Technische Universität München
Lehrstuhl für Theoretische Informationstechnik

Computation of Real-Valued Functions Over the Channel in Wireless Sensor Networks

Mario Goldenbaum

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr.-Ing. habil. Dirk Wollherr
Prüfer der Dissertation:	
	1. Univ.-Prof. Dr.-Ing. Dr. rer. nat. Holger Boche
	2. Priv.-Doz. Dr.-Ing. habil. Sławomir Stańczak, Technische Universität Berlin

Die Dissertation wurde am 11. 06. 2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 28. 10. 2014 angenommen.

Zusammenfassung

In einer Vielzahl von Sensornetzanwendungen steht nicht die Übertragung individueller Messwerte im Vordergrund, sondern die zuverlässige und effiziente Berechnung von Funktionen. Dabei kann es sich im Einzelnen um die Berechnung der maximalen Kohlenmonoxid-Konzentration innerhalb von Gebäuden aus Gründen der Brandschutzüberwachung handeln, um den mittleren Druck in einem Dampfkessel oder die minimale Feuchtigkeit in einem Gewächshaus. Der klassische Ansatz zur Lösung dieser Probleme sieht vor, dass zunächst die gesamte Messwertinformation zu einem Aggregationspunkt übertragen wird, der anschließend den gewünschten Funktionswert ermittelt. Um Kanalkollisionen während der Übertragung der Daten zu vermeiden, wird der Medienzugriff von Sensorknoten üblicherweise in der Zeit oder der Frequenz koordiniert.

In der jüngeren Vergangenheit wurde gezeigt, dass diese Vorgehensweise in höchstem Maße ineffizient sein kann, da sie vollständig ignoriert, dass die primäre Aufgabe des Netzes eben nicht in der Übertragung einzelner Messwerte besteht. Insbesondere gilt, dass für die Berechnung linearer Funktionen die Ausnutzung der Interferenz anstelle deren Unterdrückung sogar zu erheblichen Performanzgewinnen führen kann. Die obigen Beispiele deuten jedoch bereits darauf hin, dass einige der vielversprechendsten Sensornetzanwendungen die effiziente Berechnung *nichtlinearer Funktionen* erfordern.

Dementsprechend widmet sich die vorliegende Arbeit linearen und nichtlinearen Berechnungsproblemen in Sensornetzen. Der erste Teil behandelt zunächst einige grundlegende Fragestellungen, wie etwa der welche Funktionen prinzipiell durch das Ausnutzen von Interferenz berechnet werden können und wie viele Kanalbenutzungen dafür notwendig sind. Um diesbezüglich eine systematische Untersuchung zu ermöglichen werden Funkübertragungen als rauschfrei angenommen. Dabei stellt sich heraus, dass diese Überlegungen in enger Beziehung zu dem berühmten 13ten Hilbert-Problem stehen.

Der zweite Teil der Arbeit schließt Empfängerrauschen in die Betrachtungen mit ein und stellt diesbezüglich ein Übertragungsverfahren vor, das eine neuartige Datenvor- und nachverarbeitung mit nested lattice codes kombiniert. Es wird gezeigt, dass diese Kombination die zuverlässige Berechnung einer Vielzahl linearer und nichtlinearer Funktionen erlaubt mit Raten die für klassische Verfahren weitgehend unerreichbar sind.

Das gezielte Ausnutzen von Interferenz erfordert naturgemäß eine präzise Synchronisation. Dies zu gewährleisten kann in der Praxis einen unverhältnismäßig hohen Aufwand bedeuten, weshalb im letzten Teil der Arbeit ein einfaches *analoges* Übertragungsverfahren vorgeschlagen wird, das bereits mit einer groben Rahmensynchronisation auskommt. Die Performanz des Verfahrens wird analysiert und es wird erörtert wie viel Kanalinformation für akkurate Funktionswertberechnungen notwendig ist.

Abstract

A major challenge in many wireless sensor network applications consists in the reliable and efficient computation of some pre-defined function of the measurements taken by a set of spatially distributed sensor nodes. The function of interest can be, for instance, the maximum carbon-monoxide concentration in a building for fire detection, the average pressure inside a steam boiler, or the minimum humidity in a greenhouse. The standard approach for solving such problems, the so-called separation-based approach, is to transmit all the sensor readings to a nearby fusion center that computes the desired function value afterwards. As a matter of fact, however, the wireless channel is a shared broadcast medium so that a concurrent access to the common frequency spectrum by distinct nodes results in interference. Since interference makes the reliable reconstruction of individual transmit signals difficult, the channel access of sensor nodes is typically coordinated in time or frequency in order to avoid channel collisions.

It was recently shown that this approach can be highly inefficient as it ignores that the fusion center is not interested in individual sensor readings but rather in a function thereof. In particular, if the fusion center wishes to compute a linear function, exploiting interference rather than avoiding it can lead to huge performance gains. As the above mentioned examples indicate, however, many promising sensor network applications require the efficient computation of *nonlinear functions*.

Therefore, this thesis is devoted to linear and nonlinear computation problems over sensor networks. The first part deals with some fundamental questions such as, for instance, which functions are essentially computable over a wireless channel by harnessing interference and how many channel uses are needed. In order to allow for a systematic and in-depth treatment of these issues, we assume transmissions between sensor nodes and fusion centers to be noise-free. It turns out that the considerations are closely related to the famous 13th Hilbert problem.

In the second part of the thesis, we turn our attention to noisy networks and propose a corresponding computation scheme that combines a novel signal pre- and post-processing strategy with nested lattice coding. We show that this particular combination allows for the computation of a variety of linear and nonlinear functions at computation rates that are not achievable with separation-based methods.

Harnessing interference typically requires the precise synchronization of sensor nodes. In practical networks, however, it may be unreasonably costly to ensure this. Therefore, we propose in the last part a simple *analog* computation scheme that requires only coarse frame synchronization. We then analyze its performance and determine how much channel state information is needed in order to obtain accurate function values.

Acknowledgments

> *Aber Motivation hat mit Wollen keine Berührung; sie läßt sich nicht nach dem Gegensatz von Zwang und Freiheit einteilen, sie ist tiefster Zwang und höchste Freiheit.*
>
> Robert Musil, *Der Mann ohne Eigenschaften (Aus dem Nachlass)*, 1930–1943

Throughout the course of my Ph.D. studies, I had the opportunity to meet many extraordinary people. First and foremost, I thank my advisors Prof. Dr.-Ing. Dr. rer. nat. Holger Boche and PD Dr.-Ing. habil. Sławomir Stańczak for their guidance and invaluable support. Both are fascinating individuals and their energy, passion for research, vast knowledge, and enthusiasm were constant sources of inspiration and motivation, which will have a lasting effect on me.

Furthermore, I would like to thank all my colleagues at Technische Universität Berlin, Fraunhofer Institute for Telecommunications Heinrich Hertz Institute, and Technische Universität München for providing a friendly and stimulating research environment. Especially, I thank Jörg Bühler (a.k.a. Fokker), Michał Kaliszan, and Rafael F. Schaefer for the countless conversations about science and life in general. It was always a great pleasure and they have become more to me than just colleagues.

A special thanks goes to Prof. Dr.-Ing. Dieter Kraus from the Bremen University of Applied Sciences, who is mainly responsible for my path to academia as he piqued my interest in communication theory, signal theory, and mathematics already at a very early stage of my undergraduate studies.

Without those who are closest to me, I would not be where I am today. Therefore, I would like to express my sincere gratitude to my family and friends for their steady encouragement and unquestioning support. My deepest gratitude, however, goes to Sabine. This thesis would have never been possible without her endless patience and love.

Contents

List of Figures

Nomenclature

Throughout the thesis, we use lowercase italic letters to denote scalars, bold lowercase italic letters to denote vectors, and bold uppercase italic letters to denote matrices, respectively (e.g., x, $\boldsymbol{x}$, $\boldsymbol{X}$). Vectors that are involved in products with matrices or other vectors have to be always considered as column vectors, even if we write them often as row vectors (e.g., $\boldsymbol{x} = (x_1, \ldots, x_n)$, $\boldsymbol{y} = (y_1, \ldots, y_n)$, but $\boldsymbol{x}^\mathsf{T}\boldsymbol{y}$ is a scalar).

Scalar random variables are described by uppercase italic letters and their realizations by lowercase italic letters (e.g., X and x). On the other hand, vector-valued random variables are denoted as bold lowercase *sans-serif* letters and their realizations as ordinary vectors (e.g., $\mathbf{x}$ and $\boldsymbol{x}$), respectively.

The n-fold Cartesian product $\mathbb{A} \times \cdots \times \mathbb{A}$ of some set or space $\mathbb{A}$ is compactly written as $\mathbb{A}^n$ (i.e., the set of all n-tuples $(a_1, \ldots, a_n)$ with $a_i \in \mathbb{A}$).

When dealing with random elements, we assume that they are defined over some meaningful probability space $(\Omega, \mathcal{F}, \mathbb{P})$ without explicitly mentioning. Here, Ω denotes the sample space, $\mathcal{F}$ the σ-algebra of subsets of Ω, and $\mathbb{P} : \mathcal{F} \to [0, 1]$ a corresponding probability measure. For some positive integer n, we are mainly interested in the case $\Omega = \mathbb{R}^n$ along with $\mathcal{F} = \mathcal{B}^n$, where $\mathcal{B}^n$ represents the Borel σ-algebra of subsets of $\mathbb{R}^n$ [Shi96, p. 144].

It is common practice to mark the end of a proof by the symbol $\square$. As some of the examples provided in this thesis are very detailed, we mark their end in a similar manner by the symbol $\triangle$ in order to improve clarity.

General Notations

$:=$	equal by definition
$\equiv$	identically equal to (e.g., $f \equiv 1$ on $\mathbb{X} \Leftrightarrow f(x) = 1$ for all $x \in \mathbb{X}$)
$\exists$	there exists
$\forall$	for all
$\oplus_p$	addition modulo p
$\bigoplus$	summation modulo p
$g \circ f$	composition of functions f and g
$\operatorname{grad}(\cdot)$	gradient of a multivariate function (i.e., the vector of partial derivatives)
$(\cdot)^\mathrm{c}$	complement of a set
$\lvert \cdot \rvert$	the absolute value (modulus) of a real or complex number or the cardinality of a set
$\lfloor \cdot \rfloor$	floor function that maps a real number to the largest previous integer

$\lceil \cdot \rceil$	ceiling function that maps a real number to the smallest following integer
$(\cdot)!$	factorial of a natural number (i.e., $3! = 1 \cdot 2 \cdot 3$)
e	Eulers number (i.e., $\lim_{n\to\infty}(1+1/n)^n$)
i	imaginary unit (i.e., $\mathrm{i}^2 = -1$)
$\mathrm{Re}\{\cdot\}$	real part of a complex number
$\mathrm{Im}\{\cdot\}$	imaginary part of a complex number
$(\cdot)^*$	conjugate of a complex number
$(\cdot)^\mathsf{T}$	transpose of a vector or matrix
$(\cdot)^\mathsf{H}$	Hermitian transpose of a vector or matrix
$(\cdot)^{-1}$	inverse of a matrix or function
$(\cdot)^\dagger$	left inverse of a matrix
$\det(\cdot)$	determinant of a matrix
$\|\cdot\|_2$	Euclidean norm
$\|\cdot\|_\infty$	supremum norm
$\boldsymbol{I}_n$	$n \times n$ identity matrix
$\mathbf{0}$	vector of all zeros
$\mathbf{1}$	vector of all ones
$\mathrm{Vol}(\cdot)$	volume of a closed subset of an Euclidean space
$\mathbb{E}\{\cdot\}$	expected value
$\mathbb{E}_X\{\cdot\}$	expectation with respect to random variable X
$\mathbb{V}\mathrm{ar}\{\cdot\}$	variance
$\mathbb{C}\mathrm{ov}\{\cdot,\cdot\}$	covariance
$\mathbb{P}(\cdot)$	probability
$\xrightarrow{\mathrm{d}}$	convergence in distribution
$\xrightarrow{\mathbb{P}}$	convergence in probability
$\xrightarrow{\mathrm{a.s.}}$	almost sure convergence (or convergence with probability one)
$\mathcal{N}_\mathbb{R}(\cdot,\cdot)$	normal (or Gaussian) distribution
$\mathcal{N}_\mathbb{C}(\cdot,\cdot)$	proper complex normal (or Gaussian) distribution
$\mathcal{LN}(\cdot,\cdot)$	log-normal distribution
χ^2_n	Chi-squared distribution with n degrees of freedom

Sets and Function Spaces

$\mathbb{C}$	complex plane
$\mathbb{E}$	closed unit interval $[0,1]$
$\mathbb{N}$	natural numbers (i.e., $\{1,2,3,\dots\}$)
$\mathbb{R}$	real numbers
$\mathbb{R}_+$	nonnegative real numbers (i.e., $\{x \in \mathbb{R} \,\vert\, x \geq 0\}$)
$\mathbb{R}_{++}$	positive real numbers (i.e., $\{x \in \mathbb{R} \,\vert\, x > 0\}$)
$\mathbb{R}^n$	Euclidean n-space (i.e., the set of all ordered n-tuples $(x_1,\dots,x_n)$ with $x_i \in \mathbb{R}$)
$\mathbb{Z}$	integers (i.e., $\{\dots,-2,-1,0,1,2,\dots\}$)

$\mathbb{Z}_+$	nonnegative integers (i.e., $\{0, 1, 2, \dots\}$)
$\mathbb{Z}_p$	integers modulo p (i.e., $\{0, \dots, p-1\}$)
$\mathscr{C}^0(\mathbb{A})$	space of continuous functions $f : \mathbb{A} \to \mathbb{R}$
$\mathscr{F}(\mathbb{A})$	space of all functions $f : \mathbb{A} \to \mathbb{R}$
$\mathscr{L}^p(\mathbb{A})$	space of measurable functions $f : \mathbb{A} \to \mathbb{R}$ that are Lebesgue integrable to the p^{th} power, $1 \leq p < \infty$
$\mathscr{N}(\mathbb{A})$	space of nomographic functions $f : \mathbb{A} \to \mathbb{R}$
$\mathscr{N}^0(\mathbb{A})$	space of nomographic functions with continuous pre- and post-processing functions
$\mathscr{N}^0_{d,\varepsilon}(\mathbb{A})$	space of functions approximable with elements from $\mathscr{N}^0(\mathbb{A})$
$\mathcal{O}(f)$	Big-O (i.e., $\mathcal{O}(f) \coloneqq \{g \mid \exists C > 0\, \exists n_0\, \forall n > n_0 : g(n) \leq C f(n)\}$)

Special Functions

$\exp_a$	exponential function to base a
$\log_a$	logarithm to base a
erf	error function (i.e., $\operatorname{erf} : \mathbb{R} \to [-1, 1]$, $x \mapsto \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) \mathrm{d}t$)
erfc	error function complement (i.e., $\operatorname{erfc} : \mathbb{R} \to [0, 2]$, $x \mapsto \operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$)
$\mathrm{id}_{\mathbb{A}}$	identity map on set $\mathbb{A}$ (i.e., $\mathrm{id}_{\mathbb{A}} : \mathbb{A} \to \mathbb{A}$, $\mathrm{id}_{\mathbb{A}}(x) = x$, for all $x \in \mathbb{A}$)
$\mathbb{1}_{\mathbb{A}}$	indicator function on set $\mathbb{A}$ (i.e., $\mathbb{1}_{\mathbb{A}} : \mathbb{A} \to \{0, 1\}$, $\mathbb{1}_{\mathbb{A}}(x) = 1$ if $x \in \mathbb{A}$ and 0 otherwise)

Abbreviations

AWGN	additive white Gaussian noise
CDMA	code-division multiple access
CoMAC	computation over multiple-access channels
CSI	channel state information
FC	fusion center
FDMA	frequency-division multiple access
iid	independent and identically distributed
IoT	internet of things
MAC	multiple access channel
MSE	mean squared error
SNR	signal-to-noise ratio
TDMA	time-division multiple access
WBE	Welch bound equality
WMAC	wireless multiple-access channel
WSN	wireless sensor network

1

Introduction

1.1 Motivation

Over the last two decades, research and development activities in the realm of communications engineering were mainly driven by the demand for high speed wireless data access anywhere and anytime. Starting in the early 90's with data rates of not more than 10 kbit/s, we recently arrived, with the advent of the 4th cellular generation, at rates of up to 100 Mbit/s [FA14]. The lion's share of data to be transmitted is currently governed by human-generated high resolution image, audio, and video content. With the ongoing advances in microelectronics, however, wireless communication technologies increasingly find their way into other areas such as, for instance, industrial automation [Wil08, ÅGB11]. This trend not only dramatically changes the way we interact with our physical environment but also the type and amount of data to be transmitted. Accordingly, the 5th cellular generation is predicted to provide an Internet of Things (IoT) that interconnects up to 1 trillion products, machines, and devices by 2022 [Pre14]. In order to put this into practice, a plenty of smaller networks supporting various promising concepts (e.g., smart grids, smart factories, smart agriculture, eHealth) need to be joined seamlessly together. As a consequence, *wireless sensor networks* are very likely to be an integral part of the IoT.

A wireless sensor network is typically referred to as a set of low-cost devices with on-board sensors, wireless transceiver, microcontroller, memory, and power supply that are distributed over some geographical area [ZG04, SZHT07]. The devices, called sensor nodes, gather information from their environment (e.g., temperatures, pressures, humidities) and forward them to other nearby sensor nodes or a central fusion center for further processing. Unlike conventional wireless networks in which the objective is to provide high-capacity end-to-end connections for message transfer, the most exciting applications for wireless sensor networks are satisfied with low data rates but require that sensor nodes are powered by batteries. Thus, networking protocols for low-

complexity and low-energy consuming wireless connectivity are crucial for ensuring a long network lifetime. In order to enhance the network efficiency not only in terms of energy consumption but also in terms of reliability and sensing quality, it is vital to tailor wireless sensor network solutions to specific application needs [ASSC02, ZG04, SZHT07]. Recently, this paradigm shift has also been drawing more and more attention when designing next generation wireless data networks [SB14].

A vast number of wireless sensor network applications require an efficient and reliable computation of pre-defined *functions* of the sensor readings. Functions of interest can be, for instance, the maximum carbon-monoxide concentration (or temperature) in a building for fire detection, the average pressure inside a steam boiler, or the minimum humidity in a greenhouse. In this context, Giridhar and Kumar in [GK05] (see also [GK06] for an overview) take first steps towards a theory-based framework for solving such distributed computation problems with the aim of characterizing efficient application-specific computation strategies. The work is focused on complexity and protocol aspects and does not explicitly take into account the properties of wireless communication channels. A similar limitation is true for the considerations in [YSD07, SGS07, KKK08, LD13, CWL14], which are mainly concerned with determining how the computation performance or energy consumption scales with the network size, the connectivity radii of nodes, and the cardinality of the sensing range (i.e., the set from which the sensor readings are drawn). The work in [KK09, AFKZ11] provides corresponding scaling laws for wired networks whereas computation problems from a source coding perspective are addressed in [Yam82, OR01, DSME10, HS12].

Function computation in sensor networks is also envisioned as a fundamental building block for gossip and consensus algorithms, a form of distributed in-network data processing aiming at achieving some network-wide objectives based on local computations. Such algorithms, which compute a global function of sensor readings and distribute the function values among the nodes, have also attracted a great deal of attention (see [MS06, BGPS06, OSFM07, DKM+10] and references therein).[1]

What is common to all of the above referenced works is that they follow, in one way or another, a separation-based approach. Here, sensor nodes first transmit their observations to a nearby fusion center, or some other intermediate node, that computes the desired function-value *afterwards*. As a matter of fact, however, the wireless channel is a shared broadcast medium so that a concurrent access to the common frequency spectrum by distinct nodes results in interference. Since interference makes the reliable reconstruction of individual transmit signals difficult, the channel access of sensor nodes is typically coordinated in time or frequency in order to avoid channel collisions. It is shown by Nazer and Gastpar in [NG07] that this approach can be highly inefficient as it ignores that the fusion center is not interested in individual sensor readings but rather in a function thereof. In particular, if the fusion center wishes to compute a linear

[1] Of course, distributed computation problems have a long history in computer science (see [BT89] and references therein.). Here, however, we restrict ourselves to references that are primarily concerned with computation problems in wireless sensor networks that have to face strict communication constraints.

function, *exploiting interference* rather than avoiding it can lead to huge performance gains as the function of interest matches the algebraic structure of the wireless channel (see Chapter 3 for further references).

Thus, separation-based approaches treat the processes of *communication* and *computation* as completely distinct processes so that the underlying function computation (i.e., the application of interest) is not adequately taken into account in the design of communication strategies/protocols. As a consequence, wireless resources are wasted since the fusion centers are not interested in individual measurements but only in functions of them.

1.2 Contribution and Outline of the Thesis

This dissertation deals with various aspects of linear and nonlinear computation problems in wireless sensor networks, ranging from providing fundamental limits to incorporating practical constraints. Such as in [NG07], the leitmotif is to consider the interference generated by sensor nodes that concurrently access the same frequency spectrum not as a hindrance to be overcome but as a freely available computational resource. With this in mind, in **Chapter 2** we first introduce the wireless multiple-access channel as a reasonable model for concurrently transmitting nodes. Then, we provide a formal description of what is meant throughout the thesis when referring to the communication or the computation problem over a wireless multiple-access channel. Based on two simple but insightful examples, we then demonstrate that the problems can be substantially different from each other so that they should also be treated differently.

Motivated by this observation, **Chapter 3** contributes to determine the fundamental limits of real-valued function computation in sensor networks. In order to allow a corresponding in-depth and rigorous analysis, we consider an idealized wireless channel model in the sense that transmissions between sensor nodes and fusion centers are assumed to be noise-free. Based on this, in Section 3.1 we determine which functions are in principle computable by harnessing interference in networks with a single fusion center and how efficiently this can be done in terms of the number of channel uses/wireless resources. It turns out that when each node applies an appropriate pre-processing function to its sensor readings and the fusion center an appropriate post-processing function to the observed channel output, in fact *every* real-valued function can be computed with a single channel use.

The achievable efficiency, however, strongly depends on the properties that are imposed on the pre- and post-processing functions. From an implementation point of view, for instance, continuity can be a highly desirable property. We show that this property in general requires additional channel uses/wireless resources for computations over the channel. This is a consequence of a result proven by Kolmogorov in [Kol57] that solves the famous *13th Hilbert problem* stated in 1900 [Hil02]. In contrast to efficiency, it turns out that computations over the channel can always be universally performed, regardless of whether the pre-processing functions are continuous or not. Throughout the thesis,

universality refers to transmit strategies that are independent of the function to be computed so that there is no need to inform the sensor nodes once the function to be computed has changed.

Based on these results for networks with a single fusion center, in Section 3.2 we consider a generalized sensor network model consisting of multiple fusion centers, each of which aims to independently compute some function of the sensor readings of an arbitrary subset of nodes (i.e., clusters). It turns out that all the previous properties carry over to the generalized case except if pre- and post-processing functions are required to be continuous, in which case some coordination may be necessary. If, however, the fusion centers perform a simple additional post-processing step, it is shown that coordination is not required at all. In Section 3.3, we also show that the universality property is preserved against changes in topology as a result of nodes dropping out of the network (for instance due to failures or battery depletion) or new nodes joining the network. *Parts of the material in Chapter 3 are published in [7, 9, 10, 13, 16].*

In **Chapter 4**, we extend the considerations of Chapter 3 to clustered networks in which the intra-cluster communication takes place over Gaussian multiple-access channels. One of the basic facts in information theory teaches us that Gaussian channels are of finite capacity if transmit powers and bandwidths are assumed to be finite. In order to account for this, in Section 4.2 we propose an achievable computation scheme that combines the pre- and post-processing strategy of Chapter 3 with a simple quantizer and nested lattice coding. Because of their good algebraic and statistical properties, nested lattice codes are well suited for protecting sums of codewords against the channel noise, which constitutes a basic requirement for harnessing interference.

Section 4.3 is devoted to analyzing the performance of the scheme in terms of achievable computation rates, where the computation rate is defined as the number of function values that can be reliably computed per channel use. It turns out that the particular combination of analog data pre- and post-processing with nested lattice codes allows for the computation of numerous continuous real-valued functions at computation rates that are not achievable with separation-based methods. In addition to the better rate performance, the proposed scheme provides several other advantages that are essential for most wireless sensor network applications such as universality, lower decoding complexity, less coordination, and the ability to deal with maximum decoding error probabilities. *Parts of the material in Chapter 4 are published in [14, 20].*

Harnessing interference in the way of Chapter 4 typically requires the precise synchronization of sensor nodes on the symbol and phase level. In practical wireless sensor networks, however, it may be unreasonably difficult and expensive in terms of resources to ensure this. Therefore, in the first section of **Chapter 5** we propose a novel *analog* computation scheme that requires only coarse frame synchronization, which is by far easier to establish and maintain. The basic idea of the scheme consists in letting each sensor node transmit a complex-valued sequence of random symbols at a *transmit power* that is proportional to the instantaneous pre-processed sensor reading. Under some conditions and a suitable pre-processing strategy, the received energy at the fusion center equals the sum of all the transmit energies corrupted by background noise.

The application of an appropriately chosen post-processing strategy then results in an immediate estimate of the sought function value.

In order to examine how an appropriate post-processing strategy has to look like, in Section 5.2 we analyze the statistical properties of the proposed computation scheme for two canonical function examples of great practical relevance: the arithmetic mean and the geometric mean. In Section 5.3, we show the huge potential for performance gains at a wide range of operating points by comparing the scheme with two standard separation-based strategies. *Parts of the material in Sections 5.1–5.3 are published in [1, 3, 15].*

For the analog scheme in Section 5.1, it was assumed that perfect channel state information (CSI) is available at sensor nodes prior to transmissions. Because this is, similar to synchronization, difficult to provide in many wireless sensor applications, in Section 5.4 we explore the question of how much CSI is actually needed to obtain reliable function-values at the fusion center. In particular, we show that the knowledge of the channel magnitudes is sufficient to achieve the same performance as with perfect CSI. Moreover, we show that under certain conditions, CSI at sensor nodes is not needed at all provided that the fusion center has some statistical a priori knowledge. If in addition the fusion center is equipped with multiple antennas, spatial diversity can be exploited to outperform the single-antenna scheme with perfect CSI at nodes. Our findings suggest that the channel estimation effort for computational purposes can be significantly reduced. *Parts of the material in Section 5.4 are published in [2, 18].*

Finally, **Chapter 6** concludes the thesis and provides suggestions for future research directions whereas **Appendices A** and **B** recap some useful definitions and results of multivariate calculus and probability theory.

Further Results that are not Part of the Thesis

During my time as a Ph.D. student and research assistant at Technische Universität Berlin, Fraunhofer Institute for Telecommunications Heinrich Hertz Institute, and Technische Universität München, we were able to obtain a number of further results, which are not part of this thesis:

- In a work with Rudi Abi Akl and Stefan Valentin [4], we study how strongly the downlink in a multiuser orthogonal-frequency division multiplexing system with single-cell scheduling and channel estimation suffers from feedback delay. Unlike previous work, we study this degradation for optimal joint power and rate allocation under fairness constraints. Comparing the performance of the ideal case to delayed channel state information shows that adjusting the scheduler's fairness cannot mitigate the strong performance loss due to feedback delay but simple linear channel prediction is a powerful tool to do so.

- In [8], we consider power-controlled wireless multiantenna sensor networks with interference and study the general trade-off between energy consumption and

quality of service. First, we develop a model for the energy consumption of multiantenna sensor nodes and study the corresponding costs for accurate channel estimation. Then, in order to provide guidelines on the choice of strategies for different applications, we numerically compare different multiantenna strategies with the energy consumption and the achievable quality of service of a standard single-antenna system.

- In a coauthored work with Meng Zheng and Haibin Yu [11], we propose a gossip algorithm for average consensus in clustered wireless sensor networks where the nodes in each cluster exploit the interference property of the wireless channel in order to significantly decrease local averaging times. The convergence of the algorithm is proven provided that some connectivity condition between clusters is fulfilled. This serves as a preliminary step for the novel class of so-called *nomographic gossip algorithms* presented in [12], which partly allow to efficiently achieve a rapid global consensus among nodes/agents with respect to an arbitrary function of the initial states.

A complete list of publications can be found on pages 143–145.

Copyright Information

2

Communication Versus Computation

In this chapter, we first introduce the affine model of a time-discrete wireless multiple-access channel as one of the main ingredients of our considerations. Subsequently, we provide a formal description of what is meant throughout the thesis when referring to the communication or the computation problem over a wireless multiple-access channel in the context of wireless sensor networks. Based on two simple but inspiring examples, we then demonstrate that the problems can be substantially different from each other so that they should also be treated differently. Finally, these insights will lead us to further interesting questions to be addressed in later chapters.

2.1 The Wireless Multiple-Access Channel

The multiple-access channel (MAC) is one of the fundamental building blocks of network information theory as it models the scenario in which multiple users communicate with a common receiver over a common channel [Ahl71, Lia72, CT06, GK11]. As illustrated in Figure 2.1, the general model of a time-discrete MAC with $N \geq 2$ users consists of input alphabets $\mathcal{X}_1, \ldots, \mathcal{X}_N$ (not necessarily discrete and finite), an output alphabet $\mathcal{Y}$, and a conditional probability distribution

$$P_{Y|X_1,\ldots,X_N} : \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \to \mathcal{P}(\mathcal{Y}) \ ,$$

where $\mathcal{P}(\mathcal{Y})$ is used to denote the set of all probability distributions on $\mathcal{Y}$. Here, $P_{Y|X_1,\ldots,X_N}$ assigns to every choice of fixed input symbols $x_i \in \mathcal{X}_i$, $i = 1, \ldots, N$, for which $P_{X_1,\ldots,X_N}(x_1, \ldots, x_N) > 0$ a probability distribution $P_Y(y|X_1 = x_1, \ldots, X_N = x_N) = P_{Y|X_1=x_1,\ldots,X_N=x_N}(y)$ on the output $\mathcal{Y}$ with $P_{X_1,\ldots,X_N} : \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \to [0, 1]$ denoting the joint distribution of the channel inputs. Writing $P_Y(y|X_1 = x_1, \ldots, X_N = x_N)$ shortly as $P_Y(y|x_1, \ldots, x_N)$ and using the channel $n \in \mathbb{N}$ times, a MAC is said to

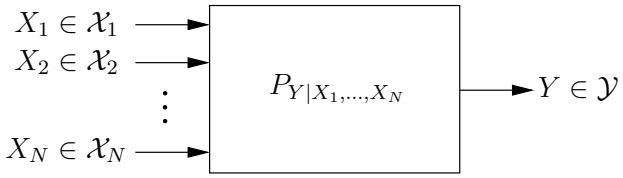


Figure 2.1: General N-user multiple-access channel.

be *memoryless* if the output $Y[t] \in \mathcal{Y}$ at some channel use $t \in \{1, \ldots, n\}$ is distributed according to

$$P(y[t] \,|\, \boldsymbol{x}_1^{(t)}, \ldots, \boldsymbol{x}_N^{(t)}, \boldsymbol{y}^{(t-1)}) = P_Y(y[t] \,|\, x_1[t], \ldots, x_N[t]) \ ,$$

where $\boldsymbol{x}_i^{(t)} \coloneqq (x_i[1], \ldots, x_i[t]) \in \mathcal{X}_i^t$ denotes the sequence of channel inputs of user i and $\boldsymbol{y}^{(t-1)} \coloneqq (y[1], \ldots, y[t-1]) \in \mathcal{Y}^{t-1}$ the sequence of all previous channel outputs. Hence, the memoryless property simply means that the current MAC output is determined by the current inputs only.

From a system-theoretic perspective, each (memoryless) MAC can alternatively be described by means of a *channel operator*

$$W : \mathcal{X}_1 \times \cdots \times \mathcal{X}_N \to \mathcal{Y}$$

that maps each choice of input symbols to a random channel output Y. A corresponding example that will be of particular interest in all that follows is the affine model of a time-discrete wireless multiple-access channel defined as follows.

Definition 2.1 (WMAC)**.** Let $N \geq 2$ and $\mathcal{X}_1 = \mathcal{X}_2 = \cdots = \mathcal{X}_N = \mathcal{Y} = \mathbb{K}$, with $\mathbb{K}$ either $\mathbb{R}$ or $\mathbb{C}$. If $W : \mathbb{K}^N \to \mathbb{K}$ such that

$$Y = W(x_1, \ldots, x_N) = \sum_{i=1}^{N} H_i x_i + Z \ , \tag{2.1}$$

then the corresponding time-discrete channel model is said to be a *wireless multiple-access channel (WMAC) with N users*. Here and hereafter, Z refers to independent additive white Gaussian noise with variance σ_Z^2 (i.e., $Z \sim \mathcal{N}_{\mathbb{K}}(0, \sigma_Z^2)$) and $H_1, \ldots, H_N$ to some random fading coefficients with joint distribution function $P_{H_1,\ldots,H_N} : \mathbb{K}^N \to [0, 1]$ that are independent of the inputs and the noise. For the degenerate case $\sigma_Z^2 = 0$ and $H_1 = \cdots = H_N \equiv 1$, W simplifies to the linear function (see Definition A.3 in Appendix A)

$$y = w(x_1, \ldots, x_N) = \sum_{i=1}^{N} x_i \ , \tag{2.2}$$

which we call the *ideal WMAC with N users*.

Remark 2.1. In the definition, the case $\mathbb{K} = \mathbb{R}$ refers to the WMAC modeled in the real passband and $\mathbb{K} = \mathbb{C}$ to the WMAC in the complex baseband (i.e., the lowpass equivalent [Lap09, PS08]), respectively.

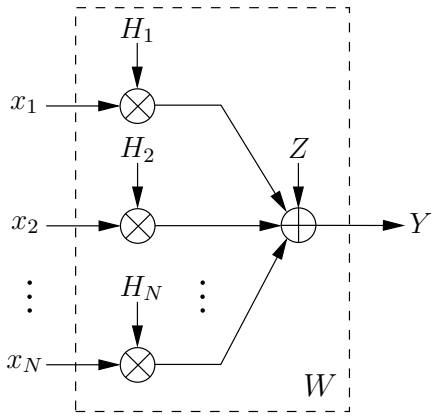


Figure 2.2: Wireless multiple-access channel with N users.

Remark 2.2. For degenerate fading coefficients (i.e., $H_i \equiv h_i$ for some $h_i \in \mathbb{K}$, $i = 1, \ldots, N$), the WMAC is also known as Gaussian MAC [GK11, p. 93]. In particular, for $\mathbb{K} = \mathbb{R}$ and for some fixed choice of input symbols we have

$$P_{Y|X_1=x_1,\ldots,X_N=x_N} = \mathcal{N}_{\mathbb{R}}\left(\sum_{i=1}^{N} h_i x_i, \sigma_Z^2\right) \in \mathcal{P}(\mathcal{Y})$$

and

$$W(x_1, \ldots, x_N) = \sum_{i=1}^{N} h_i x_i + Z \ ,$$

respectively.

The WMAC model depicted in Figure 2.2 is motivated by the fact that the wireless channel is a shared broadcast medium so that a simultaneous access of multiple users to the available frequency spectrum results in interference. In other words, the electromagnetic waves radiated in the same frequency band by a set of distributed transmitters superimpose constructively or destructively at a common nearby receiver. We refer to this in the following as the *interference property* of the wireless channel. A prominent example is the uplink of a 3rd generation cellular system in which the mobile users within a cell may transmit concurrently to the associated base station [Stü11, p. 665].

Definition 2.1 implies that the WMAC is memoryless as it is for every channel use free of intersymbol interference. This is usually the case in flat fading environments in which the coherence bandwidth of the (physical) channel is much larger than the signal bandwidth. Of course, this holds not in general since depending on the chosen transmission system and the respective propagation environment, the delay spread induced by multipath propagation may result in intersymbol interference and therefore in a channel with memory. We do not consider MACs with memory in this thesis even if some of the results could be extended to the frequency selective case.

Sequences of symbols that form the inputs of time-discrete multiple-access channels with infinite alphabets are typically constrained with respect to some nonnegative cost function.

Definition 2.2 (Input Cost Constraint)**.** Let $n \in \mathbb{N}$ be an arbitrary number of channel uses and $\varrho_n : \mathbb{K}^n \to \mathbb{R}_+$ some input cost function that is common to all the WMAC users. Then, for an arbitrary nonnegative real number P_i, the requirement

$$\varrho_n\big(x_i[1], \ldots, x_i[n]\big) \leq P_i \ ,$$

for all sequences of input symbols $(x_i[1], \ldots, x_i[n]) \in \mathbb{K}^n$, is referred to as the *input cost constraint* of user i, $i = 1, \ldots, N$.

The most relevant input cost constraints for wireless network applications are the following transmit power constraints, which are primarily considered in later chapters:

- Average transmit power constraint:

$$\varrho_n\big(x_i[1], \ldots, x_i[n]\big) = \frac{1}{n} \sum_{j=1}^{n} \big|x_i[j]\big|^2 \leq P_i \qquad i = 1, \ldots, N \ , \tag{2.3}$$

- Maximum transmit power constraint:

$$\varrho_n\big(x_i[1], \ldots, x_i[n]\big) = \max_{1 \leq j \leq n} \big|x_i[j]\big|^2 \leq P_i \qquad i = 1, \ldots, N \ . \tag{2.4}$$

Thus, we summarize that the *WMAC with input cost constraints* is, for some fixed $n \in \mathbb{N}$, completely characterized by the triple $(W, \varrho_n, \{P_i\}_{1=1}^N)$ or equivalently by the pair $(W, \{\mathcal{X}_i^{(n)}\}_{i=1}^N)$, in which

$$\mathcal{X}_i^{(n)} \coloneqq \Big\{ (x[1], \ldots, x[n]) \in \mathbb{K}^n \,\big|\, \varrho_n\big(x[1], \ldots, x[n]\big) \leq P_i \Big\} \tag{2.5}$$

denotes the constrained input space of user i, $i = 1, \ldots, N$.

Remark 2.3. Notice that in this thesis, we consider time-discrete WMACs only. This restriction relies on the celebrated Shannon sampling theorem (or rather the Whittaker-Kotel'nikov-Raabe-Shannon sampling theorem [Lük99, Mön11]) that allows to uniquely determine every measurable bandlimited signal of finite energy (i.e., elements from $\mathscr{L}^2(\mathbb{R})$ having a Fourier transform with finite support) by appropriately chosen samples.

2.2 The Communication Problem

The classical *communication problem* has a long standing history and its first systematic treatment for point-to-point channels goes back to Shannon and his groundbreaking work "A Mathematical Theory of Communication" [Sha48]. Since then, Shannon's ideas

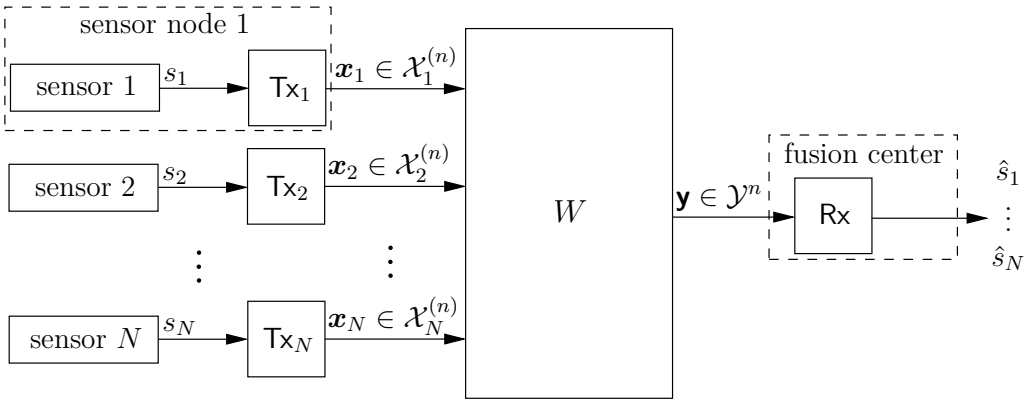


Figure 2.3: The communication problem: A fusion center wishes to reliably and efficiently reconstruct the sensor readings communicated over a WMAC by a set of distributed sensor nodes.

have been extended in various directions and in particular in the early 70s by Ahlswede, Liao, Slepian and Wolf to the problem of communicating independent and correlated sources over MACs [Ahl71, Lia72, SW73]. Instead of a general information-theoretic definition, however, we give in the following a formal description of the communication problem over a WMAC in the context of wireless sensor networks.

Towards this end, consider a wireless sensor network (WSN) consisting of a designated fusion center (FC) and $N \geq 2$ spatially distributed nodes that monitor the environment, which results in sensor readings $s_i \in \mathcal{S}$, $i = 1, \ldots, N$. Here and hereafter, $\mathcal{S} \coloneqq [s_{\min}, s_{\max}] \subset \mathbb{R}$ is used to denote some nonempty compact *sensing range* (e.g., a closed and bounded interval of temperatures, pressures, or humidities). Assuming the sensing range, also called full scale range, to be compact is justified by the fact that every commercial sensor device has a strictly limited range in which it is able to quantify observations [Fra10].[1]

Now, the corresponding distributed communication problem is as follows: Given a WMAC with fixed input cost constraints, $(W, \varrho_n, \{P_i\}_{i=1}^N)$, then the FC intends to *reliably* and *efficiently* receive $s_1, \ldots, s_N$ from the sensor nodes. To be more precise, consider Figure 2.3 and let

$$\mathsf{Tx}_i : \mathcal{S} \to \mathcal{X}_i^{(n)} \ , \ s_i \mapsto \boldsymbol{x}_i \tag{2.6}$$

be the transmitter of sensor node i, $i = 1, \ldots, N$, which maps each element of $\mathcal{S}$ to a length-n sequence of channel inputs $\boldsymbol{x}_i \coloneqq (x_i[1], \ldots, x_i[n])$. Correspondingly, the

[1] Note that a sensor is typically understood as a device that converts some physical input quantity into a corresponding analog electrical output signal (e.g., voltage, current, charge) [Fra10]. In this thesis, however, a sensor is considered as a "black box" that already provides at its output a *time-discrete* signal. When referring to sensor readings, we do not explicitly distinguish between the physical input values and the associated electrical output values. Finally, a "sensor node" is for simplicity assumed to be a sensor along with a wireless transmitter (see Figure 2.3).

mapping

$$\mathsf{Rx} : \mathcal{Y}^n \to \mathcal{S}^N \ , \ \mathbf{y} \mapsto (\hat{s}_1, \ldots, \hat{s}_N) \tag{2.7}$$

describes the receiver of the FC, where $\mathbf{y} \coloneqq (Y[1], \ldots, Y[n])$ denotes the length-n sequence of channel outputs.[2] Let $\varepsilon > 0$ be some given accuracy and

$$d_{\mathcal{S}^N} : \mathcal{S}^N \times \mathcal{S}^N \to \mathbb{R}_+ \tag{2.8}$$

some distortion measure. Then, the communication problem consists in designing the transmitters $\mathsf{Tx}_1, \ldots, \mathsf{Tx}_N$ and the receiver Rx such that

$$\mathbb{P}\Big(d_{\mathcal{S}^N}\big((\hat{s}_1, \ldots, \hat{s}_N), (s_1, \ldots, s_N)\big) < \varepsilon\Big) \geq 1 - \delta \ , \tag{2.9}$$

for some arbitrary small $\delta > 0$. Provided a solution to this problem exists, a communications engineer is generally interested in those solutions (i.e., transmitter-receiver combinations) that are efficient in the sense that (2.9) can be achieved with as little communication between the sensor nodes and the FC as possible. That is, with as little *channel uses/wireless resources* as possible.

