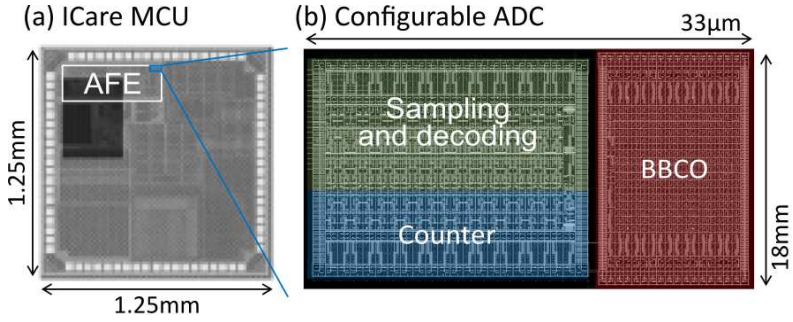

Fig. 3.6 Configurable ADC layout. The layout of the ICare MCU SoC in 22-nm FD-SOI technology (a) and configurable ADC (b)

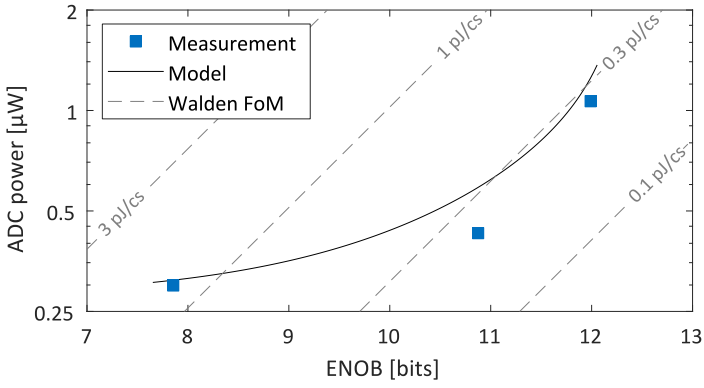

Fig. 3.7 Configurable ADC measurement. The measurement are respectively obtained by changing the configuration signals to modify the BBCO length. The solid line corresponds to the analytical model calibrated with simulation data. The lines with constant energy per conversion step, i.e., Walden FoM, are shown in dashed lines.

3.4 Simulation-based approach with numerical optimization

Using SPICE simulations to explore the performance space of a circuit is commonplace in analog circuit design. In the literature, several numerical multi-objective optimization methods have been combined with simulations to guide the exploration and converge towards the optimal Pareto front. Those methods are usually based on evolutionary algorithms, in which a set of design points evolve iteratively towards the optimum curve using bio-inspired heuristics. The NSGA-II algorithm [115] has already

Table 3.2 Analog circuit sizing with NSGA-II. The NSGA-II algorithm has been applied to the sizing of different types of circuits in the literature, often with a limited complexity.

		Circuit	# transistors	Other components	# variables	# objectives	# constraints
	This work	IA	22	2 OTA, I_b, 3R	9 / 17	2	6
Kchaou [126]	Integration, 2016	VF	11	I_b	N/S	2	1
		CCII	13	I_b	N/S	2	1
Sanabria [127]	IETE Tech Rev, 2018	CCII	13	I_b	9	2	4
Sasikumar [128]	NetACT, 2017	OA	7	V_b, 2C	9	3	4
Nicosia [129]	KBS, 2008	RF LNA	2	3R, 2L, 3C	7	2	3
		Filter	0	4 OA, 18R, 8C	20	3	13
		UWB LNA	2	2R, 3L, 2C	9	2	2
Cuate [130]	MIPRO, 2019	OTA	12	I_b	6	2	1
Takhti [131]	ISSCS, 2009	FC OA	11	V_b	16	2	4
Lopez [132]	SMACD, 2016	LDO	14	2 I_b, 3R, 2C	17	2	0
Stas [133]	NEWCAS, 2016	OTA	10	I_b	11	2	3
		Voltage ref.	4	2C	8	2	2
		DCO	12	RO	5	2	2
		LNA	3	3R	7	2	3

IA: instrumentation amplifier - VF: voltage follower - CCII: second generation current conveyor - (FC) OA: (folded-cascode) operation amplifier - RF/UWB LNA: radio frequency/ultra-wideband low-noise amplifier - OTA: operational transconductance amplifier - LDO: low-dropout regulator - DCO: digitally-controlled oscillator - I_b: bias current - V_b: bias voltage - RO: ring oscillator

been applied many times to analog circuit sizing [126, 127, 128, 129, 130, 131, 132, 133]. However, as shown in Table 3.2, the optimized circuits are usually restricted to blocks with a limited number of devices and low complexity, such as a small amplifiers. Due to the large number of time-consuming simulations, when considering more complex circuits such as a full IA, solutions are necessary to limit the number of simulation runs.

This section discusses the application of the NSGA-II multi-objective optimization algorithm to the sizing of the IA[1] in the ECG arrhythmia classification system of Chapter 2. Two solutions for reducing the number of simulations, i.e., cascaded SPICE simulations and high-level parameter selection, are introduced. The resulting Pareto front obtained by the optimization algorithm is verified experimentally and applied to the design

[1]This work only considers the sizing task for a fixed topology. Other works have proposed optimizing the topology as well, either by selection or generation [134]. This however lengthens the optimization process and may lead to circuits that are hardly interpretable by the designer, and is therefore not considered.

of a novel IA with configurable noise and power performance in TSMC 65-nm CMOS technology.

3.4.1 Instrumentation amplifier performance trade-offs

The IA is a fully-differential current-balancing topology previously shown in Fig. 2.6 in Chapter 2. It is composed of an input transconductance (G_m) that converts the differential input voltage to a proportional current variation, and a transimpedance (Z_m) that converts it back to a differential voltage at the output. The total voltage gain (A_v) of the amplifier is therefore set by the ratio of the resistances in the Z_m and the G_m, as well as by the current mirror ratio (Eq. 2.4). The operating principles of the IA are detailed in Section 2.4.1.

A reference and feasible sizing of this IA was performed manually and will be used as a starting point for the design optimization methodology presented in Section 3.4.2. Manually searching for the optimal values of the circuit parameters is a difficult task due to the numerous trade-offs between performance metrics. For example, the noise in this circuit is typically dominated by the G_m resistance R_i, the input transistor pair P1 and the G_m current source N1. To minimize the noise, one would tend to minimize R_i. However, choosing R_i too small impacts the gain due to the finite output conductance of P1 as shown in Eq. 2.4, and also increases the distortion of the signal. As this conductance depends on the bias current, it is directly linked to the power consumption which is the second performance objective. On the other hand, the resistance values R_i and R_o will be limited by the maximum silicon area. Other examples of such trade-offs involve the current ratio between G_m and Z_m (i.e., $(W/L)_{P6}/(W/L)_{P2}$ in Eq. (2.4)) which helps decreasing the power and the area but reduces the gain, or the chopping frequency that improves the noise and common-mode rejection ratio (CMRR) but impacts the bandwidth requirement and distortion. Similar reasoning applies for most circuit parameters and performance metrics.

3.4.2 Modeling via multi-objective optimization

The non-dominated sorting genetic algorithm (NSGA-II), proposed by Deb *et al.* [115] as an improvement over the previous NSGA [135], is a numerical optimization method for solving multi-objective optimization problems. As explained in Section 3.1, when multiple performance objectives are considered, several optima exist, each corresponding to a different trade-off

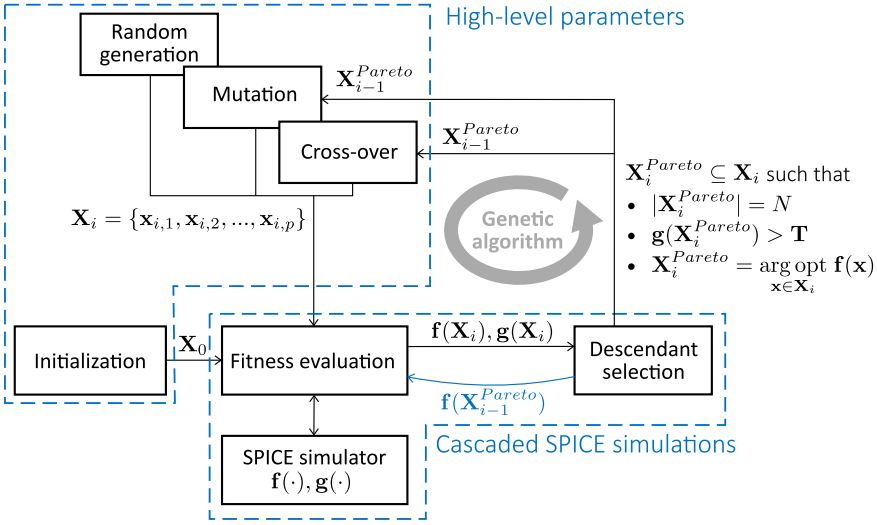

Fig. 3.8 Genetic optimization algorithm. The annotation on the arrows show the inputs and outputs of the different blocks. The proposed acceleration techniques are highlighted in blue.

between the different performance metrics. Compared to single-objective optimization that only provides a single optimal result, the output of multi-objective optimization is a set of optimal points distributed on the Pareto front. Applied to analog circuit sizing, the optimization variables are the sizing parameters of the circuit and the objectives are its performance metrics.

NSGA-II methodology for analog sizing

Like other evolutionary algorithms such as PSO [113] or MOEA/D [114], NSGA-II is an iterative process in which a population $\mathbf{X}$, made up of individuals that each correspond to a set of circuit parameters $\mathbf{x}$, evolves for several generations, as shown in Fig. 3.8. At each generation, the performance objectives and constraints are evaluated and a subset of candidates are selected to generate a new population, as detailed in the following stages:

1. Initialization The initialization phase consists in finding a non-optimized initial population $\mathbf{X}_0$ that meets the performance constraints. This is done by a random search in the design space and stops as soon as three valid individuals are found.

2. Fitness evaluation: At each ith generation, the new population $\mathbf{X}_i$ starts the genetic selection process. During fitness evaluation, all performance objectives $\mathbf{f}(\mathbf{x}_{i,j})$ and constraints $\mathbf{g}(\mathbf{x}_{i,j})$ are computed for each individual $\mathbf{x}_{i,j} \in \mathbf{X}_i$ (i and j are the generation and individual indices, respectively). To evaluate the performance metrics, the algorithm instantiates each circuit in an external SPICE simulator and runs the appropriate analyses.

3. Descendant selection: Similarly to the biological phenomenon it replicates, genetic algorithms select only the fittest subset of individuals, based on the results of the fitness evaluation. The NSGA-II algorithm is based on an elitist selection mechanism where all the best individuals are allowed to carry on to the next generation. At each generation, descendant selection first discards all individuals in the population $\mathbf{X}_i$ that do not meet the performance constraints $\mathbf{g}(\mathbf{X}_i) > \mathbf{T}$. The remaining individuals along with the selected members of the previous generation $\mathbf{X}_{i-1}^{Pareto}$ are sorted in successive nondominated fronts and their spread in the objective space is measured by the crowding distance [115]. N individuals are selected in $\mathbf{X}_i^{Pareto}$, firstly based on the rank of their nondominated front, then in descending order of distance (here we choose $N = 30$).

4. Population generation: The individuals selected at generation i, i.e., the subset $\mathbf{X}_i^{Pareto}$, are used to stochastically create the population of the next generation $\mathbf{X}_{i+1}$. Three independent processes are used to create new individuals. Mutation consists in taking an individual and slightly modify a few of its circuit parameters. Cross-over is a process where two individuals mix their parameters to create an offspring. Finally, a few randomly generated individuals are incorporated at each generation to reduce the risk of moving towards a local optimum.

Each time a population is generated, it goes back to the fitness evaluation step, until a stopping criterion is met. Stopping criteria based on the convergence of the solutions is hard to establish in multi-objective optimization problems because of the high dimensionality of the result [136]. For the sake of simplicity, the stopping criterion is often chosen as the maximum number of iterations.

Reduction of simulation runs

The drawback of genetic algorithms is that they often require a large number of simulations for solving the multi-objective optimization problem. This issue worsens when time-consuming simulations such as long tran-

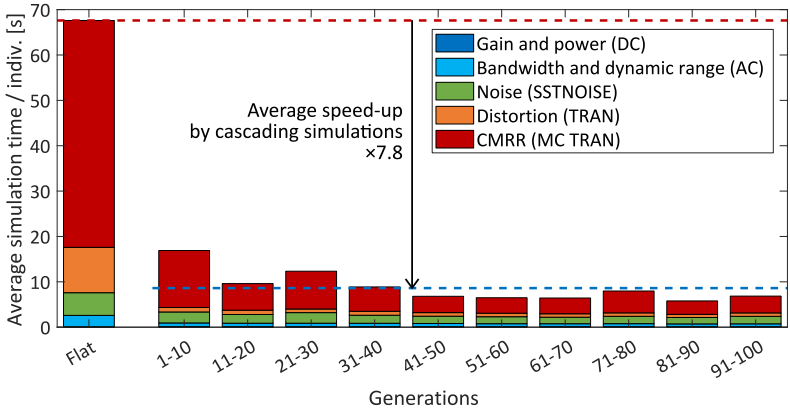

Fig. 3.9 Simulation time reduction with cascaded simulations. The first column shows the time for evaluating a single circuit candidate for all analyses. The following columns show the average time spent for each analysis throughout the generations.

sients or Monte-Carlo simulations are used and when the design space is large, typically for complex analog circuits. Two solutions are proposed in this work to mitigate the simulation time issue. The first solution relies on the fact that when many performance metrics are used as objectives or constraints, several SPICE analyses may be required with respective simulation times that differ in orders of magnitude (e.g., from 0.1 s for a DC point calculation to 50 s for Monte-Carlo analysis with transient signals). To gain simulation time, the simulations are run from the shortest to the longest and candidates are rejected as soon as requirements are not met, similarly to [117] in the case of exhaustive space search. The rejection can be done on the basis of both performance constraints that are violated or performance objectives that are not improving compared to previous generations. This allows to save simulation time on the longest analyses for individuals that are not optimal. Compared to unconditional simulations in which all analyses are performed, we observe a 7.8× reduction in the total simulation time as shown in Fig. 3.9.

Secondly, the design parameter space is reduced by relying on designer knowledge. Starting from a manual sizing performed initially, a subset of higher-level parameters are tuned instead of considering all low-level device dimensions individually. This can be done by deriving a few main equations driving the circuit performance to identify the main parameters.

Additionally, several parameters can be scaled together to keep the reference relative sizing within functional groups of devices (e.g., cascode current mirrors). Ideally, the final set of parameters should be as small as possible while still allowing enough flexibility to optimize the objective while meeting the constraints. In the case of the IA, circuit inspection shows that the circuit has at least 28 degrees of freedom. In order to narrow the parameter space, only the main design degrees of freedom are considered. Based on Equation (2.4), we select the resistances R_i and R_o, the inversion level of the input transistor pair P1 and PMOS cascaded current mirror represented by their parameter g_m/I_D [137], and the P2/P6 width ratio, as degrees of freedom. The bias current I_{ref} is also included to sweep the power consumption. As the NMOS current sources are major noise contributors, the g_m/I_D of transistor N1 is left as independent parameter, as well as its length so that the area can be modified independently. Finally, due to the low bandwidth of the IA, the chopping frequency f_{chop} has a considerable impact on the noise and is also included. Table 3.3 summarizes the list of optimization parameters and the range of admissible values. When performing a random space exploration to find candidates that meet the constraints, reducing the number of parameters from 17 to 9 increases the percentage of valid candidates by 40%. This effect is however mitigated during the optimization process as the genetic algorithm only searches in the neighbourhood of valid candidates with mutation or crossover. The improvement in convergence speed is therefore mostly apparent during the early stages of the convergence. Figure 3.10 shows the convergence based on the hypervolume (HV) metric, which is computed as the volume contained by the Pareto front in the objective space [138]. After convergence of the NSGA, the hypervolumes only differ by 0.8%. The time taken to reach 90% of the final HV is 47% shorter when considering 9 parameters instead of 17.

Optimization results and final Pareto front

Fig. 3.11 illustrates the process of convergence in the case of an IA with input-referred noise (IRN) and power consumption as performance objectives. In addition to the power-noise trade-off optimization, performance constraints for a typical biomedical application such as electrocardiogram (ECG) signal acquisition such as CMRR, distortion, or bandwidth are considered in this work, as summarized in Table 3.4. The results illustrate power-noise improvement across successive generations, with an even distribution of individuals along the Pareto curve. This algorithm can be di-

Table 3.3 IA design parameters. The selection of high-level parameters leaves 9 optimization degrees of freedom, with corresponding allowable ranges.

Parameter, x_i	**Range**
Reference current, I_{ref}	10nA - 1μA
G_m/Z_m current ratio, $(W/L)_{P6}/(W/L)_{P2}$	1 - 6
G_m resistance, R_i	1kΩ - 100kΩ
Resistance ratio, R_o/R_i	40 - 120
Inversion level of input transistor, $(g_m/I_D)_{P1}$	2 - 30
Inversion level of PMOS transistor, $(g_m/I_D)_p$	2 - 30
Inversion level of NMOS transistor, $(g_m/I_D)_n$	2 - 30
Length NMOS transistor, L_{nmos}	1μm - 20μm
Chopper frequency, f_{CLK}	1kHz - 20kHz

Table 3.4 Performance constraints. Design points that do not achieve the following requirements were rejected by the sizing algorithm.

Constraint, $g(\mathbf{x})$	**Threshold**, T
Gain	$>$ 30 V/V
Bandwidth	[0.1 - 250] Hz
Total harmonic distortion	$<$ 1%
CMRR	$>$ 80 dB
Dynamic range	$>$ 0.6 V
Active area	$<$ 20 000 μm^2

rectly extended to include more performance objectives, at the cost of an increased convergence time.

The resulting curve provides a model of the power and performance trade-off for the IA. While the model is only exact at the points selected by the algorithm, the final points show a trend which is smooth enough to be interpolated with only a limited uncertainty, in order to provide a continuous model. The desired number of points could also be increased to achieve finer resolution.

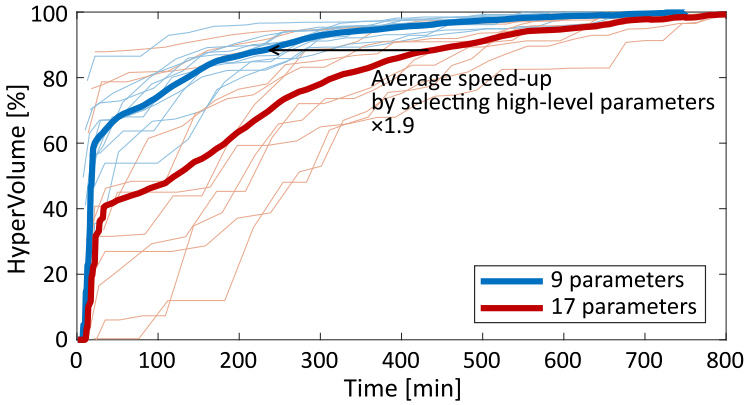

Fig. 3.10 Optimization speed-up with selection of high-level parameters. The genetic algorithm iterates for 30 generations with respectively 9 and 17 degrees of freedom. Light lines show convergence of individual runs, bold lines show the average convergence.

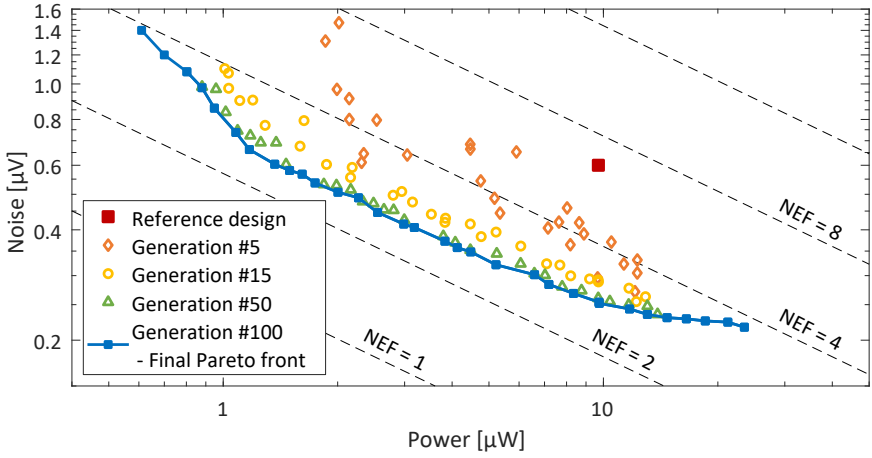

Fig. 3.11 Pareto front improvement over several algorithm iterations. The performance objectives of successive generations converge towards the Pareto-optimal curve and are displayed at several intermediate iterations. The dashed lines are the constant-NEF contours.

3.4.3 Configurable IA power-noise performance

Circuit configurability allows to dynamically change the circuit characteristics based on the current operating conditions. For example, in the case of an IA, input signals with different amplitudes or signal quality requirements may call for different IRN levels, leading to different power consumption as shown in Section 3.4.2. A trivial solution to have several noise-power trade-off choices is to implement several amplifiers on the chip and enable them separately. However, this leads to a high area overhead, essentially proportional to the number of configuration points. To avoid this overhead, a single amplifier with configurable characteristics is required. Song *et al.* [139] tuned the bias current of an amplifier to change its power consumption and IRN. However, simulation shows that this does not ensure optimal performance in all configurations as the bias conditions of the transistors vary. This work proposes to tune simultaneously several design parameters to reach optimal performance as obtained by the genetic algorithm.

Based on the Pareto curve obtained in Fig. 3.11, four design points have been selected to cover the noise-power range. Among the parameters considered in building the Pareto curve, transistor length and input transistor P1 width have been set to a fixed value, as they have converged to a near-constant value and they are more difficult to dynamically configure.

The configuration of the circuit parameters by the digital control signals C_i can be separated in three main ideas illustrated in Fig. 3.12. Firstly, the width of NMOS current sources is changed by modulating the number of transistors in parallel. Each branch is controlled by a separated bias voltage for the cascode transistor $V_{cas,n,i}$, that can be tied to ground to stop the current through that branch (Fig. 3.12.a). The generation of $V_{cas,n,i}$ is performed in the bias generation circuit where an additional transistor controlled by the digital signal C_i is used to enable the corresponding branch. All branches with C_i at high logic level appear in parallel and the reference current I_{ref} is distributed according to the width ratio. Inserting the additional control transistor in the bias generation circuit saves dynamic range in the signal branches and helps to avoid distortion. Secondly, the width of PMOS current mirrors follows the same behavior in the G_m (Fig. 3.12.b) using $V_{cas,p,i}$ as control voltage. However, unlike NMOS current sources that carry DC current, the bandwidth of PMOS current mirrors is a limiting factor for the performance. In the low-power configuration, the gates of inactive branches in the G_m and Z_m are connected to the V_o voltage,

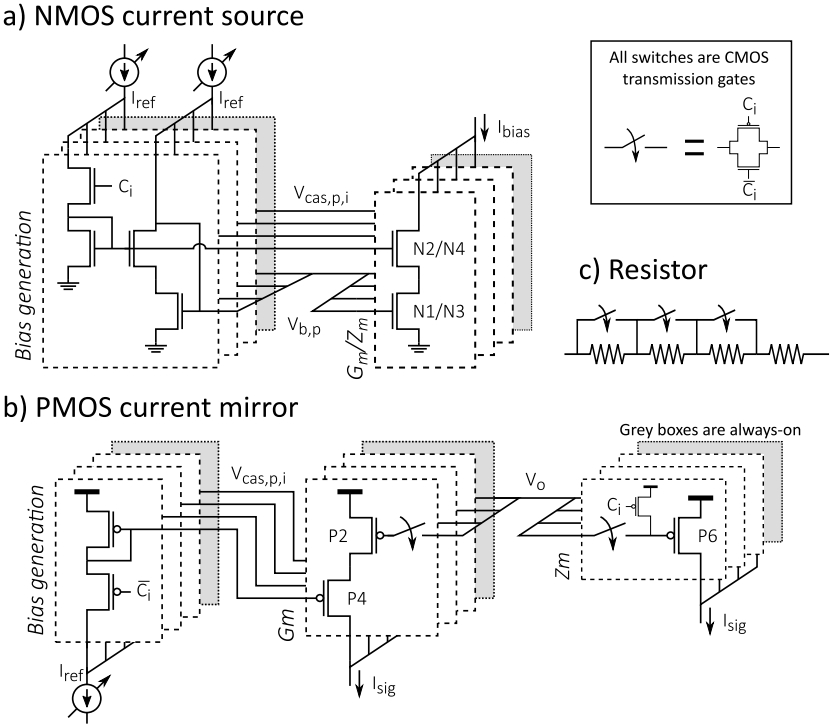

Fig. 3.12 Performance configurability implementation in the instrumentation amplifier. Digital control signals C_i allow configuring the width of transistors in current sources and current mirrors, and the resistance of polysilicon resistors. The branches in grey do not include digital control as they are always-on. Variable current source is provided off-chip.

thereby increasing the parasitic capacitance. In order to avoid this, additional switches on the gates of the deactivated branches are used to disconnect the gate from the V_o voltage, which is replaced by a much smaller source capacitance of the control transistor. Finally, changing the resistors R_i and R_o is done using transmission-gate switches to short-circuit parts of the polysilicon resistors (Fig. 3.12.c). As the resistance of R_i and R_o is quite high, the on-resistance of the switches does not impact the signal quality. Additionally, multiple frequencies of the chopper clock are generated on chip with clock dividers. The variable current reference is provided off-chip.

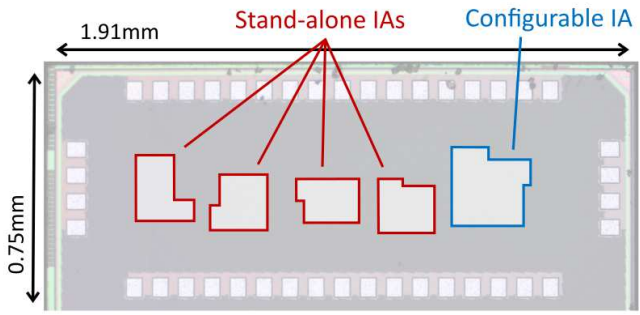

Fig. 3.13 Configurable IA prototype chip layout. The chip photograph shows the location of four stand-alone IAs in red and one configurable IA in blue. The area is approximately 0.035 and 0.055 mm^2 for stand-alone and configurable IAs, respectively.

3.4.4 Experimental validation

In order to validate the optimization methodology and amplifier configurability, a prototype chip has been built in TSMC 65nm LP CMOS. As shown in Fig. 3.13, each prototype includes five amplifiers: the first four are stand-alone amplifiers which each correspond to one selected point of the Pareto curve, and the last one is the configurable amplifier that allows modification of the power-noise trade-off using control signals.

The measured performance of the stand-alone and configurable amplifiers match the results of the genetic algorithms obtained with pre-layout simulations. The performance of the configurable IA ranges from 0.17 μV to 1 μV of RMS IRN and 23.8 μW to 0.56 μW of power consumption. The minimum NEF is 1.93. All constraints from Table 3.4 are met in all configurations. No performance penalty is introduced by the configurability.

The proposed configurable IA sized using the NSGA-II algorithm performs better in terms of noise-power trade-off compared to most state-of-the-art references. It can be noted that the works from Yaul *et al.* [140] and Song *et al.* [139] achieve lower power consumption by both using an extremely low supply voltage below 0.3 V on the input stage. This reduces the input dynamic range, which is a limiting factor in most biomedical applications.

The chip measures 1.91×0.75 mm^2. The area of the configurable IA is 0.055 mm^2, compared to around 0.035 mm^2 for each stand-alone amplifier. The limited difference is explained by the inverse relationship between re-

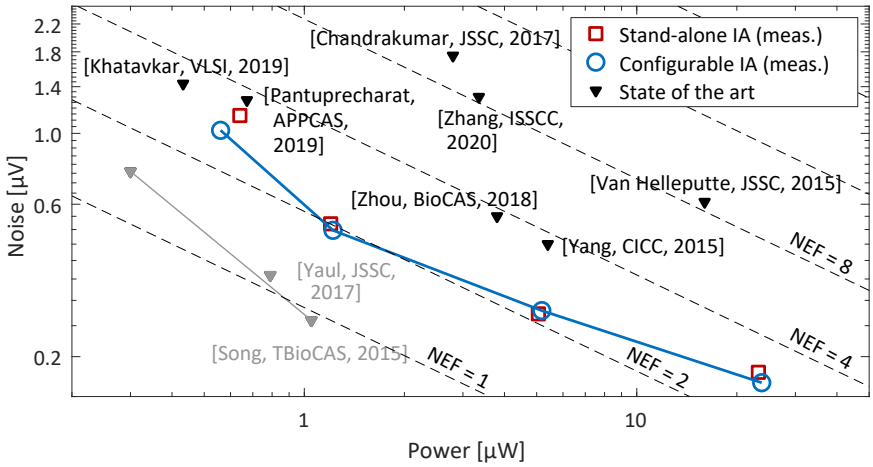

Fig. 3.14 IA performance measurement and comparison with the state of the art. Measurement results are compared for stand-alone and configurable amplifiers, as well as state-of-the-art references. The IRN from the state of the art has been scaled to a [0.5 - 100] Hz bandwidth.

sistors and transistors sizes in the different configurations according to the genetic algorithm. Indeed, while low-power configurations require small transistors, the resistance value is higher, and inversely for low-noise configurations. The configurability logic and switches incur a negligible area overhead.

3.5 Software emulation approach

The performance evaluation of the DBE, consisting in this case in the accuracy of the inference, differs significantly from the performance evaluation in the AFE. While the local performance of an analog block can generally be represented as a single metric independent from the input signals[2] (e.g., noise or transfer function of the IA, resolution of the ADC, etc.), the accuracy of the processing algorithms is highly dependent on the characteristics of the input data. The power and performance trade-off is therefore modified by two different types of parameters: extrinsic parameters that relate to the data received at the input of the DBE during its execution, and intrin-

[2]Note that this is not necessarily the case and requires the choice of metrics that feature this characteristic. For instance, inversely, the signal-to-noise ratio of an amplifier depends on the amplitude of the input signal.

sic parameters that are used for the design of the software algorithm. The extrinsic parameters typically impact the quality of the input data, which reduces the classification accuracy, for example due to a degraded separability of the classes. Due to the considerable number of possible signal variations, building a model that reflects this complexity would be unfeasible. This issue is exacerbated by the difficulty to predict the behavior of machine-learning models which are often hardly interpretable.

The intrinsic parameters considered in this work are the hyperparameters of the support-vector machine (SVM) classification model, such as the number of features and the SVM pruning threshold. The number of features can be controlled directly when training the algorithm based on the training set provided. To choose the features, we use the sequential forward floating search (SFFS) algorithm [76], similarly to Chapter 2, to provide an order list of features, which can be truncated based on the desired number of features. The number of support vectors N_{SV} cannot be controlled directly when training SVM. As a control parameter, the smoothed separable case approximation (SSCA) pruning method is used to sweep the complexity of the SVM classifier, and thus indirectly modify the number of support vectors [77]. This method consists in discarding the support vectors that are close to the decision boundary in order to increase the separability of the dataset. The pruning parameter corresponds to the width of the margins around the decision boundary: a higher value for this parameter leads to wider margins and therefore a simpler model with less support vectors.

3.5.1 Hardware-accurate fast performance evaluation

As it is not possible to sum up the DBE performance in a low-complexity model, implementing an efficient evaluation method is necessary to avoid long simulation time. First, this method should allow to take different input signals, which correspond to the output of the AFE, including signal non-idealities (noise, distortion, interference, etc.). Second, the precision of the computations should match the actual hardware execution, e.g., with the same data types and models. Third, it should take into account the hardware limitations of the target platform, such as memory or real-time constraints. Finally, the simulation time should be faster than real-time hardware execution so that the performance can be assessed on the full MIT-BIH arrhythmia database used in Chapter 2.

The simulation of digital systems can be performed with different tools which each correspond to different abstraction layers and result in varying levels of accuracy and execution time, as illustrated in Fig. 3.15. At the bottom of the abstraction hierarchy, simulation of hardware description languages (HDL) such as Verilog or VHDL provides the exhaustive logic execution at the gate level in the processor, busses and peripherals. It therefore provides the most accurate results, at the cost of an execution time thousands of times slower than real-time hardware execution for basic MCU simulation. As an alternative to HDL simulation, SystemC is a set of C++ libraries that are meant to perform transaction-level modeling (TLM), i.e., system-level execution with the different modules and interfaces. By providing models of the hardware components (e.g., Cycle Models from ARM for Cortex processors), the outputs of the system remain accurate with a higher simulation speed, though still slower than real-time hardware. Real-time simulation can be performed with instruction set simulators (ISS) which rely on instruction-accurate models such as the Fixed Virtual Platforms (FVP) from ARM. Those simulators are usually included in dedicated integrated development environments (IDE) for embedded software design. While the simulated processing results remain accurate compared to the target hardware, the performance in terms of the number of cycles is only approximated. The evaluation can also be performed at the software level by emulating the execution of the C or C++ code on a computer, without considering the specificities of the target hardware. This approach enables an execution time faster than hardware execution as the same code is executed on a platform with much more computing resources. Finally, at the algorithmic level, high-level programming languages such as MATLAB or Python are used. These offer great versatility for data processing and fast simulation but are completely agnostic with regards to the hardware execution on the embedded device.

