Title
Data-Semantics Driven Power Optimization of Wireless Medical Monitoring Devices

Author
Goudar, Vishwatma Arun

Publication Date
2013

Peer reviewed|Thesis/dissertation

University of California
Los Angeles

Data-Semantics Driven Power Optimization of Wireless Medical Monitoring Devices

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Computer Science

by

Vishwatma Arun Goudar

2013

Abstract of the Dissertation

Data-Semantics Driven Power Optimization of Wireless Medical Monitoring Devices

by

Vishwatma Arun Goudar

Doctor of Philosophy in Computer Science

University of California, Los Angeles, 2013

Professor Miodrag Potkonjak, Chair

We study techniques to improve deployment of remote wireless health monitoring Body Area Networks (BANs) that offer the promise of improved quality and availability of healthcare. Ease-of-use, safety and security have been sighted as key problems to be resolved before the scale and duration of their deployment can be extended. We focus on 4 performance factors that influence these issues: (i) Semantic Fidelity - The diagnostic quality of data measured by BANs is critical to system safety; (ii) Low-Power Operation - Reducing power consumed by BANs will result in light-weight systems with small form-factors and extended lifetimes, owing to reduced battery discharge rates and the use of smaller batteries; (iii) Fault-Tolerant Operation - Faulty sensor measurements can severely impact diagnostic quality of the data produced by the system, thereby impacting system safety; (iv) Data Sharing Security - Sharing large BAN collected medical datasets exacerbates problems related to secure control over medical data as it is shared between data owners and consumers.

Here, we study low-power sampling techniques for multi-sensor remote health monitoring BANs, where medical diagnostic metric accuracy is maximized while the number of samples taken simultaneously is constrained. Power usage is reduced both due to the decrease in number of sensors that are active at a time, and due to the reduced bandwidth required to transmit the measured signals. The techniques are extended to admit fault tolerance and fault detection capabilities while maintaining low-power operation and high semantic-fidelity

requirements. We study low-power sampling for the continuous monitoring and anomaly detection use-cases.

The use of a novel energy harvester technology in scavenging energy from foot-strikes and powering BANs in a self-sustaining manner is also studied. Several optimizations are considered to maximize the energy output of a proposed foot-strike energy harvester platform based on this novel harvester, while adhering to user-comfort requirements.

Finally, we study a robust watermarking technique to address secure data control requirements related to sharing BAN generated datasets. Three use-cases are addressed - proof-of-ownership, data tracking and content authentication. It is shown that robust and imperceptible watermarking of biomedical signals is achievable without compromising semantic fidelity of the datasets.

The dissertation of Vishwatma Arun Goudar is approved.

Qibing Pei

Jens Palsberg

Mario Gerla

Miodrag Potkonjak, Committee Chair

University of California, Los Angeles

2013

To Noya, my pillar, my beacon.

Then seek not, sweet, the "If" and "Why"

I love you now until I die.

For I must love because I live

And life in me is what you give.

- *Christopher Brennan*

Table of Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to express sincere gratitude to Dr. Miodrag Potkonjak for his generous advice, support and mentorship through all my years as a graduate student at UCLA. I have tremendously benefitted from observing and modeling his intellectual curiosity and nuanced thought, as well as, from his carrot-and-stick approach that I believe has gotten me battle ready for the trials and tribulations that will come as part of a research career.

I would also like to thank my committee members for the helpful discussions and advice regarding my dissertation research. Specifically, I would like to thank Dr. Mario Gerla for his support and encouragement. Dr. Gerla and Dr. Medy Sanadidi were especially supportive during my initial graduate research assignments, and their enthusiastic and unrestrained advice during that period is sincerely appreciated. I also owe Dr. William Kaiser a debt of gratitude for believing in me as an undergraduate researcher and gently nudging me to consider graduate school and a career in research.

My dissertation research would not have been possible without support from the labs of Dr. Majid Sarrafzadeh and Dr. Qibing Pei. Dr. Sarrafzadeh and his student Dr. Hyduke Noshadi have been tremendously generous in sharing several HERMES datasets with me and I am deeply grateful for it. Dr. Pei and his students Dr. Paul Brochu and Zhi Ren have granted me countless hours of their time and resources as I sought to understand and characterize Dielectric Elastomer behavior, and for this I am extremely thankful. I have also benefitted from the discussions and collaboration with my labmate James B. Wendt and I thank him for being my partner in crime.

I would also like to thank Dr. Dean Buonomano and Dr. Maneesh Varshney for instilling in me self-confidence in my research abilities through my final years in graduate school. As nurturing mentors, they were willing to be extremely flexible, disregard oversights and highlight small victories, which has helped me find my voice as a researcher.

Last but not least, I would like to thank my Mother, Father and Brother for their blind faith and support as I sought out a career in research. They have rooted for me since as far back as I can remember, and have stood by me through my ups and downs. For this I am

eternally grateful to them. My wife, Noya, and daughter, Adi, have borne the brunt of my long hours; Yet, they managed to instill pure joy and faith in me every moment I've spent with them. I thank them for being by my side.

Vita

2003–2004	Undergraduate Researcher, Electrical Engineering Department, University of California, Los Angeles.
2004	B.S. (Computer Science and Engineering), University of California, Los Angeles.
2004–2005	Software Engineer, Scalable Network Technologies, Los Angeles.
2005	M.S. (Computer Science), University of California, Los Angeles.
2005–2007	Product Manager, Vin Systems Technologies, Dubai, U.A.E.
2007	Teaching Assistant, Computer Science Department, University of California, Los Angeles.
2008	Research Assistant, Computer Science Department University of California, Los Angeles.
2008–present	Software Engineer, Scalable Network Technologies, Los Angeles.

Publications

M. Potkonjak and V. Goudar. *Public Physical Unclonable Functions.* Proceedings of the IEEE. 2014. *To appear.*

V. Goudar, Z. Ren, P. Brochu, M. Potkonjak and Q. Pei. *Optimizing the Output of a Human-Powered Energy Harvesting System with Miniaturization and Integrated Control.* IEEE Sensors Journal. 2014. *To appear.*

V. Goudar and M. Potkonjak. *A Semantically-Adaptive Strategy for Energy- Efficiency in Wireless Medical Monitoring Devices.* IEEE Sensors. November 2013. *To appear.*

V. Goudar, Z. Ren, P. Brochu, Q. Pei and M. Potkonjak. *Optimizing the Configuration and Control of a Novel Human-Powered Energy Harvesting System.* International Workshop on Power and Timing Modeling, Optimization and Simulation. pp. 75-82. September 2013. [Best Paper Award]

V. Goudar, Z. Ren, P. Brochu, M. Potkonjak and Q. Pei. *Driving Low-Power Wearable Systems with an Adaptively-Controlled Foot-Strike Scavenging Platform.* International Symposium on Wearable Computers. pp. 135-136. September 2013.

M. Varshney and V. Goudar. *Blue: A Unified Programming Model for Diverse Data-Intensive Cloud Computing Paradigms.* Technical Report No.130005, Computer Science Department, UCLA. April 2013.

M. Varshney and V. Goudar. *PDQCollections: A Data-Parallel Programming Model and Library for Associative Containers.* Technical Report No.130004, Computer Science Department, UCLA. April 2013.

V. Goudar and M. Potkonjak. *Fault-Tolerant and Low-Power Sampling Schedules for Localized BANs.* IEEE Journal on Emerging and Selected Topics in Circuits and Systems, Vol. 3, No. 1. pp. 86-95. March 2013.

V. Goudar and M. Potkonjak. *Spatiotemporal Assignment of Energy Harvesters on a Self-Sustaining Medical Shoe.* IEEE Sensors. pp. 1312-1315. October 2012.

J.B. Wendt, V. Goudar, H. Noshadi and M. Potkonjak. *Energy-Efficient Sampling Schedules for Body Area Networks.* IEEE Sensors. pp. 1525-1528. October 2012.

V. Goudar and M. Potkonjak. *Dielectric Elastomer Generators for Foot Plantar Pressure Based Energy Scavenging.* IEEE Sensors. pp. 1001-1004. October 2012.

V. Goudar and M. Potkonjak. *Power Constrained Sensor Sample Selection for Improved Form Factor and Lifetime in Localized BANs.* ACM Wireless Health. Article 5. October 2012.

V. Goudar, M. Cohen, M.Y. Sanadidi, M. Gerla and F. Zampognaro. *Towards Handoff Tolerance in TCP: Link Layer Handoff Detection for Optimized Data Transport.* IEEE Military Communications Conference. pp. 1-7. October 2009.

D.W. Browne, V. Goudar, H. Borgstrom, M.P. Fritz and W. Kaiser. *Antenna Actuation for Radio Telemetry in Remote Sensor Networks.* IEEE Vehicular Technology Conference. pp. 3115-3119. Fall 2004.

CHAPTER 1

Introduction

Major advances in Wireless Sensor Network (WSN) performance and application support services have wielded a host of scientific and engineering applications with robust, flexible and scalable remote monitoring and control capabilities. While early applications enlisted them to static deployment over large areas, recent applications are a product of novel deployment configurations that open WSNs up to new application domains.

For example, they are being retooled as smaller networks of wearable sensors and devices known as Body Area Networks (BANs), with applications in the medical, sport, military and security domains. Comprised of networks/arrays of physiological and/or behavioral sensors, BANs can collect detailed data about the user's body and behavior, and do so continuously. Specifically, we focus on its application to mobile patient monitoring, which may be defined as the continuous or periodic measurement and analysis of a mobile patient's biosignals from a distance by employing mobile computing, wireless communications and networking technologies [PJB12].

The detailed and diagnostically rich data collected by BANs, when used in conjunction with data from patient interviews, on-site observations and tests, can enhance diagnostic accuracy, shorten time to detection, and propel preventative healthcare. They will also provide medical researchers access to an abundance of high quality data so they may better understand the causes and symptoms of illnesses. As they mature, BANs are expected to enjoy widespread deployment owing to ability to automatically collect, process and share patient medical diagnostic metrics with authorized patient and research databases, and other data sinks in a minimally invasive and unsupervised manner.

Therefore, they are expected to have enormous impacts on the accessibility, efficiency and

quality of care for convalescing patients, geriatric patients, and those with chronic conditions through computer-assisted rehabilitation and continuous health monitoring. Improved access to healthcare by way of advances in mobile health and increased efficiency of care due to the elimination of significant barriers in tele-healthcare are strongly anticipated as well. A recent survey of post-trial assessments of such systems demonstrated their feasibility, and has underscored their potential to reduce time-to-treatment and the frequency and duration of patient hospitalization [PJB12]. Moreover, multi-sensory mobile health monitoring systems have been proposed for a variety of medical applications, including emotional health monitoring [DGS], respiratory monitoring [ESK11], general wellbeing monitoring [JBD11], etc.

1.1 BAN Performance Requirements

The authors of [HPB09] describe several challenges to the development and deployment of BANs in a variety of application domains. Deployment challenges include value to the user, ease-of-use, safety and security, among others. While the value of remote medical monitoring systems is self-evident, the other requirements are crucial to system acceptance and adoption. We now discuss the set of technical challenges in BAN design that we focus on in this work, and assess their relevance to BAN deployment.

Semantic Fidelity

Foremost among the requirements for mobile health monitoring systems is the need to provide high-quality measurements to a medical practitioner. Here, data quality isn't based on the level of detail with which the physiological signal can be spatio-temporally reconstructed from the measured samples. Rather, it is based on the system's ability to continually and accurately estimate medical diagnostic metrics and the signal features, or metrics, that they depend upon. While pertinent functional tools and tests will inform the selection of those medical diagnostic metrics, in the context of mobile health monitoring these diagnostic metrics will be estimated from signal features/metrics as sampled by multiple sensors located at key locations across the body,

possibly in different modalities (temperature, pressure, GSR, etc.). Design and performance optimization of BANs developed for remote health monitoring must, therefore, achieve their goals without compromising the fidelity of diagnostic information generated. Barring this, they will be in violation of crucial *safety* requirements. In this work, we propose BAN performance optimizations that are cognizant of the semantic fidelity of the data collected by the system.

Low-Power Operation

The authors of [HPB09] discussed several technical challenges inherent to BAN design. However, power-consumption is consistently described as a primary concern from the perspective of system usability, lifetime and maintainability, all important factors to the system's *ease-of-use*. BANs are often comprised of localized sensor arrays, powered by a single battery. While the power demands of continuously sampling an array of sensors, processing the measured physiological signals, and, transmitting extracted signal features can be substantial, these systems must be able to function unattended for extended periods.

One solution is to increasing battery capacity. While research on improving battery capacities in ongoing, the rate of this improvement has been marginal [PS05]. Consequently, increasing battery capacity implies using larger batteries which interferes with the requirement that remote health monitoring systems be light-weight and maintain a small form factor, lest they hinder/disrupt user activity. The alternative is to increase the frequency of system maintenance, which clearly affects ease-of-use. Evidently, power consumption affects system ease-of-use in more than one way, making low-power operation critical to adoption.

In this work we will study 2 complementary techniques to alleviate the power demands of remote health monitoring BANs. Reducing the battery discharge rate will leads to an increase in the battery capacity [RVR03], and hence its lifetime, thereby improving system ease-of-use. One technique will reduce power consumption by limiting the number of sensors that are operated at any time. This limits the amount of en-

ergy expended in sampling at each epoch. An added benefit of collecting less data in this manner is that less energy is expended in transmitting it to the data sink. Towards this end, several optimizations have been described in BAN literature, including sub-sampling [MKA11], sensor coverage [ZRF12], data compression and duty cycling techniques. However, in the context of health monitoring, the need for semantic fidelity imposes a strong constraint on the extent and the way in which low-power operation is achieved. We will explore a few health monitoring use-cases to study different techniques to achieve low-power operation, which are cognizant of the semantic fidelity of the sparsely collected data.

Unobtrusive energy-harvesting from various sources of human power has also been gaining traction as a means to alleviating net power consumption and even realizing self-sustaining sensors and BANs. Partly, this is due to advances in low-power design and operation, and partly it is due to advances in energy harvesting technologies. Whereas some ambient energy harvesters, such as solar or human-driven electro-magnetic harvester platforms, can yield energy on the order of magnitude that is required by most BAN applications, these introduce additional cost and usability issues, their dependence on sources of non-uniform availability notwithstanding. Instead, we argue that to a large extent, remote health monitoring is required during periods of human activity, precisely those times when energy is produced by the human body. For example, it is conceivable that sleep quality monitoring devices [HS10] may be driven by energy harvested from subject movement or breathing. Similarly, in the context of applications related to human locomotion, energy expended by human foot strikes may be scavenged. In this work, we study the use of a leading-edge energy harvester technology to scavenge the energy produced by a user's foot strikes, and drive a low-power target wearable system that involves human locomotion in a self-sustaining manner.

Fault Tolerance/Detection

Sensor faults can play a crucial role in lowering a system's semantic fidelity. Faulty sensor readings may go so far as to result in misdiagnosis or the failure to catch patient symptoms in time. Therefore, fault detection / tolerance is a crucial requirement

for system *safety* [HPB09]. Furthermore, under low-power operation, the accuracy of medical metric estimates is much more sensitive to faults. As discussed, power-efficient signal processing techniques commonly resort to taking the fewest number of sensor samples that are required to achieve acceptable levels of semantic fidelity. This is usually achieved when the information shared between such samples (mutual information) is low, thereby boosting the net information used to arrive at the medical diagnostic metric(s). Unfortunately, when the information redundancy is low, faulty samples can cause much larger errors in diagnostic metric estimates. We will discuss techniques to detect and tolerate sensor faults with the objective of minimally increasing system power consumption in the process.

Data Sharing Security

One of the most important infrastructure requirements in the domain of remote health monitoring BANs is the secure collection and dissemination of medical data of the user [HPB09]. Data security desiderata in this application domain are not limited to ensuring the confidentiality and integrity of medical data that has been logged to a data sink. Requirements also arise from the need to provide the data owner (BAN user / patient) and the data consumers (healthcare providers, insurance companies, medical research facilities) secure control over the data as it is shared between these various stakeholders. We will study a robust watermarking technique to embed security information into biosignal data such that the semantic fidelity of the data is unaffected and the watermark is not easily erased by malicious data consumers. In doing so, we address three uses-cases: proof of ownership, wherein the data owner can prove that she/his is the originator of the data; data tracking, wherein the data owner can trace unauthorized sharing of her/his biosignal data; and content authentication, wherein the data owner can prove whether the biosignal data has been maliciously altered.

	96	97	98	99		
90	91	92	93	94	95	
83	84	85	86	87	88	89
76	77	78	79	80	81	82
69	70	71	72	73	74	75
62	63	64	65	66	67	68
55	56	57	58	59	60	61
48	49	50	51	52	53	54
41	42	43	44	45	46	47
34	35	36	37	38	39	40
27	28	29	30	31	32	33
20	21	22	23	24	25	26
13	14	15	16	17	18	19
6	7	8	9	10	11	12
	1	2	3	4	5	

(a) (b) (c)

Figure 1.1: (a) The HERMES smart shoe platform [NAH10] (left), (b) The Pedar plantar mapping [Ped07] (middle), and, (c) the corresponding sensor numbering system (right).

1.2 Target Application: HERMES

While many of the techniques we shall discuss can be generalized to BANs applied to remote health monitoring, we describe and study these techniques in the context of a human balance monitoring system known as HERMES [NAH10], a wireless health monitoring smart shoe aimed at extending human instability analysis outside of the lab environment. Further, we validate our techniques based on experimental datasets collected by HERMES.

HERMES is composed of a multi-sensory array of 99 passive-resistive pressure sensors located across the shoe insole, that are used to wirelessly monitor foot plantar pressure as the subject ambulates. HERMES's plantar pressure sensors are purposefully placed at locations specified by the Pedar plantar mapping [Ped07]. Fig. 1.1 shows the HERMES platform alongside this mapping. Several functional diagnostic metrics from the GARS-M and Tinetti gait stability assessment scales [AM99] can be derived from HERMES's plantar pressure measurements, and these define the semantics of the data collected by the system.

The HERMES system is built on the MicroLEAP platform [AWB07], a wearable sensing platform for biomedical applications. It provides an 8-channel 16-bit Analog-to-Digital Converter (ADC), 8Mbits of flash memory, an MSP430 micro-controller and a Class-2 Bluetooth

radio module. Although the MicroLEAP ADC only provide 8-channels for HERMES's 99 pressure sensors, the ADC supports a sampling throughput of 100ksps and is multiplexed to achieve a 50Hz sampling rate for each HERMES sensor. With each sensor sampled at 50Hz with a 16-bit ADC, HERMES generates data at about 81kbps, including bluetooth packet overhead.

Table 1.1: Power Consumption by HERMES

Component	Power (mW)
Processor	2.69
Radio	56.49
Flash Memory	0.03
Sensors, MEMS	1.18
Sensors, Pressure	100.24
16-bit ADC	5.97
Total	*166.6*

While communication is the most demanding operation in WSNs, HERMES belongs to the class of BANs whose power draw is driven by its sensor sampling rate. Here, a node is comprised of multi-sensory array and simultaneously powers significantly more sensors than a WSN node. Further, as we have discussed, collecting data at a higher rate translates to transmitting data at a higher rate, and this increases power drawn by the radio unit. Table 1.1 breaks down the power consumption of the HERMES platform based on the requirement of its pressure sensors and that of the MicroLEAP components. The data for the MicroLEAP components were made available in [AWB07], and we have derived the sampling power draw for HERMES's plantar pressure sensors from the maximum circuit voltage and force resistance curve for the underlying $Flexiforce^{®}$ sensor [fle87]. We note that the HERMES platform is also capable of supporting an accelerometer and a gyroscope. However, since our application focus is related to plantar pressure measurements and related

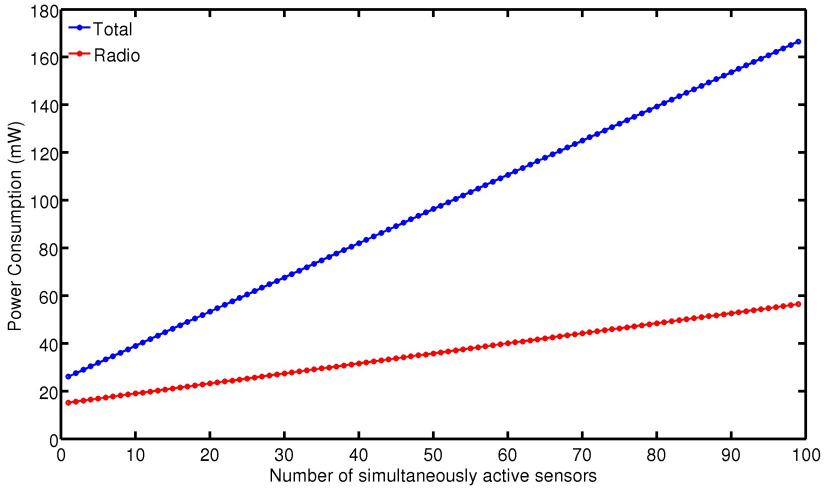


Figure 1.2: HERMES power consumption as the number of simultaneously sampled sensors is varied.

medical diagnostic metrics, we exclude these from our discussion.

Limiting the number of sensors active at any time results in a proportional decrease in power consumption by the sensor array. Further, while the MicroLEAP platform's bluetooth radio is capable of transmitting at 115.2kbps while consuming 72.74mW [AWB07], its power consumption is linearly related to the data rate it must support. Fig. 1.2 plots the relationship between the number of simultaneously active sensors and the resulting radio power consumption based on [Bal]. It also plots the total HERMES power consumption as the number of simultaneously active sensors is varied. Reducing the number of active sensors from 99 to 5 results in an 80% drop in total power consumtion.

1.3 Human Gait

Gait is defined as the way in which movement is achieved by humans with their limbs, such as walking, running, hopping, etc. The gait cycle, or stride, is a functional unit of gait defined as a single sequence of functions of one limb. It is divided into two phases, the stance phase, when the limb is in contact with the ground, and the swing phase, when the limb is in the air for advancement. As a result the gait cycle imposes structure to the plantar pressure signal

measured HERMES, and as we shall see, this will play a key role in signal segmentation. We shall also see that human gait imposes strong spatio-temporal structure to the data collected which will be capitalized on to achieve low-power, fault tolerant operation. We note that such structure is inherent to a majority of physiological and behavioral signals measured by BANs - Biosignals such as ECG, respiration, EEG exhibit segmentable and/or spatial-temporal structure that will allow for the application of power-reduction and fault-tolerance techniques such ours.

Finally, we note that the experimental datasets we use to validate our techniques are measured from walking subjects. Humans generally Walk at 1Hz and we will use this to translate power usage to energy consumption, as necessary.

The rest of our work is laid out as follows. Chapter 1 discusses a low-power sampling technique for remote health monitoring BANs, where the number of samples taken simultaneously is constrained rather than the total number of sensors sampled during operation, and the objective is to maximize diagnostic metric accuracy. The technique also admits fault tolerant operation, wherein fault detection is assumed to occur out-of-band. Chapter 2 describes low-power sampling for an alternate use-case wherein it is required that diagnostic metrics only be reported over user behaviors that are "unusual". Chapter 3 describes a low-power sampling and fault detection technique wherein the number of sensors sampled during system operation is constrained to minimize diagnostic metric accuracy. However, the problem is formulated and solved in a manner that yields fault-detection capabilities.

Chapter 4 describes properties of a novel energy harvester technology and details our work modeling its electrical and mechanical behavior, while chapter 5 studies the application of this novel energy harvester technology to harvesting energy from foot-strikes. Our goal is to power BANs from the scavenged energy in a self-sustaining manner. Several optimizations are discussed to maximize the energy output of the proposed foot-strike energy harvester platform while adhering to user-comfort requirements, to achieve the stated goal of self-sustained BAN operation.

CHAPTER 2

Fault-Tolerant and Power Constrained Sampling

2.1 Introduction

BAN usability is stymied by their intrusiveness and lifetime issues. For example, the HERMES platform [NAH10] (Fig. 1.1), comprised of a localized multi-sensory array that monitors foot plantar pressure, must be energy-efficient so it may function unattended for extended periods, and light-weight with a small form factor so as not to hinder/disrupt normal activity of the subject wearing the system. Given that such devices are usually battery operated, wherein the rate of power consumption determines the time between recharges, and that power consumption affects the battery size and therefore the form factor, power consumption is crucial to adoptability.

On the other hand, reliability issues are a by-product of physical and economic constraints. Medical sensors can be quite expensive, and over-deploying them towards redundancy and fault-tolerance on space constrained wearable platforms, such as HERMES, is infeasible. In this chapter, we aim to address these issues with a sampling paradigm that is fault-tolerant and low-power, yet affords high quality medical diagnostic measurements.

Our sampling paradigm abides by pre-specified power thresholds by procuring at most k samples at a time so that the maximum total power drawn by the sensor array is less than this threshold. Furthermore, sensing fault tolerance is realized by ensuring that up to a predefined number of the sampled sensors are allowed to fail, before all the information deduced from their samples is no longer recoverable by other means. We call this number the redundancy factor, or $redFactor$. The former constraint enables device designers to customize power usage based on requisite diagnostic performance, while the latter lets them

cater to requisite device lifetime guarantees based on the sensor fault rates, while maintaining diagnostic accuracy throughout the device lifetime.

Our approach strives to maximize the information available for diagnostic extraction under the fault-tolerance and power constraints, by (i) sampling sensors with the most contextually relevant information at the sampled time step, or epoch, (ii) utilizing domain-specific spatio-temporal relationships for a given epoch, to infer measurements at un-sampled sensors with high fidelity, (iii) sampling/inferring a sensor's measurement only if it can also be inferred by at least $redFactor$ other sensors, in case the sampled sensor fails, and, (iv) covering sensors, via measurement or inference, over the majority of epochs when their measurements are most relevant. This last approach affords us the added advantage of improving energy-efficiency by sampling no sensors at epochs deemed irrelevant. We also demonstrate that our algorithm is capable of abiding by the fault-tolerance and power threshold constraints in the context of multi-modal/diagnostic sensing, effectively distributing sensor samples across epochs and among the multiple modalities/diagnostics.

Although we present our algorithm in the context of plantar pressure monitoring systems, it is easily generalized to the monitoring of periodic behaviors, such as those common to BANs, with devices that wield localized multi-sensory arrays to collect data with strong spatio-temporal dependencies. Also, we limit ourselves to a binary fault model and assume out-of-band fault detection.

This chapter is organized as follows. Section 2.2 summarizes related work in the area of Wireless Sensor Networks (WSNs) and Body Area Sensor Networks (BANs), highlighting BAN applications and contrasting our approach with others. Section 2.3 describes the fundamentals of our application. Section 2.4 motivates our approach based on statistical observations in the context of human gait, and provides a detailed description of our algorithm and the tradeoffs therein. Section 2.5 illustrates our algorithm's performance in terms of energy-efficiency and diagnostic accuracy based on experimentally collected datasets. And section 2.7 offers conclusions for the chapter.

2.2 Related Work

WSNs have emerged as a new class of systems that leverage the power of distributed computation, sensing and actuation towards a host of scientific, military and engineering applications. However, engineering these systems is not without challenges, a topic surveyed excellently in [CES04]. As described therein, from the get go, energy has been identified as the single most important resource for this class. Furthermore, communication is recognized as the most demanding operation on this resource, and a number of designs have been proposed to curtail it [IGE00] [PHC04].

The demand for similar sensor-embedded devices in the medical domain has spawned a new subclass of WSNs known as wearable or Body Area Sensor Networks (BANs). BANs have been applied to host of medical problems such as gait analysis [BBB10], geriatric assistance [WAJ08], physiological stress inference [PRH11], and emotional health monitoring [DGS]. The challenges and opportunities arising from BANs were surveyed in [HPB09], where energy was identified as a crucial resource. As a result energy optimization techniques from WSNs have been altered to cater to BANs, including those spanning communication [GJ10] [NBG11], sensing coverage [GGG08], energy scavenging [LFS05], etc. Given that it is possible to build systems that can interpret medical diagnostics from electrical signals, attention has also been paid to the power consumption of BANs and their effects on system longevity [YZJ07].

Fault tolerance is also a key performance criterion in WSNs. Solutions in this space include self-diagnosis techniques [DLT07], as well as, co-operative fault recovery techniques [MEB09]. Often, fault-tolerant designs involve redundancy, which, when considered simultaneously with the need for lifetime maximizing/energy-efficient operation, exposes a tradeoff. For example, redundant sensor node deployment is usually accompanied by energy-efficient area coverage [MKP05]. Here the monitored area is covered, in the sensing and communication domains, by multiple distinct and minimally-sized subsets of the over-deployed set that are operated one-at-a-time, thereby achieving high sensing fidelity and redundancy, at low energy costs.

Similar ideas have migrated to the context of BANs to resolve the tradeoff between sensing fidelity and energy-efficiency, yielding sensor selection approaches wherein a subset of sensors, selected for their ability to jointly predict the measurements at all sensors, are sampled at each epoch [NDM10]. However, extending these solutions to provide fault tolerance by constructing multiple distinct sensor subsets is hindered by multiple factors. For one, a BAN sensor's coverage area rapidly changes with time. Also, a BAN designer's ability to overpopulate the wearable platform with sensors is severely limited by the wearable platform size, sensor size and cost.

In [WP11] the authors proposed an algorithm to further relax the tradeoff between sensing fidelity and "covering" sensor subset size (or power usage) by defining the fidelity only in terms of diagnostic accuracy. This focused definition of fidelity promotes the exploitation of the highly localized nature of body area networks that engenders strong spatio-temporal relationships in the sensed data. Since we no longer require high-fidelity measurements of all samples of the spatio-temporal signal segment, but only specific ones with high informational content in relation to the relevant medical diagnostics, it allows us to tackle the problem from the perspective of sample selection, instead of sensor selection, which not only motivates high-fidelity sample inference, but also enables the substitution a sample from a faulty sensor with a sample from an accurate one, yielding low-power, fault-tolerant and high-fidelity diagnostic metric estimation.

2.3 Preliminaries

2.3.1 Human Balance Monitoring via HERMES

While HERMES is capable of continuous measurement, medical professionals are interested in a set of diagnostic metrics for each stride taken by the subject, which do not require all samples for the stride. For example, the GARS-M scale [VPB96], a reliable and valid measure for documenting gait features associated with an increased risk of falling among community-dwelling, frail older persons, relies on guardedness as one of its variables. Measured as the

time between heel and toe strikes during a stride, guardedness has also been identified as an indicator of repetitive stress injuries in runners [DGW12]. However this metric only requires that each sensor be sampled until the detection of foot strike at its respective location, but not after. Similarly, elevated levels of plantar pressure, measured as the average maximum pressure over all sensors for a stride, has been identified as a causative factor in the development of many plantar ulcers in diabetic patients [CB10]. Again, the only relevant sample for each sensor is the one that provides the maximum pressure for a stride. In this paper we focus on guardedness, average maximum pressure (or maximum amplitude) and lateral pressure difference. The last diagnostic metric is an indicator of staggering in the GARS-M. We note that each of these diagnostics is an aggregate over the respective individual sensor metric measurements.

2.3.2 Sample Schedules Based on the Gait Cycle

A repeatable sampling scheme is constructed by segmenting the signal based on the phases of the gait cycle. Given that it is possible to orthogonally detect these phases with a small fixed number of sensors with the method proposed in [PKM04] or via a single large plantar pressure sensor covering the heel as outlined in [WP11], the sampling scheme produced by our algorithm is defined vis-*à*-vis the start and end of each of the two phases. Without loss of generality, we note that our selected diagnostics of guardedness, maximum amplitude and lateral pressure difference are all related to samples in the stance phase.

2.3.3 Energy Consumption

We will compare our algorithm to the CICA-based algorithm that selects a fixed subset of sensors of cardinality, k, to be sampled at every epoch [WP11]. We note that our approach offers the added advantage that fewer than k sensors may be sampled at some epochs. This translates to a further savings in energy over this algorithm. Fig. 2.1 shows a time histogram of the epochs when maximum pressure amplitude was observed during a subjects stance phase at all 99 sensor locations (ordered from heel to toes) from one of the HERMES

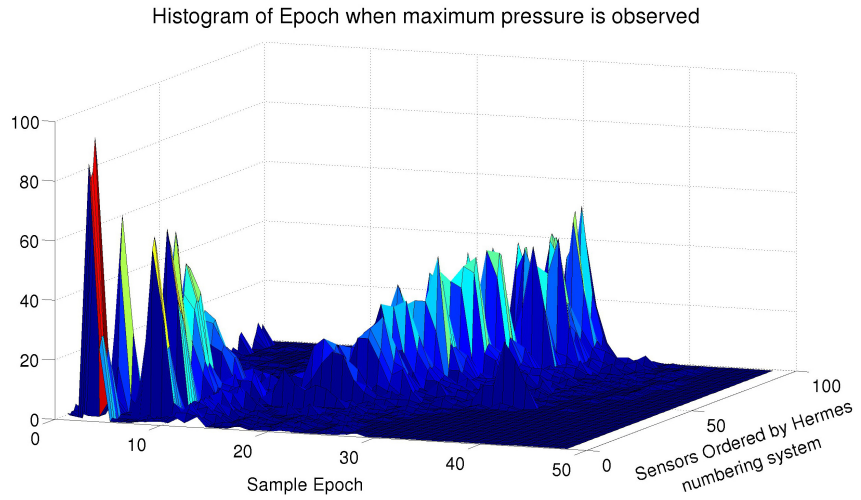


Figure 2.1: Histograms of epochs when maximum pressure was observed by HERMES Plantar Pressure Sensors.

datasets used in our study. What we see is that it is unlikely that much valuable information would be gained from sampling at some epochs (e.g. epochs 40-50, 10-20). Reducing the number of samples taken during such epochs to a value less than k yields sampling energy savings. Although sampling energy is not considered in our objective function, these savings are measured by the sampling energy savings factor, calculated as the ratio of the actual number of samples that are taken to the maximum number of samples that may be taken.

2.4 Diagnostic Driven Sample Selection

2.4.1 Algorithmic Motivation

We now discuss the motivation behind our algorithmic decisions. The stance phase of the gait cycle is divided into a number of epochs at which we would like plantar pressure to be sampled at various sensors. In order to support pre-specified constraints on maximum power usage, we limit the maximum number of samples, k, which may be taken at an epoch. A sensor would make a good choice for sampling at an epoch if it can be used to derive the readings at multiple other sensors, or predict them. This predictability of sensors may be determined by a model between readings of sensors. As observed in [WP11], linear regression

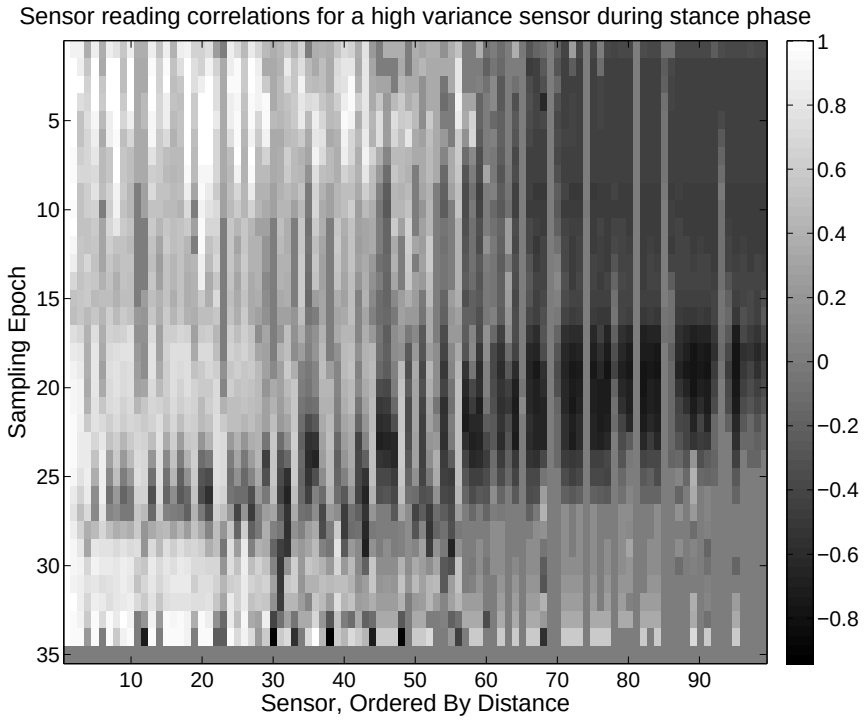


Figure 2.2: Spatio-Temporal correlations of plantar pressure measurements between a single sensor and all others on a HERMES shoe, over all epochs of the stance phase for a subject.

models do well to explain the plantar pressure readings of one sensor from another, when the mutual information is high. However, in the general case, our algorithm does not prevent the usage of any other suitable model for this purpose.