Remark 2.4. The triple $(d_{\mathcal{S}^N}, \varepsilon, \delta)$ is typically determined by the sensor network application of interest.

The standard design philosophy to solve this problem (if feasible) is to divide each transmitter (2.6) as well as the receiver (2.7) into two main parts by employing some form of Shannon's well-known separation theorem [GK11, p. 66]. This means that the first part of each transmitter consists of a source encoder, which appropriately compresses each instantaneous sensor reading into a digital message (i.e., a sequence of binary symbols), followed by a channel encoder that bijectively maps the source messages to adequate channel input sequences. At the receiving end, the system is built up in reverse order where a channel decoder first reconstructs the N source messages from the sequence of channel outputs and a source decoder subsequently provides estimates $\hat{s}_1, \ldots, \hat{s}_N$ at its output.

Remark 2.5. Even if separating source and channel coding is optimal in the point-to-point case, it is generally suboptimal in multiuser scenarios [GK11, p. 336]. In those cases, source and channel coding should be considered jointly.

It is intuitively clear that the communication problem given above is subject to a fundamental trade-off between the desired accuracy $(d_{\mathcal{S}^N}, \varepsilon, \delta)$, the channel input constraints $(\varrho_n, \{P_i\}_{i=1}^N)$, and the number of channel uses n. Therefore, finding appropriate solutions to the problem generally falls within the area of rate distortion theory where our intuition is confirmed by some strong mathematical concepts [Ber71].

[2] We deliberately decided against denoting (2.6) and (2.7) as "encoder" and "decoder" in order to avoid later confusion when introducing dedicated channel encoders and decoders.

2.3 The Computation Problem

Instead of reconstructing $s_1, \ldots, s_N$ at the FC as in the previous problem, many wireless sensor network applications require to reliably and efficiently compute some pre-defined *function* thereof [GK05, GK06]. For instance, this can be the maximum temperature (or carbon monoxide concentration) in a building for fire detection, the average pressure in a steam boiler, or the minimum humidity in a greenhouse. In all that follows, we denote this as the desired function.

Definition 2.3 (Desired Function)**.** Let $\mathcal{S} \subset \mathbb{R}$ be some compact sensing range and $s_i \in \mathcal{S}$ the sensor reading of node i, $i = 1, \ldots, N$. Then, every function

$$f : \mathcal{S}^N \to \mathbb{R} \ , \ (s_1, \ldots, s_N) \mapsto f(s_1, \ldots, s_N)$$

that has to be computed at the FC is called a *desired function*.

In accordance with the definition, whenever the sensor nodes measure some values $s_1, \ldots, s_N$, the FC would like to have a sufficiently accurate estimate $\hat{f}(s_1, \ldots, s_N)$ of the corresponding function-value $f(s_1, \ldots, s_N)$. The distributed communication problem discussed in Section 2.2 therefore transforms into a distributed *computation problem*, formally specified as follows: Let the receiver (2.7) be modified to

$$\mathsf{Rx} : \mathcal{Y}^n \to \mathbb{R} \ , \ \mathbf{y} \mapsto \hat{f}(s_1, \ldots, s_N) \ . \tag{2.10}$$

Then, given a WMAC with fixed input cost constraints, a predefined accuracy $\varepsilon > 0$ and some distortion measure

$$d_{\mathbb{R}} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_+ \ , \tag{2.11}$$

the computation problem consists in designing the transmitters (2.6) and the receiver (2.10) such that the probability

$$\mathbb{P}\Big(d_{\mathbb{R}}(\hat{f}(s_1, \ldots, s_N), f(s_1, \ldots, s_N)) < \varepsilon\Big) \geq 1 - \delta \ , \tag{2.12}$$

for $\delta > 0$ arbitrary small. Analogously to the communication problem, efficient solutions to the computation problem (if existent) are generally of particular interest, which represent solutions that achieve (2.12) using as little channel uses/wireless resources as possible. Note that developing such solutions requires to cope with a fundamental trade-off between the tuples $(d_{\mathbb{R}}, f, \varepsilon, \delta)$, $(\varrho_n, \{P_i\}_{i=1}^N)$, and the number of channel uses n.

Remark 2.6. Note that (2.12), as well as (2.9) for the communication problem, are for our purposes of sufficient generality in the sense that they contain other commonly used error criteria as a special case. For instance, let (2.11) be chosen as

$$d_{\mathbb{R}}(\hat{f}(\boldsymbol{s}), f(\boldsymbol{s})) = \left|\hat{f}(\boldsymbol{s}) - f(\boldsymbol{s})\right|^2 \ ,$$

for all $\boldsymbol{s} \coloneqq (s_1, \ldots, s_N) \in \mathcal{S}^N$. Then, by appropriately setting ε and δ in (2.12), it follows with Markov's inequality (see Theorem B.1 in Appendix B) a *mean squared error (MSE)* criterion. More precisely, it holds that

$$1 - \mathbb{P}\left(\left|\hat{f}(\boldsymbol{s}) - f(\boldsymbol{s})\right|^2 < 1\right) = \mathbb{P}\left(\left|\hat{f}(\boldsymbol{s}) - f(\boldsymbol{s})\right|^2 \geq 1\right) \leq \mathbb{E}\left\{\left|\hat{f}(\boldsymbol{s}) - f(\boldsymbol{s})\right|^2\right\} \leq \delta \ .$$

2.4 Two Insightful Examples

Having defined what is meant in this thesis when referring to the communication or the computation problem over a WMAC, we want to demonstrate that the problems substantially differ in general. Therefore, let us take a look on two simple but insightful examples.

Example 2.1. For $\mathbb{K} = \mathbb{R}$, consider the ideal WMAC (2.2) without input cost constraints[3] and let the desired function be the sum of the sensor readings, that is,

$$f(s_1, \ldots, s_N) = \sum_{i=1}^{N} s_i \; . \tag{2.13}$$

Furthermore, let the sequence of channel inputs of node i be of the form

$$\boldsymbol{x}_i = \mathsf{Tx}_i(s_i) = \big(a_i[1], a_i[2], \ldots, a_i[n]\big)^{\mathsf{T}} s_i \qquad i = 1, \ldots, N \; , \tag{2.14}$$

$(a_i[1], \ldots, a_i[n]) \in \mathbb{R}^n$, which leads due to the memoryless property of the WMAC to the channel output sequence

$$\boldsymbol{y} = \begin{pmatrix} y[1] \\ y[2] \\ \vdots \\ y[n] \end{pmatrix} = \sum_{i=1}^{N} \boldsymbol{x}_i = \underbrace{\begin{pmatrix} a_1[1] & a_2[1] & \cdots & a_N[1] \\ a_1[2] & a_2[2] & \cdots & a_N[2] \\ \vdots & \vdots & \ddots & \vdots \\ a_1[n] & a_2[n] & \cdots & a_N[n] \end{pmatrix}}_{=:\boldsymbol{A}} \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{pmatrix} \; . \tag{2.15}$$

For computing the desired function-value (2.13) at the FC, we intend to first *uniquely* infer the N unknowns $s_1, \ldots, s_N$ from $\boldsymbol{y}$. It is one of the basic facts of linear algebra that a unique solution to the nonhomogeneous linear system (2.15) exists if and only if $\boldsymbol{A} = (a_i[j]) \in \mathbb{R}^{n \times N}$ is square and of full rank. Thus, the number of channel uses has to be equal to the number of nodes in the network (i.e., $n = N$) and the transmitters (2.14) have to be designed such that the vectors $(a_1[1], \ldots, a_1[N])^{\mathsf{T}}, \ldots, (a_N[1], \ldots, a_N[N])^{\mathsf{T}}$ span $\mathbb{R}^N$. Two corresponding computation strategies are the following, provided that the FC a priori knows $\mathsf{Tx}_1, \ldots, \mathsf{Tx}_N$ and therefore $\boldsymbol{A}$.

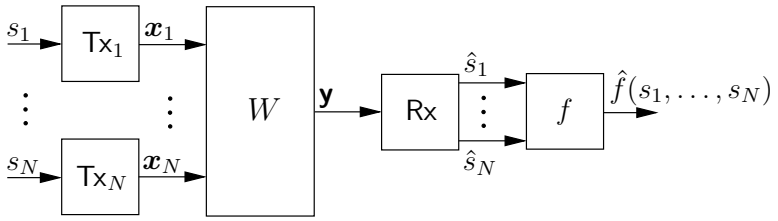
- Time Division Multiple-Access (TDMA): Let (2.14) be chosen as

  $$\mathsf{Tx}_i(s_i) = \big(\underbrace{0, \ldots, 0}_{i-1}, 1, \underbrace{0, \ldots, 0}_{n-i}\big)^{\mathsf{T}} s_i \qquad i = 1, \ldots, N \; ,$$

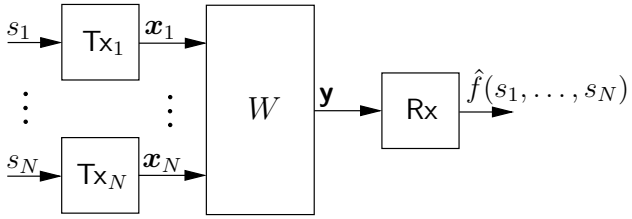
  which reflects a coordinated activation of nodes and therefore an interference avoiding strategy. As a consequence, $\boldsymbol{A}$ is equal to $\boldsymbol{I}_N$ so that the receiver (2.10) is simply given by

  $$\mathsf{Rx}(\boldsymbol{y}) = \mathbf{1}^{\mathsf{T}} \boldsymbol{y} = f(s_1, \ldots, s_N) \; . \tag{2.16}$$

[3] Because the ideal WMAC is a noiseless channel, it is not necessary here to explicitly incorporate input cost constraints as we can scale channel input and output signals arbitrarily without performance degradation.



(a) Separation-based computation



(b) Computation over MAC

Figure 2.4: Separation-based computation vs. computation over MAC.

- Code Division Multiple-Access (CDMA): For instance, let (2.14) be chosen for each i, $i = 1, \ldots, N$, such that $\boldsymbol{A}$ results in a Hadamard matrix of order N (if existent)[4] and therefore in an invertible element of $\{-1, 1\}^{N \times N}$. Then, (2.16) modifies to
$$\mathsf{Rx}(\boldsymbol{y}) = \mathbf{1}^{\mathsf{T}} \boldsymbol{A}^{-1} \boldsymbol{y} = f(s_1, \ldots, s_N) \,, \tag{2.17}$$
which represents an interference cancellation strategy.

△

The example shows that a solution to the communication problem may have the potential to also provide a solution to the computation problem. We call such strategies in all that follows *separation-based computation strategies* as they separate the process of communicating the data from the process of computing the desired function. In other words, a separation-based computation strategy first solves the communication problem and (if feasible) the computation problem afterwards. See Figure 2.4(a) for an illustration.

Example 2.2. In contrast to the previous example, consider now a single use of the ideal WMAC (i.e., $n = 1$) with transmitters chosen to be the identity function on $\mathcal{S}$:
$$x_i[1] = \mathsf{Tx}_i(s_i) = \mathrm{id}_{\mathcal{S}}(s_i) = s_i \qquad i = 1, \ldots, N \,. \tag{2.18}$$

[4] A Hadamard matrix $\boldsymbol{H}$ of order N is an element of $\{-1, 1\}^{N \times N}$ for which $\boldsymbol{H}\boldsymbol{H}^{\mathsf{T}} = N\boldsymbol{I}_N$. A necessary condition for a Hadamard matrix of order $N > 2$ to exist is that $N = 4k$, $k \in \mathbb{N}$. It is conjectured that this is also sufficient but during the writing of this thesis a proof was not known [CS10, p. 87].

Since the corresponding channel output is

$$y[1] = \sum_{i=1}^{N} x_i[1] = \sum_{i=1}^{N} s_i \ ,$$

choosing the receiver to be the identity function on $\mathbb{R}$ results in

$$\mathsf{Rx}(y[1]) = \mathrm{id}_{\mathbb{R}}(y[1]) = y[1] = f(s_1, \ldots, s_N) \tag{2.19}$$

and therefore immediately in the desired function-value (2.13). △

The moral of the above examples is that separation-based computation strategies can be highly inefficient. In particular, the strategies of Example 2.1 are capable of computing $1/N$ function-values per channel use in a distortion-free manner whereas the strategy of Example 2.2, referred to as *computation over MAC (CoMAC)*, allows to distortion-free compute one function-value per channel use. Thus, in the considered example, the efficiency is improved by a factor of N (i.e., the number of nodes in the network). The reason is that the strategy of Example 2.2 *exploits the interference* of the ideal WMAC for computing (2.13) rather than avoiding or canceling it and therefore *merges* the communication with the computation step (see Figure 2.4(b)).

Remark 2.7. For completeness, we point out that if it would be a priori known that in Example 2.1 at most $k < N/2$ of the sensor readings $s_1, \ldots, s_N$ are nonzero (i.e., the vector of sensor readings is *k-sparse*), then there would exist, at least in theory, transmitters $\mathsf{Tx}_1, \ldots, \mathsf{Tx}_N$ and a receiver Rx such that $f(s_1, \ldots, s_N)$ can be computed by a separation-based strategy with $n = 2k < N$ channel uses only. This follows from the theory of *compressive sensing*, which allows to find unique sparse solutions of underdetermined systems of equations [FR13, Ch. 2].

At first glance it may seem somewhat surprising that interference can be helpful in order to obtain better performance: in the field of wireless network design it is typically regarded as a hindrance to be overcome. Essential for the fact, however, that harnessing the interference leads in Example 2.2 to an N-fold increase in efficiency compared with the strategies of Example 2.1 is that the desired function *perfectly matches* the algebraic structure of the ideal WMAC. According to Definition 2.1, the algebraic structure of the ideal WMAC is $(\mathbb{K}, w)$, which is either $\mathbb{R}$ or $\mathbb{C}$ together with the N-ary operation $w : \mathbb{K}^N \to \mathbb{K}$, $(x_1, \ldots, x_N) \mapsto w(x_1, \ldots, x_N) = \sum_{i=1}^{N} x_i$.[5] Under a perfect match, we therefore understand in view of (2.18)

$$\forall (s_1, \ldots, s_N) \in \mathcal{S}^N : w(x_1, \ldots, x_N) = f(s_1, \ldots, s_N) \ . \tag{2.20}$$

In short, the WMAC itself computes the desired function.

[5] It can be shown that $(\mathbb{K}, w)$ forms an algebraic structure that is in the field of universal algebra known as a "reducible N-ary group" as it is a straighforward generalization of the ordinary additive group $(\mathbb{K}, +)$ [Dud01]. Note that when referring to the linear structure (or simply the structure) of the WMAC, we exclusively mean $(\mathbb{K}, w)$.

The above insight that a perfect match between the desired function and the algebraic structure of an underlying MAC can be beneficially exploited was first made by Nazer and Gastpar in their seminal paper "Computation Over Multiple-Access Channels" [NG07]. Since their results are mainly limited to the efficient computation of linear functions over MACs having linear structure, there immediately arise a couple of further interesting questions that will be addressed in the next chapter. One of those questions is, for instance, whether there is a chance that harnessing interference can be advantageous over separation-based computation strategies even if the desired function *does not perfectly match* the structure of the WMAC.

Remark 2.8. Note that in general every function can be reliably computed over the WMAC by using, for instance, a separation-based approach. Crucial is, however, the question how efficiently this can be done in terms of the number of channel uses.

Remark 2.9. In the context of wireless sensor network design, energy efficiency is commonly considered as one of the major concerns [ASSC02, ZG04]. If we have been talking of efficiency so far, however, we had in mind achieving a certain objective with as little transmissions as possible. Therefore, a computation strategy is said to be more efficient than a competing strategy if it requires less channel uses per function-value at fixed input cost constraints. Under the simplifying assumption that the sensor nodes' energy consumption is mainly determined by the transmit energy, higher efficiency in terms of the number of transmissions implies higher efficiency in terms of energy.

3

Computation Over the Wireless Channel – Fundamental Limits

> *Das Instrument, welches die Vermittlung bewirkt zwischen Theorie und Praxis, zwischen Denken und Beobachten, ist die Mathematik; sie baut die verbindende Brücke und gestaltet sie immer tragfähiger. Daher kommt es, dass unsere ganze gegenwärtige Kultur, soweit sie auf der geistigen Durchdringung und Dienstbarmachung der Natur beruht, ihre Grundlage in der Mathematik findet.*
>
> David Hilbert, 1930

Over the last decade it has been increasingly recognized by research groups from different fields such as information theory, signal processing, and communications that the interference property of the wireless channel can be helpful in order to fulfill certain tasks much more efficiently. For instance, Gastpar and Vetterli propose in [GV03] (see also [Gas08] and [GVD06]) an analog joint source-channel communication scheme that harnesses the interference property of the Gaussian MAC in order to estimate some parameter of interest from a collection of sensor readings that are corrupted by independent Gaussian sensing noise. The approach outperforms separation-based strategies as the Gaussian MAC itself computes the optimal linear estimation function (i.e., the arithmetic mean) if the nodes are allowed to transmit concurrently in the same frequency band. This observation initiated many research activities that extended the analog joint source-channel approach to more general distributed estimation problems such as those in [MT06,BHSN07,XCLG08,BTS10,BTS12], whereas references [DS98, LD07, MNT07, LS07, TD11, LED12, BSTS12] are devoted to the detection counterparts.

The basic idea of *physical-layer network coding* is also to benefit from the interference caused by transmitters that actively share the same spectrum. Unlike the traditional network coding paradigm applied across data packets on the network layer [ACLY00, LYC03, KM03], physical-layer network coding generates linear combinations of packets instantly "in the air" while on the way to their receivers, which are then simply forwarded to the intended destinations [ZLL06, PY06, KGK07, NG11b]. This principle can be also applied to *two-way relaying*, also known as bidirectional relaying, in which users exchange messages with each other by means of a relay that is located somewhere in between [WNPS10, NCL10, NG11a].

Within the realm of *resource allocation for wireless networks*, Stańczak, Wiczanowski, and Boche propose in [SWB05] (see also [SWB07]) an interesting analog transmission scheme that harnesses interference in order to efficiently obtain linear combinations of some local measurements over a network of interfering fading channels. Knowing the linear combinations then allows for the distributed computation of a transmit power allocation (across users) that maximizes some aggregate network utility.

In a nutshell, each of the above referenced works makes, in one way or another, use of interference to compute at a designated receiver a *function* that is dictated by some underlying application. Starting from this generic perspective, Nazer and Gastpar in [NG07] lay the information-theoretic foundation of computing arbitrary functions over arbitrary MACs. As already mentioned at the end of Chapter 2, however, their main results are concerned with the special case of reliably computing linear functions over MACs having a linear structure. In this regard, they show that in many cases, the performance gains over separation-based computation strategies are proportional to N (i.e., the number of concurrently active nodes/users). Less information-theoretic but no less interesting, in [KKF10] Keller, Karamchandani, and Fragouli provide conditions under which linear and specific nonlinear functions are computable over error-free linear vector-channels with finite alphabets. A similar holds true for [KKFF13].

In order to allow a systematic and in-depth analysis, in this chapter we follow a similar approach and consider the ideal WMAC model of Definition 2.1 in greater detail.[1] In particular, we answer the following fundamental questions:

(Q1) Which functions are efficiently computable over an ideal WMAC?

(Q2) What is the highest possible computation efficiency expressed in terms of the number of channel uses that are necessary to compute such functions?

(Q3) What are the properties of corresponding computation networks with regard to node complexity and coordination effort?

As Example 2.2 and the results in [NG07, KKF10] already suggest, (Q1) and (Q2) have a trivial answer when linear functions are to be computed. Many promising WSN applications, however, require the efficient computation of *nonlinear functions* [GK06]

[1] True to the motto "something that does not work in the ideal case will, a fortiori, does not work in the noisy case".

so that in Section 3.1 we first examine what is achievable when desired functions are not perfectly matched to the structure of the ideal WMAC. It turns out that these efforts are closely related to the famous *13th problem* posed by David Hilbert in 1900 [Hil00]. Certainly a surprising fact that not only draws attention to a beautiful mathematical theory that is perhaps a little less known in communications engineering but also leads to some surprising and fruitful insights.

Based on the results of Section 3.1, in Section 3.2 we consider a generalized sensor network model consisting of multiple FCs, each of which aims at independently computing some function of the sensor readings of an arbitrary subset of nodes (i.e., clusters). Thus, we attempt to answer questions (Q1)–(Q3) in the context of clustered sensor networks, where the intra-cluster communication is modeled by ideal WMACs. Special attention is paid to those networks in which clusters overlap so that the WMACs interfere with each other.

As already mentioned, sensor nodes are typically subject to strict resource constraints such as finite energy and processing capabilities and often lack the support of an established infrastructure [ZG04]. Therefore, in order to maintain the basic functionality, a WSN as whole has to independently deal with variations in network topology due to sensor nodes that drop out of the network (e.g., battery depletion, link failures) or new sensor nodes that join the network. Accordingly, we analyze in Section 3.3 how variations in network topology affect the results of Sections 3.1 and 3.2. Finally, Section 3.4 concludes the chapter.

Convention

In what follows, $\mathscr{F}(\mathbb{A})$ denotes the space of all real-valued functions $f : \mathbb{A} \to \mathbb{R}$ with arbitrary domain $\mathbb{A}$ and $\mathscr{C}^0(\mathbb{B})$ the space of real-valued continuous functions defined over some topological space $\mathbb{B}$, respectively. Without loss of generality, the sensing range, $\mathcal{S}$, is in this chapter for simplicity assumed to be equal to the closed unit interval $\mathbb{E} \coloneqq [0, 1]$.

3.1 Computation Over Ideal WMACs

Recall from the definition of the N-user ideal WMAC in Section 2.1 of Chapter 2 that the signal received at the FC is of the form

$$y[j] = w\big(x_1[j], \ldots, x_N[j]\big) = \sum_{i=1}^{N} x_i[j] \qquad j = 1, \ldots, n \,, \tag{3.1}$$

where we assume $w : \mathbb{R}^N \to \mathbb{R}$ (see Remark 2.1). In order to answer questions (Q1) and (Q2) in a satisfying way, in Sections 3.1.1 and 3.1.2 we first examine what is fundamentally possible with a single channel use (i.e., $n = 1$), whereas computations with $n > 1$ channel uses are considered in detail in Section 3.1.3.

3.1.1 Computation With a Single Channel Use

Karamchandani, Niesen, and Diggavi show in [KND13] that if there is a *structural mismatch* between a given ideal MAC and some desired function (i.e., condition (2.20) is not satisfied for a given w), then separation-based computation is almost always optimal with respect to the number of channel uses. In other words, for almost all mismatched combinations of MACs and desired functions, an attempt of exploiting interference to increase the computation efficiency fails.[2]

The result of Karamchandani et al. may appear a little discouraging but at least for the ideal WMAC in which we are interested, we provide in the following a kind of remedy by employing transmitters and receivers that are able to appropriately settle mismatches. More formally, each node in the network transforms its sensor readings prior to transmission by an individual pre-processing function defined as follows:

Definition 3.1 (Pre-Processing Functions)**.** We define the univariate function

$$\varphi_i : \mathbb{E} \to \mathbb{R} \ , \ s_i \mapsto \varphi_i(s_i) \tag{3.2}$$

to be a *pre-processing function* of node i, $i = 1, \ldots, N$.

In addition, the FC treats the received symbols with a certain post-processing function:

Definition 3.2 (Post-Processing Function)**.** Let $y \in \mathbb{R}$ be the output of the ideal WMAC at an arbitrary channel use. Then, we define the univariate function

$$\psi : \mathbb{R} \to \mathbb{R} \ , \ y \mapsto \psi(y) \tag{3.3}$$

to be a *post-processing function*.

According to these definitions, we simply choose the *computation transmitter* of node i to

$$\mathsf{Tx}_i(s_i) = x_i[1] = \left(\mathrm{id}_{\mathbb{R}} \circ \varphi_i\right)(s_i) = \varphi_i(s_i) \qquad i = 1, \ldots, N \tag{3.4}$$

and the *computation receiver* to

$$\mathsf{Rx}(y[1]) = \left(\mathrm{id}_{\mathbb{R}} \circ \psi\right)(y[1]) = \psi(y[1]) \ , \tag{3.5}$$

respectively.[3] It is obvious that this particular strategy allows the FC to compute every desired function $f : \mathbb{E}^N \to \mathbb{R}$ for which there exist N pre-processing functions as well as a post-processing function such that f can be represented for all $(s_1, \ldots, s_N) \in \mathbb{E}^N$ as

$$f(s_1, \ldots, s_N) = \psi\left(\sum_{i=1}^{N} \varphi_i(s_i)\right) \ . \tag{3.6}$$

[2] "Almost always" in the sense that exploiting interference is optimal only for a set of mismatched combinations of MACs and desired functions that is of measure zero.

[3] It may seem exaggerated to additionally mention mappings Tx_i and Rx in (3.4) and (3.5) as they are identical to the pre- and post-processing functions. In later chapters, however, Tx_i, $i = 1, \ldots, N$, and Rx consist of further components.

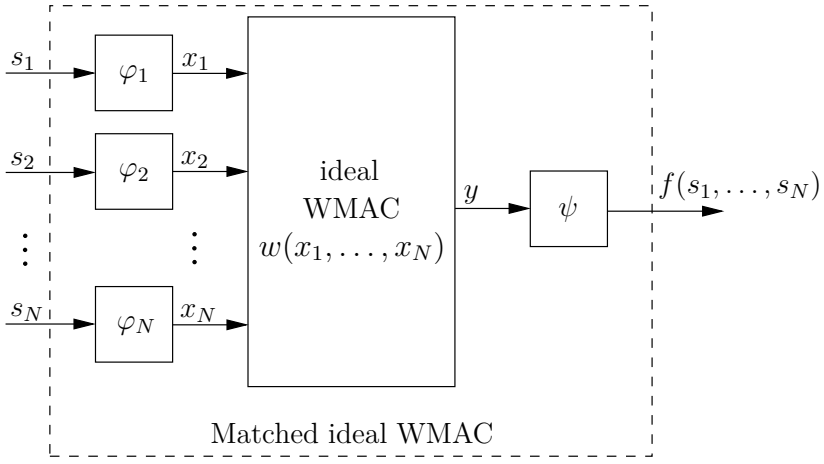


Figure 3.1: The ideal WMAC *matched* to the desired function f by appropriately chosen pre- and post-processing functions $\varphi_1, \ldots, \varphi_N, \psi$. The resulting overall channel has sensor readings as inputs and provides the desired function-value at its output.

Remark 3.1. The pre- and post-processing functions transform the ideal WMAC such that the resulting overall channel is *matched* to the desired function, which relaxes the stringent notion of a perfect match in (2.20) to

$$\forall (s_1, \ldots, s_N) \in \mathbb{E}^N : \mathsf{Rx}(w(x_1, \ldots, x_N)) = f(s_1, \ldots, s_N) \, . \tag{3.7}$$

See Figure 3.1 for an illustration.

It is an interesting coincidence that the function space consisting of all $f \in \mathscr{F}(\mathbb{E}^N)$ having a representation (3.6) is in the realm of multivariate calculus and functional analysis known as the space of *nomographic functions* [Buc79]. In what follows, we denote this space as $\mathscr{N}(\mathbb{E}^N)$, that is,

$$\begin{aligned} \mathscr{N}(\mathbb{E}^N) \coloneqq \Bigg\{ f : \mathbb{E}^N \to \mathbb{R} \,\Bigg|\, & \exists (\varphi_1, \ldots, \varphi_N, \psi) \in \mathscr{F}(\mathbb{E}) \times \cdots \times \mathscr{F}(\mathbb{E}) \times \mathscr{F}(\mathbb{R}) \\ & \forall (s_1, \ldots, s_N) \in \mathbb{E}^N : f(s_1, \ldots, s_N) = \psi \left(\sum_{i=1}^{N} \varphi_i(s_i) \right) \Bigg\} \, . \end{aligned} \tag{3.8}$$

The elements from $\mathscr{N}(\mathbb{E}^N)$ are called nomographic functions because they are the basis of *nomographs*.[4] Nomographs are graphical aids that have primarily been used before the digital age to manually solve complicated equations [Eps58, Eve82, Doe09]. Figure 3.2, for instance, depicts the simple nomograph of an affine function of two

[4] Originating from "nomos (*νόμος*)", the Greek term for "law".

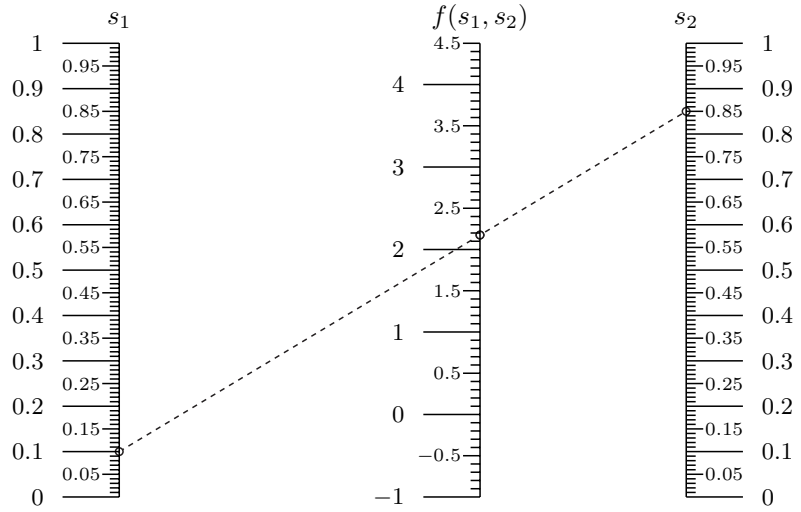


Figure 3.2: Nomograph of function $f(s_1, s_2) = \psi(\varphi_1(s_1) + \varphi_2(s_2)) = 2s_1 + 3.5s_2 - 1$, with pre-processing functions $\varphi_1(s_1) = 2s_1$, $\varphi_2(s_2) = 3.5s_2$, and post-processing function $\psi(y) = y - 1$. In order to read off a specific function-value from the middle scale one has to simply connect the corresponding arguments located on the outer scales by a straight line, as it is demonstrated for $f(0.1, 0.85) = 2.175$.

variables drawn with Leif Roschier's freeware PyNomo [Ros09]. Another popular example is the Smith chart, which is often used for circuit design in microwave engineering [Poz05, p. 64].

Remark 3.2. In favor of historical correctness, we point out that the notion "nomographic function" is not unique as the polish mathematician Mieczysław Warmus gave an early definition in his postdoctoral dissertation [War59] (see also [Eve82, Sec. 4.4]) that is different from that in (3.6). To those who are interested, we provide his idea of a nomographic function in Definition A.13, Appendix A.

Example 3.1 (Nomographic Functions)**.** Let $s_i \in \mathbb{E}$, $i = 1, \ldots, N$.

- Arithmetic Mean: $f(s_1, \ldots, s_N) = \frac{1}{N} \sum_{i=1}^{N} s_i$, with $\varphi_i(s) = s$, $i = 1, \ldots, N$, and $\psi(y) = y/N$.
- Weighted Mean: $f(s_1, \ldots, s_N) = \sum_{i=1}^{N} \lambda_i s_i / \sum_{i=1}^{N} \lambda_i$, with $\varphi_i(s) = \lambda_i s$, $\lambda_i \geq 0$, $i = 1, \ldots, N$, and $\psi(y) = y/(\sum_{i=1}^{N} \lambda_i)$.
- Harmonic Mean: $f(s_1, \ldots, s_N) = N/\sum_{i=1}^{N} s_i^{-1}$, with $\varphi_i(s) = s_i^{-1}$, $i = 1, \ldots, N$, and $\psi(y) = N/y$.
- Polynomial: $f(s_1, \ldots, s_N) = a_0 + a_1 s_1^{\ell_1} + a_2 s_2^{\ell_2} + \cdots + a_N s_N^{\ell_N}$, with $\varphi_i(s_i) = a_i s_i^{\ell_i}$, $\ell_i \in \{0, \ldots, N\}$ arbitrary, $i = 1, \ldots, N$, and $\psi(y) = y + a_0$.

- Euclidean Norm: $f(s_1, \ldots, s_N) = \sqrt{s_1^2 + \cdots + s_N^2}$, with $\varphi_i(s) = s^2$, $i = 1, \ldots, N$, and $\psi(y) = \sqrt{y}$.
- Number of Active Nodes: $f(s_1, \ldots, s_N) = N$, with $\varphi_i(s) \equiv C$ for some $C \in \mathbb{R}$, $i = 1, \ldots, N$, and $\psi(y) = y/C$.

△

A closer look at the examples reveals that the pre-processing functions depend on the particular choice of f so that a desirable property is the following.

Definition 3.3 (Universality)**.** A set of fixed pre-processing functions $\varphi_1, \ldots, \varphi_N$ is said to be *universal* with respect to some function space $\mathscr{A}(\mathbb{E}^N)$ if and only if they allow to represent *every* $f \in \mathscr{A}(\mathbb{E}^N)$ in the form of (3.6) by a proper choice of the post-processing function ψ.

Remark 3.3. Universality in the sense of Definition 3.3 is a highly desirable property for all-purpose computation networks because it mainly defines the communication structure within the network as well as to what extent coordinating feedback is required.

Definition 3.3 along with the observation that (3.8) is exactly the space of nomographic functions leads us to the following powerful theorem, which completely determines the functions that are computable over an ideal WMAC and therefore entirely answers questions (Q1) and (Q2).

Theorem 3.1. *Let $N \in \mathbb{N}$ be arbitrary. Then, every desired function $f : \mathbb{E}^N \to \mathbb{R}$ is universally computable over an ideal WMAC with a single channel use and zero distortion.*

Proof. The assertion is an immediate consequence of a result proven by Sprecher in [Spr65], which states that *every* $f \in \mathscr{F}(\mathbb{E}^N)$ is nomographic (i.e., $\mathscr{N}(\mathbb{E}^N) = \mathscr{F}(\mathbb{E}^N)$). In order to make the interested reader familiar with the corresponding proof technique, we present an expanded and slightly modified version of Sprecher's proof in Appendix 3.A.1 at the end of the chapter. □

The denotation "universally computable" refers to the fact that the pre-processing functions can be chosen at sensor nodes to be universal and therefore *independent* of the function to be computed. For the communication between the FC and the nodes this means that *no additional feedback*[5] is necessary to inform the nodes when the FC modifies the desired function during network operation. The FC decides by an appropriate choice of ψ which f shall be computed. The universality property therefore offers the potential to significantly reduce in practice the amount of coordination as well as the hardware complexity of sensor nodes. In Section 3.2, the property will play a key role in improving the efficiency of more general computation networks.

[5] Additional to the feedback that is mandatory in practical wireless systems (e.g., for providing channel state information).

Remark 3.4. In the meaning of Section 2.3, *every* computation problem can be solved by using the ideal WMAC only once.

Note that in Theorem 3.1, there are no restrictions on pre- and post-processing functions imposed so that they can be arbitrary. Since continuity can be a useful property for practical implementations as well as for appropriately evaluating the computation error (2.12) in later chapters, it is interesting to ask if Theorem 3.1 is also true when pre- and post-processing functions are required to be continuous. In this regard, we denote in contrast to (3.8) the space of nomographic functions with the additional property

$$(\varphi_1, \ldots, \varphi_N, \psi) \in \mathscr{C}^0(\mathbb{E}) \times \cdots \times \mathscr{C}^0(\mathbb{E}) \times \mathscr{C}^0(\mathbb{R}) \tag{3.9}$$

as $\mathscr{N}^0(\mathbb{E}^N)$.

In order to answer the question, we first point to the fact that the pre-processing functions constructed in the proof of Theorem 3.1 (see (3.40) in Appendix 3.A.1) are universal but merely continuous almost everywhere.[6] An essential part of the proof is then to show that this choice of pre-processing functions results in a bijective correspondence

$$(s_1, \ldots, s_N) \mapsto \sum_{i=1}^{N} \varphi_i(s_i) \ .$$

That this is necessary and sufficient for universality is stated by the following theorem.

Theorem 3.2. *Let $N \geq 2$ be arbitrary and*

$$g : \mathbb{E}^N \to \Upsilon \ , \ (s_1, \ldots, s_N) \mapsto g(s_1, \ldots, s_N) = \sum_{i=1}^{N} \varphi_i(s_i) \ , \tag{3.10}$$

with $\Upsilon \subseteq \mathbb{R}$ denoting the range of g. Then, $\varphi_1, \ldots, \varphi_N$ are universal pre-processing functions with respect to $\mathscr{F}(\mathbb{E}^N)$ if and only if g is bijective.

Proof. The proof is deferred to Appendix 3.A.2 at the end of the chapter. □

From the proofs of Theorems 3.1 and 3.2, we conclude that for (3.10) to be bijective, the pre-processing functions have to be chosen such that for all $\boldsymbol{s}^{(1)} \coloneqq (s_1^{(1)}, \ldots, s_N^{(1)}) \in \mathbb{E}^N$ and $\boldsymbol{s}^{(2)} \coloneqq (s_1^{(2)}, \ldots, s_N^{(2)}) \in \mathbb{E}^N$, $\boldsymbol{s}^{(1)} \neq \boldsymbol{s}^{(2)}$,

$$\sum_{i=1}^{N} \varphi_i(s_i^{(1)}) \neq \sum_{i=1}^{N} \varphi_i(s_i^{(2)}) \tag{3.11}$$

holds (see Figure 3.3), which means that the ranges of the pre-processing functions have to be appropriate. To illustrate that this is possible in general, we consider the special case $N = 2$ and construct a field $\mathbb{K} \subset \mathbb{R}$ that has the cardinality of the continuum

[6] A pre-processing function is "continuous almost everywhere" if it is discontinuous only on a set of Lebesgue measure zero.

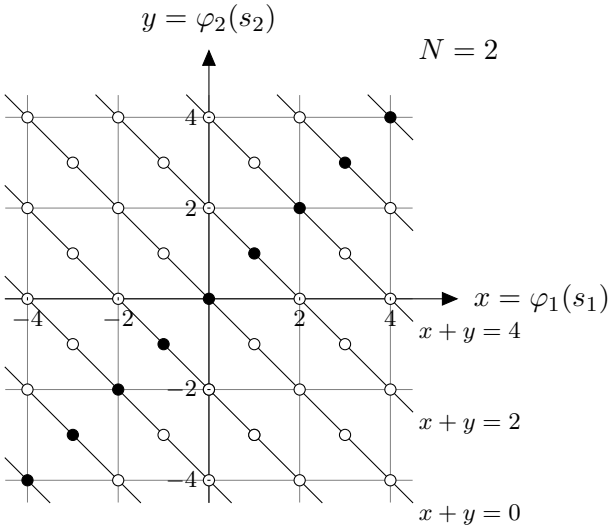


Figure 3.3: A line grid of points $\varphi_1(s_1) + \varphi_2(s_2) = \text{const.}$ illustrating the bijectivity requirement of the function g. The pre-processing functions have to be chosen such that $\forall (s_1^{(1)}, s_2^{(1)}) \neq (s_1^{(2)}, s_2^{(2)}) : \varphi_1(s_1^{(1)}) + \varphi_2(s_2^{(1)}) \neq \varphi_1(s_1^{(2)}) + \varphi_2(s_2^{(2)})$. That means in the depicted example that only the black points are allowed in the range of g whereas the white ones have to be avoided.

without containing every real number.[7] More precisely, we consider some real number α that is not in $\mathbb{K}$ (i.e., $\alpha \in \mathbb{R} \backslash \mathbb{K}$) and define $\mathbb{K}$ as the range of φ_1. Furthermore, we define the range of φ_2 to be the field $\frac{1}{\alpha}\mathbb{K}$. Then, for every $(s_1^{(1)}, s_2^{(1)}), (s_1^{(2)}, s_2^{(2)}) \in \mathbb{E}^2$ with $(s_1^{(1)}, s_2^{(1)}) \neq (s_1^{(2)}, s_2^{(2)})$, it follows

$$\varphi_1(s_1^{(1)}) + \varphi_2(s_2^{(1)}) \neq \varphi_1(s_1^{(2)}) + \varphi_2(s_2^{(2)}) \ .$$

Would this not be the case,

$$\underbrace{\varphi_1(s_1^{(1)}) - \varphi_1(s_1^{(2)})}_{\in \mathbb{K}} = \varphi_2(s_2^{(2)}) - \varphi_2(s_2^{(1)}) = \frac{1}{\alpha} \underbrace{\left(y^{(2)} - y^{(1)}\right)}_{\in \mathbb{K}}$$

would follow and thus $\alpha \in \mathbb{K}$, which is a contradiction since $\alpha \in \mathbb{R} \backslash \mathbb{K}$.

This simple example already reveals that the necessary *separation* of all points in the range Υ of g can unfortunately never be achieved for every $f \in \mathscr{F}(\mathbb{E}^N)$ with *continuous* pre-processing functions so that g and ψ are discontinuous in general. This observation is completed by the following theorem.

[7] In [vN28], von Neumann constructs an example of such a field without using the axiom of choice.

Theorem 3.3. *There exists a nonempty compact subset* $\mathbb{D}$ *of* $\mathbb{E}^N$ *such that for all* $f \in \mathscr{C}^0(\mathbb{E}^N)$ *vanishing on and only on* $\mathbb{D}$ *there exists* $\varepsilon = \varepsilon(f) > 0$ *such that*

$$\inf_{f' \in \mathscr{N}^0(\mathbb{E}^N)} \sup_{(s_1,\ldots,s_N)\in\mathbb{D}} \left| f(s_1,\ldots,s_N) - f'(s_1,\ldots,s_N) \right| \geq \varepsilon \, .$$

That is, the space of nomographic functions with continuous pre- and post-processing functions, $\mathscr{N}^0(\mathbb{E}^N)$*, is* nowhere dense *in the space of continuous functions* $\mathscr{C}^0(\mathbb{E}^N)$*.*

Proof. The constructive proof for arbitrary N is given by Buck in [Buc82]. For the special case $N = 2$, however, the theorem was previously proven by Arnol'd in [Arn57]. □

Remark 3.5. In simple terms, the theorem states that for continuous desired functions having $\mathbb{D}$ as its level set there do not exist (with a neighborhood) nomographic representations in which pre- and post-processing functions are continuous.