In this work, the inference performance evaluation is performed by the emulation of the C software code on a Linux server. Compared to an evaluation using a high-level programming language, the simulation time is slightly higher, but it allows to take into account some specificities of the software such as fixed-point calculations. As this emulation does not take into account the hardware constraints of the embedded device such as real-time execution or memory capacity, nor does it provide power or execution time estimates, those are computed using analytical regression models based on reference data points obtained from HDL simulation and logic

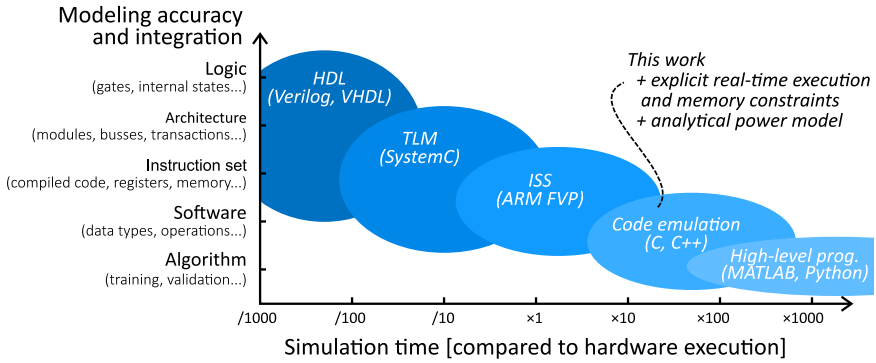

Fig. 3.15 DBE simulation tools. The orders of magnitude for execution time depend significantly on the simulated hardware, with regards to its scale or operation frequency, and on the simulation environment such as available computing resources, or target simulation accuracy.

synthesis. The evaluation framework shown in Fig. 3.16 is built in Python. The central part of this model is a Python C extension module which explicitly executes the software code of the arrhythmia classification system. This C module takes as inputs the samples to be processed, the software code and a model for the SVM classifier. For the arrhythmia classification use case, the samples are obtained from an ECG database, which can be possibly modified with analog non-idealities according to extrinsic parameters. The software code is the C code intended for execution on the embedded device, but compiled for emulation on the Linux server. The model of the SVM classifier includes support vectors, weights and hyperparameters and is trained according to the values of the intrinsic parameters. At the output of the C module, the labels inferred by the algorithm are compared with the reference annotations from the database to calculate the accuracy results. The SVM classification model is obtained by training a model with the features extracted by the C software from a training set. The training is performed in Python using the scikit-learn library [141]. While the training is performed in floating points, it is converted back to fixed-point for execution in the embedded device.

In order to check the hardware limitations which are not included natively by the C emulation approach, the constraints are calculated based on the characteristics from the trained SVM model. The total memory size depends on the baseline code size and the RAM space required for execution

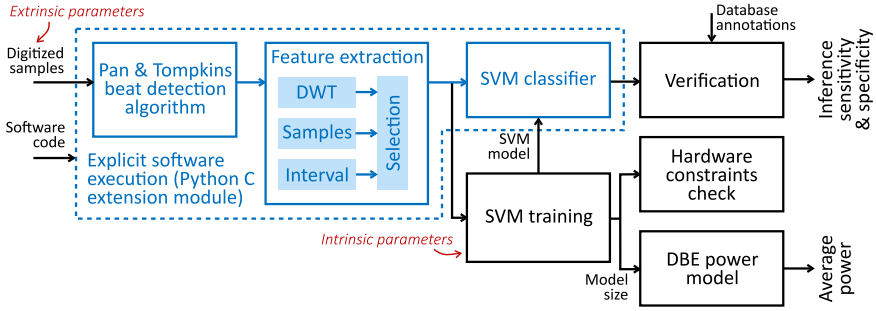

Fig. 3.16 DBE performance simulation framework. The framework is implemented in Python and allows to evaluate the inference performance for a given input dataset, according to the value of design parameters.

which both remain unchanged, and the SVM model data which depends on the training. The memory size variation is therefore directly obtained from the trained SVM model, and it is compared with the remaining memory capacity. Regarding the real-time execution constraint, the hardware execution time depends on the number of iterations required to compute the SVM decision function and on the data rate at the input. As those parameters influence linearly the execution time, a simple model calibrated with a few HDL simulations can be generated, which allows to check the constraint. This analytical approach to evaluate hardware constraints is only possible thanks to the regularity of the SVM model and parameters.

Fig. 3.17 shows the accuracy evaluation obtained by the DBE simulation tool. Sweeping the SVM pruning parameter and the number of training features allows to modify the classification model complexity, reflected by the increasing number of support vectors and features. The number of support vectors decreases by design with the pruning parameter, and increases with the number of features as the dimensionality of the separation boundary increases and requires more points to define it. Classification sensitivity and positive predictive values drop when the model complexity decreases too much and is therefore not capable of properly representing the classes, or when it increases too much and overfits the training data. The developed tool also checks the memory and real-time constraints, which is important in particular for complex models.

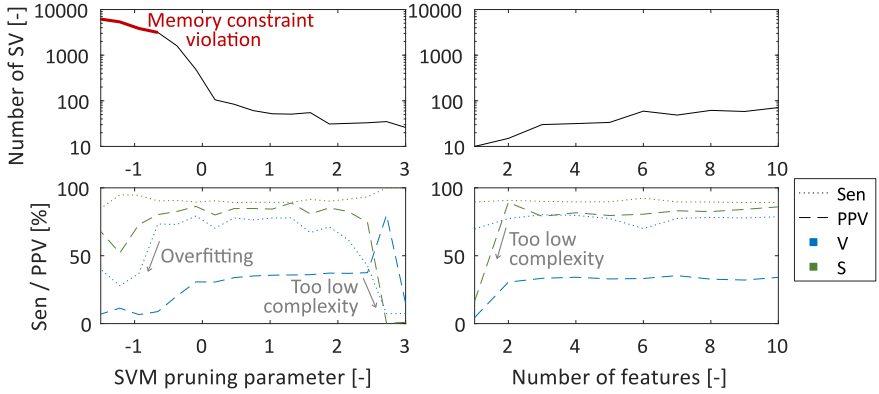

Fig. 3.17 DBE classification accuracy model. When changing the DBE parameters such as SVM pruning threshold or number of features, the accuracy degrades due to too high or too low model complexity. With low SVM pruning, the number of SVs increases dramatically and overflows the memory space.

3.5.2 Analytical power model

The C emulation approach proposed in the previous section does not provide the power consumption associated to software execution, nor the execution time which can be used as a proxy. Instead, a model of the power consumption based on known software parameters is used. The power consumption of the DBE is defined by the software code that implements the processing task, and the data rate at the input. The execution time and power of the software code depends on the complexity of the algorithm, which can be measured by the number of selected features N_{feat} and the number of support vectors N_{SV}. These parameters are available in the performance evaluation framework presented previously after training the model. The input data rate is given by the sampling frequency of the ADC F_s.

The total DBE energy consumption is equal to the sum of the energy required to perform the different sub-tasks: beat detection, feature extraction and SVM classification. For the beat detection, implemented by the algorithm from Pan and Tompkins [71] which mainly consists in a few filtering steps and an adaptive threshold, the total energy depends on the number of samples at the input, i.e., the sampling frequency F_s, and the energy per sample E_{det}. For the feature extraction, given that all computations for the

DWT and RR interval measurements need to be performed as soon as at least one such feature is selected, changing the number of features does not impact the associated energy. This is not a significant issue given that it only accounts for 16% of the DBE energy. For the SVM classifier using an RBF kernel, both the number of features and the number of support vectors impact the total energy per detected beat. The number of features modifies the calculation of the squared Euclidean distance, whose computation cost is proportional to the number of features:

$$d_i(\mathbf{x})^2 = ||\mathbf{x} - \mathbf{x}^{SV,i}||^2 = \sum_{j=1}^{N_{feat}} (x_j - x_j^{SV,i})^2. \tag{3.12}$$

The number of support vectors proportionally changes the calculation of the kernel function:

$$F(\mathbf{x}) = \sum_{i}^{N_{SV}} \alpha_i e^{-\gamma d_i(\mathbf{x})^2}. \tag{3.13}$$

The average power is therefore given by the following equation:

$$P_{DBE} = F_s E_{det} + F_{HR} \left(E_{FE} + N_{SV}(E_{dec} + N_{feat} E_{dist}) \right), \tag{3.14}$$

where F_{HR} is the heart rate, E_{FE} is the energy for performing feature extraction on a detected beat, E_{dec} is the energy per support vector to compute the SVM decision function and E_{dist} is the energy per feature and per beat to compute the Euclidean distance.

Fig. 3.19 illustrates the DBE power model by sweeping the number of features and support vectors. The constants in Eq. 3.14 are extracted from a few calibration points obtained with the place-and-route tool, Cadence Innovus, using activity annotation data from HDL simulations and standard-cell library characterization, as shown in Fig. 3.18.

3.6 Comparison of the approaches

In this chapter, several methods for evaluating and modeling the power and performance trade-offs of circuit blocks have been presented. While a unified approach applicable to any circuit block may be more convenient and appealing, the types and characteristics of circuits encountered in sensor systems are so diverse that it is difficult to find a solution that would fit them all. The different methods presented above can thus be chosen to

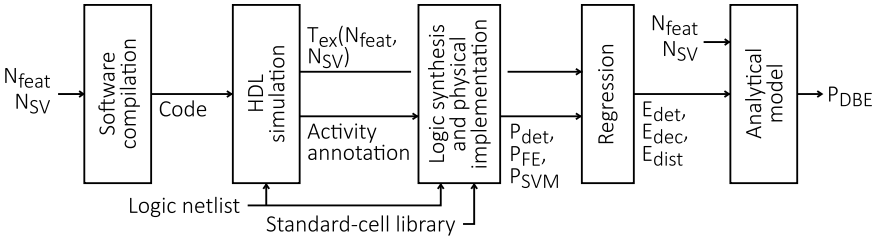

Fig. 3.18 DBE model calibration flow. The constants in the DBE analytical model are calibrated using data from digital simulation and implementation. P_{det}, P_{FE} and P_{SVM} are the average power consumption during the beat detection, feature extraction and classification steps.

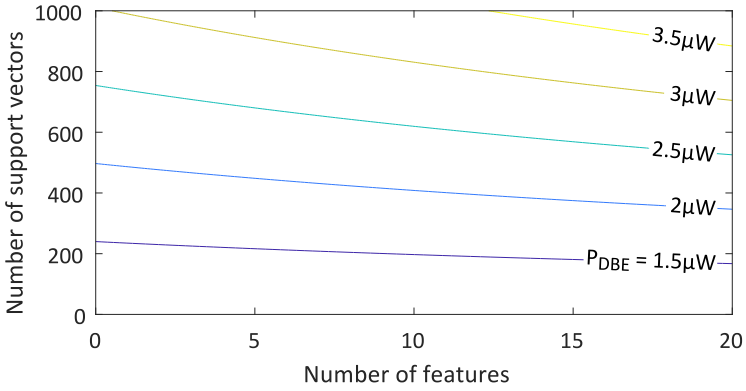

Fig. 3.19 DBE power model. The graph shows a contour plot of the DBE power consumption depending on the intrinsic parameters.

fit the context and requirements for distinct target circuits and models. To ease this choice, we discuss in this section some of the qualitative advantages and bottlenecks of those approaches, as summarized in Table 3.5.

The analytical approach has the main advantage of being the fastest to generate, at least for simple models. These preliminary models can then be iteratively refined to include second-order effects. This approach relies on the least number of simulations, limited to setting the constants in the equations and verifying the validity of the model. The major disadvantage of this method is that the quality of the model directly depends on the knowledge of the circuit. For that reason, it can hardly be applied to IP macros that are used as black boxes, or very complex circuits whose

Table 3.5 Comparison of modeling approaches

	Analytical approach	Genetic algorithm	Software emulation
Low number of simulations	✓	✗	✗
Explainability of the trade-offs	✓	✗	✗
Fast model evaluation	✓	✓	✗
Concurrent optimization	✗	✓	✗
No circuit knowledge required	✗	✓	✓
Allows modeling complex behaviors	✗	✓	✓
Model reacts to on input data	✗	✗	✓

behaviour is difficult to model. However, some simple models can be generated early in the design without having a functional implementation of the circuit, for example by considering general figures of merit (energy per conversion step for an ADC or noise efficiency factor of an IA).

The numerical optimization approach, implemented in this work with multi-objective algorithms, has the advantage of automatically extracting the trade-off curve represented by the set of Pareto-optimal points. In addition to modeling the trade-offs, this method also importantly performs an optimization of the circuit performance. It does not explicitly require designer knowledge about the circuit and is capable of generating relatively complex models. However, despite the improvements proposed in Section 3.4.2, the large number of circuit simulations required can take a significant amount of time.

The last approach of explicit emulation of the circuit execution is mainly useful when the previous approaches are unfeasible. This can be due to the dependency of circuit performance on input signals, as presented in the DBE. While the other approaches can provide the performance information instantaneously once they have been generated, the circuit emulation always relies on time-consuming simulations, even if they are accelerated on more powerful hardware.

It is important to note that the association between circuits and modeling methods we presented in this chapter are not the only possibility. For example, the noise and power trade-off in the instrumentation amplifier could have been modelled analytically by considering a constant NEF model. While it already provides an acceptable approximation, it can be seen that, compared to the more detailed model of Fig. 3.11, such analytical model would not be able to consider the limitations close to the boundaries of the design space. The other way around, the ADC could have been optimized using the numerical optimization approach with a genetic algorithm, or directly simulated similarly to the DBE. While it would have required more simulation time and reduced the understanding of the model, some additional second-order effects may have been included in the model.

3.7 Conclusion

Several methods for modeling the power and performance trade-offs of digital and analog circuits were presented, which all have their own advantages and limitations. They vary with each other regarding the achieved level of accuracy, simulation time, or required designer knowledge. In this chapter, three different types of circuits, from analog, mixed-signal and digital domains were evaluated with the respective approaches. This shows that the different approaches can apply to a wide range of circuits and design contexts. The models produced by these methods can then be used by the designer to select the appropriate power and performance trade-off for the system, possibly using hierarchical design methodology for making an informed choice. When several circuit alternatives are considered, for example with different topologies, comparing their trade-off curves also provides a more exhaustive comparison with regards to the range of performance that can be achieved.

In Chapter 4, the approaches presented to extract the power-performance trade-offs of the different circuit blocks are included in a system-level optimization methodology. By estimating the impact of local performance on the system accuracy and observing the related local power consumption, this methodology allows to find the optimum for the total power consumption that meets the global performance requirements.

4

Hierarchical cross-domain system optimization

4.1 Mixed-signal system optimization

Modern mixed-signal systems-on-chip (SoCs) such as microcontrollers, image sensors, or RF subsystems are usually designed in a hierarchical fashion. The top-level system is divided into several functions such as control, signal acquisition, data processing or communication, which are carried out by analog (amplifiers, oscillators, voltage references), digital (processor, logic accelerator) or mixed-signal circuits (analog-to-digital converters, phase-locked loops, comparators). Each circuit, possibly subdivided into smaller functional units, can then be designed at the device level by choosing the appropriate circuit topology and sizing. Many computer-aided design (CAD) tools have been implemented to help with the synthesis, layout, simulation or verification of these blocks, making it possible to deal with increasingly complex designs. While many aspects of the design at the circuit layer have been largely automated, especially for digital circuits, the system-level design still relies mostly on the designer knowledge. For instance, setting the performance specifications for individual circuits inside the system remains a critical and challenging task. Indeed, the importance of choosing these appropriately comes from the fact that

overall performance achieved by the system is a function of the block-level performance of all constituting circuits. On the one hand, relaxing local performance requirements allows to reduce the power consumption of these blocks, but may lead to a non-functional system. On the other hand, an overconstrained design often comes at the cost of power consumption overheads. Meanwhile, the challenge of setting these performance levels resides in the specificities of modern mixed-signal SoCs. For instance, the complex relationships between block-level and top-level performance can hardly be inferred directly. Digital and analog circuit design operate on different time and amplitude scales, which is reflected by their different sets of CAD tools. Full system-level performance evaluation through flat simulation is generally not possible due to the high number of devices.

While previous work in hierarchical system design focused on a single digital or analog domain, this chapter proposes a cross-domain methodology suitable for complex mixed-signal systems. It aims specifically at minimizing the total power consumption of biomedical sensor systems by choosing optimal performance requirements for its internal blocks. This methodology consists of three steps: (1) extracting power-versus-performance curves for the constituting circuits, (2) building a macromodel of the system performance and (3) using a numerical constrained optimization algorithm to find optimal system design parameters. The use of a macromodel, including here both analog and digital circuits, for abstracting the system operation allows fast performance evaluation at the system level. Gradient-based numerical optimization is used to efficiently navigate the system design space and find the global power minimum that meets the performance constraints. Assuming that performance trade-off curves have already been extracted for each circuit using for instance the approaches presented in Chapter 3, this chapter focuses on steps (2) and (3). As a demonstration of the proposed method, we apply it to the optimization of a cardiac arrhythmia classification system, with the goal of minimizing its total power consumption while maintaining a high inference accuracy. This system includes an analog-only circuit—the instrumentation amplifier (IA) in the analog front-end (AFE)—, a digital circuit with embedded software—the beat detection, feature extraction and classification tasks in the digital back-end (DBE)—, and a mixed-signal circuit—the analog-to-digital converter (ADC) in the AFE. Fig. 4.1 illustrates the optimization problem, in which the performance of the individual circuits can be tuned using a few high-level parameters : the IA bias current I_b, the

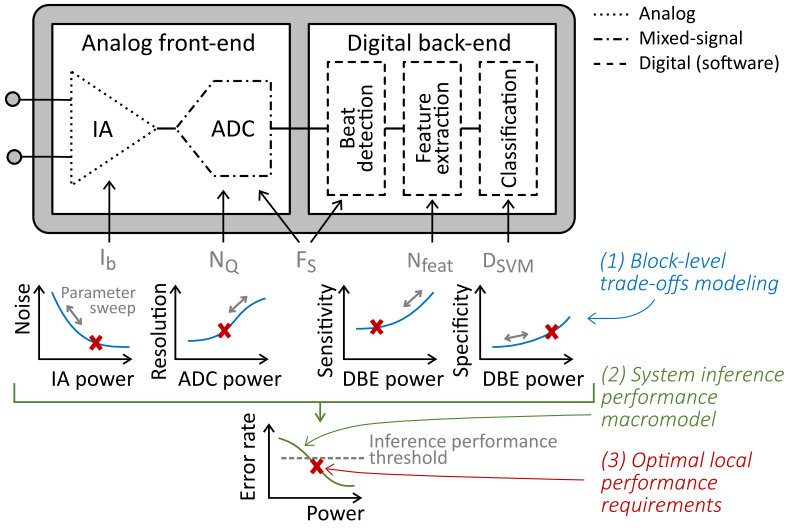

Fig. 4.1 Mixed-signal system optimization problem. High-level design parameters are used to tune the block-level performance trade-offs in blue. The optimal performance requirements in red, related to a corresponding set of design parameters, are found so that the total power in green is minimized while the global inference error rate is not excessively degraded. The baseline system design and operating conditions are similar to the system presented in Chapter 2.

number of bits N_Q in the ADC, the sampling frequency F_s, the number of features N_{feat} in the DBE and the support-vector machine (SVM) pruning threshold D_{SVM}.

This chapter presents the proposed methodology and its application to the cardiac arrhythmia classification system from Chapter 2. In Section 4.2, the optimization problem is described and formalized mathematically. Several optimization methodologies for hierarchical system optimization from the state of the art are compared and the main challenges specific to complex mixed-signal systems are highlighted. The proposed optimization methodology and implementation are detailed in Section 4.3. Starting from the performance models extracted in Chapter 3, a power and performance macromodel of the system is implemented in Python. The sequential least-square programming (SLSQP) algorithm chosen for the numerical optimization is then briefly presented and applied to this model to find the global design optimum to minimize power consumption. The results of

the system optimization and the performance gains obtained are outlined in Section 4.4. Finally, Section 4.5 discusses the advantages and limitations of the proposed methodology.

4.2 Optimization of a cardiac arrhythmia classification system

The optimization problem addressed in this work is the minimization of the power consumption of a cardiac arrhythmia classification system, similar to the one proposed in Chapter 2. The total power consumption is optimized by reducing the power consumption of its individual circuits in both analog and digital domains. However, generally speaking, reducing the power consumption of a circuit has a negative impact on its local performance. At the global level, a constraint is set on the inference performance which must remain close to the baseline performance obtained at a high power consumption, therefore limiting the local performance degradation. As the whole system design space is extremely large, the trade-off between power and performance is tuned in this work by a few selected high-level design parameters. In this section, the problem is detailed and formalized, before comparing several system design methodologies from the state of the art.

4.2.1 Mathematical formalization of the problem

The goal of the optimization problem is to minimize the power consumption of the system while maintaining the accuracy above a given threshold. Mathematically, this goal can be formalized as a single-objective minimization problem with inequality constraints, which can be written in the following way [63]:

$$\begin{aligned} &\min_{\mathbf{x}\in\Omega} f(\mathbf{x}), \\ \text{with } & f : \mathbb{R}^D \to \mathbb{R} \text{ and } \mathbf{x} = (x_1, x_2, ..., x_D) \in \Omega \subset \mathbb{R}^D \\ \text{subject to } & g_i(\mathbf{x}) \geq 0,\ i = 1, ..., M \\ & x_i^L \leq x_i \leq x_i^U,\ i = 1, ..., D \end{aligned} \tag{4.1}$$

with $\mathbf{x}$ the optimization variables, $f(\mathbf{x})$ the minimization objective function, $g_i(\mathbf{x})$ the constraint functions, Ω the parameter space, D the number of degrees of freedom and M the number of constraints. The origin pa-

rameter space Ω is defined here by the lower and upper bounds for each parameter, respectively denoted as x_i^L and x_i^U.

Applied to the cardiac arrhythmia use case, the vector $\mathbf{x}$ corresponds to the set of design parameters used to tune the performance level of each block. In this work, 5 parameters are chosen across the analog and digital domains ($D = 5$): the bias current I_b in the IA, the quantization resolution N_Q and sampling frequency F_s in the ADC, the number of features N_{feat}, and the SVM pruning threshold D_{SVM} in the classification software.

The total average power consumption of the system is simply expressed as the sum of the average power consumption of its internal blocks

$$\mathrm{P}(\mathbf{x}) = \mathrm{P}_{IA}(\mathbf{x}) + \mathrm{P}_{ADC}(\mathbf{x}) + \mathrm{P}_{DBE}(\mathbf{x}) + \mathrm{P}_0 \tag{4.2}$$

where $\mathrm{P}(\mathbf{x})$ is the system power consumption, P_{IA}, P_{ADC}, and P_{DBE} are the power consumption of the different blocks which depend on the design parameters $\mathbf{x}$, and P_0 is a constant power consumption which includes other parts of the system that are not subject to the optimization. In the rest of this work, P_0 is discarded as it does not impact the optimization task. The associated objective function to be minimized is chosen as

$$f(\mathbf{x}) = \log_{10}\big(\mathrm{P}(\mathbf{x})\big) \tag{4.3}$$

where a logarithm is used for scaling to ease the numerical optimization.

The performance of the arrhythmia classification system is measured by the accuracy of the output inferred beat classes. We focus specifically on the arrhythmia classes V and S, and not the "normal" N class, given to the large class imbalance in the dataset and the higher applicative importance of the arrhythmia events. The metrics for classification error rates are the false negative rate (FNR, also equal to 100% - sensitivity) and false discovery rate (FDR, equal to 100% - positive predictive value). The overall accuracy obtained at the system output depends on the local performance of the blocks:

$$\begin{aligned}
\mathrm{FNR}_V(\mathbf{x}) &= h_{V,FNR}(p_{IA}(\mathbf{x}), p_{ADC}(\mathbf{x}), p_{DBE}(\mathbf{x})) \\
\mathrm{FDR}_V(\mathbf{x}) &= h_{V,FDR}(p_{IA}(\mathbf{x}), p_{ADC}(\mathbf{x}), p_{DBE}(\mathbf{x})) \\
\mathrm{FNR}_S(\mathbf{x}) &= h_{S,FNR}(p_{IA}(\mathbf{x}), p_{ADC}(\mathbf{x}), p_{DBE}(\mathbf{x})) \\
\mathrm{FDR}_S(\mathbf{x}) &= h_{S,FDR}(p_{IA}(\mathbf{x}), p_{ADC}(\mathbf{x}), p_{DBE}(\mathbf{x}))
\end{aligned} \tag{4.4}$$

where $h(\cdot)$ are a priori unknown functions that map block-level with system-level performance, and p_{IA}, p_{ADC}, and p_{DBE} are the individual block performance levels. Four constraint functions are chosen so that the respective inference error rates do not increase by a certain percentage above their baseline, as defined by the following equations:

$$\begin{cases} \mathrm{FNR}_V(\mathbf{x}) \leq (1+\alpha)\mathrm{FNR}_V^0 \\ \mathrm{FDR}_V(\mathbf{x}) \leq (1+\alpha)\mathrm{FDR}_V^0 \\ \mathrm{FNR}_S(\mathbf{x}) \leq (1+\alpha)\mathrm{FNR}_S^0 \\ \mathrm{FDR}_S(\mathbf{x}) \leq (1+\alpha)\mathrm{FDR}_S^0 \end{cases} \Rightarrow \begin{cases} g_1(\mathbf{x}) = \log_{10}\left(\dfrac{(1+\alpha)\mathrm{FNR}_V^0}{\mathrm{FNR}_V(\mathbf{x})}\right) \geq 0 \\ g_2(\mathbf{x}) = \log_{10}\left(\dfrac{(1+\alpha)\mathrm{FDR}_V^0}{\mathrm{FDR}_V(\mathbf{x})}\right) \geq 0 \\ g_3(\mathbf{x}) = \log_{10}\left(\dfrac{(1+\alpha)\mathrm{FNR}_S^0}{\mathrm{FNR}_S(\mathbf{x})}\right) \geq 0 \\ g_4(\mathbf{x}) = \log_{10}\left(\dfrac{(1+\alpha)\mathrm{FDR}_S^0}{\mathrm{FDR}_S(\mathbf{x})}\right) \geq 0 \end{cases} \tag{4.5}$$

where the parameter α represents the maximum degradation from the baseline error rates FNR_V^0, FDR_V^0, FNR_S^0 and FDR_S^0. A typical value for α is 50%.

4.2.2 State of the art for system design and optimization

The problem of choosing block-level specifications to ensure that circuit design leads to the expected system performance is a typical issue in hierarchical system design. Several design methodologies have been proposed in the literature, as shown in Fig. 4.2(a). Those methodologies are mostly derived from the two most common bottom-up and top-down approaches, on each extremity of the illustration.

On the left, in the bottom-up approach, the designer starts designing the constituting blocks with only a limited set of a priori specifications [45, 47]. It is only when the blocks are assembled together that the overall performance can be evaluated, often leading to subsequent design re-spins if the performance is not sufficient compared to the expected specifications. The alternative is to over-design the individual blocks from the start, which likely results in sub-optimal designs. Despite these disadvantages, bottom-up design is still commonly used in practice [142]. The multi-objective bottom-up (MUBU) methodology extends the bottom-up approach by considering several circuit designs with a parametrized circuit sizing to represent a wider range of circuit performance [47]. Instead of propagating a single design point per block at the higher level as in the original bottom-

up approach, several points corresponding to different performance trade-offs are chosen to represent the Pareto performance of the circuit. The trade-offs are then combined at the system level, which provides a wider range of achievable performance and reduces the likelihood of not meeting the specifications. However, the iterative combination of multiple circuit performance levels across the hierarchy can lead to a large number of possible combinations, which is time-consuming when all options are considered, even when macromodels are used to emulate the block performance.

On the right side of Fig. 4.2, the top-down approach, sometimes referred to as the top-down constraint-driven (TDCD) method in analog design [143, 144, 145], consists in starting from the system-level specifications and progressing downwards in the hierarchy to define the performance levels of the blocks at the lower level. At each level, these required performance levels become the constraints for designing low-level blocks. The main difficulty lies in the translation of specifications from the higher-level performance to the levels below, especially when the interactions between the blocks are complex and non-linear. This translation can be done analytically [145, 146] or using simulations [45]. For simulation-based approaches, parametrized behavioral macromodels of the different circuits are used to evaluate the impact of local performance on the global system. Once the specifications have been propagated, the lower-level designs are then implemented to meet the specifications. However, these specifications may not always be feasible, which then requires to go up the hierarchy to modify the expected performance or architecture. Despite the difficulty of translating specifications through the hierarchy, the advantage of the top-down approach is that the final solution often ends up closer to desired system specifications than in bottom-up design. Other works included a preliminary feasibility modeling step to evaluate the regions of the circuit performance space that can be achieved, in order to constraint the specification translation [48]. However, modeling the whole performance space of several circuits is a tedious tasks, especially when the complexity of the circuit increases with many parameters and performance metrics.

In between bottom-up and top-down methodologies, other hybrid approaches have been proposed as attempts to ensure feasibility of the design and optimality of the results with respect to the target system specifications. One of these is the concurrent constraint-driven methodology that aims at setting simultaneously the block-level specifications and lower-level design parameters to ensure feasibility [147]. To do so, a numeri-

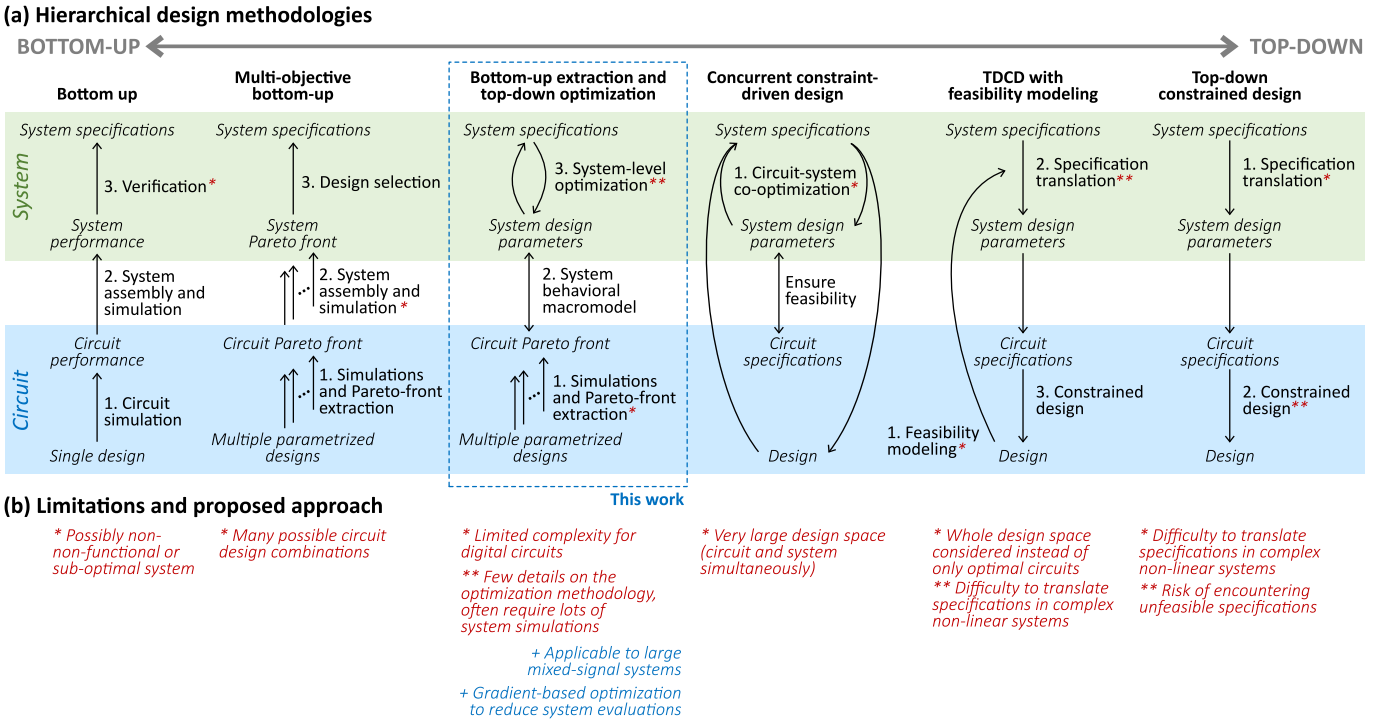

Fig. 4.2 State of the art of hierarchical system design methodologies. References are ordered along the bottom-up to top-down axis from left to right (a). The arrows represent the inference direction during the design process. The main limitations are highlighted by red asterisks and detailed at the bottom (b).

cal optimizer is used to minimize a cost function which includes both the achieved top-level performance obtained with a set of circuit specifications, and the difference between the target and actual performance at the block-level to make sure that they match. In this approach, both system and circuit levels are optimized together, meaning that the optimization variables include the local performance requirements and block-level design parameters. This quickly leads to a very large number of degrees of freedom which makes the optimization difficult for larger designs.

Finally, some works combined the bottom-up extraction of circuit Pareto performance, similarly to the MUBU approach, with a top-down optimization of the system-level design. For instance, Yu and Li [148] designed a phase-locked loop (PLL) by relying on a bottom-up regression approach to model yield-aware Pareto fronts of analog and digital circuits, a behavioral model of the system in Verilog-A to simulate the top-level PLL performance and a global optimization at the system-level with multilevel coordinate search. Zou *et al.* [149] also demonstrated the design of a PLL, using a linear model of the Pareto fronts of analog circuits, Verilog-A behavioral modeling and top-level optimization. These approaches are however dedicated to circuits which are mostly analog with small digital domains. Only a few details are provided with regards to the implementation of the optimization method, which is a decisive factor of the final design.