Fig. 2.2 shows the Spearman rank-order correlation coefficients between one sensor, s, and all 99 others over all epochs of the stance phase from a dataset collected by the HERMES platform. The 99 sensors on the x-axis are ordered by their physical distance from sensor s. In the spatial domain, we generally observe that nearby sensors are well correlated with the sensor s compared to distant ones. In the temporal domain, we observe that the level of correlation varies significantly across epochs, even for nearby sensors. The explanation for this hinges on the realization that the correlation between sensors varies by sub-phases of the stance phase.

Although we describe this insight for sensor s, we have observed that it generalizes to most sensors in all our datasets. Furthermore, each data point in Fig. 2.2 depicts the correlation

between measurements of two sensors at the same epoch. However, as observed in [NDM10], relative time shifting of the signals from 2 sensors greatly improves their correlation. As a result, we capture the ability of a sensor's measurement in one epoch to predict others in $senPred$, which is defined in equation (2.1).

$$senPred(e, s) = \{(e', t) \mid modErr(s_e, t_{e'}) < \epsilon_{modErr}\} \tag{2.1}$$

Here the function $modErr$ returns the model error, while predicting the measurements at sensor t in epoch e' from those of sensor s in epoch e, where e' may occur before or after e. Predictability is observed when the estimate is accurate to within ϵ_{modErr}. In our case, $modErr$ returns the root mean squared relative linear regression error.

Further, it is important that we place emphasis on sensors with readings that are more relevant to the medical diagnostic being computed. Given that the diagnostics of interest are aggregates of the metric measurements at each sensor, sensors with a larger spread, e.g. with high entropy, are more valuable when we attempt to reconstruct the medical diagnostic with a partial sets of readings. For example, assuming that we can use the maximum amplitude of some subset of sensors to estimate the average maximum amplitude over all sensors for each stride, the maximum amplitude of sensors whose values tend to vary more across strides is more valuable. For metric m at sensor s, this is measured as $spread_m(s)$, which in our case returns the entropy of m as measured at s. While the variance or inter-quartile range could be used instead of entropy, we observed superior results with entropy for our datasets owing to the mostly unimodal distribution of sensor measurements.

In arriving at the value of a metric m at a sensor s, it is reasonable to sample/predict the measurements of s at only those epochs when we would expect it to yield the value of m. In other words, to figure out the maximum amplitude at s, we would like to sample/predict s's measurement only at the epochs when s is likely to observe maximum amplitude. Given a training dataset, we construct a distribution over epochs, e, of the likelihood when a sensor s observes its metric value m and represent this as $distr_m(e, s)$ (see Fig. 2.1). Sensors will then be sampled according to $distr_m$ allowing us to get away with not sampling any sensor

in some epochs. For example, if $distr_m(e, s)$ indicates that metric m is never seen by s at epoch 30, rather that m tends to occur around epoch 5, we do not need to sample s for m at epoch 30 due to its statistical improbability of occurrence at that epoch. This effect is further enhanced by the spatio-temporal correlations, as a sensor may be sampled at one epoch to recreate the measurements of other sensors at other epochs, vacating the necessity to sample at these other epochs.

This approach to maximizing the quality of diagnostic information extracted, while limiting the maximum number of sensors, k, sampled at an epoch is summarized here:

1. Inferring readings with the most contextually relevant information, which corresponds to measurements at sensors with the most entropy, at epochs when they are likely to the metric. ($spread_m$)

2. Utilizing spatio-temporal relationships for a given epoch, to infer readings at unsampled sensors with high fidelity. ($senPred$)

3. Covering sensors, via measurement or inference, over a majority of the epochs when they are most likely to achieve maximum amplitude. ($distr_m$)

4. Forfeiting concurrent coverage of sensors whose maximum amplitude is strongly correlated.

The benefits of these directives are highlighted by Fig. 2.3. From the likelihood distribution of 4 representative color-coded sensors over 8 epochs (Fig. 2.3a), we see that the narrow width of the distributions allows us to gain information by forgoing inference of a sensor's reading at most epochs which creates space for others. For example, $s4$ need be sampled only at epochs $e5$ thru $e7$ leaving space for others to be sampled at $e1$ thru $e4$. It also obviates sampling at irrelevant epochs, such as $e8$, thereby improving energy-efficiency. Furthermore, inference based on spatio-temporal correlations (Fig. 2.3c) enables the recreation of multiple readings from a single sample, including those at epochs other than the sampled one, vacating the necessity to sample at these other epochs. In our example, sample $(e4, s1)$ predicts readings $(e3, s3)$ and $(e4, s3)$.

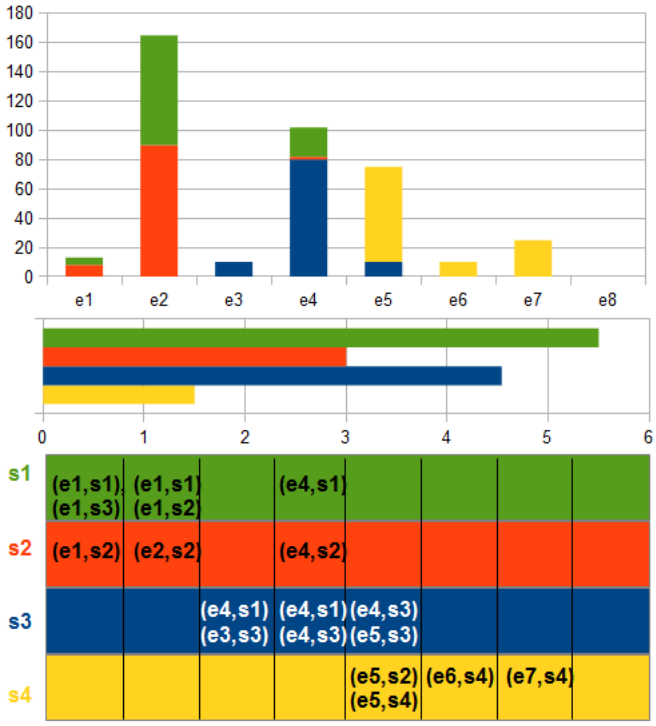


Figure 2.3: (a) Max. Amplitude Likelihood Distribution (top); (b) Max. Amplitude entropy (middle); (c) Spatio-temporal inference map that lists the set of samples that may be used to infer each reading (bottom).

Finally, sensing fault-tolerance is provided by applying one sensor's sample towards prediction of another's, only if the latter's sample is also predictable by $redFactor$ samples from sensors that are distinct from each other and the original predictor. For example, in order for it to be acceptable that s's measurement at epoch e will infer t's measurement at epoch e', there must exist a set of sensors $(s_1, , s_{redFactor})$ distinct from s, each of which produce at least one sample capable of predicting t's measurement at e'. It is important to note here that the user-imposed power-constraints must simultaneously be adhered to; with the premise that sensor s's failure will result in one of the $redFactor$ samples being used to recover t's reading at epoch e', it must be ensured that this replacement sample must occur in an epoch when, in s's absence, at most $k-1$ sensors are sampled.

In the following subsections, we describe an offline learning approach that combines these considerations. The described approach produces a customized sampling schedule based on

a training dataset of readings, at all sensors and over a few gait cycles of the subject. The resulting schedule may then be applied towards fault-tolerant and power constrained diagnostic estimation for the subject.

2.4.2 Power Constrained Sample Selection

We now describe the power constrained sample selection algorithm (PCSS) in Algorithm 1, which determines the sensors to be sampled at each epoch. The merit of a sensor s's assignment to an epoch e is measured by a heuristic $sampContr(e, s)$. The inputs to the algorithm are $senPred$, $distr_m$, and $spread_m$ that are described in the previous subsection, and the maximum samples per epoch, k. The algorithm returns $sensSamp$, which stores the sensors to be sampled at each epoch, and $predMap$ that maps each sensor t and the epoch e'' at which its value will be predicted, to the sensor s' sampled at epoch e' whose value will be used to predict (t, e''). For reasons that will be made clear in subsection 2.4.4, PCSS also takes $sensSamp$ and $predMap$ as inputs that can be assumed to be empty for now. Note that we are only interested in a sensor's measurement at epochs when the metric observation likelihood matrix, $distr_m$, indicates that the likelihood is non-trivial (lines 3 thru 7), and when the measurement hasn't already been covered (lines 13 thru 16). This latter point is noteworthy as we do not want to value sensor s at an epoch e for its ability to predict sensor t's readings at an important epoch e', if $predMap(e', t)$ has already been covered by a different pair (e'', s') that was selected into $sensSamp$. Furthermore, we are not interested in sampling sensors at epochs when k samples have already been scheduled (line 4). Finally, sensor selection is terminated when either no unassigned epochs remain in $sensSamp$, or no merited assignments may be made (lines 8 to 10).

The valuation heuristic $sampContr(e, s)$ combines, for each (e', t) that it covers, the spread of sensor t's readings via function f, and, the metric observation likelihood at e' via function g, where the likelihood matrix is adjusted at each iteration based on sampling assignments (line 15). The valuation also takes into consideration, via function h, the extent to which t has been covered at epochs with non-trivial metric observation likelihood, the

Algorithm 1 PCSS: Power Constrained Sample Selection

Input: $senPred$; $distr_m$; $spread_m$; k, the maximum sensors to sample per epoch; $sensSamp$, a $numEpoch \times k$ matrix of sensors to sample; $predMap$, a $numEpoch \times numSensor$ matrix indicating the sample (e, s) whose value will be used to predict measurements at (e', t)

Output: $sensSamp$; $distr_m$; $predMap$

1: **while** True **do**

2: $sampContr(e, s) \leftarrow 0 \ \forall$ epochs e and sensors s

3: **for** $e \leftarrow 1$ to $numEpochs, s \leftarrow 1$ to $numSensors$ **do**

4: **if** $sensSamp(e, k') = NULL$ for some $k' \leq k$ **then**

5: $$sampContr(e, s) \leftarrow \sum_{(e',t) \in senPred(e,s)} \begin{pmatrix} f(\overline{spread_m}(t)) \\ \times g(distr_m(e', t)) \\ \times h(\sum_{e''} distr_m(e'', t)) \end{pmatrix}$$

6: **end if**

7: **end for**

8: **if** $sampContr(e, s) = 0 \ \forall$ epochs e and sensors s **then**

9: Break

10: **end if**

11: $\max_{e,s}(sampContr) = sampContr(e', s')$

12: $sensSamp(e', k') \leftarrow s'$, where k' is the first index $\leq k$ with non-zero value at e'

13: **for** $(e'', t) \in senPred(e', s')$ **do**

14: $predMap(e'', t) \leftarrow (e', s')$

15: $distr_m(e'', t) \leftarrow 0$

16: **end for**

17: **end while**

18: Return $sensSamp$, $distr_m$ and $predMap$

motivation for which will be discussed in the subsection 2.4.4. The functions f, g and h are used to assign relative importance to their operands in the computation of the valuation heuristic. We use power functions with different exponents to express the relative importance.

2.4.3 Optimal Fault Tolerant Coverage

As described thus far, the PCSS algorithm attempts to cover the most relevant samples under pre-specified power-constraints. However, this coverage is not fault-tolerant. Towards this end, we augment the algorithm with a call to the subroutine outlined in Algorithm 2, while determining the coverage of a sample (line 5 of PCSS). This subroutine called Optimal Fault

Algorithm 2 OFTC: Optimal Fault Tolerant Coverage

Input: $senPred$; $distr_m$; $spread_m$; $sensSamp$, a $numEpoch \times k$ matrix of sensors to sample; s, the sampled sensor; e, the sampling epoch; $redFactor$, the redundancy factor

Output: $covSet$, the optimal set of samples (e', t') covered by (e, s) in a fault-tolerant manner under $sampContr$; $bestSetVal$, the heuristic value of $covSet$ under $sampContr$; $redSet$, containing a set of samples, for each (e', t') in $covSet$, that will cover (e', t') if sensor s fails

1: $ftPred(t'') \leftarrow NULL \ \forall$ sensors t''
2: $bestSetVal \leftarrow -\infty$
3: $redSet''(e', t') \leftarrow$ set of samples (e'', t'') corresponding to distinct sensors $t'' \neq s$ scheduled in sensSamp s.t. $(e', t') \in senPred(e, s) \cap senPred(e'', t'')$
4: **for all** sensors $t'' \neq s$ **do**
5: $\quad ftPred(t'') \leftarrow senPred(e, s) \cap senPred(e, t'')$
6: **end for**
7: $combFtSet \leftarrow redFactor$-combinations of all t'' s.t. $ftPred(t'') \neq NULL$
8: **for all** set of sensors $ftSet \in combFtSet$ **do**
9: $\quad redSet'(e', t') \leftarrow redSet''(e', t') \ \forall$ epochs e' and sensors t'
10: $\quad$ **for all** $t'' \in ftSet$ **do**
11: $\quad\quad$ **for all** $(e', t') \in ftPred(t'')$ **do**
12: $\quad\quad\quad$ **if** $(e''', t'') \notin redSet'(e', t') \ \forall$ epochs e''' **then**
13: $\quad\quad\quad\quad redSet'(e', t') \leftarrow redSet'(e', t') \cup (e, t'')$
14: $\quad\quad\quad$ **end if**
15: $\quad\quad$ **end for**
16: $\quad$ **end for**

17: $redSet'(e',t') \leftarrow NULL \forall$ epochs e' and sensors t' s.t. $cardinality(redSet'(e',t')) < redFactor$

18: $setVal \leftarrow \sum_{(e',t')s.t.redSet'(e',t')\neq NULL} \begin{pmatrix} f(\overline{spread_m}(t')) \\ \times g(distr_m(e',t')) \\ \times h(\sum_{e'''} distr_m(e''',t')) \end{pmatrix}$

19: **if** $bestSetVal < setVal$ **then**

20: $bestSetVal \leftarrow setVal$

21: $redSet \leftarrow redSet'$

22: $covSet \leftarrow (e',t')$ s.t. $redSet(e',t') \neq NULL$

23: **end if**

24: **end for**

25: Return $covSet$, $bestSetVal$ and $redSet$

Tolerant Coverage (OFTC), maximizes the value of the sample set covered by each sample (e, s) (line 3 of PCSS) under the constraints that each of the covered samples (e', t) is not only well predicted by the model (by being part of $senPred(e, s)$), but also well predicted by $redFactor$ distinct other sensors that provide redundant coverage. Here, the value of each covered sample is derived as described in the composition of *sampContr* in the previous section (line 5 of PCSS and line 18 of OFTC).

The maximum-valued sample set covered by a sample (e, s), or *covSet*, is a subset of all samples that could possibly be covered by (e, s), or $senPred(e, s)$. Furthermore, it is a subset which can simultaneously be covered by $redFactor$ distinct other sensors, while obeying the power constraints on the maximum number of samples per epoch, k, even if sensor faults occur. A valuable observation is that each of the samples in *covSet* can admit redundant coverage by a distinct set of samples, as long as the power-constraint is honored. For example, a sample (e', t') selected into the *covSet* of sample (e, s) could have redundant coverage from (e'', s') that is already selected into *sensSamp*, while the sample (e''', t''), which cannot be accurately inferred from (e'', s'), could also be selected into the *covSet* of sample (e, s) due to redundant coverage from a different sample selected into *sensSamp*. This is the basis for computing the redundant coverage from the samples in *sensSamp* for

all samples in $senPred(e, s)$ (line 3). Note that there is no power cost incurred by this form of collaborative redundant coverage.

If the redundant coverage obtained thusly doesn't satisfy the $redFactor$ cardinality constraint for some samples in $senPred(e, s)$, backup samples must be available in epoch e, so they may replace the faulty sensor s's sample without incurring additional power costs. This type of redundant coverage is captured in ftPred (lines 4-6). In order to then pick the maximum-valued subset of $senPred(e, s)$ that can be covered with redundancy and under power constraints, all $redFactor$-combinations of sensors that can provide redundant coverage for samples in $senPred(e, s)$, while themselves being sampled in epoch e, are iterated over (lines 7-24). In doing so, it is ensured that each covered sample admits redundant coverage from a distinct set of sensors (lines 3, 4, 12).

Support for fault-tolerance, via a call to the OFTC algorithm, involves related changes to the PCSS algorithm. These include taking $redFactor$ and $redSet$, the redundant covering set for each sample covered in $predMap$, as additional inputs and returning $redSet$ as an additional output. Most importantly, the computation of $sampContr$ (line 5 of PCSS) is replaced by a call to OFTC with $senPred$, $distrm$, $spreadm$, $sensSamp$, s, e and $redFactor$ as inputs. $sampContr(e, s)$ is then represented by the value of $bestSetVal$ returned by OFTC. Finally, the set of samples accounted for as covered (lines 13-16 of PCSS) are limited to the OFTC return value of $covSet$ for (e', s'), while the return value of $redSet$ is merged into the corresponding local copy in PCSS.

2.4.4 Coverage Improving Iterative Refinement

While the PCSS algorithm is driven by the ability of samples to predict others, the metric observation likelihood of the covered sensors, and, their metric spreads, it does not ensure that the selected sensors are well covered. Put differently, each sensor that we decide to cover must be covered in most, if not all, epochs when they are likely to observe the metric, barring which we increase the likelihood of using erroneous values in the composition of the diagnostic and increase its error. This is precisely why, once the PCSS algorithm has

decided to cover a sensor, function h increases its contribution to the valuation of sensors that predict it; the lower the sum of likelihood of metric observation over epochs that haven't been covered, the more we would like to ensure that it gets covered. h is particularly helpful in breaking ties when multiple sensors with similar spreads and are vying for an epoch where their metric observation likelihood is similar.

However, if the PCSS algorithm chooses to cover too many distinct sensors in the first

Algorithm 3 CIIR: Coverage Improving Iterative Refinement

Input: $senPred$; $distr_m$; $spread_m$; k, the maximum sensors to sample per epoch; $equiv_m$, a map of equivalent sensors; $redFactor$, the redundancy factor

Output: $sensSamp$, a $numEpoch \times k$ matrix of sensors to sample; $predMap$, a $numEpoch \times numSensor$ matrix indicating the sample (e, s) whose value will be used to predict measurements at (e', t); $redSet$, a $numEpoch \times numSensor$ matrix of indicating the set of samples that provide backup for the prediction of measurements at (e', t)

$\tau_{cov} \leftarrow \tau_{mincov}$
$predMap(e, s) \leftarrow NULL$ $\forall$ epochs e, sensors s
$sensSamp(e, k') \leftarrow NULL$ $\forall$ epochs e, $k' \in [1, k]$
$redSet(e, s) \leftarrow NULL$ $\forall$ epochs e, sensors s
for $k' \leftarrow 1$ to k **do**
  $senPred' \leftarrow senPred$
  while $\tau_{cov} \leq \tau_{maxcov}$ **do**
    Run $PCSS$ with input $predMap, sensSamp, redSet, senPred', spread_m, distr_m$ and k' and store the return values in $predMap', redSet', distr'_m$ and $sensSamp'$
    $coverage(s) \leftarrow 1 - \sum_{e \in [1, numEpochs]} distr'_m(e, s)$ $\forall$ sensors s
    for all epochs e, sensors s s.t. $coverage(s) < \tau_{cov}$ **do**
      $predMap'(e, s) \leftarrow NULL$
      $redSet'(e, s) \leftarrow NULL$
      $distr'_m(e, s) \leftarrow distr_m(e, s)$
    end for

15: **for all** epochs e, $k'' \in [1, k]$ **do**

16: **if** $predMap'(e', s') = NULL \ \forall (e', s')$ covered by $sensSamp'(e, k'')$ **then**

17: $sensSamp'(e, k'') \leftarrow NULL$

18: **end if**

19: **end for**

20: **for all** epochs e, sensors s, sensors t **do**

21: **if** $coverage(t) < \tau_{cov}$ **then**

22: Remove (t, e') from $senPred'(e, s) \ \forall$ epochs e'

23: **end if**

24: **end for**

25: $\tau_{cov} \leftarrow \tau_{cov} + \delta$

26: **end while**

27: **for all** epochs e, sensors s, sensors r s.t. $coverage(r) \geq \tau_{maxcov}$ **do**

28: **if** $t \in equiv_m(r)$ **then**

29: Remove (t, e') from $senPred(e, s)$

30: **end if**

31: **end for**

32: $predMap \leftarrow predMap'$

33: $sensSamp \leftarrow sensSamp'$

34: $redSet \leftarrow redSet'$

35: $distr_m \leftarrow distr'_m$

36: **end for**

37: Return $sensSamp$, $predMap$ and $redSet$

few iterations, it is likely that it will be difficult to ensure high coverage of all these sensors in the end, due to an increase in competition among the covered sensors for each epoch. This effect is exacerbated as k increases; as more samples may be scheduled per epoch, PCSS's initial focus is to choose to cover sensors with high entropy at epochs where the metric observations likelihood is high. By the time the algorithm alters course to increase

coverage, the competition may be too high. We resolve this by running the PCSS algorithm in the context of an iteratively refinement strategy for sensor coverage, that we call CIIR and outline in Algorithm 3.

The CIIR algorithm eliminates those sensors from consideration for coverage at each successive run of the PCSS algorithm, which have failed to meet a coverage threshold, τ_{cov}, over previous runs (lines 9 thru 24). This is achieved by removing them from the $senPred$ structure and a corresponding cleanup of the $predMap$, $redSet$ and $sensSamp$ structures. This reduces competition and subsequently improves coverage for the remaining sensors. Recall that the PCSS algorithm uses the $senPred$ structure to determine valuation and assign coverage in lines 5 and 14 thru 17, respectively, of the PCSS algorithm. To support this iterative improvement of coverage, the CIIR algorithm introduces 2 parameters τ_{mincov} and τ_{maxcov} that provide the range of coverage threshold values to be iterated over. Here, τ_{maxcov} serves as a tuning knob between accuracy via coverage of more sensors, versus, accuracy via improved precision of metric estimates at the covered sensors. Also, τ_{mincov} supplies the starting point for iterative improvement and may be set to the minimum coverage value observed, in *coverage* (line 9), after the first iteration.

To further improve coverage quality when multiple samples may be taken in an epoch, the CIIR algorithm calls the PCSS algorithm with stepwise increments, k', of the maximum number of samples per epoch, k (lines 5 and 8). Between such increments, sensors selected for coverage must be covered satisfactorily. Consequently, the covered sensors and sampling schedule constructed for one value of k' must be maintained as a subset of the covered sensors and sampling schedule for subsequent values of k' (lines 8 and 32 thru 35). In addition to improving coverage, this approach allows us the opportunity to keep from simultaneous covering sensors whose metric values are well correlated, or equivalent. The structure $equiv_m$ maps each sensor to the set of sensors that it is equivalent to. For our application, we define equivalence of sensors s and t by a low root mean squared relative linear regression error for the metric values at s and t. To avoid concurrent coverage of equivalent sensors, $senPred$ is updated, between increments of k', by filtering out all sensors that are equivalent to each covered sensor (lines 27 thru 31). Lastly, the point of having the PCSS algorithm make

sampling decisions for any value up to k' (line 4 of the PCSS algorithm), is that epochs left empty by PCSS when run with one value of k' may come in handy at a subsequent value of k'.

The CIIR algorithm terminates to produce the sampling scheme, in *sensSamp*, the resulting map of sensor measurements to predict, in $predMap$, and a multi-map of $redFactor$ backup samples from distinct sensors for each predicted measurement, in $redSet$. The individual metric value at each covered sensor is derived from its sampled/predicted measurements, in a manner identical to the case when its measurements are available for all epochs. For example, the maximum amplitude at a sensor s is computed as the maximum over all samples of s that were predicted or actually observed over the stride. Note that, by construction, we expect these sparsely sampled/predicted values to produce highly accurate metrics, since (i) the coverage of the relevant epochs for a sensor with respect to the metric is near complete, and (ii) for the predicted samples, the mapping in $predMap$ derives from the strongly linked samples in $senPred$.

The overall diagnostic metric is then derived from the metric values at the sensors for which they are available. Note that subject to the degree of coverage achievable under the power constraints, metric values may not be available at all sensors. Similar to [WP11], we use a multiple linear regression model to estimate the overall diagnostic metric for the metric values at the available subset of sensors. In case of sensor faults, $redSet$ is consulted to alter $predMap$, and possibly *sensSamp*, so that diagnostic metric composition can be continued at the same level of accuracy.

2.4.5 Power Constrained Multi-Modal/Diagnostic Sampling

Our algorithm may easily be extended to the multi-modal or multiple diagnostic metric cases, such as the composition of guardedness and average maximum amplitude from plantar pressure readings. In order to accomplish this, we first note that there must be one set each, of the following structures, per metric: $senPred$, *sensSamp*, $predMap$, $redSet$, $sampContr$ and $coverage$. We construct a vector of metric weights that will be used to switch between

the metrics in the main while loop of the PCSS algorithm (lines 1 thru 17). A weight, $w1$, for maximum amplitude and, $w2$, for guardedness will assign up to $w1/(w1 + w2)$ of the assigned epochs to samples for the maximum amplitude metric and up to $w2/(w1 + w2)$ of the assigned epochs to samples for guardedness. However, fewer epochs may be assigned to a metric, if only those many valuable samples may be obtained. The weights are derived from the width of the likelihood curves $distr_m$ for each metric; if a metric requires samples from more epochs, it deserves a higher weight. The assignments of epochs to samples for individual metrics are made in tandem. In addition, PCSS is modified to allow for the same sensor to be sampled at a given epoch for multiple diagnostic metrics. In this case, the construction of the redundant covering set, in $redSet$, to replace such a sample must consider the predicted measurements for each such diagnostic metric.

2.5 Results

We discuss the performance of our algorithm in the context of 5 plantar pressure datasets obtained from distinct subjects with distinct gait profiles, wearing the HERMES shoe while walking. The number of strides per dataset varies from 7 to a couple hundred. We divide each dataset into 2 subsets, ascribing 80% for training and 20% for testing. All the results discussed here are for the testing subsets. Also, note that each dataset includes data for the left and right foot which are treated independently.

2.5.1 Sensor Selection and Prediction

We assess performance from the perspectives of the quality of sensors selected, energy-efficiency of the proposed sampling schemes and the accuracy in metric prediction. The presented summaries are for $k = 1$ and $redFactor = 0$, wherein separate sampling schemes were constructed for each of the metrics. Tables 2.1 and 2.1 summarize the performance for the maximum amplitude and guardedness metrics, respectively.

We also observe that we are predicting sensors with substantial information - the average entropy of the predicted sensors as a percentage of maximum entropy is fairly high, especially

Table 2.1: Sensor Selection Performance For Maximum Amp.

Dataset	Sensors Sampled	Sensors Predicted	Mean % of Max. Entropy	Sampling Energy Savings Factor
Left 1	13	10	87	1.85
Right 1	11	9	96	2.08
Left 2	12	3	77	1.56
Right 2	8	2	88	1.78
Left 3	8	2	80	1.53
Right 3	6	3	80	2.5
Left 4	8	4	91	1.96
Right 4	4	2	95	2.0
Left 5	15	3	85	5.55
Right 5	9	2	89	5.26

Table 2.2: Sensor Selection Performance For Guardedness

Dataset	Sensors Sampled	Sensors Predicted	Mean % of Max. Entropy	Sampling Energy Savings Factor
Left 1	13	19	59	2.0
Right 1	10	12	73	2.32
Left 2	4	2	65	2.5
Right 2	6	2	67	2.0
Left 3	7	3	75	1.27
Right 3	2	8	70	1.0
Left 4	8	3	62	1.41
Right 4	2	6	74	1.0
Left 5	2	3	89	1.0
Right 5	2	6	87	1.0

in the case of the maximum amplitude metric. This bias is owed to the fact that the weight, in *sampContr*, placed on the spread in function f is much higher, relative to g and h, in the case of maximum amplitude. These weights were selected to achieve better diagnostic accuracy.

Further, we observe that despite us being able to predict many sensors, the proposed sampling schemes are able to provide sampling energy savings up to a factor of 5.55 with the mean savings hovering around 2 and 1.5 for each of the metrics, when compared to CICA [WP11]. The sampling energy savings factor is calculated based on the number of epochs during which fewer than k samples are taken, and only considers savings in the energy expended in taking samples. Note here, that the CICA-based algorithm samples the same sensor(s) at every epoch. So, it is clear, in that case, that only a single sensor's metric would have been available, given the power constraint. While the quality of the metric for this sensor would be perfect, Fig. 2.4 shows that in a majority of the datasets, our algorithm's accuracy of prediction, as described by the coefficient of determination, is high (close to 1.0) for all predicted sensors. Low accuracy is observed for the first dataset due to the fact that it is small, comprised of 5 training strides and 2 testing strides.

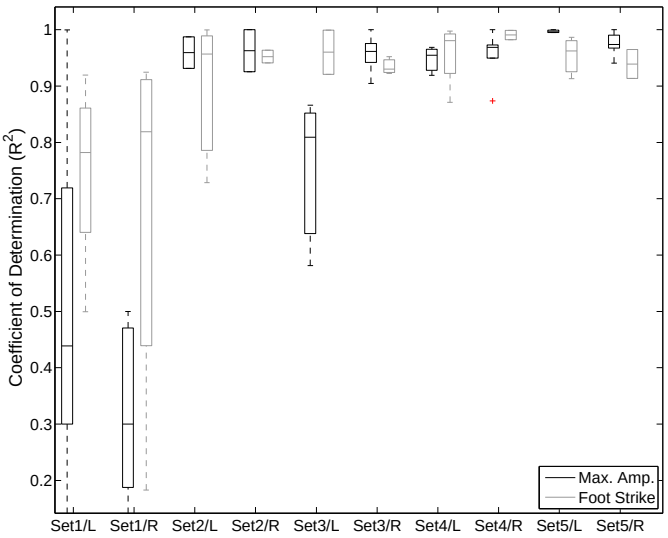


Figure 2.4: Coeff. of Determination (R2) for predicted sensors.

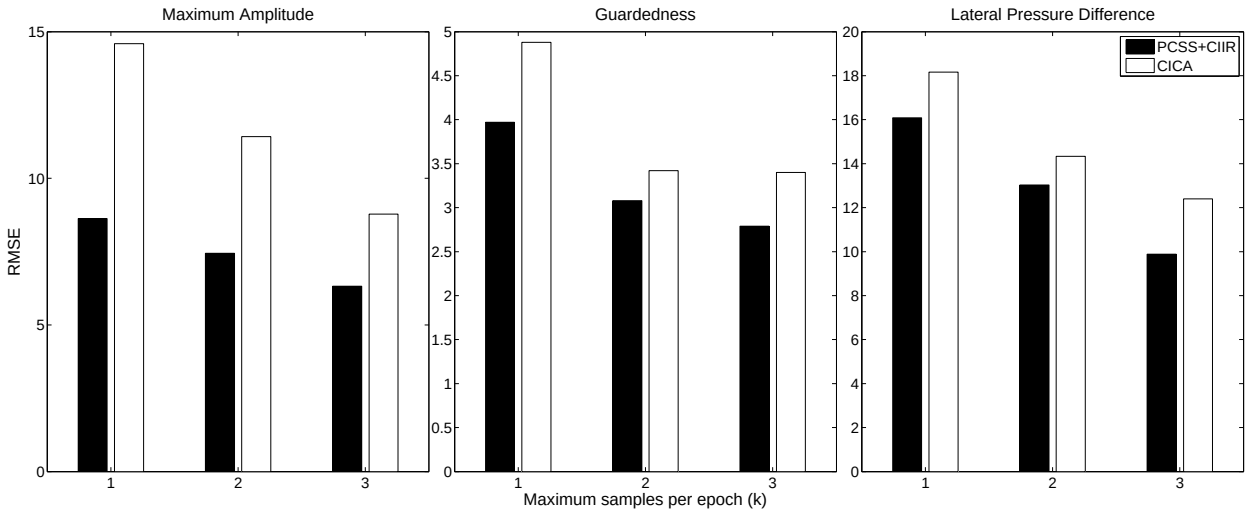


Figure 2.5: RMSE of max amplitude, guardedness and lateral pressure difference diagnostics for PCSS+CIIR vs. CICA [WP11].

2.5.2 Accuracy of Diagnostic Composition

Next, we compared the energy-efficiency and diagnostic accuracy of our algorithm for the average maximum amplitude, guardedness and lateral pressure difference diagnostics, to the CICA-based algorithm in [WP11], with $redFactor$ set to 0. Here, we measure accuracy in terms of the Root Mean Square Error (RMSE) in each of the diagnostics metrics, where the RMSE is taken jointly over all datasets. For both the algorithms, all metrics are predicted

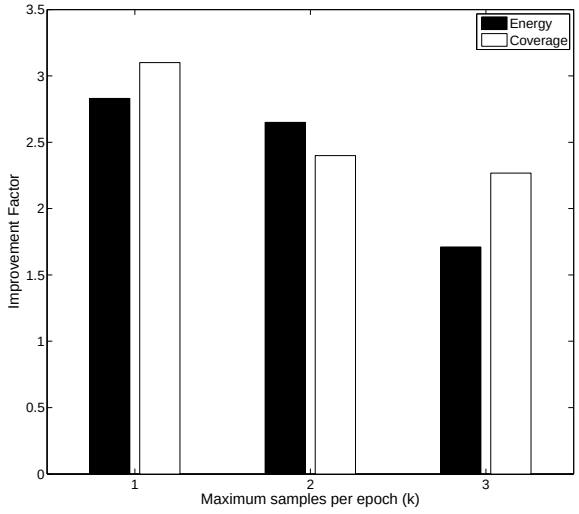


Figure 2.6: Sensor Coverage and Energy Savings Factors for different k.

from a single sampling scheme that is distinct for each dataset. Fig. 2.5 plots the observed RMSE of the 2 algorithms, averaged over all datasets, for each of the diagnostic metrics, and, for values of k between 1 and 3. The improvement over CICA ranges between 10% and 40%. We observe significant improvements in sampling energy savings as well as sensor coverage for different values of k over CICA (Fig. 2.6). Here, the sensor coverage factor is just the ratio of the total number of sensors whose metric values are available, to k. Both savings factors are a weighted average over all datasets, where the weights are the number of strides per dataset. This helped adjust for the large deviations observed in the first dataset (tables 2.1 and 2.2).

We are able infer metric values at over twice the number of sensors sampled by CICA. The sampling energy savings vary between 43% and 65% rising up to 74% for one of the datasets when k is 2. However, despite being able to infer metric values for many more sensors than k, we do not observe such large improvements in diagnostic accuracy. Whereas CICA picks the most orthogonal set of high-relevance sensors to construct the diagnostic metrics, our algorithm does not exert such fine grained control on the relationship between sensors selected for coverage. We also observe that our improvement over CICA for different values of k vary across diagnostics. This is due to the difference in the way diagnostic metrics are balanced by the two approaches while selecting a sensor for coverage. CICA is directly data driven and bases its sensor selection, at each iteration, on the improvement in error seen for each diagnostic. Based on our problem formulation, we have chosen instead to weigh sample selection for different metrics based on their relative likelihood distribution across the epochs in a stride.