Example 3.2 (Geometric Mean)**.** Let the desired function be the "geometric mean"

$$f : \mathbb{E}^N \to \mathbb{R} \, , \; (s_1,\ldots,s_N) \mapsto f(s_1,\ldots,s_N) = \left(\prod_{i=1}^{N} s_i \right)^{\frac{1}{N}} \, . \tag{3.12}$$

In the following, we prove that (3.12) does not belong locally to the closure of $\mathscr{N}^0(\mathbb{E}^N)$. To this end, let

$$\mathbb{D} = \left\{ (s_1,\ldots,s_N) \in \mathbb{E}^N \;\middle|\; \prod_{i=1}^{N} s_i = 0 \right\}$$

denote the closed subset of $\mathbb{E}^N$ on (and only on) which (3.12) vanishes. Furthermore, let

$$\mathfrak{E} \coloneqq \Big\{ (1, \underbrace{0\ldots,0}_{N-1}), (0,1,\ldots,0), \ldots, (0,0,\ldots,1) \Big\} \cup (0,\ldots,0) \subset \mathbb{E}^{N+1}$$

and suppose that there exist continuous pre- and post-processing functions such that

$$\forall (s_1,\ldots,s_N) \in \mathbb{E}^N : \left(\prod_{i=1}^{N} s_i \right)^{\frac{1}{N}} = \psi \left(\sum_{i=1}^{N} \varphi_i(s_i) \right) \, .$$

Now, each pair of points taken from $\mathfrak{E}$ can be connected by a polygonal line lying entirely in $\mathbb{D}$ as well as by a polygonal line lying in the complement $\mathbb{D}^{\mathrm{c}} \coloneqq \mathbb{E}^N \backslash \mathbb{D}$, except at the end points (i.e., the elements of $\mathfrak{E}$). Suppose that at the end points the continuous function[8] $\sum_{i=1}^{N} \varphi_i(s_i)$ takes different values ξ_i, $i = 1,\ldots,N+1$. Then, this leads to a contradiction because $\sum_{i=1}^{N} \varphi_i(s_i)$ would take the intermediate values $\frac{\xi_k+\xi_\ell}{2}$, $k,\ell = 1,\ldots,N+1$, $k \neq \ell$, on $\mathbb{D}$ *and* on $\mathbb{D}^{\mathrm{c}}$ so that

$$\psi \left(\frac{\xi_k + \xi_\ell}{2} \right) = 0 \quad \text{and} \quad \psi \left(\frac{\xi_k + \xi_\ell}{2} \right) > 0$$

[8] Every finite sum of continuous functions is continuous [Rud76, Thm. 4.9].

simultaneously.[9] We therefore conclude that $\sum_{i=1}^{N} \varphi_i(s_i)$ takes the same value at each $(s_1, \dots, s_N) \in \mathfrak{E}$, that is,

$$\varphi_1(1) + \varphi_2(0) + \cdots + \varphi_N(0) = \varphi_1(0) + \varphi_2(1) + \cdots + \varphi_N(0) = \cdots$$
$$\cdots = \varphi_1(0) + \varphi_2(0) + \cdots + \varphi_N(1) \, .$$

By a simple calculation, it follows that

$$\varphi_1(0) + \cdots + \varphi_N(0) = \varphi_1(1) + \cdots + \varphi_N(1)$$

as well as

$$\psi(\varphi_1(0) + \cdots + \varphi_N(0)) = \psi(\varphi_1(1) + \cdots + \varphi_N(1)) \, .$$

This, however, is in contradiction to

$$0 = f(0, \dots, 0) \neq f(1, \dots, 1) = 1$$

so that we conclude that there *do not exist* continuous functions $\varphi_1, \dots, \varphi_N, \psi$ ensuring that the "geometric mean" (3.12) belongs to $\mathscr{N}^0(\mathbb{E}^N)$. △

3.1.2 Approximation With a Single Channel Use

According to Theorem 3.3, requiring pre- and post-processing functions to be continuous significantly reduces the amount of functions that are computable over an ideal WMAC with a single channel use. A strict prerequisite of this statement is that the computations are distortion free, which means in terms of Section 2.3 that for some given distortion measure $d_{\mathbb{R}}$ it holds that

$$d_{\mathbb{R}}(f(s_1, \dots, s_N), \hat{f}(s_1, \dots, s_N)) \equiv 0 \, .$$

However, if we tolerate some small distortion $\varepsilon > 0$, the problem turns into a multivariate approximation problem that is determined by what we call a *nomographic approximation*.

Definition 3.4 (Nomographic Approximations)**.** Let $\varepsilon > 0$ be arbitrary but fixed. Then, we define

$$\mathscr{N}_{d,\varepsilon}^{0}(\mathbb{E}^N) \coloneqq \Bigg\{ f \in \mathscr{F}(\mathbb{E}^N) \,\Bigg|\, \exists (\varphi_1, \dots, \varphi_N, \psi) \in \mathscr{C}^0(\mathbb{E}) \times \cdots$$
$$\cdots \times \mathscr{C}^0(\mathbb{E}) \times \mathscr{C}^0(\mathbb{R}) : d_{\mathbb{R}}\Big(f(s_1, \dots, s_N), \psi\Big(\sum\nolimits_i \varphi_i(s_i)\Big)\Big) \leq \varepsilon \Bigg\},$$

as the space of *approximable nomographic functions* with respect to the distortion measure $d_{\mathbb{R}}$ and accuracy ε. If $f \in \mathscr{N}_{d,\varepsilon}^{0}(\mathbb{E}^N)$ for all $(s_1, \dots, s_N) \in \mathbb{E}^N$, we write $f(s_1, \dots, s_N) \approx \psi(\sum_i \varphi_i(s_i))$.

[9] From the intermediate value theorem [Rud76, Thm. 4.23] we know that a continuous real-valued function defined on a continuum takes all values between any pair of given points.

An adequate characterization of the spaces $\mathscr{N}_{d,\varepsilon}^{0}(\mathbb{E}^N)$ is still a difficult problem. However, if we go back, for instance, to Example 3.2, then there exists for every $\varepsilon > 0$ a $p_0 = p_0(\varepsilon)$ such that

$$\forall p \geq p_0 \, \forall (s_1, \ldots, s_N) \in \mathbb{E}^N : \left| \left(\prod_{i=1}^{N} s_i \right)^{\frac{1}{N}} - \exp_{\mathrm{e}} \left(\frac{1}{N} \sum_{i=1}^{N} \log_{\mathrm{e}} \left(s_i + \frac{1}{p} \right) \right) \right| \leq \varepsilon \, .$$

The desired function "geometric mean" is therefore the uniform limit of the sequence of nomographic functions

$$\left\{ \psi \left(\sum_{i=1}^{N} \varphi_i^{(p)}(s_i) \right) = \exp_{\mathrm{e}} \left(\frac{1}{N} \sum_{i=1}^{N} \log_{\mathrm{e}} \left(s_i + \frac{1}{p} \right) \right) \right\}_{p \in \mathbb{N}} , \tag{3.13}$$

with pre-processing functions

$$\varphi_i^{(p)}(s) = \log_{\mathrm{e}} \left(s + \frac{1}{p} \right) \qquad i = 1, \ldots, N$$

and post-processing function

$$\psi(y) = \exp_{\mathrm{e}} \left(\frac{y}{N} \right) ,$$

respectively. Roughly speaking, the "geometric mean", which is indeed an element of the function space $\mathscr{N}(\mathbb{E}^N)$ but unfortunately not of $\mathscr{N}^0(\mathbb{E}^N)$ (see Example 3.2), can be uniformly approximated with arbitrary precision by a nomographic function that consists of continuous pre- and post-processing functions. Therefore, when choosing the distortion measure as

$$\begin{aligned} d_{\mathbb{R}}(f(s_1, \ldots, s_N), f'(s_1, \ldots, s_N)) &= \sup_{(s_1, \ldots, s_N) \in \mathbb{E}^N} \left| f(s_1, \ldots, s_N) - f'(s_1, \ldots, s_N) \right| \\ &= \left\| f - f' \right\|_{\infty} , \end{aligned} \tag{3.14}$$

the "geometric mean" belongs to $\mathscr{N}_{\|\cdot\|_{\infty},\varepsilon}^{0}(\mathbb{E}^N)$ for every $\varepsilon > 0$.

Remark 3.6. The example "geometric mean" demonstrates that the WMAC is apparently able to *multiply* so that the wireless channel can be regarded as a "computer" that is capable of performing all basic arithmetic operations.

With these thoughts in mind, the following proposition is an immediate consequence of Definition 3.4.

Proposition 3.1. *Let $\varepsilon > 0$ be some desired accuracy and $d_{\mathbb{R}}$ an arbitrary but fixed distortion measure. Then, every $f \in \mathscr{N}_{d,\varepsilon}^{0}(\mathbb{E}^N)$ can be computed over the ideal WMAC within accuracy ε (with respect to $d_{\mathbb{R}}$) by a single channel use.*

Example 3.3 (Nomographic Approximations)**.** Let $\varepsilon > 0$ be arbitrary but fixed, the distortion measure $d_{\mathbb{R}}$ as in (3.14), and $p_0 = p_0(\varepsilon)$ be chosen such that $f \in \mathscr{N}_{d,\varepsilon}^0(\mathbb{E}^N)$ for all $p \geq p_0$.

- Cosine of the Product: $f(s_1, \ldots, s_N) = \cos\big(\prod_{i=1}^N s_i\big) \approx \psi\big(\sum_{i=1}^N \varphi_i(s_i)\big)$, with $\varphi_i(s) = \log_{\mathrm{e}}(s + 1/p_0)$, $i = 1, \ldots, N$, and $\psi(y) = (\cos \circ \exp_{\mathrm{e}})(y) = \cos(\mathrm{e}^y)$.
- Monomial: $f(s_1, \ldots, s_N) = a s_1^{\ell_1} s_2^{\ell_2} \cdots s_N^{\ell_N} \approx \psi\big(\sum_{i=1}^N \varphi_i(s_i)\big)$, for arbitrary $a \in \mathbb{R}$ and $\ell_i \in \mathbb{Z}$, with $\varphi_i(s) = \log_{\mathrm{e}}\big(s^{\ell_i} + 1/p_0\big)$, $i = 1, \ldots, N$, and $\psi(y) = \exp_{\mathrm{e}}(ay)$.
- Maximum Value: $f(s_1, \ldots, s_N) = \max_{1 \leq i \leq N}\{s_i\} \approx \psi\big(\sum_{i=1}^N \varphi_i(s_i)\big)$, with $\varphi_i(s) = s^{p_0}$, $i = 1, \ldots, N$, and $\psi(y) = y^{\frac{1}{p_0}}$.[10]
- Minimum Value: $f(s_1, \ldots, s_N) = \min_{1 \leq i \leq N}\{s_i\} \approx \psi\big(\sum_{i=1}^N \varphi_i(s_i)\big)$, with $\varphi_i(s) = s^{-p_0}$, $i = 1, \ldots, N$, and $\psi(y) = y^{-\frac{1}{p_0}}$.

△

Even if the desired functions in Example 3.3 are not universal nomographic approximations, especially the maximum and minimum value can be relevant for alarm-driven sensor network applications. To make this more precise, consider a wireless sensor network that is used for fire detection in a factory by periodically computing the maximal temperature or carbon monoxide concentration at a FC and comparing the result with a predefined threshold. Due to safety reasons, this has certainly to be done with a minimum transmission delay, which can be achieved by letting all nodes transmit their pre-processed measurements simultaneously to approximate the maximum over the channel in the sense of Example 3.3. Since this requires only a single channel use, the delay is N-times smaller than that of a separation-based approach, in which all the sensor readings are communicated to the FC (see Section 2.3).

Remark 3.7. Note that the functions given in Examples 3.1 and 3.3 do not consist of universal pre-processing functions, from which it is obvious that such nomographic representations are not necessarily unique.

3.1.3 Computation With Multiple Channel Uses

In the previous two subsections, we have seen that the continuity of pre- and post-processing functions crucially impacts the space of computable functions. In particular, if the pre- and post-processing functions are required to be continuous, a single use of the ideal WMAC is not sufficient to compute every $f \in \mathscr{C}^0(\mathbb{E}^N)$. To get a more complete understanding of this behavior, we extend our considerations in this subsection to multiple channel uses (i.e., $n > 1$). This allows for further harnessing the interference property of the WMAC and simultaneously having more degrees of freedom.

[10] This is based on the well-known fact that $\forall (s_1, \ldots, s_N) \in \mathbb{E}^N : \lim_{p \to \infty} \big(\sum_{i=1}^N s_i^p\big)^{1/p} = \max_{1 \leq i \leq N} s_i$ uniformly.

Towards this end, let $n > 1$ and the computation transmitter (3.4) be modified to

$$\mathsf{Tx}_i(s_i) = (x_i[1], \ldots, x_i[n]) = (\varphi_{i1}(s_i), \ldots, \varphi_{in}(s_i)) \qquad i = 1, \ldots, N \;, \tag{3.15}$$

with $\varphi_{ij} \in \mathscr{C}^0(\mathbb{E})$, $j = 1, \ldots, n$, being some pre-processing functions to be specified later. Then, the corresponding length-n sequence of ideal WMAC outputs follows to

$$y[1] = \sum_{i=1}^{N} \varphi_{i1}(s_i), y[2] = \sum_{i=1}^{N} \varphi_{i2}(s_i), \ldots, y[n] = \sum_{i=1}^{N} \varphi_{in}(s_i) \;. \tag{3.16}$$

If the FC applies an individual post-processing function $\psi_j \in \mathscr{C}^0(\mathbb{R})$ to each of the outputs, a meaningful computation receiver can be

$$\mathsf{Rx}(y[1], \ldots, y[n]) = \Psi\Big(\psi_1(y[1]), \ldots, \psi_n(y[n])\Big) \;, \tag{3.17}$$

where the purpose of the continuous function $\Psi : \mathbb{R}^n \to \mathbb{R}$ is to appropriately combine the sequence of post-processed ideal WMAC outputs. Therefore, providing (3.15) and (3.17) with a proper choice of pre- and post-processing functions allows the FC to precisely compute every element of

$$\begin{aligned} \Big\{ f \in \mathscr{F}(\mathbb{E}^N) \,\Big|\, & \exists (\zeta_1, \ldots, \zeta_n) \in \mathscr{N}^0(\mathbb{E}^N) \times \cdots \times \mathscr{N}^0(\mathbb{E}^N) \, \exists \Psi \in \mathscr{C}^0(\mathbb{R}^n) \\ & \forall (s_1, \ldots, s_N) \in \mathbb{E}^N : f(s_1, \ldots, s_N) = \Psi\Big(\zeta_1(s_1, \ldots, s_N), \ldots, \zeta_n(s_1, \ldots, s_N)\Big) \Big\} \;. \end{aligned} \tag{3.18}$$

In order to characterize the rather abstract space (3.18) in greater detail by determining essential properties of its elements (e.g., continuity, differentiability), there are different approaches. For instance, we could fix the number of channel uses to some $n > 1$ and then try to construct elements with certain additional properties. On the other hand, we could also assume some fixed, and perhaps well-understood, function space (e.g., $\mathcal{C}^0(\mathbb{E}^N)$) and then try to figure out how to choose n (if feasible) such that every member of this space can be represented as on the right-hand side of (3.17).

It is somewhat surprising that the latter approach is closely related to the 13$^{\text{th}}$ of the famous list of 23 problems published by David Hilbert in 1900 [Hil00, Vit04].[11] The original problem involves the study of solutions of algebraic equations and Hilbert conjectured that a solution of the general equation of degree seven cannot be represented as a superposition of continuous functions of two variables. In his own words [Hil00, pp. 280–281]:

> ... Wahrscheinlich ist nun die Wurzel der Gleichung 7ten Grades eine solche Function ihrer Coefficienten, die nicht zu der genannten Klasse nomographisch construierbarer Functionen gehört, d. h. die sich nicht durch eine

[11] Hilbert presented a selection of the problems in a lecture held at the 1900 International Congress of Mathematicians in Paris [Rei96, p. 82].

> endliche Anzahl von Einschachtelungen von Functionen zweier Argumente erzeugen läßt. Um dieses einzusehen, wäre der Nachweis dafür nötig, *daß die Gleichung 7ten Grades*
>
> $$f^7 + xf^3 + yf^2 + zf + 1 = 0$$
>
> *nicht mit Hilfe beliebiger stetiger Functionen von nur zwei Argumenten lösbar ist.* Daß es überhaupt analytische Functionen von drei Argumenten x, y, z giebt, die nicht durch endlich-malige Verkettung von Functionen von nur zwei Argumenten erhalten werden können, davon habe ich mich, wie ich noch bemerken möchte, durch eine strenge Ueberlegung überzeugt.

In the jargon of wireless communications engineering, Hilbert's conjecture states that for every *finite* number of channel uses, n, the computation of all continuous desired functions by harnessing the interference property of the WMAC is not possible. Fortunately, the conjecture is disproven by Kolmogorov in his landmark paper [Kol57]. We use a remarkable refinement of Kolmogorov's result to state the following theorem, which gives, in analogy to Theorem 3.1, a complete answer to the question which functions are generally computable over the WMAC when pre- and post-processing functions are required to be continuous.

Theorem 3.4. *Let $N \geq 2$ be arbitrary. Then, every continuous desired function of N variables is universally computable over an ideal WMAC with $2N + 1$ channel uses, continuous pre- and post-processing functions, and zero distortion.*

Proof. The proof follows from [BG09], where it is constructively shown that every $f \in \mathscr{C}^0(\mathbb{E}^N)$ is representable as

$$f(s_1, \ldots, s_N) = \sum_{j=1}^{2N+1} \zeta_j(s_1, \ldots, s_N) \,, \tag{3.19}$$

with $\zeta_j \in \mathscr{N}^0(\mathbb{E}^N)$, $j = 1, \ldots, 2N + 1$, defined as

$$\zeta_j(s_1, \ldots, s_N) \coloneqq \psi_j \left(\sum_{i=1}^{N} \alpha_i \varphi(s_i + (j-1)\beta) \right) \,. \tag{3.20}$$

Here, $\varphi : \mathbb{E} \to \mathbb{R}$ is a continuous and monotone increasing function and $\{\alpha_i\}_{i=1}^N, \beta$ are appropriate nonnegative real constants. Only the post-processing functions $\psi_j \in \mathscr{C}^0(\mathbb{R})$, $j = 1, \ldots, 2N + 1$, depend on f, whereas the $N(2N + 1)$ continuous pre-processing functions $\varphi_{ij}(s_i) \coloneqq \alpha_i \varphi(s_i + (j-1)\beta)$ do not. □

Representation (3.19) reveals that Ψ in (3.17) can be chosen to be simply the sum over the post-processed ideal WMAC outputs,

$$\Psi\Big(\psi_1(y[1]), \ldots, \psi_{2N+1}(y[2N+1])\Big) = \sum_{j=1}^{2N+1} \psi_j(y[j]) \,,$$

so that the function space (3.18) is equal to $\mathscr{C}^0(\mathbb{E}^N)$.

Remark 3.8. Representations in the form of (3.19) along with (3.20) are called *Kolmogorov's superpositions.*

A geometric interpretation of Theorem 3.4 is the following [Lor66, p. 169], [Kha97, p. 13]. Using $2N+1$ distinct wireless resources (channel uses) results in a continuous and bijective[12] correspondence

$$(s_1,\ldots,s_N) \mapsto \begin{pmatrix} y[1] \\ y[2] \\ \vdots \\ y[2N+1] \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N} \alpha_i \varphi(s_i) \\ \sum_{i=1}^{N} \alpha_i \varphi(s_i+\beta) \\ \vdots \\ \sum_{i=1}^{N} \alpha_i \varphi(s_i+2N\beta) \end{pmatrix} \in \Gamma \,, \tag{3.21}$$

between sensor readings and ideal WMAC output-signals. As a result of the fact that $\mathbb{E}^N$ is a compact space, the range of this mapping, $\Gamma \subset \mathbb{R}^{2N+1}$, is compact as well. In other words, (3.21) describes a *homeomorphism* between $\mathbb{E}^N$ and Γ so that $\mathbb{E}^N$ is continuously embedded into the Euclidean space $\mathbb{R}^{2N+1}$.[13] Thus, there exists a bijection between all continuous functions $f(s_1,\ldots,s_N)$ defined on $\mathbb{E}^N$ and all continuous functions $F(y[1],\ldots,y[2N+1])$ defined on Γ (i.e., between $\mathscr{C}^0(\mathbb{E}^N)$ and $\mathscr{C}^0(\Gamma)$).

Due to the proof technique used by Kolmogorov in [Kol57] for solving David Hilbert's 13$^{\text{th}}$ problem, the statement of Theorem 3.4 that every continuous desired function can be computed with $2N+1$ channel uses by exploiting interference initially provides a sufficient condition on the number of channel uses. In algebraic topology it was previously known from the Menger-Nöbeling theorem [HW48, Thm. V2] that every compact space of finite dimension lower or equal to N is homeomorphic to a compact subset of $\mathbb{R}^{2N+1}$.[14] However, the fact that every space of dimension N can be topologically embedded into a compact space of dimension lower than $2N+1$ [HW48, Thm. V6] (e.g., $\mathbb{E}^N$ can be embedded into $\mathbb{R}^N$ by the identity map) suggests that the number of channel uses in Theorem 3.4 can perhaps be decreased to save wireless resources. Unfortunately, it is proven by Sternfeld in [Ste85] that this is not possible, which inevitably leads us to the following proposition.

Proposition 3.2. *Let the transmitters at sensor nodes be chosen as in (3.15) and the receiver at the FC as in (3.17), respectively, with pre- and post-processing functions required to be continuous. Then, in order to universally compute every $f \in \mathscr{C}^0(\mathbb{E}^N)$ at zero distortion by harnessing the interference, $n \geq 2N+1$ channel uses are necessary.*

[12] Note that according to Theorem 3.2, bijectivity is necessary and sufficient for universality.

[13] A function between two topological spaces is said to be a "homeomorphism" if it is continuous, bijective, and if its inverse is continuous as well [HW48, p. 159]. Two topological spaces are therefore "homeomorphic" if there exists a homeomorphism between them, which means, from a topological perspective, that they are the same. The surface of a sphere in three-dimensional Euclidean space, for instance, is homeomorph to the surface of a three-dimensional cube as they can be converted into each other by simply pushing and stretching.

[14] The "dimension" of an arbitrary topological space $\mathbb{X}$ is defined to be the least $m \in \mathbb{Z}$ such that every $x \in \mathbb{X}$ has arbitrary small neighborhoods whose boundaries are of dimension less than m [HW48, p. 24].

Remark 3.9. From [VK67], for instance, one can conclude that if we further restrict the pre-processing functions to be *continuously differentiable*, then Theorem 3.4 is no longer valid (i.e., $2N+1$ channel uses are not sufficient to universally compute every continuous desired function of N variables over the channel by harnessing interference). Roughly speaking, restricting the choice of pre- and post-processing functions costs wireless resources or rather reduces the space of functions that are computable by means of the wireless channel.

Although for $n \geq N$ there exist separation-based approaches (see Example 2.1) in order to appropriately compute desired functions over sensor networks (e.g., in an ideal TDMA protocol the entire analog sensor readings are conveyed interference free to the FC), using $n = 2N+1$ channel uses can lead to huge performance gains in more general network topologies, which is shown in the next subsection. Note for completeness that when dropping the demand for universality, $n = 2N+1$ can sometimes be reduced even if the function of interest is not in $\mathscr{N}^0(\mathbb{E}^N)$.

Example 3.4. Let $N = 2$ and consider the continuous desired function

$$f : \mathbb{E}^2 \to \mathbb{R} \ , \ (s_1, s_2) \mapsto s_1^2 + s_1 s_2 + s_2^2 + 2s_1 + s_2 \ .$$

Furthermore, let

$$U_\varepsilon \coloneqq \left\{ (s_1, s_2) \in \mathbb{E}^2 \,\middle|\, |s_1| \leq \varepsilon, |s_2| \leq \varepsilon \right\}$$

denote a compact neighborhood of the point $(0,0) \in \mathbb{E}^2$, for some $\varepsilon > 0$. Then, it can be shown that [Buc82]

$$\forall\, 0 < \varepsilon < 10^{-2} : \inf_{g \in \mathscr{N}^0(\mathbb{E}^2)} \|f - g\|_{U_\varepsilon} > \frac{\varepsilon^3}{10} \ ,$$

with $\|f\|_{U_\varepsilon} \coloneqq \max_{(s_1,s_2)\in U_\varepsilon} |f(s_1, s_2)|$. That is, there does not exist a nomographic function with continuous pre- and post-processing functions that approximates f with arbitrary precision. On the other hand, however, it can be easily seen that f can be represented as the superposition of two nomographic functions:

$$f(s_1, s_2) = f_1(s_1, s_2) + f_2(s_1, s_2) = \exp\left(\log(s_1 + 2) + \log(s_2 + 3)\right) + (s_1^2 - s_1 + s_2^2 - s_2 - 6) \ ,$$

where obviously $f_1, f_2 \in \mathscr{N}^0(\mathbb{E}^2)$. Thus, in this particular example, the sum of two nomographic functions suffices whereas the corresponding Kolmogorov's superposition would require to add up $2N + 1 = 5$ elements from $\mathscr{N}^0(\mathbb{E}^2)$. In simple terms, there is something between $n = 1$ and $n = 2N + 1$. △

3.2 Computation Over Clustered Networks

In our previous considerations, the spatially distributed sensor nodes communicated over a WMAC with a single FC so that we implicitly assumed a simple star network

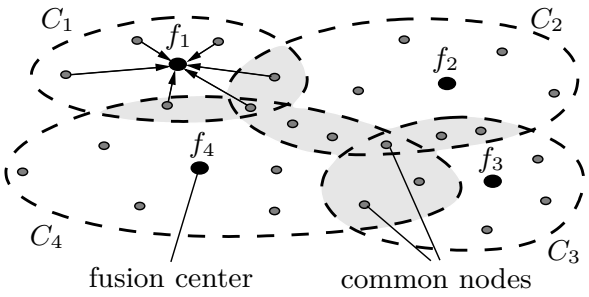


Figure 3.4: A clustered wireless sensor network consisting of $N = 25$ nodes and $L = 4$ clusters for computing some desired functions $f_1, \ldots, f_4$ at the FCs. Nodes belonging to some overlap $C_\ell \cap C_m$, $\ell \neq m$, are called common nodes.

topology. If the distances are large, however, this can be comparatively costly in terms of energy consumption as the nodes have to increase their transmit powers to compensate the higher path attenuation. In such cases, re-organizing the network into clusters turns out to be a reasonable approach to prolong the network lifetime (in some sense) while maintaining connectivity [HCB02, YF04, YKR06]. On the other hand, one could think of scenarios in which different FCs with individual tasks aim at exploiting the observations of all the nodes in the network but are limited to local neighborhoods (i.e., node clusters) due to reachability constraints. Therefore, we extend our considerations in this section to a more general network topology and assume that the network is organized into $L \in \mathbb{N}$ clusters, where the set of nodes belonging to cluster ℓ is denoted by C_ℓ, $\ell = 1, \ldots, L$. In particular, we focus on those clustered networks in which for each ℓ there exists at least one $m \neq \ell$ such that $C_\ell \cap C_m \neq \varnothing$. Furthermore, each cluster is assumed to have a designated FC that acts as the cluster head (see Figure 3.4 for a qualitative example).

In order to describe the intra-cluster communication, we continue to use the model of a real-valued ideal WMAC so that the symbol received by FC ℓ at channel use j can be written as

$$y_\ell[j] = \sum_{i \in C_\ell} x_i[j] \qquad \ell = 1, \ldots, L \, . \tag{3.22}$$

Instead of reconstructing individual sensor readings from the sequence of channel outputs (3.22), each FC, say FC ℓ, aims at reliably and efficiently computing some desired function

$$f_\ell : \mathbb{E}^{|C_\ell|} \to \mathbb{R} \, , \; (s_{\ell_1}, \ldots, s_{\ell_{|C_\ell|}}) \mapsto f_\ell(s_{\ell_1}, \ldots, s_{\ell_{|C_\ell|}}) \tag{3.23}$$

thereof, where $|C_\ell|$ denotes the number of nodes in cluster ℓ (i.e., $\bigcup_{\ell=1}^{L} |C_\ell| = N$). This is obviously equivalent to the problem of efficiently computing the vector-valued

function

$$\boldsymbol{f} : \mathbb{E}^N \to \mathbb{R}^L \ , \ (s_1, \ldots, s_N) \mapsto \boldsymbol{f}(s_1, \ldots, s_N) = \begin{pmatrix} f_1\big(s_{1_1}, \ldots, s_{1_{|C_1|}}\big) \\ f_2\big(s_{2_1}, \ldots, s_{2_{|C_2|}}\big) \\ \vdots \\ f_L\big(s_{L_1}, \ldots, s_{L_{|C_L|}}\big) \end{pmatrix} , \qquad (3.24)$$

over a sensor network that is organized into L clusters. In order to treat this problem, we start in Section 3.2.1 with the case where only continuous pre- and post-processing functions are allowed whereas in Section 3.2.2 no restrictions on pre- and post-processing functions are imposed.

Remark 3.10. The main difference to the previously studied problem is that the ideal WMACs (3.22) interfere with each other in the sense that the common nodes (see Figure 3.4) can be heard by more than one FC.

3.2.1 Continuous Pre- and Post-Processing Functions

From Proposition 3.2, we recap that in order to universally compute every continuous function of $N \geq 2$ sensor readings by means of interference and with continuous pre- and post-processing functions, at least $n = 2N + 1$ channel uses are necessary. The reason is that the N-dimensional space of sensor readings, $\mathbb{E}^N$, has to be homeomorphically mapped onto a compact set Γ of dimension $2N + 1$ (see (3.21)). Since this can be achieved by using the ideal WMAC $2N + 1$ times, we interpret it as collecting the dimensions of Γ via distinct wireless resources. If the FC, however, can only receive signals from a subset of the N nodes (i.e., some summands on the right hand side of (3.21) are missing), then the image (3.21) is *not necessarily* in Γ, which in turn implies that not every $F \in \mathscr{C}^0(\Gamma)$ is computable. This is exactly what happens in a clustered network, in which the nodes cannot reach all FCs so that each component of $\boldsymbol{f}$ depends only on a subset of the nodes. See Figure 3.4 for an illustration.

Fortunately, due to the structural properties figured out in Section 3.1.3, this can be easily and independently resolved at each FC. To demonstrate this, let φ_{ij}, $i = 1, \ldots, N$; $j = 1, \ldots, 2N + 1$, be $N(2N + 1)$ continuous functions such that according to the proof of Theorem 3.4 every continuous $f : \mathbb{E}^N \to \mathbb{R}$ can be represented as

$$f(s_1, \ldots, s_N) = \sum_{j=1}^{2N+1} \psi_j \left(\sum_{i=1}^{N} \varphi_{ij}(s_i) \right) = \sum_{j=1}^{2N+1} \psi_j \left(\sum_{i=1}^{N} \alpha_i \varphi(s_i + (j-1)\beta) \right) \qquad (3.25)$$

through a proper choice of the continuous functions $\psi_1, \ldots, \psi_{2N+1}$. Suppose that each node in the network is uniquely assigned one of the function sets $\{\varphi_{ij} \in \mathscr{C}^0(\mathbb{E})\}_{j=1}^{2N+1}$, $i = 1, \ldots, N$, to generate the individual transmitters (3.15). Now, summarize the ideal

WMAC outputs (3.22) to the vectors

$$\boldsymbol{y}_\ell = \begin{pmatrix} y_\ell[1] \\ y_\ell[2] \\ \vdots \\ y_\ell[2N+1] \end{pmatrix} = \begin{pmatrix} \sum_{i\in C_\ell} x_i[1] \\ \sum_{i\in C_\ell} x_i[2] \\ \vdots \\ \sum_{i\in C_\ell} x_i[2N+1] \end{pmatrix} = \begin{pmatrix} \sum_{i\in C_\ell} \alpha_i\varphi(s_i) \\ \sum_{i\in C_\ell} \alpha_i\varphi(s_i+\beta) \\ \vdots \\ \sum_{i\in C_\ell} \alpha_i\varphi(s_i+2N\beta) \end{pmatrix} ,$$

$\ell = 1,\dots,L$, and note that they are generally *not contained in* Γ. However, if we consider the shifted versions

$$\boldsymbol{z}_\ell \coloneqq \boldsymbol{y}_\ell + \boldsymbol{\gamma}_\ell \, , \tag{3.26}$$

with

$$\boldsymbol{\gamma}_\ell = \begin{pmatrix} \gamma_\ell[1] \\ \gamma_\ell[2] \\ \vdots \\ \gamma_\ell[2N+1] \end{pmatrix} \coloneqq \begin{pmatrix} \sum_{i\notin C_\ell} \alpha_i\varphi(0) \\ \sum_{i\notin C_\ell} \alpha_i\varphi(\beta) \\ \vdots \\ \sum_{i\notin C_\ell} \alpha_i\varphi(2N\beta) \end{pmatrix} , \tag{3.27}$$

then unlike the original $\boldsymbol{y}_\ell$, the shifted $\boldsymbol{z}_\ell = (z_\ell[1],\dots,z_\ell[2N+1])$ is always contained in Γ, for all $\ell = 1,\dots,L$.

Remark 3.11. It is important to realize that $\boldsymbol{\gamma}_\ell \in \mathbb{R}^{2N+1}$ is, for each ℓ, a constant and therefore independent of the sensor readings.

In contrast to (3.25), however, none of the desired functions in (3.23) depends on all N sensor readings. But this is not a limitation since we are able to conclude from (3.26) and (3.27) a simple post-processing strategy described as follows. After receiving at channel use j, $j = 1,\dots,2N+1$, the symbol $y_\ell[j]$, FC ℓ adds the correction term $\gamma_\ell[j]$, applies an appropriate post-processing function $\psi_{\ell j} \in \mathscr{C}^0(\mathbb{R})$, that is,

$$\psi_{\ell j}(z_\ell[j]) = \psi_{\ell j}(y_\ell[j]+\gamma_\ell[j]) = \psi_{\ell j}\left(\sum_{i\in C_\ell} \alpha_i\varphi(s_i+(j-1)\beta) + \gamma_\ell[j]\right) \, , \tag{3.28}$$

and stores this intermediate result in a buffer. Then, if all $2N+1$ channel output-symbols are received and post-processed, the FC finally sums up the respective buffer content, which results in the function-value

$$f_\ell(s_{\ell_1},\dots,s_{\ell_{|C_\ell|}}) = \sum_{j=1}^{2N+1} \psi_{\ell j}\left(\sum_{i\in C_\ell} \underbrace{\alpha_i\varphi(s_i+(j-1)\beta)}_{=\varphi_{ij}(s_i)} + \gamma_\ell[j]\right) . \tag{3.29}$$

According to this, the receiver structure of FC ℓ is given by

$$\mathsf{Rx}_\ell(\boldsymbol{y}_\ell) = \sum_{j=1}^{2N+1} \psi_{\ell j}(y_\ell[j]+\gamma_\ell[j]) \qquad \ell = 1,\dots,L \tag{3.30}$$

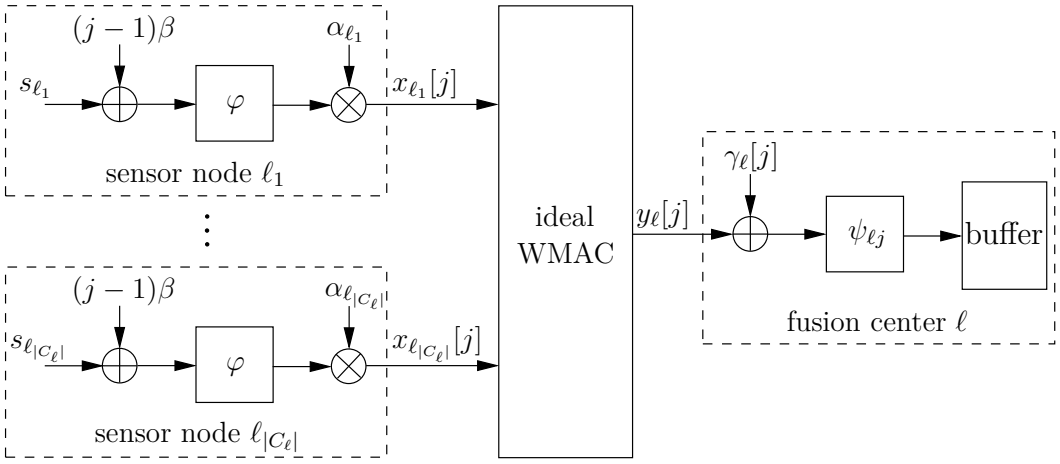


Figure 3.5: Block diagram of computations in cluster ℓ, $\ell = 1, \ldots, L$, at channel use j, $j = 1, \ldots, 2N+1$. The cluster consists of $|C_\ell|$ transmitting sensor nodes and a FC. After $2N+1$ transmissions, the FC sums up the buffered post-processed receive-signals, which results in the desired function value $f_\ell(s_{\ell_1}, \ldots, s_{\ell_{|C_\ell|}})$.

from which we infer for the representation of the desired vector-valued function (3.24)

$$\boldsymbol{f}(s_1, \ldots, s_N) = \begin{pmatrix} \mathsf{Rx}_1(\boldsymbol{y}_1) \\ \mathsf{Rx}_2(\boldsymbol{y}_2) \\ \vdots \\ \mathsf{Rx}_L(\boldsymbol{y}_L) \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{2N+1} \psi_{1j}(z_1[j]) \\ \sum_{j=1}^{2N+1} \psi_{2j}(z_2[j]) \\ \vdots \\ \sum_{j=1}^{2N+1} \psi_{Lj}(z_L[j]) \end{pmatrix} . \tag{3.31}$$

A corresponding block diagram that illustrates the computation in cluster ℓ for some channel use j is depicted in Figure 3.5.

Remark 3.12. Since the pre-processing functions are independent of the components of $\boldsymbol{f}$, the FCs determine by appropriately choosing the $L(2N+1)$ post-processing functions $\{\psi_{\ell j}\}$ which continuous functions $f_1, \ldots, f_L$ are to be universally computed.

Remark 3.13. Note that the constants α_i in (3.28) need not to be different for all N nodes in the network so that they can be reused in different clusters without any kind of arrangement between them.

We can conclude that with the post-processing described above, all properties of Section 3.1.3 carry over to *arbitrary* clustered sensor networks, which we summarize in the following theorem:

Theorem 3.5. *Let $N \geq 2$ be the total number of nodes in a sensor network that is organized into $L \in \mathbb{N}$ node clusters $C_1, \ldots, C_L$. Suppose that*

$$\forall \ell \in \{1, \ldots, L\}\, \exists m \neq \ell : C_\ell \cap C_m \neq \varnothing$$

and that the intra-cluster communication takes place over ideal WMACs (3.22). Then, by harnessing interference, every vector-valued function $\boldsymbol{f} : \mathbb{E}^N \to \mathbb{R}^L$ *that is composed of functions* $f_\ell \in \mathscr{C}^0(\mathbb{E}^{|C_\ell|})$, $\ell = 1, \ldots, L$, *can be computed with* $2N + 1$ *channel uses, continuous pre- and post-processing functions, and zero distortion.*

3.2.2 Arbitrary Pre- and Post-Processing Functions

If no restrictions on pre- and post-processing functions are imposed, the situation is much less complicated than in the last subsection, since from Theorem 3.1 we conclude that already a *single* concurrent transmission of the nodes in each cluster is sufficient to *universally* compute every $\boldsymbol{f} \in \mathscr{F}(\mathbb{E}^{|C_1|}) \times \cdots \times \mathscr{F}(\mathbb{E}^{|C_L|})$. In order to see this, consider the L coupled ideal WMAC output-symbols

$$y_\ell[1] = g_\ell(s_{\ell_1}, \ldots, s_{\ell_{|C_\ell|}}) = \sum_{i \in C_\ell} \varphi_i(s_i) \tag{3.32}$$

received by the FCs, which are mappings $g_\ell : \mathbb{E}^{|C_\ell|} \to \Upsilon_\ell \subset \mathbb{R}$, $\ell = 1, \ldots, L$. Then, the post-processing at the FCs merely consists in the application of appropriate post-processing functions such that every $\boldsymbol{f} \in \mathscr{F}(\mathbb{E}^{|C_1|}) \times \cdots \times \mathscr{F}(\mathbb{E}^{|C_L|})$ can be represented as

$$\boldsymbol{f}(s_1, \ldots, s_N) = \begin{pmatrix} \psi_1(y_1[1]) \\ \psi_2(y_2[1]) \\ \vdots \\ \psi_L(y_L[1]) \end{pmatrix} . \tag{3.33}$$

What, however, is the difference to the previous case where continuous pre- and post-processing functions are required?

We conclude from Theorem 3.2 that a necessary and sufficient condition to universally compute every vector-valued function (3.33) is that the functions g_ℓ, $\ell = 1, \ldots, L$, defined in (3.32), are bijective. Since this can never be achieved for every $\boldsymbol{f}$ with continuous pre- and post-processing functions, it was necessary in Section 3.1.3 to appropriately embed $\mathbb{E}^N$ into a higher dimensional space, resulting in a bijection between $\mathscr{C}^0(\mathbb{E}^N)$ and $\mathscr{C}^0(\Gamma)$ instead. If pre- and post-processing functions are allowed to be discontinuous, however, such an embedding is dispensable.

3.2.3 Performance Comparison

In order to evaluate the properties of wireless computation networks that harness interference rather than avoiding it, in this section we highlight the advantages of the computation approach depicted in Figure 3.5 over standard TDMA protocols.