While those methods have all been studied in the case of analog system design, they have not been applied to more complex mixed-signal systems with large digital circuits involving signal processing. As many methods rely on the modeling of direct input-output relationships specific to analog circuits, they are hardly applicable in the case of more complex digital circuits, especially when unpredictable machine-learning algorithms are used. In addition, most methodologies are also limited to finding a design that meets given system specifications, but do not include the optimization of the system-level performance.

4.2.3 Challenges

Setting local performance requirements in order to optimize performance at the system level is not a straightforward task, especially in complex mixed-signal circuits which have not been addressed in previous works. Indeed, the application of hierarchical design methodologies to cross-domain optimization encounters several specific challenges.

System complexity Firstly, while the power consumption of the system is only the sum of the power consumption of its constituents, the system-level performance is often a complex non-linear function of the local performance from each block. While analog circuits can generally be modeled by first-order input-output relationships, the complexity dramatically increases when digital processing algorithms such as machine learning classification are used. For example, in the cardiac arrhythmia classification system described in Chapter 2, the relationship between the IA intrinsic noise and the classification accuracy of the SVM can hardly be known without extensive simulations.

Multiple circuit domains Secondly, sensor systems with embedded signal processing usually consist in mixed-signal systems with circuits in both the analog and digital domains. The simulation tools that are used for each domain are not the same, as they operate on different abstraction levels and target different scales, both in time and amplitude. On the one hand, analog and mixed-signal circuit design usually relies on SPICE simulators that compute the voltage and currents at the transistor level. On the other hand, digital tools simulate logic states in the circuit at the register-transfer level (RTL) or gate level. Finally, algorithm design is usually performed at a higher level of abstraction, using numerical computing software such as Matlab or Python, which rarely takes into account directly the physical hardware limitations of the target platform. This variety of design environments and contexts makes the simultaneous simulation of all system components practically unfeasible without additional abstraction layers. While macromodels have often been used to emulate the block-level behavior in a single low-level simulation, the implementation of these models in a time-efficient manner with accurate performance evaluation is more difficult when multiple domains are considered simultaneously.

Simulation time Thirdly, even at the block level, performance simulation can be time-consuming, depending on the type of simulation that is required. For instance, long transient or Monte-Carlo simulations for analog circuits, simulations of large digital circuits for complex tasks with millions of clock cycles, or validation of machine-learning algorithms on large datasets can each require a lot of time, up to several minutes, hours or days long in some cases. The number of performance evaluations must therefore remain limited.

The proposed methodology answers the previous challenges in the following way. The system complexity is managed by a simulation-based ap-

proach that allows to emulate the non-idealities of the different blocks and their interactions. The exploration of the design space and its optimization is performed by a numerical algorithm. The analog and digital domains are abstracted concurrently in a global performance model that simulates efficiently the analog transfer functions and non-idealities, along with the digital processing algorithm. Local models of the local power and performance trade-offs of the different blocks are built and used in the system model to reduce the simulation time at the circuit level.

4.3 Proposed optimization methodology

The proposed methodology, illustrated in Fig. 4.1, is hierarchical and divided in three steps. The first step consists in the bottom-up modeling of local power and performance trade-offs of the blocks in the target system, similarly to the MUBU approach. Different methods to perform this performance space modeling have been described in Chapter 3. The second step aims at modeling the global system power consumption and inference performance as a function of the local performance characteristics of the individual blocks extracted in the first step. Using a simulation-based approach, a cross-domain macromodel is implemented to take into account both the analog behavior and non-idealities, and the digital signal processing. The third step uses this performance macromodel and applies a numerical optimization algorithm to find the optimum set of design parameters. As the first modeling step has already been explored in Chapter 3, Sections 4.3.1 and 4.3.2 will describe the macromodel implementation and numerical optimization steps, respectively.

4.3.1 System power and performance macromodel

In order to evaluate the power consumption and inference performance at the system level, the respective block-level characteristics need to be combined together. On the one hand, the total power consumption of the system can be easily evaluated as the sum of the power for each block (Equation 4.2). On the other hand, the global inference performance, taking into account both the non-idealities in the AFE and the classification performance in the DBE, is a complex function of the local block performance, which can hardly be estimated a priori (Equation 4.5). When analytical expressions of system-level performance cannot be obtained and flat simulation of the whole system is not possible due to the large number of devices,

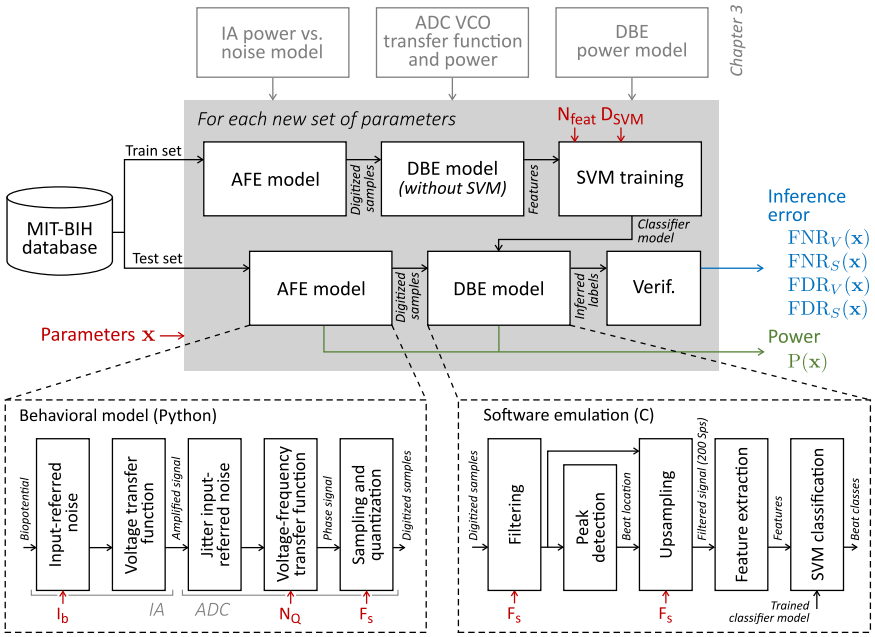

Fig. 4.3 System power and performance macromodel. The system performance is obtained by applying behavioral models of the different components to the database signals. The power consumption is obtained from the models based on the data of the test set. System design parameters can be modulated in their respective blocks. The light boxes correspond to the model calibration data obtained from simulation, typically SPICE, and only need to be extracted once at the beginning.

macromodels have been extensively used in the literature to emulate the circuit blocks behavior and estimate the system performance through simulation [45]. The challenge of these macromodels is the trade-off between complexity, with regards to both implementation and simulation, and precision. For example, a baseline model for an analog amplifier may emulate basic gain characteristics, but discard non-idealities such as distortion, noise or limited bandwidth. The level of details of the circuit macromodels impacts the precision of the system performance evaluation. While macromodels have been widely used for analog or small mixed-signal circuits, they are more rarely used for digital circuits.

A numerical model of the system, shown in Fig. 4.3, is built to simulate the system performance. This model combines the behavioral models of

the AFE, including several non-idealities, the hardware-aware execution of the embedded C code, and the training and evaluation of the machine-learning classification algorithm. The model wrapper is implemented in Python for its suitability for data processing. As shown in Fig. 4.3, the MIT-BIH arrhythmia database is used as input[1], given its frequent use in the arrhythmia classification literature (see Chapter 2). In the AFE, the acquisition of the database signals are simulated using behavioral models of the different blocks, including analog non-idealities. In the IA, the thermal and flicker input-referred noise levels are modeled, as well as the transfer function including signal distortion. Common-mode rejection is not included in the model as it requires a template for common-mode artefacts (such as powerline interference or motion), which is not trivial to define. For the ADC, the voltage-controlled oscillator jitter and voltage-to-frequency transfer function simulate the oscillator frequency signal. This signal is then integrated, sampled and quantized to emulate the ADC behavior. The constants of the numerical model are set so that behavioral characteristics are matched with the simulation results of the IA and ADC.

In the DBE, the accuracy cannot be directly modeled as a function of the DBE design parameters due to its dependence on input signals. The software is therefore explicitly emulated using a Python C extension module that allows to execute custom C code, as explained in Section 3.5. The simulation of the system inference accuracy is performed in two stages for training and testing the SVM classifier, according to the data division scheme in [74]. First, the signals from the training part of the database are subject to the AFE model. Using the digitized samples at the output of the AFE, the DBE model extracts the heartbeat features only, without performing the classification. The SVM classifier is trained on this set of features. Secondly, the testing signals from the database go through the AFE and DBE model, including the updated classifier model, to extract the inferred class labels. The FNR and FDR metrics are computed by comparing the model output with the database annotations.

In order to evaluate the system power and performance as functions of the block-level performance (Eq. 4.2 and 4.4), the block-level models are parametrized. The level of IA noise is modified by setting the amplitude of the generated random white and pink noise added at the IA model input.

[1]The signals in the MIT-BIH arrhythmia database were digitized at 360 Hz with a resolution of 11 bit over a 10 mV range. This recording performance sets the maximum signal quality that can be modeled by the AFE.

The corner frequency of flicker noise is kept unchanged. The power consumption associated to the IA is obtained by interpolating the Pareto curve obtained in Section 3.4. The quantization resolution of the ADC is modified by controlling the gain of the BBCO transfer function for the voltage-to-frequency conversion. The sampling frequency is directly modified in the sampling step. When changing the frequency, the BBCO transfer function is also modified to fit the target quantization resolution (Eq. 3.1 in Section 3.3). The ADC power is calculated based on Eq. 3.10 in Section 3.3, with the constant BBCO power consumption and the dynamic power in the counter proportional to the average BBCO output frequency. In the DBE, the filters used for beat detection are modified with the sampling frequency to maintain their respective cut-off frequencies. When the sampling frequency is reduced, the digitized signal is upsampled after the detection stage and before feature extraction and classification. This allows the feature extraction stage to remain unchanged and avoids multiple time-consuming feature selections. The number of features and the pruning of the SVM are included in the training phase of the classifier, and impact the power and performance of the DBE through the modified model. The power model takes into account the data rate, the number of features, and the number of support vectors (Eq. 3.14).

4.3.2 Optimization algorithm

The simulation of the system power and performance is time-consuming, despite the use of a macromodel as described in the previous section. This long execution time is due to the AFE behavioral model that must be applied to the 21 hours of ECG signals, the execution of the C software for beat detection and arrhythmia classification executed on the whole database, and the training of the SVM. Previous approaches for system design that rely on an exhaustive grid search of the high-level parameters are therefore unfeasible because of the thousands of evaluations required. In order to limit the number of system evaluations and converge as efficiently as possible towards the optimal design point, a numerical optimization tool is used. This section briefly summarizes the main features of the selected algorithm, and details the implementation of the system optimization process in connection with the system evaluation model.

Algorithm selection and description

The algorithm must solve the problem described in Equation 4.1, using the system macromodel to evaluate the objective and constraints functions,

i.e., $f(\mathbf{x})$ and $\mathbf{g}(\mathbf{x})$. This problem is characterized by a single, non-linear, and continuous objective function with inequality constraints. The derivative of the function is not directly available, but it can be estimated numerically using finite differences. There are many optimization algorithms available, each featuring different characteristics with regards to the type of problems they solve, the optimality of the solutions, the assumptions made on the problem, or the convergence speed. The two main classes of algorithms are the local algorithms that use or approximate function derivatives to find the best search direction, and global algorithms that rely on heuristics to explore the design space [150]. Global optimization algorithms, most of which consist in evolutionary algorithms, are very versatile as they usually make few assumptions on the problems and can also fit many different use cases such as multi-objective optimization or noisy objective functions. However, as most of them are population-based, they often require many evaluations to converge towards the optimum. On the other hand, local algorithms usually require access to the function gradient or its estimate, and they are more sensitive to noise and local optima. When they can be applied, they however often provide guarantees regarding the optimality and convergence speed, which is useful when the evaluations are computationally expensive, as in this problem [150, 151].

Gradient-based solutions generally work by following successive descent directions for the objective function starting from an initial estimate, until a local minimum is reached. In unconstrained optimization problems, a necessary condition at a local optimum of a function is that the gradient of this objective function is equal to zero. In the case of problems with inequality constraints, this condition may not be met when the optimum lies on the boundary of the feasible region, i.e., the region where the constraints are respected $g_i(\mathbf{x}) \geq 0$. The necessary conditions for a local optimum $\mathbf{x}^*$, known as the Karush-Kuhn-Tucker (KKT) conditions, therefore become

$$\nabla f(\mathbf{x}^*) + \sum_{i=1}^{M} \lambda_i \nabla g_i(\mathbf{x}^*) = 0 \text{ and } \lambda_i g_i(\mathbf{x}^*) = 0, \forall i, \tag{4.6}$$

where λ_i are the Lagrange multipliers ($\lambda_i \geq 0$) [151]. When one or several inequality constraints are active, i.e., $g_i(\mathbf{x}^*) = 0$, these KKT conditions allow the gradient of the objective to be non-zero and parallel to a linear combination of the active constraints gradients. In order to solve the constrained optimization problem and find the $\mathbf{x}^*$ solution of Eq. 4.6, an equivalent problem consists in finding a saddle point of the Lagrangian function

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}) \,. \tag{4.7}$$

While the Lagrangian provides a convenient approach for handling inequality constraints, finding the saddle point remains a challenging task in the case of nonlinear functions $f(\mathbf{x})$ and $g_i(\mathbf{x})$ [150].

Among local gradient-based methods, the most popular class of algorithms for constrained optimization problems is the sequential quadratic programming (SQP) algorithm [63]. As most gradient-based methods, SQP works by iteratively converging towards the minimum by choosing successive step directions $\mathbf{s}_k$ and sizes α_k:

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{s}_k \tag{4.8}$$

SQP uses the Lagrangian approach (Eq. 4.7) to manage inequality constraints. To deal with nonlinearity issues, a quasi-Newton optimization approach is used, meaning that it considers at each step a local quadratic approximation of the Lagrangian function in order to find the optimal step direction $\mathbf{s}_k$. It can be shown that finding the step direction that minimizes this approximation is equivalent to solving the following quadratic problem with linear constraints [63]

$$\begin{aligned} &\min_{\mathbf{s}_k} \ \frac{1}{2} \mathbf{s}_k^T B_k \mathbf{s}_k + \nabla f(\mathbf{x}_k)^T \mathbf{s}_k, \\ &\text{subject to } \nabla g(\mathbf{x}_k)^T \mathbf{s}_k + g(\mathbf{x}_k) \geq 0, \end{aligned} \tag{4.9}$$

where B_k approximates the Hessian matrix of the Lagrangian

$$B_k \approx \nabla \mathcal{L}^2_{xx}(\mathbf{x}_k, \boldsymbol{\lambda}_k) \,. \tag{4.10}$$

The optimization subproblem of Eq. 4.9 is a standard form of quadratic programming problem, which benefits from many resolution algorithms. One such implementation of the SQP optimization algorithm, called sequential least squares programming (SLSQP), is available in the Python SciPy libraries [152]. This implementation solves the quadratic programming problem with a linear least squares subproblem [63]. Contrary to the Newton method, quasi-Newton methods do not explicitly calculate the Hessian of the function $\nabla \mathcal{L}^2_{xx}(\mathbf{x}_k, \boldsymbol{\lambda}_k)$ which is computationally expensive, but rather estimate it iteratively with a matrix B_k. In SLSQP, this estimation is performed by Powell's update for constrained optimization of

the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [63]. When $\mathbf{x}_k$ is close to the optimum, the quadratic problem of Eq. 4.9 provides the optimal direction with step size $\alpha_k = 1$. Far from the optimum, step sizes improving global convergence can be found by solving a line search minimization problem with a penalty function for the constraints:

$$\min_{\alpha_k} f(\mathbf{x}_k + \alpha_k \mathbf{s}_k) + \rho \sum_{i=1}^{M} \max\big(0, g_i(\mathbf{x}_k + \alpha_k \mathbf{s}_k)\big), \tag{4.11}$$

with ρ the penalty parameter.

Optimization framework implementation

The optimization framework is implemented in Python using the SLSQP algorithm from SciPy [152] as shown in Fig. 4.4. Wrapper functions are used to interface the Python macromodel described in Section 4.3.1 and compute the functions $f(\mathbf{x})$ (Eq. 4.3) and $g_i(\mathbf{x})$ (Eq. 4.5). When the optimization algorithm calls the objective or constraint function for a given set of parameters, the macromodel evaluates the power and performance metrics. Previous results from the macromodel are saved to avoid multiple runs for the same parameter sets and ensure the coherence between power and constraint evaluations.

In order to have consistent scales for all free parameters, the design parameters are normalized between $x_i^L = 0$ and $x_i^U = 1$. The normalization is performed such that the lower bound x_i^L generally corresponds to the low-performance low-power limit and the higher bound x_i^U to high-performance high-power. The respective bounds used for each parameter are summarized in Table 4.1.

When no gradient function is provided, the SLSQP implementation uses by default a two-point finite difference approximation, which requires one additional call per parameter to the function wrappers and is time-consuming. In particular, the training of the SVM is a major contributor to the simulation time (57%). However, when modifying the parameters that impact the training features by only a small amount as in the finite difference estimation, the changes to the SVM model are not significant. This observation is leveraged to reduce gradient estimation time by skipping model training and keeping the model from the central point. These multiple training phases are however still required for the parameters that directly impact the SVM model, such as the number of features or the SVM pruning parameter.

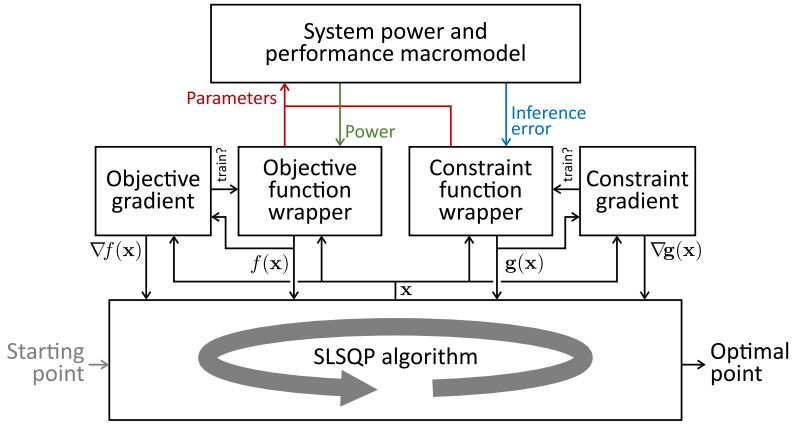

Fig. 4.4 System optimization framework. The SciPy SLSQP algorithm implementation uses wrapper functions for the objective and constraint functions, and their custom gradient estimation computed with finite differences. The power and performance simulation is performed using the parametrized system macromodel. Training flags are added to the custom finite difference gradient estimation to skip unnecessary training.

Table 4.1 Optimization parameter ranges. After normalization, the lower x_i^L and higher bounds x_i^U respectively correspond to the low-performance low-power and the higher bound to high-performance high-power limits.

Parameter	$x_i^L = 0$	$x_i^U = 1$
IA input-referred noise* [μV RMS]	0.6	30
ADC quantization resolution [bits]	7	14
ADC sampling frequency [Hz]	10	200
Number of features [-]	1	10
SVM pruning threshold [-]	3	-1.5

* Normalized using a logarithmic scale

4.4 Results

The proposed methodology for the minimization of the system power consumption was applied to the arrhythmia classification system. This section first describes the simulation results obtained with the system power and performance macromodel. The proposed implementation is characterized

with regards to the simulated trade-off results, simulation time and sensitivity to implementation choices. Secondly, the results obtained with the optimization algorithm are detailed with a focus on the convergence accuracy and speed. The application specifications such as target inference performance or input signal amplitude are also modified to evaluate the changes in the returned optima. Finally, the power reduction obtained in the arrhythmia classification use case is discussed.

4.4.1 Power and performance simulation

Exhaustive grid simulation of the parameter space is unfeasible due to the simulation time of the macromodel. Instead, the impact of each parameter on the performance and power of the system can be evaluated separately by sweeping each of them individually from a reference point, as shown in Fig. 4.5. As expected, the power decreases when each parameter goes closer to zero. Simultaneously, the inference performance degrades, as shown with the increasing error rates. For the SVM pruning parameter, the performance also degrades at values close to one, due to overfitting when the number of support vectors increases too much. The noise on the power and error curves is mainly due to the random noise added in the AFE and the sensitivity of the SVM training to small variations in the training set.

A macromodel of the system that allows to simulate the combined effects of several parameters is specifically useful if those parameters have interdependent effects on the objective or constraint functions. Indeed, inversely, the problem could be easily decomposed in several one-dimensional subproblems, in which each parameter could be evaluated and optimized independently. In the arrhythmia classification system, regarding the objective function, each block contributes independently to the total power consumption and a global model provides limited additional insights. The accuracy constraints however consist in more complex relationships between parameters, which strongly benefit from such model for their evaluation. Fig. 4.6 provides an example of how the boundary of the feasible space is curved when considering AFE parameters, thus showing that the global optimum may be different from the independent univariate optima.

The use of macromodels primarily allows to accelerate the evaluation of the system performance compared to flat simulation which becomes unfeasible for larger systems. Yet, the execution time of the model is constrained

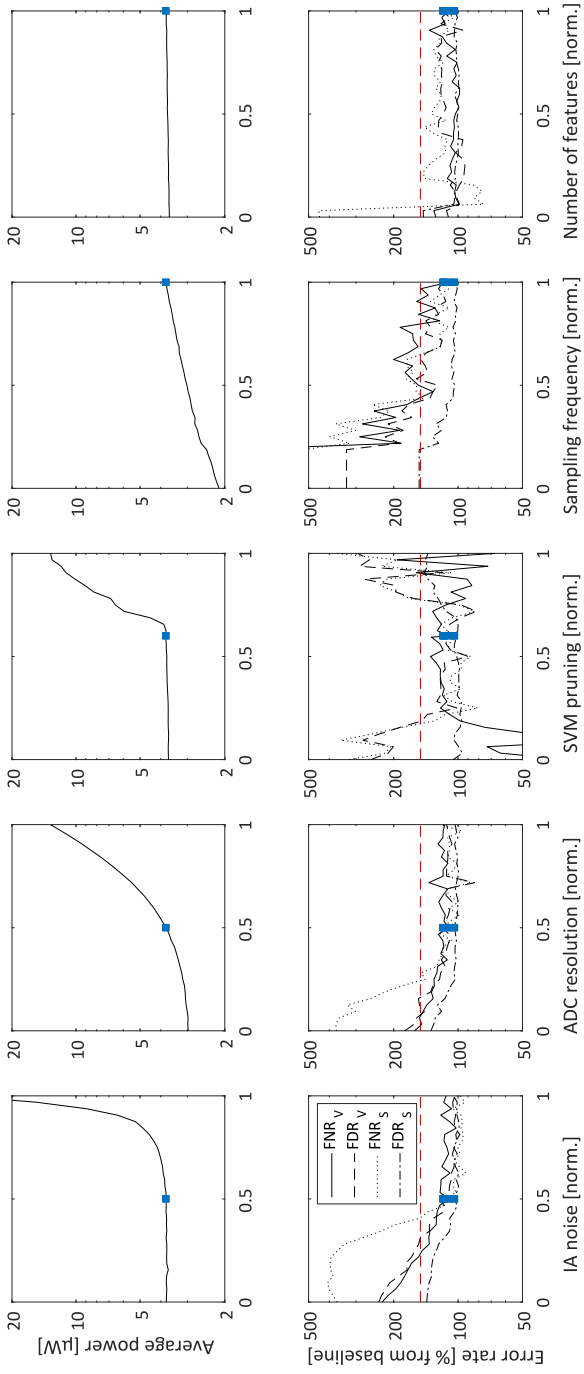

Fig. 4.5 Independent parameter sweep of the macromodel. All parameter sweeps are performed around the reference point (5-μV RMS IA IRN, 10.5-bit ADC resolution, 0.3 SVM pruning threshold, 200-Hz sampling frequency and 10 features per classifier). A new SVM model is trained for each parameter value.

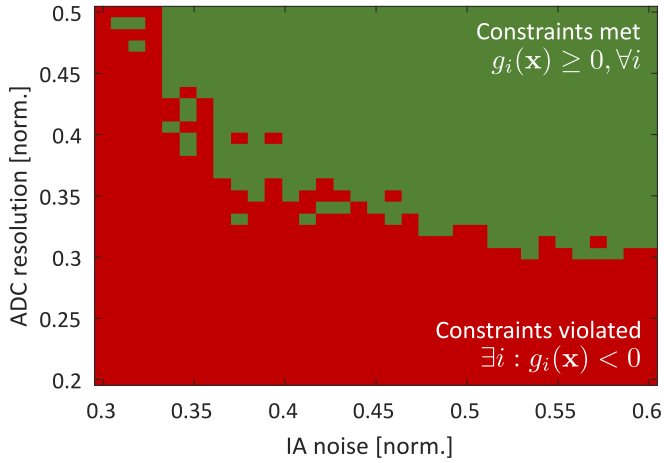

Fig. 4.6 Feasible space with AFE parameter sweep. Green and red points represent the AFE parameter sets that meet or violate the constraints, respectively. Other parameters are set to their reference value. The curved boundary of the feasible space shows that the constraints do not apply independently to the different parameters.

by the precision of the signal quality simulation and the large amount of data required to evaluate machine-learning performance. The macromodel simulation time for the whole MIT-BIH arrhythmia database (42 records of 30 minutes) is shown in Fig. 4.7(a), with the respective shares of AFE and DBE models, and SVM training.

Due to the explicit simulation of noise in the model, the simulated accuracy performance is not completely deterministic. The sources of uncertainty in the model are the additive random noise in the AFE for the IA intrinsic noise and the ADC BBCO jitter, as well as the SVM training which is sensitive to small differences in the training data. At each call of the macromodel, a different realization of the noise and therefore a different estimated performance is obtained, as illustrated in Fig. 4.8 when the evaluation of a reference point is repeated multiple times. This simulation noise amounts to a standard deviation of 0.3% for the power consumption and 2 to 7% for the error rate.

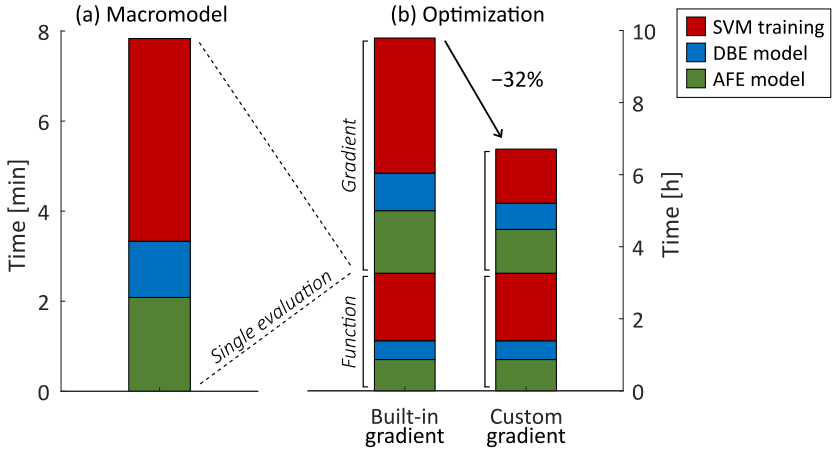

Fig. 4.7 System simulation (a) and optimization (b) time. The system evaluation is repeated for different parameter sets during optimization to obtain the function values and gradients. Avoiding unnecessary training during gradient evaluation reduces the total optimization time. SLSQP algorithm computation time during optimization is negligible. The execution time is measured on a single 2-GHz Intel Xeon E5-2620 CPU.

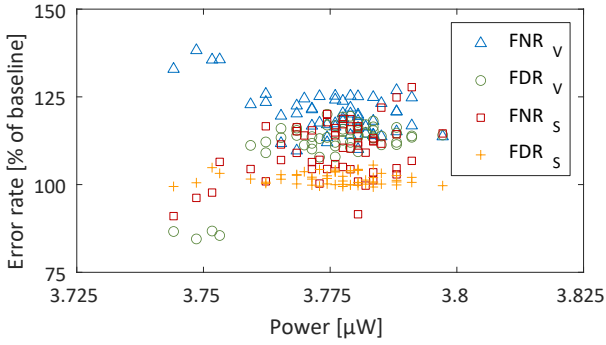

Fig. 4.8 System performance evaluation noise. The performance macromodel was run 50 times at the reference point. Each point represents a single evaluation, for one of the accuracy metrics.

4.4.2 System power optimization

The optimization algorithm described in the previous section was executed to find the optimum for the power-accuracy trade-off of the system using the macromodel. A single run requires on average 25 function and 10 gra-

dient estimations, and takes 6.7 hours to complete, as shown in Fig. 4.7(b). The optimization was run 55 times to check the robustness of the results. Among those runs, 26 failed, which is either due to the number of SLSQP iterations which reached the maximum value of 15 (15%), or a false local minimum far away from the true optimum (27%). Figure 4.9 shows the distribution of the design parameters, objective and constraints for the 29 successful runs. These results show that despite some variability in the optimum parameters, these runs converged close to a median value of 3.3 μW, with the error rate constraints met in most cases. Three runs showed a slight violation of the constraints ($< 11\%$) due to the algorithm tolerance. The distribution of the parameters shows that their characteristics in the system model influences how they can be optimized by the algorithm. In the case of the AFE parameters (IA noise and ADC resolution), most points gather around the median values, respectively 0.55 (i.e., 5.2 μV RMS of noise) and 0.34 (i.e., 9.4 bits of resolution). This can be interpreted in the model (Fig. 4.5) by the higher sensitivity of the objective function, measured by their higher partial derivatives. Regarding the constraints, the simulation noise is lower compared to the function derivative. On the opposite, the SVM pruning parameter is distributed between the values of 0.2 and 0.7 (SVM pruning parameter ranging from -0.15 to 2.1). In the model, the partial derivative is much lower close to the optimum parameters as the number of support vectors evolves very slowly below a value of 0.6. The noise on the constraint function is relatively larger compared to the function which is almost flat in this range. Finally, the noise and rapid degradation of the constraints for the sampling frequency and number of features forces them to stay close to their starting value of 1.

The different evolutions of the design parameters during the optimization is also reflected in their respective share of the power breakdown in Fig. 4.10. The AFE parameters start from a high power consumption due to the initial high signal quality, and are degraded by the optimization algorithm. The power consumption of the DBE only evolves very slightly from its starting value. At the end, the power consumption of the AFE and DBE is well balanced. The actual savings for the total power, here divided by a factor $\times 6$, are very dependent on the choice of the starting point and the target conditions of the system, which define the optimal design. A starting point with high performance is chosen to ensure feasibility of the constraints at the start of the optimization. As this starting point corresponds to a highly pessimistic power consumption, these power savings are likely

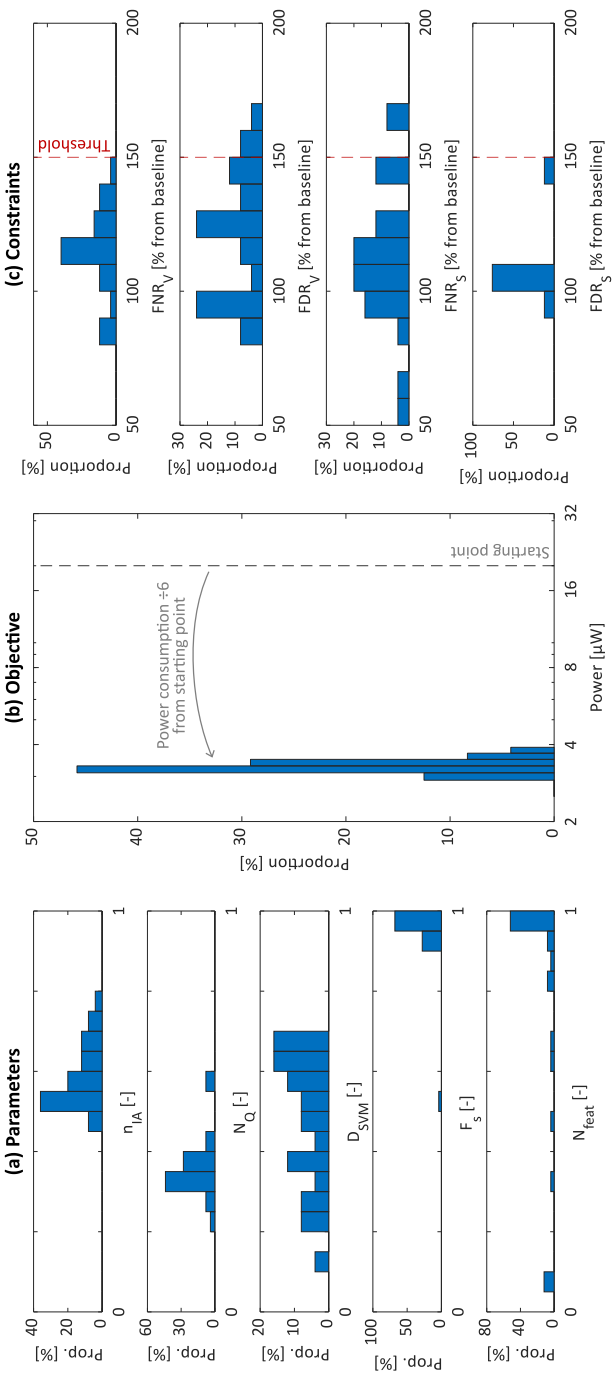

Fig. 4.9 System optimization results. The panels show the distributions of the final values for the normalized design parameters (a), power consumption (b), and error rate constraints (c). Only the runs that successfully converged are included in the figure.