2.5.3 Performance under Fault Tolerance

Finally, we compare the energy-efficiency and diagnostic accuracy of our algorithm under different values of $redFactor$ for each of the diagnostic metrics. As before, accuracy is measured in terms of the RMSE over all datasets for each diagnostic metric, based on a single sampling scheme per dataset. However, in order to determine the diagnostic accuracy

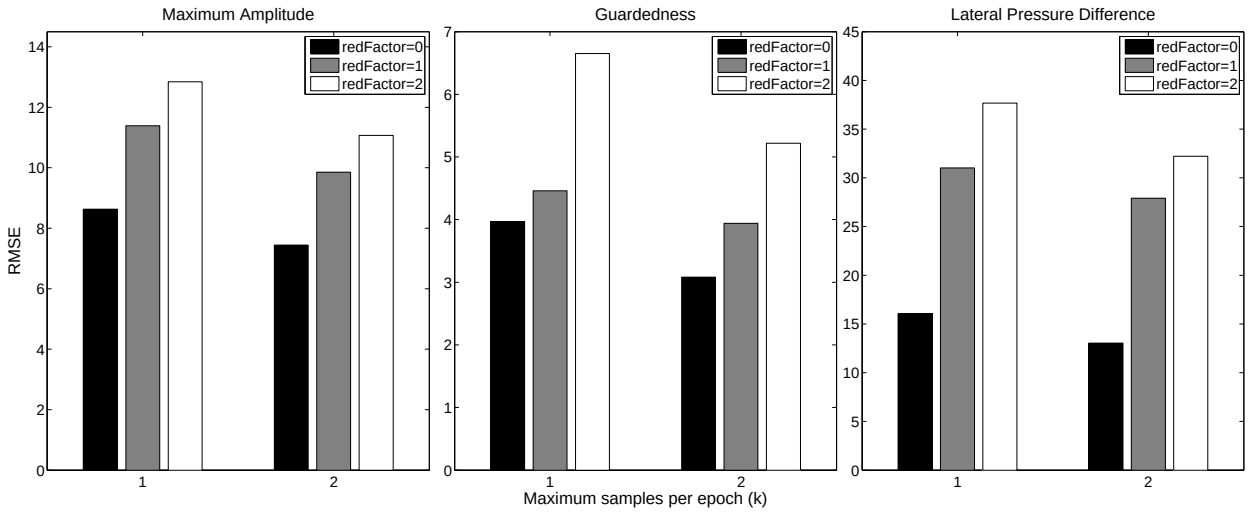


Figure 2.7: RMSE of max amplitude, guardedness and lateral pressure difference diagnostics for different levels of fault-tolerance.

in case of sensor faults, we must measure the error under such conditions. We accomplish this by removing a random sensor from $sensSamp$ and $predMap$, replacing its samples with those from $redSet$ and computing the error based on the resulting $predMap$. If $redFactor$ is greater than 1, this procedure is repeated $redFactor$ times. Fig. 2.7 shows the resulting RMSE for each metric, averaged over 100 such testing trials, for values of k set to 1 and 2, and values of $redFactor$ ranging from 0 to 2.

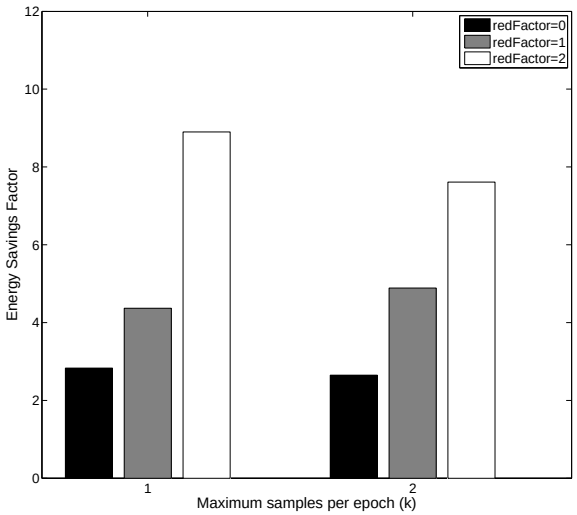


Figure 2.8: Energy Savings Factors for different levels of fault-tolerance.

In all cases, we see that the error increases with $redFactor$. As the need for fault-tolerance rises, there are fewer sensors that can be covered due to the increased standards of coverage under immutable spatio-temporal redundancy. Fig. 2.8 provides further evidence to this effect as we see a rise in the sampling energy savings factor (over CICA) with an increasing $redFactor$, owing to reduced coverage. However, we also notice that the error decreases with k regardless of $redFactor$, thereby allowing the device designer to cater to accuracy requirements under lifetime constraints by relaxing the power constraints.

2.6 Discussion

We now discuss assumptions, limitations, variations and future work for our approach. In this paper, we have outlined our algorithms as an offline learning approach, where we envision that the sensor sampling subsystem is customized to the activities of a user, such as her walking, running, jumping, climbing, etc., with a specific sampling schedule generated for each activity based on a short training dataset of readings at all sensors for her various activities. Alternatively, our algorithms may be used in an online fashion; if the accuracy of the predictions is evaluated at random or periodically, by sampling all sensors over a stride and comparing the predicted diagnostic value(s) to the actual value(s), the schedule may be discarded when such error appears to stay unreasonably high. At this time, the system may sample all sensors at all epochs for a few strides until a sufficient training set has been constructed, so the algorithms may be applied to yield a better sampling schedule. However, the assumption here is that the change is subject behavior that has led to the degradation of the original sampling schedule will be long lived so that the energy expended in collecting the training set may be amortized.

Also, while we have observed that our algorithms are generalizable beyond plantar pressure measurements, they do require that the behavior under measurement has semantic structure and can be easily segmented, so that a sampling schedule may be applied repeatedly. We maintain that, in addition to being localized in nature, BAN measurements in the medical domain are usually segmentable, be it for plantar pressure, ECGs or respiration.

Our approach is not suitable for measurements of behavior that is not segmentable. Finally, as discussed in the previous section, while there are significant gains from metric inference owing to spatio-temporal relationships in the data, when these relationships deviate our method produces larger errors.

2.7 Conclusion

We have presented an approach for diagnostic accuracy optimizing sample selection under fault-tolerance and power constraints on wearable BANs. Our formulation works based on the observations that the localized nature of measurements by BANs makes them rich in spatio-temporal redundancy, and that metric values are observed at sensors over a small subset of the entire behavior cycle making it prudent to sample only at those epochs. We have evaluated our approach on datasets collected experimentally with HERMES, for multiple medical diagnostic metrics based on plantar pressure measurements, wherein we are able to produce sensor fault-tolerant schedules with significant sampling energy savings and favorable diagnostic accuracy.

CHAPTER 3

Improving Energy-Efficiency for the Semantic Outlier Use Case

3.1 Introduction

While wireless medical monitoring promises vast improvements in applicability and quality of healthcare afforded via mobile patient monitoring systems, energy utilization of such systems pose challenges to large-scale and long term deployment [HPB09]. A large sub-class of such systems are comprised of multiple sensors that continuously measure physiological signals transmit them to a base station. To extract diagnostically meaningful information from the measurements, signal features or metrics are extracted from the signal measured at each sensor, and the metrics are combined to form one or more medically-relevant diagnostic metrics for the subject.

As we have discussed in earlier chapters, compressive sensing techniques are available that can improve energy-efficiency of this sub-class of systems by a significant margin. By carefully limiting the sampled sensors and their sampling times while maximizing the quality of information elicited with respect the metrics being monitored, these techniques are able to produce large savings in energy consumption from sampling and communication while maintaining high fidelity in the diagnostic metrics produced [GP12b] [WMN12].

While such techniques correctly characterize the quality of biosignals sensed based on their semantic relevance, the determination of which sensors and sampling epochs produce semantically rich information is often based on statistical techniques and will therefore optimize for the typical case. Rather, we contend that use-cases exist wherein metric observations

are much more important to healthcare providers when they are semantically atypical, as these are likely to be crucial in reducing diagnosis times by aiding in early detection of medical conditions. In contrast, techniques that maximize semantic accuracy over typical observations may entirely mis-estimate the diagnostic metric values during atypical subject behaviors, delaying diagnosis and making the monitoring device a safety hazard.

In this paper, we present a sampling technique that provides high-fidelity access to semantically atypical metric observations while constraining average energy consumption. In other words, the objective is to construct a sampling schedule that detects that an atypical subject behavior is underway as early as possible, and the stated energy constraints enforce the average number of data samples that may be collected and transmitted over typical subject behaviors. Once detected, all sensors are sampled continuously until the end of the signal segment is encountered, and all collected data is transmitted to the base station. The richer dataset available for the atypical subject behavior can now be used to generate more accurate diagnostic metric estimates.

Energy savings arise from identifying typical metric observations as such with fewer biosignal samples at a subset of the available sensors, and precluding transmission of these typical estimates. These savings then offset the higher energy costs arising from increasing the number of sampled sensors and the number of samples taken when the metric is expected to be atypical, thereby producing and transmitting more accurate metric estimates at such times. We validate the proposed method in the context of HERMES and compare its performance to that in chapter 2 for atypical metric observations.

3.2 Related Work

A number of signal-processing techniques have crossed over from the domain of WSNs to that of BANs with the goal of reducing power/energy consumption while maintaining signal fidelity. Sensor coverage techniques limit the number of sensors active simultaneously, while ensuring that the targeted area is "covered" without discontinuity. Duty cycling, a form of sensor subsampling, limits the amount of time each sensor node is active, thereby reducing

energy consumption. Similar signal processing techniques have crossed over to the domain of BANs [GP12a]. However, these techniques focus on reconstruction and semantic accuracy of an global aggregate signal rather than spatial fidelity of the signal.

Outlier detection techniques have also migrated from WSN to BAN applications, albeit in the context of fault detection. A fault detection technique for BANs was described in [KP11] wherein the relationship between sensors was modeled to detect outliers. The authors of [SLM13] applied spatial and temporal analyses to BAN sensor samples in order to weed out anomalous measurements and reduce false alarms. In contrast, we apply outlier detection techniques to remedy the shortcomings of energy-efficient statistical signal processing techniques for BANs. We achieve this by focusing on early detection of statistical anomalies to improve the fidelity of semantic conclusions that may be drawn from the measurements over such anomalous behaviors.

3.3 Gait Stability Monitoring

We explore and validate our algorithms in the context of gait stability monitoring. Human gait monitoring enables early detection of several conditions, including stress injuries in runners, plantar ulcers in diabetics and risk of falling in elder populations, to name a few [GP12b]. For example, the average maximum pressure metric has been identified as a causative factor in the development of diabetic plantar ulcers [CB10]. In this case, identification of subject actions that are likely to increase the likelihood of such plantar ulcers, based on the average maximum pressure diagnostic metric, is key. When an energy-efficient sampling scheme applies few sensors/samples to estimate the diagnostic metric, it is important that anomalous behaviors are identified early so that accurate diagnostic estimates can be produced. Failing this, it is possible that those anomalous behaviors increase the likelihood of plantar ulcers, but are often mis-characterized by the diagnostic metric estimate generated by the energy-efficient sampling scheme.

We explore and validate the performance of our algorithm for this average maximum amplitude diagnostic metric based on foot plantar pressure datasets experimentally collected

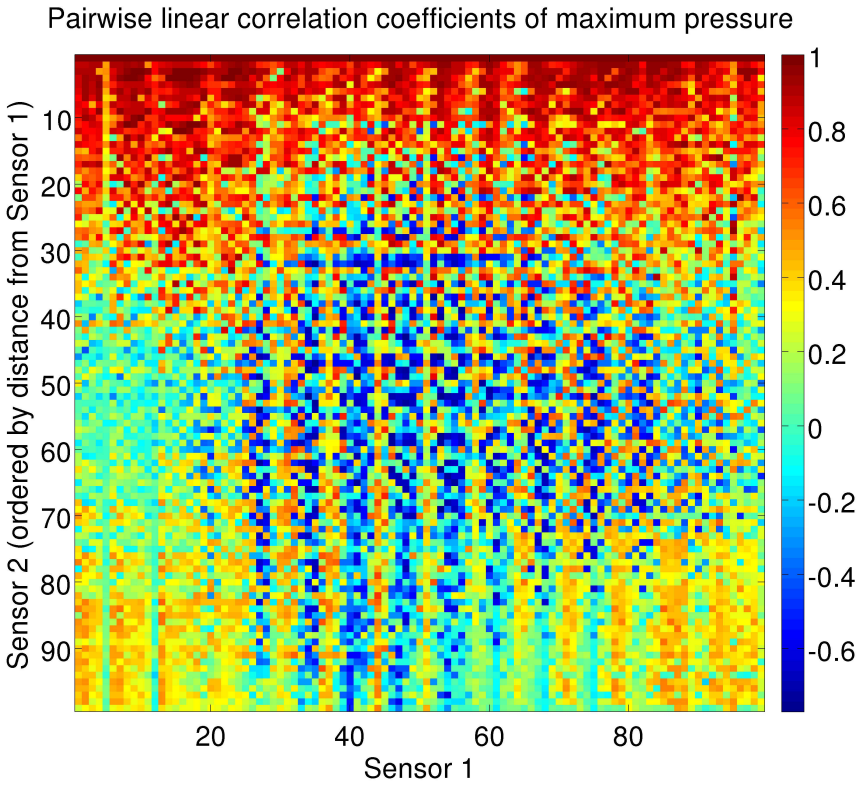


Figure 3.1: Pair-wise Correlation Map between Sensor Observations of the Max. Pressure Metric over a Foot Plantar Pressure Dataset.

with HERMES. Given the reading from HERMES's sensors, the average maximum pressure metric may be measured as the average, over all sensors, of the maximum pressure observed for a stride. The signal metric at each sensor is, therefore, the maximum pressure it observes over a footstep.

Based on a dataset of a user's plantar pressure profile collected over several footsteps by HERMES, Fig. 3.1 illustrates a map of the linear correlation coefficients for the maximum pressure metric between pairs of sensors, where we observe strong linear correlations among groups of sensors. This observation generalized to our other datasets as well. It is this high mutual semantic information among subsets of sensors that is exploited by compressive sensing schemes to limit the number of sensors and samples measured and communicated, while retaining high medical metric fidelity. For this reason, our algorithm checks for reductions in the level of mutual information within such subsets to identify user footsteps that appear

atypical/anomalous and are likely to produce low-fidelity medical metric estimates. As discussed, this is resolved by increasing the number of active sensors and samples collected to improve metric estimate fidelity, once an anomaly is detected.

3.4 Sample Selection Algorithm

3.4.1 Algorithmic Motivations

In order to predict atypical metric observations in an energy-efficient manner, it is important to do so accurately with few false positives and few false negatives. A large number of false positives will lead to an un-necessary increase in energy usage and many false negatives will lead to loss of information regarding atypical metric measurements. Further, it is also important to produce a detection as early as possible while staying within the energy constraints that apply over typical steps. This translates to the selection of samples at each stride such that the energy expended in sampling is within the stated constraints and such that the selected samples are able to yield rapid detection.

We identify atypical metric observations as those during which sensors that typically produce observations with high mutual semantic information fail to do so. The reasoning is as follows: With compressive sensing approaches, sensors with high mutual semantic information are unlikely to be sampled together due to the low information gain in sampling one having sampled the other. However, if such sensors begin to produce uncorrelated signal metric values, it is likely that the fidelity of the diagnostic metric estimate, which is based on the biosignal metric value of the sampled sensors but not the others, will deteriorate owing to a loss in semantic information corresponding to the metric value of the non-sampled sensor. To predict the loss of correlation, our algorithm samples sensors with high mutual information and constructs statistical models to estimate one from the others. If an estimated metric value differs significantly from the actual metric value then an outlier is detected. When the number of detected outliers exceeds a threshold, the diagnostic metric is expected to be atypical and all sensors are sampled at all the remaining epochs. This, then, allows for

a more accurate metric estimate to be produced based on the larger number of available samples.

As we have stated, the prediction of atypical metric observations must happen as accurately as possible under constraints on the energy usage, and must happen as early as possible. To fulfill the former requirement, sensors are assigned weights based on how relevant their metric values are to the aggregate diagnostic metric. Given, a training dataset of typical user footsteps, the feature weight, fw, of sensor s_i, is the coefficient of correlation between s_i's feature value (maximum pressure) and the aggregate metric (average maximum pressure). Given that the energy constraints are likely to limit the number of sensors that may be sampled, the feature weights will allow the algorithm to prioritize sensors as it constructs the sampling schedule.

To increase the number of sensors that may be sampled under the energy constraints, we limit the number of samples required for a sensor to determine its metric value for each user stride. This will enable sampling of other sensors at other epochs. The training dataset is used to deduce the sampling epochs, $distr_m(s_i)$, over which s_i's metric value is most likely to be observed and are known as the epochs corresponding to its metric distribution. If scheduled, s_i is sampled only at those epochs, and its metric value is deduced from the resulting samples.

To fulfill the second requirement of predicting atypical metric observations as early as possible, the algorithm separates the sampling epochs over a stride into sections to be evaluated independently. Sensors whose metric distribution epochs lie in the earliest section of a stride are evaluated first, and so on. Although sensors are evaluated in the order of their feature weights, independent evaluation of each section allows for lower feature weight sensors to be assigned to the sampling schedule before higher feature weight sensors, if their metric distributions correspond to an earlier section. This tradeoff enables early prediction of atypical metric observations, if the sampled sensors in the earlier sections produce semantic outliers.

3.4.2 Early Semantic Outlier Detector Scheduling Algorithm

Given that sampling sections are evaluated individually, there arises a need to apportion the energy constraint among the sections. To achieve this, each section is granted a portion of the energy budget that is proportional to the level of contention observed in the section. The level of contention of each sampling epoch in a section is determined based on the feature weights of sensors vying for that epoch. A section's average contention is represented by the average contention across epochs in a section. The ratio between the average contention of the sections determines apportionment of the energy budget. This notion of contention is central to the construction of the sampling schedule in the following sense: assignment of one sensor to the sampling schedule reduces the likelihood some get assigned (or may preclude them entirely). Therefore, higher contention sections must be budgeted more samples, and consequently more energy.

The ordering of sample selection based on sections dictates that there be no sensors whose metric whose distribution lies in 2 or more sections. If this must happen, we exclude such sensors from assignment. As a result, the section boundaries are chosen in such a manner to minimize total feature weight of sensors that are excluded in this manner. As it turns out, the bio-mechanical notion of phases of a stride naturally lends itself to the notion of epoch sections over a stride - "loading response" is the phase when the heel touches the ground, "mid-stance" when the foot is grounded, and "terminal stance" when the weight shifts to the ball of the foot and the toes. Consequently, we set the number of sections to 3 and select the section boundaries by starting them at estimates of the boundaries between the each bio-mechanical phase of a stride, and then performing a line search to locate section boundaries wherein the sum of the feature weights of excluded sensors is minimum.

The sample schedule construction algorithm proceeds as follows. Sections are evaluated in sequence from earliest to latest. Over each section, sensors whose metric distributions the section includes are evaluated in decreasing order of their feature weights. The assignment of a sensor, s_i, subsumes the assignment of the other sensors s_j, whose metric values will be used to estimate that of s_i. This set of sensors, $assignments$, for a given s_i is computed

based on their ability to predict s_i's metric value well. The algorithm moves on to the next section, either when all sensors corresponding to the current section have been assigned, or if the remaining energy budget for the section no-longer allows the assignment and estimation of any sensor's metric values.

Algorithm 4 outlines the multiple regression algorithm used to arrive at the set of sensors,

Algorithm 4 Multiple Regression for Semantic Outlier Detection

1: Add s_i to $assignments$
2: Exclude sensors that can no longer be assigned
3: $val(s_j) \leftarrow 0 \; \forall$ sensors s_j
4: $selSet \leftarrow \{s_j | s_j \in assignments - \{s_i\} \wedge max(distr_m(s_j)) \leq max(distr_m(s_i))\}$
5: $assgnDone \leftarrow False$
6: **while** True **do**
7: **if** $selSet = NULL$ **then**
8: **if** $assgnDone = False$ **then**
9: $selSet \leftarrow \{s_j | s_j \notin assignments \wedge max(distr_m(s_j)) \leq max(distr_m(s_i))\}$
10: $assgnDone \leftarrow True$
11: **else**
12: Return NULL
13: **end if**
14: **end if**
15: **for all** $s_j \in selSet$ **do**
16: $val(s_j) \leftarrow$ (reduction in $R_{sq}(s_i)$ if feature at s_j is included as an independent variable)
17: **if** $val(s_j) < \epsilon$ **then**
18: $val(s_j) \leftarrow 0$
19: continue
20: **end if**
21: **end for**

22: Select $s_k = s_j$ s.t. $val(s_k) = \max_j(val(s_j))$

23: **if** $val(s_k) = 0$ **then**

24: Return NULL

25: **end if**

26: Add s_k to *assignments*

27: Exclude sensors that can no longer be assigned

28: **if** $R_{sq}(s_i)$ based on the independent variables selected this far $\geq R_{sq_{min}}$ **then**

29: Return *assignments*

30: **end if**

31: **end while**

assignments, that will be used to estimate the currently evaluated s_i's metric value. The algorithm proceeds in two passes over the sensors available to predict s_i's metric value. In the first pass, when *assgnDone* is $False$, only those sensors are considered to predict s_i's metric value which have already been selected for sampling themselves (line 4). These sensors have been selected in the current or earlier sections, to have their metric value predicted or to predict others' metric value. Once this set has been exhausted, the set sensors that have not yet been selected are considered to improve the prediction of s_i's metric value (lines 7-11). In either pass, to enable early prediction the algorithm estimates s_i's metric values based only on that of sensors whose $distr_m$ occur at or before s_i's (lines 4 and 9).

In each pass, the algorithm iteratively evaluates each sensor that may be considered for the pass, for the estimation of s_i's metric values and assigns a heuristic values to its possible assignment. The sensor with the highest heuristic value is then selected for assignment at the end of an iteration (lines 22-27). A sensor's heuristic value considers the extent to which it improves the coefficient of determination (R_{sq}) in predicting s_i's metric value. Further, a sensor is only considered as an independent variable for the estimation, if it contributes to the improvement of the estimate by a minimum threshold value $R_{sq_{min}}$ (lines 17-20). This enables the production of robust regression estimates.

Finally, the algorithm returns with the assignments made, once a minimum requisite

coefficient of determination has been achieved in the estimation of s_i's metric values (lines 28-30). No assignments are made if this quality of estimation cannot be achieved, or if the energy budget has been depleted (lines 11-13, 23-25). As stated earlier, in the latter case the sensor with the next highest value after s_i, is considered for prediction.

3.5 Results

We validate our adaptive algorithm on six experimentally collected datasets of foot plantar pressure measurements, each comprising of a few hundred footsteps. The datasets correspond to different subjects and were collected with the HERMES platform. Each dataset was divided into a training subset comprised of 80% of the footsteps, all of them typical, with the remaining 20% used for testing. Half of testing subset consisted of typical steps, where the subject walked forward normally, and the other half of atypical steps where the subject shuffled her/his feet. We compared performance to the compressive sensing technique described in chapter 2. To keep the results comparable, the energy budgets for our algorithm were based on a maximum number of samples given by 5 times the number of sampling epochs over a footstep. In our evaluations, ϵ was set to 0.05 and $R_{sq_{min}}$ was set to 0.95.

Fig. 3.2 compares the metric accuracy of our algorithm over semantically atypical steps when the threshold number of outlier detections required to categorize a step as atypical is set to 2, to the metric accuracy in chapter 2 under similar energy constraints. Clearly, our proposed strategy produces substantially more accurate metric estimates for atypical footsteps over 6 all datasets (increase of between 57% and 62% in prediction accuracy). The error that remains, arises from the delay in categorizing the step as semantically atypical.

Fig. 3.3 compares the accuracy that can be achieved for atypical steps, to the change in energy costs which arise when the threshold number of outlier detections is varied. This comparison of energy costs to that in chapter 2, however, has been made over the entire testing set rather than over semantically atypical steps only. Here, with a small threshold, the average energy costs over the testing set increases due to larger sampling and transmission

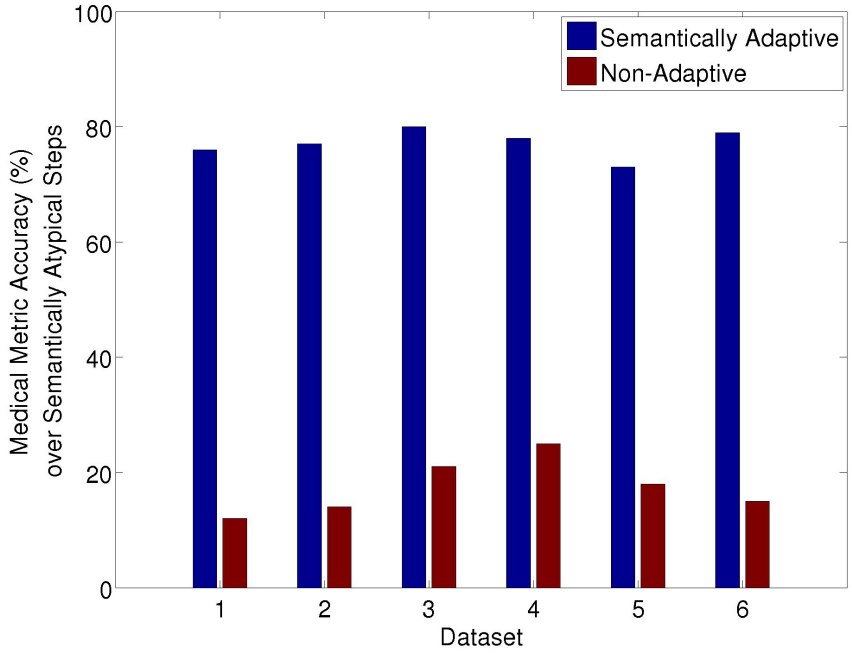


Figure 3.2: Accuracy of Average Max. Pressure Metric estimates, averaged over semantically atypical steps in each dataset, with and without the proposed adaptive sampling algorithm.

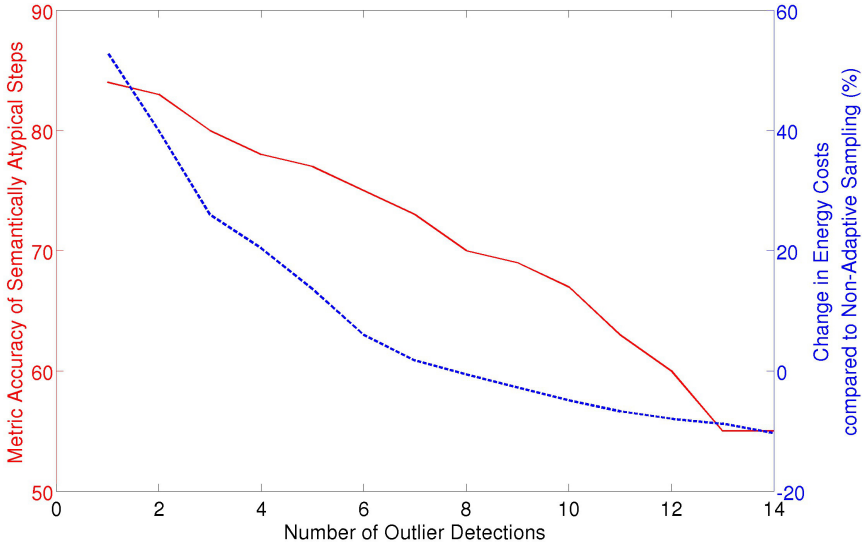


Figure 3.3: Average change in energy costs and metric accuracy for semantically atypical steps, as the number of outlier detections necessary to classify a step as atypical is changed. Values are averaged over all datasets.

related energy expenditure over atypical steps, and a similar increase over mis-classified typical steps. At the same time, this higher false positive rate also allows for a higher detection rate resulting in high average accuracy. In contrast, large thresholds result in decreased energy costs due to non-transmission over correctly classified typical steps. This low false positive rate provides for a higher false negative rate and a related reduction in average accuracy. Depending on the desired increase in accuracy, a net decrease in energy costs of up to 10% was achievable in comparison to an already efficient compressive sensing algorithm while still achieving more accurate metric estimates over atypical steps.

CHAPTER 4

Low-Power Operation while Enabling Semantic Faults Detection

4.1 Introduction

Limiting power consumption of BANs to improve usability and maintainability is important. However, foremost among the requirements for mobile monitoring systems is the need to provide high-quality measurements to a medical practitioner. Here, data quality is based on the system's ability to continually and accurately estimate medical diagnostic metrics and the signal features, or metrics, that they depend upon. In order to accurately estimate said metrics while keeping the power/energy usage to a minimum, duty-cycling, data compression and sub-sampling techniques have been proposed that also account for diagnostic metric accuracy [NBG11][WP11]. As discussed in chapter 1, an undesirable consequence of such techniques is the increased risk of mis-diagnosis / mis-estimation of the patient's condition owing to the increased sensitivity of the medical diagnostic metrics to sensor faults due to the reduced redundancy in the collected data.

Therefore, there is a need to design BANs that can produce accurate estimates of medical diagnostic metrics while keeping system power consumption to a minimum and simultaneously detect when the signal metric estimate is based on faulty observations. In this chapter, we propose a Singular Value Decomposition (SVD) based technique to select a minimal subset of physiological sensors that can collectively achieve requisite accuracy levels in diagnostic metric estimation, thereby minimizing application bandwidth usage and power consumption.

Furthermore, it aids a step-wise regression technique, which also admits a minimal in-

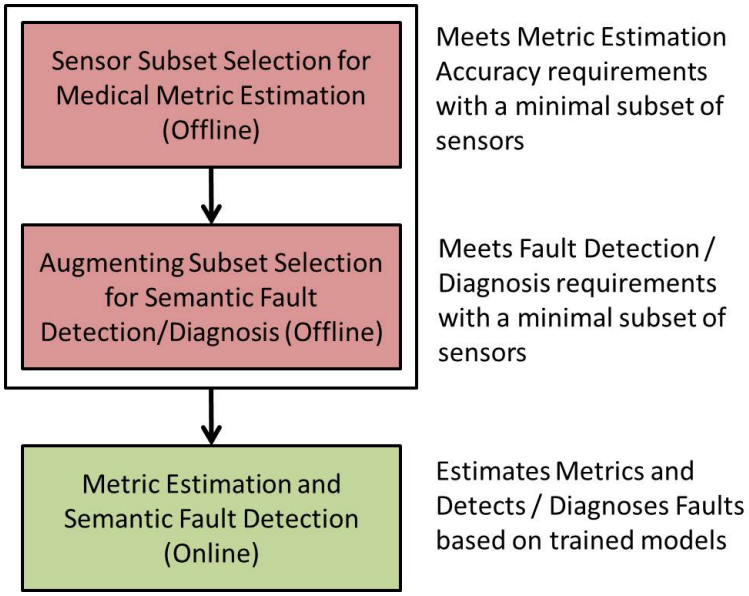


Figure 4.1: Schematic of Proposed Low-Power Medical Metric Estimation and Semantic Fault Detection Techniques.

crease to the application bandwidth and power, while incorporating the ability to detect and diagnose semantic faults in the sensors that contribute to the diagnostic metric estimates. Only semantic faults are of interest to the medical practitioner, and it is therefore sufficient to strictly target such faults. In focusing our optimization on semantic information in the signal, i.e. features/metrics of the sampled physiological signal that are relevant to diagnostic metric estimation, our techniques curtail the increase in power consumption necessary to achieve desired accuracy levels with fault diagnosability. We illustrate the set of algorithms that we shall describe, along with their purpose, in Fig. 4.1.

We validate these algorithms on eight experimentally collected datasets of foot plantar pressure measurements for gait stability monitoring. The datasets correspond to different subjects and were collected with the HERMES platform. The validation investigates the ability of our proposed techniques to accurately estimate various functional test metrics from the GARS-M and Tinetti gait stability assessment scales, and diagnose faulty estimates therein, while accomplishing this with a minimal subset of sensors thereby yielding substantial application bandwidth and power reductions.

4.2 Related Work

A number of signal-processing techniques have been proposed in the domain of WSNs to limit bandwidth requirements and reduce power/energy consumption. Data compression techniques, such as data aggregation, pipelined and distributed compression, reduce the amount data as it traverses the network from source to sink, often limiting the reduction in information content of the data after compression, thereby reducing the energy and bandwidth consumed in communication while maintaining data fidelity [FRW07][SGY11]. Sensor coverage techniques limit the number of sensors active simultaneously in spatially dense networks, while ensuring that the area "covered" by the selected sensors matches the target area without spatial discontinuity. Duty cycling, a form of sensor subsampling, limits the amount of time each sensor node is active, thereby reducing energy consumption [BWC13]. Similar signal processing techniques have crossed over to the domain of BANs [NBG11][HLR10]. However, these techniques focus on reconstruction and semantic accuracy of an aggregate global signal rather than spatial fidelity of the signal. The authors of [KMG08] offered a sensor subset selection algorithm to minimize the loss in fidelity of arbitrary functions of the spatial signal sampled by a WSN, under constraints on the number of simultaneously active sensors. The solution was shown to approximate the global minimum when the function exhibits sub-modularity.

A host of fault-tolerance techniques have been put forth in the domain of WSNs as well. In general, these use voting-based mechanisms to detect spatial outliers. However, a distinction can be made based on whether a WSN's goal is to detect an event and localize it, or achieve a spatio-temporal map of a physical signal. In the former case, each sensor makes a binary decision that must be voted on in its neighborhood. In the latter case, one of many proposed parametric, non-parametric or bayesian techniques must first be applied to model the relationships between sensors and weed out outliers (by majority vote) before computing a fault-tolerant signal aggregate over a neighborhood [ZMH10]. Whereas these traditional WSN applications capitalize on a dense deployment that offers ample spatial redundancy to achieve fault tolerance, BANs can neither afford such levels of deployment density, nor can

local correlations be assumed to always exist.

Nevertheless, algorithms have been proposed for fault tolerance in BANs under certain restrictions. The authors of [LLN12] propose a collaborative data quality estimation algorithm wherein independent estimates of data quality by each sensor are combined by a vote. A fault detection technique for BANs was described in [KP11] wherein the relationship between sensors was modeled to detect outliers, although that technique was limited to accelerometer measurements. To the best of our knowledge, we are the first to put forth a general technique in the context of BANs, to simultaneously minimize loss in semantic fidelity and be able to detect and diagnose semantic faults, under constraints on the bandwidth consumption or power usage.

4.3 Preliminaries

4.3.1 Health Monitoring System Model

Signal processing techniques are commonly applied to extract signal features/metrics from raw BAN sensor samples, as it vastly improves the informational content of the data points used to reason about the aggregate system output. Fig. 4.2 illustrates a common BAN processing configuration, where several sensors measure physiological signals at their location. The signal measurements are then wirelessly transmitted to a less resource-constrained base station, e.g. a smart phone. Relevant metrics, which are often computationally expensive to derive (translating to high power usage), are extracted and post-processed at the base station to produce an aggregate diagnostic metric estimate. Finally, this estimate is dispatched to a data center for logging into the patient's medical records. It is also possible for the BAN to include multiple platforms, each with a set of sensors, connected wirelessly to the base station. We submit that by activating only a subset of sensor in the BAN, power and bandwidth utilization by the sensing platform(s) is reduced. Further, applying the fault-detection algorithm at the base station allows the system to determine the validity of the signal metric set and the resulting diagnostic metric estimate in a manner that is frugal with

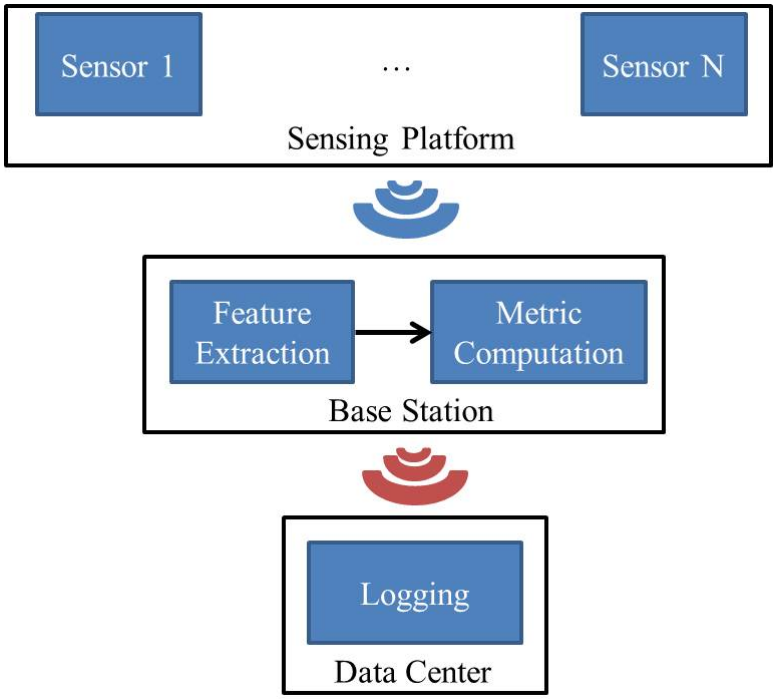


Figure 4.2: Model of Body Area Network (BAN) under study.

communication resources.