If no restrictions on pre- and post-processing functions are imposed, we conclude from Theorem 3.1 as well as from Section 3.2.2 that in fact *every* function on each FC can be universally computed without significant coordination. This can be achieved by harnessing the natural interference property of the wireless channel (i.e., without

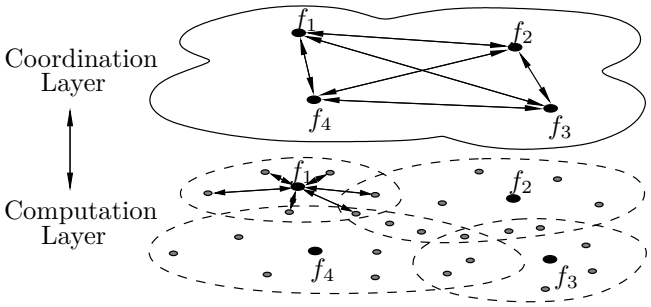


Figure 3.6: The clustered sensor network example from Figure 3.4 with an additional coordination layer that coordinates the medium-access of clusters such as in a standard TDMA approach. This requires a bidirectional interface between both layers as well as wireless communication links between the nodes and the FC in each cluster (represented by arrows with two heads).

interference avoidance). The required number of channel uses is of the order $\mathcal{O}(1)$ and therefore independent of the number of nodes and clusters. Alternatively, when a standard TDMA protocol is employed to compute functions at FCs in a clustered wireless sensor network, besides the orthogonalized medium-access of the nodes in each cluster, the clusters themselves have to be appropriately separated in time, which requires a significant amount of coordination (see Figure 3.6 for an illustration). A naive TDMA protocol would therefore need at least $L \max_{1\leq\ell\leq L} |C_\ell|$ separated transmissions to convey the entire raw sensor readings *interference-free* to the L FCs, which subsequently compute the desired functions $f_1, \ldots, f_L$. Thus, it requires $\mathcal{O}(L \max_{1\leq\ell\leq L} |C_\ell|)$ channel uses. In contrast, we conclude from Section 3.2.1 that $\mathcal{O}(2N+1)$ channel uses are sufficient for computing every $\boldsymbol{f} \in \mathscr{C}^0(\mathbb{E}^{|C_1|}) \times \cdots \times \mathscr{C}^0(\mathbb{E}^{|C_L|})$ if pre- and post-processing functions are required to be continuous. Obviously, the number $2N+1$ does not scale directly with the number of clusters, L, so that significant performance gains are possible for $L > 1$. Moreover, an additional coordination layer as in Figure 3.6 is not needed since all clusters can transmit *concurrently*.

The attentive reader might note that according to Theorem 3.4, already $\mathcal{O}(2|C_\ell|+1)$ channel uses are sufficient to universally compute in each cluster, say cluster ℓ, every $f_\ell \in \mathscr{C}^0(\mathbb{E}^{|C_\ell|})$ by harnessing interference. Since

$$\max_{1\leq\ell\leq L} \{2|C_\ell|+1\} \leq 2N+1$$

holds for arbitrary clustered networks, seemingly further channel uses could be saved. Because of the couplings between clusters due to the common nodes, however, this would require a constant adaptation of the pre-processing on the common sensor nodes and therefore a large amount of coordination. In order to illustrate this, recall from the geometric interpretation of Theorem 3.4 that there exist homeomorphisms

$$(s_{\ell_1}, \ldots, s_{\ell_{|C_\ell|}}) \mapsto (y_\ell[1], \ldots, y_\ell[2|C_\ell|+1])$$

between $\mathbb{E}^{|C_\ell|}$ and $\Gamma_\ell \subset \mathbb{R}^{2|C_\ell|+1}$ that enable each cluster to compute every $f_\ell \in \mathscr{C}^0(\mathbb{E}^{|C_\ell|})$ because of the existence of Kolmogorov's superpositions

$$f_\ell(s_{\ell_1}, \ldots, s_{\ell_{|C_\ell|}}) = \sum_{j=1}^{2|C_\ell|+1} \psi_{\ell j} \left(\sum_{i \in C_\ell} \varphi_{ij}^{(\ell)}(s_i) \right) \qquad \ell = 1, \ldots, L \,. \tag{3.34}$$

Unfortunately, since the compact sets Γ_ℓ will differ in general, the pre-processing functions in (3.34) also depend on ℓ. For nodes whose transmit signals can only be received by a single FC it does not matter. Nodes that can be heard by more than one FC (i.e., the common nodes between clusters), however, have to adapt their pre-processing functions in dependency of the FC that they want to address. Such as in the case of standard TDMA, this in turn would require the separated activation of clusters (see Figure 3.6) so that in total $\mathcal{O}(\sum_{\ell=1}^{L}(2|C_\ell|+1))$ channel uses are required.

Remark 3.14. Note that the TDMA protocol considered above is said to be "standard" in the sense that it takes no advantage of the underlying network topology. If some of the clusters are disjoint (see Figure 3.4 for an example), for instance, it is clear that the number of channel uses could be reduced by employing a TDMA protocol that activates disjoint clusters concurrently. This, however, would further increase the coordination effort.

3.2.4 A Short Note About Additional Interference

At the beginning of Section 3.2 one particular motivation for the transmission model (3.22) was that due to the connectivity radii of sensor nodes, the overlap between clusters is determined by the spatial position of FCs. Thus, the FCs compute functions of subsets of freely accessible measurements. The results of Section 3.2, however, also remain valid in a scenario in which distinct computation sensor networks interfere with each other. In order to illustrate this, consider, without loss of generality, the example depicted in Figure 3.7, which consists of two interfering sensor networks deployed to compute the individual desired functions $f_1 : \mathbb{E}^{|C_1|} \to \mathbb{R}$ and $f_2 : \mathbb{E}^{|C_2|} \to \mathbb{R}$ at the two FCs. More precisely, let C_1 and C_2 denote the finite sets of nodes belonging to network one and two, respectively, and let the corresponding sensor readings be summarized in the vectors $\boldsymbol{s}_1 = (s_{1_1}, \ldots, s_{1_{|C_1|}}) \in \mathbb{E}^{|C_1|}$, $\boldsymbol{s}_2 = (s_{2_1}, \ldots, s_{2_{|C_2|}}) \in \mathbb{E}^{|C_2|}$. The difference to our previous studies (cf. (3.32)) is that the overlap region denoted as C_0 contains nodes from both networks so that uncoordinated transmissions may result in the symbol

$$y_\ell[1] = \sum_{i \in C_\ell} \varphi_i(s_i) + \underbrace{\sum_{\substack{i \in C_k \cap C_0 \\ k \neq \ell}} \varphi_i(s_i)}_{\text{interference}} \,, \tag{3.35}$$

received at FC ℓ, $\ell \in \{1, 2\}$.

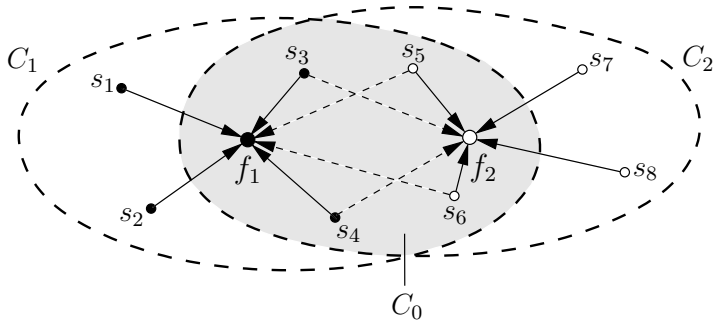


Figure 3.7: Two interfering (i.e., overlapping) wireless sensor networks. Sets C_1 and C_2 summarize the nodes belonging to network one (black dots) and network two (white dots), respectively, whereas the dashed regions are the connectivity radii of the corresponding FCs. The signals transmitted by nodes in the shaded overlap C_0 are received at both FCs and therefore generate additional (i.e., unwanted) interference.

First, it seems that the mutual interference prevents the adequate computation of arbitrary functions

$$f_1(\boldsymbol{s}_1) = \psi_1\left(\sum_{i\in C_1}\varphi_i(s_i)\right) \quad , \quad f_2(\boldsymbol{s}_2) = \psi_2\left(\sum_{i\in C_2}\varphi_i(s_i)\right) \ .$$

However, if the first FC has knowledge about the interfering nodes in $C_2 \cap C_0$ (i.e., about their pre-processing) and the other FC about the nodes belonging to $C_1 \cap C_0$, respectively, the problem is equivalent to the problem of computing functions of fewer variables as available. This knowledge provided, there always exist post-processing functions $\tilde{\psi}_1, \tilde{\psi}_2 \in \mathscr{F}(\mathbb{R})$ such that $\tilde{\psi}_1(y_1[1]) = f_1(s_{1_1}, \ldots, s_{1_{|C_1|}})$, for all $\boldsymbol{s}_1 \in \mathbb{E}^{|C_1|}$ and all $\{s_i \in \mathbb{E}\}_{i\in C_2\cap C_0}$ as well as $\tilde{\psi}_2(y_2[1]) = f_2(s_{2_1}, \ldots, s_{2_{|C_2|}})$, for all $\boldsymbol{s}_2 \in \mathbb{E}^{|C_2|}$ and all $\{s_i \in \mathbb{E}\}_{i\in C_1\cap C_0}$. From Theorem 3.2, we already know that to achieve this, the pre-processing functions have to be chosen in both networks such that every choice of sensor readings $\boldsymbol{s}_\ell^{(1)}, \boldsymbol{s}_\ell^{(2)} \in \mathbb{E}^{|C_\ell|}$, with $\boldsymbol{s}_\ell^{(1)} \neq \boldsymbol{s}_\ell^{(2)}$, leads to separated receive signals $y_\ell^{(1)}[1] \neq y_\ell^{(2)}[1]$, $\ell = 1, 2$.

Remark 3.15. If it is not possible to provide knowledge about the nodes of interfering sensor systems to the respective FCs, the unwanted part of the interference in (3.35) has to be treated as an additional noise term.

3.3 Robustness Against Changes in Topology

In Theorem 3.1, we have shown that every $f \in \mathscr{F}(\mathbb{E}^N)$ is computable over an ideal WMAC with a single channel use. The surprising fact is that the computations are

universal (see Definition 3.3) and thus the pre-processing functions are independent of the choice of f. However, the corresponding proof in Appendix 3.A.1 depends on the number N of nodes in the network. As a consequence, transmitting sensors nodes have to adapt their pre-processing functions if the network topology changes (i.e., the universality is *not robust* against modified N), which would be highly undesired in sensor networks. Hence, we want to know if this holds in general. Therefore, we analyze in this section the robustness of universality against variations in network topology due to sensor nodes that drop out of the network or new sensor nodes that join the network.

3.3.1 Dropped Out Nodes

Let us first consider the case in which a number of sensor nodes drops out of the network due to, for instance, failures or battery depletion. The question is whether the universality property is preserved when an arbitrary subset of nodes leaves the network. The following theorem provides the answer:

Theorem 3.6. *The universality of computing arbitrary desired functions over an ideal WMAC with a single channel use is* robust *against dropped nodes. That is, there exists a set of fixed pre-processing functions* $\{\varphi_1, \ldots, \varphi_N\}$ *that is universal with respect to* $\mathscr{F}(\mathbb{E}^N)$ *such that the subset* $\{\varphi_1, \ldots, \varphi_m\}$ *is universal with respect to* $\mathscr{F}(\mathbb{E}^m)$*, for all* $m < N$*.*

Proof. The proof is postponed to Appendix 3.A.3 at the end of the chapter. □

The Theorem implies the fact that it is not necessary to update the remaining active sensor nodes (i.e., the pre-processing functions) if some nodes drop out of the network, which has the potential to significantly improve the flexibility of computation networks.

Remark 3.16. It should be emphasized that even if according to Theorem 3.6 the pre-processing functions do not depend on the desired function and the number of active nodes, the post-processing function does. In other words, if nodes drop out of the network, the pre-processing functions at the remaining nodes remain the same whereas the post-processing functions have to be updated by the FCs.

3.3.2 Additional Nodes

We now consider the opposite case where an existing sensor network for computation purposes is enlarged by adding a finite number of active nodes. More precisely, assume that we connect $J - N \in \mathbb{N}$, $N < J < \infty$, transmitting sensor nodes to the network in order to universally compute every desired function

$$f : \mathbb{E}^J \to \mathbb{R}\ ,\ (s_1, \ldots, s_N, \ldots, s_J) \mapsto f(s_1, \ldots, s_J)\ ,$$

of the measurements. Then, the following theorem answers the question whether the universality property is preserved if the existing N active nodes were already able to universally compute every $f \in \mathscr{F}(\mathbb{E}^N)$.

Theorem 3.7. *The universality of computing arbitrary desired functions over an ideal WMAC with a single channel us is* robust *against a fixed enlargement of the network.*

Proof. The proof is deferred to Appendix 3.A.4 at the end of the chapter. □

In accordance to the theorem, it is not necessary to update the existing transmitting nodes (i.e., the pre-processing) if the network is enlarged by adding further active transmitting sensor nodes.

Note that the term "fixed" in Theorem 3.7 as well as the idea of the proof refers to the fact that in the current form, the robustness of universality holds if the original network was already designed for $J > N$ nodes but only N nodes are deployed to the measuring field. Then, adding up to $J - N$ nodes during network operation has no impact on the previous N nodes. The more general case in which the original network was designed for at most N nodes but extended to J nodes afterwards is therefore still an open problem.

We would like to emphasize, however, that this limits the practical significance of Theorem 3.7 only marginally. A robust network for computation purposes can always be designed without knowing the exact number $J-N$ in advance by choosing J sufficiently large (according to the application needs) and using only N out of J nodes in practice.

Remark 3.17. Such as in the case where nodes drop out of the network, only the post-processing function has to be updated if new nodes are added to the network.

Remark 3.18. Note that even if Theorems 3.6 and 3.7 refer to single-cluster networks, they remain valid for arbitrary clustered networks. The only small difference is that if *common* nodes drop out of the network, then all affected FCs have to update their post-processing functions. On the other hand, when continuous pre- and post-processing functions are employed in order to compute some continuous desired function with $2N+1$ channel uses, the FCs have to additionally adjust the correction terms (3.27) by appropriately adding further constants that correspond to the dropped out nodes.

3.4 Summary and Conclusions

In this chapter, we studied the problem of computing functions at FCs in clustered wireless sensor networks where nodes transmit simultaneously to harness the interference property of the wireless channel. By applying appropriate pre-processing functions on sensor readings and post-processing functions on the superimposed signals received by the FCs, in addition to linear functions even nonlinear functions are computable over the channel.

If no restrictions on pre- and post-processing functions are imposed, we have shown that in fact *every* function can independently be computed at each FC, where the number of required channel uses is of the order $\mathcal{O}(1)$. A standard TDMA, however, requires $\mathcal{O}(L \max_{1\leq\ell\leq L} |C_\ell|)$ channel uses. The latter scales with the number L of clusters and

the number of nodes belonging to the largest cluster so that huge performance gains are possible when performing computations over the channel instead.

Although implementing continuous pre- and post-processing functions in practice is generally less complex, a corresponding restriction generates the need for an additional post-processing step at FCs to ensure the computability of at least *every continuous* function of sensor readings. However, requiring pre- and post-processing functions to be continuous generally needs additional channel uses. In particular, we have shown that the number of required channel uses in order to simultaneously compute some continuous functions at distinct FCs is then of the order $\mathcal{O}(2N+1)$. Since this is proportional to the number N of nodes but not to the number of clusters, the proposed computation scheme still offers significant performance gains in comparison to a standard TDMA.

A remarkable property of the considered approach is that the computations can always be universally performed. Here, "universal" means that the pre-processing functions are independent of the functions to be computed at the FCs so that they do not need to be updated if these functions change. This implies that the feedback overhead between nodes and FCs can be reduced since corresponding coordination is not needed. Therefore, the architecture of sensor nodes for computation purposes is universal and of reduced complexity, which makes them cheap and easy to handle. In this regard, we have shown that the universality property is preserved even if the network topology varies because of nodes that leave or enter the network. The results of this chapter form the basis for the computation schemes presented in the upcoming two chapters.

An interesting open problem results from our considerations in Section 3.1.2, which can be stated as follows: Given a distortion measure $d_{\mathbb{R}}$, some fixed accuracy $\varepsilon > 0$ and some desired function $f \in \mathscr{C}^0(\mathbb{E}^N)$. Then, find the best *nomographic approximation* $\hat{f}^\star \in \mathscr{N}^0_{d,\varepsilon}(\mathbb{E}^N)$ to f (if existent), that is,

$$\hat{f}^\star \in \operatorname*{argmin}_{\hat{f} \in \mathscr{N}^0_{d,\varepsilon}(\mathbb{E}^N)} d_{\mathbb{R}}(f, \hat{f}) \; .$$

The sequence (3.13), for instance, solves this problem for f being the "geometric mean" and $d_{\mathbb{R}}$ the supremum norm (3.14). Such solutions have the charm that continuous desired functions that normally require at least N channel uses for a distortion free computation can be approximated with merely a single channel use by harnessing the interference (see Proposition 3.1).

Another open problem is to find a proof to Theorem 3.7 that is more general. The current proof unfortunately requires to start with an arbitrary large $J > N$ in order to guarantee universality when up to $J - N$ nodes are added to the network.

Appendix 3.A Proofs

3.A.1 Proof of Theorem 3.1

To prove the theorem, we have to show that

$$\exists(\varphi_1,\ldots,\varphi_N)\in\mathscr{F}(\mathbb{E})\times\cdots\times\mathscr{F}(\mathbb{E})\,\forall f\in\mathscr{F}(\mathbb{E}^N)\,\exists\psi\in\mathscr{F}(\mathbb{R}):f\in\mathscr{N}(\mathbb{E}^N)\;.$$

Towards this end, we first have to prove that there exists a fixed set of pre-processing functions that are universal with respect to $\mathscr{F}(\mathbb{E}^N)$ (see Definition 3.3). That is, there exist $\varphi_1,\ldots,\varphi_N$ such that the function

$$g:\mathbb{E}^N\to\mathbb{R}\;,\;(s_1,\ldots,s_N)\mapsto g(s_1,\ldots,s_N)=\sum_{i=1}^{N}\varphi_i(s_i) \tag{3.36}$$

fulfills

$$\forall(s_1,\ldots,s_N)\neq(s'_1,\ldots,s'_N):g(s_1,\ldots,s_N)\neq g(s'_1,\ldots,s'_N) \tag{3.37}$$

and therefore separates all points of $\mathbb{E}^N$. To achieve this, we use the following lemma from [Kul08, Thm. 5.2].

Lemma 3.1. *Let $q\geq 2$ be some natural number. Then, every $s\in\mathbb{R}$ is uniquely represented by a q-adic expansion*

$$s=(-1)^{\eta}\sum_{k=-v}^{\infty}\frac{w_k}{q^k}\;, \tag{3.38}$$

where $\eta\in\{0,1\}$, $w_k\in\{0,\ldots,q-1\}$, and $w_k\neq q-1$ for infinitely many k.

Since the lemma states that there is a bijection between $\mathbb{R}$ and (3.38), we conclude that there is, for every fixed $N\geq 2$ and each i, $i=1,\ldots,N$, a bijection between the set of sensor readings $\mathbb{E}$ and the infinite series

$$s_i=\sum_{k=1}^{\infty}w_{ik}N^{-k}\qquad i=1,\ldots,N \tag{3.39}$$

as long as $w_{ik}\neq N-1$ for infinitely many k, unless $w_{ik}=N-1$ for all k.

With this in mind, let us define the pre-processing functions as

$$s_i\mapsto\varphi_i(s_i)=N^{1-i}\sum_{k=1}^{\infty}w_{ik}N^{-Nk}\qquad i=1,\ldots,N\;, \tag{3.40}$$

which is nothing else than representing elements in $\mathbb{E}$ to base N^N. Since every sequence $\{w_{ik}\}_{k\in\mathbb{N}}$ uniquely determines both s_i and $\varphi_i(s_i)$, we have that (3.40) is bijective, for

all $i = 1, \ldots, N$. With this choice of pre-processing functions, it follows for g in (3.36)

$$\begin{aligned} g(s_1, \ldots, s_N) &= \sum_{i=1}^{N} N^{1-i} \sum_{k=1}^{\infty} w_{ik} N^{-Nk} \\ &= N^{1-N} \sum_{k=1}^{\infty} \underbrace{\sum_{i=1}^{N} w_{ik} N^{N-i}}_{=:v_k} N^{-Nk} \, . \end{aligned} \tag{3.41}$$

Now, we have to show that (3.41) is a *unique* representation of real numbers to base N^N as well. This requires to prove that the condition $w_{ik} \neq N-1$ for infinitely many k, unless $w_{ik} = N-1$ for all k, implies $v_k \neq N^N - 1$ for infinitely many k, unless $v_k = N^N - 1$ for all k. Accordingly, we conclude from the definition of v_k in (3.41)

$$\begin{aligned} 0 \leq \sum_{i=1}^{N} w_{ik} N^{N-i} &\leq (N-1) \sum_{i=1}^{N} N^{N-i} && (3.42) \\ &= (N-1) N^{N-1} \sum_{i=0}^{N-1} \left(\frac{1}{N} \right)^i && (3.43) \\ &= \left(N^N - N^{N-1} \right) \cdot \frac{1 - \left(\frac{1}{N} \right)^N}{1 - \frac{1}{N}} && (3.44) \\ &= N^N - 1 \, , \end{aligned}$$

where the right-hand side of (3.42) follows from $w_{ik} \in \{0, \ldots, N-1\}$, for all i and k, and (3.44) from the fact that the sum in (3.43) is a partial sum of a geometric series. Since $w_{ik} \neq N-1$ for infinitely many k, we have

$$0 \leq v_k < N^N - 1 \tag{3.45}$$

for infinitely many k, except when $w_{ik} = N-1$ for all i and k, so that the infinite series (3.41) indeed uniquely represents real numbers to base N^N. The next step in proving the theorem is therefore to show that g as defined in (3.41) is *bijective* (i.e., g fulfills (3.37)).

In order to fulfill (3.37), the equality

$$g(s_1, \ldots, s_N) = N^{1-N} \sum_{k=1}^{\infty} v_k N^{-Nk} = N^{1-N} \sum_{k=1}^{\infty} v'_k N^{-Nk} = g(s'_1, \ldots, s'_N) \tag{3.46}$$

should only hold when $v_k = v'_k$ for all $k \in \mathbb{N}$. Suppose that (3.46) is fulfilled. Then, this is equivalent to

$$v_1 - v'_1 = \sum_{k=2}^{\infty} (v'_k - v_k) N^{N(1-k)} \, . \tag{3.47}$$

As an intermediate step note that

$$\left|\sum_{k=2}^{\infty}(v'_k - v_k)N^{N(1-k)}\right| \leq \sum_{k=2}^{\infty}|v'_k - v_k|N^{N(1-k)} \tag{3.48}$$

$$\leq (N^N - 1)\sum_{k=2}^{\infty}N^{N(1-k)} \tag{3.49}$$

$$= (N^N - 1)\left[\sum_{k=0}^{\infty}N^{N(1-k)} - (1 + N^N)\right] \tag{3.50}$$

$$= (N^N - 1)\left[\frac{N^N}{1 - 1/N^N} - (1 + N^N)\right] \tag{3.51}$$

$$= 1\ ,$$

with equality if and only if $|v'_k - v_k| = N^N - 1$ for all $k \geq 2$. Here, the right-hand side of (3.48) follows from the triangle inequality,[15] (3.49) from (3.45), and (3.51) from the fact that the sum in (3.50) is a converging geometric series. The inequality, however, is strict since otherwise the uniqueness condition $v_k \neq N^N - 1$ for infinitely many k, unless $v_k = N^N - 1$ for all $k \in \mathbb{N}$, would lead together with (3.47) to the contradiction

$$|v'_1 - v_1| = N^N - 1 = 1\ .$$

Thus, we have

$$|v'_1 - v_1| < 1\ . \tag{3.52}$$

Now, to prove that (3.46) is indeed only possible if $v_k = v'_k$ for all $k \in \mathbb{N}$, we proceed inductively. It is obvious that $|v'_1 - v_1|$ is either a natural number or zero so that we conclude from (3.52) that $|v'_1 - v_1|$ is equal to zero and therefore $v_1 = v'_1$. For the inductive step, suppose that $v_k = v'_k$ for all $k \leq \ell$, with $\ell \in \mathbb{N}$ arbitrary. Then, it follows from (3.46)

$$v_{\ell+1} - v'_{\ell+1} = \sum_{k=\ell+2}^{\infty}(v'_k - v_k)N^{N(\ell+1-k)}$$

and hence $|v_{\ell+1} - v'_{\ell+1}| \leq 1$. The above argument concerning the basis case shows that $v_{\ell+1} = v'_{\ell+1}$ also holds. Since ℓ was arbitrarily chosen, this completes the induction.

Now, let

$$v'_k = \sum_{i=1}^{N}w'_{ik}N^{N-i}\ .$$

Then, the fact that $v_k = v'_k$ for all $k \in \mathbb{N}$ allows us to write

$$0 \leq w_{ik} = w'_{ik} + \left[N\sum_{i=1}^{N-1}(w'_{ik} - w_{ik})N^{N-i-1}\right] \leq N - 1\ .$$

[15] It is easy to verify that the triangle equality applies to an infinite series that converges.

However, except when $w_{ik} = w'_{ik}$ for all $i = 1, \ldots, N$, this is impossible because w'_{ik} is nonnegative and the number in brackets is an integer multiple of N. As a consequence, $v_k \neq v'_k$ unless $w_{ik} = w'_{ik}$ for all $i = 1, \ldots, N$ so that (3.41) *bijectively maps* $\mathbb{E}^N$ onto the interval $[0, N^{1-N}]$.

For the final step note that for every $g(s_1, \ldots, s_N) \in [0, N^{1-N}]$ and every $\psi \in \mathscr{F}([0, N^{1-N}])$,

$$\psi(g(s_1, \ldots, s_N)) = \psi\left(\sum_{i=1}^{N} N^{1-i} \sum_{k=1}^{\infty} w_{ik} N^{-Nk}\right)$$

is uniquely defined. Hence, choosing ψ for some given $f \in \mathscr{F}(\mathbb{E}^N)$ such that

$$\forall (s_1, \ldots, s_N) \in \mathbb{E}^N : \psi(g(s_1, \ldots, s_N)) = f(s_1, \ldots, s_N)$$

shows that every desired function has a nomographic representation, which proves the theorem.

3.A.2 Proof of Theorem 3.2

The proof is a generalization of an idea from [Bie32].

"$\Longleftarrow$": Let $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)} \in \mathbb{E}^N$, with $\boldsymbol{s}^{(1)} \neq \boldsymbol{s}^{(2)}$, and $f \in \mathscr{F}(\mathbb{E}^N)$ be arbitrary. Assuming that g is bijective, it follows that $g(\boldsymbol{s}^{(1)}) \neq g(\boldsymbol{s}^{(2)})$ and from the fact that $\mathbb{E}^N$ has the cardinality of the continuum, Υ has the cardinality of the continuum as well. Now, let $g(\boldsymbol{s}) = s \in \Upsilon$ and $\boldsymbol{g}^\star$ be a vector-valued function such that $\boldsymbol{g}^\star \circ g = \mathrm{id}_{\mathbb{E}^N}$, that is,

$$\boldsymbol{g}^\star : \Upsilon \to \mathbb{E}^N \ , \ s \mapsto \boldsymbol{g}^\star(s) = \begin{pmatrix} g_1^\star(s) = s_1 \\ \vdots \\ g_N^\star(s) = s_N \end{pmatrix} \ .$$

Then, we conclude

$$\begin{aligned} f(s_1, \ldots, s_N) &= f\big(g_1^\star(s), \ldots, g_N^\star(s)\big) \\ &= \big(f \circ \boldsymbol{g}^\star\big)(s) \\ &= \psi(s) = \psi\big(g(\boldsymbol{s})\big) \\ &= \psi\left(\sum_{i=1}^{N} \varphi_i(s_i)\right) \ , \end{aligned}$$

with $\psi \coloneqq f \circ \boldsymbol{g}^\star$.

"$\Longrightarrow$": Suppose that g is not bijective. Then, there exist at least two points $\boldsymbol{s}^{(1)}, \boldsymbol{s}^{(2)} \in \mathbb{E}^N$ such that $\boldsymbol{s}^{(1)} \neq \boldsymbol{s}^{(2)}$ but $g(\boldsymbol{s}^{(1)}) = g(\boldsymbol{s}^{(2)})$ as well as an $f \in \mathscr{F}(\mathbb{E}^N)$ with $f(\boldsymbol{s}^{(1)}) \neq f(\boldsymbol{s}^{(2)})$. This, however, leads to a contradiction because of

$$f\big(\boldsymbol{s}^{(1)}\big) = \psi\big(g(\boldsymbol{s}^{(1)})\big) = \psi\big(g(\boldsymbol{s}^{(2)})\big) = f\big(\boldsymbol{s}^{(2)}\big) \ ,$$

from which follows that $\varphi_1, \ldots, \varphi_N$ are not universal pre-processing functions in the sense of Definition 3.3.

3.A.3 Proof of Theorem 3.6

Let $N \geq 2$ be arbitrary and $\varphi_1, \ldots, \varphi_N \in \mathscr{F}(\mathbb{E})$ be fixed *universal* pre-processing functions to compute every $f(s_1, \ldots, s_N) \in \mathscr{F}(\mathbb{E}^N)$. Furthermore, let $\mathcal{I} \neq \varnothing$ be some subset of $\mathcal{I}_N \coloneqq \{1, \ldots, N\}$. Then, we have to prove that $\{\varphi_k\}_{k \in \mathcal{I}}$ are also universal pre-processing functions to compute every $f(s_1, \ldots, s_k) \in \mathscr{F}(\mathbb{E}^k)$, $k \in \mathcal{I}$. Since the problem is permutation invariant, the numbering of nodes does not matter. Hence, we assume $\mathcal{I} = \mathcal{I}_m = \{1, \ldots, m\}$ with $m < N$ (since otherwise there is nothing to prove).

If we proceed inductively, we have to show that $\varphi_1, \ldots, \varphi_{N-1}$ are universal pre-processing functions for nodes $\{1, \ldots, N-1\}$ to compute every $f(s_1, \ldots, s_{N-1}) \in \mathscr{F}(\mathbb{E}^{N-1})$. If this is successful, we arrive in $N - m$ steps at $\mathcal{I}_m \subset \mathcal{I}_N$.

We prove the induction hypothesis by contradiction. Assume $\varphi_1, \ldots, \varphi_{N-1}$ are *not* universal pre-processing functions. Then, due to Theorem 3.2, the function

$$(s_1, \ldots, s_{N-1}) \mapsto \sum_{i=1}^{N-1} \varphi_i(s_i)$$

is not bijective and hence there exist at least two points $\boldsymbol{s}^{(1)} \coloneqq (s_1^{(1)}, \ldots, s_{N-1}^{(1)}) \in \mathbb{E}^{N-1}$ and $\boldsymbol{s}^{(2)} \coloneqq (s_1^{(2)}, \ldots, s_{N-1}^{(2)}) \in \mathbb{E}^{N-1}$, $\boldsymbol{s}^{(1)} \neq \boldsymbol{s}^{(2)}$, such that

$$\sum_{i=1}^{N-1} \varphi_i\big(s_i^{(1)}\big) = \sum_{i=1}^{N-1} \varphi_i\big(s_i^{(2)}\big) \, .$$

Now, we choose an arbitrary $\hat{s}_N \in (0, 1)$ and consider the points

$$\hat{\boldsymbol{s}}^{(1)} = \begin{pmatrix} \hat{s}_1^{(1)} \\ \vdots \\ \hat{s}_N^{(1)} \end{pmatrix} \coloneqq \begin{pmatrix} \boldsymbol{s}^{(1)} \\ \hat{s}_N \end{pmatrix} \quad \text{and} \quad \hat{\boldsymbol{s}}^{(2)} = \begin{pmatrix} \hat{s}_1^{(2)} \\ \vdots \\ \hat{s}_N^{(2)} \end{pmatrix} \coloneqq \begin{pmatrix} \boldsymbol{s}^{(2)} \\ \hat{s}_N \end{pmatrix} \, .$$

Of course $\hat{\boldsymbol{s}}^{(1)} \neq \hat{\boldsymbol{s}}^{(2)}$ and therefore

$$\begin{aligned} \sum_{i=1}^{N} \varphi_i\big(\hat{s}_i^{(1)}\big) &= \sum_{i=1}^{N-1} \varphi_i\big(s_i^{(1)}\big) + \varphi_N\big(\hat{s}_N\big) \\ &= \sum_{i=1}^{N-1} \varphi_i\big(s_i^{(2)}\big) + \varphi_N\big(\hat{s}_N\big) = \sum_{i=1}^{N} \varphi_i\big(\hat{s}_i^{(2)}\big) \end{aligned}$$

contradicts the universality of $\varphi_1, \ldots, \varphi_N$ when N nodes are active, which proves the preservation of universality for $\mathcal{I}_N \to \mathcal{I}_{N-1}$. Proceeding essentially along the same lines shows that the property is preserved for $\mathcal{I}_{N-1} \to \mathcal{I}_{N-2} \to \cdots \to \mathcal{I}_1$.

3.A.4 Proof of Theorem 3.7

The proof follows immediately from the proof of Theorem 3.6 by considering all subsets $\mathcal{I}_m = \{1, \ldots, m\}$ of $\mathcal{I}_J = \{1, \ldots, J\}$, with $N \leq m < J$.

Starting with the assumption that $\{\varphi_k\}_{1 \leq k \leq J}$ are fixed universal pre-processing functions to compute every $f \in \mathscr{F}(\mathbb{E}^J)$, the induction arrives in $J - N$ steps at $m = N$ so that $\{\varphi_k\}_{1 \leq k \leq m}$ are universal pre-processing functions to compute every $f \in \mathscr{F}(\mathbb{E}^m)$, for all $N \leq m \leq J$.

4

Reliable Computation Over Clustered Gaussian Networks

In this chapter, we extend the considerations of the previous chapter to clustered sensor networks in which the intra-cluster communication takes place over Gaussian MACs (see Remark 2.2). One of the basic facts in multiuser information theory declares that Gaussian MACs are *finite capacity* channels if transmit powers and bandwidths are assumed to be finite [GK11, p. 98]. Communicating arbitrary real values with infinite precision is therefore not possible in any practically relevant manner. To account for this, we propose an achievable computation scheme in which each node first quantizes its real-valued pre-processed sensor readings into a digital message followed by a nested lattice encoder. On the receiving end, upon successfully decoding *the sum* of their intended messages, the FCs apply the corresponding individual post-processing functions in order to obtain reliable estimates of the sought function values.

It turns out that this combination of analog data pre- and post-processing with nested lattice coding allows for the computation of numerous nomographic functions at computation rates that are within certain limits not achievable with separation-based methods. The computation rate is thereby defined as the number of function values that can be reliably computed per channel use. Furthermore, if some finite number of different nomographic functions is allowed to be computed over the channel one after another, then even every continuous function of the sensor readings can be handled. In addition to the improved rate performance, the proposed scheme provides several other advantages that are essential for wireless sensor network applications such as universality, lower decoding complexity, less coordination, and the ability to deal with maximum decoding error probabilities.

A remarkable fact is that the gains over separation-based strategies are achievable by employing the same quantizer along with the same linear channel code at each sensor node. Already in the late 70's, Körner and Marton observed that in specific multi-

terminal scenarios, pure random codes can be significantly outperformed by codes with certain structure [KM79] (see also [NG08a, Zam11]). In particular, they consider the problem of distributively compressing two correlated binary sources such that their modulo-two sum can be losslessly recovered at a decoder. If in this connection both terminals employ the same linear source encoder, the resulting compression rate region strictly contains the Slepian-Wolf region, which is based on the random binning principle [GK11, p. 260].

In [NG07], Nazer and Gastpar carry over the result of Körner and Marton to the scenario in which a single receiver wishes to reliably compute some linear function of distributed sources over a linear finite-alphabet MAC. It turns out that even if the sources are statistically independent, huge performance gains are possible if transmitters utilize linear source and channel codes instead of ordinary random ensembles. Regarding the computation of linear functions over Gaussian MACs (or more generally WMACs), the aforementioned class of *nested lattice codes* are well suited because in addition to their linear structure they achieve the capacity of the point-to-point additive white Gaussian noise (AWGN) channel [NG07, WNPS10, NG11a, NA11, SV12]. This finding, based on the seminal articles [EZ04, ELZ05], serves as a fundamental building block for efficiently computing functions by means of interference over more general networks. For instance, Zhan et al. examine in [ZPGS13] the computation of linear functions over unreliable wired and wireless multi-hop networks that consist of multiple destinations. Closely related to our considerations in this chapter, however, are the studies in [WJG13, JWG13]. There, Wang, Jeon, and Gastpar deal with the problem of computing type-threshold functions of finite-alphabet sources over a network of FCs that are connected with a set of sensor nodes via Gaussian MACs. In contrast, we are interested in computing *real-valued continuous* functions of real-valued sensor readings over noisy clustered networks.

Accordingly, we formally introduce our clustered Gaussian sensor network model in Section 4.1, followed by a corresponding coding scheme in Section 4.2. We then study in Section 4.3 the performance of the scheme in terms of achievable computation rates and conclude the chapter with a short summary in Section 4.4.

Convention

For some positive integer p, $\mathbb{Z}_p = \{0, \ldots, p-1\}$ denotes the set of integers modulo p, $\oplus_p$ addition modulo p, and $\bigoplus$ summation modulo p. The volume of a closed subset $\mathbb{A}$ of the Euclidean space $\mathbb{R}^n$ is described by $\mathrm{Vol}(\mathbb{A})$ and $\log_2^+(x) \coloneqq \max\{\log_2(x), 0\}$.

4.1 Network Model and Problem Statement

Consider the clustered sensor network model of Section 3.2 expanded in the following manner. The spatially distributed nodes periodically monitor the environment resulting in sequences of sensor readings $\{s_i[t] \in \mathcal{S}\}_{t\in\mathbb{N}}$, $i = 1, \ldots, N$, where $\mathcal{S} \subset \mathbb{R}$ again denotes some compact sensing range and t, inter alia, a discrete time or the index of a memory

cell. Each node, say node i, maps its sensor readings to a length-n sequence of transmit symbols, $x_i[1], \ldots, x_i[n]$, subject to the average input cost constraint (2.3), that is,

$$\forall i \in \{1, \ldots, N\} : \varrho_n(x_i[1], \ldots, x_i[n]) = \frac{1}{n} \sum_{m=1}^{n} x_i^2[m] \leq P \tag{4.1}$$

for some $P > 0$. Then, the real-valued symbol received by FC ℓ at channel use $m \in \{1, \ldots, n\}$ is given by[1]

$$Y_\ell[m] = \sum_{i \in C_\ell} x_i[m] + Z_\ell[m] \qquad \ell = 1, \ldots, L \,, \tag{4.2}$$

which is (3.22) complemented by independent and identically distributed (iid) AWGN of variance $\sigma_Z^2 > 0$ (i.e., $Z_\ell \sim \mathcal{N}_{\mathbb{R}}(0, \sigma_Z^2)$ for all ℓ). Thus, the communication in each cluster is modeled by a special WMAC, namely, a Gaussian MAC with unit fading coefficients (see Remark 2.2). In all that follows, we call such a network a *clustered Gaussian sensor network*.

Now, the problem to be solved in this chapter is to efficiently compute, at some pre-defined accuracy, arbitrary *continuous desired functions*

$$f_\ell : \mathcal{S}^{|C_\ell|} \to \mathbb{R} \,, \; (s_{\ell_1}, \ldots, s_{\ell_{|C_\ell|}}) \mapsto f_\ell(s_{\ell_1}, \ldots, s_{\ell_{|C_\ell|}}) \qquad \ell = 1, \ldots, L \tag{4.3}$$

over a clustered Gaussian sensor network. This is challenging due to the following reasons.

(i) The common nodes (i.e., nodes that belong to some overlap between adjacent clusters) can be heard by more than one FC, which results in interference between clusters.[2]

(ii) The superposition of channel input symbols is corrupted by Gaussian noise.

To account for these facts, we need to devise an achievable computation scheme that combines an adaptive data pre- and post-processing with a coding strategy that fundamentally differs from those designed for standard message transfer. In particular, to address (ii), we employ a nested lattice code that is well suited for protecting sums of codewords, whereas (i) is accounted for by exploiting the universality of Kolmogorov's representations of continuous functions such as examined in the previous chapter.

4.2 Coding Scheme

The basis for the achievable computation scheme proposed in this section will be the pre- and post-processing strategy given in Section 3.2.1 (see Figure 3.5). In order to facilitate the understanding, we briefly recap this strategy in Section 4.2.1 whereas its extension to Gaussian MACs is presented in Sections 4.2.2 and 4.2.3.

[1] For ease of exposition, the considerations in this chapter are made under the assumption of real-valued Gaussian MACs. The extension to the complex baseband is straightforward.

[2] Recall that each cluster overlaps with at least one of its neighboring clusters, that is, $\forall \ell \in \{1, \ldots, L\} \, \exists \ell' \neq \ell : C_\ell \cap C_{\ell'} \neq \varnothing$ (see Figure 3.4 for an example).

4.2.1 Outline

Let φ_{ij}, $i = 1, \ldots, N$ and $j = 1, \ldots, 2N+1$, be a set of universal continuous pre-processing functions. Then, in accordance with (3.29) and Proposition 3.2, every continuous desired function (4.3) to be computed at FC ℓ can be expressed as

$$f_\ell(s_{\ell_1}[t], \ldots, s_{\ell_{|C_\ell|}}[t]) = \sum_{j=1}^{2N+1} \psi_{\ell j}\left(\sum_{i \in C_\ell} \varphi_{ij}(s_i[t]) + \gamma_{\ell j}\right) \qquad \ell = 1, \ldots, L \tag{4.4}$$

through an appropriate choice of the continuous post-processing functions $\psi_{\ell j}$. Just as in Section 3.2.1, the terms $\gamma_{\ell j} \in \mathbb{R}$ are defined as

$$\gamma_{\ell j} \coloneqq \sum_{i \notin C_\ell} \varphi_{ij}(0) \qquad j = 1, \ldots, 2N+1\,, \tag{4.5}$$

which are independent of the sensor readings.

Suppose that each sensor node is uniquely assigned one of the sets of pre-processing functions $\{\varphi_{ij} \in \mathscr{C}^0(\mathcal{S})\}_{j=1}^{2N+1}$, $i = 1, \ldots, N$. Moreover, suppose that FC ℓ a priori knows the set $\{\psi_{\ell j} \in \mathscr{C}^0(\mathbb{R})\}_{j=1}^{2N+1}$ of post-processing functions as well as the set of constants $\{\gamma_{\ell j}\}_{j=1}^{2N+1}$. Then, the computation scheme described in detail in the following subsections can shortly be outlined as follows:

- With sensor nodes transmitting concurrently in the same frequency band, each FC, say FC ℓ, reliably reconstructs for every $t \in \mathbb{N}$ the sequence

  $$\sum_{i \in C_\ell} \varphi_{i1}(s_i[t]), \ldots, \sum_{i \in C_\ell} \varphi_{i,2N+1}(s_i[t])$$

  of superimposed pre-processed sensor readings.