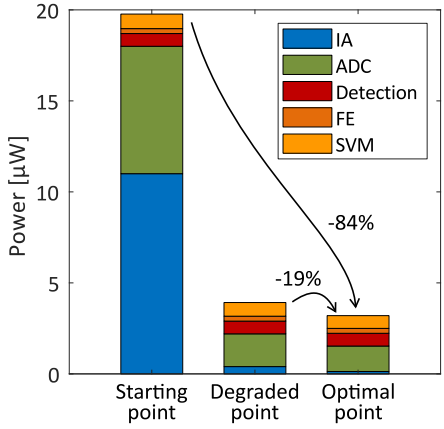

Fig. 4.10 Power optimization of respective blocks. The power distribution is shown before and after optimization. A slightly degraded point is added for a reasonable estimation of the power savings expected in a practical application.

overestimated compared to a practical application of the optimization. Another comparison is therefore provided with a point which has a slightly degraded power consumption, and a higher performance level (IA IRN $\times 2$, ADC resolution +1 bit and SVM pruning parameter +0.5).

In order to visually verify the optimality of the convergence, the parameters are swept independently around the final point in Fig. 4.11. The results show that for all parameters, the final point is close to the crossing point of the accuracy constraints, especially for the FNR of the S and V classes. While power consumption could still decrease aside those constraints, the optimization is therefore stopped by the active constraints. The noise on the functions evaluations close to the boundary explains the variability of the final points across multiple runs.

To show how the optimization algorithm reacts to changes in the operating conditions of the system, which impact the accuracy constraints, the amplitude of the input signal is modified before executing the algorithm. The optimization results displayed in Fig. 4.12 show that the optimization adapts by improving the AFE performance accordingly when the amplitude drops. In the DBE, no significant difference between runs can be observed.

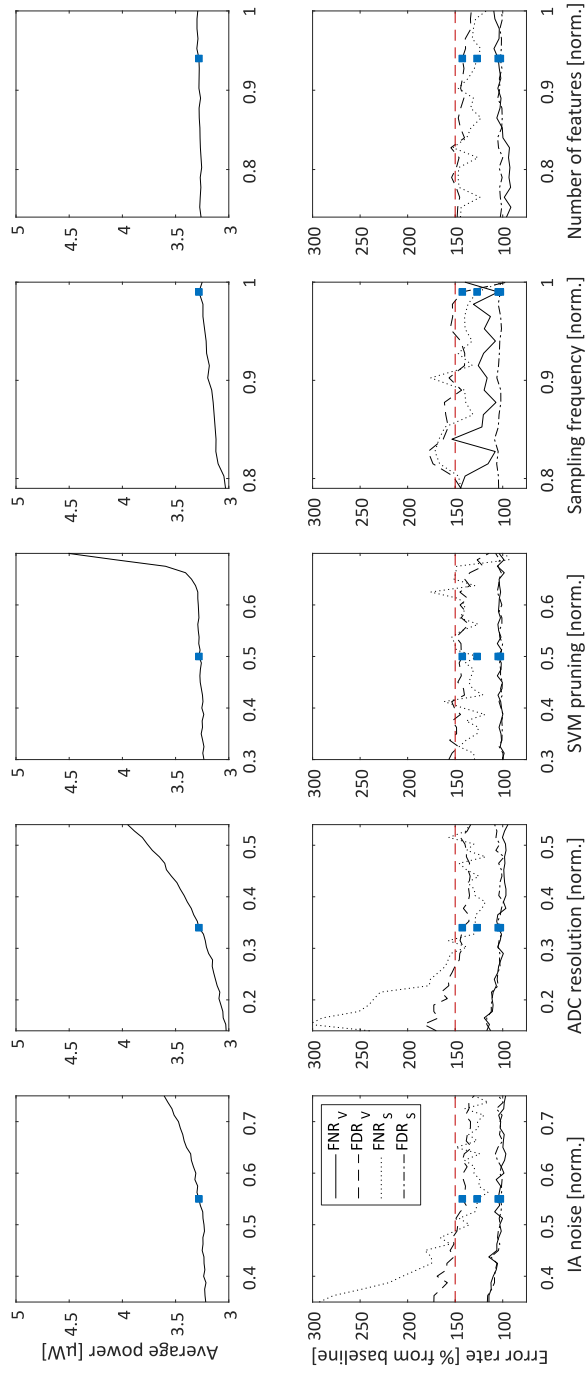

Fig. 4.11 Local parameter sweep. Parameters are swept around the median optimal point provided by the optimization runs: 5.2 µV RMS of IA noise, 9.4 bits of resolution, SVM pruning parameter of 0.75, 200-Hz sampling frequency and 10 features per classifier.

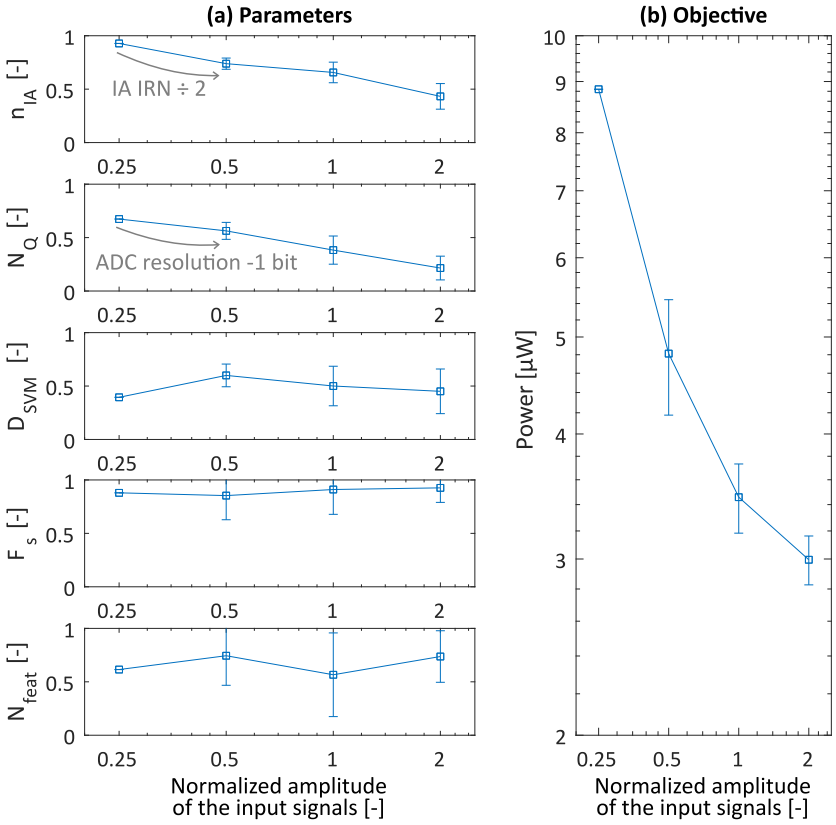

Fig. 4.12 Optimization results with signal amplitude sweep. Several gains are applied directly to the database signals and the optimization algorithm is executed. The markers show the mean of the parameters (a) and objective (b) across successful optimization runs. The error bars indicate the standard deviation.

In Section 4.2.1, the allowable error rate degradation α used to set the performance constraint threshold is set to 50%. The impact of this parameter on the convergence result depends on the gradient of the constraint functions close to the optimum. As observed in Fig. 4.5, the error rate increases rapidly close to the optimum, which means that the performance threshold has a limited impact. By changing the value of α from 50 to 100%, the achieved power is reduced by less than 0.1 µW. However, an excessively low value of α, e.g., below 25%, pushes the threshold limit too close to the simulation noise level and induces false local optima.

4.5 Discussion

While the results presented in the previous section show that the proposed methodology allows to find the global system optimal design, there are two main technical limitations encountered with this approach. The first challenge is the time taken for the optimization. As shown in Section 4.4.1, the evaluation of the objective and constraint functions are on average 7.8 minutes long. The numerical optimization requires on average 25 evaluations for the objective and constraints, and 10 for their gradient. While this number of evaluations remains relatively low compared to other optimization algorithms, a single run takes around 6.7 hours with the optimized gradient estimation.

The second challenge is the noise of the evaluation that limits the accuracy of the optimization result. The SLSQP requires the gradients of the objective and constraints in order to find the best search direction. As the gradient information is not directly available, it is estimated using finite differences. However, the noise of the function evaluations degrades the precision of this estimation. To avoid this issue, the size of the finite difference step can be increased to reduce the impact of the noise, at the cost of a lower resolution in the parameter space. Despite this, as the noise is explicitly present in the model, some local optima may appear and generate false positives in the convergence criteria. Reducing the tolerance of the stopping criterion or averaging the evaluation results on multiple runs may reduce these issues, but will increase the execution time. To ensure the validity of the result and avoid the need for multiple runs of the optimization process, averaging of the objective and constraint function values and gradients at the final point could be performed to check whether constraints are met and optimality is achieved.

The proposed methodology was applied in this chapter on the arrhythmia classification system, based on the implementation in Chapter 2. The optimization was performed by choosing only five free design parameters. While these parameters were chosen to maximize the tunability of the system power and performance trade-offs, there are many other parameters that come into the design of a complex sensor system. This limited number of free parameters narrows the range of possible optimization of the system. For instance, firstly, the minimum power consumption that can be reached by tuning these parameters is limited by the fixed power consumption sources, such as the AFE minimum current, the rest of the DBE

software, or other hardware blocks such as the CPMU. In the proposed model, the minimum power consumption that can be reached is 1.37 μW. Secondly, the fixed value of discarded design parameters may influence the optimal design point for the chosen free parameters. For example, the type of features that are chosen in the baseline design influences the number of features and support vectors. Additional design parameters can be added in the proposed methodology, but they require the implementation of their parametrization in the macromodel and possibly longer convergence time for the optimization algorithm.

The benefits brought by the proposed optimization methodology depends on the use case to which it is applied. As shown in the previous section, the power reduction obtained by the optimization is linked to the distance between the starting and final design points. If the starting point is already well optimized, the possible improvements are limited. Likewise, the final power consumption that can be obtained depends on the constraints and stimuli that are applied to the system. Finally, the impact of the free parameters on the objective and constraint functions also limit the optimization potential. In terms of partial derivatives, a good optimization parameter has a high derivative for the objective function in order to reduce it, but low derivatives for the constraints so that it is not limited too early. The largest power gains are therefore obtained when the starting design is under-optimized and there is still margin for optimizing the power consumption, and the free parameters have a large impact on the power consumption. Unfortunately, these characteristics can often only be assessed a posteriori. In the case of a starting point whose power consumption is already close to the optimum, the methodology could be used instead to optimize the accuracy of the system, with similar limitations. In the proposed system, the accuracy is limited by the intrinsic separability of the classes in the datasets, as observed in the state of the art for inter-patient arrhythmia classification.

The methodology presented in this PhD suffers from a few limitations. First, it is only applicable when there are two levels in the hierarchy. As the optimization method works by combining the trade-offs of lower-level circuits with the performance specifications of the level above, it cannot be applied across more hierarchy levels. If more levels are present in the design, additional steps are required so that the methodology can still be applied. For example, the performance trade-offs could be propagated from lower to higher levels of the hierarchy, for example using the MUBU approach.

Alternatively, the translation of top-level specifications can be used to infer requirements for the blocks, such as in the normal top-down approach, and thus apply the numerical optimization at a lower level. Despite this limitation, low-power sensor devices rarely contain many levels of hierarchy, and the presented architecture with a small AFE and machine-learning processing is a representative signal data path in such devices.

A second limitation is the lack of explicit considerations for reliability of the performance. Electronic devices can fail for many reasons such as manufacturing variability, operating conditions such as temperature or supply voltage drops, intrinsic noise, which all reduce the yield if the margins are not large enough. Circuit design therefore often takes those additional constraints into account by simulating several process, voltage and temperature corners, or by running Monte-Carlo simulations. To include these issues in the proposed methodology, the sensitivity of the Pareto-front to variability could be taken into account using dedicated methods [148]. Those robustness-aware methods however often increase the simulation time.

Finally, the SLSQP optimization method that has been chosen is only suitable for continuous parameters. In the target use case, the number of features however only takes discrete values. This limitation has been avoided by considering discrete parameters as continuous ones in the optimizer. While non-integers values are allowed from the optimizer perspective, they are rounded for the evaluation in the model. Additionally, the gradient finite difference estimation needs to be performed with differences corresponding to at least one discrete step. This alternative however only works when the number of steps for the discrete parameters is high enough and the discrete values are ordered, so that it can be approximated as continuous. In the case of unordered categorical parameters, which can occur for example when choosing between different circuit topologies, the optimization methodology can be applied to each individually and used to compare the optimal results.

4.6 Conclusion

In this chapter, a new optimization methodology for system design was proposed. It relies on a three-step approach: power and performance trade-off modeling at the block level, implementation of a cross-domain system-level performance evaluation framework, and numerical optimization of

Table 4.2 Power reduction with circuit and system design. The system optimized in this chapter is compared to the wireless transfer of the raw ECG signal, as discussed in Chapter 1, and the baseline design of Chapter 2.

	Transmission of raw signal	**Local data processing**	**System optimization**
Reference	Chap. 1	Chap. 2	Chap. 4
Technology	28nm FD-SOI	28nm FD-SOI	28nm FD-SOI
System optimization	✗	✗	✓
Embedded DBE	✗	✓	✓
TX data rate	3 kbps	< 4 bps	< 4 bps
Power [b]			
AFE	2.1 µW	2.1 µW	1.5 µW
DBE	-	2.1 µW	1.8 µW
TX [a]	12 µW	16 nW	16 nW
Total	*14.1 µW*	*4.2 µW*	*3.3 µW*
Silicon area			
AFE	0.02 mm^2	0.02 mm^2	0.02 mm^2
DBE	-	0.39 mm^2	0.39 mm^2
TX [a]	0.61 mm^2	0.61 mm^2	0.61 mm^2
Total	*0.63 mm^2*	*1.02 mm^2*	*1.02 mm^2*

[a] Assuming a Bluetooth Low-Energy TX front-end with 4 nJ/bit and 0.61 mm^2, similar to [15] in 22-nm CMOS.
[b] Sleep power and DC/DC conversion losses not considered.

the total power consumption. Compared to usual methodologies, it improves the optimality of the final system design, while limiting the simulation overhead of exploring the system parameter space which can become large for complex SoCs. Additionally, this methodology is also suitable for cross-domain optimization by combining in a single framework analog and digital circuits.

As a first step in the study of this proposed methodology, it has been applied to the arrhythmia classification use case presented in the previous chapters. In this context, the optimization results demonstrate an improvement up to a factor $\times 6$ compared to non-optimized designs. This gain however strongly depends on multiple factors such as the baseline design as well as the numerous degrees of freedom for the optimization.

Applied to the proposed optimization task, a more reasonable comparison point shows a 19% improvement for the optimized design. It is important to note that the ECG arrhythmia classification system had already been well optimized at the circuit level in Chapter 2, which therefore leaves less room for optimization at the system level. The results following circuit- and system-level design and optimization are shown in Table 4.2.

While this approach provides a solution to find the optimum system design, it requires significant work for the implementation of the block-level and system-level models. However, once the models of basic blocks have been generated, the methodology can be transferred to alternative use cases with a reduced implementation time. As a future step, the findings obtained on this first use case should therefore be extended to other systems with different specifications and balance between analog and digital domains. Many types of sensor systems could benefit from the presented approach, either for biomedical monitoring, or in other environmental or industrial contexts. In addition, the methodology could also be used to evaluate design choices for the system, for instance to compare the SVM classifier with different types of machine-learning algorithms such as neural networks or random forests. This allows a fair comparison for each algorithm, in terms of both their power consumption, classification accuracy, and robustness to degraded signal quality. Other future goals would be to mitigate the current technical limitations of simulation time and evaluation noise.

5

Environmental sustainability of mHealth

5.1 Mobile health

As described in Chapter 1, eHealth is an emerging technological ecosystem which aims at improving the quality and efficiency of healthcare thanks to information and communication technologies (ICT) [153]. In the same way that ICT has deeply impacted many other sectors (e.g., work, retail, communication), eHealth may represent a significant shift in paradigm for the healthcare sector in the near future [154].

eHealth includes a wide array of technological solutions as shown in Fig. 5.1. Within its subcategories, we focus on mobile health (mHealth), which refers to the “medical and public health practice supported by mobile devices" [2], sometimes overlapping with telemedicine. mHealth devices mainly include software health applications for smartphones or other portable devices, and custom hardware wearable devices that are used to monitor biophysical or biochemical parameters. Examples of mHealth technologies include smartphone apps for sleep or stress reporting, mobile telemedicine and health prevention services, fitness trackers or devices for remote monitoring of blood sugar levels, cardiac disorders or neurological conditions. The wearable devices presented in Chapter 1, and the use case

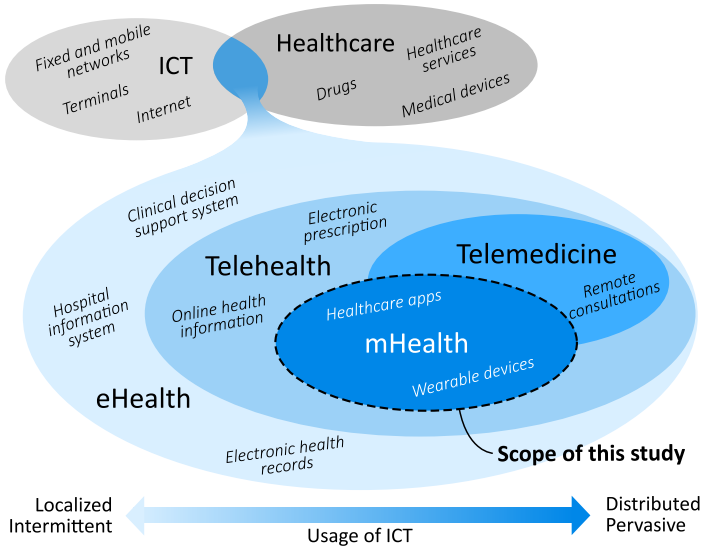

Fig. 5.1 eHealth technological ecosystem and subcategories. eHealth stands at the intersection of ICT and healthcare and may be subdivided in more specific concepts. Examples of technological applications appear in italic. This chapter focuses on the category of mobile health technologies.

of the cardiac arrhythmia classification system designed in Chapter 2, are examples of mHealth technologies. While this chapter focuses on mHealth and wearable devices, some of the discussions also include eHealth as they share many characteristics and impacts on the healthcare system. The classification illustrated in Fig. 5.1 also corresponds to an increasing level of pervasiveness of ICT devices, both temporally and spatially. Indeed, while some eHealth examples such as clinical decision support software or health digital databases remain constrained to the hospital computer infrastructure, mHealth and telemedicine rely on electronic devices being distributed in many households and sometimes used continuously. It is this property of ubiquity of mobile devices and remote communication that enables a significant shift in the healthcare paradigm [154]. Indeed, over the last decades, healthcare has been focused on curing diseases when they appeared, mainly thanks to doctors knowledge in an infrastructure centered around the hospital. As it is now possible to interact with the patient and collect data outside the usual medical settings, new possibilities are open for preventing diseases before their manifestation, relying

on data to provide new insights on medical disorders, in an infrastructure that is more widely distributed [155]. However, this ubiquity also introduces new challenges due to its vast deployment and its consequences on the healthcare system.

mHealth is expected to play different roles in healthcare management: (1) prevention, by promoting healthy lifestyles; (2) monitoring, by enabling early screening and frequent health assessment; (3) treatment, with a personalized and more effective care; and (4) support, by ensuring proper information and communication for the patient and caregiver [156]. Until now, mHealth and eHealth have been mostly considered as a solution to compensate for the limitations of the healthcare sector. For example, they have been presented as an answer to issues such as the lack of access to health services in low-income countries [2, 157, 158], the shrinking workforce in the health sector [159], the ageing population in the Western countries [160], or the increasing weight of chronic diseases [161]. They are also expected to improve cost efficiency in healthcare and clinical efficacy, though this is still under debate [162, 163]. More generally, mHealth is often considered as a tool for carving the path towards a more sustainable healthcare system. Sustainability is usually divided in three interconnected pillars which are environmental, social and economic sustainability, each featuring multiple dimensions [164]. The economic pillar looks at the creation of economic value, financial viability, proper governance, risk management and prosperity. The social pillar includes the promotion of equity, the respect of human rights, or the enhancement of the well-being of stakeholders. The environmental sustainability is evaluated with regard to several aspects, e.g., the emission of greenhouse gases (GHG) and climate change, the protection of biodiversity and ecosystems, the generation of waste or the depletion of natural resources. Despite the growing body of research on the question, it is still unclear whether mHealth actually enhances these different dimensions of sustainability for the healthcare sector [JP3].

This chapter focuses on the environmental aspects of sustainability and on how mHealth technologies can contribute to reducing the ecological impacts of the healthcare sector. In general, a technology has direct effects on the environment which are due to its life cycle impacts, that is, for its pro-

duction, during its use phase or at its end of life. It can also induce indirect effects, i.e., increase or decrease the impacts of other goods or services due to a modification of behaviors, organizations or processes. With eHealth and mHealth, the switch from physical to ICT-supported services has significant implications for the environment, both direct and indirect. For instance, while often referred to as virtualization, this change of medium is actually supported by a large distributed infrastructure. This is particularly true for mHealth which relies on the massive deployment of portable and wearable devices, as well as the underlying Internet network and datacenters, which all have direct environmental impacts. Through the digitalization process, eHealth and mHealth also have indirect impacts, such as the reduction of travel thanks to telemedicine or the reduction of the paper consumption with electronic records. Overall, these considerations have been mostly optimistic, under the commonly shared assumption that digital solutions provide opportunities to reduce the environmental impacts [54, 57]. However, the evidence on the actual impacts of these technologies remains limited and the balance between positive and negative impacts is still hard to evaluate [165].

By extending the work presented in [JP3] on the ICT aspects and the environmental sustainability, this chapter aims at reviewing and commenting on the current state of environmental considerations around mHealth solutions in the scientific literature. This review shows that several critical characteristics of mHealth technologies are poorly taken into account, especially regarding their indirect impacts. The LES model proposed by Hilty and Aebischer [51] is therefore used as a framework to address these limitations and discuss the larger systemic impacts of mHealth technologies. This chapter is organized as follows. First, in Section 5.2, a systematic literature review evaluates which aspects of the environmental sustainability of mHealth and medical devices have been the most explored by the scientific community. Section 5.3 discusses the shortcomings of these studies with a focus on the specificities of mHealth technologies compared to traditional medical devices. In particular, the lack of an approach that takes into account their systemic impacts is highlighted. Finally, Section 5.4 leverages the LES framework originating from the field of ICT to highlight and sort the potential systemic impacts which are currently under-evaluated in the literature. Examples of presumptive effects of mHealth drawn from the literature are used to illustrate the proposed framework.

5.2 Literature review

In this section, we highlight how environmental impacts have been considered for mHealth technologies in the scientific literature. Evaluating the environmental sustainability of mHealth technologies is a challenging task that involves direct and indirect levels of impacts, different scopes (components, devices and systems) and many indicators of environmental harm. The analysis of the selected articles aims at understanding from which perspectives sustainability has been looked at and what aspects have been most studied.

5.2.1 Methodology

In order to provide a detailed account on how environmental sustainability is considered in the scientific literature, we conducted a systematic literature review. The PRISMA methodology [166] was used to sort and report the results, as detailed in Section 5.2.2. The selection of relevant abstracts and articles was performed by the author. Three databases were searched to extract relevant studies: Scopus (multidisciplinary), PubMed (biomedical and life sciences) and IEEE Xplore (engineering and technology). The searched phrase was built as a combination of a keyword linked to mobile health and another linked to environmental sustainability:

$$(<\textit{mHealth \#1}> \text{OR} <\textit{mHealth \#2}> \text{OR} \ldots) \text{ AND } (<\textit{Environment \#1}> \text{OR } \ldots) \quad (5.1)$$

Due to the limited number of papers focusing specifically on mHealth, the search was expanded to the related fields of medical devices and telemedicine. The complete list of keywords used is as follows:

- Mobile health
  - eHealth
  - mHealth / m-health / mobile health
  - wearable device / wearable electronics / wearables
  - medical device
  - telemedicine / telehealth
- Environmental sustainability
  - environmental sustainability
  - environmental impact / ecological impact
  - ecological footprint

- life-cycle analysis / life-cycle assessment / LCA
- eco design / ecodesign
- carbon footprint / GHG / greenhouse gas
- climate change

Section 5.2.3 provides an overall description of the articles selected in this review. Given the wide variety of topics covered in the articles, their presentation is structured in several clusters, which were defined in a bottom-up fashion based on the similarity of the presented topics and research approaches. The description of the articles focuses on how environmental sustainability is considered in the articles and which research approach is used to address it. The exhaustive list of the selected papers and associated topic clusters are listed in Appendix A.

A content analysis was then performed in Section 5.2.4 to assess quantitatively which themes are the most present in the articles. This analysis consisted in counting how many articles mentioned given themes. The targeted topics for this analysis are the scope of the environmental study, the level of impacts considered and the environmental indicators. The scope indicates which parts of the healthcare system is evaluated, i.e., either the medical device, its internal components or the broader system. The level of impacts relates to the direct life cycle impacts or indirect induced effects of the technology. The environmental indicators list the types of impacts which are considered, e.g., carbon footprint, ecosystem pollution, or resource depletion. The papers included in the analysis were each reviewed individually and linked to the themes that they mention. To avoid any subjective appreciation when assessing the significance of a theme mention, all theme references are counted as a single hit per article, regardless of their importance in the text.

5.2.2 Search results

The database search resulted in a total of 537 unique articles. The titles and abstracts were screened to exclude those that do not fit the research question. More specifically, the articles whose scope do not include healthcare systems, medical or ICT devices, or environmental sustainability were rejected. After retrieval of the full texts, a second assessment for relevance was made based on the same criteria, which returned 59 results. In addition, 4 documents from the grey literature that were not in the searched databases were added: the NHS zero-carbon strategy [167], the *Plan de transformation de l'économie française* report from the Shift Project [165], a

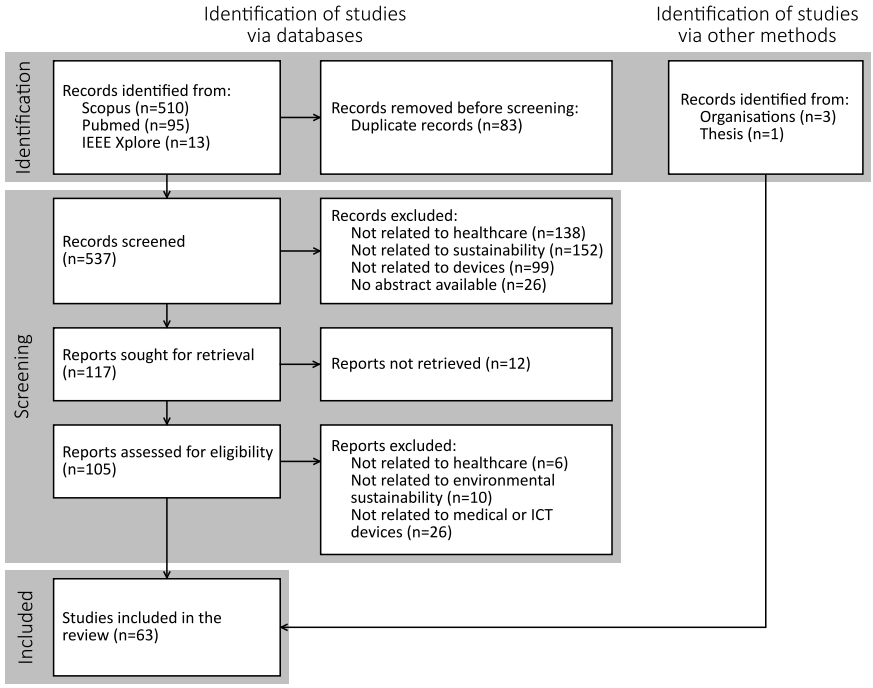

Fig. 5.2 Identification of relevant studies with the PRISMA methodology [166]. The articles identified in the database search were iteratively sorted according to the exclusion criteria to provide the final list of studies for the following analyses (cfr. Appendix A). The number of papers excluded and related criteria are specified in the chart.

Master thesis on wearable cardiac monitoring devices [28] and a publication from the H2020 UBORA project [168]. A total of 63 studies are therefore included in the following analyses, as reported in Fig. 5.2.

5.2.3 Structured presentation of the selected articles

In this section, the main ideas presented in the selected articles are described. Given the large variety of subjects and research domains present in the list of articles, this description focuses on discussing how environmental impacts are considered and from which perspective they are being addressed. The articles are structured in different clusters which correspond to the main topics and research approaches, that were built in a bottom-up fashion during the review of the articles. The number of clusters and

their boundaries were decided arbitrarily to balance the level of details and the ease of understanding. The main clusters are the environmental assessment of medical devices, the design of new components for wearable devices, the reduction of travel-related emissions with telemedicine, methodologies for environmentally-conscious design, sustainable business models, top-down healthcare system assessment and environmental considerations of stakeholders. When necessary, subcategories in the clusters are also presented to highlight specific differences. The exhaustive list of articles and their classification of articles in the different clusters can be found in Table A.1 in Appendix A.

Medical device assessment

The most represented topic in the selected articles (n=18, 29%) is the assessment of impacts of medical devices. Those articles generally aim at understanding the absolute or relative impacts of medical devices. The precision of the assessment ranges from a general analysis of the environmental impacts and the associated design challenges [169], to a more specific and quantitative impact evaluation for a given device [170, 171]. The reference method for the quantification of environmental impacts of a device is the life-cycle assessment (LCA), defined in the ISO14044 standard [172], which relies on a precise material inventory and information from databases to evaluate several impact categories (e.g., global warming, ozone depletion, ecotoxicity, human toxicity). This analysis is however tedious to apply and suffers from a lack of data on environmental impacts and significant variability in the results. It is therefore still rarely applied and only a handful of examples exist in the literature. The types of devices assessed in those studies vary largely, but focus mainly on devices used in the hospitals, such as dialysis equipment [173] or catheters [171]. A more general review of LCA and eco-design studies for medical devices is provided by Sousa *et al.* [174]. They conclude that despite the growing importance of sustainability in the medical device industry, the number of studies is still quite limited.

Within the assessment of medical devices, some categories are more represented. For instance, single-use devices, or their reusable alternative, have received more attention. Some examples include laryngoscopes [169, 175, 176], dialyzer cartridges [177], dental burs [170], biopolymer-based devices [178], catheters [179], or other surgery equipment [180]. The interest for reusable alternatives can be explained by the large quantities of medical waste generated annually by the medical sector and the associated

concerns with its sustainability.

Wearable devices for mobile health are however assessed in only 6 of the selected studies. This number is quite low given that mHealth is the focus of this study. The subjects of these studies include a wearable computer [181], a cardiac monitoring device [28], a smartwatch [59], and e-textiles [60, 182, 183]. Three of these studies include precise evaluation of the direct impacts with LCA: Ma *et al.* [59] on a smartwatch, Kokare [28] on rigid and flexible cardiac monitoring devices, and van der Velden *et al.* [183] on a garment with embedded electronics. Lankey *et al.* [181] propose qualitative guidelines and a list of criteria for the design of a wearable computer, i.e., equivalent to a personal digital assistant. Gurova *et al.* [60] discuss the general impacts of e-textiles by merging insights from the fashion and ICT industries. Finally, Köhler *et al.* [182] focus on the issues that arise with the end of life and disposal of e-textiles, due to the combination of electronic and textile waste streams.

Component design

Among selected articles, another significant group focused on designing subparts of the device (n=16, 25%), e.g., power supply, flexible substrate or electronic components, with the goal of improving their sustainability thanks to new materials or designs. Even though they are presented in the articles as components of wearable devices, they can also belong to other types of electronic devices.

In the selected studies, 6 aim at providing a sustainable source of energy thanks to the use of energy harvesting. In the case of wearable devices, several sources of energy are available for supplying the device, if a suitable conversion circuit is present. In most studies, energy is harvested from the human body, either in the form of mechanical energy from motion [184, 185, 186] or chemical energy from fluids [187, 188]. In the work of Mukherjee and Dahiya [189], the energy is harvested from the ambient light using photovoltaic cells. In most cases, the amount of energy that can be harvested often remains limited and requires strong reduction in the device power consumption.

Other studies have proposed new materials or fabrication techniques for the design of the components, usually to reduce the end-of-life impacts, or to enable sustainable sourcing and avoid resource depletion. Some of these solutions rely on bio-inspired materials such as silk for piezoelectric sensors or actuators [190, 191], protein-nanowire humidity sensors [192], flex-

ible substrates made of biopolymers [193], or DNA-based memories [194]. Other synthetic materials have been studied for their biodegradability [195, 196], Earth abundance [197], or flexibility [196, 198]. Finally, Rahman *et al.* [199] evaluated additive manufacturing to design touch sensors in order to reduce waste in production.