4.3.2 Sensor Fault Model

Sensor faults occur due to a variety of reasons such as sensor aging, mis-calibration, random noise, etc. While it is difficult to determine the cause of a fault, a few types of faults are commonly observed with sensors. These include short faults, where anomalous readings are present for a short period, constant faults, where the sensor reading is stuck at an anomalous value for some duration, long faults related to long duration noisy measurements, and offset faults, wherein the reading consistently differs from the true value by a fixed amount [SGG10]. Drift and transformation faults have also been observed where the true value is consistently transformed by a function before it is presented as the sensor's reading [KPS03].

In furtherance of our goal of semantic accuracy, we take a semantic-oriented view of faults. First, this means that we solely focus on faults that affect extracted signal metrics and treat other faults, such as those that occur outside of the signal segment of interest, as

irrelevant. Second, we allow the device designer to define how far a signal metric may deviate from its true value before it must be classified as a fault. Towards this end, we focus on a specific type of drift fault while training the statistical models underlying the fault-detection algorithm. The models are trained to estimate each sensor's signal metrics to within T% of the true value with high confidence. If the sensor's observation of a signal metric lies outside of the prediction bounds, it is marked inconsistent. Short, constant or long faults may be subsumed by our fault model, assuming the faulty readings lie outside the prediction bounds. Observations from faulty sensors that happen to be within an acceptable range of the true value are assumed to pose no threat to metric accuracy, and can be safely overlooked.

4.3.3 Assumptions

Some amount of redundancy must exist across the sensors of the monitoring system in order to achieve accurate estimates of diagnostic metrics with a minimal subset of sensors, and to detect faulty estimates. We assume that each sensor's signal metric measurement is approximately linearly dependent on that of some subset of all sensors, not including itself. In other words, given a set of N sensors, the metric p at sensor j can be closely estimated as a linear combination of the values for p at some subset of sensors 1 thru $j-1$ and $j+1$ thru N. In mathematical form,

$$y_{jp} \approx \sum_{k=[1,N]\setminus j} (w_{pk} \times y_{kp}) \quad \forall p, j \tag{4.1}$$

where y_{kp} and w_{pk} are the signal metric values and linear coefficients, respectively.

Also, for training to produce robust and meaningful output, we assume that the training datasets are free of faulty measurements. In sections 4.4 and 4.5, we describe the sensor subset selection algorithms for the case where the medical metric is computed based on a single metric p at all sensors. In section 4.6, we outline generalizations to these algorithms to support cases where multiple diagnostic metrics are to be computed from different signal metrics.

4.4 Sensor Subset Selection For Accurate Metric Estimation

We would like to use the fewest possible sensors in estimating the medical diagnostic metric to within an accuracy threshold of ϵ. Note that the diagnostic metric is an aggregate function over the signal metric values at all sensors. To achieve high-fidelity estimate of the diagnostic metric with the fewest sensors, we formulate the problem as the need for the smallest subset S of sensors, which we shall refer to as the sensing subset, that sufficiently explains the signal metric values of the sensors not in S. We will use the signal metric values of the sensors in S to estimate those of the sensors not in S, all of which will be combined while computing the diagnostic metric estimate. In the absence of faults, the computed diagnostic metric value is an estimate as it is based on the true signal metric values for some sensors and estimates for others. Formulated in this manner, the best subset S of a given cardinality is the one that will provide the closest estimates for sensors not in S. Furthermore, as we shall see in the following section, by capitalizing on the dependencies between signal metric values of different sensors, this formulation also facilitates bandwidth-limited, low-power and fault-tolerant coverage of the sensors in S.

Singular Value Decomposition (SVD) is a technique that factorizes a matrix E into the product of a unitary matrix U, a diagonal matrix Σ, and the transpose of another unitary matrix V. The diagonal values σ_i of Σ are known as the singular values of E and offer a unique property via the Eckhart-Young theorem. Given that the rank of a matrix is the number of independent columns in it, this theorem states that using the r largest singular values in Σ, one can construct an approximation of E with rank r, which we refer to as $\tilde{E}$. Furthermore, this is a least-squares approximation that minimizes $\sum_{t,j}(\tilde{e}_{tj} - e_{tj})^2$, where e_{tj} and $\tilde{e}_{tj}$ are elements of E and $\tilde{E}$, respectively.

Given a training set of signal values at each sensor, let Y be a matrix where column y_j correspond to the sensor j's values at different instances. Algorithm 5 outlines the sensing subset selection procedure. It starts out by standardizing each sensor's value in Y (line 1). Next, it incrementally builds the sensing subset, S, by adding that sensor at each iteration of the outer loop (lines 3-23), whose inclusion into S most improves the explanation of

signal metric values for sensors not in S. To accomplish this, each sensor, k, not in S is independently evaluated for inclusion into S, in the inner loop (lines 5-12).

Here, the columns of Y corresponding to the sensors in the subset under evaluation, $(S \cup k)$, are skewed by a large multiplicative factor, Φ (line 7). The resulting matrix, $\tilde{Y}$, is factorized and approximated to rank $|S| + 1$ (lines 8-10), where S is the sensing subset constructed thus far. By weighing the values of the sensing subset under evaluation with a large multiplicative factor, the SVD based approximation of Y is biased to minimize their error. A byproduct is that the least-squares approximation error for sensors not in the subset under evaluation is reduced to the extent that it linearly depends on the values of subset under evaluation. The sensor that most improves the estimation error in this manner is added to S (lines 13-14).

All errors in Algorithm 5 are evaluated by the L2-norm, i.e. with the Root Mean Squares

Algorithm 5 Sensing Subset Selection

Input: Input: Y, training dataset; ϵ, requisite error bound; $F()$, function that computes diagnostic metric estimates from signal metrics

Output: S, sensing subset of sensors to be used

1: $y_i \leftarrow \frac{y_i - MEAN(y_i)}{STDDEV(y_i)}$ $\forall i \in [1, N]$
2: $S \leftarrow NULL$
3: **for** $j \in [1, N]$ **do**
4: $\quad er_k \leftarrow \infty$ $\forall k \in [1, N]$
5: $\quad$ **for** $k \in [1, N] \backslash S$ **do**
6: $\quad\quad \tilde{Y} \leftarrow Y$
7: $\quad\quad \tilde{y_i} \leftarrow \tilde{y_i} \times \Phi$ $\forall i \in S \cup \{k\}$
8: $\quad\quad [U, \Sigma, V^T] \leftarrow SVD(\tilde{Y})$
9: $\quad\quad$ Zero out the smallest $(N - j)$ diagonal elements in Σ
10: $\quad\quad \tilde{Y} \leftarrow U \Sigma V^T$
11: $\quad\quad er_k \leftarrow RMS(\tilde{y_i} - y_i$ $\forall i \in [1, N] \backslash (S \cup \{k\}))$
12: $\quad$ **end for**

13: $p \leftarrow \{i | er_i \leq er_k \ \forall k \in [1, N]\}$

14: $S \leftarrow S \cup \{p\}$

15: $X \leftarrow Y$

16: **for** $k \in [1, N] \backslash S$ **do**

17: $x_k \leftarrow$ linear regression estimate for signal metric value of k based on signal metric values of sensors in S

18: **end for**

19: $e \leftarrow RMS(F(X) - F(Y))$

20: **if** $e \leq \epsilon$ **then**

21: Return S

22: **end if**

23: **end for**

24: Return NULL

(RMS) function. We also note that the algorithm assumes that the metric computation function, F, weighs each sensor's signal metric equally. However, it can easily be adapted for the case of a weighted aggregation, but applying similar weights to the columns of Y after standardization.

4.5 Sensor Subset Augmentation For Fault Detection

We achieve the ability to detect semantic faults with an additional subset of sensors that are selected to boost the redundancy of the signal metrics of the selected sensing subset. This redundancy will be exploited to glean inconsistencies in the sensing subsets' observations of true signal metric values, and ultimately used to detect faulty observations. We will refer to this additional subset of sensors as the augmenting subset. Note that the need for redundancy is also inherent in our formulation of the sensing subset selection problem. In other words, the larger the subset of sensors, Z (not in S), that a sensor, s, shares mutual information with, the more likely s is to get selected into the sensing subset. Consequently, it is also more likely that s's signal metric values will be well estimated by some subset of

sensors in Z.

Assuming s is selected into the sensing subset, faulty observations of its signal metric value may then be detected by cross-checking them with independent estimates by a subset of Z. However, if such a validation results in an inconsistency, although it is likely that a fault has occurred, it is unclear if the fault lies at s or at one or more sensors in Z. Depending on the application and deployment settings, fault detection may be sufficient to warrant maintenance of the system, including testing the sensors and replacing/recalibrating the faulty sensor(s). Alternatively, it may be necessary that the fault be diagnosed and the faulty sensor(s) be identified. However, a tradeoff emerges that demands more sensors for fault diagnosis than for fault detection.

Theorem 1. *At most f faulty sensors can be identified if each sensor is accurately estimated by $f + 1$ disjoint subsets of sensors excluding itself.*

In other words, if we limit ourselves to identifying at most 2 faulty sensors, then s can be identified as a faulty sensor if it is characterized as such by 3 disjoint subsets of sensors, each of which do not include s. The theorem assumes that all estimations are accurate such that the signal metric value of a faulty sensor will be inconsistent with all its estimates, including those of other faulty sensors. That is, there will be no collusion between faulty sensors. Estimation accuracy will be discussed in detail in the following section. Furthermore, in the worst case, an accurate sensor will be inconsistent with f estimates wherein one faulty sensor exists in each of its estimating subsets. Yet, there is still one estimate that is consistent leading to the sensor being identified as accurate. This rationale can be wielded to prove Theorem 1 by contradiction.

Proof of Theorem 1. Case 1: Validity of Accurate Sensor

For an accurate sensor, s, to be agreed upon as invalid, all $f + 1$ disjoint subsets that estimate it must deem it invalid. Since each estimating subset that does not itself include a faulty sensor produces an accurate estimate of s, this must mean that each of the $f + 1$ estimating subsets includes at least one faulty sensor given that they are disjoint. But we have assumed that there are at most f faulty sensors leading to a contradiction.

Case 2: Invalidity of Faulty Sensor

For a faulty sensor, t, to be agreed upon as valid, at least 1 of the $f+1$ disjoint subsets that estimate it must deem it valid. Each estimating subset that does not itself include a faulty sensor must produce an accurate estimate of the the true signal metric value at t. Therefore, all of these subsets must deem t's measurement as invalid. Further, estimating subsets that do include a faulty sensor are assumed not to collude with t's faulty measurement and may produce erroneous estimates of t's true signal metric value, but ones that do no agree with t's faulty estimate. Since all estimating subsets for t disagree with it's signal metric measurement, there is contradiction with the requirement that at least 1 of the estimating subsets deem t's measurement as valid.

□

Clearly, the larger f is, the larger the augmenting subset is likely to be. Furthermore, if faults must be identified for all selected sensors (sensing and augmenting subsets), constructing a small augmenting subset towards our stated goal of bandwidth-limited, low-power operation is likely to become intractable. We contain this problem by employing Theorem 1 to construct an augmenting set wherein each sensor only in the sensing subset is accurately estimated by $f + 1$ disjoint subsets of the augmenting subset. This will enable diagnosis of faults in the sensing subset and is sufficient for metric accuracy in the following sense: if the faulty sensor lies in the sensing set, the metric estimate is unusable; But faults in the augmenting subset do not affect metric estimates, and can be ignored. Of course, any inconsistent estimate should signal the need for system maintenance and repair. But until this can happen, assuming at most f faulty sensors, the system will continue to provide accurate metric estimates, when available. Under these considerations, our objective is to select the smallest augmenting subset that collectively and accurately estimates the signal metric values of each sensor in the sensing subset while being able to identify at most f faulty sensors in the sensing subset.

4.5.1 Selecting an Augmenting Subset for One Sensor

We begin with the case where the sensing subset contains a single sensor and f is zero, which is taken to mean that fault detection is required, but not diagnosis. The solution will be generalized in the following section. As stated in section 4.3.3, we take a semantic-oriented view of faults as defined by the device designer. As we will see, a more restrictive definition demands more bandwidth and power.

Allowing the designer to provide the least restrictive definition of a fault will, therefore,

Algorithm 6 Augmenting Subset Selection for Single Sensor

Input: Y, training dataset; j, sensor whose feature values must be estimated; p_{rem}, threshold p-value to remove a sensor from the augmenting subset; T, threshold prediction error bound; *init*, an initial augmenting subset; D, sensors to exclude from augmenting subset

Output: A, augmenting subset of sensors to be used

1: $A \leftarrow NULL$
2: $init \leftarrow init \backslash (\{j\} \cup D)$
3: **while** $A \neq init$ **do**
4: $\quad A \leftarrow init$
5: $\quad init \leftarrow removeSensor(Y, j, A, p_{rem})$
6: **end while**
7: **repeat**
8: $\quad a_{max} \leftarrow PB$ for $R(y_j, y_A)$
9: $\quad$ **while** $a_{max} > 2T$ **do**
10: $\quad\quad k_{add} \leftarrow 0$
11: $\quad\quad$ **for** $k \in [1, N] \backslash (A \cup \{j\} \cup D)$ **do**
12: $\quad\quad\quad a_{curr} \leftarrow PB$ for $R(y_j, y_{A \cup k})$
13: $\quad\quad\quad$ **if** $(a_{curr} - a_{max}) > 0$ **then**
14: $\quad\quad\quad\quad a_{max} \leftarrow a_{curr}$
15: $\quad\quad\quad\quad k_{add} \leftarrow k$
16: $\quad\quad\quad$ **end if**

end for

if $k_{add} = 0$ **then**

Break

end if

$A \leftarrow A \cup \{k_{add}\}$

end while

$A_{prev} \leftarrow A$

$A \leftarrow removeSensor(Y, j, A_{prev}, p_{rem})$

if $A_{prev} = A$ **then**

Break

end if

until maximum iteration count

Return A

proc $removeSensor(Y, j, A, p_{rem})$

$r_{max} \leftarrow p_{rem}$

$k_{rem} \leftarrow 0$

for $k \in A$ **do**

$r_{curr} \leftarrow$ p-value for F-test under null-hypothesis that k has 0 coefficient in $R(y_j, y_A)$

if $r_{curr} > r_{max}$ **then**

$k_{rem} \leftarrow k$

$r_{max} \leftarrow r_{curr}$

end if

end for

$A \leftarrow A \backslash \{k_{rem}\}$

Return A

lower the bandwidth and power requirements. Based on our assumption of linear dependence between sensor metric values, the value of each sensor in the sensing subset will be corroborated by linear regression estimates from sensors in the augmenting subset. A fault

is then defined based on the prediction bound for the linear regression estimate over the training dataset. The prediction bound PB is defined such that:

$$probability(R(y_s, y_A) - PB(y_s, y_A) \leq y_s^* \leq R(y_s, y_A) + PB(y_s, y_A)) = 1 - \alpha \tag{4.2}$$

for a new signal metric observation y_s^* at some sensor s in the sensing subset whose value is estimated by the signal metric values at some sensors in the augmenting subset, A. The probability that the new value observed at s exceeds the regression estimate, based on A, by more than PB is bounded by α. The prediction bound is different from the confidence bound which is based on confidence in the regression parameter estimates rather than confidence in the predicted value itself.

In our formulation, the designer defines the maximum percentage T by which the observed value can deviate from the true signal metric value before it is deemed inconsistent with confidence $1 - \alpha$. This enforces the maximum value of PB to $2T$. Algorithm 6 describes a step-wise regression procedure to select a minimal augmenting subset for sensor j in the sensing subset. It achieves this via multiple regression, $R()$, between feature values at j and potential sensors in the augmenting subset (lines 8, 12, 33). Sensors outside of the augmenting subset are sequentially evaluated for addition into it (lines 11-17), and at each iteration, the sensor that contributes most to the improvement in the prediction bound is added (line 21). This continues until the target prediction bound is met, or if no sensor improves the prediction bound (lines 9, 18-20).

The algorithm also prevents over-fitting by evaluating each sensor in the augmenting subset via a statistical F-test, rejecting the ones most likely to result in over-fitting (lines 32-39). When no sensor in the augmenting subset passes the F-test and the requisite prediction bound is met, the algorithm will have converged.

Algorithm 7 Augmenting Subset Selection for Multiple Sensor

Input: Y; p_{rem}; T; S, sensing subset; f, fault diag. threshold Output: A, augmenting subset of sensors to be used

1: $A \leftarrow NULL$
2: $A_{max} \leftarrow NULL$
3: $P_{j,k} \leftarrow NULL \ \forall j \in S, k \in [0, f]$
4: **for** $k \in [0, f]$ **do**
5: $C \leftarrow NULL$
6: **for** $i \in [1, |S|]$ **do**
7: $j_{max} \leftarrow 0$
8: **for** $j \in S \backslash C$ **do**
9: $D \leftarrow (P_{j,m} \ \forall m \in [0, k])$
10: $A_{curr} \leftarrow ALGORITHM6(Y, j, p_{rem}, T, A \cup S, D)$
11: **if** $CARD(A_{max} \cup S) < CARD(A_{curr} \cup A \cup S)$ or $j_{max} = 0$ **then**
12: $A_{max} \leftarrow A_{curr} \cup A$
13: $j_{max} \leftarrow j$
14: $\check{P} \leftarrow A_{curr}$
15: **end if**
16: **end for**
17: $A \leftarrow A_{max}$
18: $P_{j_{max},k} \leftarrow \check{P}$
19: $C \leftarrow C \cup \{j\}$
20: **end for**
21: **end for**
22: Return A

4.5.2 Selecting a Common Augmenting Subset for all Sensors

To generalize the algorithm from the previous section, we describe how it can be extended to support a sensing subset of cardinality greater than 1 and diagnose of a bounded number

of faults. Algorithm 7, which subsumes algorithm 6, outlines these generalizations. The first goal is achieved by incrementally building the augmenting subset by considering the requirements of each sensor in the sensing subset. The sensor that increases the cardinality, $CARD()$, of the subset of sensors to be activated (sensing and augmenting subsets) by the largest amount is considered most demanding and given highest preference (lines 11-15, 17-19). Subsequent sensors in the sensing subset are catered to in a similar order of preference. However at each iteration, the augmenting subset constructed thus far, together with the sensing subset, are offered as a "seed" of initial sensors to be considered for estimation of the sensors awaiting coverage (line 10). This yields a collaborative approach for minimal augmenting subset construction. Set C keeps track of the members of the sensing subset for which estimating subsets have been selected.

Finally, fault diagnosis for up to f faults is achieved in an outer loop, by incrementally building $f + 1$ disjoint estimation subsets for all sensors in the sensing subset (line 4). To ensure disjointness, the subsets of sensors estimating a sensor j's metric values, thus far, are excluded from consideration for each subsequent estimation subset of j, via a parameter D that combines all such estimating subsets (line 9). Her set $P_{j,k}$ denotes $(k+1)^{th}$ independent estimating subset that has been selected for sensor j in the sensing subset.

4.5.3 Online Semantic Fault Detection

The previous sections have described the offline training algorithms to determine the sensing and augmenting subsets that will be activated in the BAN. Here, we briefly outline the online algorithm that shall be used in semantic fault detection. Each sensor in the activated subset will continuously compute its signal metrics and wirelessly transmit them to the base station. At the base station, the diagnostic metric estimates are computed based on the signal metric values of the sensing subset. Before transmitting these metric estimates to the data sink however, the base station further computes $f + 1$ signal metric estimates for each sensor in the sensing subset from its estimating subset. If the measurement received from a sensor in the sensing subset is deemed inconsistent with all its estimates based on

the prediction bounds for the respective regression models, all related diagnostic metric measurements for the time segment are marked faulty before being transmitted to the data sink.

4.6 Multiple Feature Support

The training algorithms described thus far can be extended to handle the case where each sensor produces multiple features that are required to estimate one or more medical diagnostic metrics. In this case, common sensing and augmenting subsets are drawn. With regards to the sensing subset, SVDs are computed independently on the signal metric matrices for each signal metric (line 8, algorithm 5), and least-squares approximation errors are constructed jointly across all signal metrics for each evaluated sensing subset (line 11, algorithm 5). Also, the accuracy threshold for each metric is evaluated separately (line 19, algorithm 5), and convergence occurs only if thresholds for all metrics are met. Finally, the approximation errors for different signal metrics are weighed by the distance between the accuracy achieved thus far by dependent diagnostic metrics, and their corresponding accuracy thresholds (line 11, algorithm 5), giving preference to straggling metrics.

Similarly, a common augmenting set is drawn by modification to algorithm 6. Here, each sensor is evaluated for addition to the augmenting subset based on the combined improvement of prediction bounds over all features (lines 9-22, algorithm 6). Likewise, a sensor will only be considered for removal from the augmenting set, if it passes the F-test for all features (lines 32-39, algorithm 6).

4.7 Validation

We validate the proposed techniques in the context of gait stability monitoring with HERMES. Recall that reducing the number of sensors that are simultaneously active will significantly reduce both the application bandwidth and system power requirements. Here, a given point in solution space will yield a $100 \times \frac{1-|A \cup S|}{99}\%$ reduction in bandwidth required

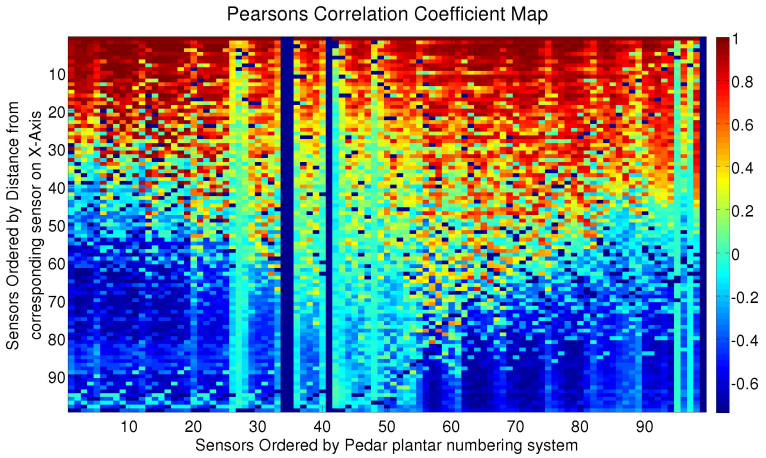


Figure 4.3: Linear Correlation Coefficients between HERMES Plantar Pressure Sensors.

(excluding bluetooth overhead). The related reduction in HERMES power usage was plotted in Fig. 1.2.

Based on a dataset of a user's plantar pressure profile collected over several strides by the HERMES platform, Fig. 4.3 illustrates a map of the linear correlation coefficients between pairs of sensors, where we observe strong linear correlations among groups of sensors. A similar map was observed over all our datasets, giving credence to our linear dependence assumption discussed in section 4.3.3.

Features extracted from the measurements of HERMES' sensors can be used to compute several functional metrics. For example, the foot contact metric of the GARS-M gait assessment scale [AM99] evaluates the heel strike angle, which can be estimated by HERMES based on the difference in time between heel strike and forefoot strike. Similarly, the staggering metric from the GARS-M scale determines laterally directed loss of balance, and can be monitored based on the plantar pressure difference between lateral areas of the foot. Other metrics that can be gleaned include the inter foot-strike interval that is relevant to the step symmetry and continuity measures of the Tinetti gait assessment tool [AM99]. Finally, the spatial average of peak plantar pressure is of interest in monitoring diabetic patients for plantar ulcers.

To validate the ability of our techniques to achieve accurate estimation of these metrics

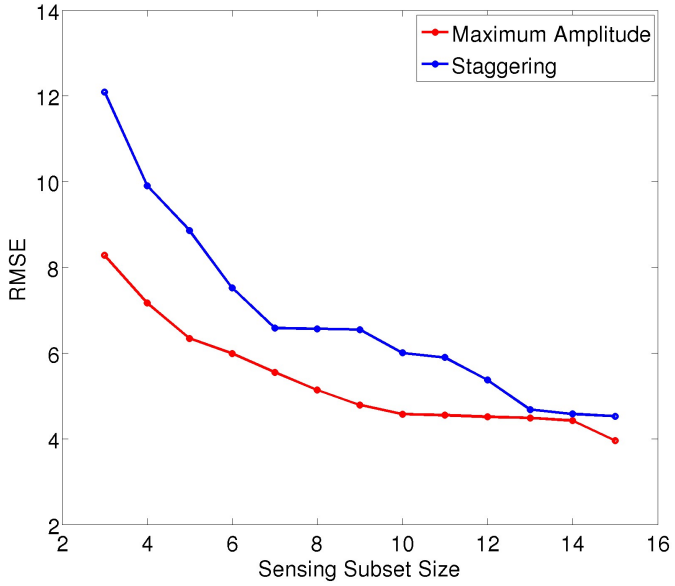


Figure 4.4: RMSE of max. amplitude related metrics for different sensing subset sizes; $T = 50\%$.

and detect faulty feature measurements involved in this estimation while activating few sensors, our algorithms were applied to several datasets collected by HERMES. Each dataset spanned a few hundred steps per subject that were collected while she/he was walking. The proposed training algorithms were used to compute minimal sensing and augmenting subsets per subjects with 80% of each dataset. The rest of each dataset was then used for testing, the outcomes for which are presented in the following sections. For all test, we set α to 0.1 and Φ to 10.

4.7.1 Metric Estimate Quality

The pressure signals from each training set were used to compute 2 signal metrics per sensor - the maximum amplitude observed at the sensor for each step and the time elapsed, since the start of a step, before pressure is observed at the sensor. The amplitude signal metric was used to estimate the maximum amplitude and staggering diagnostic metrics, while the contact time signal metric was used for the foot contact and inter-foot-strike interval (or step duration) diagnostic metrics.

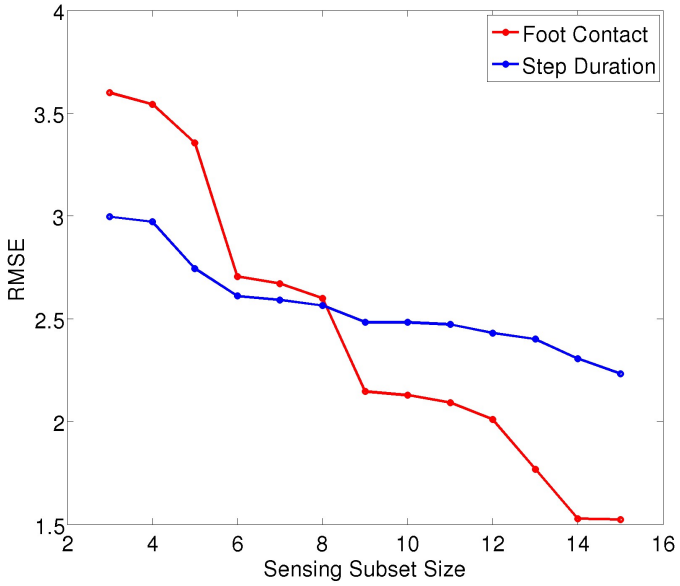


Figure 4.5: RMSE of contact time related metrics for different sensing subset sizes; $T = 50\%$.

Fig. 4.4 and 4.5 show the accuracy of these metric estimates as a function of the size of the sensing subsets, which were jointly selected for both signal metrics. The RMS error values shown are solely based on the testing sets for each dataset, and have been averaged over testing portions of all 8 of them. We observe more than a 50% reduction in semantic error between sensing subsets of size 1 and 15 for all but the step duration diagnostic metric, which observes a 20% reduction. With 15 sensors, we are using less than $1/6^{th}$ of the sensors, which would almost yields an 85% reduction in bandwidth required. This illustrates the ability of algorithm 5 to substantially increase semantic quality as bandwidth/power constraints are relaxed.

4.7.2 Augmenting Subset Selection

Fig. 4.6 plots the size of the augmenting subsets selected for various sensing subset sizes and fault diagnosability criteria. The figure distinguishes between augmenting subsets picked solely for the amplitude signal metric, solely for the contact time signal metric and jointly for both metrics. Here, we have set the signal metric estimation accuracy threshold T to 50% for all cases. As expected from our discussion based on theorem 1, diagnosing more

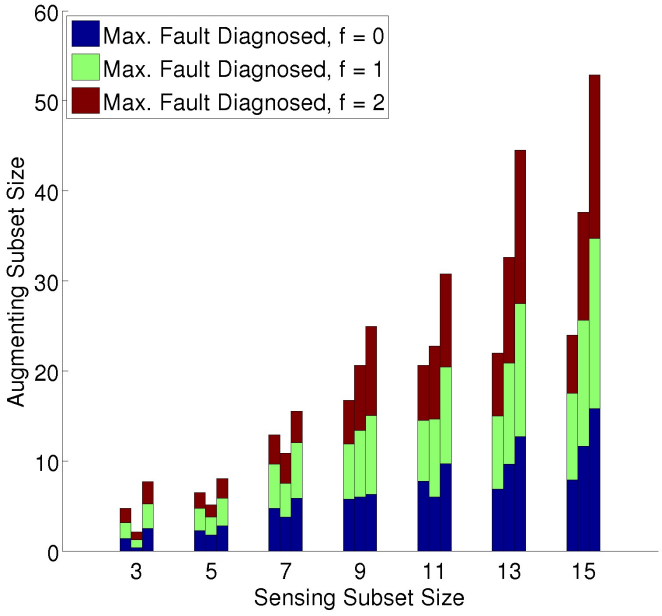


Figure 4.6: Size of augmenting sets as the sensing subset size and max. number of faults diagnosable, f, were varied. Bar groups represent max. augmenting subset sizes for max. amplitude only, contact time only, and both features combined; $T = 50\%$.

faults required more augmenting sensors. Note that the stacked bars are incremental, not cumulative.

It was also interesting to note that the contact time signal metric required fewer augmenting sensors when the sensing subset is small and the opposite was true when the sensing subset is large. This can be attributed to the manner in which algorithm 5 picks sensors to the sensing subset by minimizing aggregate errors over non-selected sensors, and to the levels of redundancy inherent in the two signal metrics. For the contact time feature, it appears that the first few sensors selected share a lot of mutual information with a few sensors that are quickly selected into the augmenting set. In contrast, the amplitude feature leads to the selection of sensors, at least initially, that share less mutual information across several other sensors, leading to a larger augmenting subset.

Finally, we observed that when the signal metric requirements are combined, a larger augmenting subset is picked. This is to be expected as the two features may not share

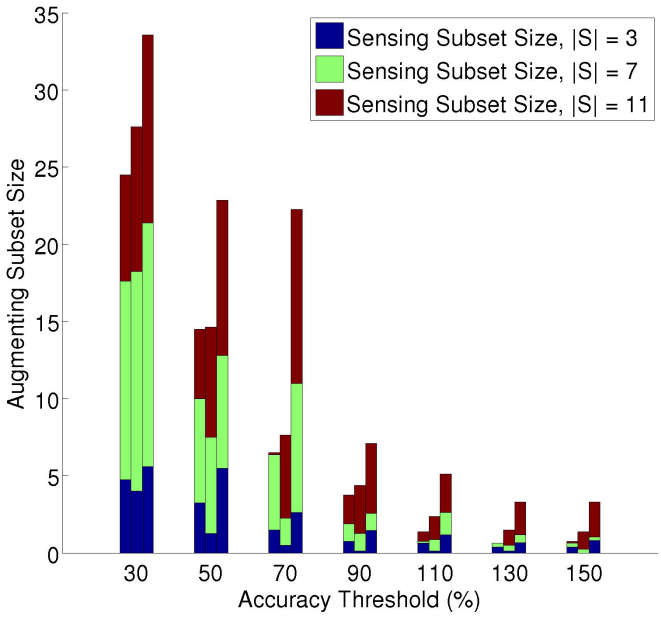


Figure 4.7: Augmenting subset sizes over a range of feature estimation accuracy requirements. Stacks represent different sensing subset sizes, groups are similar to Fig. 4.6; $f = 1$.

information in the same way across sensors. Nevertheless, the increase in the augmenting subset size was always less than 30% over the larger of the two features. Finally, depending on the desired level of accuracy, the total power (bandwidth) consumption was reduced between 28%(33%) and 77%(90%), while admitting diagnosis of up to 2 faults.

We further probed the performance of the augmenting subset selection algorithm by measuring the size of the subset as the signal metric estimation accuracy threshold, T, was varied. Performance is shown, in Fig. 4.7, for 3 sensing subset sizes. As expected, a smaller accuracy threshold results in a larger augmenting set. If the allowed variation is less than 30%, then the features of the sensing subset must be very closely estimated, leading to augmenting subsets that are almost $1/3^{rd}$ of all sensors when the sensing subset has 11 sensors. This is still more than a 47% (55%) reduction in system power (application bandwidth). Interestingly, it appears that an increasing accuracy threshold does not cause significant variation in the percentage changes across sensing subset sizes. As it turns out, there is sufficient redundancy regardless of the sensing subset size and the proposed algorithms are

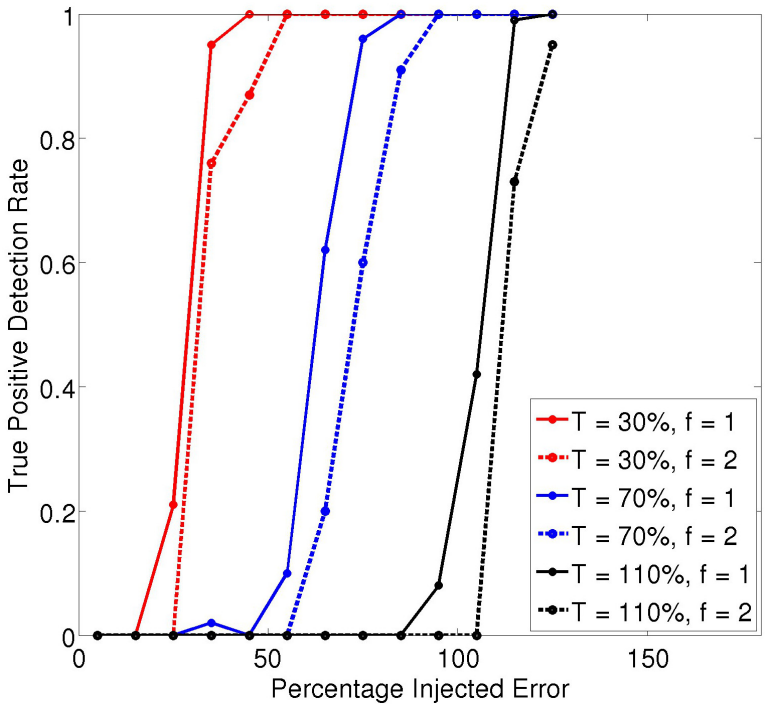


Figure 4.8: Fault-Detection performance of the selected subsets for different levels of feature errors introduced to the testing set; $|S| = 10$

able to capture this redundancy. Finally, if the signal metric requirements are combined, as before, a larger augmenting subset results.

4.7.3 Fault Detection Performance

By injecting faults in a controlled manner into the signal metrics of the testing dataset, we were able to validate the ability of our proposed techniques to detect them. Here, we set the sensing subset size to 10. A given percentage of error was introduced at each sensor in the sensing subsets (drift fault). These errors were introduced one at a time for the case of $f = 1$, and two at a time when $f = 2$. Fig. 4.8 plots the true positive detection rate averaged over all sensors in the sensing subsets of all tested datasets, as the percentage error introduced was varied.