- The FCs add the constants (4.5), apply their post-processing functions

  $$\psi_{\ell 1}, \ldots, \psi_{\ell,2N+1}$$

  and finally sum up all intermediate results to yield the desired function values (4.4).

Remark 4.1. We would like to emphasize again that by Theorem 3.4, the pre-processing functions in each cluster do not depend on the choice of f_ℓ, $\ell = 1, \ldots, L$, so that they do not need to be updated if the desired functions change during network operation. The same applies in case of a changing network topology (see Theorems 3.6 and 3.7).

In contrast to the ideal case discussed in the previous chapter, realizing the above two-step procedure in a reliable manner requires the application of coding techniques because we have to deal with the additive noise in (4.2). A generic information-theoretic

framework that is well suited for this kind of problems is *computation coding*, which is introduced by Nazer and Gastpar in [NG07]. We adapt their notions of computation code, achievable computation rate, and computation capacity to our specific needs as follows.

Definition 4.1 (Computation Code)**.** Let $T, n \in \mathbb{N}$, and $P > 0$ be chosen arbitrarily and let $W^{(n)} : \mathcal{X}^{(n)} \to \mathbb{R}^n$ be the n^{th} extension of some given WMAC W where

$$\mathcal{X}^{(n)} \coloneqq \Big\{ (x[1], \dots, x[n]) \in \mathbb{R}^n \,|\, \varrho_n(x[1], \dots, x[n]) \leq P \Big\}$$

denotes, with respect to some fixed input cost function ϱ_n (see Definition 2.2), the constrained channel input alphabet. An (f, T, n) *computation code* for W consists of:

- A desired function $f : \mathcal{S}^N \to \mathbb{R}$.
- N transmitters (encoding functions)

  $$\mathsf{Tx}_i : \mathcal{S}^T \to \mathcal{X}^{(n)} \,,\ (s_i[1], \dots, s_i[T]) \mapsto \boldsymbol{x}_i \coloneqq (x_i[1], \dots, x_i[n]) \tag{4.6}$$

  that map T sensor readings to n channel input symbols.
- A receiver (decoding function)

  $$\mathsf{Rx} : \mathbb{R}^n \to \mathbb{R}^T \,,\ (Y[1], \dots, Y[n]) \mapsto \Big(\hat{f}\big(s_1[1], \dots, s_N[1]\big), \dots, \hat{f}\big(s_1[T], \dots, s_N[T]\big) \Big)$$

  that assigns T estimates of desired function-values to each length-n sequence of channel output symbols.

The performance of a computation code is typically determined in terms of an achievable computation rate, which specifies how many function values can be computed per channel use within a predefined accuracy.

Definition 4.2 (Computation Rate)**.** Let $f \in \mathscr{F}(\mathcal{S}^N)$ be some fixed desired function, $\hat{f}$ a corresponding estimate at the FC, and W a given WMAC. Furthermore, let $d_{\mathbb{R}}$ be some distortion measure and $\varepsilon > 0$ an arbitrary but fixed accuracy. Then, with respect to $(d_{\mathbb{R}}, \varepsilon)$, $R^{\mathrm{C}}(f, d_{\mathbb{R}}, \varepsilon) \in \mathbb{R}_+$ is said to be an *achievable computation rate* for f if for every rate

$$R' \coloneqq \frac{T}{n} < R^{\mathrm{C}}(f, d_{\mathbb{R}}, \varepsilon) \qquad \frac{\text{function values}}{\text{channel use}}$$

and every $\delta > 0$ there exists an (f, nR', n) computation code for W such that the error probability fulfills for n sufficiently large

$$\mathbb{P}\left(\bigcup_{t=1}^{T} \Big\{ d_{\mathbb{R}}\Big(\hat{f}\big(s_1[t], \dots, s_N[t]\big), f\big(s_1[t], \dots, s_N[t]\big) \Big) < \varepsilon \Big\} \right) \geq 1 - \delta \,. \tag{4.7}$$

Definition 4.3 (Computation Capacity)**.** For given $f \in \mathscr{F}(\mathcal{S}^N)$ and $(d_{\mathbb{R}}, \varepsilon)$, the *computation capacity* of a WMAC W is defined to be the supremum over all achievable computation rates.

Remark 4.2. As already pointed out in Chapter 2, the main difference to the standard information-theoretic setting is given by (4.7), which clearly states that we only require to reliably decode function values rather than individual sensor readings.

For the remainder of this chapter, let $d_{\mathbb{R}}$ be chosen to be the supremum norm $\|\cdot\|_\infty$. Then, in connection with Definitions 4.1 and 4.2, we need L computation codes (f_ℓ, T, n) (i.e., one for each cluster) that guarantee

$$\mathbb{P}\left(\bigcup_{t=1}^{T}\bigcup_{\ell=1}^{L}\left\{\left\|\hat{f}_\ell - f_\ell\right\|_\infty < \varepsilon\right\}\right) \geq 1-\delta \tag{4.8}$$

for some given $\varepsilon, \delta > 0$, and sufficiently large block length n. Therefore, in the following two subsections we devise N transmitters (4.6) as well as L receivers

$$\mathsf{Rx}_\ell : \mathbb{R}^n \to \mathbb{R}^T\ ,\ \mathbf{y}_\ell \coloneqq (Y_\ell[1], \ldots, Y_\ell[n]) \mapsto \left(\hat{f}_\ell(\boldsymbol{s}_\ell[1]), \ldots, \hat{f}_\ell(\boldsymbol{s}_\ell[T])\right), \tag{4.9}$$

$\boldsymbol{s}_\ell[t] \coloneqq (s_{\ell_1}[t], \ldots, s_{\ell_{|C_\ell|}}[t])$, that are capable of achieving (4.8) at computation rates that are, to some extent, not achievable with separation-based strategies. Since the transmitters and receivers have to respect the particular pre- and post-processing strategy outlined at the beginning of this subsection as well as the linear structure of the channel, we decompose them into multiple components each.

4.2.2 Data Pre- and Post-Processing

As already mentioned at the beginning of the chapter, the Gaussian MAC is a finite capacity channel (see Theorem 4.2 below) so that we have to first quantize the pre-processed sensor readings. Since the sensing range $\mathcal{S}$ is a compact interval, it follows that the range of each pre-processing function is a compact interval as well and we denote these sets by $\Pi_{ij} \subset \mathbb{R}$ (i.e., $\forall s \in \mathcal{S} : \varphi_{ij}(s) \in \Pi_{ij}$ for all i, j) in the following.[3] As a consequence, the union

$$\Pi \coloneqq \bigcup_{i=1}^{N}\bigcup_{j=1}^{2N+1}\Pi_{ij}$$

is compact and by

$$\pi_{\max} \coloneqq \max_{\xi\in\Pi}|\xi|\ , \tag{4.10}$$

we denote the unique maximal element in absolute value.

Remark 4.3. To keep the notation simple, we assume in the following that the elements of Π are nonnegative. This is without loss of generality as Π can be shifted to the nonnegative reals by adding $\pi_{\max}$ to every $\xi \in \Pi$.

[3] Please recall that the pre-processing functions are continuous.

We assume that all nodes in the network employ the same *quantizer*

$$Q : \Pi \to \{0,1\}^b \tag{4.11}$$

that forms for each $t \in \mathbb{N}$ the $2N+1$ length-b binary representations

$$\boldsymbol{w}_{ij}[t] \coloneqq Q\big(\varphi_{ij}(s_i[t])\big) \qquad i = 1, \ldots, N \;, \tag{4.12}$$

where b is some positive integer to be specified below. To better understand how quantizer Q works, recall first from Lemma 3.1 in Section 3.A that every $\xi \in \Pi$ has a unique dyadic expansion

$$\xi = \sum_{r=-v}^{\infty} \frac{w_r}{2^r} = \lim_{\eta\to\infty} \sum_{r=-v}^{\eta} \frac{w_r}{2^r} \;,$$

with $w_r \in \{0,1\}$ and $w_r \neq 1$ for infinitely many r, unless $w_r = 1$ for all r. Observe that v depends on the largest integer part of ξ. With this in mind, consider for each $i = 1, \ldots, N$ and $j = 1, \ldots, 2N+1$ the instantaneous approximation (in the sense given in the proof of Lemma 4.4 below)

$$\varphi_{ij}(s_i[t]) \approx \tilde{\varphi}_{ij}(s_i[t]) = \sum_{r=-v}^{\eta} \frac{w_{ij}^{(r)}[t]}{2^r} \tag{4.13}$$

by terminating the dyadic expansion. Then, setting $b \coloneqq \eta + v + 1$ with

$$v \coloneqq \lfloor \log_2(\pi_{\max}) \rfloor \tag{4.14}$$

fixed, quantizer Q simply forms the length-b binary representations in (4.12) by extracting the digits from (4.13).

Each quantizer is followed by the same *source encoder*

$$\mathcal{E}_1 : \{0,1\}^{bT} \to \mathbb{Z}_p^k \;, \tag{4.15}$$

which combines $T \in \mathbb{N}$ of the binary representations (4.12) to a length-k message over $\mathbb{Z}_p$:

$$\boldsymbol{w}_{ij} = \mathcal{E}_1(\boldsymbol{w}_{ij}[1], \ldots, \boldsymbol{w}_{ij}[T]) \;, \tag{4.16}$$

$i = 1, \ldots, N$ and $j = 1, \ldots, 2N+1$. Here and hereafter, k is a natural number and p is assumed to be prime.[4] See Fig. 4.1 for a block diagram.

Now, in order to compute the desired function (4.4) over the Gaussian MAC (4.2), the ℓ^{th} FC first needs for each fixed t, $t = 1, \ldots, T$, reliable estimates of the $2N+1$ inner sums

$$\tilde{g}_{\ell j}(s_{\ell_1}[t], \ldots, s_{\ell_{|C_\ell|}}[t]) \coloneqq \sum_{i \in C_\ell} \tilde{\varphi}_{ij}(s_i[t]) \;. \tag{4.17}$$

[4] We construct the encoder (4.15) explicitly in the proof of Theorem 4.1.

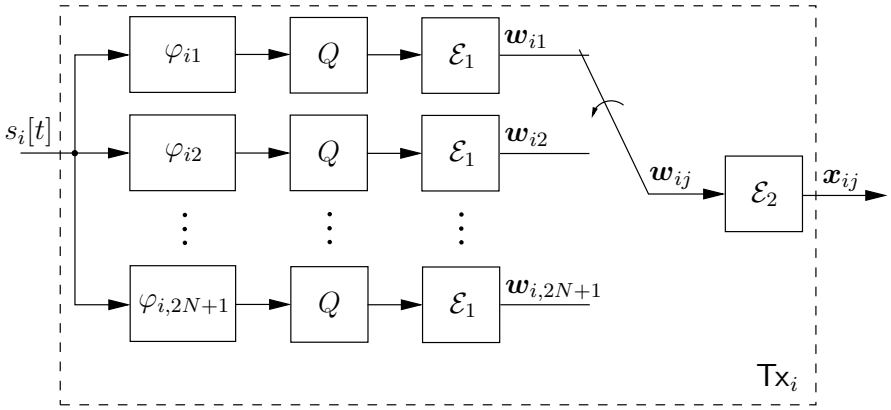


Figure 4.1: Block diagram of the i^{th} computation-transmitter, Tx_i, $i = 1, \ldots, N$, consisting of adequate data pre-processing as well as of nested lattice encoding.

As long as p is in (4.16) chosen sufficiently large, this can be achieved by reliably computing the *modulo p sums* of the messages (4.16):

$$\boldsymbol{g}_{\ell j} \coloneqq \bigoplus_{i \in C_\ell} \boldsymbol{w}_{ij} \; . \tag{4.18}$$

Remark 4.4. Requiring p to be sufficiently large is necessary in order to avoid a wraparound in (4.18).

Once FC ℓ knows $\{\boldsymbol{g}_{\ell j}\}_{j=1}^{2N+1}$, a *source decoder*

$$\mathcal{D}_1 : \mathbb{Z}_p^k \to \{0, \ldots, N\}^{bT}$$

first decomposes each $\boldsymbol{g}_{\ell j}$ into modulo p sums of the corresponding binary representations (4.12):

$$\mathcal{D}_1(\boldsymbol{g}_{\ell j}) = (\boldsymbol{g}_{\ell j}[1], \ldots, \boldsymbol{g}_{\ell j}[T]) \coloneqq \left(\bigoplus_{i \in C_\ell} \boldsymbol{w}_{ij}[1], \ldots, \bigoplus_{i \in C_\ell} \boldsymbol{w}_{ij}[T] \right) \; , \tag{4.19}$$

$j = 1, \ldots, 2N+1$ and $\ell = 1, \ldots, L$. Afterwards, the *inverse quantizer*

$$Q^{-1} : \{0, \ldots, N\}^b \to N\Pi \coloneqq \{\xi_1 + \cdots + \xi_N \,|\, \xi_1 \in \Pi, \ldots, \xi_N \in \Pi\} \tag{4.20}$$

evaluates the right hand side of (4.13) for each t, $t = 1, \ldots, T$, at the digits $\boldsymbol{g}_{\ell j}[t] \in \mathbb{Z}_p^k$. Then, adding the constants $\gamma_{\ell j}$ (see (4.21)), applying the post-processing functions $\{\psi_{\ell j}\}_{j=1}^{2N+1}$, and summing over all intermediate results provides FC ℓ with an approximation of (4.4) given by

$$\tilde{f}_\ell(s_{\ell_1}[t], \ldots, s_{\ell_{|C_\ell|}}[t]) \coloneqq \sum_{j=1}^{2N+1} \psi_{\ell j} \left(\sum_{i \in C_\ell} \tilde{\varphi}_{ij}(s_i[t]) + \gamma_{\ell j} \right) \qquad \ell = 1, \ldots, L \; . \tag{4.21}$$

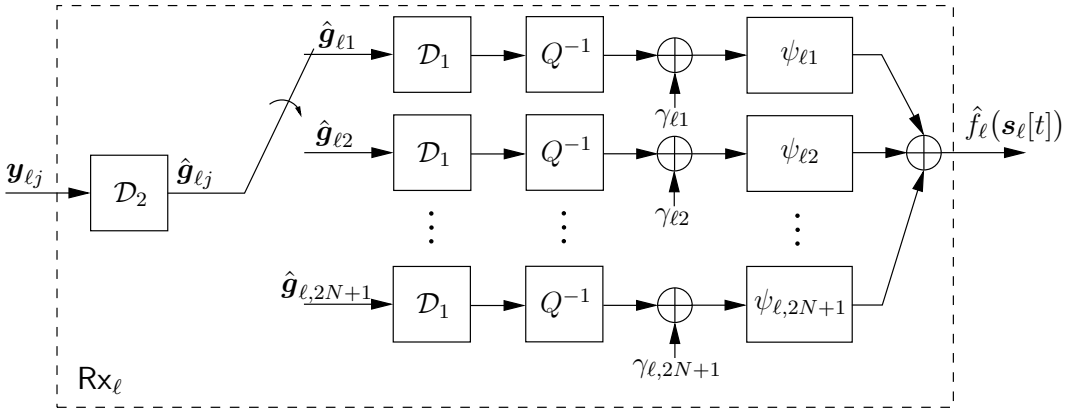


Figure 4.2: Block diagram of the ℓ^{th} computation-receiver, Rx_ℓ (i.e., of FC ℓ, $\ell = 1, \ldots, L$), consisting of adequate data post-processing as well as of nested lattice decoding.

The corresponding quantization error, $\|f_\ell - \tilde{f}_\ell\|_\infty$, crucially depends on the explicit choice of the quantization parameter b (see (4.11)). As this will also have a significant impact on the achievable computation rate, the relationship between b and some given accuracy $\varepsilon > 0$ is provided in Section 4.3.[5]

See Figure 4.2 for a block diagram of the described data post-processing.

Remark 4.5. Strictly speaking, Q^{-1} such as defined in (4.20) is not the inverse function of (4.11) because Q maps to $\{0,1\}^b$ whereas Q^{-1} has $\{0,\ldots,N\}^b$ as its domain. In order to illustrate what is actually meant by the symbol Q^{-1}, let us consider a very simple two-node example (i.e., $N = 2$, $L = 1$) in which we intend to compute the real sum $s_1 + s_2 = 3 + 2 = 5$ of sensor readings over the channel. In this regard, we choose without loss of generality $k = 3$ and $p = 3$. Then, adding the quantized sensor readings $\boldsymbol{w}_1 = Q(s_1) = Q(3) = 011$ and $\boldsymbol{w}_2 = Q(s_2) = Q(2) = 010$ modulo 3 results in $\boldsymbol{g} = \boldsymbol{w}_1 \oplus_3 \boldsymbol{w}_2 = 011 \oplus_3 010 = 021$. Putting this back into a dyadic expansion provides $Q^{-1}(\boldsymbol{g}) = Q^{-1}(021) = 0 \cdot 2^2 + 2 \cdot 2^1 + 1 \cdot 2^0 = 5$ as desired.

Remark 4.6. The two-node example considered in Remark 4.5 can also be used to demonstrate the significance of the condition that p has to be sufficiently large. Let $p = 2$ and everything else as before. Then, it follows that $Q^{-1}(\boldsymbol{g}) = Q^{-1}(\boldsymbol{w}_1 \oplus_2 \boldsymbol{w}_2) = Q^{-1}(001) = 1 \neq 5 = s_1 + s_2$.

4.2.3 Nested Lattice Coding

Reading through Section 4.2.2 reveals that the crucial step in achieving reliable computations is the protection of (4.18) against Gaussian noise. In order to ensure this,

[5] Note that for the general case, the quantization error has to be evaluated with regard to the distortion measure $d_{\mathbb{R}}$.

we employ sequences of nested lattice codes from [NG11a, NDG11] as they possess favorable structural properties. Towards this end, we first briefly recap some necessary notions on nested lattices from [EZ04, ELZ05, For03, NG11a].

Basic Facts and Definitions

Definition 4.4 (Lattice)**.** An n-dimensional *lattice* Λ is a discrete additive subgroup of the Euclidean space $\mathbb{R}^n$ that is closed under addition and subtraction (i.e., $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in \Lambda \Rightarrow \boldsymbol{\lambda}_1 \pm \boldsymbol{\lambda}_2 \in \Lambda$). For every lattice Λ there exists a full-rank generator/basis matrix $\boldsymbol{G} \in \mathbb{R}^{n \times n}$ so that

$$\Lambda = \{\boldsymbol{\lambda} = \boldsymbol{G}\boldsymbol{\mu} \,|\, \boldsymbol{\mu} \in \mathbb{Z}^n\} =: \boldsymbol{G}\mathbb{Z}^n \ .$$

Definition 4.5 (Lattice Quantizer)**.** A *quantizer* associated with a lattice Λ is a map that assigns every point $\boldsymbol{\mu} \in \mathbb{R}^n$ to the nearest lattice point in Euclidean distance, that is,

$$Q_\Lambda : \mathbb{R}^n \to \Lambda \ , \ \boldsymbol{\mu} \mapsto Q_\Lambda(\boldsymbol{\mu}) = \operatorname*{argmin}_{\boldsymbol{\lambda} \in \Lambda} \|\boldsymbol{\mu} - \boldsymbol{\lambda}\|_2 \ ,$$

where ties are dissolved in a systematic way.

Definition 4.6 (Voronoi Region)**.** The *fundamental Voronoi region* of some n-dimensional lattice Λ, denoted as $\mathcal{V}$, is the set of all points in $\mathbb{R}^n$ that quantize to the zero vector:

$$\mathcal{V} \coloneqq \{\boldsymbol{\mu} \in \mathbb{R}^n \,|\, Q_\Lambda(\boldsymbol{\mu}) = \boldsymbol{0}\} \ .$$

Definition 4.7 (Modulo Operation)**.** The *modulo operation* with respect to a lattice Λ provides, for every $\boldsymbol{\mu} \in \mathbb{R}^n$, the quantization error

$$[\boldsymbol{\mu}] \bmod \Lambda \coloneqq \boldsymbol{\mu} - Q_\Lambda(\boldsymbol{\mu}) \ ,$$

which is always in $\mathcal{V}$.

Definition 4.8 (Moments)**.** The *second moment* (per dimension) of some lattice $\Lambda \subset \mathbb{R}^n$ is defined as

$$\sigma^2(\Lambda) \coloneqq \frac{1}{n} \frac{\int_{\mathcal{V}} \|\boldsymbol{x}\|_2^2 \, \mathrm{d}\boldsymbol{x}}{\mathrm{Vol}(\mathcal{V})} \ , \tag{4.22}$$

where $\mathrm{Vol}(\mathcal{V}) = \int_{\mathcal{V}} \mathrm{d}\boldsymbol{x}$ denotes the volume of the fundamental Voronoi region of Λ, whereas the *normalized second moment* is defined as

$$G(\Lambda) \coloneqq \frac{\sigma^2(\Lambda)}{\mathrm{Vol}(\mathcal{V})^{2/n}} \ . \tag{4.23}$$

Definition 4.9 (Goodness)**.** Let $\{\Lambda^{(n)}\}$ be a sequence of lattices indexed by their dimension and $\mathsf{z} \sim \mathcal{N}_{\mathbb{R}}(\boldsymbol{0}, \sigma_Z^2 \boldsymbol{I}_n)$ multivariate Gaussian noise. Then, $\{\Lambda^{(n)}\}$ is said to be *good for AWGN channel coding* if

$$\mathbb{P}\big(\mathsf{z} \notin \mathcal{V}^{(n)}\big) \to 0$$

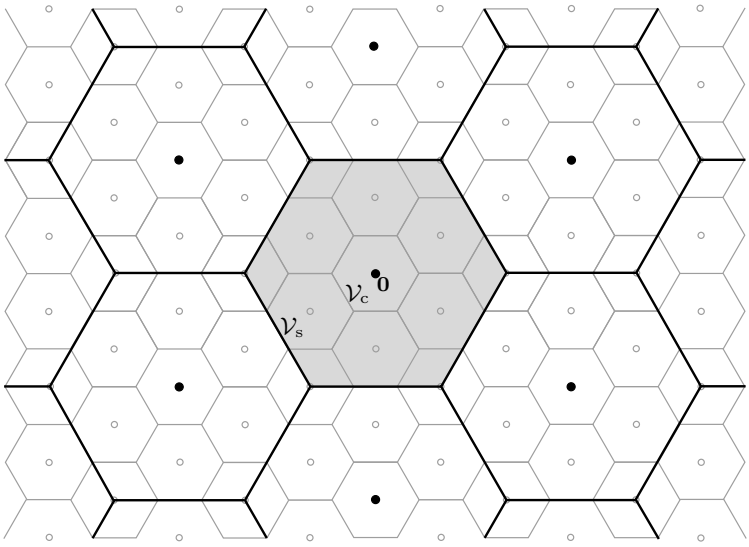


Figure 4.3: Part of a nested hexagonal lattice $\Lambda_s \subset \Lambda_c$ in Euclidean space $\mathbb{R}^2$ with $\mathcal{V}_s$ the fundamental Voronoi region of the shaping lattice Λ_s (black dots) and $\mathcal{V}_c$ the fundamental Voronoi region of the coding lattice Λ_c (white dots).

exponentially fast with growing n whenever

$$\mathrm{Vol}\big(\mathcal{V}^{(n)}\big)^{2/n} > 2\pi\mathrm{e}\sigma_Z^2 \ .$$

On the other hand, the sequence $\{\Lambda^{(n)}\}$ is said to be *good for shaping* if

$$\lim_{n\to\infty} \log_2\big(2\pi\mathrm{e}G(\Lambda^{(n)})\big) = 0 \ .$$

Definition 4.10 (Nested Lattices)**.** A lattice Λ_s is *nested* in some lattice Λ_c if Λ_s is a sublattice of Λ_c (i.e., $\Lambda_s \subset \Lambda_c$). The lattice Λ_s with fundamental Voronoi region $\mathcal{V}_s$ is called *shaping lattice* whereas Λ_c with fundamental Voronoi region $\mathcal{V}_c$ is called *coding lattice*.

Figure 4.3 depicts an example of a two-dimensional nested hexagonal lattice pair, in which

$$\Lambda_c = \boldsymbol{G}\mathbb{Z}^2 = \begin{pmatrix} \sqrt{3}/2 & 0 \\ 1/2 & 1 \end{pmatrix} \qquad \text{and} \qquad \Lambda_s = 3\boldsymbol{G}\mathbb{Z}^2 \ .$$

The modulo operation such as defined in Definition 4.7 has the following useful properties.

Lemma 4.1. *For all $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}^n$, all $\alpha \in \mathbb{R}_+$, and every pair of nested lattices $\Lambda \subset \Lambda' \subset \mathbb{R}^n$, the modulo operation of Definition 4.7 fulfills:*

a) $[\boldsymbol{\mu} + \boldsymbol{\nu}] \bmod \Lambda = \big[[\boldsymbol{\mu}] \bmod \Lambda + [\boldsymbol{\nu}] \bmod \Lambda\big] \bmod \Lambda$ (distributivity)

b) $[Q_{\Lambda'}(\boldsymbol{\mu})] \bmod \Lambda = [Q_{\Lambda'}([\boldsymbol{\mu}] \bmod \Lambda)] \bmod \Lambda$ (commutativity)

c) $\alpha[\boldsymbol{\mu}] \bmod \Lambda = [\alpha\boldsymbol{\mu}] \bmod \alpha\Lambda$.

Proof. The proof is provided in Appendix 4.A.1 at the end of the chapter. □

Remark 4.7. Note that part a) of the lemma implies

$$[\boldsymbol{\mu} + \boldsymbol{\nu}] \bmod \Lambda = \big[[\boldsymbol{\mu}] \bmod \Lambda + \boldsymbol{\nu}\big] \bmod \Lambda = \big[\boldsymbol{\mu} + [\boldsymbol{\nu}] \bmod \Lambda\big] \bmod \Lambda \, .$$

In the context of nested lattices, we distill from Erez and Zamir's main theorem in [EZ04] the following lemma, which will be essential for proving our results in Section 4.3.

Lemma 4.2 (Erez-Zamir)**.** *There exists a sequence of nested lattices* $\{\Lambda_{\mathrm{s}}^{(n)} \subset \Lambda_{\mathrm{c}}^{(n)}\}$ *indexed by their dimension in which* $\{\Lambda_{\mathrm{s}}^{(n)}\}$ *is simultaneously good for AWGN channel coding and shaping and* $\{\Lambda_{\mathrm{c}}^{(n)}\}$ *for AWGN channel coding.*

Definition 4.11 (Nested Lattice Code)**.** Given some pair of n-dimensional nested lattices $\Lambda_{\mathrm{s}} \subset \Lambda_{\mathrm{c}}$, a *nested lattice code* $\mathcal{C}^{(n)}$ is defined as

$$\mathcal{C}^{(n)} \coloneqq \Lambda_{\mathrm{c}} \cap \mathcal{V}_{\mathrm{s}} \tag{4.24}$$

with rate

$$R_{\mathcal{C}} = \frac{1}{n} \log_2\big(|\mathcal{C}^{(n)}|\big) = \frac{1}{n} \log_2 \left(\frac{\mathrm{Vol}(\mathcal{V}_{\mathrm{s}})}{\mathrm{Vol}(\mathcal{V}_{\mathrm{c}})} \right) \, . \tag{4.25}$$

Remark 4.8. The essential algebraic property of a nested lattice code is linearity, which means that each sum of lattice codewords modulo the shaping lattice is a codeword itself:

$$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathcal{C}^{(n)} \Rightarrow \left[\sum_{i=1}^{N} \boldsymbol{x}_i \right] \bmod \Lambda_{\mathrm{s}} \in \mathcal{C}^{(n)} \, . \tag{4.26}$$

Channel Encoding

In order to protect the modulo sums in (4.18) against Gaussian receiver noise, each sensor node employs the same n-dimensional nested lattice code $\mathcal{C}^{(n)}$ based on a nested lattice pair taken from Lemma 4.2. The shaping lattice is scaled such that the second moment equals the transmit power constraint (i.e., $\sigma^2(\Lambda_{\mathrm{s}}) = P$). Thus, each node is equipped with the same *channel encoder* (see Figure 4.1)

$$\mathcal{E}_2 : \mathbb{Z}_p^k \to \mathcal{C}^{(n)} \subset \mathbb{R}^n \, , \tag{4.27}$$

which maps each message, $\boldsymbol{w}_{ij}$, to a length-n lattice codeword. That is, for each $j = 1, \ldots, 2N+1$,

$$\boldsymbol{x}_{ij} = (x_{ij}[1], \ldots, x_{ij}[n]) = \mathcal{E}_2(\boldsymbol{w}_{ij}) \qquad i = 1, \ldots, N \, .$$

Due to the scaling of the shaping lattice, each codeword meets the average power constraint and the message rate (4.25) (in bits per channel use) is

$$R = \frac{k}{n} \log_2(p) \ . \tag{4.28}$$

In what follows, we assume that the encoder (4.27) is a bijection that preserves linearity:

$$\mathcal{E}_2^{-1} \left(\left[\sum_{i \in C_\ell} \mathcal{E}_2(\boldsymbol{w}_{ij}) \right] \bmod \Lambda_\mathrm{s} \right) = \bigoplus_{i \in C_\ell} \boldsymbol{w}_{ij} \ . \tag{4.29}$$

The existence of nested lattice codebooks with bijective linearity preserving encoders is guaranteed by the following lemma.

Lemma 4.3 (Nazer-Gastpar)**.** *Let $\Lambda_\mathrm{s} \subset \Lambda_\mathrm{c}$ be a nested lattice pair taken for some fixed $n \in \mathbb{N}$ from the sequence of Lemma 4.2 and let $\mathcal{C}^{(n)} = \Lambda_\mathrm{c} \cap \mathcal{V}_\mathrm{s}$ be the corresponding nested lattice code. Then, there exists an encoder (4.27) that is bijective and satisfies (4.29).*

Proof. Since the lemma provides the crucial fact that we can map real sums of lattice points back to modulo p sums of messages, we reproduce in Appendix 4.A.2 a slightly modified version of Nazer and Gastpar's proof (i.e., Lemmas 5 and 6 of [NG11a]) for the sake of completeness. □

Channel Decoding

After the Gaussian MACs have been used by the sensor nodes n times, the ℓ^{th} FC is aware of the length-n receive vector

$$\mathbf{y}_{\ell j} = \sum_{i \in C_\ell} \boldsymbol{x}_{ij} + \mathbf{z}_{\ell j} \ , \tag{4.30}$$

where $\mathbf{z}_{\ell j} \sim \mathcal{N}_\mathbb{R}(\mathbf{0}, \sigma_Z^2 \boldsymbol{I}_n)$, $j = 1, \ldots, 2N+1$ (see Figure 4.2). To obtain estimates of the modulo p sums (4.18), the FC applies a channel decoder

$$\mathcal{D}_2 : \mathbb{R}^n \to \mathbb{Z}_p^k$$

that consists of an *Euclidean nearest neighbor decoder* (also known as lattice decoder [EZ04]) followed by the inverse of the channel encoding function. Thus, by (4.29) we have

$$\hat{\mathbf{g}}_{\ell j} = \mathcal{D}_2(\mathbf{y}_{\ell j}) = \mathcal{E}_2^{-1} \left(\left[Q_{\Lambda_\mathrm{c}}(\mathbf{y}_{\ell j}) \right] \bmod \Lambda_\mathrm{s} \right) \ . \tag{4.31}$$

Obviously, the nearest neighbor decoder quantizes a receive vector onto the coding lattice and then reduces the outcome to the shaping lattice in order to guarantee that the resulting lattice point is a valid codeword (see Figure 4.4). Inserting (4.30) in (4.31)

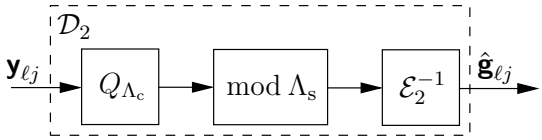


Figure 4.4: Block diagram of the Euclidean nearest neighbor decoder (lattice decoder), $\mathcal{D}_2$, which is part of the receiver at FC ℓ (i.e., Rx_ℓ, $\ell = 1, \ldots, L$) depicted in Figure 4.2. At its output, the decoder provides an estimate of the modulo p sum of messages.

shows together with Lemma 4.1 that

$$\begin{aligned}\hat{\mathbf{g}}_{\ell j} &= \mathcal{E}_2^{-1}\left(\left[Q_{\Lambda_c}\left(\sum_{i\in C_\ell} \boldsymbol{x}_{ij} + \mathbf{z}_{\ell j}\right)\right] \bmod \Lambda_s\right) \\ &= \mathcal{E}_2^{-1}\left(\left[Q_{\Lambda_c}\left(\left[\sum_{i\in C_\ell} \boldsymbol{x}_{ij}\right] \bmod \Lambda_s + \mathbf{z}_{\ell j}\right)\right] \bmod \Lambda_s\right) \\ &= \mathcal{E}_2^{-1}\Big(\big[Q_{\Lambda_c}(\boldsymbol{x}_{\ell j} + \mathbf{z}_{\ell j})\big] \bmod \Lambda_s\Big)\,, \end{aligned} \tag{4.32}$$

where $\boldsymbol{x}_{\ell j} \coloneqq \left[\sum_{i\in C_\ell} \boldsymbol{x}_{ij}\right] \bmod \Lambda_s$. Because of (4.26), we have $\boldsymbol{x}_{\ell j} \in \mathcal{C}^{(n)}$, for all ℓ, j, so that (4.30) is essentially a codeword corrupted by Gaussian noise. The computation over a clustered Gaussian sensor network can therefore be seen as L point-to-point links in which each transmitter, say transmitter ℓ, aims at reliably communicating the codewords $\boldsymbol{x}_{\ell 1}, \ldots, \boldsymbol{x}_{\ell,2N+1}$ to its intended receiver.

Decoding Error Probability

Let $\delta > 0$ be arbitrary. Then, the modulo p sums of messages are said to be decoded with error probability δ if

$$P_e^{(n)} \coloneqq \mathbb{P}\left(\bigcup_{\ell=1}^{L}\bigcup_{j=1}^{2N+1}\{\hat{\mathbf{g}}_{\ell j} \neq \boldsymbol{g}_{\ell j}\}\right) \leq \delta\,. \tag{4.33}$$

To demonstrate that this can be considered as a *maximum probability of error*, we establish in the following the upper bound

$$P_e^{(n)} \leq \sum_{\ell=1}^{L}\sum_{j=1}^{2N+1} \mathbb{P}(\hat{\mathbf{g}}_{\ell j} \neq \boldsymbol{g}_{\ell j}) \leq L(2N+1)\mathbb{P}(\mathbf{z} \notin \mathcal{V}_c)\,, \tag{4.34}$$

with $\mathbf{z} \sim \mathcal{N}_{\mathbb{R}}(\mathbf{0}, \sigma_Z^2 \boldsymbol{I}_n)$. Towards this end, note that each node chooses for each j one out of p^k codewords so that at FC ℓ, the modulo-p sums $\boldsymbol{g}_{\ell j}$ can take on at most

$$U_\ell \coloneqq \binom{p^k + |C_\ell| - 1}{|C_\ell|} = \frac{(p^k + |C_\ell| - 1)!}{|C_\ell|!(p^k - 1)!}$$

different values $\boldsymbol{g}_{\ell j}^{(u)}$, $u = 1, \ldots, U_\ell$. Thus, the conditional probability of error given that $\boldsymbol{g}_{\ell j}^{(u)}$ is the correct sum leads with (4.32) to

$$\begin{aligned}\lambda_{\ell j}^{(u)} &\coloneqq \mathbb{P}\Big(\hat{\mathbf{g}}_{\ell j} \neq \boldsymbol{g}_{\ell j} \,\Big|\, \boldsymbol{g}_{\ell j} = \boldsymbol{g}_{\ell j}^{(u)}\Big) \\ &= \mathbb{P}\Big(\big[Q_{\Lambda_\mathrm{c}}(\boldsymbol{x}_{\ell j} + \mathbf{z}_{\ell j})\big] \bmod \Lambda_\mathrm{s} \neq \mathcal{E}_2(\boldsymbol{g}_{\ell j}) \,\Big|\, \boldsymbol{g}_{\ell j} = \boldsymbol{g}_{\ell j}^{(u)}\Big) \\ &= \mathbb{P}\left(\mathbf{z}_{\ell j} \notin \mathcal{V}_\mathrm{c}\right) \ . \end{aligned} \tag{4.35}$$

Observe that (4.35) is independent of u, which follows from the symmetry of the coding lattice Λ_c. Then, upper bounding the total probability as

$$\mathbb{P}\left(\hat{\boldsymbol{g}}_{\ell j} \neq \boldsymbol{g}_{\ell j}\right) \coloneqq \sum_{u=1}^{U_\ell} \lambda_{\ell j}^{(u)} \mathbb{P}\left(\boldsymbol{g}_{\ell j} = \boldsymbol{g}_{\ell j}^{(u)}\right) \leq \max_{1 \leq u \leq U_\ell} \lambda_{\ell j}^{(u)} = \mathbb{P}\left(\mathbf{z}_{\ell j} \notin \mathcal{V}_\mathrm{c}\right) \tag{4.36}$$

shows with (4.35) that the decoding error probability at FC ℓ, $\ell = 1, \ldots, L$, is essentially a maximum probability of error for each $j \in \{1, \ldots, 2N+1\}$. Since the receiver noise is iid across time and FCs, (4.36) is equal for all codewords so that (4.34) follows with Theorem B.2 of Appendix B (i.e., with the union bound).

Remark 4.9. Note that in the network model given in Section 4.1, we did not introduce a probability distribution on the sensor readings, which requires the decoding error probability to be small for every codeword and thus for every choice of $\{\varphi_{ij}(s_i) \in \Pi\}$. According to (4.34), this can be ensured because if $\mathbb{P}(\mathbf{z} \notin \mathcal{V}_\mathrm{c}) \leq \frac{\delta}{L(2N+1)}$, we have $P_e^{(n)} \leq \delta$, which justifies to consider (4.33) as a maximum probability of error.

4.3 Achievable Computation Rates

In the following, our objective is to characterize the computation rates achievable with the scheme proposed in the previous section. To gain first insights, we start with a single cluster network in Section 4.3.1, followed by the general case in Section 4.3.2. Towards this end, we first prove a lemma that determines the quantization error caused by the approximation of the desired function (4.4) by (4.21).

Lemma 4.4. *Let $(f_1, \ldots, f_L) \in \mathscr{C}^0(\mathcal{S}^{|C_1|}) \times \cdots \times \mathscr{C}^0(\mathcal{S}^{|C_L|})$ be some choice of L Kolmogorov's superpositions. Then, each f_ℓ can be uniformly approximated with arbitrary precision $\varepsilon > 0$ if the common quantizer (4.11) is configured with sufficiently large $b = b(f_1, \ldots, f_L, \|\cdot\|_\infty, N, \varepsilon)$. That is,*

$$\forall \varepsilon > 0 \, \exists b_0 \, \forall b \geq b_0 \, \forall \ell \in \{1, \ldots, L\} : \sup_{\boldsymbol{s}_\ell \in \mathcal{S}^{|C_\ell|}} \left| f_\ell(\boldsymbol{s}_\ell) - \tilde{f}_\ell(\boldsymbol{s}_\ell) \right| < \varepsilon \ ,$$

where $\boldsymbol{s}_\ell \coloneqq (s_{\ell_1}, \ldots, s_{\ell_{|C_\ell|}})$.

Proof. Observe that an expansion in the way of (4.13) represents along with (4.14) the pre-processed sensor readings up to precision

$$\left|\varphi_{ij}(s) - \tilde{\varphi}_{ij}(s)\right| < 2^{-\eta} = 2^{-b+v+1} \leq \pi_{\max} 2^{-b+1} \,, \tag{4.37}$$

for all $s \in \mathcal{S}$, $i = 1, \ldots, N$, and $j = 1, \ldots, 2N+1$. Hence, we can bound the accuracy of the sum of pre-processed sensor readings (4.17) by virtue of the triangle inequality to

$$\left|\sum_{i \in C_\ell} \varphi_{ij}(s_i) - \sum_{i \in C_\ell} \tilde{\varphi}_{ij}(s_i)\right| \leq \sum_{i \in C_\ell} \left|\varphi_{ij}(s_i) - \tilde{\varphi}_{ij}(s_i)\right| < |C_\ell| \pi_{\max} 2^{-b+1} \,, \tag{4.38}$$

for all $\boldsymbol{s}_\ell \in \mathcal{S}^{|C_\ell|}$, $\ell = 1, \ldots, L$, and $j = 1, \ldots, 2N+1$.

Since the constants in (4.5) are bounded, we conclude from (4.38) along with the Heine-Cantor Theorem (see Theorem A.2 in Appendix A) that[6]

$$\sup_{\boldsymbol{s}_\ell \in \mathcal{S}^{|C_\ell|}} \left|\psi_{\ell j}\left(\sum_{i \in C_\ell} \varphi_{ij}(s_i) + \gamma_{\ell j}\right) - \psi_{\ell j}\left(\sum_{i \in C_\ell} \tilde{\varphi}_{ij}(s_i) + \gamma_{\ell j}\right)\right| < \varepsilon_{\ell j}(b) \,, \tag{4.39}$$

for some $\varepsilon_{\ell j}(b) > 0$ and for all ℓ, j. Now, let $\varepsilon > 0$ be arbitrary but fixed. Then, there exists $b_0 = b_0(f_1, \ldots, f_L, \|\cdot\|_\infty, N, \varepsilon)$ such that for all $b \geq b_0$

$$\max_{\ell, j} \varepsilon_{\ell, j}(b) < \frac{\varepsilon}{2N+1} \,.$$

As a consequence, we have for all $\boldsymbol{s}_\ell \in \mathcal{S}^{|C_\ell|}$, $\ell = 1, \ldots, L$, and $b \geq b_0$

$$\left|f_\ell(\boldsymbol{s}_\ell) - \tilde{f}_\ell(\boldsymbol{s}_\ell)\right| \leq \sum_{j=1}^{2N+1} \left|\psi_{\ell j}\left(\sum_{i \in C_\ell} \varphi_{ij}(s_i) + \gamma_{\ell j}\right) - \psi_{\ell j}\left(\sum_{i \in C_\ell} \tilde{\varphi}_{ij}(s_i) + \gamma_{\ell j}\right)\right| < \varepsilon \,,$$

which proves the lemma. □

Remark 4.10. In words, the quantization parameter $b_0 = b_0(f_1, \ldots, f_L, \|\cdot\|_\infty, N, \varepsilon)$ denotes the smallest number of bits with which $f_1, \ldots, f_L$ can be represented within accuracy ε. We point out, however, that it is not a particular property of the scheme presented in Section 4.2 that b_0 generally also depends on the number of nodes. In fact, all computation schemes that approximate a real-valued multivariate function by quantizing its arguments suffer from this. Hence, we drop the corresponding indication in what follows.