Travel reduction with telemedicine

As mentioned in the methodology in Section 5.2.1, the search includes articles on telemedicine, whose applications are often related to mHealth. Due to the exclusion criteria, only those that discussed the equipment used were selected, which resulted in a list of 10 articles (n=10, 16%). Within those articles, the modalities for remote consultation varied: videoconferencing [200, 201, 202, 203], telephone [204], or either of them [58, 205, 206, 207, 208]. From an environmental perspective, the focus of all these papers stands on the indirect reduction of the carbon footprint thanks to the reduction of travel. Regardless of the modalities used, all papers concluded that the savings from travel reduction largely exceeded the direct impacts of the infrastructure. The amplitude of the savings depends however significantly on several factors such as the type of consultations (general or specialized medicine), the country, the settings (urban or rural), or the modes of transportation.

Environmentally-conscious design methodologies

In-depth environmental assessment is time-consuming and data-intensive. It is thus impractical to iteratively perform such evaluation during the design of a medical device. To bypass this issue, some authors have worked on tools to guide the design process and take environmental considerations into account at each step (n=3, 5%). The information provided by those design tools is often less comprehensive than LCA, but they can be directly applied by the designer.

Alternative assessment methods have been proposed to replace LCA during the design for faster iteration. Moultrie *et al.* [209] developed an evaluation grid specifically tailored for medical devices, in which several indicators are qualitatively evaluated to identify the main bottlenecks in the design. Another common tool is the checklist, as discussed by Unger *et al.* [210], that allows to quickly verify that some environmental criteria are met in hospitals, such as encouraging recycling programs or sustainable device sourcing.

One of the main difficulties in the design of medical devices is the large number of constraints that apply such as regulations, safety, medical efficacy or cost. Mak and Kritchanchai [211] investigated those different constraints, ranked them according to their importance and provided design recommendations. Hede *et al.* [212] proposed to incorporate sustainability in a decision-making framework, in order to help prioritize sustainability criteria with other constraints.

Business models

Maintenance, repair, recycling or reconditioning are practices that enhance the sustainability of devices by prolonging their life cycle or giving them a second life. Those are however not straightforward to apply in a cost-conscious healthcare industry, in addition to the conflicting constraints of safety and medical efficacy. Circular business models that incentivize sustainable practices have been studied and adapted for this particular sector. While business models impact primarily the economic aspects of sustainability, they have consequences for the environmental dimension as well, which is why they are included in this review (n=5, 8%).

As the medical device ecosystem is quite diverse and different devices often have distinct characteristics and specifications, several types of alternative business models exist to fit each use case. Guzzo *et al.* [213] reviewed different types of circular business models, such as rental, end-of-life equipment collection, reprocessed devices or equipment as a service, and mapped them to different device characteristics. Other authors studied how some models apply to specific healthcare use cases and how they improve sustainability. For example, product-as-a-service systems [214, 215], in which the client pays for the use of a product rather than its property, usually encourage the provider to maximize the life cycle of the device by improving maintenance and reconditioning. Zlamparet *et al.* [216] studied the technological and economic advantages of recycling and remanufacturing imaging equipment and servers. Finally, the UBORA project [168] evaluated the potential for open-source devices in the medical industry, as well as their impacts on environmental sustainability.

Top-down healthcare system assessment

While most works look at sustainability in a bottom-up approach by starting from selected use cases, five studies included in the analysis proceed in a top-down approach by taking a macroscopic view at healthcare systems (n=5, 8%). The aggregate results that are provided are usually coarser,

but they allow to understand the global picture of environmental impacts in healthcare and come up with practical guidelines to mitigate those impacts. Among the most urgent objectives regarding environmental sustainability, the reduction of GHG emissions is necessary in the healthcare sectors to comply with the Paris agreement which aims at limiting global warming well below 2 °C above pre-industrial levels [217]. Sector-wide GHG emission reduction strategies have therefore been proposed for the healthcare system. Sherman *et al.* [218] reviewed the healthcare emissions research and summarized the main emission reduction strategies and research gaps. Plans at the national level have also been proposed by organizations in different countries. In the selected articles, two of those plans are included, for England [167] and France [165], respectively. These plans consist in an overview of the current annual emissions divided by sources as well as propositions, usually at the political and regulatory level, to reduce those emissions. The interest of these documents in this review is twofold. First, the potential savings enabled by telemedicine and mHealth for different emission sources can be compared with their respective share in the overview of the total emissions, illustrated in Fig. 5.3. Second, digitalization, which usually includes telemedicine and mHealth, is recognized as an important driver for sustainability, e.g., by using virtual procedures instead of those currently taking place in physical settings, such as remote consultations or electronic records. In these documents, the importance of taking into account the footprint due to the direct impacts of this increasing digital infrastructure is also highlighted.

In two studies, the authors specifically tackled the evaluation of the environmental impacts of eHealth, including the impacts of several applications. Holmner *et al.* [54] evaluated qualitatively the potential environmental benefits of several applications that are part of a larger eHealth system, such as remote consultations, remote health education, electronic prescriptions or remote diagnosis. Di Giacomo and Håkansson [57] quantitatively assessed the GHG reduction following the implementation of an eHealth network for electronic prescription and referral. Both studies mentioned the reduction in travel as the main environmental benefit from eHealth, compared to paperwork reduction for instance, while direct impacts from ICT infrastructure were negligible in the studied applications.

Stakeholders environmental considerations

In order to understand how environmentally-friendly practices are considered in the healthcare sector, some authors performed interviews directly

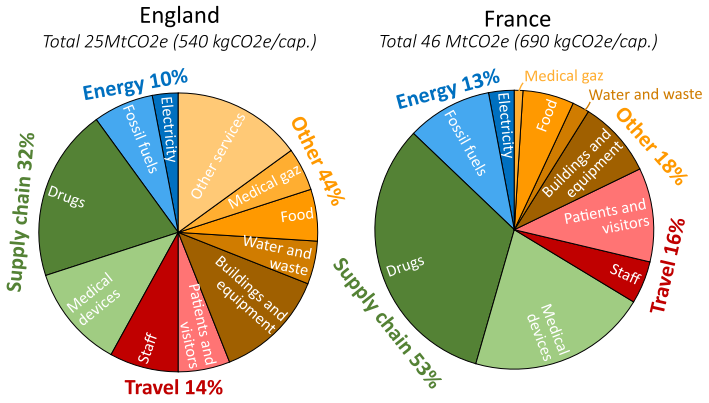

Fig. 5.3 National carbon footprints for the healthcare sector (adapted from [165] and [167], with supplementary data from [219]). The main emission source categories are represented with similar colors, with shades to distinguish subcategories. The scopes for the emission sources each chart may not be exactly similar and they cannot be compared directly.

with stakeholders (n=2, 3%). Moultrie *et al.* [220] interviewed designers from medical device manufacturers to assess the extent of environmental measures in their design work, and the drivers and barriers to their implementation. Ordway *et al.* [221] investigated hospital practices to reduce waste by interviewing hospital staff. While stakeholders are generally willing to improve the environmental impact of the sector, they face significant constraints to do so. From a design point of view, the medical industry is seen as risk-averse, cost-conscious and subject to a lot of stringent regulations. The environmental issues are therefore much lower in the list of priorities. In both studies, the lack of information or education was also designated as a barrier to the implementation of more sustainable practices.

Other

A few studies (n=3, 5%) appeared as outliers in the pool of selected papers in the sense that their scope was not deemed similar to other studies. For instance, Cosgrove *et al.* [222] proposed a framework to evaluate the energy reduction in medical devices production facilities. Alshqaqeeq *et al.* [223] performed a literature review of the carbon footprint assessment of patient care alternatives in hospital services. Finally, McGain *et al.* [224] discussed qualitatively the environmental impacts of a catherization laboratory and

an operating room, with a focus on the opportunities and barriers for improvement.

5.2.4 Quantitative content analysis

In order to get a clearer understanding of what aspects of sustainability are mostly considered in the literature, and hence also identify those that are under-studied, a quantitative content analysis is performed. This analysis aims at classifying the selected articles in three dimensions: the scope, the level of impacts and the environmental indicators. Along each dimension, categories have been created and associated with the articles, with possibly one additional level of hierarchy to highlight important subcategories. For a given dimension, each study is linked to at least one category, and possibly more than one if several corresponding themes are discussed in the article. The extracted themes and corresponding occurrences are reported respectively in Figs. 5.4, 5.5 and 5.6. The exhaustive association of themes and articles is provided in Table A.1 in Appendix A.

Scope of the environmental considerations

The scope of the environmental considerations corresponds to the part of the healthcare system whose environmental impacts are discussed, and has been divided in three categories. Firstly, the *device* category corresponds to the studies that look at the impacts of whole individual medical devices, such as a laryngoscope or a smartwatch. Secondly, the *system* category corresponds to the aggregated impacts of several devices, usually in the context of an application. Thirdly, the *component* category includes studies that focus on subparts of a device, such as the sensor or power supply.

It can be seen in Fig. 5.4 that most of the selected articles belong to the *device* category, which is the focus of the review. However, only 6 of these correspond to wearable devices and, when specified, a larger proportion focuses on hospital equipment and single-use devices for medical intervention. The *system* category mainly consists in articles on telemedicine. A few papers discussed the global impacts of eHealth, but none focused on the impacts of mHealth as a system. In the *component* category, the energy harvesting circuits, sensors and flexible substrate are the most represented.

Level of environmental impacts

The level of environmental impacts is here divided in two categories. The *direct* impacts are related to the life cycle of the devices, which includes

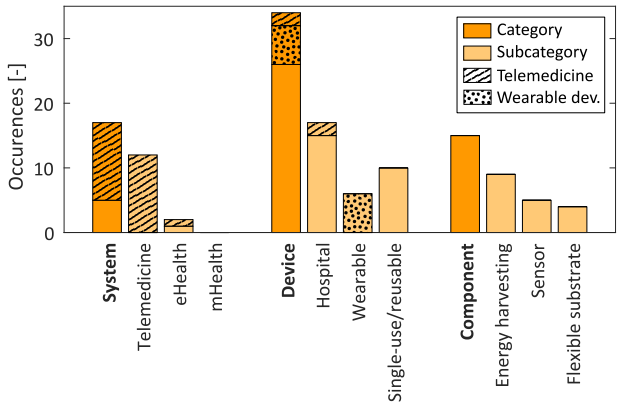

Fig. 5.4 Occurences of scopes in selected articles. The scopes correspond to healthcare system components whose environmental impacts are considered. Relevant subcategories are displayed to highlight significant groups. Studies on telemedicine and wearable device are highlighted.

the impacts of their production (material extraction, manufacturing, transport), use phase (energy consumption, maintenance) and end of life (disposal, recycling, reconditioning). The *indirect* impacts are the changes in other impact sources enabled by the use of the devices. In the selected articles, those consist mainly in positive impacts such as the reduction of emissions from transportation and optimization of healthcare procedures.

Fig. 5.5 shows that a majority of the papers analyze the *direct* impacts of the medical devices. All life cycle stages are discussed, with a slightly higher proportion for the end of life stage. The articles associated with the *indirect* impacts are predominantly focused on telemedicine and the induced reduction in travel. Two studies [204, 205] also discuss the optimization of healthcare procedures thanks to remote triage of the patients. Finally, four articles briefly discussed potential negative indirect impacts due to obsolescence [60, 165, 182] or rebound effects [58, 165].

Environmental indicators

The environmental impacts of healthcare are multifold and require several indicators to be fully evaluated. This is reflected by the following categories of indicators. *Waste* relates to the amount of discarded devices at their end of life and includes the alternative ways to reduce its impact such

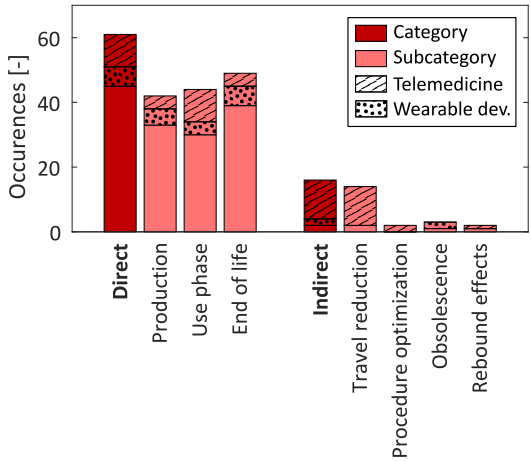

Fig. 5.5 Occurences of impact levels in selected articles. The impact levels indicate in what ways the devices generate effects on the environment, either directly through their life cycle or indirectly through their use. Relevant subcategories are displayed to highlight significant groups. Studies on telemedicine and wearable device are highlighted.

as recycling, refurbishing, reusing or using biodegradable materials. The *pollution* can be of different forms depending on the types of emissions and impacts on the humans or ecosystem. The *energy* category includes the energy consumption at all stages of the life cycle, as well as the attempts to reduce it. The *carbon footprint* corresponds to all emissions of GHG that contribute to climate change. Finally, the depletion of non-renewable resources and the usage of water are included in the *resources* category.

As shown in Fig. 5.6, the waste category is the most represented one with 71% of the selected articles mentioning it. The generation of waste has indeed been reported as one of the main concerns in the medical device industry. It is also likely the most visible one for the different stakeholders compared to other categories of impacts that are more difficult to visualize. The other categories are all significantly present in the studies, with between 44 and 60% of the articles. The carbon footprint is slightly overrepresented in the studies on telemedicine that evaluate the emission savings due to transportation. Within each category, different qualitative or quantitative evaluation methods may be used for a given indicator.

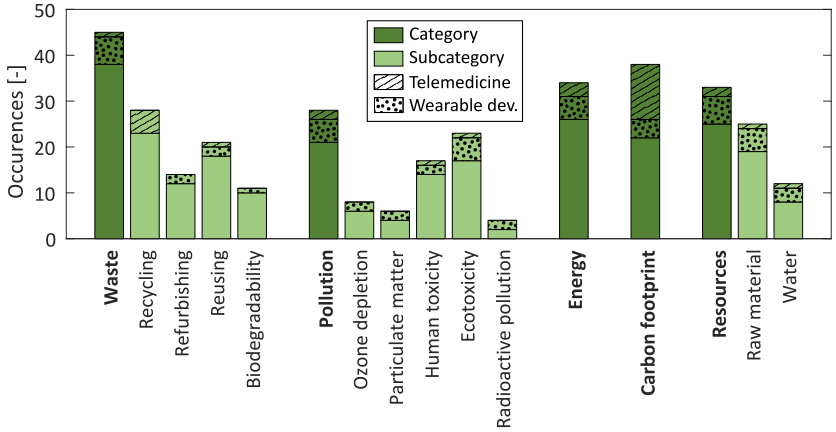

Fig. 5.6 Occurences of environmental indicators in selected articles. The environmental indicators correspond to the aspects of the environment that are impacted and how they are quantified. Relevant subcategories are displayed to highlight significant groups. Studies on telemedicine and wearable device are highlighted.

5.2.5 Limitations

The systematic literature review suffers from a few limitations. First, the review of the abstracts and articles could only be performed by a single reader due to practical reasons. Additional reviewers would likely reduce the risk of subjective bias in the assessment of the selection criteria. Second, in the selection and analysis of the articles, all studies were considered equal and included in the analysis, regardless of their scientific quality and the amount of evidence presented. Because of the large variety of research areas present in the articles, an evaluation of research quality would have required expertise in all those different domains. Finally, it is possible that some studies were missed due the limited selection of the search keywords. Indeed, as the research question was purposefully broad to include as many view points as possible, it is difficult to frame the issue with a reasonable number of key-words. For example, as only a single generic term "medical device" to complement the search on mHealth, other studies using more specific terms may have been missed. However, the selection of the articles returned 63 individual items, which already provides an extensive and representative account of the literature.

Regarding the content analysis, a first limitation is the choice and definition of the classification dimensions (scopes, levels and indicators) which were chosen arbitrarily and thus subject to bias. Secondly, as mentioned in the methodology, all mentions of the themes were counted equally, with exactly zero or one hit per article. While this avoids a subjective assessment of the importance of each theme, it represents less accurately the weight of a given theme in the literature. Indeed, this counting method does not take into account whether a theme is the focus of the study or only a small part of the discussion. Finally, similarly to the article selection, the content analysis could only be performed by one researcher, which increases the risk of bias when associating words in the texts to themes.

5.3 Shortcomings of mHealth environmental impact assessment

Based on the description of the articles in Section 5.2.3 and the observations from the content analysis in Section 5.2.4, some conclusions can be made with regards to the current considerations for environmental sustainability of mHealth. Section 5.2.3 showed that environmental sustainability for mHealth, and more largely telemedicine and medical devices, has been looked at from many different perspectives. While the multiplicity of research approaches is beneficial with regards to the complexity of the question, all these perspectives often seem rather disconnected and do not provide a consistent understanding of the impacts of these new technologies. In Section 5.2.4, we discussed some specific aspects of environmental impacts assessment, i.e., the scopes, levels and indicators, and highlighted the distribution of studies along these axes. This division helps emphasize that some characteristics of environmental impacts have been more studied than others.

The research on sustainability in the healthcare sector is still relatively new and it is therefore normal that significant research questions remain unanswered. However, this lack of research is even more pronounced for mHealth, which can be seen as an issue for two reasons. Firstly, regarding the devices themselves, wearable devices differ in many respects from traditional medical devices. Indeed, their composition relies primarily on electronics whose production and end of life have significant impacts. For instance, the LCA of an e-textile by van der Velden *et al.* [183] showed that electronic components are responsible for 71% of the device production im-

pacts. Due to their portability requirements, they are also subject to additional constraints such as their power consumption or size, which requires more tightly integrated components. Compared to traditional medical devices that are mostly prescribed or used by doctors, there are many devices associated to mHealth, such as fitness trackers, smartwatches, or simple monitoring devices for sleep or stress, that are currently mostly bought on the patient's own initiative. Finally, they operate outside usual medical settings and procedures, which means that they are not subject to the same medical and regulatory oversight. Similarly to other electronic devices, the wide variety of mHealth devices should be reflected in the evaluations, as their respective bottlenecks may differ [27]. Secondly, regarding the mHealth system, the current drive towards a more sustainable healthcare system is expected to rely partly on eHealth. When discussing the environmental impacts at the system level, most authors focus on the positive effects and often consider all new technologies indiscriminately, without discerning the different devices, applications and contexts. As mentioned in the national strategic plans [165, 167], the sustainability potential of digitalization can only occur if it is explicitly taken into account to avoid adverse effects. As these technologies are only currently emerging, the opportunity can be seized to implement sustainable practices and regulations from the start, but first require understanding the use case in which they can actually bring benefits.

Within direct impacts, the previous analyses showed that sustainability in the healthcare sector is largely centered around the issues of waste and end of life. This may be due to the frequent use of single-use devices and plastics to avoid contaminations, and the higher visibility of waste for hospital staff compared to other indicators. However, the relative importance of each life cycle stage in terms of impacts can significantly vary from one device to another. For instance, due to their technological complexity, the production of electronics often accounts for a significant part of the total impacts in sensor devices [27, 29]. The evaluation of mHealth solutions therefore needs to take into account all life cycle stages, looking at the various indicators of environmental harm. Additionally, direct impacts due to the devices life cycle are not the only effects that a new technological solution can have on the environment. Indirect effects induced by the use of a technology can, in some cases, be larger than the direct impacts. In the literature, the discussion around these indirect impacts are largely dominated by the travel reduction obtained thanks to telemedicine. They are therefore

mostly focused on positive impacts and revolve around the issues of carbon footprint. However, indirect impacts can also be negative, e.g., when they induce an increase in the demand for another resource, and involve different indicators than GHG emissions. In our review, only four article included negative indirect impacts. Three mentioned the increased obsolescence for e-textiles [60, 182] or digital devices [165]. Purohit *et al.* and the Shift Project report also succinctly discussed the possibility of rebound effects [58, 165].

Due to the shift in paradigm enabled by these technologies, eHealth and mHealth can induce deep transformations in the healthcare system. In order to anticipate the sustainability aspects of these emerging technologies, their effects at the macroscopic systemic level must be taken into account. An exhaustive sustainability evaluation should therefore consider the direct life-cycle impacts of a technology and the positive and negative effects induced by the use of the technology, but also the broader structural changes to the healthcare infrastructure, usage, institutions and economy.

5.4 Systemic approach to mHealth sustainability

As discussed in Section 5.3, when considering the long-term global impacts of a new technology, it is important to consider both its direct and indirect impacts, whose relative proportion varies from one application to another. Indirect impacts are especially important for ICT-based solutions which have the potential to indirectly influence the consumption of other related resources through, e.g., behavior modification, process optimization or virtualization. A commonly cited example is the reduction in the amount of resources for paper consumption thanks to the use of electronic communication and archiving. The nature of these indirect impacts varies in their direction, either improving or degrading the overall environmental footprint, and their action scale, at the level of an individual or society. Hilty and Aebischer [51] have proposed the LES model to conceptualize these impacts along multiple levels. In their model, the *Life-cycle (LC)* impacts relate to the direct impacts that occur during the production, use or disposal of ICT devices, similarly to the definition used in the previous sections. Indirect impacts are divided in two separate levels. The *Enabling (E)* impacts are the changes induced by the usage of the devices at the micro level, with technological, behavioral or organizational modifications. These modifications may have a positive impact, through substitution (ICT

replaces another resource) or optimization (ICT enables the reduction of another related resource), or a negative impact, through induction (ICT stimulates the use of another resource) or obsolescence (ICT reduces the lifespan of another resource). For mHealth, this relates for example to the reduction of patient travel, the optimization of health procedures, or the increased healthcare demand due to medicalization. The *Structural (S)* impacts correspond to macroscopic changes in the economic structures and institutions, enabled by the ICT devices. Long-term modifications in the healthcare infrastructure or in the role of doctors are examples of these dynamics due to mHealth. Like enabling impacts, structural impacts can also go in either positive or negative directions regarding the overall footprint evolution. The different impact categories of the LES model are usually conveniently displayed in a grid as the one in Table 5.1.

Building on the limitations drawn from the literature review presented in Section 5.3, this section relies on the LES framework to evaluate more extensively the impacts of mHealth. To do so, the different levels of the framework are illustrated with presumptive examples of positive and negative effects that mHealth may have on the environment, all listed in Table 5.1. For a few of the proposed examples (in blue in the table), the link between mHealth systems and environmental impacts has already been explicitly mentioned in the literature. For the others (in black), the link from mHealth to environmental impacts has not been discussed explicitly and is therefore extrapolated in this work from the wider literature on mHealth. The intentionally normative distinction between positive and negative impacts present in the original framework has been kept as it highlights the balance between the optimistic expectations and possible downsides. All items in Table 5.1 are detailed in the rest of this section. While the discussion aims at providing a large overview of the possible impacts, it cannot be completely exhaustive due to the wide range of possible impacts, especially as we progress towards complex system-level consequences of mHealth technologies.

5.4.1 Life-cycle impacts

As observed in the literature review, *Life-cycle* impacts, i.e., direct impacts, are by far the main point of attention for medical devices. However, the evaluations that are provided often remain only qualitative or incomplete. For a thorough evaluation of the direct impacts of a technology, life-cycle assessment (LCA) studies, as described in the ISO14040 standard, provide

Table 5.1 Life cycle-Enabling-Structural impact model applied to mHealth, based on the model by Hilty and Aebischer [51]. Presumptive examples of environmental effects from mHealth illustrate the categories. For examples highlighted in blue, the link between mHealth and environmental impacts has been explicitly mentioned in the literature, while it is extrapolated for examples in black.

	Positive	Negative
Life cycle	*Not applicable by definition*	**[LC1$^-$]** More integrated devices require more advanced technologies with potentially higher impacts. **[LC2$^-$]** Contamination risks makes circular end of life options difficult. **[LC3$^-$]** mHealth is expected to be deployed at a large scale. **[LC4$^-$]** More ICT infrastructure will be required to support eHealth.
Enabling	**[E1$^+$]** Telehealth reduces travel and use of other resources. **[E2$^+$]** Remote monitoring can reduce the frequency and duration of hospital stays. **[E3$^+$]** Prevention and early diagnosis could reduce the amount of procedures. **[E4$^+$]** More targeted, personalized and possibly more efficient evidence-based medicine optimizes medical procedures and reduces drugs, procedures, etc.	**[E5$^-$]** Demand for healthcare increases due to medicalization and adverse effects. **[E6$^-$]** Lifetime of electronic and wearable devices is shortened by obsolescence (hardware, software or psychological) and fast innovation cycles.
Structural	**[S1$^+$]** ICT helps promote more sustainable business models in the healthcare sector.	**[S2$^-$]** Rebound effects of energy/carbon savings countervail the potential efficiency improvements.
	[S3] Remote health services lead to the centralization of specialized medical infrastructure. **[S4]** Practices and regulations are transferred between healthcare and ICT sectors. **[S5]** Changing care paradigm restructures the healthcare institutions.	

the most extensive results. In the review, only 11 studies reported LCA results for medical devices [28, 59, 170, 171, 174, 175, 178, 180, 183, 186, 210]. Within these, two of them focused on wearable devices: the article by Ma and Chan on a smartwatch [59] and the thesis of Kokare on cardiac monitoring devices [28]. Despite the increasing efforts, LCA studies remain difficult to perform due to the length of the process, the required expertise and the lack of available data, especially for electronic devices.

Several characteristics of mHealth devices point towards an increase of their direct impacts compared to traditional medical devices. **[LC1$^-$]** At the unit scale, mHealth devices often consist in complex devices: flexible and biocompatible materials for skin-worn electronics, devices seamlessly integrated into garments, advanced technologies for small form factor, etc. These technologies all involve complex material streams that make their fabrication and disposal increasingly difficult to handle. For instance, Köhler *et al.* [182] highlighted future issues with the disposal schemes of e-textiles due to the tight integration of electronics into the fabrics. **[LC2$^-$]** In some cases, these issues with more complex streams might be further amplified by the safety precautions required for devices that have been contaminated though contact with tissues or biological fluids. Those hinder the implementation of circular business models that reduce end-of-life impacts by enhancing recycling or refurbishing rates [213].

While the previous effects refer to the relative impacts of individual devices, the overall environmental impacts depend on the absolute footprint of the devices, which is proportional to the scale of the deployment. For instance, as discussed in Chapter 1, the production carbon footprint of a signal wearable device is limited (around a few $kgCO_2e$'s, with a lot of variability), but must be multiplied by the number of devices produced. As mHealth is built as a distributed infrastructure, the amount of electronic equipment for mHealth is therefore likely to increase by design. **[LC3$^-$]** While applications for smartphones mainly rely on existing devices, mHealth also includes a large number of wearable connected sensors that will add to the existing ICT infrastructure. This market, often referred to the Internet of Medical Things (IoMT) in reference to the global Internet of Things (IoT), is expected to grow very rapidly with millions of new connected devices [62]. This growth in the number of devices is not only due to economic reasons, but is also sometimes required by design. Indeed, for instance, lots of devices are necessary to justify the shared networks and infrastructure which must be installed to connect them. Similarly, the arti-

ficial intelligence algorithms often used in combination with these devices require lots of data originating from many devices, in order to improve their own accuracy. **[LC4$^-$]** Wearable devices usually require additional infrastructure in order to store, transmit and process the huge quantities of data generated. As the ICT infrastructure of healthcare, including dedicated computers or datacenters, has not been designed to manage all this data, it will also need to grow in the future. In telemedicine-related studies, this impact of the ICT infrastructure has often shown to be negligible, but the scope was usually only limited to videoconference or telephone call during the time of the remote consultation, at a limited scale. Those conclusions will likely therefore change if telehealth becomes more mainstream and other uses start to emerge.

5.4.2 Enabling impacts

In the scientific literature, indirect *Enabling* impacts have been mostly studied from the perspective of reduced travel thanks to telehealth. **[E1$^+$]** In the review of Section 5.2, 11 studies that discussed telemedicine reported a decrease in the carbon footprint obtained by substituting face-to-face appointments with remote consultations [58, 200, 201, 202, 203, 204, 205, 206, 207, 208, 223]. While most of these studies focus on specialized medicine associated with longer travel distances, the carbon break-even point between indirect travel emissions and direct ICT footprint has been estimated at only a few kilometers (16 km [207], 7.2 km [201]).

Other reductions have also been reported in those studies, though less frequently. **[E2$^+$]** One example is the information obtained through ICT devices which can be used to optimize the use of healthcare resources. Audebert *et al.* [205] mentioned that telehealth and telemonitoring could be used as a basis to reduce unnecessary hospital admissions. Indeed, on top of the reduced travel, large amounts of resources and energy are consumed at the hospital, which can be avoided by relying on data obtained remotely. Similarly, Wang *et al.* [204] proposed using preoperative triage remotely to assess whether the patient is fit for surgery travelling to the hospital. Thanks to telemedicine, additional savings are obtained from the reduced use of physical resources such as personal protective equipment [207] or paper [206] associated with physical consultations.

To the best of our knowledge, other positive indirect effects of mHealth have not been discussed in the literature, yet present a high potential for reducing environmental impacts. Indeed, when considering the carbon

footprint, as shown in Fig. 5.3, many emission sources contribute to the global footprint. Among those, travel which receives most of the attention only accounts for 14 to 16% of the total [165, 167]. Relying on mHealth to optimize specifically the other emission sources such as the supply chain or hospital resources may prove equally or more effective in reducing the sector impacts. **[E3$^+$]** One way to do so is to increase preventive medicine. It has been mentioned as a lever to reduce the healthcare footprint, thanks to the reduction of resource-hungry curative medical procedures [165, 167]. By making it possible to interact more frequently with the patients, telehealth has been considered as an interesting tool to promote healthy behaviors and prevent medical issues [225, 226, 227]. **[E4$^+$]** It has also been reported that up to 40% of care is either useless or even harmful [228]. In addition to the effects on the health, cost and comfort for the patient, this also induces additional environmental impacts. Artificial intelligence, relying in part on mHealth sensor data, could be used to pave the way towards a more evidence-based and personalized medicine in order to bridge this gap [229]. Given the large proportion of environmental impacts originating from the supply chain (Fig. 5.3), this data-driven optimization could induce significant savings.

Enabling impacts can however also be negative when they cause an increase in the environmental footprint. Induction effects are an example of such negative indirect impact which occur when the use of a technology stimulates the consumption of similar or different resources. **[E5$^-$]** For mHealth, induction effects could appear as an increase in the demand for healthcare. For instance, adverse health effects may occur due to the use of eHealth applications. Behne and Teuteberg [230] reviewed some of the negative health impacts of these technologies such as increased stress, reduced quality of sleep, or overcompensation behaviours. Additionally, even in the absence of actual health issues, the demand in healthcare may increase due to the medicalization of previously non-medical aspects of the patients life. Baker [231] hypothesized that the feedback obtained from a wearable device would encourage the wearer to self-correct toward a norm that may not be well defined. While the ensuing actions could be beneficial, e.g., exercising more or improving their diet, it could also lead to unnecessary medical consultations.

[E6$^-$] Finally, connected objects with embedded electronics tend to be subject from shorter innovation cycles, which leads to increased obsolescence [232, 233]. Obsolescence can occur in multiple ways: components of

the system failing prematurely (hardware obsolescence), software updates not being provided for older systems (software obsolescence), or features from new devices lessening the appeal for previous versions (psychological obsolescence). This effect has also been reported for wearable devices such as hearing aids [234], e-textiles [60, 182] or handheld assistant devices for nurses [235]. This shortened life-cycle increases the quantities of devices that must be replaced, along with the associated fabrication and end-of-life impact overhead.

5.4.3 Structural impacts

The level of *Structural* impacts refers to macroscopic, systemic and long-term changes to institutions or economic structures. Due to their more complex nature, they are often more difficult to describe and predict. This part is therefore more of a prospective exercise. **[S1$^+$]** At the business level, ICT has been shown to be a key component in the transition towards more sustainable production and consumption patterns, by using the connectivity properties to enhance material circularity. In a healthcare industry that is still very reliant on disposable single-use devices, more circular business models could emerge to reduce the amounts of waste, with the support of digital devices and mHealth [213, 236]. However, digital systems are not necessarily associated with these more sustainable patterns, for example as they have also contributed to increase consumption in some contexts.

[S2$^-$] Rebound effects, also know as the Jevons paradox, take place at the macro-level when the resources (energy, time, money, etc.) saved by improving efficiency are reinvested and increase the demand for similar or different services, i.e. direct or indirect rebound effects. Overall, this effect curbs the gains that would be expected from the efficiency improvement alone and, in some cases, even leads to an increase in the absolute amount of resources spent [53], which are called backfire cases. For instance, even though many studies on the travel-related emissions of telemedicine assume that all other factors remain equal, Purohit *et al.* mention the possibility of rebound effects in that context, as time and cost savings may be reinvested in other activities [58]. While this effect has not been specifically studied for telemedicine, some studies on telework show that the expected GHG reductions are significantly mitigated by such second-order effects [237]. More generally, due to the behavioral and societal factors, it is difficult to predict if and how the energy, time and cost savings from mHealth and telemedicine will be reinvested in other areas and thus

reduce potential abatement [238].