As expected, more stringent feature estimation accuracy requirements lead to earlier fault detections. Of course, this improvement comes at the cost of a higher number of augmenting

sensors, as observed in the earlier section. The error level at which faults are detected are quite close to the trained accuracy threshold values, with some faults being detected at lower injected error rates. However, we also observe that an increase in the number of diagnosable faults to 2, produces a similar but slightly degraded result. Note that there are more simultaneous errors in this case, but this is not the reason for the reduction in performance. Instead, each fault must be simultaneously diagnosed by 3 estimating sets and this requirement causes some faults to go undetected. In contrast, we observed that the number of false positives decrease as the number of diagnosable faults is increased. However, in all cases tested, this number was less than 8% and is expected to closely relate to α.

4.8 Discussion

While we have posed a solution to the problem of erroneous functional metric estimation based on faulty sensors, a competing solution may be to use higher quality components that are built better and likely to last longer. Although, this can impact the device's cost, it may be a good solution for certain applications. For example, an EEG based application will typically use few sensors that may not correlate well with each other, and are quite susceptible to noise. Using better components here may be the way to go. On the other hand, biokinetic applications that involve acceleration, orientation and force measurements will typically exhibit good correlation across specific locations in the body and may be supported better by an approach such as ours, if the device cost must be kept low.

On a related note, our proposed approach has been shown to work well when trained for a single activity such as walking. However, it is unclear how the relationships between different sensor locations generalize across activity types. It may well be the case that the redundancy available across sensors sharply declines as multiple actions are considered simultaneously while training. A natural solution here would be to leverage proposed activity and transition detection techniques, and train models separately for each activity. It is unclear how easily these techniques can be integrated into our approach and made to abide by the low application bandwidth, low-power objective.

Finally, our work has been based on the linear dependence assumption laid out in section 4.3.3. However, non-linear/non-parametric versions of the techniques proposed herein may be used to address the case where the dependencies are non-linear. Further, our proposed solution requires customization for each user. Methods to generalize the training such that a common solution may be applied to a host of subjects would be greatly beneficial.

4.9 Conclusion

We have proposed a novel set of techniques to achieve bandwidth-sensitive and low-power yet accurate estimates of medical diagnostic metrics from mobile health monitoring systems. Our key contribution is the incorporation of semantic fault diagnosis with minimal increases to the bandwidth and power usage. Validating the performance of our approach on several plantar pressure datasets, we show that the signal metric estimation algorithm aids the fault detection technique to yield increasingly accurate estimates of multiple functional metrics with fault diagnosability, while reducing the application power consumption (bandwidth) by up to 77%(90%). Stringent fault diagnosis requirements were also shown to be met under significant power savings.

CHAPTER 5

Energy Harvesting with Dielectric Elastomers

5.1 Introduction

Major strides are being made towards limiting power consumption rates. However in the context of BANs, the desire for high quality data constrains the sampling frequency and coverage that can be afforded. Nevertheless, these developments are sufficient to close the gap between the energy requirements of BANs and the energy that may be unobtrusively harvested from various sources of human power. Given that devices powered by bulky batteries pose the risk of changing user behavior and modifying the very signals they are supposed to unobtrusively measure as well as impede usability with their need for frequent battery recharge/replacement, self-sustaining human powered sensors and BANs are gaining notice.

In furtherance of this vision of self-sustaining ubiquitous body area sensing, there is a need for high energy density harvesters that are also unobtrusive. While electro-magnetic harvesters have the potential to produce significant amounts of energy, they are bulky and composed of multiple moving components that hinder their usability and energy density. Recently, a new class of harvesters have emerged with the promise of superior transduction ability and minimal impact on user comfort. Dielectric Elastomers (DEs) are high energy-density rubber-like materials that possess the ability to behave as energy generators, actuators and sensors. Their ability to yield 5 to 40 times the energy density of piezoelectrics [KPP11] makes them more effective transducers. Their soft rubber-like nature further enables near-transparency of the harvesting sub-system to the user. Similar to piezoelectrics, they also possess the ability to behave as energy transducers, actuators and sensors.

However, while competing technologies transduce ambient energy by exploiting underlying physical phenomena (e.g. seebeck, photovoltaic and piezoelectric effects), DEs transduce mechanical energy based on an electrostatic principle. DEs are electrostatic transducers that use the input mechanical work against the electrostatic forces in the "capacitor" to vary capacitance towards transduction. Therefore, the DE "capacitor" must be charged before the mechanical work may be transduced, making its configuration and control key to the levels of energy that it may harvest.

In this chapter, we summarize the electrical and mechanical properties of DEs before describing their operation. We will also discuss our work in experimentally characterizing the mechanical behavior of DEs and mathematically modeling its electrical behavior. The resulting models are applied in the following chapter that describes configuration and control optimizations towards maximizing the output of a foot-strike powered energy harvesting system, with the ultimate goal of self-sustained operation of low-power mobile health monitoring devices such as HERMES.

5.2 Related Work

Detailed reviews of energy scavenging at miniature scale cite research that propose harvesters for a myriad of ambient energy sources including solar energy, ambient RF energy, fluid flow, air flow, vibration in walls, bridges, car engines and chassis, and jet engine housing [PS05] [MYR08]. Human powered transduction is also reviewed and spans heat transfer from the skin, vibration from foot strikes, movement of knee joints, inertia from backpacks and change in blood pressure. A wide array of human motion has been proposed for exploitation including cranking, shaking, pushing, pumping, pulling as well as the isometric forces of squeezing and pushing.

The reviews further cite recommended applications; Solar power has long been proposed as a means to power homes, charge cell phones and design self-sustaining PDAs and watches. Newer applications for energy harvesting include a variety of sensors, data-loggers and transceivers, in addition to more intrusive implantable devices for insulin administration,

real time ECG processing and pacemaking. Self-sustaining designs for structural monitoring and fine-grained temperature sensing and control have been laid out as well.

Energy harvesting mechanisms may be classified into ambient radiation, electromagnetic, electrostatic and piezoelectric transduction, among others [MYR08]. Whereas solar radiation transducers offer the most mature and impactful solutions, their placement requirements and the sporadic availability of light to BANs hampers adoptability. Electromagnetic generators are most effective and widely popular at the macro-scale, however, they are extremely obtrusive and difficult to miniaturize for the purposes of BANs. RF radiation, thermoelectric and conventional air-gap electrostatic harvesters are limited in their output capacity (μW to low mW range for BANs [PS05]) with electrostatic harvesters further hindered by their composition of moving parts. Finally, piezoelectrics have higher energy densities as compared to electrostatic and electromagnetic transducers, and are also well suited for harvesting the vibrational forces commonly experienced by BANs [MYR08].

Dielectric elastomers are a relatively new entrant to the class of miniaturized generators. An excellent survey of the material properties relevant to its transduction mechanism, various proposed transducer configurations, capabilities in comparison to other common transducers, recent applications as well as operational boundaries and lifetime issues are detailed in [KPP11]. It is noted that dielectric elastomer generators (DEGs) require charging at high voltage so they may achieve their output potential; however self-priming circuits have been proposed that use an inverse charge pump to convert some of the DEG voltage boost into charge, incrementally increasing the source voltage from 10V to the kV range [MOC10].

In [GMS10], the authors expound on the the transduction mechanism and operational modes of the material. They also derive the maximum net output and energy gains achievable for each of the operational modes at the operational boundaries of the material. As we shall see in section 5.3, although DEGs are electrostatic transducers, their material properties distinguish them from conventional transducers in this class. Owing to their superior properties, researchers have proposed novel transducer configurations to retrofit them to a host of applications initially proposed for piezoelectrics.

5.3 Dielectric Elastomer Generators

5.3.1 Material Properties

DEs are deformable yet incompressible insulating polymer films that can be built from a variety of materials. The most commonly used materials are acrylics and silicones due to their high electric permittivity (ε_r), operational boundaries, elasticity and relatively low mechanical and electrical losses. With electric permittivities 3 to 10 times that of air-gap electrostatic transducers, DEs can store and transduce more energy. They can be quite soft compared to piezoelectrics and their operational boundaries usually change when stretched, entertaining the possibility of extracting more energy from the film as it is stretched. DEs also exhibit non-linear stiffness behavior and can be viscoelastic, depending on the composite material.

As described in [KPP11], their relatively high elastic energy density means that they can store more energy when deformed for the same amount (mass and volume) of material, yielding less obtrusive and more productive transducers. They are also well suited for direct transduction as they do not need an inertial mass akin to most electrostatic transducers, making them highly efficient despite unpredictable driving forces, especially at low frequencies, both of which are characteristic of human motion.

Given their high resistive and permittive properties, DEs make for remarkable variable capacitors. Also, the fact that DEs are soft and flexible not only diminishes their mechanical complexity, but also allows for the use of micro-scale layers of compliant electrodes, such as carbon grease, simplifying fabrication and allowing for the active area to keep up with the deformation of the DE. This is in contrast to conventional electrostatic transducers that are limited by their larger rigid electrodes.

5.3.2 Transduction Mechanism

Fig. 5.1 shows the operation of a DE in the uniform biaxial transducer configuration. This configuration allows for a uniform deformation in the x and y directions when pressure is

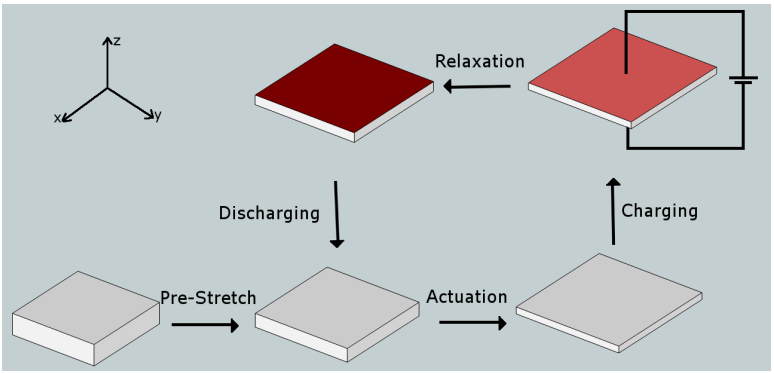


Figure 5.1: Dielectric elastomer generators transduction cycle.

applied, with a proportionate reduction in material thickness (z-direction) due to material incompressibility.

As shown in Fig. 5.1, the operation of a DE as a generator occurs in 4 stages. The DE is first stretched by the mechanical pressure applied to it in the **actuation stage**, thereby increasing its capacitance. In the **charging stage**, electrical charge is be added to the elastomer's surface at a fixed voltage V, via compliant electrodes. This generates a charge on the DE surface proportional to its capacitance.

Throughout the **transduction stage** that follows, the amount of charge Q on the DE is held constant. Here, as the applied mechanical pressure reduces to zero, the elastic forces in the DE relax and are transduced into electrostatic force. As the DE relaxes, its capacitance decreases leading to an increase in its voltage and the electrical energy it stores. This may be conceptualized as an increase in electrical energy in the film when it relaxes, as like charges on the same surface are brought close together, and opposite charges are drawn apart. The conversion continues until the material completely relaxes, or, the increased electrostatic forces are able to maintain the film in a stretched state, albeit at a lower stretch. Over the **discharging stage**, the DE is discharged, its voltage rectified and energy added to the battery. While transduction of DEs can also occur at constant voltage, it has been analytically shown that this leads to a lower amount of net energy transduced, when all other factors are held the same.

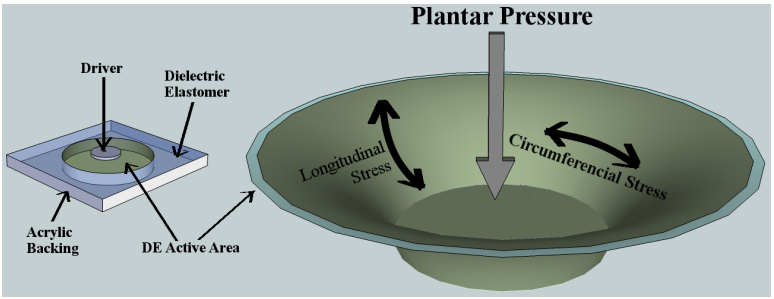


Figure 5.2: Proposed DE transducer configuration.

$$C_\lambda = \frac{\varepsilon_0 \varepsilon_r \lambda^2 A}{\frac{d}{\lambda^2}} \tag{5.1a}$$

$$= \lambda^4 C_0 \tag{5.1b}$$

$$\Delta E_{DE} = \frac{Q^2}{2C_0}\left(\frac{1}{\lambda^4_{final}} - \frac{1}{\lambda^4_{init}}\right) \tag{5.1c}$$

$$= \frac{C_{\lambda_{init}} V^2}{2}\left(\frac{\lambda^4_{init}}{\lambda^4_{final}} - 1\right) \tag{5.1d}$$

The relationship between a DEs capacitance and stretch can be characterized as follows. If the length and width of the elastomer film each increase by a factor of λ when stretched, the area will have increased by a factor of λ^2. As an incompressible material, the volume must stay constant to cause a decrease in thickness by a factor of λ^2, leading to an increase in capacitance by a factor of λ^4 (equation (5.1b)). The reason behind the high energy density of DEs is underscored by equations (5.1b) thru (5.1d), where ΔE_{DE} is the net energy output of the DE. If it fully relaxes, the electrical energy in the film will have increased to a factor of λ^4 of the input electrical energy. This is in contrast to a maximum λ^2 factor increase in energy in conventional electrostatic transducers, a limitation of their rigid structure.

Finally, we note that electrostatic force may be maintained in the film during the transduction stage, at constant voltage or constant charge. In [GMS10], the authors show that the constant charge operational mode yields larger net output in electrical energy for the same input stress. This is in contrast to air-gap electrostatic devices that are more productive in the constant voltage mode. As a result, we choose constant charge relaxation for our study.

5.4 DE Transducer Configuration and Model

Fig. 5.2 shows the transducer configuration that we propose for our application. It involves a thin acrylic layer, slightly larger in area than a pressure sensor use by HERMES, with a hole in its center to accommodate the active area. The DE film is pre-stretched and laminated onto this acrylic backing. When pressure is applied to the film via a driver component, the DE observes a deformation with a change in its area and a proportionate reduction in its thickness. A harvester will have a circular active area, 1cm in diameter, with multiple layers of DE film that are each 5mm thick and prestretched to 300% by 300%. The reason we have chosen these dimensions will be discussed in the next chapter. The harvesters are made of the 3M manufactured VBH4905 acrylic DE, the material properties for which are outlined in table 5.1 [3M11]. The table also lists relevant properties of the associated compliant electrode (carbon grease).

Note that pre-stretching a DE has been observed to increase its dielectric breakdown strength via increased permittivity. While this also translates to improved capacitances (5.1a), and therefore energy outputs, pre-stretch also makes DEs stiffer and shortens their operational lifetimes. For this reason, we choose a 300% pre-stretch, an acceptable level that does not push the DE generators to their operational limit for pressures exerted by the human body.

Based on the operational behavior that will result from the proposed configuration, we

Table 5.1: Material Properties for 3M VHB4905.

MATERIAL PROPERTY	VALUE
Dielectric Const.	4.5
Bulk Resistivity	$3.1 \times 10^{15} \Omega.cm$
Surface Resistance	$1.0 \times 10^{16} \Omega.cm/m^2$
Density	$960 kg/m^3$
Bulk Resistivity (carbon grease)	$1.17 \Omega.cm$

now detail mathematical models to describe DE behavior during the actuation and transduction stages. These models will help us arrive at the initial (λ_{init}) and final (λ_{init}) stretches in the DE at the end of the actuation and transduction stages, respectively. These values may then be plugged in to equations (5.1a)- (5.1c) to arrive at the net energy output. Essentially, we have modified the experimentally validated model described in [JBC10] for a uniform biaxial stretch deformation, to address the conical deformation in our configuration that involves non-uniform stretching.

5.4.1 DE Geometry Characterization

To analytically model DE behavior over the actuation stage, we must relate the elastic forces experienced in the DE when mechanical pressure is externally applied to it, or its stress, to the stretch it experiences. This requires a characterization of its geometry when stretched. In contrast to the DE transduction in Fig. 5.1 where the DE stretch is equal and uniform along the x and y axes, the proposed configuration (Fig. 5.2) involves a non-uniform stretch that is also unequal in the circumferential and longitudinal directions.

Although the area of the DE when in the stretched state may be precisely deduced via controlled measurements of capacitance, the shape of the DE and therefore its stress is difficult to deduce. To overcome this issue, we characterize its stretch based on the minimum achievable area for a given vertical displacement of the DE. Since the energy output of a DE depends primarily on it stretch (equation (5.1c)), a minimum area-based characterization will provide a lower bound for the energy output. Fig. 5.3(a) shows a cross-section of the DE while stretched.

The lateral surface of the stretched DE in Fig. 5.2 can be deconstructed as a surface of revolution of this curve in the x-z plane (Fig. 5.3(a)), about the z-axis. In the x-z plane, this curve experiences boundaries at $(a, 0)$ and (b, h), where a and b are the radii of the driver and DE active area, respectively. The minimum surface of revolution for such a curve about the z-axis is a catenoid. However, some values of h can yield no solution for the parameters of a catenoid. Therefore, we tackle this problem numerically.

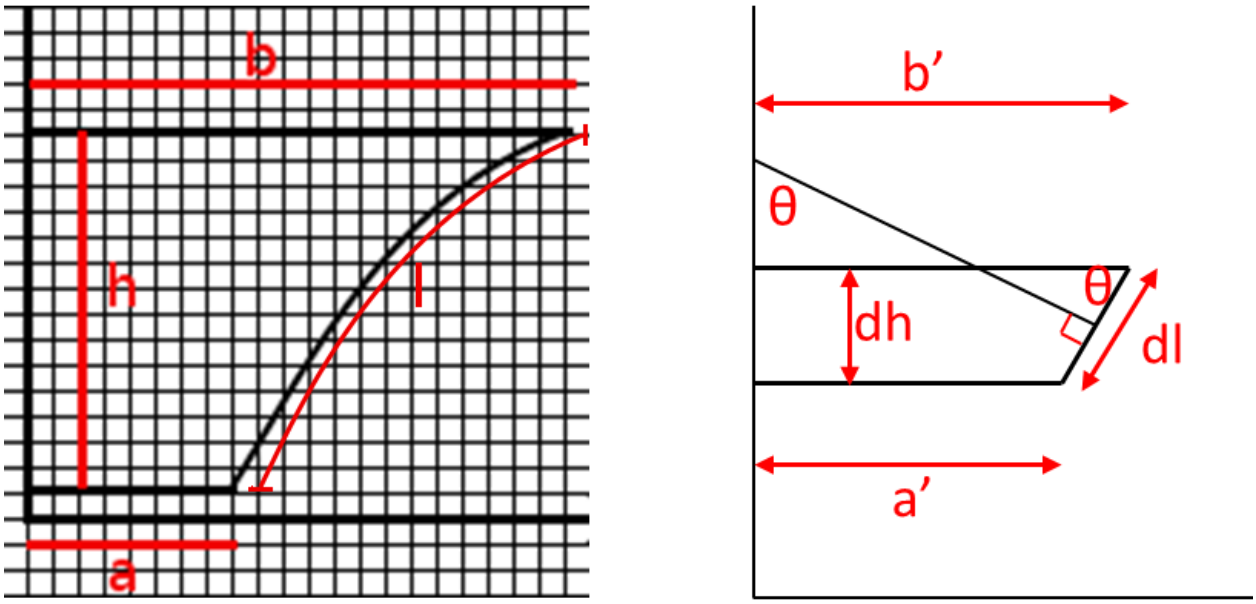


Figure 5.3: (a) Longitudinal cross-section of transducer configuration (left), and (b) Single section of configuration in the longitudinal plane.

As illustrated in Fig. 5.3(a), if a grid with separation dh is laid across the longitudinal cross-section of the stretched DE, the lateral surface area of the stretched DE is the sum of the lateral surface area of its section between each row in the grid. As $dh \rightarrow 0$, each section approaches a conical frustrum illustrated in Fig. 5.3(b). Under the same condition, it is acceptable to assume that the end-points of the top and bottom edges (b' and a', respectively) of each section must occur at the grid's intersection points. A search may then be performed by changing the end-point locations of each section's edges along the grid. The search space is pruned based on the following: (i) The continuity of the sections must be ensured, i.e. the length of bottom edge of a section must be the same as that of the top edge of the section immediately below it; (ii) The length of bottom edge of a section must always be less than or equal to that of its top edge, if minimum lateral surface area is sought. Boundary conditions at the top and bottom of the DE cross-section also apply. To solve the problem numerically, the search is repeated for steadily decreasing values of dh, until the reduction in lateral surface area between consecutive runs is no more than some ϵ. This procedure yields a distinct lateral surface area (A_{lat}) and longitudinal length (l) for a given

vertical displacement h.

$$\lambda_l = \frac{l+a}{b} \tag{5.2a}$$

$$\lambda_l \lambda_c = \frac{A_{lat} + \pi a^2}{\pi b^2} \tag{5.2b}$$

$$\lambda_t = \frac{1}{\lambda_l \lambda_c} \tag{5.2c}$$

$$d' = \frac{d}{\lambda_t} \tag{5.2d}$$

As shown in Fig. 5.2, pressure applied vertically onto the DE induces stresses in the circumferential (σ_c) and longitudinal (σ_l) directions. These stresses yield stretching in the longitudinal (λ_l), circumferential (λ_c) and transverse (λ_t) directions. The stretches in each direction are derived based on the DE geometry as outlined in equations (5.2a) - (5.2c). Equation (5.2d) relates the thickness of the DE in the stretched state (d') to that in the relaxed state (d).

To relate the externally applied force (F_p) to stress induced in the DE, we note that at equilibrium, the vertical component of the longitudinal stress (σ_{l-vert}) must equal F_p. Non-uniformity in stress is accounted for by calculating the equivalent mechanical forces in the DE as an average over the longitudinal length of the stretched DE (equation (5.3a)). Here A'_{cross} is the average circumferential cross-sectional area of the stretched DE, not considering the cross-section of its base. It is derived from the volume of the conical frustrum shell (V_{sec}) at a given height h'(equation (5.3d)). The volume of this shell may simply be computed as the difference in volumes of an inner and outer conical shell separated by the thickness d' of the stretched DE.

$$F_p = \frac{\int \sigma_{l-vert} A_{cross}\, dl}{l} \tag{5.3a}$$

$$= \frac{\int \sigma_l A'_{cross} \sin\theta\, dl}{l} \tag{5.3b}$$

$$= \sigma_l \frac{\int A'_{cross} \sin\theta\, dl}{l} \tag{5.3c}$$

$$= \sigma_l \sum_{h'=0}^{h} \frac{V_{sec}(h') \sin\theta(h')}{dl(h')} \tag{5.3d}$$

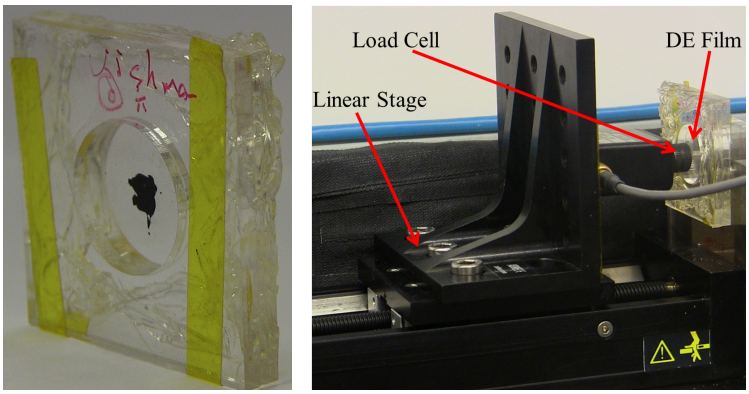
$$A_{cross_{vert}} = \sum_{h'=0}^{h} \frac{V_{sec}(h') \sin\theta(h')}{dl(h')} \tag{5.3e}$$

Equation (5.3e) gives the vertical projection of the average circumferential cross-sectional area ($A_{cross_{vert}}$) of the stretched DE. The average circumferential cross-sectional area of the stretched DE including that of its base, A_{cross}, can be derived in a similar manner to $A_{cross_{vert}}$ by additionally admitting the transverse cross-section at the base of the stretched DE, and losing the $sin(\theta)$ from equation (5.3e) which indicates that it is a vertical component. From the relationships laid out thus far, it can be deduced that the lateral surface area (A_{lat}), longitudinal length (l) and average cross-sectional areas (A_{cross}, $A_{cross_{vert}}$) are functions of h and are monotonic ones at that. This has been verified by simulation and can be used to generate injective maps between any of them.

5.4.2 DE Behavior During Actuation Stage

By characterizing the DE's geometry when stretched in the proposed configuration, we are able to arrive at the longitudinal stress (σ_l) induced in the DE film when an externally applied force (F_p) yields a deformation with vertical displacement h. We have also discussed the model to relate this displacement to the stretches observed by the film (equations (5.2a)-(5.2c)).

As noted earlier, the behavioral model of a DE generator in the actuation stage arises from a description of the *relationship* between the stress that the DE film experiences and the stretch that results. However, DEs are neo-hookean materials with non-linear stress-stretch relationships. As is the convention with such materials, we modeled the elastic behavior of



(a) (b)

Figure 5.4: (a) Prototype DE harvester, and, (b) Experimental Setup for DE prototype stress-stretch measurements.

a DE in the proposed configuration with the Ogden hyper-elastic material model [Ogd72]. This model parametrically relates the mechanical stress in hyper-elastic materials to their resulting stretch, via a partial differential equation and captures the non-linearity therein.

To generate an accurate model, we experimentally derived the parameters of the Ogden curve for a prototype DE generator. The formula for this curve is expressed in equation (5.4), where μ_p and α_p are parameters of the Ogden model. Fig. 5.4 shows the experimental setup for these stress-stretch measurements - A linear stage (Zaber linear stage model A-LSQ300A-E01) was used to move a driver affixed load cell (Transducer Techniques MDB Series) into the active area of the DE prototype that was firmly fastened in place. The displacement h in the DE film was measured as the displacement in the linear stage, and the input mechanical force F_p was measured by the load cell. A non-linear least squares fit of the measurements produced the parameters of the Ogden model specific to the proposed configuration.

$$\sigma_l = \sum_{p=1}^{3} \mu_p \left(\lambda_l^{\alpha_p} - \frac{1}{(\lambda_l \lambda_c)^{\alpha_p}} \right) \tag{5.4}$$

For a given force applied vertically to the DE, we can now deduce the stresses experienced by the DE. From the Ogden stress-stretch curve, we arrive at the resulting stretch and hence the vertical displacement in the DE and values of related geometrical variables. Fig. 5.5 plots

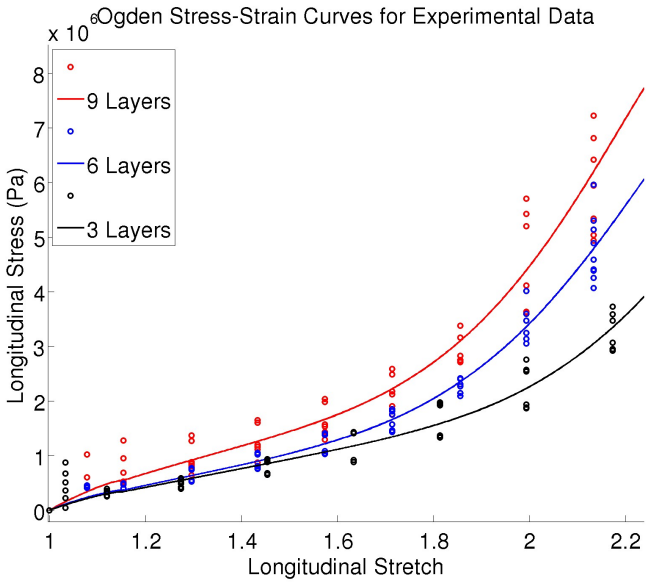


Figure 5.5: Experimentally collected measurements for the stress-stretch relationships of DE harvesters prototypes at different number of layers, and the corresponding best fit curves.

the stress-stretch data as well as the best fit Ogden curves for our experimental measurements for 3 DE configurations, one with 3 DE layers, a second with 6 layers and a third with 9 layers. The related Ogden parameters are listed in table 5.2. From Fig. 5.5 we observe that the configuration with more layers must experience more stress to achieve similar levels of stretch.

Table 5.2: Best fit Ogden parameters for experimentally collected stress-stretch data.

Parameter	3 Layers	6 Layers	9 Layers
μ_1	85209	999892	139347
α_1	-10.2	1.3	9.4
μ_2	-287447	61773	96859
α_2	-4.4	-10.1	-10.2
μ_3	102749	90022	-989128
α_3	9.5	9.2	-1.9

5.4.3 DE Behavior During Transduction Stage

We describe transduction model of a DE generator in the proposed configuration, which we used to compute the net energy produced. In equation (5.1c), whereas λ_{init} is arrived at with the Ogden model solely from the mechanical pressure applied at the time of charging, λ_{final} depends on the interaction between the elastic and electrostatic forces acting on the DE film throughout the transduction stage until the time at which it is discharged. To model DE behavior over the transduction stage, we modified the transduction model for uniform biaxial stretch described in [JBC10] and addressed the geometry involved in the proposed deformation.

On charging the DE at stretch λ_{init}, the net force that leads to a further change in stretch is expressed in terms of the balance between the elastic, electrostatic, gravitational and residual mechanical pressure being applied to the DE. The equations listed below describe this "force-balance" relationship that is central to arriving at the steady-state stretch, when multiple stresses act on the DE film. It expresses the net force that leads to a change in stretch, in terms of the mass and acceleration during this change (LHS of equation below). This force is achieved as a balance between the elastic forces (first and second terms) and the sum of the electrostatic forces (third term), the gravitational forces (forth term) and additional vertical forces (fifth term), acting on the DE. Note that the 5^{th} term represents external forces acting on the DE over the course of transduction. As the stretch decreases (increases), the elastic forces decrease (increase) and electrostatic forces increase (decrease), ultimately leading to a steady state.

$$\begin{aligned} mb\frac{\partial^2\lambda_l}{\partial t^2} = -A_{cross}\Bigg[&\mu_1\left(\lambda_l^{\alpha_1} - \frac{1}{(\lambda_l\lambda_c)^{\alpha_1}}\right) \\ &+ \mu_2\left(\lambda_l^{\alpha_2} - \frac{1}{(\lambda_l\lambda_c)^{\alpha_2}}\right) \\ &+ \mu_3\left(\lambda_l^{\alpha_3} - \frac{1}{(\lambda_l\lambda_c)^{\alpha_3}}\right)\Bigg] - 2\pi bdp' \\ &+ \frac{Q^2}{\pi^2\varepsilon_0\varepsilon_r b^4(\lambda_l\lambda_c)^2} + \frac{A_{cross}mg}{A_{cross_{vert}}} + \frac{A_{cross}F'}{A_{cross_{vert}}} \end{aligned} \tag{5.5}$$

Here, F' is the externally applied pressure on the DE at the instant when the equation is evaluated, and p' is the residual stress in the DE due to pre-straining and is derived while fitting the Ogden model. Recall that μ_p and α_p are parameters of the Ogden model that capture the hyper-elastic properties of the DE. Also, m is the mass of the DE, g is standard gravity, d is the DE thickness after it is pre-stretched but before it is operated on, Q is the charge applied, and, $\frac{A_{cross_{vert}}}{A_{cross}}$ is the fraction of the force vector that corresponds to its vertical component. This force balance equation is solved with a Runge-Kutta forth order method to produce the steady state stretch value. And λ_{final} is the steady-state stretch value at the commencement of the discharging stage.

$$E_{harv} = LossFact_{ckt}(\Delta E_{DE} - E_{loss_{elec}}) \tag{5.6}$$

The net energy output ΔE_{DE} may now be derived as in equation (5.1c). Whereas that equation represents equi-biaxial stretch, it can be transformed to yield the energy output for the bulge configuration by replacing λ^2 with $\lambda_l \lambda_c$. The net energy harvested (E_{harv}) is proportional to the difference between the net energy output of the DE and the electrical losses incurred during charging and transduction (equation (5.6)). These losses $E_{loss_{elec}}$ include conduction and dielectric losses in the material (Equation (5.7)) and through a 1micron thick compliant electrode, and impedance-based losses in the charging circuitry (1kΩ impedance). In calculating the net energy harvested we consider the DC-DC conversion inefficiencies in the rectification circuit $LossFact_{ckt}$. Finally, since DEs do not heat up significantly in generator mode [JBC10], dielectric losses have been ignored.

$$P_{cond_B} = \frac{V^2}{R_b} \quad P_{cond_S} = R_s I_s^2 \tag{5.7}$$

Fig. 5.6 shows the input pressure observed near the ball of the foot of a subject wearing the HERMES shoe, over the stance phase of a single stride. It also shows the net energy output, based on the above model, that would results from operating a DE harvester at this location. Here, the DE is assumed to be charged at 8kV. The energy output depicted at each epoch, is the net energy produced if the DE were charged at that time step and

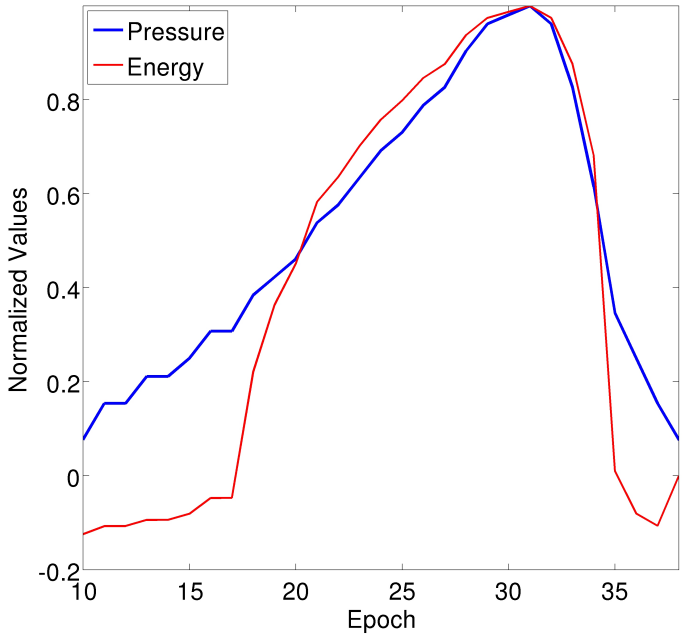


Figure 5.6: The input pressure and net output energy profile at a single pressure sensor's location for a sample step in the HERMES plantar pressure dataset.

discharged at the end of the stance phase. The inherent non-linearities in the relationship between the input pressure and the net energy output are apparent here. Furthermore, even if the input pressure is the same, the energy output isn't necessarily so. If charged as the pressure rises, the sum of losses over the following epochs will be more than if charged, at the same pressure, as the pressure falls.

CHAPTER 6

Optimizing the Configuration and Control of a DE Generator based Foot-Strike powered Harvesting System

6.1 Introduction

Having studied the benefits and behavior of DE harvesters, we turn our attention to their application in powering BANs. Specifically, in the context of low-power wearable systems that involve human locomotion, we propose the use of a leading-edge energy harvester technology to scavenge the energy produced by a user's foot strikes and drive a target wearable system in a self-sustaining manner.

Foot strikes have been estimated to yield significant amounts of energy - a 154-pound person can produce upwards of 7W per foot strike [PS05]. However, the portion of this that can be scavenged is limited by two factors. First, current energy transduction technologies are characterized by energy densities that are insufficient to capture such quantities of energy. Second, state-of-the-art transducer configurations are unable to transduce significant proportions of human energy output without causing discomfort to the user. For example, a hydraulic-amplified piezo-electric based foot-strike energy harvesting system was proposed in [ABG95], capable of producing 675mW per foot-strike while adding significant heft to the shoe, thereby altering the user's gait. In contrast, the authors of [SP01] proposed a piezo-electric based foot-strike energy harvesting system that interfered minimally with the user's gait and was capable of producing 10-20mW.

In this chapter, we propose a DE harvester array configuration for improved transduction

efficiency of the mechanical energy from foot pressure. We leverage the users plantar pressure profile to control the design of individual transducers in the array, such that the energy yield is more uniformly high across the foot despite spatial variations in the pressure profile. High efficiency is also achieved via direct transduction of input mechanical energy. Finally, its mechanical simplicity keeps system design and maintenance complexity to a minimum.