[6] Because every finite sum of compact spaces is compact, it follows from the compactness of the Π_{ij} (i.e., the ranges of pre-processing functions) that the ranges of the sums $\sum_{i \in C_\ell} \varphi_{ij}(s_i)$, $\ell = 1, \ldots, L$; $j = 1, \ldots, 2N+1$, are compact as well.

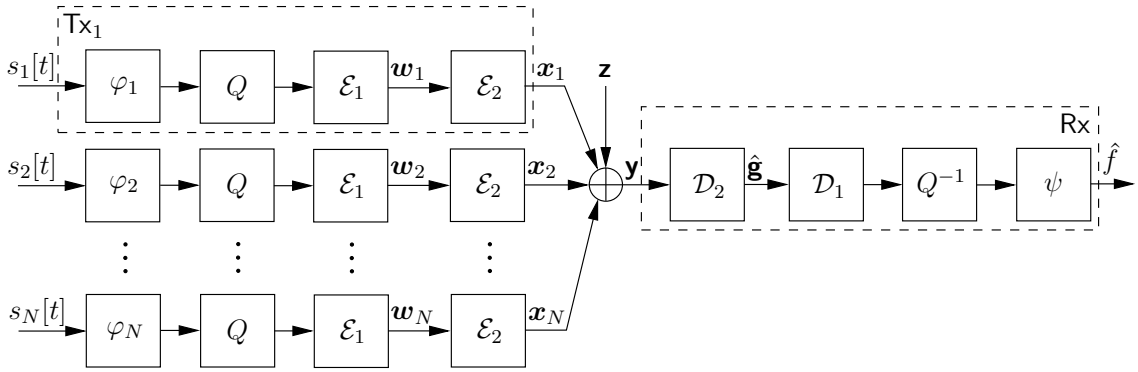


Figure 4.5: Block diagram of the entire transmission chain for computing a nomographic function over a Gaussian multiple-access channel.

Remark 4.11. Due to Remark 4.9, (4.7) represents a maximum error probability, which is therefore independent of the statistics of sensor readings. Because of the considerations in Lemma 4.4, it can therefore be written as

$$\mathbb{P}\left(\bigcup_{t=1}^{T}\left\{\hat{f}(\boldsymbol{s}[t]) \neq \tilde{f}(\boldsymbol{s}[t])\right\}\right) ,$$

with $\tilde{f}$ being the quantized version of f (see 4.21).

4.3.1 The Single Cluster Case

Consider a single cluster consisting of N nodes, which means that $L = 1$ and $|C_1| = N$.

Nomographic Functions

We start with the computation of a single nomographic function such as depicted in Figure 4.5. Hence, (4.30) simplifies to

$$\mathbf{y} = \sum_{i=1}^{N} \boldsymbol{x}_i + \mathbf{z} \tag{4.40}$$

whereas the decoding error probability (4.33) reduces to $P_e^{(n)} = \mathbb{P}(\hat{\mathbf{g}} \neq \boldsymbol{g})$.

The following theorem, which is an extension of [NDG11, Thm. 2], provides an achievable rate at which elements from $\mathscr{N}^0(\mathcal{S}^N)$ can be reliably computed through harnessing the interference in (4.40). Note that according to (4.19), the estimate $\hat{f}$ of some given $f \in \mathscr{N}^0(\mathcal{S}^N)$ is defined to be

$$\hat{f}(\boldsymbol{s}[t]) = (\psi \circ Q^{-1})(\hat{\mathbf{g}}[t]) \qquad t = 1, \ldots, T \, . \tag{4.41}$$

Theorem 4.1. *Given $f \in \mathscr{N}^0(\mathcal{S}^N)$, let $\hat{f}$ be its estimate defined by (4.41). Let $\varepsilon > 0$ be some given desired accuracy and $b_0(f, \|\cdot\|_\infty, \varepsilon)$ be specified as in Lemma 4.4. Then,*

$$R^{\mathrm{C}}(f, \|\cdot\|_\infty, \varepsilon) = \frac{\frac{1}{2}\log_2^+\left(\frac{P}{\sigma_Z^2}\right)}{b_0(f, \|\cdot\|_\infty, \varepsilon) + \log_2(N)} \tag{4.42}$$

is, under the distortion measure $\|\cdot\|_\infty$, an achievable computation rate for f and ε.

Proof. The proof is deferred to Appendix 4.A.3 at the end of the chapter. □

Remark 4.12. Note that in accordance with the proof of the theorem, (4.42) can even be slightly improved if the bound in (4.66) is applied instead of (4.67).

Proposition 3.1 immediately leads to the following corollary.

Corollary. *Let the computation accuracy, ε, be chosen such that $\varepsilon \geq \varepsilon'$ for some arbitrarily small $\varepsilon' > 0$. Then, Theorem 4.1 also applies to every nomographic approximation $f \in \mathscr{N}^0_{\|\cdot\|_\infty, \varepsilon'}(\mathcal{S}^N)$.*

The theorem reveals that if the channel is noisy, the number of required channel uses per function value depends on the function to be computed, which is in stark contrast to the results obtained for the ideal WMAC in Chapter 3. The obvious reason is that the nonzero noise variance forces us to perform quantization at the sensor nodes in order to represent function-values up to a certain accuracy by finite bit strings. In simple terms, "bad" functions require more bits for sufficiently accurate representations (i.e., $b_0(f, \|\cdot\|_\infty, \varepsilon)$ has to be larger) whereas "good" functions require less. In the following, we present some examples of functions that are reliably computable over the Gaussian MAC with rate (4.42).

Example 4.1 (Arithmetic Mean)**.** Let the desired function be the arithmetic mean $f(s_1, \ldots, s_N) = \frac{1}{N}\sum_{i=1}^N s_i$ (see Example 3.1), $\mathcal{S} = \mathbb{E}$, and $\varepsilon > 0$ some desired accuracy. Then, we conclude from (4.38)

$$\frac{1}{N}\left|\sum_{i=1}^N s_i - \sum_{i=1}^N \tilde{s}_i\right| \leq \frac{1}{N}\sum_{i=1}^N |s_i - \tilde{s}_i| < \pi_{\max} 2^{-b+1} = 2^{-b+1}\,,$$

where the last equality follows from the fact that the pre-processing functions are identities on $\mathbb{E}$ (i.e., $\pi_{\max} = \max_i \max_{s\in\mathbb{E}} |\varphi_i(s)| = \max_{s\in\mathbb{E}} \mathrm{id}_{\mathbb{E}}(s) = 1$). Therefore, in order to satisfy $2^{-b+1} \leq \varepsilon$, we require

$$b \geq b_0 = \left\lceil \log_2\left(\frac{1}{\varepsilon}\right)\right\rceil + 1 \quad \text{bits}\,.$$

△

Example 4.2 (Geometric Mean)**.** Let the desired function be the geometric mean $f(s_1,\dots,s_N) = \left(\prod_{i=1}^N s_i\right)^{1/N} = \exp_{\mathrm{e}}(\frac{1}{N}\sum_{i=1}^N \log_{\mathrm{e}}(s_i))$, $\mathcal{S} = [s_{\min},1]$ for some $0 < s_{\min} < 1$, and $\varepsilon > 0$. Note that f is an element of $\in \mathscr{N}^0([s_{\min},1]^N)$ but not of $\mathscr{N}^0(\mathbb{E}^N)$ (see Example 3.2). Now, let $g(\boldsymbol{s}) \coloneqq \frac{1}{N}\sum_{i=1}^N \log_{\mathrm{e}}(s_i)$ for some fixed $\boldsymbol{s} = (s_1,\dots,s_N) \in [s_{\min},1]^N$ and $\tilde{g}(\boldsymbol{s})$ the quantized version. Then, due to the mean value theorem there exists between every pair of points $g(\boldsymbol{s})$ and $\tilde{g}(\boldsymbol{s})$ a $\xi(\boldsymbol{s})$ such that with (4.37)

$$\left|\mathrm{e}^{g(\boldsymbol{s})} - \mathrm{e}^{\tilde{g}(\boldsymbol{s})}\right| = \mathrm{e}^{\xi(\boldsymbol{s})}|g(\boldsymbol{s}) - \tilde{g}(\boldsymbol{s})| < \mathrm{e}^{\xi(\boldsymbol{s})}\pi_{\max}2^{-b+1} \; . \tag{4.43}$$

Without loss of generality, let $\tilde{g} \leq \xi \leq g$ so that (4.43) can be upper bounded as

$$\begin{aligned}
\mathrm{e}^{\xi(\boldsymbol{s})}\pi_{\max}2^{-b+1} &\leq \mathrm{e}^{g(\boldsymbol{s})}\pi_{\max}2^{-b+1} \\
&= f(\boldsymbol{s})\pi_{\max}2^{-b+1} \\
&\leq \max_{\boldsymbol{s}\in[s_{\min},1]^N} f(\boldsymbol{s})\pi_{\max}2^{-b+1} \\
&= \pi_{\max}2^{-b+1} \\
&= \frac{1}{N}|\log_{\mathrm{e}}(s_{\min})|2^{-b+1} \; ,
\end{aligned} \tag{4.44}$$

where (4.44) follows from $\pi_{\max} = \max_{s\in[s_{\min},1]} |\log_{\mathrm{e}}(s)| = |\log_{\mathrm{e}}(s_{\min})|$. With regard to this bound, we require

$$b \geq b_0 = \left\lceil \log_2\left(\frac{|\log_{\mathrm{e}}(s_{\min})|}{N\varepsilon}\right)\right\rceil + 1 \quad \text{bits}$$

in order to represent the geometric mean within accuracy ε. △

Example 4.3 (Euclidean Norm)**.** Let the desired function be the Euclidean norm $f(s_1,\dots,s_K) = \sqrt{s_1^2 + \cdots + s_N^2}$ (see Example 3.1), $\mathcal{S} = \mathbb{E}$, and $\varepsilon > 0$. Furthermore, let $g(\boldsymbol{s}) \coloneqq \sum_{i=1}^N s_i^2$ for some fixed $\boldsymbol{s} = (s_1,\dots,s_N) \in \mathbb{E}^N$ and $\tilde{g}(\boldsymbol{s})$ the corresponding quantized version. It is a matter of fact that the square root is a $(1,1/2)$-Hölder continuous function over $\mathbb{R}_+$ (see Definition A.8 along with Example A.2 for the definition of Hölder continuity) so that

$$\left|\sqrt{g(\boldsymbol{s})} - \sqrt{\tilde{g}(\boldsymbol{s})}\right| \leq \sqrt{|g(\boldsymbol{s}) - \tilde{g}(\boldsymbol{s})|} \leq \left(\sum_{i=1}^N |\varphi(s_i) - \tilde{\varphi}(s_i)|\right)^{1/2} < \sqrt{N\pi_{\max}}2^{-(b-1)/2} \; .$$

Since $\pi_{\max} = \max_{s\in\mathbb{E}} s^2 = 1$, we require

$$b \geq b_0 = \left\lceil \log_2\left(\frac{N}{\varepsilon^2}\right)\right\rceil + 1 \quad \text{bits}$$

in order to represent the Euclidean norm within accuracy ε. △

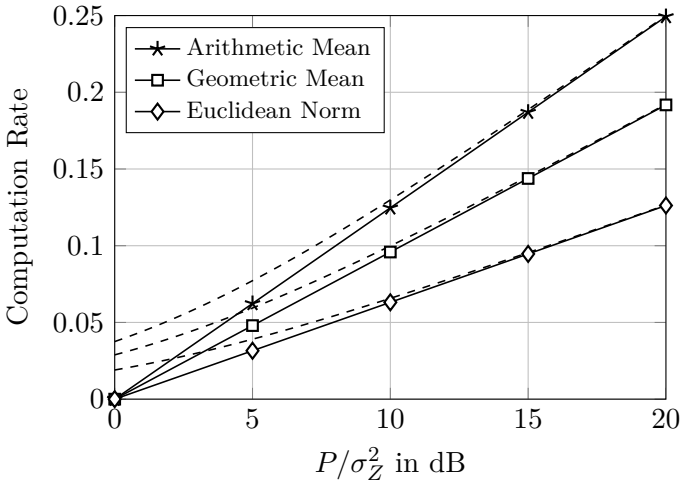


Figure 4.6: Achievable computation rates in a cluster with $N = 5$ nodes for the nomographic functions of Examples 4.1–4.3, where the aimed computation accuracy is set to $\varepsilon = 10^{-3}$. The dashed upper bounds correspond to (4.47).

Figure 4.6 depicts for $N = 5$, $s_{\min} = 10^{-20}$, and $\varepsilon = 10^{-3}$ the achievable computation rates of Examples 4.1–4.3 (i.e., (4.42)). It turns out, for instance, that at a signal-to-noise ratio (SNR) of 15 dB, the "arithmetic mean" can be computed approximately 1.3 times faster than the "geometric mean" and approximately 2 times faster than the "Euclidean norm", respectively.

Consider now the standard separation-based computation approach in which the FC reliably decodes all quantized sensor readings individually from the Gaussian MAC output in order to compute the desired function-values afterwards. Then, the corresponding rate performance is limited by the MAC capacity region, which is for the general Gaussian MAC (see Remark 2.2) given by the following fundamental theorem [GK11, p. 98].

Theorem 4.2. *The capacity region of the real-valued Gaussian MAC with N users consists of the set of all rate tuples $(R_1, \ldots, R_N) \in \mathbb{R}_+^N$ for which*

$$\forall \mathcal{I} \subseteq \{1, \ldots, N\} : \sum_{i \in \mathcal{I}} R_i \leq \frac{1}{2} \log_2 \left(1 + \frac{\sum_{i \in \mathcal{I}} |h_i|^2 P_i}{\sigma_Z^2} \right) \quad \frac{\text{bits}}{\text{channel use}} , \tag{4.45}$$

where $P_i \in \mathbb{R}_+$, $|h_i|^2$, and $\sigma_Z^2 > 0$ denote the transmit power constraint of user i, the channel gain between user i and the receiver, and the noise variance, respectively.

The region (4.45) forms a polymatroid in Euclidean space $\mathbb{R}^N$ with $N!$ vertices [TH98]. Each vertex can be achieved by using Gaussian single-user codebooks in combination with successive cancellation decoding that is performed in the decoding order that corresponds to the preferred vertex.

As a special case of Theorem 4.2, the capacity region of the Gaussian MAC considered in this section (see (4.40)) follows to

$$\forall \mathcal{I} \subseteq \{1, \dots, N\} : \sum_{i \in \mathcal{I}} R_i \leq \frac{1}{2} \log_2 \left(1 + \frac{|\mathcal{I}| P}{\sigma_Z^2} \right) ,$$

where we have taken into account that the transmit power constraint is uniformly P. From this, we conclude that the best computation rate achievable with any separation-based approach is limited to

$$R^{\mathrm{C}}(f, \|\cdot\|_\infty, \varepsilon) = \frac{\frac{1}{2N} \log_2 \left(1 + \frac{NP}{\sigma_Z^2} \right)}{b_0(f, \|\cdot\|_\infty, \varepsilon)} . \tag{4.46}$$

Comparing (4.46) with (4.42) reveals that many linear and nonlinear functions of sensor readings can be reliably computed, with the coding scheme of Section 4.2, at a rate that is significantly higher than every rate achievable with separation, except for small SNRs (see Section 4.3.3 for a more detailed discussion).

It is easy to see that (4.42), for $P/\sigma_Z^2 \to \infty$, achieves an upper bound given by the normalized single-user AWGN capacity:

$$\bar{R}^{\mathrm{C}}(f, \|\cdot\|_\infty, \varepsilon) \coloneqq \frac{\frac{1}{2} \log_2 \left(1 + \frac{P}{\sigma_Z^2} \right)}{b_0(f, \|\cdot\|_\infty, \varepsilon) + \log_2(N)} . \tag{4.47}$$

Up to the writing of this thesis, however, it was unknown whether this bound can also be achieved for finite SNRs. See Figure 4.7 for an example.

Remark 4.13. The additional logarithmic term in the denominators of (4.42) and (4.47) is the penalty for avoiding wraparounds in the modulo p addition of the messages (see Remark 4.4).

Kolmogorov's Superpositions

Although Examples 4.1–4.3 demonstrate that $\mathscr{N}^0(\mathcal{S}^N)$ contains many functions of practical relevance, it has to be emphasized that by Theorem 3.3, $\mathscr{N}^0(\mathcal{S}^N)$ is a nowhere dense subset of all continuous functions. By Theorem 3.4, however, every continuous function of N variables can be composed of $2N + 1$ elements from $\mathscr{N}^0(\mathcal{S}^N)$. Now, we use this fact in order to provide the computation rate that is achievable for reliably computing Kolmogorov's superpositions with the scheme depicted in Figures 4.1 and 4.2. Given some $f \in \mathscr{C}^0(\mathcal{S}^N)$, the corresponding estimate is of the form

$$\hat{f}(\boldsymbol{s}[t]) = \sum_{j=1}^{2N+1} (\psi_j \circ Q^{-1})(\hat{\mathbf{g}}_j[t]) \qquad t = 1, \dots, T . \tag{4.48}$$

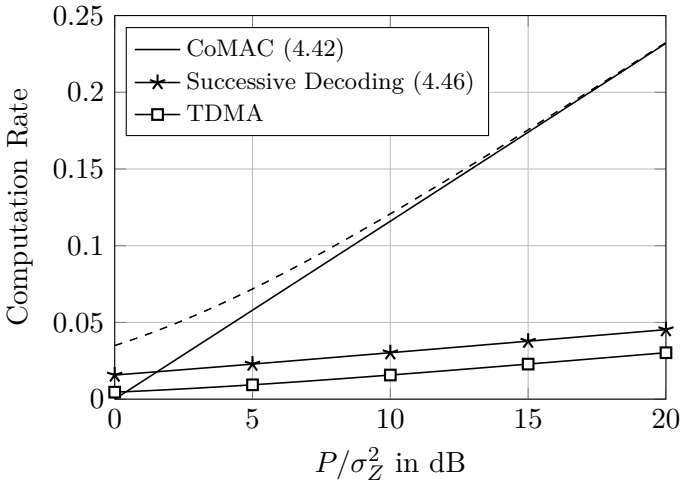


Figure 4.7: Achievable computation rates in a cluster with $N = 10$ nodes where the FC wants to evaluate the arithmetic mean (see Example 4.1) within accuracy $\varepsilon = 10^{-3}$, which requires $b_0(f, \|\cdot\|_\infty, \varepsilon) = 11$ bits. The dashed upper bound represents the single-user AWGN capacity normalized by $11 + \log_2(10)$ whereas TDMA refers to naive time-sharing between nodes [GK11, p. 96].

Theorem 4.3. *Given $f \in \mathscr{C}^0(\mathcal{S}^N)$, let $\hat{f}$ its estimate defined by (4.48). Let $\varepsilon > 0$ be some given desired accuracy and $b_0(f, \|\cdot\|_\infty, \varepsilon)$ be specified as in Lemma 4.4. Then,*

$$R^{\mathrm{C}}(f, \|\cdot\|_\infty, \varepsilon) = \frac{\frac{1}{4N+2} \log_2^+\left(\frac{P}{\sigma_Z^2}\right)}{b_0(f, \|\cdot\|_\infty, \varepsilon) + \log_2(N)} \tag{4.49}$$

is, under the distortion measure $\|\cdot\|_\infty$, an achievable computation rate for f and ε.

Proof. Representing f as its Kolmogorov's superposition (see Theorem 3.4) suggests that it can be computed at the FC by successively computing the corresponding $2N + 1$ nomographic functions over the Gaussian MAC. Hence, given some fixed $\varepsilon > 0$, choose $b = b_0(f, \|\cdot\|_\infty, \varepsilon)$ in accordance with Lemma 4.4 sufficiently large such that the quantization error is smaller than ε. Now, due to (4.34), we have for the decoding error probability

$$P_e^{(n)} \leq \sum_{j=1}^{2N+1} \mathbb{P}(\hat{\mathbf{g}}_j \neq \boldsymbol{g}_j) \,.$$

Therefore, the theorem follows from Theorem 4.1 by taking into account that for each j, $\mathbb{P}(\hat{\mathbf{g}}_j \neq \boldsymbol{g}_j)$ goes to zero exponentially fast in the block length n as long as (4.49) is fulfilled. □

Remark 4.14. Comparing (4.49) with (4.42) illustrates that when harnessing the superposition property of the Gaussian MAC, universality with respect to the number of

computable functions and the pre-processing strategy costs additional wireless resources. See Section 4.3.3 for a more detailed discussion.

4.3.2 The Multiple Cluster Case

Consider now the general network model introduced in Section 4.1, in which N sensor nodes are divided into L overlapping clusters C_ℓ with the objective to compute at FC ℓ, $\ell = 1, \ldots, L$, some desired function $f_\ell \in \mathscr{C}^0(\mathbb{E}^{|C_\ell|})$ of the associated sensor readings.

Nomographic Functions

From Section 3.2.1 we know that when restricted to nomographic functions with continuous pre- and post-processing functions, the pre-processing functions can never be chosen to be universal. This is due to the fact that each cluster overlaps with at least one of its neighboring clusters. Therefore, the clusters have to be activated in a time-division manner whenever the functions to be computed at adjacent FCs are different. As a consequence, the average computation rate achievable in cluster ℓ, $\ell = 1, \ldots, L$, under a naive time-division strategy that schedules clusters in time follows from Lemma 4.4 and Theorem 4.1 to[7]

$$R_\ell^{\mathrm{C}}(f_1, \ldots, f_L, \|\cdot\|_\infty, \varepsilon) = \frac{\frac{1}{2L}\log_2^+\left(\frac{P}{\sigma_Z^2}\right)}{b_0(f_1, \ldots, f_L, \|\cdot\|_\infty, \varepsilon) + \log_2\big(\max_\ell |C_\ell|\big)} \,. \tag{4.50}$$

In contrast to the single cluster case, the rates depend on $f_1, \ldots, f_L$ as the error probability (4.7) extends in the multi-cluster case to

$$\mathbb{P}\left(\bigcup_{\ell=1}^{L}\bigcup_{t=1}^{T}\left\{\sup_{\boldsymbol{s}_\ell[t]\in\mathbb{E}^{|C_\ell|}}\left|\hat{f}_\ell(\boldsymbol{s}_\ell[t]) - f_\ell(\boldsymbol{s}_\ell[t])\right| > \varepsilon\right\}\right) \,.$$

This means that the computation accuracy has to be within ε for all ℓ so that b_0 depends on $f_1, \ldots, f_L$ (see Remark 4.10).

Following a similar reasoning for a separation-based approach results in the achievable computation rate

$$R_\ell^{\mathrm{C}}(f_1, \ldots, f_L, \|\cdot\|_\infty, \varepsilon) = \frac{\frac{1}{2L|C_\ell|}\log_2\left(1 + \frac{|C_\ell|P}{\sigma_Z^2}\right)}{b_0(f_1, \ldots, f_L, \|\cdot\|_\infty, \varepsilon)} \qquad \ell = 1, \ldots, L \,, \tag{4.51}$$

which is significantly smaller than (4.50), except for small ratios P/σ_Z^2.

Remark 4.15. Observe that the rate in (4.50) is independent of ℓ and therefore equal for all clusters.

[7] It is assumed that the time is divided into L slots of equal duration.

Remark 4.16. In clustered networks in which $C_\ell \cap C_{\ell'} \neq \varnothing$ for all ℓ, ℓ', (4.50) and (4.51) cannot be increased by, for example, more clever time-sharing. The reason is that the common nodes transmit continuously and would therefore violate the average power constraint when increasing their transmit powers by a factor of L. If, on the other hand, some of the clusters are disjoint (see Figure 3.4 for an example), the rates could be improved by designing a time-division protocol that activates these clusters simultaneously. This, however, would further increase the coordination effort and is part of future work.

Kolmogorov's Superpositions

Now, let us consider the general case of computing arbitrary continuous functions over arbitrary clustered Gaussian sensor networks. Since the pre-processing functions in (4.21) are universal and therefore independent of f_ℓ, the function that FC ℓ, $\ell = 1, \ldots, L$, computes is determined by the choice of the post-processing functions $\psi_{\ell 1}, \ldots, \psi_{\ell, 2N+1}$ only. As a consequence, the pre-processing and lattice encoding is fixed, and therefore an additional protocol for coordinating the activation of clusters, as it was required for achieving (4.50) and (4.51), is not necessary. Therefore, the computation rate achievable with the scheme of Section 4.2 follows from Theorem 4.3 to

$$R_\ell^{\mathrm{C}}(f_1, \ldots, f_L, \|\cdot\|_\infty, \varepsilon) = \frac{\frac{1}{4N+2} \log_2^+\left(\frac{P}{\sigma_Z^2}\right)}{b_0(f_1, \ldots, f_L, \|\cdot\|_\infty, \varepsilon) + \log_2\left(\max_\ell |C_\ell|\right)}\,, \tag{4.52}$$

for all $\ell = 1, \ldots, L$.

4.3.3 Discussion of the Results

The results for the single cluster case in Section 4.3.1 show that when harnessing the superposition property of the Gaussian MAC, nomographic functions with continuous pre- and post-processing functions can be computed significantly faster than with any separation-based strategy, except for small SNRs (see Figure 4.7 for an example). On the other hand, when considering the computation of arbitrary continuous functions of the sensor readings, the corresponding computation rates scale down by a factor of $2N + 1$ (see (4.49)).

In a network of multiple clusters, as considered in Section 4.3.2, the computation rate achievable at all nodes when considering an individual continuous nomographic function in each cluster is reduced by a factor of L (see (4.50) and (4.51)) since additional coordination is necessary in the form of time sharing between clusters. In contrast, due to the universality of pre-processing functions and the particular data post-processing strategy described in Section 4.2.2, the rate at which a different Kolmogorov's superposition can be computed in each cluster is given by Theorem 4.3, *regardless* of the coupling between clusters.

In the domain of wireless sensor networks, achieving high rates is generally not the only concern. Due to limited energy and processing capabilities, computation schemes

of low complexity are also of particular interest. Considering the results of Section 4.3.2 from this perspective reveals that the proposed computation scheme has not only in the case of continuous nomographic functions several advantages over separation-based approaches. For example, when computing a set of individual Kolmogorov's superpositions in a clustered network, any coordination of nodes or clusters is not necessary as it would be the case for continuous nomographic functions and separation-based approaches. Especially for large networks with many clusters this may lead to significant savings in complexity such that computing Kolmogorov's superpositions (i.e., continuous functions) over the channel can be an option even if the achievable computation rate is not maximal. It is clear from the structure of nomographic functions, and therefore Kolmogorov's superpositions, that the computation of only one-variable functions is required at the FCs, which can be less demanding than computing the multivariate desired function given the entire set of raw sensor readings such as in the case of separation based computation.

If the underlying application is satisfied with the computation of continuous nomographic functions, then in addition to the superior rate performance, the scheme proposed in this chapter has a significantly lower decoding complexity, which is essentially the complexity of a single-user lattice decoder. As a consequence, the decoding complexity in cluster ℓ, $\ell = 1, \ldots, L$, is $|C_\ell|$-fold less than for separation-based computation in which the FC has to reliably decode all the sensor readings gathered in cluster C_ℓ. The latter has also the drawback of a higher sensitivity regarding decoding errors since already a single wrongly decoded sensor reading results in a faulty function-value. We would like to emphasize that the computation rates presented in this chapter are all achievable under a maximum probability of error criterion as it is indispensable for most sensor network applications.

When additionally using common randomness at sensor nodes and FCs in combination with minimum mean square error estimation prior to decoding [NG11a, WNPS10, NA12], slightly higher computation rates could be achieved than those presented in Theorems 4.1 and 4.3. However, this would have the drawback that only average error probabilities could be handled together with uniformly distributed sensor readings.

4.4 Summary and Conclusions

In this chapter, we considered the reliable computation of arbitrary continuous real-valued functions of the measurements in clustered Gaussian sensor networks. Towards this end, we proposed a coding scheme that combines a suitable data pre- and post-processing strategy with a simple quantizer and nested lattice coding. Such as in Chapter 3, the pre- and post-processing strategy matches the desired function to the algebraic structure of the channel whereas the remaining components are concerned with the Gaussian receiver noise. Based on this scheme, it has been found that when harnessing the interference property of the underlying Gaussian MAC, a certain subset of all continuous functions (i.e., the space of nomographic functions with continuous pre-

and post-processing functions) can be computed at considerably higher rates than those achievable with an approach that intends to decode all associated sensor readings at the FCs for computing the function-values afterwards. Since many continuous functions of practical relevance are nomographic, the result extends the known results for the computation of linear functions to numerous nonlinear functions.

When the computation of *arbitrary* continuous functions is desired, then the presented approach requires the successive computation of multiple nomographic functions, which scales down the achievable computation rates accordingly. Even though these rates can be inferior to those achievable with standard multiple-access schemes, the proposed approach provides several other advantages that are indispensable in many sensor network applications such as lower decoding complexity, less coordination and the ability of controlling maximum error probabilities. As a consequence, the results of this chapter partially carry over the results of Chapter 3 to noisy networks.

Note that the clustered Gaussian sensor network model considered in this paper assumes that the channel gains between nodes and FCs are all equal to one (see (2.1)). For many applications, however, the propagation conditions are more challenging as corresponding wireless transmissions may be subject to fading effects. In networks with non-overlapping clusters this is not a big issue as nodes could invert their channels by employing channel state information. In contrast, when clusters allowed to overlap some nodes can be heard by more than one FC, which generally results in different channel gains. Since this can have a detrimental impact on the computation rate performance, it has to be figured out in future work how to appropriately cope with this.

Another direction for future work could be the following. Note that the computation capacity such as defined in Definition 4.3 refers, for some given f and $(d_{\mathbb{R},\varepsilon})$, to the supremum of all achievable rates of a given WMAC W. On the other hand, one could think of defining it with respect to some given function space (e.g., $\mathscr{C}^0(\mathcal{S}^N)$, $\mathscr{N}^0(\mathcal{S}^N)$) in the sense that it denotes, for given $(d_{\mathbb{R},\varepsilon})$, the supremum of all achievable rates at which *every function* of that function space can be reliably computed. This could somehow be related to Kolmogorov's notion of ε-entropy, which is defined as follows [Lor66, pp. 150].

Definition 4.12 (ϵ-Entropy)**.** Let $\varepsilon > 0$ and $\mathscr{G} \subset \mathscr{X}$ be an arbitrary function space that is compact with respect to some fixed norm $\|\cdot\|$ so that there exists an ε-net for $\mathscr{G}$, that is, a finite set of functions $\mathfrak{N}_{\mathscr{G},\varepsilon} = \{\hat{g} \in \mathscr{X}\}$ such that

$$\sup_{g\in\mathscr{G}} \min_{\hat{g}\in\mathfrak{N}_{\mathscr{G},\varepsilon}} \|g - \hat{g}\| \leq \varepsilon \ .$$

Let $N_\varepsilon(\mathscr{G})$ denote the smallest such ε-net. Then, the *Kolmogorov ε-entropy* for $\mathscr{G}$ is defined as

$$H_\varepsilon(\mathscr{G}) = \log_2 N_\varepsilon(\mathscr{G}) \ .$$

The ε-entropy therefore denotes the least number of bits that are required in order to describe every $g \in \mathscr{G}$ within accuracy $\varepsilon > 0$. Since the simple quantizer employed in the coding scheme of Section 4.2 has to be adapted every time a desired function in the

network changes, it would be interesting to consider universal quantizers. It appears likely that the ε-entropy could provide corresponding bounds on the performance of such quantizers as well as to the mentioned modified notion of computation capacity.

Appendix 4.A Proofs

4.A.1 Proof of Lemma 4.1

Let $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}_+$ be arbitrary and let $\Lambda \subset \Lambda' \subset \mathbb{R}^n$ be some fixed pair of nested lattices.

a):

$$\begin{aligned}
[\boldsymbol{\mu} + \boldsymbol{\nu}] \bmod \Lambda &= \boldsymbol{\mu} - Q_\Lambda(\boldsymbol{\mu}) + \boldsymbol{\nu} - Q_\Lambda(\boldsymbol{\nu}) - Q_\Lambda(\boldsymbol{\mu} + \boldsymbol{\nu}) + Q_\Lambda(\boldsymbol{\mu}) + Q_\Lambda(\boldsymbol{\nu}) \\
&= [\boldsymbol{\mu}] \bmod \Lambda + [\boldsymbol{\nu}] \bmod \Lambda - Q_\Lambda(\boldsymbol{\mu} - Q_\Lambda(\boldsymbol{\mu}) + \boldsymbol{\nu} - Q_\Lambda(\boldsymbol{\nu})) \qquad (4.53) \\
&= [\boldsymbol{\mu}] \bmod \Lambda + [\boldsymbol{\nu}] \bmod \Lambda - Q_\Lambda([\boldsymbol{\mu}] \bmod \Lambda + [\boldsymbol{\nu}] \bmod \Lambda) \\
&= \big[[\boldsymbol{\mu}] \bmod \Lambda + [\boldsymbol{\nu}] \bmod \Lambda\big] \bmod \Lambda \,,
\end{aligned}$$

where (4.53) follows from the fact that $Q_\Lambda(\boldsymbol{\mu})$ and $Q_\Lambda(\boldsymbol{\nu})$ are points in Λ.

b):

$$\begin{aligned}
[Q_{\Lambda'}(\boldsymbol{\mu})] \bmod \Lambda &= \big[Q_{\Lambda'}(\boldsymbol{\mu}) - Q_\Lambda(\boldsymbol{\mu})\big] \bmod \Lambda \qquad (4.54) \\
&= \big[Q_{\Lambda'}(\boldsymbol{\mu} - Q_\Lambda(\boldsymbol{\mu}))\big] \bmod \Lambda \qquad (4.55) \\
&= \big[Q_{\Lambda'}([\boldsymbol{\mu}] \bmod \Lambda)\big] \bmod \Lambda \,,
\end{aligned}$$

where (4.54) follows from a) along with the fact that $Q_\Lambda(\boldsymbol{\mu}) \in \Lambda$ and (4.55) because $\Lambda \subset \Lambda'$ implies $Q_\Lambda(\boldsymbol{\mu}) \in \Lambda'$.

c):

$$\begin{aligned}
\alpha[\boldsymbol{\mu}] \bmod \Lambda &= \alpha\boldsymbol{\mu} - \alpha Q_\Lambda(\boldsymbol{\mu}) \\
&= \alpha\boldsymbol{\mu} - \alpha \operatorname*{argmin}_{\boldsymbol{\lambda} \in \Lambda} \|\boldsymbol{\mu} - \boldsymbol{\lambda}\|_2 \\
&= \alpha\boldsymbol{\mu} - \operatorname*{argmin}_{\boldsymbol{\lambda} \in \Lambda} \|\alpha\boldsymbol{\mu} - \alpha\boldsymbol{\lambda}\|_2 \qquad (4.56) \\
&= \alpha\boldsymbol{\mu} - \operatorname*{argmin}_{\boldsymbol{\lambda} \in \alpha\Lambda} \|\alpha\boldsymbol{\mu} - \boldsymbol{\lambda}\|_2 \\
&= \alpha\boldsymbol{\mu} - Q_{\alpha\Lambda}(\alpha\boldsymbol{\mu}) \\
&= [\alpha\boldsymbol{\mu}] \bmod \alpha\Lambda \,,
\end{aligned}$$

where (4.56) follows from the absolute homogeneity of $\|\cdot\|_2$.

4.A.2 Proof of Lemma 4.3

The proof by Erez and Zamir in [EZ04] on the existence of sequences of good nested lattices (good in the sense summarized in Lemma 4.2) is based on a lattice construction known as *Construction A* [CS10,Loe97]. We follow this construction in order to generate a lattice codebook that is based on a good pair of nested lattices and then show that the corresponding encoder satisfies (4.29).

Towards this end, let p be prime, $\ell \in \{1, \ldots, L\}$, and $n, k, T \in \mathbb{N}$ be arbitrary but fixed. Furthermore, let $\boldsymbol{G}_\mathrm{s} \in \mathbb{R}^{n \times n}$ denote the full-rank generator matrix of a given

(good) shaping lattice (i.e., $\Lambda_\mathrm{s} = \boldsymbol{G}_\mathrm{s}\mathbb{Z}^n$). Then, Construction A mainly consists in generating a codebook over $\mathbb{Z}_p$ at random that is subsequently embedded into the Euclidean space $\mathbb{R}^n$. The construction works as follows:

1) Draw a matrix $\boldsymbol{G}_\mathrm{c} \in \mathbb{Z}_p^{n\times k}$ at random with elements uniformly iid over $\mathbb{Z}_p$.

2) Use $\boldsymbol{G}_\mathrm{c}$ to generate a linear codebook over $\mathbb{Z}_p$ as follows:

$$\tilde{\mathcal{C}}^{(n)} \coloneqq \left\{ \tilde{\boldsymbol{x}} = \boldsymbol{G}_\mathrm{c}\boldsymbol{w} \,|\, \boldsymbol{w} \in \mathbb{Z}_p^k \right\} ,$$

   where all operations are carried out modulo p.

3) Embed $\tilde{\mathcal{C}}^{(n)}$ into $\mathbb{R}^n$ by

$$\tilde{\Lambda}_\mathrm{c} \coloneqq p^{-1}\tilde{\mathcal{C}}^{(n)} + \mathbb{Z}^n .$$

4) Generate the coding lattice Λ_c by rotating $\tilde{\Lambda}_\mathrm{c}$ with $\boldsymbol{G}_\mathrm{s}$ so that $\Lambda_\mathrm{s} \subset \Lambda_\mathrm{c}$:

$$\Lambda_\mathrm{c} \coloneqq \boldsymbol{G}_\mathrm{s}\tilde{\Lambda}_\mathrm{c} .$$

5) Define the nested lattice codebook:

$$\mathcal{C}^{(n)} \coloneqq \Lambda_\mathrm{c} \cap \mathcal{V}_\mathrm{s} .$$

According to this procedure, the corresponding nested lattice encoder is of the form

$$\mathcal{E}_2 : \mathbb{Z}_p^k \to \mathcal{C}^{(n)} \ , \ \boldsymbol{w} \mapsto \mathcal{E}(\boldsymbol{w}) = \left[p^{-1}\boldsymbol{G}_\mathrm{s}\boldsymbol{G}_\mathrm{c}\boldsymbol{w}\right] \bmod \Lambda_\mathrm{s} . \tag{4.57}$$

In order to proof the properties of (4.57) stated in the lemma, we focus on showing (4.29) while skipping the bijectivity part (see [NG11a, Lem. 5] for a proof). In this context, let

$$\boldsymbol{x} \coloneqq \left[\sum_{i\in C_\ell} \boldsymbol{x}_i \right] \bmod \Lambda_\mathrm{s} ,$$

with $\boldsymbol{x}_i = \mathcal{E}(\boldsymbol{w}_i) \in \mathcal{C}^{(n)}$. Observe that $\boldsymbol{G}_\mathrm{s}^{-1}\Lambda_\mathrm{s} = \mathbb{Z}^n$ due to the fact that $\boldsymbol{G}_\mathrm{s}$ (i.e., the generator matrix of Λ_s) is of full rank. Thus, we have

$$\begin{aligned}
\left[\boldsymbol{G}_\mathrm{s}^{-1}\boldsymbol{x}\right] \bmod \mathbb{Z}^n &= \left[\boldsymbol{G}_\mathrm{s}^{-1} \sum_{i\in C_\ell} \boldsymbol{x}_i - \boldsymbol{G}_\mathrm{s}^{-1} Q_{\Lambda_\mathrm{s}} \left(\sum_{i\in C_\ell} \boldsymbol{x}_i \right) \right] \bmod \mathbb{Z}^n && (4.58)\\
&= \left[\boldsymbol{G}_\mathrm{s}^{-1} \sum_{i\in C_\ell} \boldsymbol{x}_i \right] \bmod \mathbb{Z}^n && (4.59)\\
&= \left[p^{-1}\boldsymbol{G}_\mathrm{c} \sum_{i\in C_\ell} \boldsymbol{w}_i - \boldsymbol{G}_\mathrm{s}^{-1} Q_{\Lambda_\mathrm{s}} \left(p^{-1}\boldsymbol{G}_\mathrm{s}\boldsymbol{G}_\mathrm{c}\boldsymbol{w}_i \right) \right] \bmod \mathbb{Z}^n && (4.60)\\
&= \left[p^{-1}\boldsymbol{G}_\mathrm{c} \sum_{i\in C_\ell} \boldsymbol{w}_i \right] \bmod \mathbb{Z}^n , && (4.61)
\end{aligned}$$

where (4.59) and (4.61) follow from the fact that

$$\forall \boldsymbol{\mu} \in \mathbb{R}^n : Q_{\Lambda_\mathrm{s}}(\boldsymbol{\mu}) \in \Lambda_\mathrm{s} \Rightarrow \boldsymbol{G}_\mathrm{s}^{-1} Q_{\Lambda_\mathrm{s}}(\boldsymbol{\mu}) \in \mathbb{Z}^n$$

and (4.60) by incorporating (4.57). Now, observe that

$$\forall \boldsymbol{w}_1, \ldots, \boldsymbol{w}_{|C_\ell|} \in \mathbb{Z}_p^k \, \exists \boldsymbol{v} \in \mathbb{Z}_+^n : \boldsymbol{G}_\mathrm{c} \sum_{i \in C_\ell} \boldsymbol{w}_i = \boldsymbol{G}_\mathrm{c} \bigoplus_{i \in C_\ell} \boldsymbol{w}_i + p\boldsymbol{v} \; ,$$

due to the modulo p arithmetic, so that multiplying (4.58) by p results, along with part c) of Lemma 4.1, in

$$p \left[\boldsymbol{G}_\mathrm{s}^{-1} \boldsymbol{x} \right] \bmod \mathbb{Z}^n = \left[\boldsymbol{G}_\mathrm{c} \bigoplus_{i \in C_\ell} \boldsymbol{w}_i + p\boldsymbol{v} \right] \bmod p\mathbb{Z}^n = \boldsymbol{G}_\mathrm{c} \bigoplus_{i \in C_\ell} \boldsymbol{w}_i \; . \tag{4.62}$$

Erez, Litsyn and Zamir show in [ELZ05, Sec. III] that matrices resulting from step 1) of Construction A are, for sufficiently large n, of full rank with probability that is close to one. Therefore, assume $\boldsymbol{G}_\mathrm{c}$ is full-rank so that its left inverse

$$\boldsymbol{G}_\mathrm{c}^\dagger \coloneqq \left(\boldsymbol{G}_\mathrm{c}^\mathsf{T} \boldsymbol{G}_\mathrm{c} \right)^{-1} \boldsymbol{G}_\mathrm{c}^\mathsf{T}$$

exists. Multiplying the right-hand side of (4.62) by $\boldsymbol{G}_\mathrm{c}^\dagger$ yields

$$\boldsymbol{G}_\mathrm{c}^\dagger \boldsymbol{G}_\mathrm{c} \bigoplus_{i \in C_\ell} \boldsymbol{w}_i = \bigoplus_{i \in C_\ell} \boldsymbol{w}_i \; .$$

As a final step, we merely have to put all the above pieces together in order to obtain the inverse

$$\mathcal{E}^{-1}(\boldsymbol{x}) = p \boldsymbol{G}_\mathrm{c}^\dagger \left[\boldsymbol{G}_\mathrm{s}^{-1} \boldsymbol{x} \right] \bmod \mathbb{Z}^n \; ,$$

which proves that (4.57) preserves linearity and concludes the lemma.