Regarding some other systemic effects, it can be extremely difficult to assess whether the global impact will end up as positive or negative. **[S3]** For example, Cornaggia *et al.* showed that telemedicine puts a financial strain on rural hospital due to the redistribution of services towards more specialized urban hospitals [239]. This financial hardship may cause bankruptcy for smaller hospitals and lead to centralization of care in larger facilities. On the one hand, this could lead to economies of scale and improved energy performance in the hospitals. On the other hand, as telemedicine cannot fully replace face-to-face appointments, typically for physical examination or treatment, this centralization could increase the length of individual journeys, which is a form of rebound effect as telemedicine is intended to limit these journeys.

Structural effects also include changes at the institutional level. For example, institutions, regulatory bodies or insurance organisations already play an important role in the biomedical industry. They decide which devices are deemed safe enough to be commercialized, whether social security or private insurance cover the medical costs, what the optimal guidelines are for medical personnel intervention, and so on. When new technologies emerge, changes in the laws and procedures may be required to adapt to their specificities. **[S4]** Regulation has often been reported as both a barrier [173, 214, 218, 224] or a driver [169, 220, 221] for implementing more sustainable healthcare solutions. For instance, in the European Union, some directives such as the ones on Waste Electrical and Electronic Equipment Directive (WEEE), Restriction of Hazardous Substances Directive (RoHS) or Ecodesign already encourage more sustainable practices, including for mHealth devices [220]. However, derogations to these regulations can be obtained for some medical devices, for example in the case of recycling of infected devices. Given the current concerns for the environment, these are likely to become increasingly stringent. At the same time, other regulatory constraints regarding, e.g., safety or privacy put an additional, although necessary, burden on the design of the devices, thus shifting the focus away from environmental sustainability.

Given their impact on the healthcare paradigm, eHealth has the potential to deeply modify the way healthcare currently operates in the long-run [240], ultimately having implications for the environmental footprint. **[S5]** Some examples of the potential profound changes are the modification of the role of doctors through the patient empowerment [241], the increas-

ing importance of technology in healthcare services [242] or the overall improvements in the health outcomes. The amplitude of these effects are still very much unknown and their future environmental impacts would be very difficult to estimate.

5.5 Conclusion

mHealth and eHealth are on the verge of profoundly transforming the healthcare system. This transformation is likely to have significant impacts on environmental sustainability, but they are still difficult to evaluate due to the limited knowledge on these issues. This chapter aimed at providing a better understanding of the global environmental impacts of mHealth using the LES framework, and highlighting the lack of a systemic vision. This has been done by first evaluating how environmental impacts are considered in the scientific literature, then highlighting the current limitations in the evaluation of these impacts, and finally proposing a systemic framework that takes into consideration a macroscopic view of the mHealth system. By looking at the potential evolutions of the mHealth system and its impacts, some prospective examples were used to illustrate these larger impacts. This is a first step towards an approach that looks further than usual considerations in the literature, which remain limited to life-cycle impacts of individual devices and travel reduction.

The more comprehensive LES framework also highlights the complexity of evaluating the overall impact of design directions. Indeed, the environmental benefit or detriment of a technology comes from the balance between several positive (substitution, optimization, structural modifications, etc.) and negative (life-cycle, induction, obsolescence, rebound effects, etc.) impacts, which can hardly be predicted. In the case of specific design choices, these induced effects can be synergistic or antagonistic, i.e., acting in similar or opposite directions, respectively. For instance, the implementation of smart wearable devices discussed in Chapter 1, which embed local processing of the data to limit the wireless transmission data rates, can reduce the average power consumption of the device, in addition to mitigating the need for infrastructure to transfer, store and process large quantities of data. On the other hand, antagonistic effects may occur, such as the more complex and integrated devices with higher production impacts, or increased obsolescence due to specialization of the device if upgradability is not included. Evaluating the balance of these choices is

not straightforward, especially when the effects span multiple levels.

Environmental sustainability considerations have been mostly lacking in the healthcare sector due to several reasons [220]. Firstly, with human health at stake, the concerns about environmental impacts have often been considered, knowingly or unknowingly, as a lesser evil and left aside in terms of priority. Secondly, the medical industry is already facing many barriers such as strict economic constraints, tedious regulatory procedures, or a strong emphasis on patient safety. In parallel to this, few economic or legal incentives exist to incite companies and institutions to change the status quo and take environmental constraints into consideration. Thirdly, an important barrier is still the lack of knowledge pertaining to environmentally-friendly practices for medical device designers, policymakers, medical staff, or patients. Despite these barriers, awareness about these issues is rising and many stakeholders are willing to act. In particular, in healthcare, it is becoming more clear that environmental impacts are directly linked to human health due to the consequences of climate change or environmental pollution, for instance. When healthcare activities contribute to an increase in environmental damage, this comes as a direct contradiction of the "first, do no harm" injunction to medical personnel.

Research around those larger environmental impacts of mHealth is required in order to design future guidelines and policies, and to ensure that mHealth leads the healthcare sector on a path towards increased sustainability. As concerns regarding environmental footprint are currently rising, the priority given to sustainability in the design and choice of new technologies is likely to rise, but requires the proper methodologies and tools. While the proposed framework provides a way to characterize and categorize the potential effects of these technologies, there is still a long road to providing actionable guidelines for their development. As the effects discussed in Section 5.4 remain hypothetical, future steps consist in their study, and quantification when possible, in relevant use cases to evaluate their relative importance. The balance between positive and negative impacts could then be assessed for different mHealth technologies and in different contexts. Indeed, while wearable technologies could provide valuable benefits in some situations, the outcomes could change in other cases due to differences in cultures, geographical and political contexts, or needs for the healthcare system. A striking example of the impact of the context on the sustainability of mHealth technologies is the difference between high- and low-income countries. In the former case, mHealth adds

to a generally efficient and robust healthcare system, whereas in the latter case it is used to compensate for the limited infrastructure. The sustainability challenges and barriers to its implementation therefore differ significantly [158, 243]. Further research into these questions would be valuable for both designers in order to optimize the impacts of their devices, and for policy-makers and institutions to properly regulate the development of these emerging technologies.

As described in Section 5.4, the indirect environmental effects of mHealth induce changes in the healthcare system as a whole. The study of these changes involves medical, economic, social, psychological, political or behavioral questions that are beyond the traditional scope of engineering. In the LES model that is used to structure the impact categories, the life cycle level is usually limited to technical considerations and is easier to quantify, which also likely explains why it has been the focal point of previous research. Enabling and structural changes are less well-defined, depend on user behavior and institutional response, and they are less prone to precise quantification. An interdisciplinary dialogue therefore seems necessary to reconcile these different characteristics and fully grasp those systemic impacts. It is also worth remembering that environmental issues are only a part of the overall sustainability question, in parallel with economic and social considerations. This multi-sided vision ensures that mHealth has the least detrimental impact on the environment, but is also socially acceptable, e.g., with fair treatment of the workforce and respect of the local and global communities, and economically viable, i.e., thanks to cost effectiveness and proper governance. These considerations further highlight the need for interdisciplinary research, in order to provide the proper tools and education to the stakeholders [JP3].

Finally, while improving efficiency and minimizing environmental impacts of the individual devices are necessary strategies, they alone will not address the systemic impacts and potential rebound effects. The concept of digital sufficiency described by Santarius *et al.* [52] provides clues on how to identify ICT solutions that foster a reduction in the absolute environmental footprint. Applied to mHealth, sufficiency consists in ensuring that new technological developments actually answer the medical needs, that their deployment and use fits the intended purpose, or that they provide low-impact alternatives to existing services. Considering that we only have finite resources at our disposal, this question of selecting only the most valuable outcomes can be seen as a sensitive issue, especially in the

context of health which is often seen as a primary concern. These issues call for a wider societal reflection on our health objectives and how technology will contribute to them.

6

Conclusion

This thesis explored the design and optimization of ultra-low-power biomedical sensor systems for mHealth technologies through the different design layers illustrated in Fig. 6.1. In Chapter 2, this exploration started at the bottom of the abstraction hierarchy with the design of the analog and digital, hardware and software components of an electronic integrated system for a wearable monitoring device. The system design targeted the use case of signal acquisition and arrhythmia classification in long-term ECG monitoring. In Chapter 3 between the circuit and system layers, the performance and power consumption of the internal blocks of the system were first modeled in order to evaluate the design trade-offs. These performance characteristics were then optimized numerically to fit the specifications of the system and minimize the total power consumption in Chapter 4. Finally, at the application level, a literature review reconsidered the sustainability of wearable devices and mHealth technologies in Chapter 5. Three main research questions related to the design of mixed-signal electronic systems for wearable devices were highlighted at the beginning of the thesis. This concluding chapter consecutively reviews these questions in the light of the results presented in the previous chapters. Some perspectives for further work are also outlined.

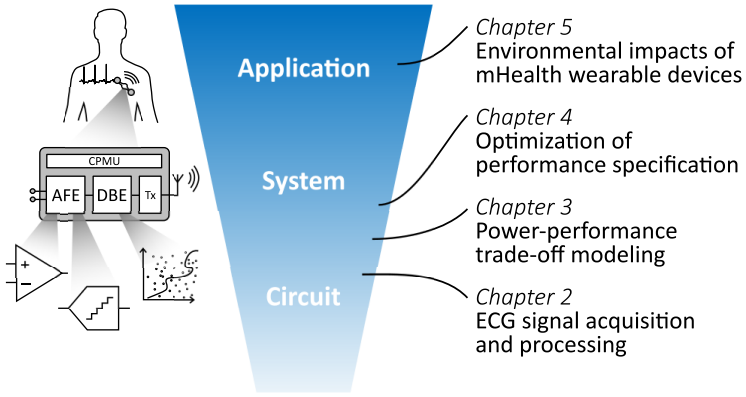

Fig. 6.1 Circuits and systems abstraction layers. The different chapters explored different parts of the abstraction hierarchy.

6.1 ULP MCU for complex embedded processing

Microcontrollers are at the core of many sensor devices as they enable on-chip processing of the acquired signals. Wearable devices for biomedical applications often rely on complex algorithms to extract the information from the raw biosignal. These algorithms require high computing resources, which tend to increase the power consumption in energy-constrained devices. While application-specific circuits are often used for their higher computing performance and energy efficiency, they often lack the versatility of MCUs. In order to bring together the upgradability of MCUs, the computing requirements of biomedical applications and the low power budget of wearable devices, the following question was raised:

Question 1 *Can ultra-low-power microcontroller units enable complex on-chip real-time signal processing such as beat-to-beat arrhythmia classification, within a limited power budget of a few µW's?*

A ULP MCU prototype, codenamed SleepRider and fabricated in 28-nm FD-SOI, was designed in Chapter 2 to demonstrate the ECG monitoring and arrhythmia classification application. The high energy efficiency of the ultra-low-voltage logic and ultra-low-power memories combined with the low power consumption of the custom AFE enable the proposed system to achieve the target use case with an average power of only 13.1 µW. The implementation of the system shows that all components of the system, both analog and digital, hardware and software, must be optimized

to fit the limited power budget. More specifically, the IA in the AFE is optimized for low noise and low power consumption while meeting the constraints of measuring biopotentials. The DBE is operated at ultra-low-voltage for minimizing the processing power, relying on adaptive back-biasing to enable a high clock frequency and compensate PVT variations, as well as ULP SRAM and low-power CPMU to further reduce the DBE energy. Finally, the classification algorithm and its software implementation are designed for a resource-constrained device, namely by pruning the algorithm to reduce the memory and computing requirements.

While the implementation proposed in the first chapter focused on arrhythmia classification in ECG monitoring devices, ULP MCUs can be useful in many different energy-constrained applications. More specifically, they are particularly fit for applications that have mid-range requirements in terms of data rate (around 0.1 to 100 ksps), frequent operation (e.g., active duty-cycle higher than 0.1%) and processing requirements (typically small machine-learning models). These applications can for example include biomedical, audio, or industrial monitoring use cases. In contrast, in applications with very low data rates and limited processing requirements, such as simple environmental sensors for light, humidity or temperature, their extremely low duty-cycle makes the power consumption in active mode negligible. The necessity of energy-efficient processing is therefore quite limited. On the opposite end of the spectrum, some applications require that the device handles very high data rates or complex data processing, such as in real-time intelligent vision systems. These applications may not fit the specifications of ULP MCUs, whose peak performance is limited by the MHz-range clock frequency and the scalar processor. Additional hardware acceleration or parallel processing are likely necessary to meet the requirements of these applications. In order to map the circuit-level performance of ULP MCUs to the applications benefiting most from their characteristics, future research would need to expand this study to different applications with various acquisition and processing tasks.

In the introduction of this thesis, it was assumed that local processing of the acquired signals thanks to embedded data processing would reduce the device power consumption and mitigate the environmental footprint of the application by reducing the amount of data transmitted through the network. Some adverse effects were highlighted such as the risk of faster obsolescence, the increased energy consumption and the impact on the production due to more complex devices. While Chapter 2 addressed

the first two issues by achieving low-energy computations with an upgradable platform, the latter is still subject to significant uncertainty. Indeed, the added DBE required to perform the signal processing impacts the production environmental footprint in two ways. First, a larger silicon area is required to include the digital circuits, which can be significantly larger than the AFE alone. In Chapter 2 for instance, the logic and memories are respectively 6× and 10× larger than the AFE. Secondly, the implementation of ULP MCUs often relies on more advanced technology nodes to benefit from the reduced energy in digital circuits, compared to usual AFEs that mostly use older nodes, as detailed in Tables 2.3 and 2.4. These more advanced nodes typically have a larger production footprint per silicon area [61, 102]. While digital circuits benefit from a higher density which balances these impacts, the continued integration of on-chip functionalities and the limited scaling of analog circuits may lead to an overall increase in the footprint [244]. Further investigations of these secondary impacts of embedding advanced processing in the circuits, in parallel to the optimization of their implementation, will be necessary to allow designers to make an informed choice with regards to the best integration strategy.

6.2 Numerical optimization in hierarchical system design

In the field of electronic system design, a lot of research focuses on the local improvement of circuit-level performance for the multiple blocks that make up electronic systems. By pushing the technical boundaries of individual circuits, the global system performance is also enhanced. However, these improvements may not be fully leveraged if the target block-level specifications do not directly serve the system-level performance objective. Multiple hierarchical design methods have been proposed and are used on a daily basis by design teams. Yet, none of these actually guarantees to reach a global system performance optimum, and few can directly apply to large mixed-signal SoCs that are often found in sensor systems. The second question addressed in this thesis was therefore:

Question 2 *How can we tune the block-level performance requirements in order to minimize the system power consumption while maintaining the inference accuracy in complex mixed-signal sensor systems?*

In Chapters 3 and 4, a hierarchical cross-domain system-level methodology was proposed to minimize the power consumption of the MCU-based arrhythmia classification system of Chapter 2. This methodology relies on

three steps: (1) the modeling of block-level power and performance tradeoffs, (2) the implementation of a system performance macromodel for efficient inference accuracy evaluation, and (3) he numerical optimization of the system based on the implemented model. Several methods for extracting Pareto performance of circuits are presented in Chapter 3, respectively for knowledge-based, simulation-based and software emulation approaches. These different methods can be used to model diverse circuits, from simpler to more complex, in analog or digital domains, as illustrated with the different circuits in the arrhythmia classification system. The second and third steps are discussed in Chapter 4. By applying the proposed methodology to the target system, the results demonstrate that our numerical optimization approach allows to find a design optimum for the system and reduce the total power consumption.

This work provides new insights on how numerical optimization can benefit system-level design. Despite its potential for optimizing system performance, the application of the methodology still shows some technical limitations which open perspectives for future research. Firstly, the long simulation time for the system macromodel, leading to an optimization time of several hours, is a hurdle for practical use of the method. The total execution time could be reduced either by implementing a faster model, by relying on parallel execution, or by reducing the number of function calls in the optimization algorithm. A faster model would mainly require working on the SVM classifier training, either by reducing the training frequency or the training time. Parallel execution requires splitting the numerical simulations in several threads, which could be done easily as the database already consists in several independent records. Finally, the number of function calls mainly depends on the type of algorithm used. Gradient-based algorithms are often considered as the most efficient in this regard. Yet, other algorithms such as surrogate model-based Bayesian optimization are worth comparing, as they do not require gradient estimation and are also meant to limit the number of function evaluations [116]. Secondly, the evaluation noise of the system macromodel interferes with the gradient-based optimization process. This tends to increase convergence time or provide false optima. As the noise is an intrinsic part of the model, the only possibility to reduce it is to rely on averaging, at the cost of increased simulation time. Alternatively, the optimization algorithm could be made more robust to noise, either using a modified SQP algorithm [245], or alternatives such as Bayesian optimization [116].

Another important limitation of the proposed methodology is the need for an accurate macromodel of the system, which includes parametrized non-idealities for the optimized blocks. This model can require significant work to implement, especially as the system grows more complex. In particular, important characteristics of the model should be its implementation and simulation time, accuracy, and robustness, which rarely go together. Design time is often a critical factor. While a designer may prefer to spend a minimum of time implementing the system model, making an extra effort to ensure that the simulation time is short enough pays back when executing the optimization process. Implementing models of the circuits that fit as close as possible their actual behavior is a challenging task, especially as there is a trade-off between model complexity and accuracy. In the end, the accuracy of the model directly impacts the accuracy of the optimization result. The robustness of the model is also an important aspect as the optimization algorithm may explore untested parts of the parameter space, which can result in unexpected and possibly wrong results. Fortunately, if circuit blocks are standard and appear in different systems, the models can be reused and the implementation work is shared. As an alternative to design-time optimization which requires accurate and robust models, post-silicon tuning can also be considered with configurable circuit blocks, at the expense of a more complex analog design for configurability implementation.

In this thesis, the proposed methodology is applied specifically to the use case of ECG arrhythmia classification. However, the method and discussions can surely be extended to other types of biomedical and non-biomedical sensors that share similar characteristics. The main feature that makes the presented approach relevant is the balance between different blocks of the system with regards to their impact on total power and overall performance. This balance is often found for instance in the acquisition and processing of low-dimensional signals with a data rate in the kHz-range. Depending on the exact use cases, several biosignal monitoring applications may fit these requirements, such as epilepsy detection on EEG signals or stress measurement from multimodal signals. Other use cases outside biomedical applications can also enter in this category such as speech processing in audio sensors or machine fault detection accelerometer-based systems. Inversely, when only one of the blocks consumes much more power than the others or is the major determinant of the overall performance, the usual approach of focusing the optimization on this single block

is the most relevant. In order to evaluate the scope of the systems which could benefit from the proposed methodology, more research is necessary to target use cases with varying importance of analog or digital circuits, different signal acquisition constraints or new processing tasks.

6.3 Global approach to mHealth environmental impacts

Environmental sustainability is a growing concern in both the ICT and healthcare sectors. In ICT research, the efforts towards the reduction of environmental impacts is still largely focused on the issues of energy efficiency and reduction of batteries. In the healthcare sector, the considerations for the environment tend to revolve mostly around the issues of waste generation and the life cycle impacts of medical devices. In both sectors, the digitalization of services is seen as a major tool for reducing impacts and paving the way towards sustainable systems. At their intersection, mHealth is a recent technological trend which carries a lot of expectations for improving the cost and medical efficiency of healthcare services, and reducing their environmental footprint. Despite the hype around these new technologies and their expected potential, there is still limited evidence regarding their actual benefits, and the efforts for evaluating their environmental sustainability are quite fragmented. The last question addressed in this thesis proposes to take a step back with regards to the best way to improve the sustainability of mHealth devices and systems:

Question 3 *How can we assess whether mHealth technologies contribute to the environmental sustainability of the healthcare system?*

By reviewing and commenting on the scientific literature, this work highlighted the lack of a consistent approach to evaluate the environmental impacts of mHealth, and specifically the lack of a framework that considers more comprehensively the indirect effects of these technologies as a system. As an answer to these shortcomings, we proposed to rely on the LES framework, which allows to explore deeper the systemic consequences for the environment of introducing mHealth technologies in the healthcare system. Numerous examples of indirect impacts, both enabling and systemic, have been suggested based on examples from the literature. This larger perspective on environmental impacts allows to better anticipate the potential adverse effects of a new technological deployment, and therefore possibly mitigate them. The long-term repercussions of this systemic impact assessment can range from device eco-design in order to foster pos-

itive impacts and avoid the negative ones, to new regulations or policies which aim at favouring applications that enhance the sustainability of the healthcare system.

While the LES framework has the advantage of structuring and opening perspectives with regards to the effects that are considered, it is not meant to provide a final exhaustive quantification of the environmental impacts. It therefore opens many new lines of research, in order to first investigate the existence and magnitude of these potential impacts in different use cases. Indeed, there exists a large variety of devices which can respectively be used in various manners or contexts, which influence the scale of environmental impacts. The sustainability assessment of mHealth technologies therefore needs to consider these dimensions. Ultimately, sustainability assessment also questions the countermeasures that can be considered to mitigate those impacts at different design, validation or regulatory levels, in order to better select and optimize these emerging technologies.

An important aspect which is emphasized with the proposed systemic approach is the need to initiate a multidisciplinary dialogue. Indeed, the dimensions of environmental sustainability involving mostly technical aspects are those that have understandably received most of the attention from the research community. In the meantime, the other dimensions, including aspects that are further away from the engineering perspectives such as psychology, economy or sociology, are often poorly evaluated, if at all. Removing barriers between research fields is however not an easy task as those do not share the same practices, evaluation metrics or vocabulary. Future research on sustainability will therefore likely require working in an interdisciplinary or transdisciplinary fashion, for building both new knowledge and new methods. For instance, one can think of the economic evaluation methodologies used to balance costs and benefits in complex markets, which may provide ways to address those problems.

6.4 From system optimization to sustainable system design

Hierarchical thinking is at the core of the engineering practice. Breaking down a complex problem into smaller ones allows to make them individually easier to understand and solve. In this thesis, this is illustrated by the division of the ECG arrhythmia classification system into several analog or digital circuits, or by an mHealth application broken down into wearable sensor systems. While this approach is very effective and has allowed

to build very complex technological ecosystems, it also has its flaws. The division of a larger system and the abstraction of its components often entail some assumptions and simplifications. Optimizing a component of a larger system on its own, without considering the implications of including it into the system, may lead to suboptimal or non-functional solutions. These issues may arise from the improper specifications of target performance, or from unexpected interactions between components. The considerations at the lower levels are thus often decoupled from the issues that arise at a higher hierarchy level. Additionally, as the technologies become more complex and the design teams grow, these issues are amplified in practice when different people handle separate levels of abstractions.

As the approach proposed in Chapter 4 tries to overcome these limitations and allows to bridge the circuit and system abstraction layers to optimize the power consumption, one may wonder how this framework can be expanded to include environmental sustainability aspects in the ecodesign of wearable systems. This objective is however not without its challenges. As a first step, additional specifications for the system design could include environmental metrics. On top of existing circuit objectives such as power consumption and accuracy, or system objectives regarding safety, cost or comfort, new targets such as GHG emissions or raw material intake could be considered. These new metrics would therefore allow to include considerations for the direct impacts of the device into the design process. However, as previously reported, the evaluation of these environmental metrics is not straightforward. Exhaustive environmental assessment such as LCA is a tedious and time-consuming process. Simplified models require characterization data which is still hard to come by and often suffer from significant variability [27].

However, as seen in Chapter 5, environmental impacts of wearable devices extends much beyond the direct life-cycle impacts. The translation of the indirect enabling and structural impacts for these applications across the abstraction hierarchy is still out of reach for several reasons. First, environmental sustainability involves many indicators which do not evolve in the same way. A design that minimizes the carbon footprint may generate more waste or require more non-renewable resources. Secondly, the methods for evaluating the indirect impacts are not yet mature due to the complexity of the relationships between the device and its effects in the society. The LES framework, which is proposed as a way to evaluate systemic impacts of mHealth, only explores the possible outcomes, but does

not provide any quantification of the impacts. Furthermore, the indirect effects of the devices are highly variable, which creates a lot of uncertainty for their quantification. Thirdly, while the design of a wearable device is only performed once, it can then be used in a variety of contexts, which will have distinct implications for the environmental footprint due to political, economic, geographical or cultural differences. For example, while a given telemedicine service can be implemented in different parts of the world, the environmental benefits strongly depend on the locations of the patients and hospitals. Finally, when systemic impacts are considered, the technical design of a device is not the only factor influencing its sustainability. The business model for its distribution or the regulations of the sector are as important as the device itself. These different aspects will need to be addressed, first by studying them separately in well-defined scopes and contexts, then by combining them in more complex models. All these aspects require new knowledge and tools, which likely involve other scientific disciplines. This calls for re-thinking design methodologies, so as to include interdisciplinary sustainability questions into the engineering practice. As we face ever more pressing environmental constraints, one such global approach would help anticipate the outcomes of emerging technologies, to avoid future adverse effects and prevent technological lock-in situations. Our role as engineers is not only to design better technologies, but also to understand and discuss how and when they should be used in order to maximize their benefits.

Bibliography

[1] C. R. Doarn, S. Pruitt, J. Jacobs, Y. Harris, D. M. Bott, W. Riley, C. Lamer, and A. L. Oliver, "Federal Efforts to Define and Advance Telehealth - a Work in Progress," *Telemedicine and e-Health*, vol. 20, no. 5, pp. 409–418, 2014.

[2] World Health Organisation, "mHealth – New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth," 2011.

[3] S. Iqbal, I. Mahgoub, E. Du, M. A. Leavitt, and W. Asghar, "Advances in Healthcare Wearable Devices," *NPJ Flexible Electronics*, vol. 5, no. 1, pp. 1–14, 2021.

[4] fitbit, "fitbit," 2022. [Online]. Available: https://www.fitbit.com/

[5] Withings, "Pulse HR," 2022. [Online]. Available: https://www.withings.com/pulse-hr

[6] Dexcom, "Continuous Glucose Monitoring," 2022. [Online]. Available: https://www.dexcom.com/

[7] CardiacSense Ltd, "Watching Over You," 2022. [Online]. Available: https://www.cardiacsense.com/

[8] Q. Wang, M. Toeters, W. Chen, A. Timmermans, and P. Markopoulos, "Zishi: a Smart Garment for Posture Monitoring," *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 3792–3795, 2016.

[9] E. Rodriguez-Villegas, S. Iranmanesh, and S. A. Imtiaz, "Wearable Medical Devices: High-Level System Design Considerations and Tradeoffs," *IEEE Solid-State Circuits Magazine*, vol. 10, no. 4, pp. 43–52, 2018.

[10] Q. Lin, S. Song, I. D. Castro, H. Jiang, M. Konijnenburg, R. Van Wegberg, D. Biswas, S. Stanzione, W. Sijbers, C. Van Hoof *et al.*, "Wearable Multiple Modality Bio-Signal Recording and Processing on Chip: a Review," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1108–1123, 2020.

[11] M. Gao, P. Wang, L. Jiang, B. Wang, Y. Yao, S. Liu, D. Chu, W. Cheng, and Y. Lu, "Power Generation for Wearable Systems," *Energy & Environmental Science*, vol. 14, no. 4, pp. 2114–2157, 2021.

[12] *MAX30009 - Low-Power, High-Performance Bioimpedance Analog Front-End*, Analog Devices, 2022, Datasheet.

[13] *ADS1198 - Low-Power, 8-Channel, 16-Bit Analog Front-End for Biopotential Measurements*, Texas Instruments, 2011, Datasheet.

[14] G. Cosoli, S. Spinsante, F. Scardulla, L. D'Acquisto, and L. Scalise, "Wireless ECG and Cardiac Monitoring Systems: State of the Art, Available Commercial Devices and Useful Electronic Components," *Measurement*, vol. 177, p. 109243, 2021.

[15] K. Shibata, H. Matsui, H. Asano, Y. Kusaka, K. Ueda, N. Matsuno, and H. Sato, "A 22nm 0.84 mm^2 BLE Transceiver With Self IQ-Phase Correction Achieving 39dB Image Rejection and on-Chip Antenna Impedance Tuning," *2022 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 65, pp. 398–400, 2022.

[16] X. Zhang, Z. Zhang, Y. Li, C. Liu, Y. X. Guo, and Y. Lian, "A 2.89μW Dry-Electrode Enabled Clockless Wireless ECG SoC for Wearable Applications," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 10, pp. 2287–2298, 2016.

[17] A. F. Yeknami, X. Wang, I. Jeerapan, S. Imani, A. Nikoofard, J. Wang, and P. P. Mercier, "A 0.3-V CMOS Biofuel-Cell-Powered Wireless Glucose/Lactate Biosensing System," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 11, pp. 3126–3139, 2018.

[18] N. Modak, D. Das, M. Nath, B. Chatterjee, G. Kumar, S. Maity, and S. Sen, "EQS Res-HBC: a 65-nm Electro-Quasistatic Resonant 5-240 μW Human Whole-Body Powering and 2.19 μW Communication SoC with Automatic Maximum Resonant Power Tracking," *IEEE Journal of Solid-State Circuits*, vol. 57, no. 3, pp. 831–844, 2022.

[19] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, D. Niyato, O. Dobre, and H. V. Poor, "6G Internet of Things: a Comprehensive Survey," *IEEE Internet of Things Journal*, 2021.

[20] J. Aslan, K. Mayers, J. G. Koomey, and C. France, "Electricity Intensity of Internet Data Transmission: Untangling the Estimates," *Journal of Industrial Ecology*, vol. 22, no. 4, pp. 785–798, 2018.

[21] J. Malmodin and D. Lundén, "The Energy and Carbon Footprint of the Global ICT and E&M Sectors 2010–2015," *Sustainability*, vol. 10, no. 9, p. 3027, 2018.

[22] D. Bol, "Intelligent Energy-Efficient Systems at the Edge of IoT - Ultra-Low-Power SoCs for Local Sensor Data Processing," *2018 IEEE International Solid-State Circuits Conference (ISSCC), Forum*, 2018.

[23] E. Covi, E. Donati, X. Liang, D. Kappel, H. Heidari, M. Payvand, and W. Wang, "Adaptive Extreme Edge Computing for Wearable Sevices," *Frontiers in Neuroscience*, p. 429, 2021.

[24] I. Pang, Y. Okubo, D. Sturnieks, S. R. Lord, and M. A. Brodie, "Detection of Near Falls Using Wearable Devices: a Systematic Review," *Journal of Geriatric Physical Therapy*, vol. 42, no. 1, pp. 48–56, 2019.

[25] Y. C. Jo, H. N. Kim, W. H. Hwang, H. K. Hong, Y. S. Choi, and S. W. Jung, "Wearable Patch Device for Uterine EMG and Preterm Birth Monitoring Applications," *2018 IEEE Region 10 Conference (TENCON)*, pp. 1127–1130, 2018.

[26] Y. Xu, J. Yang, and M. Sawan, "Trends and Challenges of Processing Measurements from Wearable Devices Intended for Epileptic Seizure Prediction," *Journal of Signal Processing Systems*, pp. 1–16, 2021.

[27] T. Pirson and D. Bol, "Assessing the Embodied Carbon Footprint of IoT Edge Devices with a Bottom-Up Life-Cycle Approach," *Journal of Cleaner Production*, vol. 322, p. 128966, 2021.

[28] S. Kokare, "Comparative Life Cycle Assessment of Cardiac Monitoring Devices: a Case Study," Master's thesis, KTH Royal Institute of Technology, 2020.

[29] L.-P. P.-V.P. Clément, Q. E. S. Jacquemotte and L. M. Hilty, "Sources of Variation in Life Cycle Assessments of Smartphones and Tablet Computers," *Environmental Impact Assessment Review*, vol. 84, p. 106416, 2020.

[30] World Health Organization, "Global Health Estimates," https://www.who.int/data/global-health-estimates, 2020, accessed: 03-01-2022.

[31] J. H. Nagel, "Biopotential Amplifiers," *The Biomedical Engineering Handbook*, vol. 2, p. 1300, 2000.

[32] S. K. Berkaya, A. K. Uysal, E. S. Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, "A Survey on ECG Analysis," *Biomedical Signal Processing and Control*, vol. 43, pp. 216–235, 2018.

[33] M. A. Serhani, H. T. El Kassabi, H. Ismail, and A. Nujum Navaz, "ECG Monitoring Systems: Review, Architecture, Processes, and Key Challenges," *Sensors*, vol. 20, no. 6, p. 1796, 2020.

[34] iRhythm Technologies, "Zio by iRhythm," 2022. [Online]. Available: https://www.irhythmtech.com/

[35] S. Shareghi, M. Tavakol, K. Lindborg, C. Alfaro Vives, and L. Spaccavento, "SEEQ Mobile Cardiac Telemetry Associated with a High Yield of Clinically Relevant Arrhythmias in Patients with Suspected Arrhythmia," *Circulation*, vol. 134, no. suppl_1, pp. A16 078–A16 078, 2016.

[36] P. B. Heart, "ePatch - Extended Wear Holter Monitoring," 2022. [Online]. Available: https://www.gobio.com/heart-monitoring/#epatch

[37] Savvy, "Monitor the activity of your heart. Anytime and anywhere." 2022. [Online]. Available: http://www.savvy.si/en/

[38] A. Rashkovska, M. Depolli, I. Tomašić, V. Avbelj, and R. Trobec, "Medical-Grade ECG Sensor for Long-Term Monitoring," *Sensors*, vol. 20, no. 6, p. 1695, 2020.

[39] J. Liu, Z. Zhu, Y. Zhou, N. Wang, G. Dai, Q. Liu, J. Xiao, Y. Xie, Z. Zhong, H. Liu *et al.*, "BioAIP: A Reconfigurable Biomedical AI Processor with Adaptive Learning for Versatile Intelligent Health Monitoring," *2021 IEEE International Solid-State Circuits Conference (ISSCC)*, vol. 64, pp. 62–64, 2021.