We also propose and discuss 2 novel adaptive closed-loop DE transduction control paradigm that exploits characteristics of the user's gait to maximize energy output, a non-parametric predictive technique and a semi-parametric predictive technique. As electrostatic transducers, DE output is heavily reliant on the control timing during its transduction cycle, which will vary between foot steps owing to variability in the user's gait. To maximize energy output, the proposed algorithms that makes use of fool pressure sensors to predict optimal control timing of each harvester in the array, based on statistical characteristics of the user's gait. The sensors are sampled to maximize the net energy harvested under sampling power constraints imposed by the power demands of the target system. We note that adaptive control of DEs has been proposed in the context of actuation [GCX08], however to the best of our knowledge, we are the first to propose an energy-maximizing adaptive control technique for DE generators.

We apply the DE behavior model from Chapter 5, in conjunction with our proposed control algorithms, to experimentally collected datasets of foot pressure of multiple users to demonstrate that our DE based foot-strike energy harvesting system can yield upwards of 120mJ per foot-strike, thereby producing enough energy to independently power a number of wearable platforms while the user walks [NAH10][CBC08][CIP06].

6.2 Preliminaries

6.2.1 Human Gait

Although gait characteristics vary across people, human locomotion, but its very nature produces commonalities across people. One such commonality is that spatial pressure profile

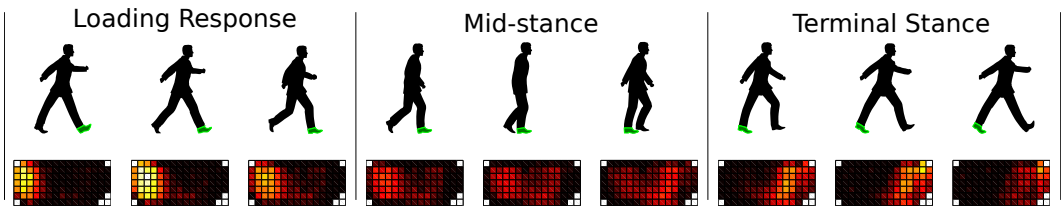


Figure 6.1: Foot pressure profiles for sub-phases of stance phase.

across the sole of the foot exhibits local correlation. Fig. 6.1 illustrates sub-phases of the stance phase along with plantar pressure profiles corresponding to those sub-phases. These profiles are averages over several strides from one of our datasets, collected with a high spatial-resolution plantar pressure monitoring system. They depict the heel at the left and toes and the right with lighter colors corresponding to higher pressure.

Three properties of plantar pressure that generalized to all our datasets may be observed from Fig. 6.1: (i) The pressure profiles present spatial correlation localized to small regions across the sole; (ii) The maximum pressure observed at different regions is different; and (iii) Different regions observe maximum pressure at different times. Even within a region of high correlation, adjacent locations may observe maximum pressure at different times. We use these properties of human gait to optimize both the configuration and control of the harvesting system.

6.3 DE Harvester Design

Fig. 6.2 depicts the high level mechanical energy harvester design for DEs that follows from their mode of transduction as described in chapter 5. Here the user's activity context and gait, as described by the spatio-temporal foot pressure profile for each stride, dictates the instantaneous input mechanical pressure to the DE. From equation (5.1c), the net energy output depends on λ_{init}, the stretch in the DE when charge is applied, and λ_{final}, the stretch in the DE when it is discharged. Therefore, the control of the charging and discharging stages are crucial to the amount of net energy harvested, and are respectively controlled by switches 1 and 2 in Fig. 6.2. The mechanical and electrical behavior of the DE control the outcome of

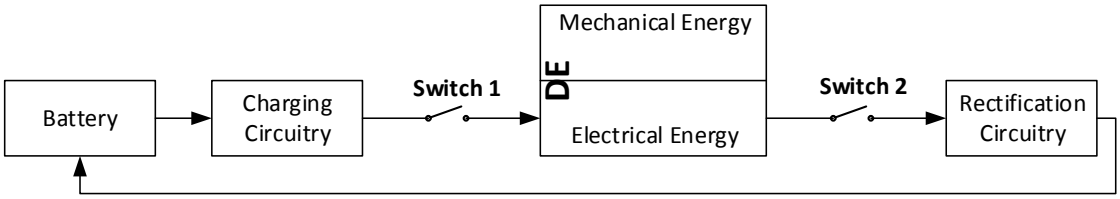


Figure 6.2: Schematic of DE Energy Harvester System.

the actuation and transduction stages, and are governed by the DE's physical configuration and material properties. Electrical losses in the DE and the rectification circuitry impact the net energy output.

In our effort to maximize the net energy harvested, we propose a physical configuration that enhances transduction efficiency. First, rather than a single DE or a few larger DE generators spanning the shoe sole, we propose an array of 99 DE micro-generators spanning the shoe sole, at locations corresponding to those in foot pressure profile diagrams in Fig. 6.1. This is precisely the reason why our experimental modeling of DE behavior in chapter 5 was based on DE micro-generators about the size of a single HERMES plantar pressure sensor.

As we shall see, the timing at which maximum pressure is observed at a location is crucial to control over the charging stage. Since each location may experience maximum pressure at different times, such a configuration will enable finer control of DE charging stages, thereby improving energy output. Second, the DEs harvesters will be operated in the conical configuration described in chapter 5, which will enable direct transduction and reduce mechanical losses. Fig. 5.2 shows a harvester in this configuration, where DE deformation is brought about by a driver component affixed to the shoe insole.

6.4 DE Harvester Array Control Parameters

Based on the description of the DE transduction mechanism in chapter 5, we identify 4 parameters that can be crucial to the net energy harvested.

6.4.1 Charge Timing

A key observation is that the initial stretch λ_{init} that figures in equation (5.1d) is a function of the timing of the applied charge. For example, if DEs at the toes are charged at the beginning of the stance phase when they experience no input mechanical pressure, λ_{init} will be close to 1 regardless of the user-applied pressure that follows during the stride. This will result in negative energy harvested due to the electrical losses. Therefore it is imperative that charge be applied as close to maximum stretch as possible to maximize the energy output. Since maximum stretch will be observed at maximum input mechanical pressure (due to monotonicity of the stress-stretch curve), the ability to predict when this maximum pressure will be observed is key to optimal control at the charging stage. In the following sections, we will explore control algorithms that makes use of a few foot pressure samples to enable such prediction.

Towards this end, the harvester platform will require electronic control during the charging and discharging stages of DE operation. Processing of the sensor samples towards DE control parameter estimation may be offloaded to the microcontroller of the target device. Specifically, in the case of HERMES, we leverage the design of HERMES which employs the low-power MicroLEAP platform [AWB07], which provides a multiplexed 8-channel 16-bit ADC, 8Mbits of flash memory, and an MSP430 micro-controller. This low-power electronic control platform will be used to control switches 1 and 2 in Fig. 6.2 based on the control algorithms that are discussed in the following sections.

6.4.2 Input Pressure

While a higher input mechanical pressure will yield a higher stretch and potentially more net energy output, this parameter is controlled by users gait over the actuation and transduction stages. However, foot-strikes yield distinct spatio-temporal properties that may be leveraged to reduce transduction in-efficiencies with an array configuration of DE micro-generators, as discussed in section 6.3.

6.4.3 Harvester Thickness

Recall that we experimentally characterized DE stress-stretch behavior in chapter 5 for 3 DE configurations, one with 3 DE layers, a second with 6 layers and a third with 9 layers. Based on these experimental measurements of DE stress-stretch behavior, we concluded that a a thinner DE will stretch more at comparable applied pressures In contrast, thicker DE films can withstand much more pressure without rupture. Therefore, thinner DE films can harvest more energy at acceptable levels of applied pressure.

Further, different regions of the foot see different levels of maximum plantar pressure (Fig. 6.1). To increase the energy yield of low input pressure regions of the foot, such as the arches, we propose an adaptive harvester placement strategy. Here, the thickness of a harvester in the array will depend on the amount of pressure observed over the users gait at its location. Based on the maximum pressure observed over a training dataset for each DE location, decisions will be made to fit high pressure locations with thicker harvesters and low pressure locations with thinner ones. Specifically, to address harvester lifetime and user-comfort issues, a location is occupied by the thinnest harvester for which the maximum observed pressure over the training dataset produces no more than a vertical displacement h of 9mm, as per the stress-stretch model. This is an acceptable level of compression for soft cushioned shoes [vci].

6.4.4 Applied Voltage

Equation (5.1d) suggests that higher applied voltages should yield higher net energy outputs, ad infinitum. However, since the applied voltage is related to the applied charge, adding too much charge to the elastomer film will create large electrostatic forces on it, which will prevent full relaxation of the film at the end of the stance phase. As λ_{final} is driven away from 1 at higher applied voltages, extremely high voltages are undesirable. Fig. 6.3 shows the net energy output predicted by the DE behavior model when a harvester is exposed to different input pressures and charged over a range of voltages. We observe that the output energy is not a monotonically increasing function of the applied voltage. At voltages beyond

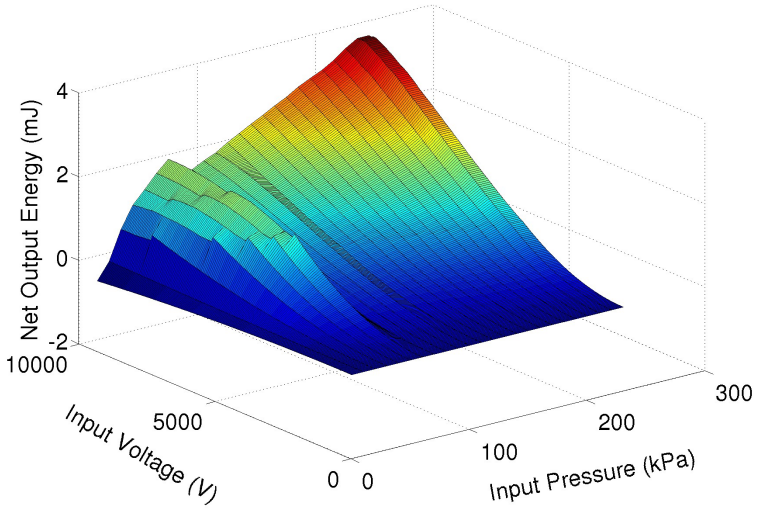


Figure 6.3: Net Energy Output (ΔE_{DE}) of harvester model for different input pressures and applied voltages.

9500V, and especially for input pressures at the higher end of the spectrum, the net energy output begins to decrease with increasing voltage owing to large electrostatic forces that are retained in the DE and prevent it from fully relaxing. As a result, we restrict system operation to 8000V.

The plot also factors in the DE thickness configuration over the plantar pressure range, as described above. For this reason, the output energy does not monotonically increase with input pressure. The energy output up to 180kPa is governed by the thinner harvesters and they produce much more energy than the 9 layer harvester at those pressures. However, sharp transitions occur when the input pressure surpasses the user-comfort / harvester lifetime thresholds for the thinner harvester. This necessitates switching to thicker harvesters leading to sharp declines in energy output.

6.5 Control of DE Harvester Array in the Charging Stage

As indicated in Fig. 6.1, the DEs must not be charged at the beginning of the stance phase if energy output is to maximized. In order to accurately predict the optimal charge timing of

each harvester, we propose the use of a few pressure sensors samples. We may then leverage the local correlations in the foot pressure profiles to make accurate predictions about the timing of maximum pressure at each harvester, thereby accurately predicting the timing of maximum stretch and maximizing net energy output.

Multiple approaches may be conceptualized for the inclusion of foot pressure sensors into the system. The DEs may be juxtaposed with passive-resistive sensors, or, the DEs may support dual-mode operation as sensors or generators. Sensor mode operation of DEs may consume up to a few nJ more than passive-resistive sensors, but not more; Since our goal is not to maximize energy output when they are operated as sensors, DEs may be charged at a low voltage to measure a change in capacitance and calculate the applied pressure therefrom. If the DE is already charged for harvesting, the changes in its capacitance can still be measured in a similar manner.

We formally define the charge timing prediction problem as follows. We assume a maximum of T epochs between the start and end of the stance phase, over which we are given the foot pressure readings at each of the harvester locations. We would like to predict the time at which maximum pressure is observed at each location in a manner that maximizes the net system energy output. Here the net system energy output is computed as the net harvester output (equation (5.6)) summed over all harvester, minus the energy cost of each pressure sample used in the charge timing prediction. Further, the following constraints apply:

- If the optimal charge timing for a DE is predicted as epoch k, the prediction must be made between epochs 1 and $k-1$.
- If the optimal charge timing for a DE is predicted as epoch k, the prediction must be based on samples taken solely between 1 and $k-1$.
- There is a maximum number of samples that may be used at each epoch, which may be different for different epochs. This constraint is governed by the battery power rating and the power draw of the target application that we are trying to power. However, samples taken for the purposes of the target application may also be applied to charge timing prediction.

6.5.1 Non-Parametric Charge Timing Prediction

The extent to which energy is harvested from a DE, depends on the system knowing or predicting when the maximum stretch is reached for the step at the DE's location (i.e. the moment when the charge should be applied). Unfortunately, different users will have different stepping patterns and pressure profiles across their soles. Furthermore, ambulation timing and step pressure varies across the steps of the same user. Thus, in order to maximize the total energy collected over the DE array, the system must be able to measure and/or predict when each harvester will reach maximum stretch during a given step. Predicting these measurements requires knowledge of harvester location, pressure distribution, and step timing. As we have stated, it follows that we can minimize error if this prediction is assisted, in real-time, by some pressure sampling data. The system must utilize a set of sensors measurements that are able to predict with high accuracy, the timing of maximum stretch at the harvesters.

Similarly, placement of the harvesters is not as simple as choosing the location with the highest average pressure. To achieve the final goal of maximizing aggregate net energy output from the DE array, we must consider all aspects of the individual harvesters that will contribute to their net output: (i) What is the expected energy output for the harvester; (ii) What is the expected prediction error for the harvester, and, (iii) What is the energy cost of prediction for the harvester? If they net energy expended in predicting a harvester's charge timing is more than the net energy it produces, then it doesn't make sense to charge the harvester at all and we will choose not to "place" or operate the harvester. The net energy output of a harvester could be lower either due to poor predictability, or because not much pressure is observed at its location on average.

Our methodology for coordinated sample selection, harvester placement and scheduling is executed in two steps. Given a training dataset, the first step builds an exhaustive pool of harvester charge-timing predictors from samples at all locations on the sole and over all epochs of the stance phase. The second step calculates a subset of harvesters, while simultaneously allocating samples to predict their charge timing, in a manner that collectively

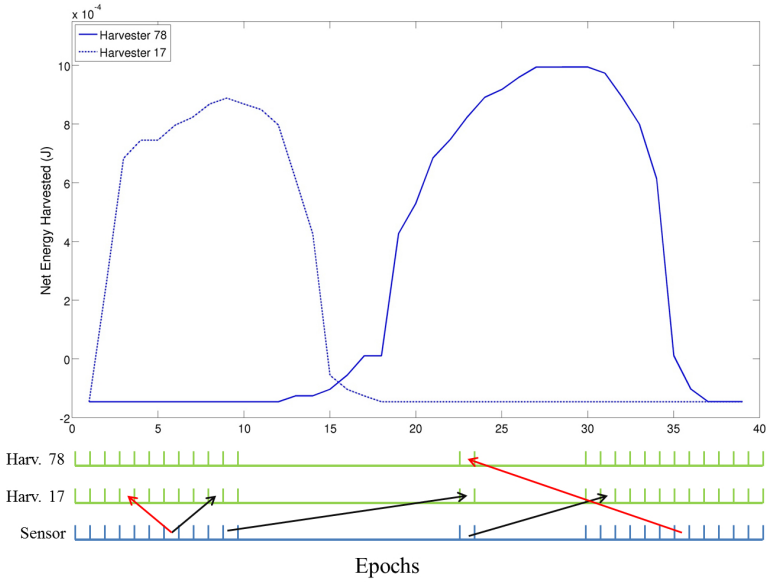


Figure 6.4: Energy profiles of two harvesters with relatively large energy outputs. Valid predictions for real-time control are shown in black, while invalid ones are shown in red.

maximizes the expected aggregate net harvested energy over the dataset.

6.5.1.1 Predicting Harvester Energy Profiles

In order to accurately predict the optimal charge timing of the DE harvesters, we enlist the local correlation property of plantar pressure. Based on this property, an aptly placed sensor, sampling at the right time should be able to accurately predict the timing of a number of harvesters in its neighborhood. Towards this end, we construct of a pool of robust statistical models between sensor samples and temporal profiles of harvester output. Here, we define sample $s_{i,j}$ as a single pressure measurement of the sensor i sampled at the epoch j. Similarly, yield $h_{p,q}$ is the net energy output of harvester p after applying a charge at epoch q and discharging at the end of the stance phase. For a stride, the models corresponding to a sample may be used to predict the entire temporal profile of a harvester's output, based on the sample value. Consequently, a sample's optimal charge timing prediction would be the timing when the predicted profile achieves maximum.

Note that our goal isn't to accurately predict the temporal profile. Fig. 6.4 illustrates the

temporal profiles for 2 harvesters (named after their location in Fig. 1.1) from a single stride by a subject. Harvester 78 sees minuscule changes in the energy scavenged between epochs 28 and 32, however, harvester 17 sees more variance around its peak output. Therefore, the quality of a predictor is not quantified by the number of epochs it is off by in predicting the instant of maximum stretch for the DE. Neither is it based on the error in its predicted value of maximum energy harvestable. Rather, the error in the predictor is defined by the difference between the amount of energy that will be scavenged, if its timing prediction is adhered to, and the actual maximum energy that can be harvested for the step. This key insight expands the solution space substantially, thereby reducing the complexity of the problem. Finally, while creating the pool of models, it is crucial that we only consider those models where the sensor sample is seen before the predicted yield, i.e. $j < q$ (Fig. 6.4). Obviously, hindsight is irrelevant in the context of real-time control.

In order to build a robust and effective statistical model between a valid <sample, yield> pair, we must consider, (i) the non-linearity in the relationship between samples and yields despite a high correlation in the pressure observed at the locations of the corresponding sensor and harvester, (ii) the temporal variability inherent in the samples, and, (iii) the delay between samples and yield. This last point was discussed in the context of Fig. 5.6 where it was shown that the net energy output could be different for the same input pressure depending on the charge timing.

In considerations of these issues, we enlist non-parametric kernel-regression to model the <sample, yield> relationships. To account for its non-linearity, the sample value is treated as a continuous independent variable whose probability density is computed with a gaussian kernel. The sample epoch is treated as an ordered factor to explain the temporal variability in the sample, and its probability density is computed with the ordered kernel described in [WR81]. Finally, the yield epoch accounts for the sample-yield delay, and is the other ordered factor in the model. The kernels are combined in a generalized product kernels approach [LR03], with kernel bandwidths of 2000, 0.25 and 0.25 respectively. These bandwidths were pre-determined to produce high accuracy models for our datasets. While the models are computationally expensive to produce, we note that this is done offline. The

Algorithm 8 Calculate AEMHPs and corresponding predictors (ILP)

Constants: $pe(s_i, h_j)$, the average harvested energy at harvester h_j based on predictions from sample s_i, C_i, the energy cost of each sample, and N the number of harvesters

Integer Variables: S_iH_j, an indicator for whether sample s_i will control harvesting at h_j, and S_i, an indicator for whether sample s_i will be used

1: Max: $\sum_{s_i} \left(\sum_{h_j} [pe(s_i, h_j).S_iH_j] - C_i.S_i \right)$

2: $\sum_{s_i} S_iH_j \leq 1$ for each h_j

3: $\sum_{h_j} S_iH_j \leq N.S_i$ for each s_i

4: $0 \leq S_i \leq 1$ for each s_i

5: $0 \leq S_iH_j \leq 1$ for each s_i and h_j

subsequent prediction required for real-time control only involves a few table lookups and few simple arithmetic operations. By repeating this prediction at each stride, our proposed control algorithm will result in charging schedules that are adaptive, depending on sensor pressure measurements, and generated in real-time. At each stride of the user, the pressure is sampled to generate a distinct harvesting schedule.

6.5.1.2 Aggregate Energy Maximizing Harvester Placement (AEMHP)

While we are now equipped with models that can predict the charge timing for individual DEs, it is still unclear as to which samples will drive each harvester. Although we observe local correlations, the exact spatial nature of these correlations will vary among users. As stated earlier, we must also consider that some harvesters generally see insufficient input pressure and produce insignificant amounts of energy, and a few others may be too variable to predict in a manner that will result in a net positive energy output. There is also a cost involved in prediction. While each harvester can be paired up with the sample that single-handedly maximizes its expected net energy output, a better configuration may be possible if the cost of sampling is taken under advisement and multiple harvesters are predicted by a single sample, even if the individual predictions are marginally lower than the best possible sample-harvester pairings.

Algorithm 9 Calculate AEMHPs and corresponding predictors (Greedy)

Input: $pe(s_i, h_j)$, the average harvested energy at harvester h_j based on predictions from sample s_i, and C_i, the energy cost of each sample

Output: $AEMHPr$, the AEMHP predictor set of selected samples, and $AEMHP(s_i)$, the AEMHP sets of harvesters for which s_i will be the predicting sample

$he(h_j) \leftarrow 0 \; \forall \; harvesters \; h_j$
loop
  $se(s_i) \leftarrow 0 \; \forall \; samples \; s_i$
  for all s_i not in $AEMHPr$ **do**
    $se(s_i) \leftarrow \sum_{h_j | pe(s_i,h_j) > he(h_j)} [pe(s_i, h_j) - he(h_j)]$ - C_i
    $s_k \leftarrow$ sample with maximum value in se
    if $se(s_k) \leq 0$ **then**
      Break
    else
      Add s_k to $AEMHPr$
      for all all h_j s.t. $pe(s_k, h_j) > he(h_j)$ **do**
        Remove h_j from $AEMHP(s_l)$ where s_l currently covers h_j, if $he(h_j) > 0$
        Add h_j to $AEMHP(s_k)$
        $he(h_j) \leftarrow pe(s_k, h_j)$
      end for
    end if
  end for
end loop
Return $AEMHPr$ and $AEMHP$

With this goal of simultaneously maximizing the net energy output, we present an ILP formulation that will ensure maximum expected aggregate net energy output (algorithm 8). Due to its computationally-intensive nature, we also formulate a greedy algorithm that runs in polynomial time and produces near optimal results (algorithm 9). Judging each sample on the average net energy that it assists a harvester in producing, at each iteration the algorithm

selects the sample that offers the best improvement over the covering set selected so far. In other words, a sample's merit is measured as the sum, over all harvesters, of improvements in energy harvested based on its predictions. The sample's merit is also penalized for the cost of sampling. At each iteration, the sample with the highest merit is selected into the covering set. The algorithm ends when no sample offers an improvement over the covering set selected thus far.

We note 2 implication of being constrained to the samples afforded by the target application, as defined in the charge timing prediction problem statement. First, samples that are readily available owing to their use in the target application will incur 0 cost (i.e. $C_i = 0$) and can be used freely if they provide improved prediction results. Second, algorithms 8 and 9 can easily altered to account for a constraint on the number of samples allowed per epoch arising from power usage constraints on the target platform. Algorithms 8 will have to include an additional constraint to this effect. Algorithm 9 will remove all samples that will violate the power usage constraint every time an assignment is made (line 10).

6.5.2 Semi-Parametric Charge Timing Prediction

Here, we take a semi-parametric statistical approach to solving the charge timing prediction problem - Given a training dataset, we apply a semi-parametric survival analysis based technique to predict optimal charge timing.Originally developed to predict machine failure times and death of biological organisms, their use has expanded to several applications requiring time-to-event prediction. Based on a training dataset, the survival rate $S(t)$ of a harvester is the probability that its optimal charge timing hasn't yet occurred at epoch t. The hazard rate $f(t)$ is the conditional probability that the optimal charge timing will occur at epoch t, under the condition that it hasn't occurred yet. The cumulative hazard rate $F(t)$ is the sum of the hazard rates less than or equal to t. And from equation (6.1a), where Δt is the duration of an epoch, it follows that it approaches negative log of the survival rate as Δt approaches 0.

Algorithm 10 ctPred: Optimal Charge Timing Prediction

Input: k value of next epoch, s_j value of sample j, β_j the cox coefficient for samples s_j, and F_0 the non-parametric estimates of the cumulative hazard function

1: $F \leftarrow F_0 e^{\sum_j \beta_j s_j}$
2: $Sr \leftarrow e^{-F}$
3: **for** $i \leftarrow k$ to end of stance phase - 1 **do**
4: $\quad Pr(i) \leftarrow Sr(i) - Sr(i+1)$
5: **end for**
6: $Ep \leftarrow$ epoch at which Pr(i) is maximum
7: **if** $Ep = k$ **then**
8: $\quad$ Charge Harvester at epoch k
9: **end if**

$$f(t) = \frac{S(t) - S(t + \Delta t)}{\Delta t S(t)} \tag{6.1a}$$

$$F(t) = \sum_{t'=0}^{t} f(t) \tag{6.1b}$$

$$= -log(S(t)) \tag{6.1c}$$

Cox regression, $coxReg()$ is a semi-parametric survival regression technique. Given a dataset of optimal charge timing for a DE over several steps, and a set of covariates x_1 thru x_n, that signify a subset of samples, the algorithm constructs a cumulative hazard rate function based on a non-parametric baseline estimate of the function (F_0), and multiplicative effects parameter composed of a linear combination of the samples x_i. The coefficients β_i are derived as maximum-likelihood estimates for on the training set.

$$F(t) = F_0(t) e^{\beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n} \tag{6.2}$$

6.5.2.1 Online Prediction

Algorithm 10 outlines the procedure used to decide whether a DE should be charged at the following epoch based on the samples available so far. The algorithm is called for each uncharged harvester and at each epoch, and is provided the samples available thus far, along with the corresponding cox regression parameters. With these inputs, the algorithm computes the cumulative hazard function for the following epoch k and converts the function to the corresponding survival rate function Sr. While the survival rate function tells us the probability that the optimal charge timing won't have occurred at each future epoch, lines 3 thru 5 of the algorithm convert this to the probability that optimal charge timing will occur at each future epoch. If the highest probability coincides with the following epoch, a decision is taken to charge the DE. The algorithm waits until just before the optimal charge timing is expected for two reasons. First, this provides additional information that is likely to lead to a better decision. Second, we observed that with fewer epochs remaining until the end of the stance phase, it increases the kurtosis of the optimal charge timing probability density function thereby improving our expectation.

6.5.2.2 Offline Optimization

Given our use of the survival analysis technique, the charge timing prediction problem stated in section 6.5 turns into a sample selection problem that offers the best predictions while abiding by the constraints therein. We solve the problem with the stepwise-regression based optimization in algorithm 11. In the absence of samples, each harvester's timing is predicted from its baseline hazard functions. Else, each sample, s_j, is evaluated via $avgImp()$ that measures the improvement s_j affords to the prediction of harvester h_i (line 9), averaged over training steps. This improvement contributes to s_j's profit sp_j (lines 10 thru 12). The prediction's improvement is measured in terms of the increase in energy output of h_i due to the addition of s_j into its predictive sample subset for epoch k. At each iteration, the most profitable sample is selected into the predictive subset (lines 14 thru 17). Note, that if a sample is available freely, due to its necessity in the target application, its cost C_j is

0. Otherwise, its cost corresponds to the energy expended in acquiring the sample. Power constraints are observed by line 19, and overfitting is prevented by applying the Akaike Information Criterion (AIC) to evaluate goodness of fit before accepting a sample into a harvester's predictive subset for an epoch (lines 6 thru 8). For the sake of brevity, we have implicitly assumed that lines 11 and 18 include mechanisms to denote which harvester and epoch each included sample will be used to predict.

6.6 Control of DE Harvester Array in the Discharging Stage

Aggregate net energy output depends on the amount of stretch experienced by the DE harvesters. However, all DEs in the array experience stretch solely during the stance phase, when the foot is in contact with the ground. Hence, the T epochs over which DE charging

Algorithm 11 Optimal Predictive Sample Subset Selection

Input: S_0, set of available samples, C, set of energy costs of samples in S_0, and H, set of harvesters

Output: S, selected predictive sample subset

1: $S \leftarrow NULL$
2: **loop**
3: $sp_j \leftarrow -C_j \ \forall \ s_j \in S_0$
4: **for all** $h_i \in H$, $s_j \in S_0$, $k > t(s_j)$ **do**
5: $CP \leftarrow \text{coxReg}(h_i, k, (S \cup s_j))$
6: **if** AIC fails on $(S \cup s_j)$ **then**
7: Continue
8: **end if**
9: $\Delta E \leftarrow \text{avgImp}(h_i, S, \text{ctPred}(h_i, k, (S \cup s_j), CP))$
10: **if** $\Delta E > 0$ **then**
11: $sp_j \leftarrow sp_j + \Delta E$
12: **end if**
13: **end for**

14: $s_k \leftarrow \{s_j \mid sp_j \geq sp_k \forall s_k \in S_0\}$

15: **if** $sp_k \leq 0$ **then**

16: Return S

17: **end if**

18: $S \leftarrow S \cup s_k$

19: $S_0 \leftarrow S_0 \backslash$ ($s_k \cup$ ($s_j \mid s_j$ violates power constraints))

20: **if** S_0 is empty **then**

21: Return S

22: **end if**

23: **end loop**

control is actualized begin at the start of the stance phase and end when the stance phase ends. Towards this end, we assume the application of one of the techniques discussed in chapter 2 to identify the transition between the stance and swing phases with a few sensors. By detecting the transition between phases, we are also able to achieve discharging stage control. In other words, all DEs may be simultaneously discharged at the end of the stance phase.

Drawing from the design in [OHB02], we have also designed the rectification circuitry for the system that will combine this simultaneously discharged energy output before sending it towards the energy store or battery. The proposed circuit (Fig. 6.5) assumes each DE, k, is connected to the harvesting circuit via transistors $T1_k$ and $T2_k$ and a diode bridge. The aggregate harvesting circuit is composed of single DC-DC converter and battery, thus keeping design costs at bay. As we have discussed, the detection of a transition between the stance and swing gait phases will be used to control all transistors, such that all DEs can discharge into the battery at once over the swing phase.

From out experimentally collected foot pressure datasets, we expect the total mechanical energy expended by foot-strikes to exhibit low variance. Therefore, we can design the DC-DC converter in Fig. 6.5 to achieve high efficiency over the narrow expected output current range for the user. It has been shown that under low input current variance, DC-DC converters can

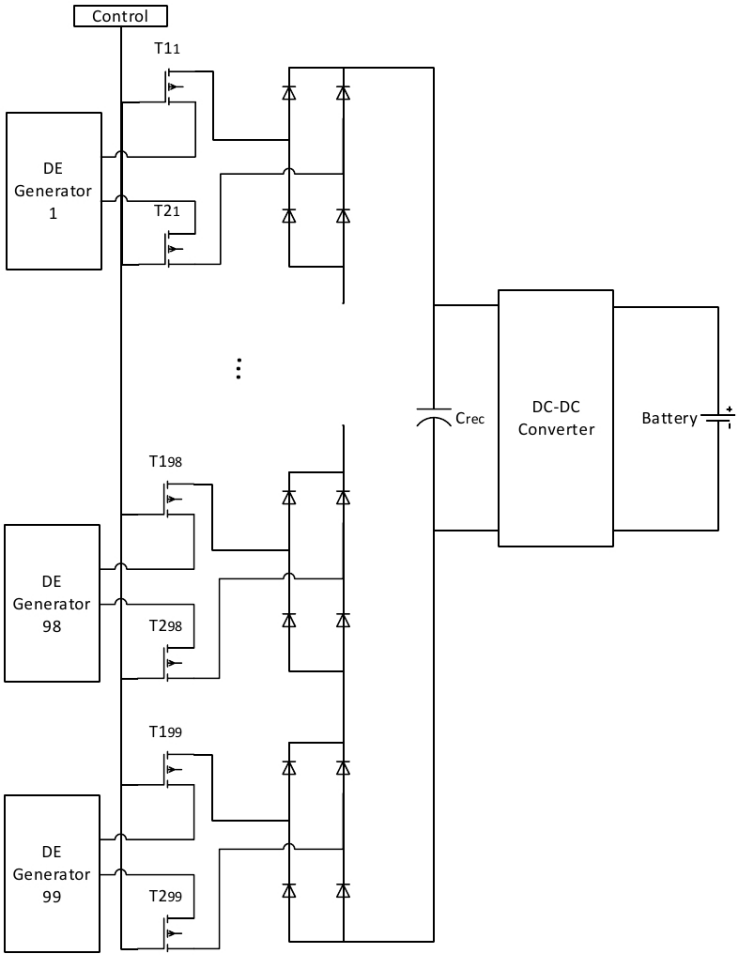


Figure 6.5: Schematic of proposed DE net energy output combination circuit.

be designed to yield efficiencies upwards of 80% [SBD01]. Therefore, we set the rectification loss factor ($LossFact_{ckt}$) to 0.8.

6.7 Performance Evaluation

We evaluate the system performance with experimentally collected foot plantar pressure datasets by HERMES from 3 users. While HERMES was used to collect the datasets we use in our evaluation, the goal of our evaluation is, in part, to assess whether the designed system can be applied to power HERMES in a self-sustaining manner as it monitor's the wearers gait. Two of the datasets correspond to the gaits of lighter individuals, one male and

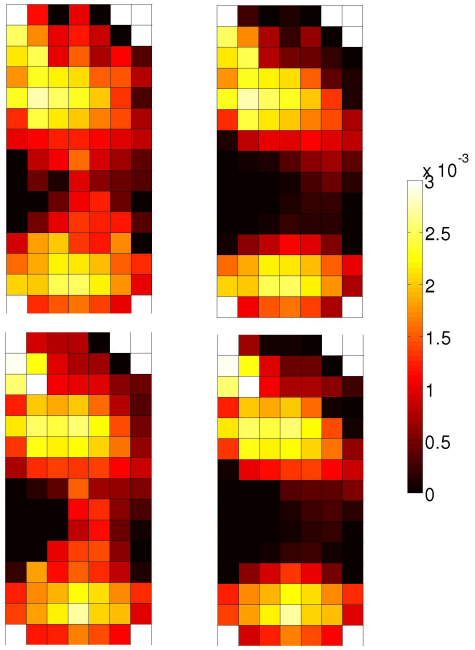


Figure 6.6: The maximum energy harvestable from plantar pressure datasets 1 and 2, averaged over strides, with uniformly thick DEs across the array (right), and those with location sensitive DE thickness (left)

one female. The third dataset corresponds to a heavier male individual. Each of the datasets offer several steps worth of data at each of the ninety-nine harvester locations. This allows us to derive performance at all harvesters and apply the predictive sample subset selection and control algorithms, thereby evaluating the system net energy output. We divide each dataset into a training subset comprised of 80% of the data and a testing subset comprised of the rest.

Fig. 6.6 presents the maximum energy scavengable (with perfect charge timing prediction), averaged over the strides in the datasets, for both feet of a single user. Here, the maximum achievable output with and without the location-specific thickness configuration are juxtaposed. While the areas that normally produce large amounts of plantar pressure produce exactly the same amounts of energy, surrounding areas, that were previously anemic in their output begin to look more promising. To test the predictive sample subset selection algorithm, we leveraged our work from chapter 2 towards power-constrained sub-sampling

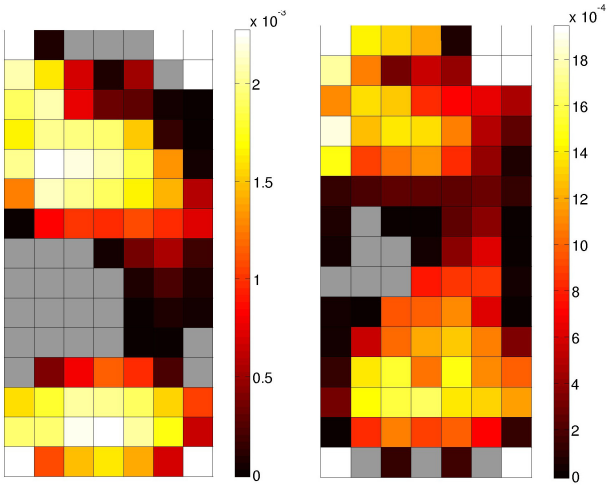


Figure 6.7: Average energy harvested by the non-parametric technique for 2 of the datasets, along with the output of the harvester placement algorithm. Non-selected harvesters have been grayed out. The number of samples dedicated to the real-time control of the selected harvesters are 13 and 11, respectively.

with HERMES. The authors therein, proposed a semantic accuracy preserving sampling strategy that reduced the energy consumption of HERMES from 166mJ to 32mJ per stride, under power constraints of at most 5 sensors sampled at a time. We leverage the output of the algorithm proposed in that work and constrained our algorithm by those samples. Additional samples were allowed to be chosen. if a power constraints of 5 samples per epoch permitted this.