4.A.3 Proof of Theorem 4.1

For some $T \in \mathbb{N}$ to be specified below, consider some sequence

$$f\big(\boldsymbol{s}[1]\big) = \psi \left(\sum_{i=1}^N \varphi_i\big(s_i[1]\big) \right), \ldots, f\big(\boldsymbol{s}[T]\big) = \psi \left(\sum_{i=1}^N \varphi_i\big(s_i[T]\big) \right)$$

of nomographic function-values with continuous pre- and post-processing functions and let

$$\tilde{f}\big(\boldsymbol{s}[1]\big) = \psi \left(\sum_{i=1}^N \tilde{\varphi}_i\big(s_i[1]\big) \right), \ldots, \tilde{f}\big(\boldsymbol{s}[T]\big) = \psi \left(\sum_{i=1}^N \tilde{\varphi}_i\big(s_i[T]\big) \right)$$

denote the corresponding approximations in accordance with (4.17). Let $\varepsilon > 0$ and and choose the quantization parameter $b = b(f, \|\cdot\|_\infty, \varepsilon)$ as in Lemma 4.4 to $b_0 = b_0(f, \|\cdot\|_\infty, \varepsilon)$ so that

$$\left\| f - \tilde{f} \right\|_\infty = \sup_{\boldsymbol{s} \in \mathcal{S}^N} \left| f(\boldsymbol{s}) - \tilde{f}(\boldsymbol{s}) \right| < \varepsilon \; .$$

In order to prove the theorem, we have to first construct the source encoder (4.15). To this end, each of the binary representations (4.12) is equivalently considered as an element of the set of integers $\{0, 1, \ldots, 2^{b_0} - 1\}$, which we denote in the following as $w_i[t]$ to avoid confusion with the vector notation. With this in mind, for each $t \in \{1, \ldots, T\}$, the sum of these integers is bounded above as

$$\sum_{i=1}^{N} w_i[t] \leq N(2^{b_0} - 1) =: q - 1 \,. \tag{4.63}$$

Now, for some $\tau \in \mathbb{N}$ to be specified below, we form the length-k messages (4.16) in the following way

$$\boldsymbol{w}_i = \left(\sum_{t=1}^{\tau} w_i[t] q^{t-1}, \ldots, \sum_{t=1}^{\tau} w_i[t + (k-1)\tau] q^{t-1} \right),$$

$i = 1, \ldots, N$, with q as defined in (4.63). Note that the sum over i of each component is bounded above as

$$\sum_{i=1}^{N} \sum_{t=1}^{\tau} w_i[t] q^{t-1} = \sum_{t=1}^{\tau} \sum_{i=1}^{N} w_i[t] q^{t-1} \leq q^{\tau} - 1 \,. \tag{4.64}$$

Hence, for every fixed q and alphabet size p (see (4.15) and (4.27)), choosing τ such that

$$q^{\tau} - 1 \leq p - 1 \tag{4.65}$$

avoids wraparounds when messages add up over the channel. Thus, with the right-hand side of (4.63) we have

$$\tau \leq \frac{\log_2(p)}{\log_2\big(2^{b_0(f, \|\cdot\|_\infty, \varepsilon)} - 1\big) + \log_2(N)} \,. \tag{4.66}$$

Now, consider the more conservative bound

$$\tau \leq \frac{\log_2(p)}{b_0(f, \|\cdot\|_\infty, \varepsilon) + \log_2(N)} \tag{4.67}$$

by ignoring the -1 in the denominator of (4.66). Then, as the number of encoded sensor readings is $T = k\tau$, we conclude for the computation rate (see Definition 4.2)

$$R' = \frac{T}{n} \leq \frac{k \log_2(p)}{n\big(b_0(f, \|\cdot\|_\infty, \varepsilon) + \log_2(N)\big)} = \frac{R}{b_0(f, \|\cdot\|_\infty, \varepsilon) + \log_2(N)} \,.$$

Letting the nodes transmit their corresponding codewords

$$\boldsymbol{x}_i = \mathcal{E}(\boldsymbol{w}_i[1], \ldots, \boldsymbol{w}_i[T]) \qquad i = 1, \ldots, N$$

simultaneously over the Gaussian MAC results, by (4.34), in the decoding error probability

$$\mathbb{P}(\hat{\mathbf{g}} \neq \boldsymbol{g}) \leq \mathbb{P}(\mathbf{z} \notin \mathcal{V}_{\mathrm{c}}^{(n)}) \, .$$

Since our coding scheme employs a code-sequence, $\{C^{(n)}\}_{n\in\mathbb{N}}$, based on a sequence of nested lattices chosen from Lemma 4.2 (see Section 4.2.3), we have that $\mathbb{P}(\mathbf{z} \notin \mathcal{V}_{\mathrm{c}}^{(n)}) \to 0$ exponentially fast in n as long as the message rate fulfills at each node

$$\begin{aligned} R = \frac{k}{n}\log_2(p) &= \frac{1}{n}\log_2\left(\frac{\mathrm{Vol}(\mathcal{V}_{\mathrm{s}}^{(n)})}{\mathrm{Vol}(\mathcal{V}_{\mathrm{c}}^{(n)})}\right) \\ &= \frac{1}{2}\log_2\left(\frac{P}{G(\Lambda_{\mathrm{s}}^{(n)})\,\mathrm{Vol}(\mathcal{V}_{\mathrm{c}}^{(n)})^{2/n}}\right) \qquad (4.68) \\ &< \frac{1}{2}\log_2^+\left(\frac{P}{\sigma_Z^2}\right) - \frac{1}{2}\log_2\Big(2\pi\mathrm{e}G(\Lambda_{\mathrm{s}}^{(n)})\Big) \, . \qquad (4.69) \end{aligned}$$

Here, (4.68) follows from Definition 4.8 and the fact that each shaping lattice $\Lambda_{\mathrm{s}}^{(n)}$ is scaled such that its second moment equals the power constraint P, whereas (4.69) is a consequence of the sequence of coding lattices $\{\Lambda_{\mathrm{c}}^{(n)}\}_{n\in\mathbb{N}}$ being good for AWGN channel coding (see Definition 4.9). Because the sequence $\{\Lambda_{\mathrm{s}}^{(n)}\}_{n\in\mathbb{N}}$ is simultaneously good for shaping (i.e., $\lim_{n\to\infty}\log_2(2\pi\mathrm{e}G(\Lambda_{\mathrm{s}}^{(n)})) = 0$), we therefore have $\mathbb{P}(\hat{\mathbf{g}} \neq \boldsymbol{g}) \to 0$ exponentially fast with growing n if

$$R < \frac{1}{2}\log_2^+\left(\frac{P}{\sigma_Z^2}\right) \, .$$

Consequently, letting T, and thus k and p, grow appropriately with n, expression

$$\mathbb{P}\left(\bigcup_{t=1}^{T}\Big\{\hat{g}(\boldsymbol{s}[t]) \neq \tilde{g}(\boldsymbol{s}[t])\Big\}\right) \qquad (4.70)$$

vanishes exponentially fast in n as well, provided that the computation rate

$$R' < \frac{\frac{1}{2}\log_2^+\left(\frac{P}{\sigma_Z^2}\right)}{b_0(f, \|\cdot\|_\infty, \varepsilon) + \log_2(N)} = R^{\mathrm{C}}(f, \|\cdot\|_\infty, \varepsilon) \, , \qquad (4.71)$$

where $\tilde{g}(\boldsymbol{s}[t]) = \sum_{i=1}^{N}\tilde{\varphi}_i(s_i[t])$ and $\hat{g}(\boldsymbol{s}[t]) = Q^{-1}(\hat{\mathbf{g}}[t])$, $t = 1, \ldots, T$, are the corresponding estimates at the FC.

Now, recall that $\mathcal{D}_1(\hat{\mathbf{g}}) = (\hat{\mathbf{g}}[1], \ldots, \hat{\mathbf{g}}[T])$ and choose $\hat{\mathbf{g}}[t] = \hat{\boldsymbol{g}}[t]$ for some fixed $t \in \{1, \ldots, T\}$ such that

$$\psi(\hat{g}(\boldsymbol{s}[t])) \neq \psi(\tilde{g}(\boldsymbol{s}[t])) \, .$$

Then, this choice implies

$$\hat{g}(\boldsymbol{s}[t]) \neq \tilde{g}(\boldsymbol{s}[t])$$

because ψ is a function. Summarizing all such outage events into the sets

$$\mathcal{A} \coloneqq \left\{ \hat{\mathbf{g}} \in \mathbb{Z}_p^k \,\middle|\, \hat{g}(\boldsymbol{s}) \neq \tilde{g}(\boldsymbol{s}) \right\} \quad \text{and} \quad \mathcal{B} \coloneqq \left\{ \hat{\mathbf{g}} \in \mathbb{Z}_p^k \,\middle|\, \psi(\hat{g}(\boldsymbol{s})) \neq \psi(\tilde{g}(\boldsymbol{s})) \right\} ,$$

we have $\mathcal{B} \subseteq \mathcal{A}$ and therefore $\mathbb{P}(\mathcal{B}) \leq \mathbb{P}(\mathcal{A})$ due to the monotonicity of probability and the measurability of ψ. Hence, we can conclude from (4.70) that for each $t \in \{1, \ldots, T\}$,

$$\mathbb{P}\Big(\psi(\hat{g}(\boldsymbol{s}[t])) \neq \psi(\tilde{g}(\boldsymbol{s}[t]))\Big) = \mathbb{P}\Big(\hat{f}(\boldsymbol{s}[t]) \neq \tilde{f}(\boldsymbol{s}[t])\Big)$$

goes to zero exponentially fast in n, regardless of the choice of $\boldsymbol{s}[t] \in \mathcal{S}^N$. Since almost sure convergence implies convergence in probability (see Theorem B.4 in Appendix B), we have for every $\delta > 0$ that

$$\sum_{t=1}^{T} \mathbb{P}\left(\sup_{\boldsymbol{s}[t] \in \mathcal{S}^N} |\hat{f}(\boldsymbol{s}[t]) - f(\boldsymbol{s}[t])| > \varepsilon \right) < \delta$$

if n is sufficiently large, which implies (4.7) due to the union bound.

From this, we conclude that the function-values $f(\boldsymbol{s}[1]), \ldots, f(\boldsymbol{s}[T])$ can be computed with high probability within accuracy ε at a computation rate that is as close to the right-hand side of (4.71) as desired. This proves the theorem.

5

Robust Analog Computation of Nomographic Functions

In light of practical constraints, a drawback of the considerations in the previous chapters can be the implicit assumption that if spatially distributed transmit symbols are put on the channel input, then the corresponding decoder observes the sum of these inputs at the channel output. Obviously, this is only satisfied with sensor nodes perfectly synchronized on the symbol and phase level. In practical wireless sensor networks, however, it may be unreasonably difficult and expensive in terms of resource consumption to ensure such a perfect synchronization [SBK05]. Hence, the question that remains is how to beneficially exploit the interference property of the wireless channel in the presence of practical impairments.

Towards this end, in this chapter we propose and analyze a novel computation scheme that requires only a *coarse frame synchronization* and is therefore robust against synchronization errors. It is a simple *analog* scheme in which

(i) each sensor node encodes its instantaneous real-valued message (sensor reading) in the transmit power of a series of random signal pulses, and

(ii) the receiver estimates the function value directly from the corresponding received real-valued sum energy.

As the WMAC output might be corrupted by fading, we assume for our initial considerations that the sensor nodes carry out a transmitter-side channel inversion. This generally requires that perfect channel state information (CSI) is available at sensor nodes prior to transmission, which is difficult to provide in many WSN applications. Therefore, in this chapter we also examine how much CSI is actually needed at the sensor nodes in order to obtain accurate function-values. In this context, we consider FCs that are equipped with multiple antennas so that spatial diversity can be utilized for enhancing the estimation quality. This is similar to the considerations in [BSTS12, JCS14],

where the authors are interested in reliably detecting some environmental parameter from the superimposed transmit signals observed by a multiantenna FC.

A computation scheme that is related to the one proposed in this chapter is given by Mohammadi, Gohari, and Aghaeinia in [MGA12]. There, each node spreads its sensor readings, modeled as binary random variables, prior to transmissions with a binary signature waveform that is generated from the parity check matrix of some linear code. Jakimovski, Schmidtke, Sigg, Weiss Ferreira Chaves, and Beigl follow a similar approach in [JSS+12,SJB12] where the spreading sequences are chosen as sufficiently long pseudo-random bit vectors instead.

The chapter is organized as follows. In Section 5.1 we present our novel analog computation scheme for estimating nomographic functions over the wireless channel and study its statistical behavior in Section 5.2. This analysis is used to define appropriate computation-receivers (i.e., estimators) for two canonical function examples of great practical importance: the arithmetic mean and the geometric mean. Numerical examples in Section 5.3 illustrate the statistical performance of the proposed scheme and compare it with TDMA and CDMA-based computation methods. The question of how much channel knowledge is actually needed at sensor nodes is then discussed in Section 5.4. Finally, Section 5.5 provides a short summary and concludes the chapter.

Convention

A random variable X that is conditioned on a realization $\boldsymbol{s}$ of some random vector $\mathbf{s}$ is shortly written as $X|\boldsymbol{s}$. The log-normal distribution is denoted as $\mathcal{LN}(\cdot,\cdot)$ whereas χ^2_m describes the Chi-squared distribution with $m \in \mathbb{N}$ degrees of freedom. The error function and error function compliment are denoted as $\operatorname{erf} : \mathbb{R} \to [-1,1], x \mapsto \operatorname{erf}(x)$ and $\operatorname{erfc} : \mathbb{R} \to [0,2], x \mapsto \operatorname{erfc}(x)$, respectively. The indicator function on some set $\mathbb{A}$ is denoted as $\mathbb{1}_{\mathbb{A}} : \mathbb{A} \to \{0,1\}$, where $\mathbb{1}_{\mathbb{A}}(x) = 1$ if $x \in \mathbb{A}$ and zero otherwise, and the imaginary unit is denoted by the symbol i.

5.1 Random Sequences with Information Bearing Transmit Powers

In this chapter, we model the sensor readings as continuous random variables $S_i \in \mathcal{S}$, $i = 1,\dots,N$, where $\mathcal{S} \coloneqq [s_{\min}, s_{\max}] \subset \mathbb{R}$ again denotes some underlying compact sensing range, and assume that the corresponding joint probability density

$$p_{S_1,\dots,S_N} : \mathcal{S}^N \to \mathbb{R}_+ \,,\ (s_1,\dots,s_N) \mapsto p_{S_1,\dots,S_N}(s_1,\dots,s_N)$$

exists. For ease of notation, we shortly write $p_{\mathbf{s}}(\boldsymbol{s})$ with $\mathbf{s} \coloneqq (S_1,\dots,S_N) \in \mathcal{S}^N$ denoting the corresponding vector-valued random variable.

Each node, say node i, employs a computation-transmitter

$$\mathsf{Tx}_i : \mathcal{S} \to \mathbb{C}^n \,,\ S_i \mapsto \mathbf{x} \coloneqq (X_i[1],\dots,X_i[n]) \tag{5.1}$$

that maps an instantaneous sensor reading to a sequence of n complex-valued channel input symbols, subject to the maximum input cost constraint (2.4), that is,

$$\forall i \in \{1,\dots,N\} : \varrho_n\big(x_i[1],\dots,x_i[n]\big) = \max_{1\le j\le n} \big|x_i[j]\big|^2 \le P \tag{5.2}$$

for some $P > 0$. In relation to the WMAC in the complex baseband (see Definition 2.1 along with Remark 2.1), the FC receives the complex-valued sequence

$$Y[j] = \sum_{i=1}^{N} H_i[j]X_i[j] + Z[j] \qquad j = 1,\dots,n \tag{5.3}$$

in which $H_i[j] \in \mathbb{C}$ and $Z \sim \mathcal{N}_{\mathbb{C}}(0,\sigma_Z^2)$ (iid over channel uses), for some $\sigma_Z^2 > 0$.

Remark 5.1. The WMAC is a symbol-synchronous channel similar to the standard synchronous CDMA channel studied for instance in [VAT99, Ver98]. We would like to emphasize, however, that the computation scheme proposed in this section does not require such a synchronous channel and the only reason for assuming perfect synchronization is to simplify the error analysis in Section 5.2 as well as the notation throughout the chapter.

Now, let $f \in \mathscr{N}^0(\mathcal{S}^N)$ be some given desired function. Then, the FC maps each block of n channel output symbols to an estimate of a desired function-value, that is,

$$\mathsf{Rx} : \mathbb{C}^n \to \mathbb{R}\ ,\ \big(Y[1],\dots,Y[n]\big) \mapsto \hat{f}(\mathbf{s})\ . \tag{5.4}$$

In what follows, we choose the distortion measure (2.11) to be

$$d_{\mathbb{R}}\big(\hat{f}(\boldsymbol{s}), f(\boldsymbol{s})\big) = \left|\frac{\hat{f}(\boldsymbol{s}) - f(\boldsymbol{s})}{f_{\max} - f_{\min}}\right|\ , \tag{5.5}$$

where $f_{\max} \coloneqq \max_{\boldsymbol{s}\in\mathcal{S}^N} f(\boldsymbol{s})$ and $f_{\min} \coloneqq \min_{\boldsymbol{s}\in\mathcal{S}^N} f(\boldsymbol{s})$.[1] In accordance with the problem formulation given in Section 2.3, our objective is to design the computation-transmitters (5.1) along with the computation-receiver (5.4) such that

$$\mathbb{P}\Big(d_{\mathbb{R}}\big(\hat{f}(\boldsymbol{s}), f(\boldsymbol{s})\big) \ge \varepsilon\Big) < \delta \tag{5.6}$$

is fulfilled for some application-dependent constants $\varepsilon, \delta > 0$ and all $\boldsymbol{s} \in \mathcal{S}^N$.[2] Note that practical systems may tolerate distortions provided that they are small enough. However, in many applications, the requirement cannot be met permanently due to, for

[1] Since $\mathcal{S}^N$ is assumed to be a compact subset of $\mathbb{R}^N$ and $f \in \mathscr{N}^0(\mathcal{S}^N)$ is continuous, it follows from Theorem A.3 that $f_{\max}$ and $f_{\min}$ exist.

[2] Even though this outage constraint is, using the distortion measure (5.5), weaker than the supremum distortion considered in the previous chapter (see for instance (4.8)), it will suffice for the needs in this chapter.

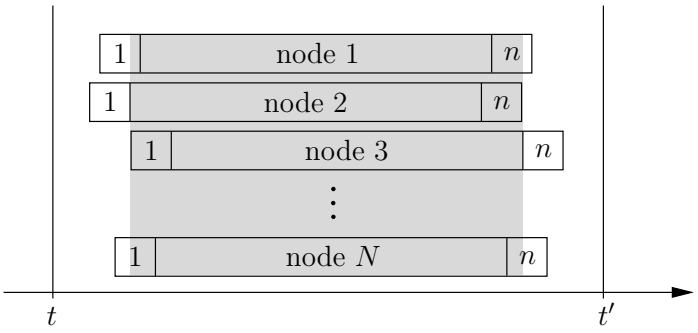


Figure 5.1: Transmit sequences of nodes sent between two arbitrary measurement times t and t' without precise symbol- and phase-synchronization. The shaded area depicts the region of maximum overlap.

instance, some random influences. In such cases, the main figure of merit is the *outage probability* (5.6).

Computing nomographic functions over (5.3) seemingly requires a receiver-side constructive superposition of the transmit signals from different sensor nodes. However, such a perfect synchronization at the symbol and phase level is notoriously difficult to realize in wireless networks and in particular in large-scale wireless sensor networks [SBK05]. Therefore, in the following we propose an *analog computation scheme*, consisting of a computation-transmitter given in Section 5.1.1 as well as a computation-receiver given in Section 5.1.2, that tolerates a *coarse frame synchronization* at the FC. This is by far easier to establish and maintain than the perfect synchronization required by traditional approaches.

The basic idea of the scheme consists in letting each sensor node transmit a distinct complex-valued sequence of length $n \in \mathbb{N}$ at a *transmit power* that depends on the pre-processed sensor reading. Under some conditions and employing a suitable pre-processing strategy, the received energy at the FC equals the sum of all the transmit energies corrupted by Gaussian noise. The application of an appropriately chosen post-processing function then results, together with some simple arithmetic calculations (to ensure certain estimation properties), in an immediate estimate of the desired function of the sensor readings. The coarse block-synchronization is needed to ensure a sufficiently large overlap of the different signal frames, and accordingly of the different transmit energies, as illustrated in Figure 5.1.

5.1.1 Analog Computation-Transmitter

Data Pre-Processing

As each pre-processed sensor reading is to be encoded in transmit power only, it is necessary to apply a suitable bijective continuous mapping

$$g_\varphi : [\varphi_{\min}, \varphi_{\max}] \to [0, P]$$

from the set of all pre-processed sensor readings onto the set of all feasible transmit powers. Here and hereafter,

$$\varphi_{\min} \coloneqq \min_{1\le i\le N} \min_{s\in\mathcal{S}} \varphi_i(s)$$
$$\varphi_{\max} \coloneqq \max_{1\le i\le N} \max_{s\in\mathcal{S}} \varphi_i(s)\ ,$$

which both exist by Theorem A.3 because the pre-processing functions are continuous. Note that even though the mapping g_φ depends on the pre-processing functions and the sensing range, it is independent of i. This is necessary as the FC does not have access to individual transmit signals but only to the WMAC output given by (5.3). We call the quantity

$$P_i \coloneqq g_\varphi(\varphi_i(S_i)) \tag{5.7}$$

the transmit power of node i and point out that it is a random variable whenever S_i is random. Moreover, we have $P_i \le P$ and thus the information to be conveyed to the FC is, for each $i \in \{1,\dots,N\}$, encoded in P_i.

Random Sequences

The transmit power (5.7) *modulates* a sequence of random symbols. In what follows, we use $\mathbf{v}_i \coloneqq (V_i[1],\dots,V_i[n]) \in \mathbb{C}^n$ to denote a sequence of random transmit symbols that are independently generated by node i. The symbols of the sequence are assumed to be of the form $V_i[j] = \exp_{\mathrm{e}}(\mathrm{i}\Theta_i[j])$, $j = 1,\dots,n$, where $\{\Theta_i[j]\}$ are continuous random phases that are uniformly iid on $[0, 2\pi)$. This implies $\|\mathbf{v}_i\|_2^2 = n$, for all $i = 1,\dots,N$, and a constant envelope (i.e., $|V_i[j]|^2 = 1$ for all i, j), which is a vital practical constraint. We have two remarks.

Remark 5.2. Instead of optimizing the sequences assigned to different nodes, we consider sequences

$$\mathbf{v}_i = \Big(\exp_{\mathrm{e}}(\mathrm{i}\Theta_i[1]),\dots,\exp_{\mathrm{e}}(\mathrm{i}\Theta_i[n])\Big) \qquad i = 1,\dots,N$$

with *random phases* and constant envelope in order to reduce the overhead for coordination and to improve scalability when compared to systems with optimized sequences. It is worth pointing out that the sequence design should be different from that for traditional asynchronous CDMA systems [Ver98, VAT99, SWB06], where the objective is to eliminate or mitigate the detrimental impact of inter-user interference. In contrast, sequences for CoMAC schemes should be designed to harness interference for a common goal, which is the computation of functions of sensor readings.

Remark 5.3. Note that the assumption of continuous random phases is not necessary for our CoMAC scheme to be implemented. Without loss of performance, the phases can take on values on any discrete subset of $[0, 2\pi)$ provided that the resulting sequences are zero-mean.

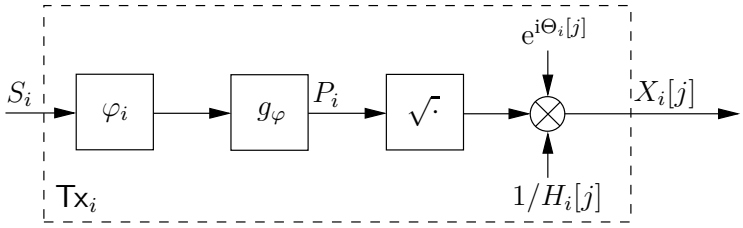


Figure 5.2: Block diagram of the analog computation-transmitter of sensor node i, $i = 1, \ldots, N$.

Transmitter-Side Channel Inversion

According to (5.3), the WMAC output, and thus the computation result, is contaminated by fading. Since a receiver-side elimination of this is generally infeasible, we suggest that each transmitter corrects the fading impact by inverting its channel to the FC. To this end, CSI is necessary at each transmitter, which can be estimated from a known pilot signal transmitted by the FC. In practical systems, the pilot signal can also be used to wake up sensor nodes and initiate the computation process. Assume $H_i[j] \neq 0$ for all i, j. Then, with CSI available at the nodes and the transmit powers given by (5.7), the WMAC input of node i at some channel use j takes the form

$$X_i[j] = \frac{\sqrt{P_i}}{H_i[j]} V_i[j] = \frac{\sqrt{g_\varphi(\varphi_i(S_i))}}{H_i[j]} \exp_{\mathrm{e}}(\mathrm{i}\Theta_i[j]) \ , \tag{5.8}$$

from which we conclude for the transmit vector if node i (see (5.1))

$$\mathbf{x} = \mathsf{Tx}_i(S_i) = \sqrt{P_i} \left(H_i[1]^{-1} \mathrm{e}^{\mathrm{i}\Theta_i[1]}, \ldots, H_i[n]^{-1} \mathrm{e}^{\mathrm{i}\Theta_i[n]} \right) \ .$$

See Figure 5.2 for a block diagram.

Remark 5.4. In Section 5.4 it is shown that dividing by the channel magnitude $|H_i[j]|$ is sufficient. Hence, estimating the channel phase is generally not necessary.

Remark 5.5. Notice that each node i with $P_i/|H_i[j]|^2 > P$ for some j cannot invert its channel under the power constraint and must therefore be excluded from transmissions associated with the current measurement instant. One possibility to mitigate the problem is to scale down all transmit powers by the same constant so that the power constraint is satisfied. Of course, this impacts the performance in noisy channels and requires some degree of coordination. We are not going to dwell on this point and assume in the following that the set of nodes is chosen such that each node can invert its own channel without violating the power constraint.

5.1.2 Analog Computation-Receiver

As mentioned before (see Remark 5.1), in order to avoid cumbersome notation and to simplify the error analysis in the next section, we assume a perfect synchronization

of signals from different nodes at the FC. The reader, however, may easily verify that the proposed computation scheme based on a simple energy estimator is insensitive to the lack of synchronization provided that a significant overlap of different signal frames (i.e., a coarse frame-synchronization) is ensured as illustrated in Figure 5.1. We also point out that the assumption of perfect synchronization has been widely used when analyzing asynchronous CDMA systems [Ver98].

Equivalent WMAC

With this assumption, the vector of receive symbols follows from (5.3) with (5.8) to

$$\mathbf{y} = (Y[1], \ldots, Y[n]) = \sum_{i=1}^{N} \sqrt{P_i}\, \mathbf{v} + \mathbf{z} \,, \tag{5.9}$$

where $\mathbf{z} \coloneqq (Z[1], \ldots, Z[n]) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma_Z^2 \boldsymbol{I}_n)$, $0 < \sigma_Z^2 < \infty$. The observation vector in (5.9) is a basis for estimating the desired function value $f(S_1, \ldots, S_N)$. Note, however, that the information about the pre-processed sensor readings is contained in the transmit powers (5.7) so that on the basis of the observation $\mathbf{y}$, the sum-energy given by

$$\|\mathbf{y}\|_2^2 = n \sum_{i=1}^{N} P_i + \underbrace{\sum_{i=1}^{N} \sum_{\substack{i'=1 \\ i' \neq i}}^{N} \sqrt{P_i P_{i'}}\, \mathbf{v}_i^{\mathsf{H}} \mathbf{v}_{i'}}_{=:\Delta_1 \in \mathbb{R}} + \underbrace{2 \sum_{i=1}^{N} \sqrt{P_i}\, \mathrm{Re}\{\mathbf{v}_i^{\mathsf{H}} \mathbf{z}\}}_{=:\Delta_2 \in \mathbb{R}} + \underbrace{\mathbf{z}^{\mathsf{H}} \mathbf{z}}_{=:\Delta_3 \in \mathbb{R}_+} \tag{5.10}$$

is a sufficient statistic for the sum $\sum_{i=1}^{N} \varphi_i(s_i)$ of pre-processed sensor readings. As a consequence, instead of (5.3), we can equivalently consider the real-valued additive noise channel (5.10), which is formally expressed as

$$W' : \mathbb{R}_+^N \to \mathbb{R}_+ \,,\ (P_1, \ldots, P_N) \mapsto W'(P_1, \ldots, P_N) = Y' = n \sum_{i=1}^{N} P_i + \Delta \,, \tag{5.11}$$

where $\Delta \coloneqq \Delta_1 + \Delta_2 + \Delta_3 \in \mathbb{R}$ and $Y' \coloneqq \|\mathbf{y}\|_2^2$. Note that in all that follows, we consider the channel in (5.11).

Signal Post-Processing

Before applying the post-processing function to the equivalent channel output Y', the receiver has to remove the influence of the function g_φ, which was used at the transmitting nodes in order to map the sensing range onto the set of feasible transmit powers. In other words, if $\Delta \equiv 0$, an application of the post-processing function has to perfectly reconstruct the sought function value, which is expected from every computation or

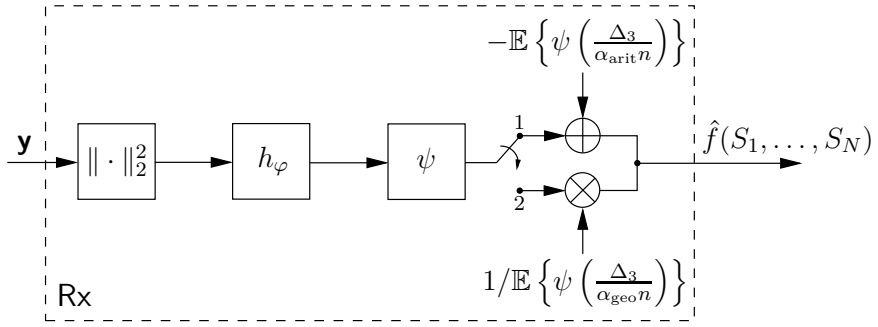


Figure 5.3: Block diagram of the analog computation-receiver for computing either the arithmetic mean (switch position 1) or the geometric mean (switch position 2). Functions h_φ and ψ depend on the choice of the desired function and should be chosen according to Definitions 5.1 and 5.2.

transmission scheme. Now, an examination of (5.11) with (5.7) shows that given g_φ, ψ and $\varphi_1, \ldots, \varphi_N$, we need to apply a function $h_\varphi : \mathbb{R} \to \mathbb{R}$ to (5.11) such that

$$\psi\left(h_\varphi\left(n\sum_{i=1}^{N} g_\varphi(\varphi_i(s_i))\right)\right) \equiv \psi\left(\sum_{i=1}^{N}\varphi_i(s_i)\right) . \tag{5.12}$$

Thus, given the respective pre- and post-processing functions, we can compute every desired function $f \in \mathscr{N}^0(\mathcal{S}^N)$ at zero distortion if $\Delta \equiv 0$ and the pair (g_φ, h_φ) satisfies (5.12). The following proposition provides a necessary and sufficient condition for the functions to fulfill (5.12).

Proposition 5.1. *Let $N \geq 2$ and $f \in \mathscr{N}^0(\mathcal{S}^N)$ be arbitrary. Then, (5.12) holds for some given $\psi, \varphi_1, \ldots, \varphi_N$ if and only if g_φ and h_φ are affine functions with $h_\varphi \equiv g_\varphi^{-1} - c$, where the constant $c \in \mathbb{R}$ depends on g_φ.*

Proof. The proof is deferred to Appendix 5.A.1 at the end of the chapter. □

In addition to the components mentioned above, the signal post-processing requires a further component that takes into account the statistics of the transformed effective noise Δ (transformed by h_φ and ψ), as it is illustrated in Figure 5.3. A more detailed explanation of this part of the computation-receiver as well as examples of the data pre-processing and the signal post-processing functions for the arithmetic mean and the geometric mean are given in Sections 5.2.2 and 5.2.3.

5.2 Outage Analysis

This section is devoted to the performance analysis of the proposed analog computation scheme in terms of the outage probability (5.6). First, we show that for sufficiently many channel uses n, the distribution of the computation noise Δ can be approximated

by a normal distribution. Since the distortion (5.5) is strongly influenced by the post-processing function ψ, and with it by the choice of the desired function f, we confine our attention in the following subsections to two special cases of great practical importance: the arithmetic mean and the geometric mean. Note that these two functions are canonical representatives of the basic arithmetic operations *summation* and *multiplication*. For both cases, we define appropriate receivers/estimators (5.4) by taking statistical properties of the transformed effective noise Δ (transformed by h_φ and ψ) into account, prove some properties and provide accurate approximations of the corresponding outage probabilities.

5.2.1 Approximation of the Effective Noise Distribution

The statistics of the effective noise in (5.11) play a key role when defining an estimator $\hat{f}$ for some desired function $f \in \mathscr{N}^0(\mathcal{S}^N)$ as well as when evaluating the performance of the proposed CoMAC scheme. Since the exact distribution of $\Delta = \Delta_1 + \Delta_2 + \Delta_3$ conditioned on the sensor readings $\mathbf{s} = \boldsymbol{s}$ is difficult to determine, we focus on suitable asymptotic approximations.

Towards this end, let us first compute the first and second statistical moments of Δ_1, Δ_2 and Δ_3. As far as Δ_1 is concerned, we have

$$\Delta_1 = \sum_{i=1}^{N} \sum_{\substack{i'=1 \\ i' \neq i}}^{N} \sum_{j=1}^{n} \sqrt{P_i P_{i'}}\, V_i^*[j] V_{i'}[j] = 2 \sum_{k=1}^{K} \sum_{j=1}^{n} \sqrt{P'_k} \underbrace{\cos(\Theta'_k[j])}_{=:C_k[j]} , \tag{5.13}$$

where $K \coloneqq N(N-1)/2$, $P'_k \coloneqq P_i P_{i'}$ and $\Theta'_k[j] \coloneqq (\Theta_{i'}[j] - \Theta_i[j]) \bmod 2\pi$ the random phase difference between nodes i and i' at channel use j. The mapping $(i, i') \mapsto k$ is obtained by

$$k = k(i, i') = i' + (i-1)N - i(i+1)/2 ,$$

$i = 1, \ldots, N-1$ and $i' = i+1, \ldots, N$, respectively.

By convolution of the densities of $\Theta_{i'}[j]$ and $\Theta_i[j]$, $\Theta'_k[j]$ is uniformly iid over $[0, 2\pi)$. Hence, the probability density of each $C_k[j]$ in (5.13) is[3]

$$p_C(c) = \frac{1}{\pi\sqrt{1-c^2}} \mathbb{1}_{(-1,1)}(c) , \tag{5.14}$$

which is symmetric around zero. This implies $\forall j, k : \mathbb{E}\{C_k[j]\} = 0$ and therefore

$$\mathbb{E}\{\Delta_1\} = 2 \sum_{k=1}^{K} \sum_{j=1}^{n} \mathbb{E}\left\{\sqrt{P'_k}\right\} \mathbb{E}\{C_k[j]\} = 0 .$$

Furthermore,

$$\mathbb{V}\mathrm{ar}\{\Delta_1\} = 4 \sum_{k=1}^{K} \sum_{j=1}^{n} \mathbb{E}\{P'_k\}\, \mathbb{V}\mathrm{ar}\{C_k[j]\} = 2n \sum_{k=1}^{K} \mathbb{E}\{P'_k\} \tag{5.15}$$

[3] Note that by the definitions, all the probability density functions and expected values in this section exist.

since $\forall j \forall k \neq k' : \mathbb{C}\mathrm{ov}\{C_k[j], C_{k'}[j]\} = 0$ and $\forall j, k : \mathbb{V}\mathrm{ar}\{C_k[j]\} = 1/2$, where the latter can be concluded by considering (5.14).

As for the second noise component Δ_2, we have

$$\Delta_2 = 2\sum_{i=1}^{N} \sqrt{P_i}\ \mathrm{Re}\{\mathbf{v}_i^{\mathsf{H}}\mathbf{z}\} = 2\sum_{i=1}^{N}\sum_{j=1}^{2n} \sqrt{P_i}\, U_{ij} Z'_j\ ,$$

where

$$U_{ij} \coloneqq \begin{cases} \cos(\Theta_i[j]) & j \text{ odd} \\ \sin(\Theta_i[j]) & j \text{ even} \end{cases} \qquad \text{and} \qquad Z'_j \coloneqq \begin{cases} \mathrm{Re}\{Z[j]\} & j \text{ odd} \\ \mathrm{Im}\{Z[j]\} & j \text{ even} \end{cases}\ .$$

Notice that $\forall j : Z'_j \sim \mathcal{N}_{\mathbb{R}}(0, \sigma_Z^2/2)$ and the probability density of U_{ij} is given by (5.14). Because U_{ij} and Z'_j are zero mean and independent for all i, j, it follows for the expectation value

$$\mathbb{E}\{\Delta_2\} = 2\sum_{i=1}^{N}\sum_{j=1}^{2n} \mathbb{E}\{\sqrt{P_i}\}\mathbb{E}\{U_{ij}\}\mathbb{E}\{Z'_j\} = 0\ .$$

Arguing along similar lines as in the case of Δ_1, the variance of Δ_2 can be easily shown to be

$$\mathbb{V}\mathrm{ar}\{\Delta_2\} = 4\sum_{i=1}^{N}\sum_{j=1}^{2n} \mathbb{E}\{P_i\}\,\mathbb{V}\mathrm{ar}\{U_{ij}\}\,\mathbb{V}\mathrm{ar}\{Z'_j\} = 2n\sigma_Z^2 \sum_{i=1}^{N} \mathbb{E}\{P_i\}\ . \tag{5.16}$$

Since $\Delta_3 = \sum_{j=1}^{n} |Z[j]|^2 \sim \chi^2_{2n}$, we finally conclude $\mathbb{E}\{\Delta_3\} = n\sigma_Z^2$ and

$$\mathbb{V}\mathrm{ar}\{\Delta_3\} = n\sigma_Z^4\ . \tag{5.17}$$

Lemma 5.1. *Δ_1, Δ_2 and Δ_3 are mutually orthogonal (in the Hilbert space of random variables with the inner product defined to be $\langle \Delta_j, \Delta_{j'} \rangle \equiv \mathbb{E}\{\Delta_j \Delta_{j'}\}$).*

Proof. Since the sensor readings, the sequence symbols and the noise are mutually independent random variables with $\mathbb{E}\{Z[j]\} = 0$ for all j, a straightforward calculation of the covariances between Δ_1 and Δ_2 as well as between Δ_2 and Δ_3 proves the lemma. □

The above derivations show that $\mathbb{E}\{\Delta\} = n\sigma_Z^2$ whereas by Lemma 5.1, the variance of Δ is the sum of the variances (5.15), (5.16) and (5.17). Thus,

$$\sigma_\Delta^2 \coloneqq \mathbb{V}\mathrm{ar}\{\Delta\} = 2n\sum_{k=1}^{K} \mathbb{E}\{P'_k\} + 2n\sigma_Z^2 \sum_{i=1}^{N} \mathbb{E}\{P_i\} + n\sigma_Z^4\ . \tag{5.18}$$

Note that when conditioned on a realization of the sensor readings (i.e., $\mathbf{s} = \boldsymbol{s}$), the variance in (5.18) yields

$$\sigma^2_{\Delta|\boldsymbol{s}} \coloneqq \mathbb{E}\left\{(\Delta - \mathbb{E}\{\Delta\})^2 \mid \mathbf{s} = \boldsymbol{s}\right\} = 2n \sum_{k=1}^{K} p'_k + 2n\sigma_Z^2 \sum_{i=1}^{N} p_i + n\sigma_Z^4 \,. \tag{5.19}$$

As mentioned at the beginning of this section, we were not able to determine the exact distribution of the effective noise Δ, which includes various terms with different distributions. However, since the number of summands $J \coloneqq N(N-1)n/2 + 2Nn + 2n$ in the definition of Δ is, already for small values of N and n, relatively large, we argue that it is well-founded to invoke the central limit theorem so as to approximate the conditional distribution by a normal distribution. The following proposition proves the corresponding convergence as $n \to \infty$.

Proposition 5.2. *Let $\Delta|\boldsymbol{s}$ be the effective noise according to (5.10) and (5.11) conditioned on the sensor readings $\mathbf{s} = \boldsymbol{s}$ with $\mathbb{E}\{\Delta \mid \mathbf{s} = \boldsymbol{s}\} = n\sigma_Z^2$, $0 < \sigma_Z^2 < \infty$, and $\sigma^2_{\Delta|\boldsymbol{s}}$ as defined in (5.19). Then, for every fixed $N, P < \infty$ and some compact sensing range $\mathcal{S}$, we have*

$$\forall \boldsymbol{s} \in \mathcal{S}^N : \frac{\Delta|\boldsymbol{s} - n\sigma_Z^2}{\sigma_{\Delta|\boldsymbol{s}}} \xrightarrow{\mathrm{d}} \mathcal{N}_{\mathbb{R}}(0,1)$$

as $n \to \infty$, where $\xrightarrow{\mathrm{d}}$ denotes the convergence in distribution.[4]

Proof. The proof is postponed to Appendix 5.A.2. □

According to Remark B.1 in Appendix B, Proposition 5.2 implies the uniform convergence of the sequence of distribution functions associated with $\{\Delta|\boldsymbol{s}\}_{n\in\mathbb{N}}$. Therefore, we can conclude that the distribution of $\Delta|\boldsymbol{s}$ can be approximated by a normal distribution provided that n is sufficiently large. This is summarized in a corollary.