[40] J. Wu, F. Li, Z. Chen, Y. Pu, and M. Zhan, "A Neural Network-Based ECG Classification Processor with Exploitation of Heartbeat Similarity," *IEEE Access*, vol. 7, pp. 172 774–172 782, 2019.

[41] P. Schönle, F. Glaser, T. Burger, G. Rovere, L. Benini, and Q. Huang, "A Multi-Sensor and Parallel Processing SoC for Miniaturized Medical Instrumentation," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 7, pp. 2076–2087, 2018.

[42] S. Song, M. Konijnenburg, R. van Wegberg, J. Xu, H. Ha, W. Sijbers, S. Stanzione, D. Biswas, A. Breeschoten, P. Vis, C. van Liempd, C. van Hoof, and N. van Helleputte, "A 769 μW Battery-Powered Single-Chip SoC With BLE for Multi-Modal Vital Sign Monitoring Health Patches," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 6, pp. 1506–1517, 2019.

[43] Y.-P. Chen, D. Jeon, Y. Lee, Y. Kim, Z. Foo, I. Lee, N. B. Langhals, G. Kruger, H. Oral, O. Berenfeld, Z. Zhang, D. Blaauw, and D. Sylvester, "An Injectable 64nW ECG Mixed-Signal SoC in 65nm for Arrhythmia Monitoring," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 1, pp. 375–390, 2015.

[44] H. Kim, S. Kim, N. Van Helleputte, A. Artes, M. Konijnenburg, J. Huisken, C. Van Hoof, and R. F. Yazicioglu, "A Configurable and Low-Power Mixed Signal SoC for Portable ECG Monitoring Applications," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, no. 2, pp. 257–267, 2013.

[45] R. A. Rutenbar, G. G. Gielen, and J. Roychowdhury, "Hierarchical Modeling, Optimization, and Synthesis for System-Level Analog and RF designs," *Proceedings of the IEEE*, vol. 95, no. 3, pp. 640–669, 2007.

[46] S. Kundu and P. Mandal, "An Efficient Method of Pareto-Optimal Front Generation for Analog Circuits," *Analog Integrated Circuits and Signal Processing*, vol. 94, no. 2, pp. 289–316, 2018.

[47] G. Gielen, T. McConaghy, and T. Eeckelaert, "Performance Space Modeling for Hierarchical Synthesis of Analog Integrated Circuits," *2005 42nd Annual Design Automation Conference (DAC)*, pp. 881–886, 2005.

[48] G. Stehr, H. Graeb, and K. Antreich, "Feasibility Regions and their Significance to the Hierarchical Optimization of Analog and Mixed-Signal

Systems," in *Modeling, Simulation, and Optimization of Integrated Circuits*. Springer, 2003, pp. 167–184.

[49] D. Bol, G. de Streel, and D. Flandre, "Can We Connect Trillions of IoT Sensors in a Sustainable Way? A Technology/Circuit Perspective (Invited)," *2015 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, pp. 1–3, 2015.

[50] M. Alioto, "From Less Batteries to Battery-Less: Enabling a Greener World Through Ultra-Wide Power-Performance Adaptation Down to pWs (Invited)," *2022 48th IEEE European Solid-State Circuits Conference (ESSCIRC)*, 2022.

[51] L. M. Hilty and B. Aebischer, "ICT for Sustainability: An Emerging Research Field," in *ICT Innovations for Sustainability*. Springer, 2015.

[52] T. Santarius, J. C. Bieser, V. Frick, M. Höjer, M. Gossen, L. M. Hilty, E. Kern, J. Pohl, F. Rohde, and S. Lange, "Digital Sufficiency: Conceptual Considerations for ICTs on a Finite Planet," *Annals of Telecommunications*, pp. 1–19, 2022.

[53] C. Gossart, "Rebound Effects and ICT: a Review of the Literature," *ICT Innovations for Sustainability*, pp. 435–448, 2015.

[54] Å. Holmner, J. Rocklöv, N. Ng, and M. Nilsson, "Climate Change and eHealth: a Promising Strategy for Health Sector Mitigation and Adaptation," *Global Health Action*, vol. 5, no. 1, p. 18428, 2012.

[55] P.-P. Pichler, I. S. Jaccard, U. Weisz, and H. Weisz, "International Comparison of Health Care Carbon Footprints," *Environmental Research Letters*, vol. 14, no. 6, p. 064004, 2019.

[56] V. K. Parida, D. Sikarwar, A. Majumder, and A. K. Gupta, "An Assessment of Hospital Wastewater and Biomedical Waste Generation, Existing Legislations, Risk Assessment, Treatment Processes, and Scenario During COVID-19," *Journal of Environmental Management*, p. 114609, 2022.

[57] P. Di Giacomo and P. Håkansson, "A Method to Measure the Reduction of CO_2 Emissions in E-Health Applications," in *User Centred Networked Health Care*. IOS Press, 2011, pp. 970–974.

[58] A. Purohit, J. Smith, and A. Hibble, "Does Telemedicine Reduce the Carbon Footprint of Healthcare? A Systematic Review," *Future Healthcare Journal*, vol. 8, no. 1, p. e85, 2021.

[59] M. M. M. Ma, Z. Zhu, and Y. C. Chan, "Environmental Impact Analysis of Smartwatch Using SimaPro8 Tools and Energy Dispersive X-Ray Spectroscopy (EDX) Technique," *2017 IEEE 19th Electronics Packaging Technology Conference (EPTC)*, pp. 1–6, 2017.

[60] O. Gurova, T. R. Merritt, E. Papachristos, and J. Vaajakari, "Sustainable Solutions for Wearable Technologies: Mapping the Product Development Life Cycle," *Sustainability*, vol. 12, no. 20, p. 8444, 2020.

[61] M. G. Bardon, P. Wuytens, L.-Å. Ragnarsson, G. Mirabelli, D. Jang, G. Willems, A. Mallik, A. Spessot, J. Ryckaert, and B. Parvais, "DTCO Including Sustainability: Power-Performance-Area-Cost-Environmental Score (PPACE) Analysis for Logic Technologies," *2020 IEEE International Electron Devices Meeting (IEDM)*, pp. 41–4, 2020.

[62] Deloitte Centre for Health Solution, "Medtech and the Internet of Medical Things - How connected medical devices are transforming health care," 2018.

[63] D. Kraft, "A Software Package for Sequential Quadratic Programming," *Forschungsbericht- Deutsche Forschungs- und Versuchsanstalt fur Luft- und Raumfahrt*, 1988.

[64] D. O. Arnar, G. H. Mairesse, G. Boriani, H. Calkins, A. Chin, A. Coats, J.-C. Deharo, J. H. Svendsen, H. Heidbüchel, R. Isa *et al.*, "Management of Asymptomatic Arrhythmias: a European Heart Rhythm Association (EHRA) Consensus Document, Endorsed by the HFA, HRS, APHRS, CASSA, and LAHRS," *EP Europace*, 2019.

[65] S. Bose, B. Shen, and M. L. Johnston, "A Batteryless Motion-Adaptive Heartbeat Detection System-on-Chip Powered by Human Body Heat," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 11, pp. 2902–2913, 2020.

[66] K.-R. Lee, J. Kim, C. Kim, D. Han, J. Lee, J. Lee, H. Jeong, and H.-J. Yoo, "A 1.02-μW STT-MRAM-Based DNN ECG Arrhythmia Monitoring SoC With Leakage-Based Delay MAC Unit," *IEEE Solid-State Circuits Letters*, vol. 3, pp. 390–393, 2020.

[67] Y. Yin, S. M. Abubakar, S. Tan, J. Shi, P. Yang, W. Yang, H. Jiang, Z. Wang, W. Jia, and S.-P. U, "A 2.63μW ECG Processor With Adaptive Arrhythmia Detection and Data Compression for Implantable Cardiac Monitoring Device," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 4, pp. 777–790, 2021.

[68] S. Mondal, C.-L. Hsu, R. Jafari, and D. A. Hall, "A Dynamically Reconfigurable ECG Analog Front-End With a 2.5× Data-Dependent Power Reduction," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 5, pp. 1066–1078, 2021.

[69] X. Tang and W. Tang, "An ECG Delineation and Arrhythmia Classification System Using Slope Variation Measurement by Ternary Second-Order Delta Modulators for Wearable ECG Sensors," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 5, pp. 1053–1065, 2021.

[70] C. J. Deepu, X. Y. Xu, D. L. T. Wong, C. H. Heng, and Y. Lian, "A 2.3μW ECG-On-Chip for Wireless Wearable Sensors," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 10, pp. 1385–1389, 2018.

[71] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230–236, 1985.

[72] A. for the Advancement of Medical lnstrumentation, "EC57:2012 Testing and Reporting Performance Results of Cardiac Rhythm and ST Segment Measurement Algorithms," 2012.

[73] PhysioNet, "PhysioBank Annotations," 2016. [Online]. Available: https://archive.physionet.org/physiobank/annotations.shtml

[74] P. de Chazal, M. O'Dwyer, and R. Reilly, "Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.

[75] G. Moody and R. Mark, "The Impact of the MIT-BIH Arrhythmia Database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[76] T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, and R. Poll, "Optimization of ECG Classification by Means of Feature Selection," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 8, pp. 2168–2177, 2011.

[77] D. Geebelen, J. A. K. Suykens, and J. Vandewalle, "Reducing the Number of Support Vectors of SVM Classifiers Using the Smoothed Separable Case Approximation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 23, no. 4, pp. 682–688, 2012.

[78] S. Höppner, H. Eisenreich, D. Walter, U. Steeb, A. S. C. Dmello, R. Sinkwitz, H. Bauer, A. Oefelein, F. Schraut, J. Schreiter, R. Niebsch, S. Scherzer, M. Orgis, U. Hensel, and J. Winkler, "How to Achieve World-Leading Energy Efficiency Using 22FDX with Adaptive Body Biasing on an ARM Cortex-M4 IoT SoC," *ESSDERC 2019 49th European Solid-State Device Research Conference (ESSDERC)*, pp. 66–69, 2019.

[79] R. G. Carvajal, J. Ramírez-Angulo, A. J. López-Martín, A. Torralba, J. A. G. Galán, A. Carlosena, and F. M. Chavero, "The Flipped Voltage Follower: a Useful Cell for Low-Voltage Low-Power Circuit Design," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 7, pp. 1276–1291, 2005.

[80] J. Kim, T.-K. Jang, Y.-G. Yoon, and S. Cho, "Analysis and Design of Voltage-Controlled Oscillator Based Analog-to-Digital Converter," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 1, pp. 18–30, 2010.

[81] C. Pochet, J. Huang, P. Mercier, and D. Hall, "A 174.7-dB FoM, 2^{nd}-order VCO-based ExG-to-Digital Front-End Using a Multi-phase Gated-Inverted Ring Oscillator Quantizer," *IEEE Transactions on Biomedical Circuits and Systems*, pp. 1–1, 2021.

[82] M. Schramme, L. Van Brandt, D. Flandre, and D. Bol, "Comprehensive Analytical Comparison of Ring Oscillators in FDSOI Technology: Current Starving Versus Back-Bias Control," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 69, no. 5, pp. 1883–1895, 2022.

[83] A. Cathelin, "Fully Depleted Silicon on Insulator Devices CMOS: the 28-nm Node Is the Perfect Technology for Analog, RF, mmW, and Mixed-Signal System-on-Chip Integration," *IEEE Solid-State Circuits Magazine*, vol. 9, no. 4, pp. 18–26, 2017.

[84] M. Baert and W. Dehaene, "A 5-GS/s 7.2-ENOB Time-Interleaved VCO-Based ADC Achieving 30.5 fJ/cs," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 6, pp. 1577–1587, 2020.

[85] K. Park, B. Cho, D. Lee, S. Song, J. Lee, Y. Chee, I. Kim, and S. Kim, "Hierarchical Support Vector Machine Based Heartbeat Classification Using Higher Order Statistics and Hermite Basis Function," *2008 Computers in Cardiology*, pp. 229–232, 2008.

[86] *Apollo3 Blue MCU - Ultra-low Power Apollo MCU Family*, Ambiq Micro, 2022, Datasheet.

[87] J. Myers, A. Savanth, P. Prabhat, S. Yang, R. Gaddh, S. O. Toh, and D. Flynn, "A 12.4 pJ/cycle Sub-Threshold, 16pJ/cycle Near-Threshold ARM Cortex-M0+ MCU with Autonomous SRPG/DVFS and Temperature Tracking Clocks," *2017 Symposium on VLSI Circuits*, pp. C332–C333, 2017.

[88] P. Prabhat, B. Labbe, G. Knight, A. Savanth, J. Svedas, M. J. Walker, S. Jeloka, P. M.-Y. Fan, F. Garcia-Redondo, T. Achuthan *et al.*, "27.2 M0N0: a Performance-Regulated 0.8-to-38MHz DVFS ARM Cortex-M33 SIMD MCU with 10nW Sleep Power," *2020 IEEE International Solid-State Circuits Conference (ISSCC)*, pp. 422–424, 2020.

[89] G. Lallement, F. Abouzeid, J.-M. Daveau, P. Roche, and J.-L. Autran, "A 1.1-pJ/cycle, 20-MHz, 0.42-V Temperature Compensated ARM Cortex-M0+ SoC with Adaptive Self Body-Biasing in FD-SOI," *IEEE Solid-State Circuits Letters*, vol. 1, no. 7, pp. 174–177, 2018.

[90] J. Lee, Y. Zhang, Q. Dong, W. Lim, M. Saligane, Y. Kim, S. Jeong, J. Lim, M. Yasuda, S. Miyoshi *et al.*, "A Self-Tuning IoT Processor Using Leakage-Ratio Measurement for Energy-Optimal Operation," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 1, pp. 87–97, 2020.

[91] E. H. T. Shad, M. Molinas, and T. Ytterdal, "Impedance and Noise of Passive and Active Dry EEG Electrodes: a Review," *IEEE Sensors Journal*, vol. 20, no. 24, pp. 14 565–14 577, 2020.

[92] W. Yang, H. Jiang, Y. Yin, and Z. Wang, "A 4-μW Analog Front End Achieving 2.4 NEF for Long-Term ECG Monitoring," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 15, no. 4, pp. 655–665, 2021.

[93] Y.-P. Hsu, Z. Liu, and M. M. Hella, "A 12.3-μW 0.72-mm^2 Fully Integrated Front-End IC for Arterial Pulse Waveform and ExG Recording," *IEEE Journal of Solid-State Circuits*, vol. 55, no. 10, pp. 2756–2770, 2020.

[94] J. Xu, M. Konijnenburg, H. Ha, R. van Wegberg, S. Song, D. Blanco-Almazán, C. Van Hoof, and N. Van Helleputte, "A 36 μW 1.1 mm^2 Reconfigurable Analog Front-End for Cardiovascular and Respiratory Signals Recording," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 4, pp. 774–783, 2018.

[95] L. Zeng, B. Liu, and C.-H. Heng, "A Dual-Loop Eight-Channel ECG Recording System With Fast Settling Mode for 12-Lead Applications," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 7, pp. 1895–1906, 2019.

[96] Y. Park, J.-H. Cha, S.-H. Han, J.-H. Park, and S.-J. Kim, "A 3.8-μW 1.5-NEF 15-GΩ Total Input Impedance Chopper Stabilized Amplifier With Auto-Calibrated Dual Positive Feedback in 110-nm CMOS," *IEEE Journal of Solid-State Circuits*, 2022.

[97] A. Villa, M. Deviaene, R. Willems, S. Van Huffel, and C. Varon, "Are We Training our Heartbeat Classification Algorithms Properly?" *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6363–6366, 2019.

[98] M. Chen, G. Wang, Z. Ding, J. Li, and H. Yang, "Unsupervised Domain Adaptation for ECG Arrhythmia Classification," *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 304–307, 2020.

[99] V. Mondéjar-Guerra, J. Novo, J. Rouco, M. G. Penedo, and M. Ortega, "Heartbeat Classification Fusing Temporal and Morphological Information of ECGs via Ensemble of Classifiers," *Biomedical Signal Processing and Control*, vol. 47, pp. 41–48, 2019.

[100] L. Guo, G. Sim, and B. Matuszewski, "Inter-Patient ECG Classification with Convolutional and Recurrent Neural Networks," *Biocybernetics and Biomedical Engineering*, vol. 39, no. 3, pp. 868–879, 2019.

[101] N. Larburu, T. Lopetegi, and I. Romero, "Comparative Study of Algorithms for Atrial Fibrillation Detection," *2011 Computing in Cardiology*, pp. 265–268, 2011.

[102] T. Pirson, T. Delhaye, A. Pip, G. Le Brun, J.-P. Raskin, and D. Bol, "The Environmental Footprint of IC Production: Meta-analysis and Historical Trends," *2022 52nd IEEE European Solid-State Device Research Conference (ESSDERC)*, 2022.

[103] S. Prakash, R. Liu, K. Schischke, L. Stobbe, and C. Gensch, "Schaffungeiner Datenbasis zur Ermittlung ökologischer Wirkungen der Produkte der

Informations-und Kommunikationstechnik (IKT)," 2013. [Online]. Available: https://www.umweltbundesamt.de/sites/default/files/medien/378/publikationen/texte_82_2013_janssen_informationstechnik_teil_c.pdf

[104] R. J. Baker, *CMOS: Circuit Design, Layout, and Simulation*, 3rd ed. John Wiley & Sons, 2010.

[105] H. Y. Koh, C. H. Sequin, and P. R. Gray, "OPASYN: a Compiler for CMOS Operational Amplifiers," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 9, no. 2, pp. 113–125, 1990.

[106] P. G. Jespers and B. Murmann, *Systematic Design of Analog CMOS Circuits*. Cambridge University Press, 2017.

[107] M. Ding, P. Harpe, G. Chen, B. Busze, Y.-H. Liu, C. Bachmann, K. Philips, and A. van Roermund, "A Hybrid Design Automation Tool for SAR ADCs in IoT," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 26, no. 12, pp. 2853–2862, 2018.

[108] Z. Gao, J. Tao, F. Yang, Y. Su, D. Zhou, and X. Zeng, "Efficient Performance Trade-Off Modeling for Analog Circuit Based on Bayesian Neural Network," *2019 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 1–8, 2019.

[109] G. G. Gielen, H. C. Walscharts, and W. M. Sansen, "Analog Circuit Design Optimization Based on Symbolic Simulation and Simulated Annealing," *IEEE Journal of Solid-State Circuits*, vol. 25, no. 3, pp. 707–713, 1990.

[110] R. Mina, C. Jabbour, and G. E. Sakr, "A Review of Machine Learning Techniques in Analog Integrated Circuit Design Automation," *Electronics*, vol. 11, no. 3, p. 435, 2022.

[111] G. Stehr, H. Graeb, and K. Antreich, "Performance Trade-Off Analysis of Analog Circuits by Normal-Boundary Intersection," *2003 40th Design Automation Conference (DAC)*, pp. 958–963, 2003.

[112] D. Mueller-Gritschneder, H. Graeb, and U. Schlichtmann, "A Successive Approach to Compute the Bounded Pareto Front of Practical Multiobjective Optimization Problems," *SIAM Journal on Optimization*, vol. 20, no. 2, pp. 915–934, 2009.

[113] S. Mallick, R. Kar, D. Mandal, and S. P. Ghoshal, "Optimal Sizing of CMOS Analog Circuits Using Gravitational Search Algorithm with Particle Swarm Optimization," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 1, pp. 309–331, 2017.

[114] Q. Zhang and H. Li, "MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition," *IEEE Transactions on Evolutionary Computation*, vol. 11, no. 6, pp. 712–731, 2007.

[115] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-Objective Optimization: NSGA-II," *International Conference on Parallel Problem Solving from Nature*, pp. 849–858, 2000.

[116] W. Lyu, F. Yang, C. Yan, D. Zhou, and X. Zeng, "Multi-Objective Bayesian Optimization for Analog/RF Circuit Synthesis," *2018 55th Annual Design Automation Conference (DAC)*, pp. 1–6, 2018.

[117] G. de Streel, D. Flandre, C. Dehollain, and D. Bol, "Towards Ultra-Low-Voltage Wideband Noise-Cancelling LNAs in 28nm FDSOI," *2015 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, pp. 1–2, 2015.

[118] D. Bol, C. Hocquet, and F. Regazzoni, "A Fast ULV Logic Synthesis Flow in Many-V_t CMOS Processes for Minimum Energy Under Timing Constraints," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 59, no. 12, pp. 947–951, 2012.

[119] P. Christ, G. Sievers, J. Einhaus, T. Jungeblut, M. Porrmann, and U. Rückert, "Pareto-Optimal Signal Processing on Low-Power Microprocessors," *SENSORS, 2013 IEEE*, pp. 1–4, 2013.

[120] X. Fafoutis, L. Marchegiani, A. Elsts, J. Pope, R. Piechocki, and I. Craddock, "Extending the Battery Lifetime of Wearable Sensors with Embedded Machine Learning," *2018 IEEE 4th World Forum on Internet of Things (WF-IoT)*, pp. 269–274, 2018.

[121] U. Jensen, P. Kugler, M. Ring, and B. M. Eskofier, "Approaching the Accuracy–Cost Conflict in Embedded Classification System Design," *Pattern Analysis and Applications*, vol. 19, no. 3, pp. 839–855, 2016.

[122] V. Konstantakos, A. Chatzigeorgiou, S. Nikolaidis, and T. Laopoulos, "Energy Consumption Estimation in Embedded Systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 4, pp. 797–804, 2008.

[123] S. Reda, A. N. Nowroz *et al.*, "Power Modeling and Characterization of Computing Devices," *Foundations and Trends in Electronic Design Automation*, vol. 6, no. 2, pp. 121–216, 2012.

[124] H. Sultan, G. Ananthanarayanan, and S. R. Sarangi, "Processor Power Estimation Techniques: a Survey," *International Journal of High Performance Systems Architecture*, vol. 5, no. 2, pp. 93–114, 2014.

[125] A. Muttreja, A. Raghunathan, S. Ravi, and N. K. Jha, "Automated Energy/Performance Macromodeling of Embedded Software," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 26, no. 3, pp. 542–552, 2007.

[126] O. B. Kchaou, A. Garbaya, M. Kotti, P. Pereira, M. Fakhfakh, and M. H. Fino, "Sensitivity Aware NSGA-II Based Pareto Front Generation for the Optimal Sizing of Analog Circuits," *Integration*, vol. 55, pp. 220–226, 2016.

[127] A. Sanabria-Borbón and E. Tlelo-Cuautle, "Sizing Analogue Integrated Circuits by Integer Encoding and NSGA-II," *IETE Technical Review*, vol. 35, no. 3, pp. 237–243, 2018.

[128] A. Sasikumar and R. Muthaiah, "Operational Amplifier Circuit Sizing Based on NSGA-II and Particle Swarm Optimization," *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)*, pp. 64–68, 2017.

[129] G. Nicosia, S. Rinaudo, and E. Sciacca, "An Evolutionary Algorithm-Based Approach to Robust Analog Circuit Design Using Constrained Multi-Objective Optimization," *Knowledge-Based Systems*, vol. 21, pp. 175–183, 2008.

[130] O. Cuate, O. Schuetze, F. Grasso, and E. Tlelo-Cuautle, "Sizing CMOS Operational Transconductance Amplifiers Applying NSGA-II and MOEAD," *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 149–152, 2019.

[131] M. Takhti, A. Beirami, and H. Shamsi, "Multi-Objective Design Automation of the Folded-Cascode OP-AMP Using NSGA-II Strategy," *2009 International Symposium on Signals, Circuits and Systems (ISSCS)*, pp. 1–4, 2009.

[132] J. López-Arredondo, E. Tlelo-Cuautle, and F. V. Fernández, "Optimization of LDO Voltage Regulators by NSGA-II," *2016 13th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, pp. 1–4, 2016.

[133] F. Stas, G. de Streel, and D. Bol, "Sizing and Layout Integrated Optimizer for 28nm Analog Circuits Using Digital PnR Tools," *2016 14th IEEE International New Circuits and Systems Conference (NEWCAS)*, 2016.

[134] S. E. Sorkhabi and L. Zhang, "Automated Topology Synthesis of Analog and RF Integrated Circuits: a Survey," *Integration*, vol. 56, pp. 128–138, 2017.

[135] N. Srinivas and K. Deb, "Multiobjective Optimization using Nondominated Sorting in Genetic Algorithms," *Evolutionary Computation*, vol. 2, no. 3, pp. 221–248, 1994.

[136] L. Marti, J. Garcia, A. Berlanga, and J. M. Molina, "A Stopping Criterion for Multi-Objective Optimization Evolutionary Algorithms," *Information Science*, vol. 367, pp. 700 – 718, 2016.

[137] F. Silveira, D. Flandre, and P. G. Jespers, "A g_m/I_D-based Methodology for the Design of CMOS Analog Circuits and its Application to the Synthesis of a Silicon-On-Insulator Micropower OTA," *IEEE Journal of Solid-Ctate Circuits*, vol. 31, no. 9, pp. 1314–1319, 1996.

[138] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, and V. G. da Fonseca, "Performance Assessment of Multiobjective Optimizers: an Analysis and Review," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 2, pp. 117–132, 2003.

[139] S. Song, M. Rooijakkers, P. Harpe, C. Rabotti, M. Mischi, A. H. van Roermund, and E. Cantatore, "A Low-Voltage Chopper-Stabilized Amplifier for Fetal ECG Monitoring with a 1.41 Power Efficiency Factor," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 9, no. 2, pp. 237–247, 2015.

[140] F. M. Yaul and A. P. Chandrakasan, "A Noise-Efficient 36 nV/$\sqrt{Hz}$ Chopper Amplifier Using an Inverter-Based 0.2-V Supply Input Stage," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 11, pp. 3032–3042, 2017.

[141] scikit-learn, "Support Vector Classification (SVC)," 2022. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

[142] S. Pandit, C. Mandal, and A. Patra, *Nano-Scale CMOS Analog Circuits: Models and CAD Techniques for High-Level Design*. CRC Press, 2018.

[143] H. Chang, A. Sangiovanlli-Vincentelli, F. Balarin, E. Charbon, U. Choudhury, G. Jusuf, E. Liu, E. Malavasi, R. Neff, and P. R. Gray, "A Top-Down, Constraint-Driven Design Methodology for Analog Integrated Circuits," *1992 IEEE Custom Integrated Circuits Conference (CICC)*, pp. 8–4, 1992.

[144] Y. Fellah, N. Abouchi, T. Tixier, A. Aubert, and L. Labrak, "Top-Down, Constraint Driven Design: Application to Sigma-Delta Modulator," *2004 IEEE International Conference on Industrial Technology (ICIT)*, vol. 1, pp. 132–136, 2004.

[145] S. Dam and P. Mandal, "Modeling and Design of CMOS Analog Circuits through Hierarchical Abstraction," *Integration*, vol. 46, no. 4, pp. 449–462, 2013.

[146] B. G. Arsintescu, E. Charbon, E. Malavasi, U. Choudhury, and W. H. Kao, "General AC Constraint Transformation for Analog ICs," *1998 35th Design and Automation Conference (DAC)*, pp. 38–43, 1998.

[147] R. Phelps, M. J. Krasnicki, R. A. Rutenbar, L. R. Carley, and J. R. Hellums, "A Case Study of Synthesis for Industrial-Scale Analog IP: Redesign of the Equalizer/Filter Frontend for an ADSL CODEC," *2000 37th Annual Design Automation Conference (DAC)*, pp. 1–6, 2000.

[148] G. Yu and P. Li, "Hierarchical Analog/Mixed-Signal Circuit Optimization under Process Variations and Tuning," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 2, pp. 313–317, 2011.

[149] J. Zou, D. Mueller, H. Graeb, and U. Schlichtmann, "A CPPLL Hierarchical Optimization Methodology Considering Jitter, Power and Locking Time," *2006 43rd Annual Design Automation Conference (DAC)*, pp. 19–24, 2006.

[150] G. Venter, *Encyclopedia of Aerospace Engineering*. AIAA, December 2010, ch. Review of Optimization Techniques.

[151] X.-S. Yang, *Engineering Optimization: an Introduction with Metaheuristic Applications*. John Wiley & Sons, 2010.

[152] SciPy, "Minimize – SLSQP," 2022. [Online]. Available: https://docs.scipy.org/doc/scipy/reference/optimize.minimize-slsqp.html

[153] H. Oh, C. Rizo, M. Enkin, and A. Jadad, "What is eHealth (3): a Systematic Review of Published Definitions," *Journal of Medical Internet Research*, vol. 7, no. 1, p. e110, 2005.

[154] B. M. Silva, J. J. Rodrigues, I. de la Torre Díez, M. López-Coronado, and K. Saleem, "Mobile-Health: a Review of Current State in 2015," *Journal of Biomedical Informatics*, vol. 56, pp. 265–272, 2015.

[155] T. Moerenhout, I. Devisch, and G. C. Cornelis, "E-health Beyond Technology: Analyzing the Paradigm Shift that Lies Beneath," *Medicine, Health Care and Philosophy*, vol. 21, no. 1, pp. 31–41, 2018.

[156] D. Malvey and D. J. Slovensky, *mHealth: Transforming Healthcare*. Springer, 2014.

[157] T. J. Betjeman, S. E. Soghoian, and M. P. Foran, "mHealth in Sub-Saharan Africa," *International Journal of Telemedicine and Applications*, vol. 2013, 2013.

[158] E. Osei and T. P. Mashamba-Thompson, "Mobile Health Applications for Disease Screening and Treatment Support in Low-and Middle-Income Countries: a Narrative Review," *Heliyon*, vol. 7, no. 3, p. e06639, 2021.

[159] L. V. Lapão, G. Dussault, World Health Organization *et al.*, "The Contribution of eHealth and mHealth to Improving the Performance of the Health Workforce: a Review," *Public Health Panorama*, vol. 3, no. 03, pp. 463–471, 2017.

[160] F. P. Tajudeen, N. Bahar, T. Maw Pin, and N. I. Saedon, "Mobile Technologies and Healthy Ageing: a Bibliometric Analysis on Publication Trends and Knowledge Structure of mHealth Research for Older Adults," *International Journal of Human–Computer Interaction*, vol. 38, no. 2, pp. 118–130, 2022.

[161] S. Hamine, E. Gerth-Guyette, D. Faulx, B. B. Green, A. S. Ginsburg *et al.*, "Impact of mhealth chronic disease management on treatment adherence and patient outcomes: a systematic review," *Journal of medical Internet research*, vol. 17, no. 2, p. e3951, 2015.

[162] M. S. Marcolino, J. A. Q. Oliveira, M. D'Agostino, A. L. Ribeiro, M. B. M. Alkmim, and D. Novillo-Ortiz, "The Impact of mHealth Interventions: Systematic Review of Systematic Reviews," *JMIR mHealth and uHealth*, vol. 6, no. 1, p. e8873, 2018.

[163] S. J. Iribarren, K. Cato, L. Falzon, and P. W. Stone, "What is the Economic Evidence for mHealth? A Systematic Review of Economic Evaluations of mHealth Solutions," *PloS one*, vol. 12, no. 2, p. e0170581, 2017.

[164] B. Purvis, Y. Mao, and D. Robinson, "Three Pillars of Sustainability: in Search of Conceptual Origins," *Sustainability Science*, vol. 14, no. 3, pp. 681–695, 2019.

[165] L. Marrauld, T. Rambaud, M. Sarfati, M. Egnell, E. Proto, H. Lesimple, and J.-N. Geist, "Décarboner la santé pour soigner durablement dans le cadre du Plan de transformation de l'économie fraçaise," The Shift Project, Tech. Rep., November 2021.

[166] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow *et al.*, "The PRISMA 2020 Statement: an Updated Guideline for Reporting Systematic Reviews," *British Medical Journal*, vol. 372, no. 71, 2021.

[167] National Health Service, "Delivering a 'Net Zero' National Health Service," 2020.

[168] A. D. Lantada, R. Rodríguez-Rivero, A. M. Romero, R. B. García, L. I. B. Sánchez, L. Di Pietro, C. De Maria, and A. Ahluwalia, "On the Sustainable Growth of the Biomedical Industry Reinvented Through Innovative Open-Source Medical Devices," in *Engineering Open-Source Medical Devices*. Springer, 2022.

[169] D. Agrawal and Z. Tang, "Sustainability of Single-Use Endoscopes," *Techniques and Innovations in Gastrointestinal Endoscopy*, vol. 23, no. 4, pp. 353–362, 2021.

[170] S. R. Unger and A. E. Landis, "Comparative Life Cycle Assessment of Reused Versus Disposable Dental Burs," *The International Journal of Life Cycle Assessment*, vol. 19, no. 9, pp. 1623–1631, 2014.

[171] H. Stripple, R. Westman, and D. Holm, "Development and Environmental Improvements of Plastics for Hydrophilic Catheters in Medical Care: an Environmental Evaluation," *Journal of Cleaner Production*, vol. 16, no. 16, pp. 1764–1776, 2008.

[172] ISO 14044:2006, "Environmental management – Life cycle assessment – Requirements and guidelines," International Organization for Standardization, Geneva (CH), Standard, 2006.

[173] S. Barbero, A. Pereno, and P. Tamborrini, "Systemic Innovation in Sustainable Design of Medical Devices," *The Design Journal*, vol. 20, no. sup1, pp. S2486–S2497, 2017.