Fig. 6.7 presents the results of the non-parametric model based harvester placement algorithm (algorithm 9) for datasets 1 and 3. These datasets correspond to the left feet of different users and illustrate the variability between users. Also note that the results correspond to the harvester platform behavior under uniformly thick DEs. The figure grays out harvesters that were not shortlisted into the DE array. Their locations mostly correspond to the arch of the foot where little input pressure with significant variance is seen, making them largely unproductive and difficult to predict. However it is noteworthy that across all datasets, only a few samples (between 10 and 15) were required to cover what appears to be a vast majority of harvesters in the DE array.

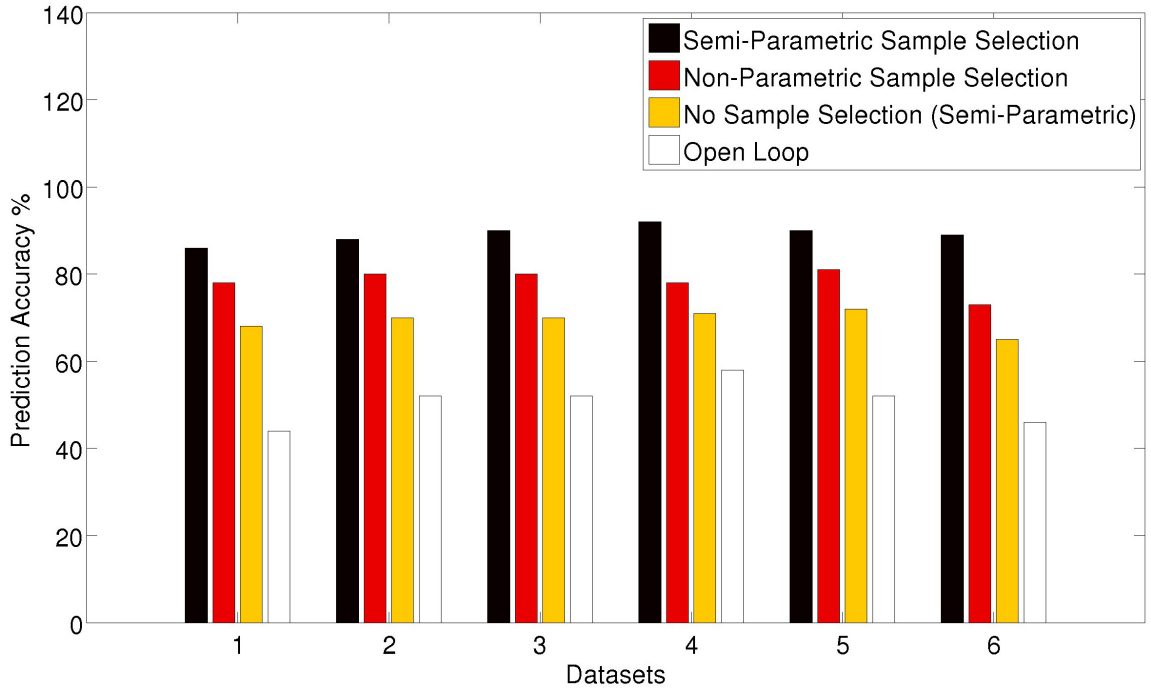


Figure 6.8: Bar Plots for the prediction accuracy of different sample selection and charge timing prediction techniques over the testing subsets of each of our datasets

Next, we assessed the necessity and the performance of different predictive algorithms. In Fig. 6.8, we compare the performance over the testing subset of 4 sampling subset selection techniques: (i) The semi-parametric adaptive control technique, (ii) the non-parametric adaptive control technique (iii) a semi-parametric adaptive system where all available samples are provided to the cox-regression algorithm, and, (iv) an algorithm that estimates charge timing of each harvester as the median epoch, over the training dataset, at which its net energy output was maximum. The comparison is shown for all 6 datasets. It is clear that the semi-parametric adaptive system yields the most energy. The prediction accuracy of the third technique over the testing subset does not rise beyond 75%. However, its accuracy over the training subset was more than 98% for each of the datasets. Clearly this is a case of overfitting. The forth technique perfoms the poorest owing to its non-adaptiveness - it fails to adapt to the variability in the user's stride and is easily outperformed by the adaptive algorithms.

The non-parametric adaptive technique performs well, predicting at about 80% accu-

racy. Yet it it outperformed by the semi-parametric approach. The reason is that the non-parametric technique does not account for localized temporal correlation between the pressure observed at neighboring pedar locations. As we discussed in chapter 2, a sample usually correlates well with the pressure observed at other sensor locations for upto a few epochs ahead of it. However, the non-parametric algorithm uses a single sample to predict harvester yield at all epochs for a harvester (i.e. the harvester's entire profile), resulting in poor prediction of harvester yield for a number of epochs. In contrast, using a different sample to predict a harvester's yield at different epochs was intentionally avoided - when the yield of a harvester at distinct epochs are predicted by distinct samples, the predicted values are likely to be largely inconsistent with each other, thereby increasing the error of prediction from the harvester's profile. Note that in comparison with the PCSS algorithm in chapter 2, this effect is significantly amplified by the non-linearity of DE behavior. The semi-parametric technique on the other hand applies multiple regression to limit the prediction errors and is able to improve prediction performance as a result.

Table 6.1 compares the average aggregate energy output of our proposed system, under both DE configurations (with and without location-specific thickness) and both adaptive prediction techniques, for each of the datasets. Owing to the superior material properties of DEs and the accuracy of our algorithm, our system outperforms the piezoelectric-driven system in [PS05] that produced 20mJ per stride. Also, across our datasets and with both our configurations, we are able to satisfy the energy demands of our target applications (32mJ). Furthermore, the semi-parametric algorithm's inbuilt penalties for overfitting produces minimal difference between performance over the training and testing subsets for all datasets ($< 1\%$ on average). The non-parametric technique does not consider the quality of fit, and yields a larger difference in performance between the testing and training subsets ($< 5\%$ on average).

Another question we sought to answer was whether there was something inherent in the samples afforded by the PCSS algorithm in chapter 2 that yielded such high outputs. As it turns out, the answer is yes. Fig. 6.9 shows the prediction accuracy from the semi-parametric technique from using between 1 and 5 samples per epoch restricted to the samples produced

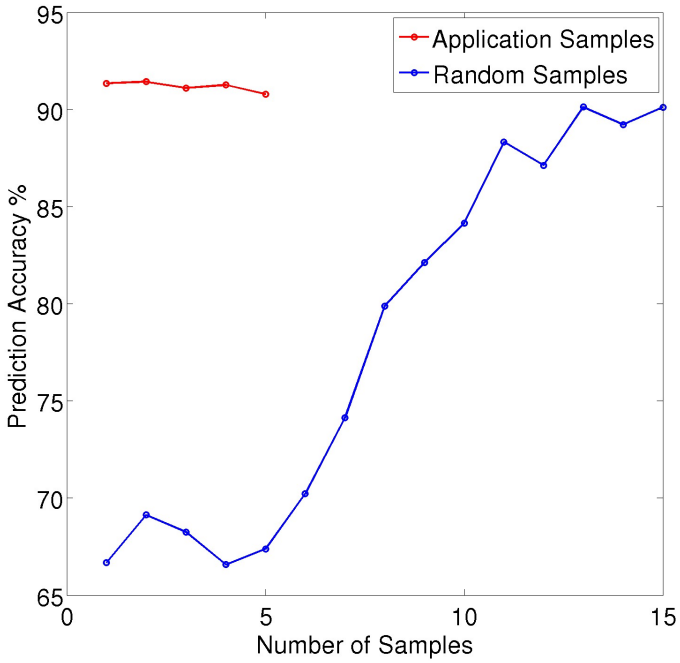


Figure 6.9: Performance of our propose predictive sample subset selection algorithm under different power constraints

by the PCSS algorithm in chapter 2. It also shows the prediction accuracy from the same technique when the samples are randomly chosen at each epoch. Note that the sample subset

Table 6.1: Performance Comparison of Foot Strike Energy Harvesting Output.

Dataset	Net Energy Output (mJ) - Layers, SP	Net Energy Output (mJ) - Layers, SP	Net Energy Output (mJ) - Uniform, NP	Net Energy Output (mJ) - Uniform, NP
(1) Left 2	107.1	97.1	91.5	83.2
(2) Right 2	120.5	109.5	98.3	90.1
(3) Left 3	86.9	77.24	68.0	62.1
(4) Right 3	81.9	69.4	69.5	60.7
(5) Left 4	90.0	80.8	71.8	65.2
(6) Right 4	80.8	66.3	71.2	59.3

selection algorithm was run on the chosen samples for all cases. Both curves correspond to averages across 3 of our datasets. The blue curve in Fig. 6.9 shows that a few random samples per epoch are insufficient in producing accurate predictions. In these cases, the prediction accuracy depends mostly on the baseline hazard function as no samples are chosen into the predictive subset. As the number of samples per epoch increases beyond 12, the results improve to recognizable levels!

6.8 Conclusion

We have designed and modeled a novel energy-harvesting platform to harvest foot-strike energy and drive low-power wearable and mobile systems. We laid out harvester configuration and control techniques to leverage characteristics of the user's gait and optimize energy output. Towards optimized control, we have proposed 2 novel and adaptive statistical charge-timing prediction techniques, including corresponding training algorithms.Finally, we have validate the performance of our algorithms over experimentally collected datasets of foot pressure produced by different users.

CHAPTER 7

Addressing Biosignal Data Sharing Security Issues with Robust Watermarking

7.1 Introduction

The rising maturity of Body Area Networks (BANs) and their underlying technologies is increasing the feasibility of their large-scale deployment over the long term in a variety of domains including health, military and sport. Self-contained, battery-powered embedded systems equipped with a diverse set of sensors, and capable of wireless communication, BANs promise continuous and nearly-unobtrusive monitoring of the wearer's physiological and biomechanical signals and the inference of her/his activity, behaviors and health therefrom. However, several requirements are still being tackled towards successful large-scale and long-term adoption of BANs. These include improvements in the design and configuration of existing systems so they may afford as yet unachieved capabilities and improved performance, such as improved sensing technologies, signal processing techniques, communication, data storage and power-management techniques [HPB09], as well as, enhanced infrastructure support for the deployed BANs and management of the data they collect.

One of the most important infrastructure requirements in the domain of remote health monitoring BANs is the secure collection and dissemination of medical data [HPB09]. Data security desiderata in this application domain are not limited to ensuring the confidentiality and integrity of medical data that has been logged to a data sink. Requirements also arise from the need to provide the data owner (BAN user / patient) and the data consumers (healthcare providers, insurance companies, medical research facilities) secure control over the data as it is shared between these various stakeholders.

Several benefits of adopting BANs for remote health monitoring exist beyond monitoring patient health while she/he is under medical observation. The detailed data that is continuously collected by the BAN is often logged to the patient's health records and can see multiple uses thereafter. The data can be used by the BAN user's healthcare providers to monitor long-term trends in the patient's health, with the temporally and diagnostically rich measurements offered by BANs enabling shortened time-to-treatment and reduced frequency and duration of hospitalization [PJB12]. It can be used for insurance purposes as well, as supporting documentation in insurance claims and as measures in assessing coverage and indemnity requirements. Data collected periodically with BANs over long periods may also be selectively shared in lieu of compensation with medical researchers conducting clinical studies.

Although these use-cases for BAN collected patient health data promise to propel preventative healthcare, several security concerns emerge from the necessity to share large amounts of medical data. For one, the infrastructure may need to provide the ability to establish that a dataset was collected from a specific patient. With this ability, the data owner (or patient) can prove ownership of her/his data if the ownership is challenged. Similarly, the data consumer can utilize this feature to track ownership of the datasets he has access to. Second, having shared her/his data with different consumers, the data owner may want to track the movement of this data even after it has changed hands several times. With this facility, the data owner can trace un-authorized sharing of her/his data. Third, data owners may be required to verify that their data has not been altered before being shared between consumers, a process known as content authentication.

While cryptography-based systems have been proposed to address these security concerns, the security of these solutions itself lies in the strength of their respective key-management protocols. Further, a malicious consumer may come into possession of the decryption keys through legitimate channels only to then illegitimately claim ownership of the data and subsequently share it with other consumers, possibly after altering it. Instead, in this chapter we offer digital watermarking as an alternative solution for these problems. Watermarking is defined as the practice of imperceptibly altering a Work to embed a message about that

Work [CMB08]. As we shall see, embedding identifying information regarding the data owner and/or summary information regarding the physiological signal into the signal itself enables proof of ownership, data tracking and content authentication. Embedding this information in the data not only obviates the need for key management but also increases overall security and reduces the transmission and storage overhead necessary to achieve data security.

Furthermore, watermarking supports these capabilities while constraining the extent to which the resulting changes to the measured physiological signal are perceivable. This translates to minimal changes to the signal waveform. More importantly, watermarks may also be applied without compromising the diagnostic quality of the medical data. Here, it is important to note that diagnostic quality of the data isnt based on the extent to which the watermarked and original data measurements differ. Rather, it is based on the consumers ability to accurately estimate medical diagnostic metrics, and the signal features that they depend upon, from the watermarked data. While pertinent functional tools and tests will inform the selection of those medical diagnostic metrics, in the context of mobile health monitoring these metrics are estimated from features of the physiological and/or biomechanical signals sampled by BANs. Therefore, as long as the data can be watermarked without altering the signal features and diagnostic metrics that are derived from it, the diagnostic value of the data is left unaffected by the watermarking process.

In this chapter, we detail a linear programming based technique to robustly watermark physiological signal data while altering it in an imperceptible manner. It is neccesary that the watermarking process produce robust watermarks so that they are not easily erased by malicious data consumers, and in this sense the robustness of a watermark is key to the level of security that the watermarking technique will provide. We study the performance of the technique in the context of multiple datasets collected by the gait stability monitoring HERMES BAN [NAH10] (see section 1.2). In this application context we study several tradeoffs inherent to watermarking algorithms including tradeoffs between embedding robustness and the amount of information that is embedded, between the extent to which imperceptibility is achieved and robustness is afforded by the watermark, and, the extent to which characteristics of the patient's physiological and biomechanical signals can impact the performance of

the watermarking process.

7.2 Related Work

The confidentiality and integrity of BAN collected patient data during exchange and storage within the network, as well as over the course of transmission outside the network for post-processing and storage at the data sink, have been identified as a key security issue for BAN deployments and addressed by a number of studies [LLR10] [LYG13]. Further, secure sharing of patient health records has been studied in the context of data encryption-based solutions [LYZ13] [LZG12]. In [LYZ13], the authors proposed a patient-centric framework that provides data access control for patient health records that are stored on semi-trusted servers. The framework provides mechanisms based on Attribute Based Encrypyion (ABE) for scalable key management, flexible access for multiple classes of data consumers, and efficient user revocation. In [PPM13], the authors propose a detailed data provenance framework to collect and share provenance metadata for patient health records, to help data consumers verify the accuracy and authenticity of the data and track its origins and changes made to it. However, the framework does not define and address a threat model.

Watermarking techniques have been thoroughly studied as a means to achieve proof of ownership and transaction tracking [CMB08]. Primarily, these techniques have been studied in the context of image, video and audio watermarking. These include least significant bit alteration methods, quantization methods and signal decomposition based methods, where the general idea is to decompose the data in the spatial and/or spectral domains and alter the coefficients of the decomposed signal in a manner that maximizes the embedded information while curtailing the visually and/or audibly perceivable changes. These watermarking techniques have been extended to the domain of medical images, as well as physiological signal measurements including ECG and EEG signals. The authors of [KF01] compare and contrast three watermarking techniques with regards to their ability to verify EEG signal integrity after noise contamination resulting from communication. The authors of [IKS11] propose an LSB watermarking scheme in support of proof-of-ownership for ECG signals. However,

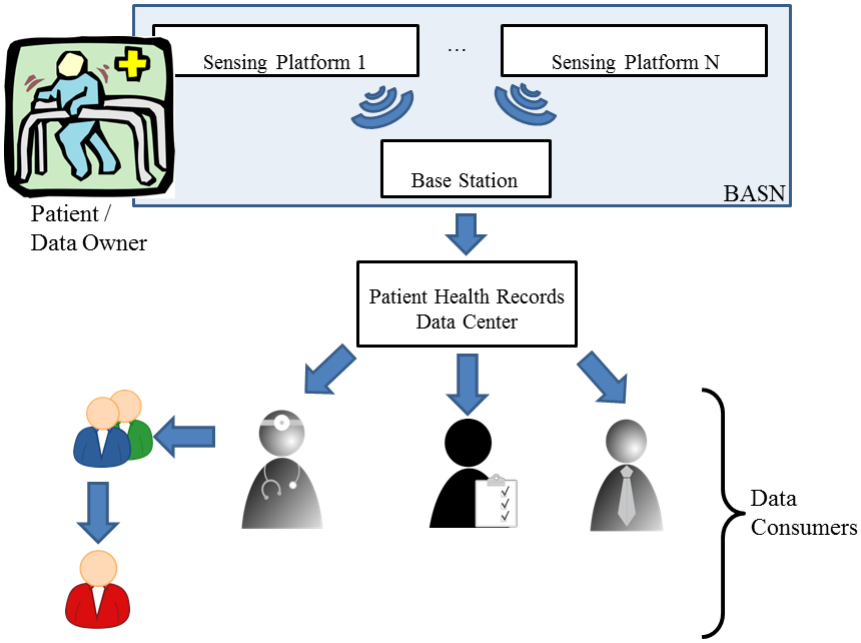


Figure 7.1: Body Area Sensor Network Architecture.

LSB watermarks provide poor robustness to malicious alterations. The authors of [KFS10] describe a spread spectrum watermarking scheme that embeds robust and imperceptible watermarks into ECG signals. However, such a scheme addresses security considerations only during communication of the data rather that over the course of sharing it. To our knowledge, our work is the first to address robustness and imperceptibility of watermarks in the context of medical data to address security issues arising from data sharing rather than communication.

7.3 Preliminaries

7.3.1 System Architecture

Fig. 7.1 illustrates a BAN architecture in the context of its data usage. A BAN typically consists of several body worn sensors/sensing platforms measuring physiological and/or bio-mechanical signals continuously and wirelessly transmitting the measurements over a short range to a base station such as a smart phone. The data is then forwarded by the base

station to a data sink that logs the it to the patient's health records. A commmon use-case is one where the sensing platforms and the base-station collaboratively apply signal processing techniques to detect "anomalous" health events and alert the patient and/or the healthcare provider(s). Aside from such use-cases involving real-time data consumption, the data may also be consumed offline by its owner, her/his healthcare providers, insurance providers, family, research orgianizations, etc. Further, the data may be shared between data consumers as well, after authorization by the data owner.

We study our watermarking technique in the context of HERMES. Typically, HERMES continously samples each of the plantar pressure sensors at 50Hz as the patient ambulates, and forwards the data to a smart phone via a low-power Bluetooth radio. Features extracted from the measurements of HERMES sensors can be used to compute several functional gait metrics. For example, the foot contact metric of the GARS-M gait assessment scale [WM99] evaluates the heel strike angle, which can be estimated by HERMES based on the difference in time between heel strike and forefoot strike. Similarly, the staggering metric from the GARS-M scale determines laterally directed loss of balance, and can be monitored based on the plantar pressure difference between lateral areas of the foot. Other metrics that can be gleaned include the inter foot-strike interval that is relevant to the step symmetry and continuity measures of the Tinetti gait assessment tool [WM99]. Finally, the spatial average of peak plantar pressure is of interest in monitoring diabetic patients for plantar ulcers. Signal processing for functional gait metric computations based on the data collected by HERMES either occurs at the base station (smart phone), or is performed offline.

7.3.2 Thread Model

In all three security use-cases that we address, namely proof of ownership, data tracking and content authentication, our goal is to thwart the effort of a malicious data consumer attempting to erase the embedded watermark and possibly replace it with his own. By successfully achieving this in the proof of ownership and data tracking use-cases, the malicious data consumer gains the ability to illegally share data that is not owned by him and cover his

tracks while doing so. In the case of content authentication, the malicious consumer can alter the data without leaving any evidence of the transformation. For example, this can be used by a disingenuous healthcare provider looking to avoid liability lawsuits, or by a duplicitous insurance provider looking to deny legitimate claims. Therefore it is important to validate the robustness claim of our watermarking procedure. Towards this end, we assume that a malicious data consumer's attempt to erase the watermark involves addition of gaussian noise to the watermarked signal in order to corrupt the watermark beyond recognition.

7.4 Robust Watermarking of Biomedical Signals

Fig. 7.2 outlines the blind watermarking system we will use to address the proof of ownership, data tracking and content authentication requirements for biomedical signal data collected by BANs. The system consists of a watermark embedder used by the data owner to imperceptibly and robustly embed a watermark, and a watermark detector can be used to test for the existence of the watermark in biomedical signal datasets, as required. The embedder uses an encoding algorithm to generate an encoding vector, P, for the watermark, w, that is most suitable for the biomedical signal vector S. Suitability of the encoding is governed by the robustness and imperceptibility afforded by it once it is embedded into the original signal. The embedder then embeds the encoding vector into the original signal vector, S, to generate an altered biomedical signal, S_w, which now includes the watermark.

Over the course of sharing the watermarked signals, a malicious consumer may further alter the signal into $S_{w,N}$ by adding noise to it in an attempt to erase the watermark. Once this version of the signal needs to be validated by checking for the existence of the data owner's watermark, the watermark detector accepts the original watermark and the corresponding encoding vector from the data owner. It then confirms (i) that the encoding vector P actually encodes the watermark w, and (ii) that the encoding vector exists in the signal under test $S_{w,N}$. If both these tests are positive, the detector issues a decision b indicating a positive detection of the watermark w in the signal. Otherwise, a negative detection is indicated. Note that encoding vector P and the original signal S are kept secret

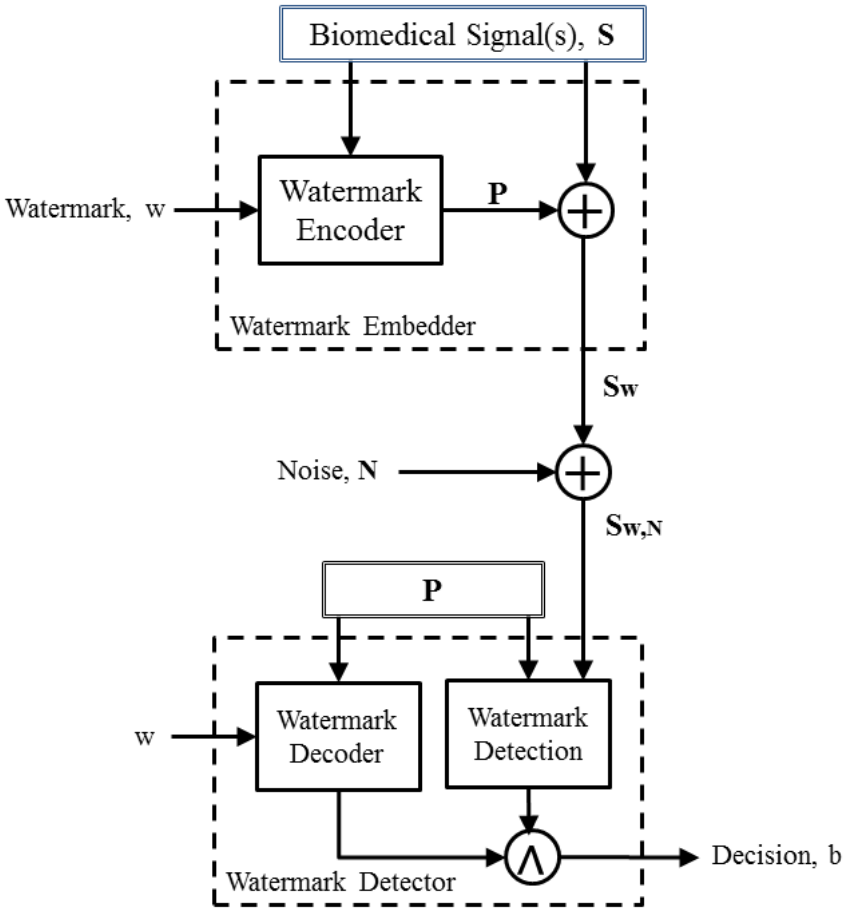


Figure 7.2: Watermarking System.

from the data consumers. Similarly, while testing detection, the tested signal $S_{w,N}$ is not disclosed to the data owner lest she manipulate the encoding vector to falsely prove that data which she does not own is actually hers. Also note that the detection system here is a blind detector. That is, the original signal S is never disclosed during the decoding process. Alternatively, we can replace it with an informed detector which is provided the original signal during the detection procedure. It can then apply the encoding algorithm to the original signal and independently generate the encoding vector P.

To enable proof of ownership, a watermark identifying the data owner is embedded into the medical signal. Proving ownership is as simple as proving that the data owner's identifier is embedded in the signal when it is tested. As long as the watermarking procedure produces robust watermarks, the watermark detector should indicate the watermark's presence. Data

tracking is achieved with the data owner embedding distinct identifiers in each shared copy of her/his biomedical signal data. For example, data shared with one's healthcare provider is embedded with a different watermark than the one shared with the insurance provider or with one's friends. If the data ends up in the possession of an unauthorized party, the data owner can iterate through the set of watermarks she/he has used and identify the legitimate but malicious data consumer that released the patient's data without her/his authorization. Finally, content authentication is achieved by embedding the data with a watermark comprised of a cryptographic hash of the original signal. On successful detection of the watermark by the watermark detector, authentication of the data is achieved by removing the watermark from the watermarked signal and comparing it to a cryptographic hash of the signal that remains. If this is, in fact, the original signal, the watermark and hash will match. If the signal has been altered, either the watermark detection will fail, or the watermark will not match the hash.

Before we delve into details regarding what is required from the watermarking procedure and how it is achieved, we highlight an important property of most biomedical signals which we leverage in our watermark encoding algorithm. Bio-medical signals are often segmentable in a straight-forward manner since they arise from human physiology. For example, ECG signals are based on heart-beats and have a distinct waveform (e.g. QRS complex in ECG) that is well-studied and whose origin is well-understood. The same is true of EEG waveforms. In the case of bio-mechanical measurements, human movement and gait usually consists of periodic and repeated sequences of movements which can be used to segment spatio-temporal bio-signals comprised of accelrometer, gyroscope and pressure/force measurements at the joints and limbs. For example, the foot plantar pressure measurements of HERMES can be attributed to the well-studied stance and swing gait phases. Stance is the phase of human gait when the foot is in contact with the ground, while swing in the phase when the foot is in the air. Therefore, the pressure measurements at all sensors can be segmented based on these phases. Consequently, we denote sensor s_i's measurement at epoch j of the segmented signal as sample s_{ij}. Encoding a watermark w into the samples can be equated to perturbing each sample s_{ij} by some amount p_{ij}:

$$s_{w_{ij}} = s_{ij} + p_{ij} \tag{7.1}$$

Now, there are three requirements for the watermarking procedure. First, the watermark should be robustly embedded in the original signal. Recall that this means the watermark is detectable even if the watermarked signal is altered maliciously or unintentionally by addition of gaussian noise. To satisfy this requirement, robust detection of the watermark is achieved with a dot-product based detection procedure. Here, detection of the watermark depends on how well correlated the watermarked signal and the watermark are. The detection metric z is calculated as:

$$z = S_t \cdot P \tag{7.2}$$

where S_t is the signal being tested for the watermark. If $S_t = S_w$, then equations (7.3a) - (7.3c) outline why the detection metric has a high value. Since P is highly correlated with itself, the value of $(P \cdot P)$ is high. In contrast, an arbitrary original signal S is unlikely to correlate well with P especially in high-dimensional space. Therefore, $(S \cdot P)$ is very likely to have a small value. Further, if S_t does not embed the watermark w that has been encoded to P, then z is highly likely to evaluate to a low value since S_t and P will exhibit low correlation. Therefore, z is most likely to evaluate to a high value only when the tested signal S_t embeds the encoded watermark P.

$$z = S_w \cdot P \tag{7.3a}$$
$$= (S + P) \cdot P \tag{7.3b}$$
$$= (S \cdot P) + (P \cdot P) \tag{7.3c}$$

This procedure also yields robust watermarks for the reasons illustrated in equations (7.4a) - (7.4c). Here, the tested signal S_t is a maliciously altered version of the watermarked signal with the goal of erasing the watermark. Assuming the malicious alteration is in the form of additive gaussian noise, the watermark should still be detectable. The reason is similar to

our discussion above - additive gaussian noise is unlikely to correlate well with the encoded watermark P. Therefore, although the addition of such noise might change the detection metric z, this change is unlikely to be very significant.

$$z = S_{w,N} \cdot P \tag{7.4a}$$
$$= (S + P + N) \cdot P \tag{7.4b}$$
$$= (S \cdot P) + (P \cdot P) + (N \cdot P) \tag{7.4c}$$

The dot-product based technique evaluates the existence of the watermark in the tested signal based on the detection metric, z. Specifically, if the metric value is greater than some threshold τ, positive detection occurs. Otherwise negative detection occurs. However, not that there is always a chance that the watermark detector will produce incorrect results. Either the detector will indicate the presence of the watermark when it does not exist, which is called a false positive. False positives can occur when the original signal happens to be well correlated with the encoded watermark by chance, driving the detection metric beyond the threshold. Alternatively, the detector may fail to detect the watermark even though the signal has includes the watermark, which is known as a false negative. This occurs when the original signal exhibits a strong negative correlation with the encoded watermark by chance, and reduces the detection metric below the threshold, thereby eluding detection. Therefore, the false positive and false negative rates play and imporant role in achieving a robust watermarking scheme and must be taken into consideration by the encoding algorithm.

The second requirement for a watermarking procedure is that the watermark must be successfully embedded in the original signal. If it cannot be embedded, the procedure must indicate this. Further, the larger the watermark that can be successfully embedded, the stronger the procedure's ability to address the use cases. For example, a larger watermark in the content authentication use-case can support a larger cryptographic hash of the original signal making for a much stronger guarantee regarding malicious alterations to the biosignal data. As we shall see, tradeoffs exist between the size of the watermark that can be successfully embedded and the robustness with which it can be embedded.

Third, the embedding should only result in imperceptible alterations of the original signal. As we have discussed in section 7.1, imperceptible embedding in the context of biomedical signals translates to impereceptible changes to the signal waveform which we shall refer to as signal waveform imperceptibility. It also translates to no loss in the diagnostic value of the signal data which we call diagnostic imperceptibility. The constraints on diagnostic imperceptibility depend on the diagnostic metrics and their underlying signal features which will be derived from the data. The goal of diagnostic imperceptibility is to ensure that the diagnostic metrics that will be derived from the watermarked signals are exactly the same as those derived from the original signals, thereby yielding no compromise to diagnostic fidelity by the watermarking process.

We now describe a linear-programming approach for watermark encoding that addresses these three requirements for a biomedical signal watermarking scheme.

7.4.1 Constraint: Embedding the Watermark

To address the successful embedding requirement, it is necessary that the watermark w be corrected encoded by the vector P. This is achieved by the constraint in equation (7.5). The watermark is encoded as a set of perturbations over all sensors and for all epochs of the segmented signals. These perturbations can take on arbitrary values, both positive and negative, subject to the other constraints in the Linear Program (LP). However, the constraint in equation (7.5) requires that the perturbations sum to the watermark value w. This is also one of the 2 conditions checked by the detector - the watermark and its encoding must be coherent.

$$\sum_{i,j} p_{ij} = w \tag{7.5}$$

7.4.2 Constraint: Signal Waveform Imperceptibility

We address imperceptibility of the changes to the original signal waveform in two ways. First, each perturbation p_{ij} should be within a small percentage of the original sample s_{ij}

(equation (7.6)). In other words, the perturbations are bounded by $[100 - \epsilon \times 100, 100 + \epsilon \times 100]\%$ of the original samples.

$$-\epsilon s_{ij} \leq p_{ij} \leq \epsilon s_{ij} \quad \forall i, j \tag{7.6}$$

The second signal waveform imperceptibility constraint applies to signal smoothness after perterbuation. Here, a bound is applied to the first derivative of the altered signal (equation (7.7a)). Equations (7.7b) - (7.7c) translate this constraint to LP form. Together, these two sets of constraints limit changes to the signal waveform.

$$|p_{ij-1} - p_{ij}| \leq \Delta \qquad \forall i, j > 1 \tag{7.7a}$$

$$p_{ij-1} - p_{ij} \leq \Delta \qquad \forall i, j > 1 \tag{7.7b}$$

$$p_{ij} - p_{ij-1} \leq \Delta \qquad \forall i, j > 1 \tag{7.7c}$$

7.4.3 Constraint: Diagnostic Imperceptibility

We address diagnostic imperceptibility in the context of the maximum amplitude and foot contact signal features of plantar pressure signals measured by HERMES. The maximum amplitude feature is computed as the maximum pressure observed at each sensor over the stance phase. The maximum pressure diagnostic metric is then derived as the mean maximum pressure taken over all sensors. Similarly, the staggering diagnostic metric is derived as the average difference between the maximum amplitude features of lateral areas of the foot sole. To prevent loss of diagnostic accuracy for these metrics, the maximum amplitude signal feature is computed from the watermarked signal (equation (7.8a)), and the maximum amplitude and staggering diagnostic metrics derived from these signal features are constrained to the corresponding metric values derived from the original signals (equations (7.8b) - (7.8c)). Here, A and B correspond to the maximum amplitude and staggering diagnostic metrics based on the original signals, and a_i and b_i are the corresponding coefficients for each sensor i.

$$\max_j(s_{ij} + p_{ij}) = M_i \qquad \forall i \tag{7.8a}$$

$$\sum_i (a_i M_i) = A \tag{7.8b}$$

$$\sum_i (b_i M_i) = B \tag{7.8c}$$

Note that equation (7.8a) is not in LP form and must be transformed before input to an LP solver (equation (7.9)). However, equation (7.9) does not provide the true maximum (M_i) of the perturbed samples. To arrive at the true maximum, m_i must be minimized, and this will be addressed in the following subsection.

$$s_{ij} + p_{ij} \leq m_i \qquad \forall i, j \tag{7.9}$$

The foot contact signal feature is derived as the sampling epoch for the segmented signal at which foot contact is sensed at a sensor. Note that foot contact at different sensors will be different. For example, contact at the heel occurs much before contact at the toes. Contact is sensed at a sensor when its pressure measurements exceed some threshold tr. To guarantee that the foot contact based diagnostic metrics, namely heel strike angle and inter-footstrike interval, are left unaltered after the watermark has been embedded, we ensure that at each sensor the perturbed signal at all epochs prior to the foot contact epoch is less than the threshold tr (equation (7.10a)), while the perturbed signal at each sensor at the actual foot contact epoch is greater than or equal to tr (equation (7.10b)). Here, the foot contact epoch derived from the original signals is denoted t_{fc}. In this way, the foot contact signal feature derived from the watermarked signal at each sensor will be identical to that derived from the original signal.

$$s_{ij} + p_{ij} < tr \qquad \forall i, j < t_{fc} \tag{7.10a}$$

$$s_{it_{fc}} + p_{it_{fc}} \geq tr \qquad \forall i \tag{7.10b}$$

7.4.4 Objective Function: Maximizing Robustness

Finally, the robustness requirement is addressed by the objective function. Note from our discussion in section 7.4 that minimizing false positives and false negatives translates to keeping $S \cdot P$ as close to 0 as possible. If it is large and positive, false positives will result. If it is large and negative, false negatives will result. In other words, $|S \cdot P|$ must be minimized. Equations (7.11a) - (7.11c) translate this objective to LP form. Note that the objective function in equation (7.11c) not only aims to limit false positive and false negative rates, but also addresses the maximum amplitude related constraint from the previous subsection.

$$\sum_{i,j} s_{ij} p_{ij} \leq X \tag{7.11a}$$

$$-\sum_{i,j} s_{ij} p_{ij} \leq X \tag{7.11b}$$

$$minimize: \quad X + \sum_{i} m_i \tag{7.11c}$$

7.5 Validation

We validate the ability of our linear programming based approach to robustly watermark biomedical signals by applying the technique to three plantar pressure datasets collected with HERMES. The datasets correspond to plantar pressure measurements of male and female subjects with different weights and gaits, collected over a few hundred footsteps as the subjects walked. The LP is applied independently to the plantar pressure signals at each footstep of each subject.