[174] A. C. Sousa, A. Veiga, A. C. Maurício, M. A. Lopes, J. D. Santos, and B. Neto, "Assessment of the Environmental Impacts of Medical Devices: a Review," *Environment, Development and Sustainability*, vol. 23, no. 7, pp. 9641–9666, 2021.

[175] J. D. Sherman, L. A. Raibley IV, and M. J. Eckelman, "Life Cycle Assessment and Costing Methods for Device Procurement: Comparing Reusable and Single-Use Disposable Laryngoscopes," *Anesthesia & Analgesia*, vol. 127, no. 2, pp. 434–443, 2018.

[176] C. Viani, M. Vaccari, and T. Tudor, "Recovering Value from Used Medical Instruments: a Case Study of Laryngoscopes in England and Italy," *Resources, Conservation and Recycling*, vol. 111, pp. 1–9, 2016.

[177] J. J. Hanson and R. W. Hitchcock, "Towards Sustainable Design for Single-Use Medical Devices," *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5602–5605, 2009.

[178] S. R. Unger, T. A. Hottle, S. R. Hobbs, C. L. Thiel, N. Campion, M. M. Bilec, and A. E. Landis, "Do Single-Use Medical Devices Containing Biopolymers Reduce the Environmental Impacts of Surgical Procedures Compared with their Plastic Equivalents?" *Journal of Health Services Research & Policy*, vol. 22, no. 4, pp. 218–225, 2017.

[179] F. McGain, S. McAlister, A. McGavin, and D. Story, "A Life Cycle Assessment of Reusable and Single-Use Central Venous Catheter Insertion Kits," *Anesthesia & Analgesia*, vol. 114, no. 5, pp. 1073–1080, 2012.

[180] S. Unger and A. Landis, "Assessing the Environmental, Human Health, and Economic Impacts of Reprocessed Medical Devices in a Phoenix Hospital's Supply Chain," *Journal of Cleaner Production*, vol. 112, pp. 1995–2003, 2016.

[181] R. Lankey, H. MacLean, and A. Sterdis, "A Case Study in Environmentally Conscious Design: Wearable Computers," *1997 IEEE International Symposium on Electronics and the Environment. ISEE-1997*, pp. 204–209, 1997.

[182] A. R. Köhler, L. M. Hilty, and C. Bakker, "Prospective Impacts of Electronic Textiles on Recycling and Disposal," *Journal of Industrial Ecology*, vol. 15, no. 4, pp. 496–511, 2011.

[183] N. M. Van der Velden, K. Kuusk, and A. R. Köhler, "Life Cycle Assessment and Eco-Design of Smart Textiles: the Importance of Material Selection Demonstrated Through E-Textile Product Redesign," *Materials & Design*, vol. 84, pp. 313–324, 2015.

[184] H. Abdi, N. Mohajer, and S. Nahavandi, "Human Passive Motions and a User-Friendly Energy Harvesting System," *Journal of Intelligent Material Systems and Structures*, vol. 25, no. 8, pp. 923–936, 2014.

[185] L. Liu, X. Guo, and C. Lee, "Promoting Smart Cities into the 5G Era with Multi-Field Internet of Things (IoT) Applications Powered with Advanced Mechanical Energy Harvesters," *Nano Energy*, vol. 88, p. 106304, 2021.

[186] M. Mahmud and S. H. Farjana, "Comparative Eco-Profiles of Polyethylene Terephthalate (PET) and Polymethyl Methacrylate (PMMA) Using Life Cycle

Assessment," *Journal of Polymers and the Environment*, vol. 29, no. 2, pp. 418–428, 2021.

[187] H. Wu, Y. Zhang, A.-L. Kjøniksen, X. Zhou, and X. Zhou, "Wearable Biofuel Cells: Advances from Fabrication to Application," *Advanced Functional Materials*, vol. 31, no. 48, p. 2103976, 2021.

[188] L. Manjakkal, L. Yin, A. Nathan, J. Wang, and R. Dahiya, "Energy Autonomous Sweat-Based Wearable Systems," *Advanced Materials*, p. 2100899, 2021.

[189] R. Mukherjee and R. Dahiya, "Life Cycle Assessment of Energy Generating Flexible Electronic Skin," *2021 IEEE International Conference on Flexible and Printable Sensors and Systems (FLEPS)*, pp. 1–4, 2021.

[190] A. Veronica and I.-m. Hsing, "An Insight into Tunable Innate Piezoelectricity of Silk for Green Bioelectronics," *ChemPhysChem*, vol. 22, no. 22, pp. 2266–2280, 2021.

[191] S. Strassburg, S. Zainuddin, and T. Scheibel, "The Power of Silk Technology for Energy Applications," *Advanced Energy Materials*, vol. 11, no. 43, p. 2100519, 2021.

[192] X. Liu, T. Fu, J. Ward, H. Gao, B. Yin, T. Woodard, D. R. Lovley, and J. Yao, "Multifunctional Protein Nanowire Humidity Sensors for Green Wearable Electronics," *Advanced Electronic Materials*, vol. 6, no. 9, p. 2000721, 2020.

[193] M. Cocchi, M. Bertoldo, M. Seri, P. Maccagnani, C. Summonte, S. Buoso, G. Belletti, F. Dinelli, and R. Capelli, "Fully Recyclable OLEDs Built on a Flexible Biopolymer Substrate," *ACS Sustainable Chemistry & Engineering*, vol. 9, no. 38, pp. 12 733–12 737, 2021.

[194] J.-Y. Lam, G.-W. Jang, C.-J. Huang, S.-H. Tung, and W.-C. Chen, "Environmentally Friendly Resistive Switching Memory Devices with DNA as the Active Layer and Bio-Based Polyethylene Furanoate as the Substrate," *ACS Sustainable Chemistry & Engineering*, vol. 8, no. 13, pp. 5100–5106, 2020.

[195] S. Yu, Z. Liu, L. Zhao, and L. Li, "Degradable, Ultra-Flexible, Transparent and Conductive Film Made of Assembling CuNWs on Chitosan," *Optical Materials*, vol. 123, p. 111752, 2022.

[196] S. Wang, J. Y. Oh, J. Xu, H. Tran, and Z. Bao, "Skin-Inspired Electronics: an Emerging Paradigm," *Accounts of Chemical Research*, vol. 51, no. 5, pp. 1033–1045, 2018.

[197] B. Yang, C. Yao, Y. Yu, Z. Li, and X. Wang, "Nature Degradable, Flexible, and Transparent Conductive Substrates from Green and Earth-Abundant Materials," *Scientific Reports*, vol. 7, no. 1, pp. 1–8, 2017.

[198] H. Dong, J. Li, J. Guo, F. Lai, F. Zhao, Y. Jiao, D. J. Brett, T. Liu, G. He, and I. P. Parkin, "Insights on Flexible Zinc-Ion Batteries from Lab Research to Commercialization," *Advanced Materials*, vol. 33, no. 20, p. 2007548, 2021.

[199] M. T. Rahman, A. Rahimi, S. Gupta, and R. Panat, "Microscale Additive Manufacturing and Modeling of Interdigitated Capacitive Touch Sensors," *Sensors and Actuators A: Physical*, vol. 248, pp. 94–103, 2016.

[200] S. Blenkinsop, A. Foley, N. Schneider, J. Willis, H. J. Fowler, and S. M. Sisodiya, "Carbon Emission Savings and Short-Term Health Care Impacts from Telemedicine: an Evaluation in Epilepsy," *Epilepsia*, vol. 62, no. 11, pp. 2732–2740, 2021.

[201] Å. Holmner, K. L. Ebi, L. Lazuardi, and M. Nilsson, "Carbon Footprint of Telemedicine Solutions - Unexplored Opportunity for Reducing Carbon Emissions in the Health Sctor," *PLoS One*, vol. 9, no. 9, p. e105040, 2014.

[202] J. Whetten, J. Montoya, and H. Yonas, "ACCESS to Better Health and Clear Skies: Telemedicine and Greenhouse Gas Reduction," *Telemedicine and e-Health*, vol. 25, no. 10, pp. 960–965, 2019.

[203] C. Masino, E. Rubinstein, L. Lem, B. Purdy, and P. G. Rossos, "The Impact of Telemedicine on Greenhouse Gas Emissions at an Academic Health Science Center in Canada," *Telemedicine and e-Health*, vol. 16, no. 9, pp. 973–976, 2010.

[204] E. Y. Wang, J. E. Zafar, C. M. Lawrence, L. F. Gavin, S. Mishra, A. Boateng, C. L. Thiel, R. Dubrow, and J. D. Sherman, "Environmental Emissions Reduction of a Preoperative Evaluation Center Utilizing Telehealth Screening and Standardized Preoperative Testing Guidelines," *Resources, Conservation and Recycling*, vol. 171, p. 105652, 2021.

[205] H. Audebert, T. Meyer, and F. Klostermann, "Potentials of Telemedicine for Green Health Care," *Frontiers in Neurology*, vol. 1, p. 10, 2010.

[206] P. M. Yellowlees, K. Chorba, M. Burke Parish, H. Wynn-Jones, and N. Nafiz, "Telemedicine Can Make Healthcare Greener," *Telemedicine and e-Health*, vol. 16, no. 2, pp. 229–232, 2010.

[207] S. Bartlett and S. Keir, "Calculating the Carbon Footprint of a Geriatric Medicine Clinic Before and After COVID-19," *Age and Ageing*, vol. 51, no. 2, pp. 1–4, 2022.

[208] M. A. Edison, M. J. Connor, S. Miah, T. El-Husseiny, M. Winkler, R. Dasgupta, H. U. Ahmed, and D. Hrouda, "Understanding Virtual Urology Clinics: a Systematic Review," *BJU International*, vol. 126, no. 5, pp. 536–546, 2020.

[209] J. Moultrie, L. Sutcliffe, and A. Maier, "A Maturity Grid Assessment Tool for Environmentally Conscious Design in the Medical Device Industry," *Journal of Cleaner Production*, vol. 122, pp. 252–265, 2016.

[210] S. R. Unger, N. Campion, M. M. Bilec, and A. E. Landis, "Evaluating Quantifiable Metrics for Hospital Green Checklists," *Journal of Cleaner Production*, vol. 127, pp. 134–142, 2016.

[211] T. Mak and D. Kritchanchai, "Factors Investigation for the Development of Medical Device towards Sustainable Product," *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 2019.

[212] S. Hede, M. J. L. Nunes, P. F. V. Ferreira, and L. A. Rocha, "Incorporating Sustainability in Decision-Making for Medical Device Development," *Technology in Society*, vol. 35, no. 4, pp. 276–293, 2013.

[213] D. Guzzo, M. Carvalho, R. Balkenende, and J. Mascarenhas, "Circular Business Models in the Medical Device Industry: Paths Towards Sustainable Healthcare," *Resources, Conservation and Recycling*, vol. 160, p. 104904, 2020.

[214] M. Fargnoli, F. Costantino, G. Di Gravio, and M. Tronci, "Product Service-Systems Implementation: a Customized Framework to Enhance Sustainability and Customer Satisfaction," *Journal of Cleaner Production*, vol. 188, pp. 387–401, 2018.

[215] N. Haber and M. Fargnoli, "Sustainable Product-Service Systems Customization: a Case Study Research in the Medical Equipment Sector," *Sustainability*, vol. 13, no. 12, p. 6624, 2021.

[216] G. I. Zlamparet, Q. Tan, A. Stevels, and J. Li, "Resource Conservation Approached with an Appropriate Collection and Upgrade-Remanufacturing for Used Electronic Products," *Waste Management*, vol. 73, pp. 78–86, 2018.

[217] United Nations Framework Convention on Climate Change, "The Paris Agreement," 2016.

[218] J. D. Sherman, C. Thiel, A. MacNeill, M. J. Eckelman, R. Dubrow, H. Hopf, R. Lagasse, J. Bialowitz, A. Costello, M. Forbes *et al.*, "The Green Print: Advancement of Environmental Sustainability in Healthcare," *Resources, Conservation and Recycling*, vol. 161, p. 104882, 2020.

[219] I. Tennison, S. Roschnik, B. Ashby, R. Boyd, I. Hamilton, T. Oreszczyn, A. Owen, M. Romanello, P. Ruyssevelt, J. D. Sherman *et al.*, "Health Care's Response to Climate Change: a Carbon Footprint Assessment of the NHS in England," *The Lancet Planetary Health*, vol. 5, no. 2, pp. e84–e92, 2021.

[220] J. Moultrie, L. Sutcliffe, and A. Maier, "Exploratory Study of the State of Environmentally Conscious Design in the Medical Device Industry," *Journal of Cleaner Production*, vol. 108, pp. 363–376, 2015.

[221] A. Ordway, J. S. Pitonyak, and K. L. Johnson, "Durable Medical Equipment Reuse and Recycling: Uncovering Hidden Opportunities for Reducing Medical Waste," *Disability and Rehabilitation: Assistive Technology*, 2018.

[222] J. Cosgrove, J. Littlewood, and P. Wilgeroth, "Development of a Framework of Key Performance Indicators to Identify Reductions in Energy Consumption in a Medical Devices Production Facility," *International Journal of Ambient Energy*, vol. 39, no. 2, pp. 202–210, 2018.

[223] F. Alshqaqeeq, M. A. Esmaeili, M. Overcash, and J. Twomey, "Quantifying Hospital Services by Carbon Footprint: a Systematic Literature Review of Patient Care Alternatives," *Resources, Conservation and Recycling*, vol. 154, p. 104560, 2020.

[224] F. McGain, N. R. Cox, S. S. Cecchin, S. McAlister, and P. R. Barach, "Sustainable Cardiac Services - From the Catheterization Laboratory to the Operating Room and Beyond," *Progress in Pediatric Cardiology*, vol. 33, no. 1, pp. 81–84, 2012.

[225] G. Rinaldi, A. Hijazi, and H. Haghparast-Bidgoli, "Cost and Cost-Effectiveness of mHealth Interventions for the Prevention and Control of Type 2 Diabetes Mellitus: a Systematic Review," *Diabetes Research and Clinical Practice*, vol. 162, p. 108084, 2020.

[226] C. Fernández, C. Maturana, S. Coloma, A. Carrasco-Labra, and R. Giacaman, "Teledentistry and mHealth for Promotion and Prevention of Oral Health: a Systematic Review and Meta-Analysis," *Journal of Dental Research*, vol. 100, no. 9, pp. 914–927, 2021.

[227] K. Jin, S. Khonsari, R. Gallagher, P. Gallagher, A. M. Clark, B. Freedman, T. Briffa, A. Bauman, J. Redfern, and L. Neubeck, "Telehealth Interventions for the Secondary Prevention of Coronary Heart Disease: a Systematic Review and Meta-analysis," *European Journal of Cardiovascular Nursing*, vol. 18, no. 4, pp. 260–271, 2019.

[228] J. Braithwaite, P. Glasziou, and J. Westbrook, "The Three Numbers You Need to Know About Healthcare: the 60-30-10 Challenge," *BMC Medicine*, vol. 18, no. 1, pp. 1–8, 2020.

[229] A. J. Nowak, "Artificial Intelligence in Evidence-Based Medicine: On the Verge of Breakthrough," in *Artificial Intelligence in Medicine*. Springer, 2022, pp. 255–266.

[230] A. Behne and F. Teuteberg, "A Healthy Lifestyle and the Adverse Impact of its Digitalization: The Dark Side of Using eHealth Technologies," *Wirtschaftsinformatik (Zentrale Tracks)*, pp. 584–599, 2020.

[231] D. A. Baker, "Four Ironies of Self-Quantification: Wearable Technologies and the Quantified Self," *Science and Engineering Ethics*, vol. 26, no. 3, pp. 1477–1498, 2020.

[232] R. Jain and J. Wullert, "Challenges: Environmental Design for Pervasive Computing Systems," *2002 8th Annual International Conference on Mobile Computing and Networking*, pp. 263–270, 2002.

[233] L. M. Hilty, C. Som, and A. Köhler, "Assessing the Human, Social, and Environmental Risks of Pervasive Computing," *Human and Ecological Risk Assessment*, vol. 10, no. 5, pp. 853–874, 2004.

[234] R. Cole, C. F. Lindsay, and F. Barker, "Reverse Exchange of Healthcare Devices: the Case of Hearing Aid Equipment in the UK," *Production Planning & Control*, vol. 29, no. 13, pp. 1045–1057, 2018.

[235] M.-H. Tu, P. Chang, and Y.-L. Lee, "Avoiding Obsolescence in Mobile Health: Experiences in Designing a Mobile Support System for Complicated Documentation at Long-Rerm Care Facilities," *CIN: Computers, Informatics, Nursing*, vol. 36, no. 10, pp. 501–506, 2018.

[236] F. Alanezi and T. Alanzi, "A Gig mHealth Economy Framework: Scoping Review of Internet Publications," *JMIR mHealth and uHealth*, vol. 8, no. 1, p. e14213, 2020.

[237] E. D. V. Cerqueira, B. Motte-Baumvol, L. B. Chevallier, and O. Bonin, "Does Working From Home Reduce CO_2 Emissions? An Analysis of Travel Patterns as Dictated by Workplaces," *Transportation Research Part D: Transport and Environment*, vol. 83, p. 102338, 2020.

[238] J. Bieser and L. Hilty, "Indirect Effects of the Digital Transformation on Environmental Sustainability: Methodological Challenges in Assessing the Greenhouse Gas Abatement Potential of ICT," *EPiC Series in Computing*, no. 52, pp. 68–81, 2018.

[239] K. Cornaggia, X. Li, and Z. Ye, "Virtual Competition and Cost of Capital: Evidence from Telehealth," *Available at SSRN 3833193*, 2022.

[240] A. André, *Digital Medicine*. Springer, 2018.

[241] V. Vo, L. Auroy, A. Sarradon-Eck *et al.*, "Patients' Perceptions of mHealth Apps: Meta-Ethnographic Review of Qualitative Studies," *JMIR mHealth and uHealth*, vol. 7, no. 7, p. e13817, 2019.

[242] Elsevier Health, "Clinician of the Future," March 2022.

[243] A. F. Shao, C. Rambaud-Althaus, N. Swai, J. Kahama-Maro, B. Genton, V. D'Acremont, and C. Pfeiffer, "Can Smartphones and Tablets Improve the Management of Childhood Illness in Tanzania? A Qualitative Study from a Primary Health Care Worker's Perspective," *BMC health services research*, vol. 15, no. 1, pp. 1–12, 2015.

[244] D. Bol, T. Pirson, and R. Dekimpe, "Moore's Law and ICT Innovation in the Anthropocene," *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 19–24, 2021.

[245] K. Schittkowski, "A Robust Implementation of a Sequential Quadratic Programming Algorithm with Successive Error Restoration," *Optimization Letters*, vol. 5, no. 2, pp. 283–296, 2011.

[246] A. Mirkouei, B. Silwal, and L. Ramiscal, "Enhancing Economic and Environmental Sustainability Benefits Across the Design and Manufacturing of Medical Devices: a Case Study of Ankle Foot Orthosis," *International Design*

Engineering Technical Conferences and Computers and Information in Engineering Conference, vol. 58165, pp. 1–9, 2017.

A

mHealth Environmental Sustainability Literature Review

A.1 List of articles

Table A.1 lists all the articles selected with the PRISMA methodology in Section 5.2. The papers are sorted according to the topic clusters described in Section 5.2.3. A short description and the themes (scope, level, and indicators) associated to each item are provided.

Table A.1 List of articles

Reference	Description	Scope	Level	Indicator
Medical device assessment				
– Hospital devices				
Barbero, 2017 [173]	Sustainability analysis of haemodialysis equipment based on systemic design	Dev (Hosp)	Dir (Use, EoL)	Waste; Pollution (ET, RP)
Stripple, 2008 [171]	Environmental evaluation of plastic materials for catheters using LCA	Dev (Hosp)	Dir (Prod, Use, EoL)	Pollution (OD, P, HT, ET); Energy; Carbon footprint; Resources (Raw)
Sousa, 2021 [174]	Review of LCA and eco-design studies of medical devices	Dev (Hosp)	Dir (Prod, Use, EoL)	Waste (Rec); Pollution (HT, ET); Energy; Carbon footprint; Resources (Raw, Water)
– Single-use or reusable devices				
Agrawal, 2021 [169]	Review of environmental, economic and social sustainability of single-use endoscopes	Dev (Hosp, SU)	Dir (Prod, Use, EoL)	Waste (Rec); Pollution (PM, HT, ET); Carbon footprint
Hanson, 2009 [177]	LCA and functional analysis of a dialyzer cartridge	Dev (Hosp, SU)	Dir (Prod, Use, EoL)	Waste (Rec, Reu, Bio); Pollution (ET); Energy; Resources (Raw)
Unger, 2014 [170]	LCA comparison of dental burs	Dev (Hosp, SU)	Dir (Prod, Use, EoL)	Waste (Rec, Reu); Pollution (OD, HT, ET); Energy; Carbon footprint; Resources (Water)
Unger, 2017 [178]	LCA of plastics and biopolymers for surgery equipment	Dev (Hosp, SU)	Dir (Prod, Use, EoL)	Waste (Bio); Pollution (OD, HT, ET); Energy; Carbon footprint; Resources (Raw)
Unger, 2016 [180]	LCA and cost assessment of reprocessed medical devices	Dev (Hosp, SU)	Dir (Prod, Use, EoL)	Waste; Pollution (HT); Carbon footprint
Sherman, 2018 [175]	LCA and cost assessment of laryngoscopes	Dev (Hosp, SU)	Dir (Prod, Use, EoL)	Waste (Rec, Ref, Reu); Pollution (OD, PM, HT, ET); Energy; Carbon footprint; Resources (Water)
McGain, 2012 [179]	LCA of venous catheter insertion kits	Dev (Hosp, SU)	Dir (Prod, Use, EoL)	Waste (Rec); Pollution (ET); Energy; Carbon footprint; Resources (Raw, Water)
Viani, 2016 [176]	End of life value recovery from used laryngoscopes	Dev (Hosp, SU)	Dir (EoL)	Waste (Rec, Reu); Energy
– Wearable devices				
Lankey, 1997 [181]	Eco-design of a wearable computer	Dev (Wear)	Dir (Prod, Use, EoL)	Waste (Rec, Ref, Reu); Pollution (ET); Energy; Resources (Raw)

Table A.1 List of articles (continued)

Reference	Description	Scope	Level	Indicator
Kokare, 2020 [28]	LCA of wearable cardiac monitoring devices	Dev (Wear)	Dir (Prod, Use, EoL)	Waste (Rec); Pollution (OD, PM, HT, ET, RP); Energy; Carbon footprint; Resources (Raw, Water)
Ma, 2017 [59]	LCA of a smartwatch	Dev (Wear)	Dir (Prod, EoL)	Waste; Pollution (OD, PM, HT, ET, RP); Energy; Carbon footprint; Resources (Raw, Water)
– E-textiles				
Gurova, 2020 [60]	Review of environmental impacts of e-textiles from the fashion and ICT standpoints	Dev (Wear)	Dir (Prod, EoL); Ind (Obs)	Waste (Rec, Ref, Reu, Bio); Pollution (ET); Energy; Carbon footprint; Resources (Raw, Water)
Köhler, 2011 [182]	Impacts of e-textiles on device end of life	Dev (Wear)	Dir (EoL); Ind (Obs)	Waste (Rec); Pollution (ET); Resources (Raw)
van der Velden, 2015 [183]	LCA of an e-textile garment	Dev (Wear)	Dir (Prod, Use, EoL)	Waste (Rec); Energy; Carbon footprint; Resources
– Other				
Mirkouei, 2017 [246]	Eco-design of an ankle foot orthosis	Dev	Dir (Prod)	Energy; Carbon footprint; Resources (Raw)
Component design				
– Energy harvesting				
Abdi, 2014 [184]	Design of a piezoelectric energy harvester from abdominal motion	Comp (EH)	Dir (Use)	Energy
Liu, 2021 [185]	Review of mechanical energy harvesters	Comp (EH)	Dir (Use)	Energy
Mahmud, 2021 [186]	LCA of PET and PMMA used in triboelectric energy harvesters of	Comp (EH)	Dir (Prod, EoL)	Waste; Pollution (OD, PM, HT, ET, RP); Energy; Carbon footprint; Resources (Raw, Water)
Wu, 2021 [187]	Review of wearable biofuel cells to harvest energy from bodily fluids	Comp (EH)	Dir (EoL)	Waste (Bio); Energy
Manjakkal, 2021 [188]	Review of sweat-based energy harvesting and storage	Comp (EH, Sens)	Dir (Use, EoL)	Waste; Energy
Mukherjee, 2021 [189]	Evaluation of environmental impacts of electronic skin with photovoltaic energy harvesting	Comp (EH)	Dir (Use, EoL)	Waste (Rec, Ref, Bio); Pollution (ET)

Table A.1 List of articles (continued)

Reference	Description	Scope	Level	Indicator
– New material				
Cocchi, 2021 [193]	Design of organic LEDs and flexible biopolymer substrate	Comp (Disp, Flex)	Dir (Prod, EoL)	Waste (Rec, Bio); Resources (Raw)
Dong, 2021 [198]	Review of aqueous rechargeable zinc-ion batteries	Comp (Supp)	Dir (Prod, EoL)	Waste (Rec); Energy
Lam, 2020 [194]	Design of a DNA-based memory device	Comp (Mem)	Dir (Prod, EoL)	Waste (Bio); Resources (Raw)
Liu, 2020 [192]	Design of a humidity sensor with protein nanowires	Comp (Sens)	Dir (Prod, EoL)	Waste; Energy; Resources (Raw)
Yu, 2022 [195]	Design of a biodegradable transparent conductive film	Comp (Flex)	Dir (Prod, EoL)	Waste (Rec); Resources (Raw)
Yang, 2017 [197]	Design of a biodegradable transparent conductive film with sustainable materials	Comp (Flex)	Dir (Prod, EoL)	Waste (Rec, Bio); Resources (Raw)
Wang, 2018 [196]	Review of skin-inspired electronic substrates	Comp (Flex)	Dir (Use, EoL)	Waste (Ref, Bio)
Veronica, 2021 [190]	Review of silk-based piezoelectric devices	Comp (EH, Sens, Act)	Dir (Prod, Use, EoL)	Waste (Bio); Energy; Resources (Raw)
Strassburg, 2021 [191]	Review of silk-based electronic devices	Comp (EH, Sens)	Dir (Prod, Use, EoL)	Waste (Bio); Carbon footprint; Resources (Raw)
Rahman, 2016 [199]	Design of capacitive touch sensors with additive manufacturing	Comp (Sens)	Dir (Prod)	Waste; Pollution (HT)
Telemedicine				
Audebert, 2010 [205]	Qualitative environmental impact analysis of telemedicine in neurology	Syst (TM)	Ind (Travel, Opti)	Energy; Carbon footprint
Yellowlees, 2010 [206]	Qualitative environmental impact analysis of telemedicine and eHealth in the US	Syst (TM, eH)	Dir (Use, EoL); Ind (Travel)	Waste (Rec, Reu); Pollution (HT); Energy; Carbon footprint; Resources (Raw)
Bartlett, 2022 [207]	Carbon footprint assessment of virtual geriatry consultations	Dev (Hosp); Syst (TM)	Dir (Prod,Use, EoL); Ind (Travel)	Carbon footprint
Blenkinsop, 2021 [200]	Carbon footprint assessment of virtual consultations in epilepsy services	Syst (TM)	Ind (Travel)	Waste; Pollution (ET); Carbon footprint; Resources (Water)
Edison, 2020 [208]	Review of telemedicine in urology	Syst (TM)	Ind (Travel)	Carbon footprint
Holmner, 2014 [201]	Carbon footprint assessment of virtual consultations in different specialized services	Syst (TM)	Dir (Prod,Use, EoL); Ind (Travel)	Carbon footprint
Whetten, 2019 [202]	Carbon footprint assessment of virtual consultations in neurology	Syst (TM)	Dir (Use); Ind (Travel)	Carbon footprint

Table A.1 List of articles (continued)

Reference	Description	Scope	Level	Indicator
Wang, 2021 [204]	Carbon footprint assessment of telehealth for preoperative screening	Syst (TM)	Dir (Use); Ind (Travel, Opti)	Carbon footprint
Purohit, 2021 [58]	Review of carbon footprint assessment in telemedicine applications	Syst (TM)	Dir (Use); Ind (Travel, Reb)	Carbon footprint
Masino, 2010 [203]	Carbon footprint assessment of virtual consultations in different specialized services	Syst (TM)	Dir (Use); Ind (Travel)	Carbon footprint
Design methodologies				
Hede, 2013 [212]	Implementation of a multicriteria hierarchical decision model for medical device design	Dev	Dir (Prod, Use, EoL)	Waste (Rec, Ref, Reu); Pollution; Resources (Raw)
Mak, 2019 [211]	Analysis of the factors influencing medical device design using analytical hierarchy process	Dev	Dir (Prod, Use, EoL)	Waste (Rec, Ref, Reu); Pollution (OD, HT, ET); Energy; Carbon footprint
Unger, 2016 [210]	Evaluation of green practices checklists in hospitals	Dev (Hosp); Syst	Dir (Prod, Use, EoL)	Waste (Reu); Energy; Carbon footprint; Resources (Water)
Moultrie, 2016 [209]	Implementation of a qualitative assessment grid for environmental impacts of medical devices	Dev	Dir (Prod, Use, EoL)	Waste (Rec, Ref, Reu); Pollution (HT, ET); Energy; Carbon footprint; Resources (Raw)
Business models				
Fargnoli, 2018 [214]	Implementation of a product-service system business model for haemodialysis equipment	Dev (Hosp)	Dir (Prod, Use, EoL)	Waste (Ref, Reu)
Haber, 2021 [215]	Implementation of a product-service system business model for molecular diagnosis equipment	Dev (Hosp)	Dir (EoL)	Waste (Rec, Ref, Reu)
Guzzo, 2020 [213]	Review of circular business models in the medical device industry	Dev (Hosp)	Dir (Use, EoL)	Waste (Rec, Ref, Reu); Pollution (ET); Energy; Resources
Zlamparet, 2018 [216]	Evaluation of upgrade and remanufacturing business model for used electronic products	Dev (Hosp)	Dir (EoL)	Waste (Rec, Reu); Energy; Carbon footprint; Resources (Raw)
Ahluwalia, 2022 [168]	Environmental impact assessment of open-source medical devices	Dev (Hosp)	Dir (Prod, Use, EoL)	Waste (Rec, Ref, Reu); Carbon footprint; Resources (Raw)
Top-down healthcare system assessment				
Sherman, 2020 [218]	Review of environmental sustainability research in the healthcare sector	Dev (Hosp); Syst	Dir (Prod, Use, EoL)	Waste (Rec, Ref, Reu); Pollution (HT, ET); Carbon footprint

Table A.1 List of articles (continued)

Reference	Description	Scope	Level	Indicator
NHS, 2020 [167]	Decarbonization propositions for the English National Health Service	Syst	Dir (Prod, Use, EoL); Ind (Travel)	Waste (Ref, Reu); Energy; Carbon footprint
Shift Project, 2021 [165]	Decarbonization propositions for the French healthcare sector	Syst	Dir (Prod, Use, EoL); Ind (Travel, Obs, Reb)	Waste (Rec, Reu); Energy; Carbon footprint; Resources (Water)
Di Giacomo, 2011 [57]	Assessment of carbon emission reduction in eHealth applications	Syst (TM, eH)	Dir (Prod, Use); Ind (Travel)	Waste (Rec, Reu); Pollution (ET); Carbon footprint
Holmner, 2012 [54]	Review of climate change mitigation and adaptation eHealth strategies in the healthcare sector	Syst (eH)	Dir (Prod, Use, EoL)	Energy, Carbon footprint; Resources
Stakeholders environmental considerations				
Ordway, 2018 [221]	Interviews of healthcare staff regarding opportunities for reuse and recycling of medical equipment	Dev	Dir (EoL)	Energy; Carbon footprint
Moultrie, 2015 [220]	Interviews of medical device designers regarding eco-design	Dev	Dir (Prod, Use, EoL)	Waste; Pollution (HT, ET)
Other				
Cosgrove, 2018 [222]	Evaluation of energy savings in medical device production facilities	Dev	Dir (Prod)	Energy
Alshqaqeeq, 2020 [223]	Review of GHG emission reduction alternatives for medical services and equipments	Dev (Hosp); Syst (TM)	Dir (Prod, Use, EoL); Ind (Travel)	Carbon footprint
McGain, 2012 [224]	Analysis of environmental sustainability improvement opportunities in paediatric cath lab and operating room	Dev (Hosp);	Dir (Prod, Use, EoL)	Waste (Rec, Reu); Pollution (HT); Energy; Carbon footprint; Resources (Raw, Water)

Dev: Device - Hosp: Hospital - Wear: Wearable - Comp: Component - EH: Energy harvesting - Sens: Sensor - Disp: Display - Flex: Flexible substrate - Supp: Power supply - Mem: Memories - Act: Actuator - Syst: System - TM: Telemedicine - eH: eHealth

Dir: Direct impacts - Prod: Production - Use: Use phase - EoL: End of life - Ind: Indirect impacts - Travel: Travel reduction - Opti: Care optimization - Obs: Obsolescence - Reb: Rebound effects

Rec: Recycling - Ref: Refurbishing - Reu: Reusing - Bio: Biodegradability - OD: Ozone depletion - PM: Particulate matter - HT: Human toxicity - ET: Ecotoxicity - RP: Radioactive pollution - Raw: Raw material