Fig. 7.3 shows the watermarked signals from 2 sensors over a single footstep of one of the plantar pressure datasets, where w is 3300, ϵ is 0.1, Δ is $\delta \times max(s_{ij-1}, s_{ij})$, and δ is 0.05. Note that we will continue with this definition of Δ for the rest of the chapter. Several observation with regards to the requirements for the watermarking procedure are apparent in Fig. 7.3. The signal waveforms are very similar before and after watermarking for both sensors. This observation is also apparent for aggregated measurements over all sensors

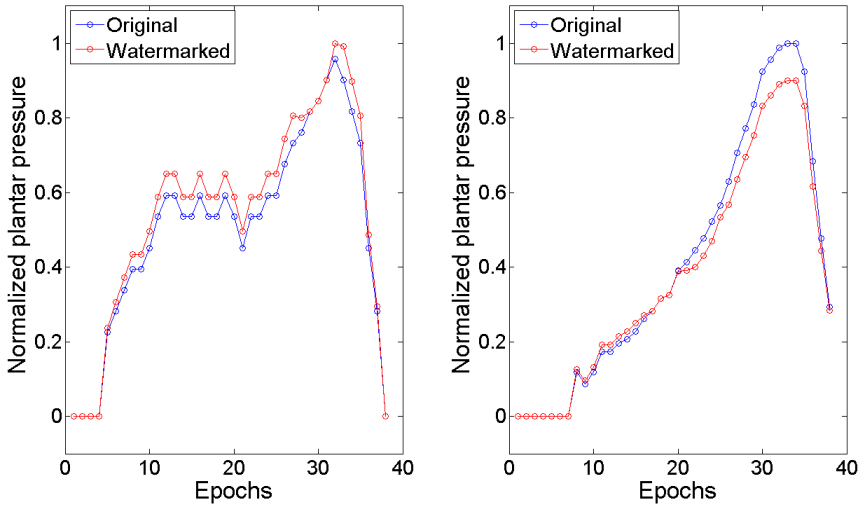


Figure 7.3: Original vs. Watermarked plantar pressure measurements for 2 sensors over the stance phase of a single footstep from dataset 1.

(Fig. 7.4). Further, Fig. 7.3 also shows that the diagnostic imperceptibility requirement for the foot contact signal feature dictates minimal changes to the waveforms at the initial epochs prior to foot contact, which we observe in both plots. In contrast, the LP solver achieves imperceptibility of the maximimum amplitude signal feature in a different way. The maximum amplitude of the watermarked signal of some sensors are less than that of the original signal (Fig. 7.3, right plot), while the maximum amplitude of the watermarked signal of other sensors are more than that of the original signal (Fig. 7.3, left plot). However, these difference are balanced out to achieve the overall effect of identical diagnostic metrics before and after watermarking.

Next, we address the false positive and false negative rates of our watermarking procedure in the context of watermarking all plantar pressure signals for all footsteps in dataset 1, when w was set to 4000, ϵ to 0.1 and δ to 0.05. Fig. 7.5 plots histograms of the detection metrics before watermarking (i.e. $z = S \cdot P$) and after watermarking (i.e. $z = S_w \cdot P$). As expected, the detection metric is generally larger after watermarking. Further the value of the detection threshold τ dictates the false positive and false negative rates. If $\tau = 0$, the false positive rate is 100% while the false negative rate is 0%. In contrast, if $\tau = 200000$, the false negative rate is 100% while the false positive rate is 0%. Varying τ in the range $[0, 200000]$ results

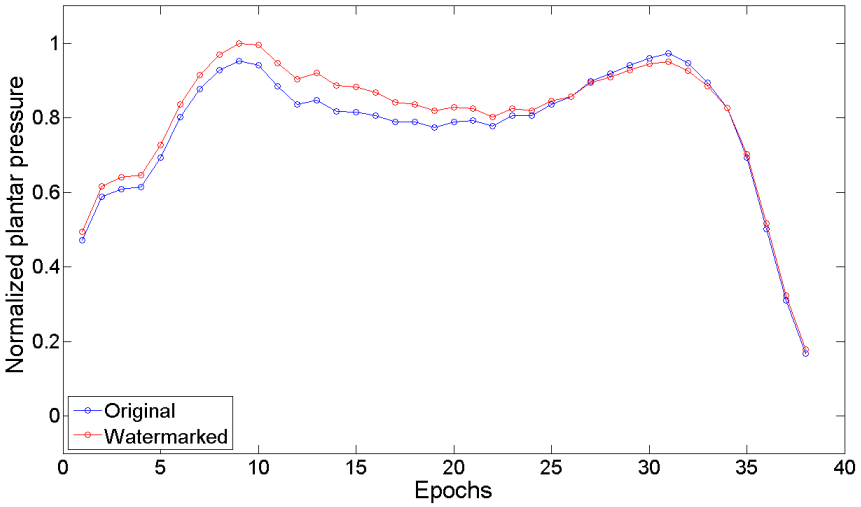


Figure 7.4: Original vs. Watermarked plantar pressure measurements aggregated over all 99 sensors over the stance phase of a single footstep from dataset 1.

in a range of false positive rates between 100% and 0% and false negative rates between 0% and 100%. To achieve a balance between the two, for the rest of the chapter, we select a τ that minimizes the product of the false positive and false negative rates.

Based on this derivation of τ, Fig. 7.6 plots the false positive, false negative and embedding failure rates for all datasets when ϵ was set to 0.1, δ to 0.05, and w was varied. Smaller

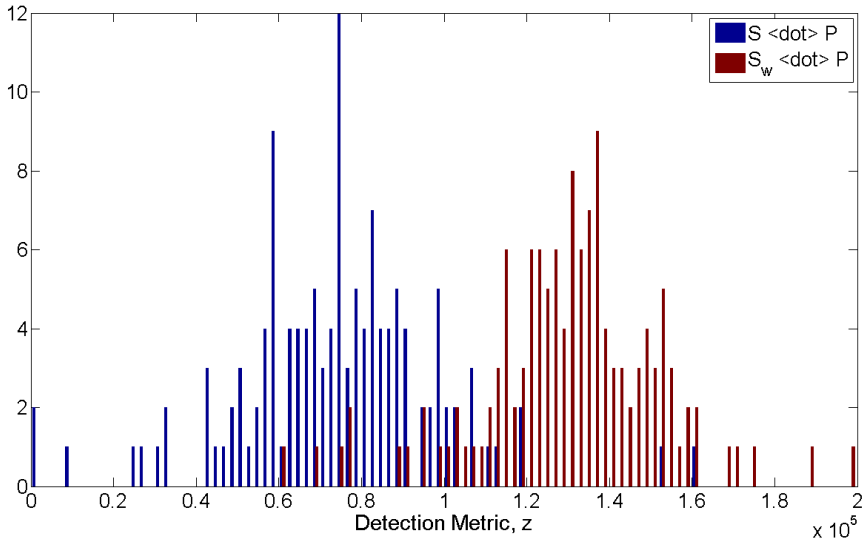


Figure 7.5: Histograms of detection metric before and after embedding encoded watermark.

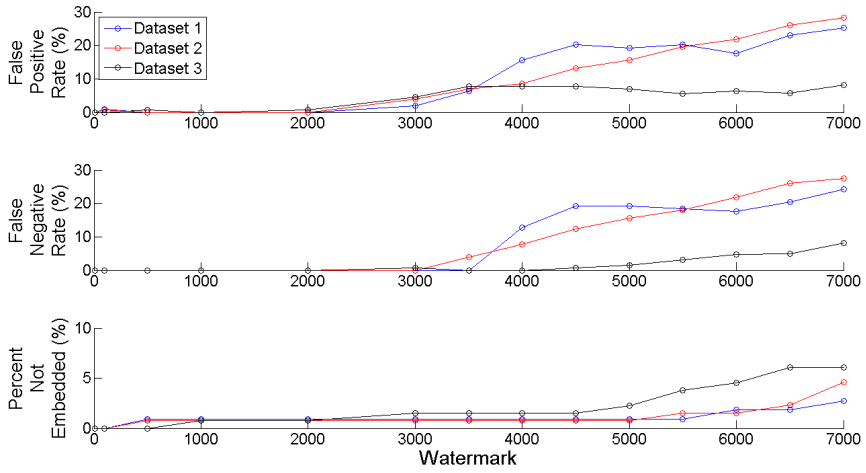
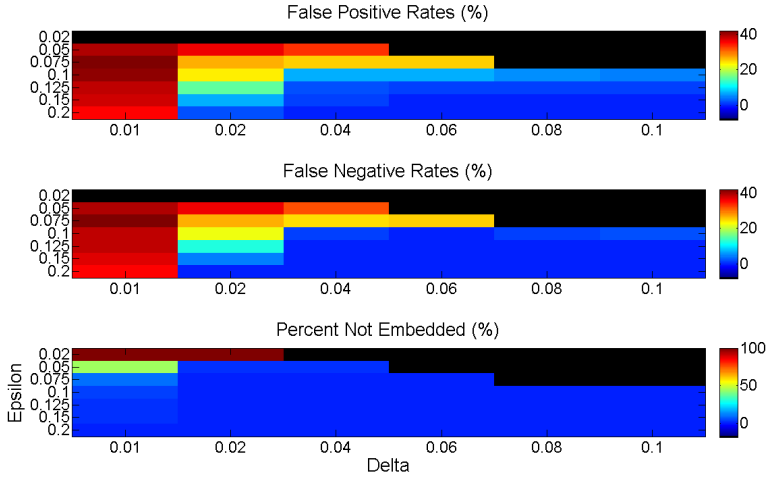


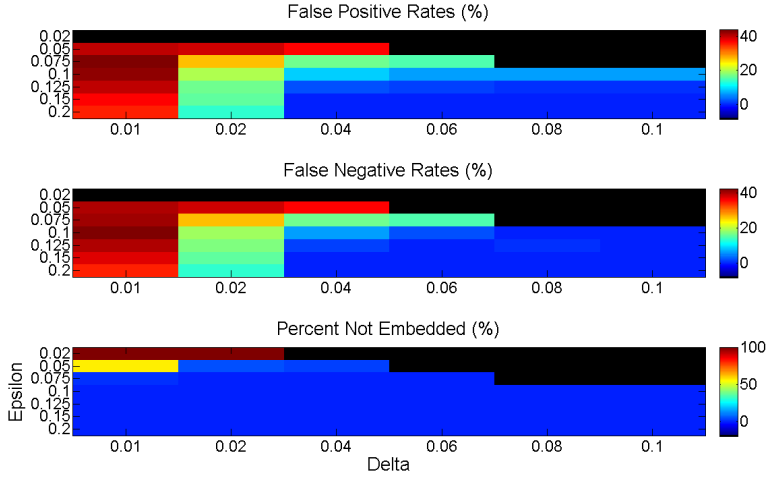
Figure 7.6: False Positive, False Negative and Embedding failure rates for different watermark (w) values.

values of w translate to a requirement of robustly and imperceptibly embedding fewer bits into the original signal resulting is lower false positive, false negative and embedding failure rates. As the amount of information to be embedded increases, it becomes more difficult to do so under the imperceptibility constraints. As a result, the procedure's ability to embed the watermark decreases leading to an increase in the embedding failure rate. More importantly, as the solution space shrinks, the encoded watermark P becomes more correlated with the original signal S resulting in poorer false positive and false negative rates. This illustrates the tradeoff between embedding robustness and the size of the watermark that can be embedded. We also observe that the extent and manner to which this tradeoff occurs depends heavily on the patient and her/his gait. While dataset 1 sees a sharp decline in robustness as w increases past 3500, the decline in robustness for dataset 2 is much more gradual. Also, dataset 3 appears to embed much more robustly compared to the other two datasets, however it also exhibits a higher likelihood of failed embedding.

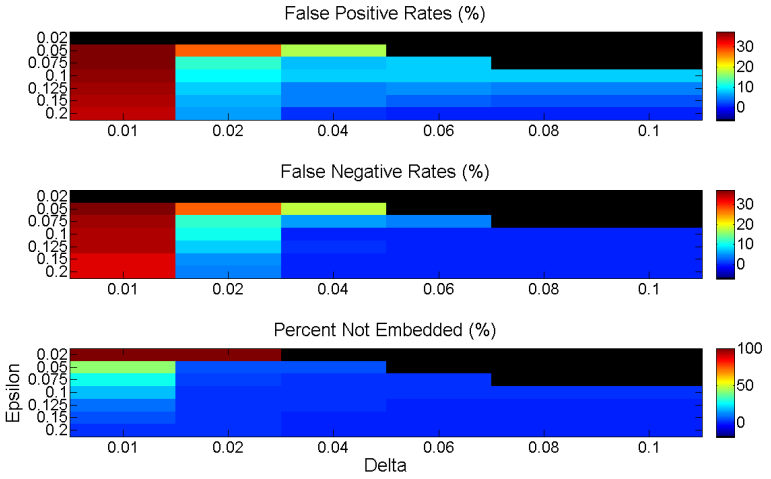
We also studied the impact of the signal waveform imperceptibility constraints on the robustness and embedding performance of the watermarking procedure. Fig. 7.7 illustrates the false positive, false negative and embedding failure rates as ϵ and δ were varied while keeping w fixed at 3500. Here, note that when $\delta > \epsilon$ the smoothness constraint has no effect. In other words, this is the same as setting $\delta = \epsilon$. Therefore, robustness and embedding



(a)



(b)



(c)

Figure 7.7: False Positive, False Negative and Embedding failure rates for different signal waveform imperceptibility constraints (ϵ, δ).

performance when $\delta > \epsilon$ has been disregarded. As expected, robustness and embedding performance improves as the signal waveform imperceptibility constraints are relaxed. We also observe that the performance is much more sensitive to the smoothness constraint (δ) than the perturbation bound constraint (ϵ). Finally, it continues to be true that robustness and embedding performance depends, to a large extent, on the patient and her/his gait, regardless of the signal waveform imperceptibility constraints.

7.6 Discussion

One drawback of our approach is that a sizeable LP has to be solved for each segment of the measured biosignal. Although LPs are polynomial-time solvable, this can be further simplified if the LP may be solved once on a statistical or mathematical model of the patient's biosignal segments, with a constant or linear-time embedding procedure applied to each segment based on the LP's solution. The fact that embedding performance relies so heavily on the properties of the patient's gait gives credence to the fact that such a model may be constructed. However, it may be difficult to ensure that all constraints may be satisfied for each individual segment, when the encoding procedure is driven by a model. This approach is currently under study.

Another advantage of such an approach is that the since the encoding algorithm is run one time (or a few times) on a model of the biosignal segments, a computationally expensive encoding procedure such an Integer Linear Program (ILP) or an Non-Linear Program (NLP) may also be used, provided there is a substantial benefit to doing so. Two such benefits are immediately apparent. First, the complexity of the encoding can be increased in a manner that increases the amount of information that can be embedded. For example, instead of requiring p_{ij} to sum to w, $|p_{ij}|$ can be required to sum to w thereby dramatically increasing the size of watermarks that may be embedded. More complex signal decomposition based techniques can also be addressed. Second, tighter diagnostic imperceptibility constraints can be constrcuted and more complex diagnostic metrics can be easily addressed. For example, when limited to an LP, the foot contact signal feature must be identical at each sensor before

and after watermarking. With an ILP, the constraint can be set up such that the signal features need not match, but the resulting diagnostic metrics must match. Less restrictive constraints will lead to improved robustness and embedding performance.

7.7 Conclusion

In this chapter, we have addressed security concerns arising from the sharing of BAN collected biosignal measurements with and between medical data consumers. Specifically, three use-cases are addressed: proof of ownership, data tracking and content authentication. We propose an LP formulation to encode watermarks in a robust manner while adhering to imperceptibility constraints that limit significant alterations to the signal waveform and prevent loss of diagnostic fidelity by the embedding process. We study the performance of this technique in the context of foot plantar pressure datasets collected by the HERMES BAN. Results of our validation studies indicate that 8-bit watermarks can be robustly and imperceptably embedded into signal segments sampled by 99 sensors for upto 1 second at 50Hz.

References

[3M11] 3M. "3M VHB^{TM} Tapes.", June 2011.

[ABG95] J. F. Antaki, G. E. Bertocci, E. C. Green, A. Nadeem, T. Rintoul, R. L. Kormos, and B. P. Griffith. "A gait-powered autologous battery charging system for artificial organs." *ASAIO journal (American Society for Artificial Internal Organs : 1992)*, **41**(3), 1995.

[AM99] Wendy K. Anemaet and Michelle E. Moffa-Trotter. "Functional Tools for Assessing Balance and Gait Impairments." *Topics in Geriatric Rehabilitation*, **15**(1):66–83, 1999.

[AWB07] L.K. Au, W.H. Wu, M.A. Batalin, D.H. Mclntire, and W.J. Kaiser. "MicroLEAP: Energy-aware Wireless Sensor Platform for Biomedical Sensing Applications." In *Proceedings of Biomedical Circuits and Systems Conference (BIOCAS)*, pp. 158 –162, nov. 2007.

[Bal] Rahul Balani. "Energy Consumption Analysis for Bluetooth, WiFi and Cellular Networks." Technical report, Electrical Engineering Department, University of California, Los Angeles.

[BBB10] Adam T. Barth, Benjamin Boudaoud, Jeff S. Brantley, Shanshan Chen, Christopher L. Cunningham, Taeyoung Kim, Harry C. Powell, Jr., Samuel A. Ridenour, John Lach, and Bradford C. Bennett. "Longitudinal high-fidelity gait analysis with wireless inertial body sensors." In *Wireless Health 2010*, pp. 192–193, 2010.

[BWC13] Md Zakirul Alam Bhuiyan, Guojun Wang, Jiannong Cao, and Jie Wu. "Energy and Bandwidth-Efficient Wireless Sensor Networks for Monitoring High-Frequency Events." In *Proceedings of IEEE International Conference on Sensing, Communication and Networking*, 2013.

[CB10] Peter R. Cavanagh and Sicco A. Bus. "Off-loading the diabetic foot for ulcer prevention and healing." *Journal of Vascular Surgery*, **52**(3, Supplement):37S – 43S, 2010.

[CBC08] Tanzeem Choudhury, Gaetano Borriello, Sunny Consolvo, Dirk Haehnel, Beverly Harrison, Bruce Hemingway, Jeffrey Hightower, Pedja Klasnja, Karl Koscher, Anthony Lamarca, Jonathan Lester, James Landay, Louis Legrand, Ali Rahimi, Adam Rea, and Danny Wyatt. "The Mobile Sensing Platform: An Embedded System for Capturing and Recognizing Human Activities." *IEEE Pervasive Computing*, **7**(2):32–41, 2008.

[CES04] D. Culler, D. Estrin, and M. Srivastava. "Guest Editors' Introduction: Overview of Sensor Networks." *Computer*, **37**(8):41–49, 2004.

[CIP06] Enrico Costanza, Samuel A. Inverso, Elan Pavlov, Rebecca Allen, and Pattie Maes. "eye-q: Eyeglass peripheral display for subtle intimate notifications." In *Proceedings of Mobile HCI*, 2006.

[CMB08] Ingemar Cox, Matthew Miller, Jeffrey Bloom, Jessica Fridrich, and Ton Kalker. *Digital Watermarking and Steganography*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2 edition, 2008.

[DGS] Robert F. Dickerson, Eugenia I. Gorlin, and John A. Stankovic. "Empath: a continuous remote emotional health monitoring system for depressive illness.".

[DGW12] A.I. Daoud, G.J. Geissler, F. Wang, J. Saretsky, Y.A. Daoud, and D.E. Lieberman. "Foot Strike and Injury Rates in Endurance Runners: a retrospective study." *Medicine & Science in Sports & Exercise*, 2012.

[DLT07] Min Ding, Fang Liu, Andrew Thaeler, Dechang Chen, and Xiuzhen Cheng. "Fault-tolerant target localization in sensor networks." *EURASIP J. Wirel. Commun. Netw.*, **2007**(1):19–19, January 2007.

[ESK11] Emre Ertin, Nathan Stohs, Santosh Kumar, Andrew Raij, Mustafa al'Absi, and Siddharth Shah. "AutoSense: unobtrusively wearable sensor suite for inferring the onset, causality, and consequences of stress in the field." In *Proceedings of the International Conference on Embedded Networked Sensor Systems*, pp. 274–287, 2011.

[fle87] "Tekscan, FlexiForce Sensors (A201 Model). http://www.tekscan.com/flexible-force-sensors.", 1987.

[FRW07] E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi. "In-network aggregation techniques for wireless sensor networks: a survey." *Wireless Communications, IEEE*, **14**(2):70–87, 2007.

[GCX08] Todd A. Gisby, Emilio P. Calius, Shane Xie, and Iain A. Anderson. "An adaptive control method for dielectric elastomer devices." *Proceedings of SPIE Electroactive Polymer Actuators and Devices*, **6927**:69271C–69271C–8, 2008.

[GGG08] H. Ghasemzadeh, E. Guenterberg, K. Gilani, and R. Jafari. "Action coverage formulation for power optimization in body sensor networks." In *Design Automation Conference, Asia and South Pacific*, pp. 446–451, 2008.

[GJ10] H. Ghasemzadeh and R. Jafari. "Data Aggregation in Body Sensor Networks: A Power Optimization Technique for Collaborative Signal Processing." In *Sensor Mesh and Ad Hoc Communications and Networks (SECON), IEEE Communications Society Conference on*, pp. 1–9, 2010.

[GMS10] Christian Graf, J Maas, and D Schapeler. "Energy harvesting cycles based on electro active polymers." *Proceedings of SPIE*, **7642**:764217, 2010.

[GP12a] V. Goudar and M. Potkonjak. "Energy-efficient sampling schedules for Body Area Networks." In *Sensors, IEEE*, pp. 1–4, 2012.

[GP12b] Vishwa Goudar and Miodrag Potkonjak. "Power Constrained Sensor Sample Selection for Improved Form Factor and Lifetime in Localized BANs." In *Proceedings of the 3rd Conference on Wireless Health*, 2012.

[HLR10] Nan Hua, Ashwin Lall, Justin Romberg, Jun (Jim) Xu, Mustafa al'Absi, Emre Ertin, Santosh Kumar, and Shikhar Suri. "Just-in-time sampling and pre-filtering for wearable physiological sensors: going from days to weeks of operation on a single charge." In *Wireless Health 2010*, pp. 54–63, 2010.

[HPB09] M.A. Hanson, H.C. Powell, A.T. Barth, K. Ringgenberg, B.H. Calhoun, J.H. Aylor, and J. Lach. "Body Area Sensor Networks: Challenges and Opportunities." *Computer*, **42**(1):58 –65, jan. 2009.

[HS10] Enamul Hoque and John A. Stankovic. "Monitoring quantity and quality of sleeping using WISPs." In *IPSN*, pp. 370–371. ACM, 2010.

[IGE00] Chalermek Intanagonwiwat, Ramesh Govindan, and Deborah Estrin. "Directed diffusion: a scalable and robust communication paradigm for sensor networks." In *Proceedings of the 6th annual international conference on Mobile computing and networking*, MobiCom '00, pp. 56–67, New York, NY, USA, 2000. ACM.

[IKS11] A. Ibaida, I. Khalil, and R. van Schyndel. "A low complexity high capacity ECG signal watermark for wearable sensor-net health monitoring system." In *Computing in Cardiology, 2011*, pp. 393–396, 2011.

[JBC10] C Jean-Mistral, S Basrour, and J-J Chaillout. "Modelling of dielectric polymers for energy scavenging applications." *Smart Materials and Structures*, **19**(10):105006, 2010.

[JBD11] Val Jones, Richard Bults, Rene De Wijk, Ing Widya, Ricardo Batista, and Hermie Hermens. "Experience with using the sensewear BMS sensor system in the context of a health and wellbeing application." *Int. J. Telemedicine Appl.*, **2011**:1–18, January 2011.

[KF01] X. Kong and Rui Feng. "Watermarking medical signals for telemedicine." *Information Technology in Biomedicine, IEEE Transactions on*, **5**(3):195–201, 2001.

[KFS10] S. Kaur, O. Farooq, R. Singhal, and B.S. Ahuja. "Digital Watermarking of ECG Data for Secure Wireless Commuication." In *Recent Trends in Information, Telecommunication and Computing (ITC), 2010 International Conference on*, pp. 140–144, 2010.

[KMG08] Andreas Krause, H. Brendan McMahan, Carlos Guestrin, and Anupam Gupta. "Robust submodular observation selection." *Journal of Machine Learning*, pp. 2761–2801, 2008.

[KP11] Duk-Jin Kim and B. Prabhakaran. "Motion fault detection and isolation in Body Sensor Networks." In *Pervasive Computing and Communications (PerCom), IEEE International Conference on*, pp. 147–155, 2011.

[KPP11] Roy D Kornbluh, Ron Pelrine, Harsha Prahlad, Annjoe Wong-Foy, Brian McCoy, Susan Kim, Joseph Eckerle, and Tom Low. "From boots to buoys: promises and challenges of dielectric elastomer energy harvesting." *Polymer*, **7976**(1), 2011.

[KPS03] F Koushanfar, M. Potkonjak, and A. Sangiovanni-Vincentelli. "On-line fault detection of sensor measurements." In *Sensors. Proceedings of IEEE*, volume 2, pp. 974–979, 2003.

[LFS05] V. Leonov, P. Fiorini, S. Sedky, T. Torfs, and C. Van Hoof. "Thermoelectric MEMS generators as a power supply for a body area network." In *Solid-State Sensors, Actuators and Microsystems, International Conference on*, pp. 291–294, 2005.

[LLN12] S.I. Lee, C. Ling, A. Nahapetian, and M. Sarrafzadeh. "A mechanism for data quality estimation of on-body cardiac sensor networks." In *Consumer Communications and Networking Conference (CCNC), 2012 IEEE*, pp. 194–198, 2012.

[LLR10] Ming Li, Wenjing Lou, and Kui Ren. "Data security and privacy in wireless body area networks." *Wireless Communications, IEEE*, **17**(1):51–58, 2010.

[LR03] Q. Li and J. Racine. "Nonparametric estimation of distributions with categorical and continuous data." *journal of multivariate analysis*, **86**(2):266–292, 2003.

[LYG13] Ming Li, Shucheng Yu, Joshua D. Guttman, Wenjing Lou, and Kui Ren. "Secure Ad Hoc Trust Initialization and Key Management in Wireless Body Area Networks." *ACM Trans. Sen. Netw.*, **9**(2):18:1–18:35, April 2013.

[LYZ13] Ming Li, Shucheng Yu, Yao Zheng, Kui Ren, and Wenjing Lou. "Scalable and Secure Sharing of Personal Health Records in Cloud Computing Using Attribute-Based Encryption." *Parallel and Distributed Systems, IEEE Transactions on*, **24**(1):131–143, 2013.

[LZG12] Jong Hyun Lim, Andong Zhan, Evan Goldschmidt, JeongGil Ko, Marcus Chang, and Andreas Terzis. "HealthOS: A Platform for Pervasive Health Applications." In *Proceedings of the Second ACM Workshop on Mobile Systems, Applications, and Services for HealthCare*, pp. 4:1–4:6, 2012.

[MEB09] Z. Merhi, M. Elgamel, and M.A. Bayoumi. "A Lightweight Collaborative Fault Tolerant Target Localization System for Wireless Sensor Networks." *Mobile Computing, IEEE Transactions on*, **8**(12):1690–1704, 2009.

[MKA11] H. Mamaghanian, N. Khaled, D. Atienza, and P. Vandergheynst. "Compressed Sensing for Real-Time Energy-Efficient ECG Compression on Wireless Body Sensor Nodes." *IEEE Transactions on Biomedical Engineering*, **58**(9):2456–2466, 2011.

[MKP05] S. Megerian, F Koushanfar, M. Potkonjak, and M.B. Srivastava. "Worst and best-case coverage in sensor networks." *Mobile Computing, IEEE Transactions on*, **4**(1):84–92, 2005.

[MOC10] Thomas McKay, Benjamin OBrien, Emilio Calius, and Iain Anderson. "Self-priming dielectric elastomer generators." *Smart Materials and Structures*, **19**(5), 2010.

[MYR08] P.D. Mitcheson, E.M. Yeatman, G.K. Rao, A.S. Holmes, and T.C. Green. "Energy Harvesting From Human and Machine Motion for Wireless Electronic Devices." *Proceedings of the IEEE*, **96**(9):1457 –1486, sept. 2008.

[NAH10] Hyduke Noshadi, Shaun Ahmadian, Hagop Hagopian, Jonathan Woodbridge, Foad Dabiri, Navid Amini, Majid Sarrafzadeh, and Nick Terrafranca. "HERMES - Mobile Balance and Instability Assessment System." In *Proceedings of BIOSIGNALS*, pp. 264–270, 2010.

[NBG11] Sidharth Nabar, Ayan Banerjee, Sandeep K. S. Gupta, and Radha Poovendran. "Resource-efficient and reliable long term wireless monitoring of the photoplethysmographic signal." In *Proceedings of the Conference on Wireless Health*, pp. 1–10, 2011.

[NDM10] Hyduke Noshadi, Foad Dabiri, Saro Meguerdichian, Miodrag Potkonjak, and Majid Sarrafzadeh. "Energy optimization in wireless medical systems using physiological behavior." In *Wireless Health 2010*, WH '10, pp. 128–136. ACM, 2010.

[Ogd72] R. W. Ogden. "Large Deformation Isotropic Elasticity - On the Correlation of Theory and Experiment for Incompressible Rubberlike Solids." *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, **326**(1567):565–584, 1972.

[OHB02] G.K. Ottman, H.F. Hofmann, A.C. Bhatt, and G.A. Lesieutre. "Adaptive piezoelectric energy harvesting circuit for wireless remote power supply." *IEEE Transactions on Power Electronics*, **17**(5):669–676, 2002.

[Ped07] "Novel.de, Pedar. http://www.novel.de/.", 2007.

[PHC04] Joseph Polastre, Jason Hill, and David Culler. "Versatile low power media access for wireless sensor networks." In *Proceedings of the International conference on Embedded networked sensor systems*, pp. 95–107, 2004.

[PJB12] Pravin Pawar, Val Jones, Bert-Jan F. van Beijnum, and Hermie Hermens. "A framework for the comparison of mobile patient monitoring systems." *Journal of Biomedical Informatics*, **45**(3):544–556, 2012.

[PKM04] I.P.I. Pappas, T. Keller, S. Mangold, M.R. Popovic, V. Dietz, and M. Morari. "A reliable gyroscope-based gait-phase detection sensor embedded in a shoe insole." *Sensors Journal, IEEE*, **4**(2):268–274, 2004.

[PPM13] A. Prasad, R. Peterson, S. Mare, J. Sorber, K. Paul, and D. Kotz. "Provenance framework for mHealth." In *Communication Systems and Networks (COMSNETS), 2013 Fifth International Conference on*, pp. 1–6, 2013.

[PRH11] K. Plarre, A. Raij, S.M. Hossain, A.A. Ali, M. Nakajima, M. Al'absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, Daniel Siewiorek, A. Smailagic, and L.E. Wittmers. "Continuous inference of psychological stress from sensory measurements collected in the natural environment." In *Information Processing in Sensor Networks, International Conference on*, pp. 97–108, 2011.

[PS05] Joseph A. Paradiso and Thad Starner. "Energy Scavenging for Mobile and Wireless Electronics." *IEEE Pervasive Computing*, **4**(1):18–27, January 2005.

[RVR03] R. Rao, S. Vrudhula, and D.N. Rakhmatov. "Battery modeling for energy aware system design." *Computer*, **36**(12):77–87, 2003.

[SBD01] T. Simunic, L. Benini, and G. De Micheli. "Energy-efficient design of battery-powered embedded systems." *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, **9**(1):15–28, 2001.

[SGG10] Abhishek B. Sharma, Leana Golubchik, and Ramesh Govindan. "Sensor faults: Detection methods and prevalence in real-world datasets." *ACM Trans. Sen. Netw.*, **6**(3):1–39, June 2010.

[SGY11] Lu Su, Jing Gao, Yong Yang, Tarek F. Abdelzaher, Bolin Ding, and Jiawei Han. "Hierarchical aggregate classification with limited supervision for data reduction in wireless sensor networks." In *Proceedings of the International Conference on Embedded Networked Sensor Systems*, pp. 40–53, 2011.

[SLM13] O. Salem, Yaning Liu, and A. Mehaoua. "A lightweight anomaly detection framework for medical wireless sensor networks." In *Wireless Communications and Networking Conference, IEEE*, pp. 4358–4363, 2013.

[SP01] Nathan S. Shenck and Joseph A. Paradiso. "Energy Scavenging with Shoe-Mounted Piezoelectrics." *IEEE Micro*, **21**(3):30–42, May 2001.

[vci] "Optimal Performance Index System.".

[VPB96] J M VanSwearingen, K A Paschal, P Bonino, and J F Yang. "The modified Gait Abnormality Rating Scale for recognizing the risk of recurrent falls in community-dwelling elderly adults." *Physical Therapy*, **76**(9):994–1002, 1996.

[WAJ08] Winston Wu, Lawrence Au, Brett Jordan, Thanos Stathopoulos, Maxim Batalin, William Kaiser, Alireza Vahdatpour, Majid Sarrafzadeh, Meika Fang, and Joshua Chodosh. "The SmartCane system: an assistive device for geriatrics." In *Proceedings of the ICST 3rd international conference on Body area networks*, BodyNets '08, pp. 2:1–2:4. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008.

[WM99] A.K. Wendy and M.E. Moffa-Trotter. "Functional tools for assessing balance and gait impairments." *Topics in Geriatric Rehabilitation*, **15**(1):66–83, 1999.

[WMN12] James Wendt, Saro Meguerdichian, Hyduke Noshadi, and Miodrag Potkonjak. "Semantics-driven sensor configuration for energy reduction in medical sensor networks." In *Proceedings of ACM/IEEE international symposium on Low power electronics and design*, pp. 303–308, 2012.

[WP11] James B Wendt and Miodrag Potkonjak. "Medical diagnostic-based sensor selection." In *Sensors, 2011 IEEE*, pp. 1507 –1510, oct. 2011.

[WR81] Min-Chiang Wang and John van Ryzin. "A Class of Smooth Estimators for Discrete Distributions." *Biometrika*, **68**(1):pp. 301–309, 1981.

[YZJ07] Le Yan, Lin Zhong, and Niraj K. Jha. "Energy comparison and optimization of wireless body-area network technologies." In *Proceedings of the ICST 2nd international conference on Body area networks*, BodyNets '07, pp. 8:1–8:8. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007.

[ZMH10] Yang Zhang, N. Meratnia, and P. Havinga. "Outlier Detection Techniques for Wireless Sensor Networks: A Survey." *Communications Surveys Tutorials, IEEE*, **12**(2):159–170, 2010.

[ZRF12] Piero Zappi, Daniel Roggen, Elisabetta Farella, Gerhard Tröster, and Luca Benini. "Network-Level Power-Performance Trade-Off in Wearable Activity Recognition: A Dynamic Sensor Selection Approach." *ACM Trans. Embed. Comput. Syst.*, **11**(3):68:1–68:30, September 2012.