Corollary 5.1. *If the block length, n, is sufficiently large, $\Delta|\boldsymbol{s}$ is close to $\tilde{\Delta}|\boldsymbol{s} \sim \mathcal{N}_{\mathbb{R}}(n\sigma_Z^2, \sigma^2_{\Delta|\boldsymbol{s}})$ in distribution.*

Remark 5.6. It turns out by a simple but lengthy calculation that for finite N, P, σ_Z^2, the conditions of the Berry-Esseen theorem (see Theorem B.8) are satisfied, which allows us to conclude, with regard to (B.3), that the convergence rate in Proposition 5.2 is of the order $\mathcal{O}(1/\sqrt{n})$. Our numerical experiments in Section 5.4.3 demonstrate that the approximation stated in Corollary 5.1 is already justified for small values of n and most cases of practical interest.

Remark 5.7. The equivalent channel given in (5.11) is asymptotically, in the block length n, a Gaussian MAC.

[4] See Definition B.5 in Appendix B.

5.2.2 Arithmetic Mean Analysis

First, we define a suitable *arithmetic mean receiver* based on the observation of the equivalent channel output Y' (see (5.11)). Subsequently, we analyze the corresponding outage performance.

Definition 5.1 (Arithmetic Mean Receiver)**.** Let f be the desired function "arithmetic mean" (see Example 3.1) and let the expected value $\mathbb{E}\{\psi(\Delta_3/(\alpha_{\text{arit}}n))\} = \sigma_Z^2/(\alpha_{\text{arit}}N)$ be known a priori to the FC, where $\alpha_{\text{arit}} \coloneqq \frac{P}{s_{\max}-s_{\min}}$. Then, given some block length $n \in \mathbb{N}$, the estimate $\hat{f}_n(\mathbf{s})$ of $f(\mathbf{s})$ is defined to be

$$\hat{f}_n(\mathbf{s}) = \mathsf{Rx}(Y') \coloneqq \psi(h_\varphi(Y')) - \mathbb{E}\left\{\psi(\Delta_3/(\alpha_{\text{arit}}n))\right\} \ . \tag{5.20}$$

Writing $n\sum_i g_\varphi(\varphi_i(s_i)) = n\sum_i p_i =: y'$, we have

- *Data pre-processing*: $\forall i \in \{1,\dots,N\} : \varphi_i(s) = s$, $g_\varphi(s) = \alpha_{\text{arit}}(s - s_{\min})$, $\varphi_{\min} = s_{\min}$, $\varphi_{\max} = s_{\max}$,
- *Signal post-processing*: $h_\varphi(y') = \frac{y'}{n\alpha_{\text{arit}}} + Ns_{\min} =: y''$ and $\psi(y'') = y''/N$.

The arithmetic mean receiver is depicted in Figure 5.3 with the switch in position 1.

Now, we provide two statements in order to show that the arithmetic mean receiver of Definition 5.1 is meaningful as it has the two most desired properties: unbiasedness and consistency.

Proposition 5.3. *The computation-receiver of Definition 5.1 is* unbiased. *That is, we have*

$$\forall n \in \mathbb{N}\, \forall \boldsymbol{s} \in \mathcal{S}^N : \mathbb{E}\left\{\hat{f}_n(\mathbf{s}) \,\middle|\, \mathbf{s} = \boldsymbol{s}\right\} = f(\boldsymbol{s}) \ .$$

Proof. The proof is given in Appendix 5.A.3. □

Proposition 5.4. *Let $N, P, \sigma_Z^2 < \infty$ be arbitrary but fixed and let $\hat{f}_1, \hat{f}_2, \dots$ be a sequence generated from (5.20). Then, the arithmetic mean receiver of Definition 5.1 is* consistent. *That is,*

$$\forall \boldsymbol{s} \in \mathcal{S}^N : \hat{f}_n(\boldsymbol{s}) \xrightarrow{\mathbb{P}} f(\boldsymbol{s}) \ ,$$

where $\xrightarrow{\mathbb{P}}$ denotes convergence in probability (see Definition B.3 in Appendix B).

Proof. The proof is deferred to Appendix 5.A.4. □

Proposition 5.4 permits us to conclude that for every given pair $\varepsilon, \delta > 0$ there exists $n_0 = n_0(\varepsilon, \delta)$ such that the outage constraint (5.6) is fulfilled for all $n \geq n_0$ and all $\boldsymbol{s} \in \mathcal{S}^N$. Since this is only a qualitative statement based on the rather loose bound (5.46) provided in the proof of Proposition 5.4, it tells little about how rapidly the outage probability vanishes with growing n. It turns out that a better approach is to invoke Proposition 5.2 in order to approximate the outage probability by using a transformed normal distribution. Note that as we have $f_{\max} = s_{\max}$ and $f_{\min} = s_{\min}$,

the distortion of the arithmetic mean conditioned on the sensor readings $\mathbf{s} = \boldsymbol{s}$ has the form (see (5.5))

$$|D_n|\boldsymbol{s}| \coloneqq d_{\mathbb{R}}(\hat{f}_n(\boldsymbol{s}), f(\boldsymbol{s})) = \left| \frac{\hat{f}_n(\boldsymbol{s}) - f(\boldsymbol{s})}{s_{\max} - s_{\min}} \right| = \left| \frac{\Delta|\boldsymbol{s} - n\sigma_Z^2}{nNP} \right| . \tag{5.21}$$

The Mann-Wald theorem (see Theorem B.5 of Appendix B) guarantees that for every real-valued continuous function h, one has $h(X_n) \xrightarrow{\mathrm{d}} h(X)$ whenever $X_n \xrightarrow{\mathrm{d}} X$. We can therefore conclude from Corollary 5.1 that for sufficiently large values of n, $D_n|\boldsymbol{s}$ can be approximated (in distribution) by a random variable

$$\tilde{D}_n|\boldsymbol{s} \sim \mathcal{N}_{\mathbb{R}}\left(0\,, \frac{\sigma^2_{\Delta|\boldsymbol{s}}}{(nNP)^2}\right)$$

with conditional distribution function

$$P_{\tilde{D}_n|\boldsymbol{s}} : \mathbb{R} \to [0,1]\ ,\ \xi \mapsto P_{\tilde{D}_n|\boldsymbol{s}}(\xi) \coloneqq \mathbb{P}\left(\tilde{D}_n \le \xi \,|\, \mathbf{s} = \boldsymbol{s}\right) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{nNP\xi}{\sigma_{\Delta|\boldsymbol{s}}\sqrt{2}}\right)\right] .$$

Since the absolute value is a continuous function as well, we obtain for arbitrary $\varepsilon > 0$ and sufficiently large n

$$\begin{aligned} \mathbb{P}\Big(d_{\mathbb{R}}(\hat{f}(\mathbf{s}), f(\mathbf{s})) \ge \varepsilon \,|\, \mathbf{s} = \boldsymbol{s}\Big) &\approx \mathbb{P}\left(|\tilde{D}_n| \ge \varepsilon \,|\, \mathbf{s} = \boldsymbol{s}\right) \\ &= 1 - P_{\tilde{D}_n|\boldsymbol{s}}(\varepsilon) + P_{\tilde{D}_n|\boldsymbol{s}}(-\varepsilon) \\ &= \operatorname{erfc}\left(\frac{nNP\xi}{\sigma_{\Delta|\boldsymbol{s}}\sqrt{2}}\right) , \end{aligned} \tag{5.22}$$

where we used the fact that $\operatorname{erf}(-x) = -\operatorname{erf}(x)$ for all $x \in \mathbb{R}$.

Now, one could examine the accuracy and the behavior of (5.22) for different values of n and $\boldsymbol{s}$. We find it more convenient, however, to consider, in the numerical examples Section 5.3, the average with regard to some density $p_{\mathbf{s}}(\boldsymbol{s})$, that is, the total probability

$$\mathbb{P}\left(|D_n| \ge \varepsilon\right) = \mathbb{E}_{\mathbf{s}}\left\{\mathbb{P}\Big(d_{\mathbb{R}}(\hat{f}(\mathbf{s}), f(\mathbf{s})) \ge \varepsilon \,|\, \mathbf{s}\Big)\right\} \approx \int_{\mathcal{S}^N} \operatorname{erfc}\left(\tfrac{nNP\xi}{\sigma_{\Delta|\boldsymbol{s}}\sqrt{2}}\right) p_{\mathbf{s}}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}\ .^{5} \tag{5.23}$$

Remark 5.8. Note that by compactness of $\mathcal{S}$, it follows from Proposition 5.4 along with the theorem on dominated convergence (see Theorem B.6) that

$$\lim_{n\to\infty} \mathbb{E}_{\mathbf{s}}\left\{\mathbb{P}\Big(d_{\mathbb{R}}(\hat{f}(\mathbf{s}), f(\mathbf{s})) \ge \varepsilon \,|\, \mathbf{s}\Big)\right\} = \mathbb{E}_{\mathbf{s}}\left\{\lim_{n\to\infty} \mathbb{P}\Big(d_{\mathbb{R}}(\hat{f}(\mathbf{s}), f(\mathbf{s})) \ge \varepsilon \,|\, \mathbf{s}\Big)\right\} = 0\ .$$

[5] The right-hand side is a compact notation for the multiple Riemann integral

$$\int_{\mathcal{S}} \cdots \int_{\mathcal{S}} \operatorname{erfc}(\cdot) p_{S_1,\ldots,S_N}(s_1,\ldots,s_N)\,\mathrm{d}s_1 \cdots \mathrm{d}s_N\ ,$$

which exists as the integrand is bounded.

5.2.3 Geometric Mean Analysis

As in the preceding subsection, we first define a *geometric mean receiver* including the required data pre-processing and signal post-processing functions.

Definition 5.2 (Geometric Mean Receiver)**.** Let the sensing range be $\mathcal{S} = [s_{\min}, s_{\max}]$ with $s_{\min} > 0$ and f the desired function "geometric mean" be defined as in Example 3.2. Furthermore, let the expected value $\mathbb{E}\{\psi(\Delta_3/(\alpha_{\text{geo}}n))\}$ be a priori known to the FC (see Lemma 5.2 below) with $\alpha_{\text{geo}} \coloneqq \frac{P}{\log_{\text{e}}(s_{\max})-\log_{\text{e}}(s_{\min})}$. Then, given $n \in \mathbb{N}$, the estimate $\hat{f}_n(\mathbf{s})$ of $f(\mathbf{s})$ is defined to be

$$\hat{f}_n(\mathbf{s}) = \mathsf{Rx}(Y') \coloneqq \frac{\psi(h_\varphi(Y'))}{\mathbb{E}\left\{\psi(\Delta_3/(\alpha_{\text{geo}}n))\right\}} = f(\mathbf{s})\frac{\psi(\Delta/(\alpha_{\text{geo}}n))}{\mathbb{E}\left\{\psi(\Delta_3/(\alpha_{\text{geo}}n))\right\}} \,. \tag{5.24}$$

Writing $n\sum_i g_\varphi(\varphi_i(s_i)) = n\sum_i p_i =: y'$, we have

- *Data pre-processing*: $\forall i \in \{1,\dots,N\} : \varphi_i(s) = \log_{\text{e}}(s)$, $\varphi_{\min} = \log_{\text{e}}(s_{\min})$, $\varphi_{\max} = \log_{\text{e}}(s_{\max})$, and $g_\varphi(\log_{\text{e}}(s)) = \alpha_{\text{geo}}(\log_{\text{e}}(s) - \log_{\text{e}}(s_{\min}))$.
- *Signal post-processing*: $h_\varphi(y') = \frac{y'}{n\alpha_{\text{geo}}} + N\log_{\text{e}}(s_{\min}) =: y''$ and $\psi(y'') = \exp_{\text{e}}\left(\frac{y''}{N}\right)$.

The geometric mean receiver is shown in Figure 5.3 with the switch in position 2.

As mentioned in the definition, our computation-receiver requires to know the expected value $\mathbb{E}\{\psi(\Delta_3/(\alpha_{\text{geo}}n))\}$ in advance, which is explicitly given in part (i) of the following lemma. Part (ii) is used in the proof of Proposition 5.5.

Lemma 5.2. *Let α_{geo} be such as in Definition 5.2. Suppose that $\sigma_Z^2 < \alpha_{\text{geo}}Nn$. Then,*

(i) $\lambda_n \coloneqq \mathbb{E}\left\{\psi(\Delta_3/(\alpha_{\text{geo}}n))\right\} = \left(\frac{\alpha_{\text{geo}}Nn}{\alpha_{\text{geo}}Nn-\sigma_Z^2}\right)^n$

(ii) $\lim\limits_{n\to\infty}\lambda_n = \exp_{\text{e}}\left(\frac{\sigma_Z^2}{\alpha_{\text{geo}}N}\right)$.

Proof. The proof is deferred to Appendix 5.A.5. □

We point out that the expected value λ_n exists if $\sigma_Z^2 < \alpha_{\text{geo}}Nn$ holds, which is usually fulfilled in practical situations and therefore assumed in what follows.

With (5.24), the distortion of the geometric mean conditioned on the sensor readings $\mathbf{s} = \boldsymbol{s}$ becomes

$$|D_n|\boldsymbol{s}| \coloneqq d_{\mathbb{R}}(\hat{f}_n(\boldsymbol{s}), f(\boldsymbol{s})) = \left|\frac{\hat{f}(\boldsymbol{s}) - f(\boldsymbol{s})}{f_{\max} - f_{\min}}\right| = \left|\frac{1}{\gamma(\boldsymbol{s})}\Xi|\boldsymbol{s} - \beta(\boldsymbol{s})\right| = \left|\beta(\boldsymbol{s})\left(\frac{\Xi|\boldsymbol{s}}{\lambda_n} - 1\right)\right| \,, \tag{5.25}$$

where we used the following notation: $f_{\max} = s_{\max}$, $f_{\min} = s_{\min}$, $\beta(\boldsymbol{s}) \coloneqq f(\boldsymbol{s})/(s_{\max} - s_{\min})$, $\gamma(\boldsymbol{s}) \coloneqq \lambda_n/\beta(\boldsymbol{s})$ and $\Xi|\boldsymbol{s} \coloneqq \psi(\Delta|\boldsymbol{s}/(\alpha_{\text{geo}}n))$.

Note that the receiver of Definition 5.2 is not necessarily unbiased but it offers the advantage of a simple implementation in practical systems. In contrast, the receiver

$$\hat{f}'_n(\mathbf{s}) = \mathsf{Rx}'(Y') \coloneqq \frac{\psi(h_\varphi(Y'))}{\mathbb{E}\left\{\psi(\Delta/(\alpha_{\text{geo}}n))\right\}} \tag{5.26}$$

is unbiased but not applicable in practice because, opposed to the expected value in (5.24), $\mathbb{E}\left\{\psi(\Delta/(\alpha_{\text{geo}}n))\right\}$ depends on the effective noise and thus on the distribution of the sensor readings, which is usually unknown at the FC. Although (5.24) is not unbiased, the following proposition shows that it is consistent.

Proposition 5.5. *For some fixed $N, P_Z, \sigma_Z^2 < \infty$ and arbitrary $\boldsymbol{s} \in \mathcal{S}^N$, the sequence $\hat{f}_1, \hat{f}_2, \ldots$ of geometric mean receivers proposed in Definition 5.2 is* consistent*.*

Proof. The proof can be found in Appendix 5.A.6. □

Remark 5.9. Notice that the proposition implies that the proposed geometric mean receiver (5.24) is asymptotically unbiased, that is, we have

$$\forall \boldsymbol{s} \in \mathcal{S}^N : \lim_{n\to\infty} \mathbb{E}\left\{\hat{f}_n(\mathbf{s}) \,|\, \mathbf{s} = \boldsymbol{s}\right\} = f(\boldsymbol{s}) \,.$$

As a consequence, (5.24) and (5.26) are asymptotically equivalent. Additionally, the proposition implies, in conjunction with the compactness of $\mathcal{S}$ and Theorem B.6, that

$$\lim_{n\to\infty} \mathbb{P}(|D_n| \geq \varepsilon) = \lim_{n\to\infty} \mathbb{E}_{\mathbf{s}}\left\{\mathbb{P}\Big(|D_n| \geq \varepsilon \,|\, \mathbf{s}\Big)\right\} = \mathbb{E}_{\mathbf{s}}\left\{\lim_{n\to\infty} \mathbb{P}\Big(|D_n| \geq \varepsilon \,|\, \mathbf{s}\Big)\right\} = 0 \,.$$

Unfortunately, $\mathbb{P}(|D_n| \geq \varepsilon)$ cannot be exactly evaluated because we are not able to determine the exact distribution function of $|D_n| = |\frac{1}{\gamma(\mathbf{s})}\Xi - \beta(\mathbf{s})|$. For this reason, as in the preceding subsection, we approximate the distribution of $\Xi|\boldsymbol{s}$ by a transformed normal distribution since in contrast to the arithmetic mean case $\Xi|\boldsymbol{s}$ depends nonlinearly on the conditioned effective noise $\Delta|\boldsymbol{s}$.

Lemma 5.3. *Let $N < \infty$, $\mathcal{S}$ be compact and n sufficiently large. Then, $\Xi|\boldsymbol{s}$ can be approximated (in distribution) by a random variable $\tilde{\Xi}|\boldsymbol{s} \sim \mathcal{LN}(\mu_\Xi, \sigma^2_{\Xi|\boldsymbol{s}})$, where $\mu_\Xi \coloneqq \sigma_Z^2/(\alpha_{\text{geo}}N)$ and $\sigma^2_{\Xi|\boldsymbol{s}} \coloneqq \sigma^2_{\Delta|\boldsymbol{s}}/(\alpha_{\text{geo}}Nn)^2$, respectively.*

Proof. The proof is deferred to Appendix 5.A.7. □

With Lemma 5.3 in hand, we are now in a position to provide an approximation of $\mathbb{P}(|D_n| \geq \varepsilon)$, which is given in the following proposition.

Proposition 5.6. *Consider the proposed geometric mean receiver (5.24) and suppose that $|D_n|$ is the corresponding distortion. For some $\boldsymbol{s} \in \mathcal{S}^N$, let μ_Ξ and $\sigma^2_{\Xi|\boldsymbol{s}}$ be given by Lemma 5.3 and let $\beta(\boldsymbol{s}), \gamma(\boldsymbol{s}) > 0$ such as defined in (5.25). Then, for n sufficiently large, the outage probability $\mathbb{P}(|D_n| \geq \varepsilon)$ can be approximated, for arbitrary $\varepsilon > 0$, by*

$$\mathbb{P}(|D_n| \geq \varepsilon) \approx \mathbb{P}\Big(|\tilde{D}_n| \geq \varepsilon\Big) = \int_{\mathcal{S}^N} \mathbb{P}\Big(|\tilde{D}_n| \geq \varepsilon \,|\, \mathbf{s} = \boldsymbol{s}\Big) p_{\mathbf{s}}(\boldsymbol{s}) \,\mathrm{d}\boldsymbol{s} \tag{5.27}$$

with $\mathbb{P}(|\tilde{D}_n| \geq \varepsilon \,|\, \mathbf{s} = \boldsymbol{s})$ *equal to*

$$\frac{1}{2}\left[2 + \operatorname{erf}\left(\frac{\log_{\mathrm{e}}(\rho^-(\boldsymbol{s},\varepsilon)) - \mu_\Xi}{\sqrt{2}\,\sigma_{\Xi|\boldsymbol{s}}}\right) - \operatorname{erf}\left(\frac{\log_{\mathrm{e}}(\rho^+(\boldsymbol{s},\varepsilon)) - \mu_\Xi}{\sqrt{2}\,\sigma_{\Xi|\boldsymbol{s}}}\right)\right] \tag{5.28}$$

for $0 < \varepsilon < \beta(\boldsymbol{s})$ *and equal to*

$$\frac{1}{2}\operatorname{erfc}\left(\frac{\log_{\mathrm{e}}(\rho^+(\boldsymbol{s},\varepsilon)) - \mu_\Xi}{\sqrt{2}\,\sigma_{\Xi|\boldsymbol{s}}}\right) \tag{5.29}$$

for $\beta(\boldsymbol{s}) \leq \varepsilon < \infty$*, respectively, where* $\rho^-(\boldsymbol{s},\varepsilon) \coloneqq \gamma(\boldsymbol{s})(\beta(\boldsymbol{s}) - \varepsilon)$ *and* $\rho^+(\boldsymbol{s},\varepsilon) \coloneqq \gamma(\boldsymbol{s})(\beta(\boldsymbol{s}) + \varepsilon)$*.*

Proof. The proof is postponed to Appendix 5.A.8. □

In Section 5.3.1, we choose a particular density $p_{\mathbf{s}}(\boldsymbol{s})$ and evaluate (5.27) numerically to indicate the accuracy of the approximation for different network parameters.

5.3 Numerical Examples

The objective of this section is twofold. First, in Section 5.3.1 we show that the approximations of Section 5.2 are sufficiently accurate, and second, in Section 5.3.2 we compare the proposed analog computation scheme with a TDMA-based and a CDMA-based scheme (i.e., two separation-based strategies) in order to show the huge potential for performance gains in typical sensor network operating points.

As a basis, we consider a classical environmental monitoring scenario in which the FC is interested in the arithmetic mean or geometric mean of temperature measurements carried out by a number of sensor nodes that are distributed over some geographical area. Towards this end, we assume that all nodes are equipped with a low-power temperature sensor operating in a typical sensing range $\mathcal{S} = [-55\,^\circ\mathrm{C}, 130\,^\circ\mathrm{C}]$ [STM09].

5.3.1 Approximation Accuracy

To assess the accuracy of the approximated outage probabilities, we consider two scenarios: one in which the FC estimates the arithmetic mean, and another where the geometric mean is desired. Accordingly, we compare the approximations (5.23) and (5.27) with Monte Carlo evaluations of the true outage probability $\mathbb{P}(|D_n| \geq \varepsilon)$ based on 10^4 realizations. Note that for both simulation examples, the power constraint P and the noise variance σ_Z^2 have been chosen in agreement with commercial IEEE 802.15.4 compliant sensor platforms [Ins07].

Example 5.1 (Arithmetic Mean)**.** Let the number of channel uses chosen to be $n = 25, 50, 150, 250$, the number of nodes as $N = n$, and the sensor readings uniformly iid on the interval $[1\,^\circ\mathrm{C}, 30\,^\circ\mathrm{C}]$, which is a subset of the above-specified sensing range $\mathcal{S}$. The resulting experimental data is depicted in Figure 5.4. △

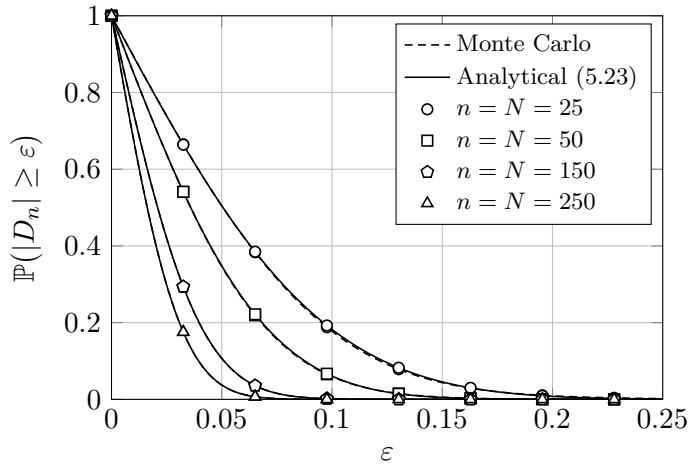


Figure 5.4: Monte Carlo evaluation of the outage probabilities (10^4 simulation runs) versus analytical results for the desired function "arithmetic mean" and different values $n = N$.

The plots in Figure 5.4 show that the approximation (5.23) closely matches the true outage probability $\mathbb{P}(|D_n| \geq \varepsilon)$ for all $\varepsilon > 0$. Notice that already for relatively few channel uses n differences between the analytical expression and the Monte Carlo simulations are negligible. Furthermore, the plots confirm the consistency statement of Proposition 5.4 because the probability curves tend to the ordinate axis with growing n.

Example 5.2 (Geometric Mean)**.** Let $\mathcal{S}' \coloneqq [0.5\,^\circ\mathrm{C}, 130\,^\circ\mathrm{C}] \subset \mathcal{S}$ be the restricted sensing range for desired function "geometric mean", the sensor readings uniformly iid on $[1\,^\circ\mathrm{C}, 30\,^\circ\mathrm{C}] \subset \mathcal{S}'$, and all other simulation parameters as in Example 5.1.[6] The resulting experimental data is depicted in Figure 5.5. △

Similar as for Example 5.1, the plots in Figure 5.5 indicate that expression (5.27), in conjunction with (5.28) and (5.29), approximates the true outage probability sufficiently accurate with a deviation that is negligible also for small values of n. Even though the geometric mean receiver (5.24) is applicable in practice, it has the drawback that unbiasedness is achieved only asymptotically as the number of channel uses, n, tends to infinity. To quantify this drawback, Figure 5.5 also depicts Monte Carlo evaluations of $\mathbb{P}(|D_n| \geq \varepsilon)$ based on the unbiased (but impractical) computation receiver (5.26). Note that the drawback vanishes quickly with increasing n, which confirms the statements of Proposition 5.5 and Remark 5.9.

Remark 5.10. Propositions 5.4 and 5.5 as well as Examples 5.1 and 5.2 demonstrate that the block length n (i.e., the number of channel uses) is the crucial design pa-

[6] Notice that $\mathcal{S}'$ is necessary since the geometric mean cannot be continuously extended onto the entire sensing range $\mathcal{S} = [-55\,^\circ\mathrm{C}, 130\,^\circ\mathrm{C}]$.

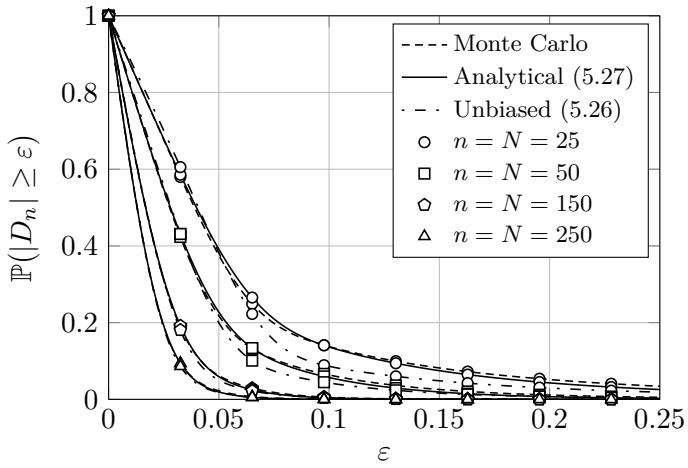


Figure 5.5: Monte Carlo evaluation of the outage probabilities (10^4 simulation runs) versus analytical results for the desired function "geometric mean" and different values $n = N$.

rameter that determines the trade-off between computation accuracy and achievable computation rate.

5.3.2 Comparisons with TDMA and CDMA

The numerical examples in the preceding subsection indicate the general behavior of the proposed analog computation architecture without concrete evidence regarding the computation performance compared to standard separation-based strategies. Therefore, in this subsection we demonstrate the advantage of the proposed analog computation architecture over idealized uncoded TDMA and CDMA schemes. In both cases, each node quantizes its sensor readings uniformly over $\mathcal{S}$ with $k \in \mathbb{N}$ bits and transmits them to the FC as bipolar symbol streams. The FC then reconstructs all individual readings in order to compute the sought function value afterwards.

TDMA

To ensure fairness between CoMAC and TDMA, with fixed degrees of freedom (e.g., bandwidth, symbol duration), both schemes should induce the same costs per function value computation with respect to transmit energy and transmit time. Therefore, let $T \in \mathbb{R}_{++}$ be the common symbol duration and let $P_{\mathrm{TDMA},i} \in \mathbb{R}_{++}$ denote the TDMA transmit power of node i, $i = 1, \ldots, N$. Then, the transmission times per function value are $T_{\mathrm{CoMAC}} = nT$ and $T_{\mathrm{TDMA}} = kNT$ whereas the transmit energies can be written as $E_{\mathrm{CoMAC},i} = nP_iT$ and $E_{\mathrm{TDMA},i} = kP_{\mathrm{TDMA},i}T$, respectively. Now, from the fairness conditions $T_{\mathrm{CoMAC}} = T_{\mathrm{TDMA}}$ and $E_{\mathrm{CoMAC},i} = E_{\mathrm{TDMA},i}$, for all i, it follows $n = kN$ for

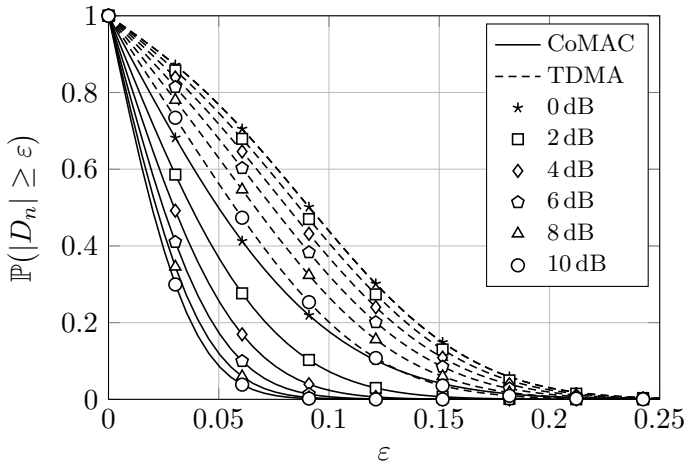


Figure 5.6: CoMAC vs. TDMA: outage probabilities for computing the "arithmetic mean" in a network of $N = 25$ nodes, quantization with $k = 10$ bits (in the case of TDMA), $n = kN$ channel uses, and $\mathrm{SNR}^{\mathrm{dB}}_{\mathrm{TDMA}} = 0, 2, 4, 6, 8, 10$.

the number of channel uses and

$$P_{\mathrm{TDMA},i} = \frac{nP_i}{k} = \frac{ng_\varphi(\varphi_i(S_i))}{k} \qquad i = 1, \ldots, N$$

for the required TDMA transmit powers.

In addition to the fairness aspect, an adequate comparison requires the determination of a common system operating point, which can be done in terms of an average SNR. Assume for simplicity that the sensor readings, S_i, are iid on $\mathcal{S}$ for all $i = 1, \ldots, N$ such that the average received TDMA-SNR per node can be defined as[7]

$$\mathrm{SNR}_{\mathrm{TDMA}} \coloneqq \frac{2n\mathbb{E}\{P_1\}}{k\sigma_Z^2} \,. \tag{5.30}$$

Example 5.3 (Small Network Size)**.** Let $N = 25$, $k = 10$ bits, the number of channel uses $n = kN$, and let P and σ_Z^2 be chosen such that

$$\mathrm{SNR}^{\mathrm{dB}}_{\mathrm{TDMA}} \coloneqq 10\log_{10}(\mathrm{SNR}_{\mathrm{TDMA}}) \in \{0, 2, 4, 6, 8, 10\} \,.$$

Furthermore, let the sensor readings be uniformly iid over $[5\,^\circ\mathrm{C}, 30\,^\circ\mathrm{C}]$, which is a proper subset of the sensing range, and let the desired function be the "arithmetic mean". The corresponding simulation data is depicted in Figure 5.6. △

Example 5.4 (Medium Network Size)**.** Let $N = 250$, the desired function be "geometric mean" with $\mathcal{S}' = [1\,^\circ\mathrm{C}, 130\,^\circ\mathrm{C}] \subset \mathcal{S}$, and let all other simulation parameters as in Example 5.3. The corresponding simulation data is shown in Figure 5.7. △

[7] The factor of 2 in (5.30) and (5.31) results from the fact that the considered TDMA- and CDMA-based approaches use, in contrast to the analog computation scheme presented in this chapter, only one real dimension per transmission.

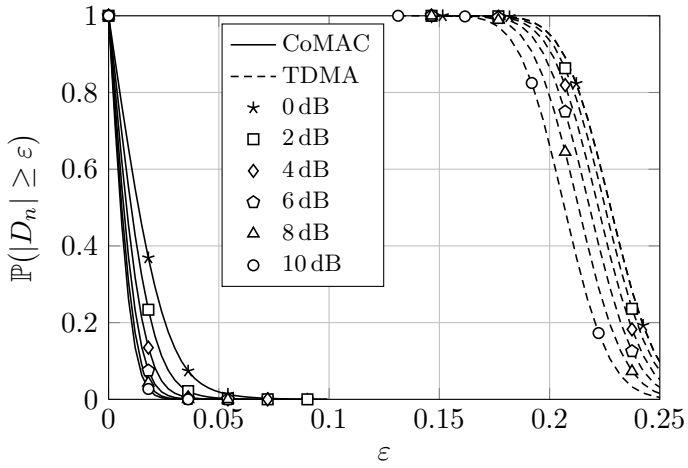


Figure 5.7: CoMAC vs. TDMA: outage probabilities for computing the "geometric mean" in a network of $N = 250$ nodes, quantization with $k = 10$ bits (in the case of TDMA), $n = kN$ channel uses, and $\mathrm{SNR}_{\mathrm{TDMA}}^{\mathrm{dB}} = 0, 2, 4, 6, 8, 10$.

Figures 5.6 and 5.7 indicate the huge potential of the proposed analog CoMAC scheme for efficiently computing linear and nonlinear functions over the wireless channel. In both examples, CoMAC entirely outperforms TDMA with respect to the computation accuracy for different network parameters. It should be clear that the shown performance gains can be traded off for better computation rate or higher energy efficiency.

Remark 5.11. It is important to emphasize that the shown performance gains are quite conservative since the simulated TDMA scheme was idealized in many ways. For example, a realistic TDMA would require an established protocol stack with considerable amount of overhead per frame (e.g., header, synchronization information, check sum) so that the overall TDMA transmission time would extend to $T_{\mathrm{TDMA}} = (k+k')NT$ for some $k' \in \mathbb{N}$.

CDMA

In contrast to the TDMA-based approach in which the sensor readings are transmitted to the FC in an interference-free manner, the nodes can employ the direct-sequence CDMA principle [Rap02, pp. 331-333] to also transmit concurrently in the same frequency band. Therefore, to compare a CDMA-based approach with our analog CoMAC scheme, for CDMA let each node spread its quantized sensor readings prior to transmission with an individual fixed bipolar Welch-bound-equality (WBE) sequence of length $\tilde{n} \in \mathbb{N}$. In particular, we assume that the WBE sequences are constructed according to [MM93] from some maximum-length sequence.[8] The FC estimates all sen-

[8] A *maximum-length sequence* (or m-sequence) is a binay pseudo-noise sequence that is generated by means of a linear feedback shift register with m states, $m \in \mathbb{N}$. They owe their name from the fact

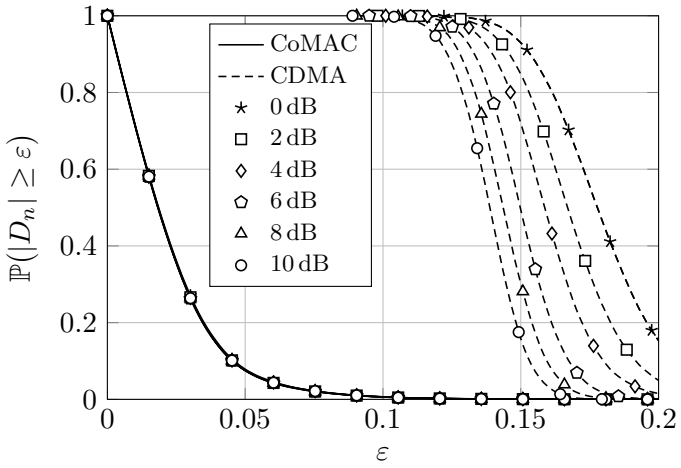


Figure 5.8: CoMAC vs. CDMA: outage probabilities for quantization with $k = 10\,\text{bits}$ (in the case of CDMA), $N = 256$ nodes, $n = 150$ channel uses, $\text{SNR}_{\text{CDMA}}^{\text{dB}} = 0, 2, 4, 6, 8, 10\,\text{dB}$, and desired function "geometric mean".

sor readings from the received signal using the standard CDMA matched-filter receiver [Ver98, pp. 56-60], which in conjunction with WBE sequences becomes an optimal linear minimum mean square error receiver.

To make the comparison fair, we again assume that both schemes induce the same costs per function-value computation, from which it follows that CoMAC has to use the channel exactly $n = k\tilde{n}$ times. This in conjunction with equally set transmit powers (i.e., $P_{\text{CDMA},i} = P_i = g_\varphi(\varphi_i(S_i))$, $i = 1, \ldots, N$), and S_i uniformly iid over $[5\,^\circ\text{C}, 30\,^\circ\text{C}]$ results in equal transmit energy consumptions per transmitted symbol and in the average CDMA-SNR (per node)

$$\text{SNR}_{\text{CDMA}} \coloneqq \frac{2\mathbb{E}\{P_1\}}{\sigma_Z^2} \tag{5.31}$$

at the output of the matched filter.

Example 5.5. Let $N = 256$, $k = 10\,\text{bits}$, $\tilde{n} = 15$ ($\Rightarrow n = 150$), and the desired function be the "geometric mean". Furthermore, let P and σ_Z^2 chosen such that

$$\text{SNR}_{\text{CDMA}}^{\text{dB}} \coloneqq 10\log_{10}(\text{SNR}_{\text{CDMA}}) \in \{0, 2, 4, 6, 8, 10\}\ ,$$

and all other simulation parameters such as in Example 5.4. The corresponding achievable outage probabilities are depicted in Figure 5.8. △

The plots show that in the considered range of typical sensor network operating points, our analog CoMAC scheme also entirely outperforms a computation approach

that they are periodic with period $2^m - 1$, which corresponds to the maximal number of different shift register states apart from the state where all bits are zero [PS08, pp. 461-463].

that is based on the CDMA principle.[9] The reader should notice that even if the proposed CoMAC scheme seems to be quite similar to the CDMA concept from a transmitter-side perspective, the huge performance gains result from estimating the desired function *immediately* from the WMAC output instead of trying to reconstruct all individual sensor readings by multiuser detection.

Remark 5.12. In the terminology of CDMA represents Example 5.5 an overloaded case (i.e., $n < N$). This was considered because signature waveforms can always be chosen to be mutually orthogonal whenever $n \geq N$ so that the resulting outage performance would at best be the same as for TDMA (see Figures 5.6 and 5.7), and would therefore not provide new insights. Note that the overloaded case was the reason for letting the nodes spread their quantized sensor readings with WBE sequences since for $n < N$ they are optimal in the sense that they can achieve the lower bound on the sum of cross-correlations between spreading sequences [MM93].

5.4 On the Channel Estimation Effort

A crucial assumption in the previous sections was that perfect CSI is available at the sensor nodes prior to transmissions, which is called "Full CSI" in what follows. Because "Full CSI" is difficult to provide in many WSN applications, we explore in this section the question of how much CSI is actually needed to obtain reliable function estimates. In particular, we show that the knowledge of the channel magnitudes (called "Modulus CSI") is sufficient to achieve the same performance as with "Full CSI". Moreover, we show that under certain conditions, no CSI is needed at the sensor nodes provided that the FC has some a priori statistical knowledge. If in addition the FC is equipped with multiple antennas, it is shown that spatial diversity can be exploited to outperform the single-antenna scheme with "Full CSI". Our findings suggest that the channel estimation effort at the sensor nodes can significantly be reduced.

5.4.1 Multiantenna Fusion Centers

We slightly extend the system model of Section 5.1 in order to incorporate multiple antennas at the FC. Accordingly, let the FC be equipped with $M \in \mathbb{N}$ antennas so that the complex-valued time-discrete signal received at antenna element m can be written as

$$Y_m[j] = \sum_{i=1}^{N} H_{mi}[j] X_i[j] + Z_m[j] \qquad j = 1, \ldots, n \; . \tag{5.32}$$

Here and hereafter, $H_{mi}[j] \neq 0$ is used to model the complex-valued frequency-flat fading channel from node i to antenna m and $Z_m \sim \mathcal{N}_{\mathbb{C}}(0, \sigma_Z^2)$ the m^{th} receiver noise

[9] The CoMAC plots in Figure 5.8 coincide with each other due to the fact that the CoMAC outage performance is, in the considered parameter range, mainly determined by the cross-correlations between the random transmit sequences and less by the receiver noise.

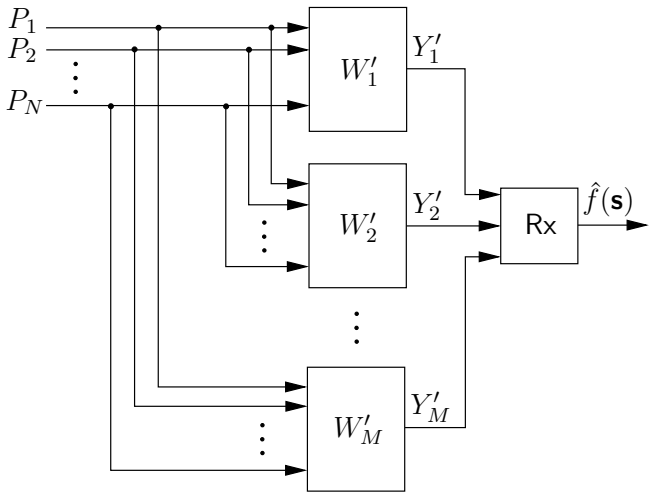


Figure 5.9: N sensor nodes communicate with a FC that is equipped with M antenna elements, which can be modeled as M parallel (equivalent) WMACs $W'_1, \ldots, W'_M$, where $P_i \in [0, P]$ for all $i = 1, \ldots, N$.

process (iid over time and antennas). If we collect the receive symbols into vectors

$$\mathbf{y}_m := (Y_m[1], \ldots, Y_m[n]) \in \mathbb{C}^n \qquad m = 1, \ldots, M \ ,$$

the communication between the sensor nodes and the FC can alternatively modeled by M parallel WMACs (cf. (5.11))

$$W'_m : \mathbb{R}_+^N \to \mathbb{R}_+ \ , \ (P_1, \ldots, P_N) \mapsto Y'_m := \|\mathbf{y}_m\|_2^2 = \sum_{i=1}^{N} \sum_{j=1}^{n} \big|H_{mi}[j]\big|^2 P_i + \Delta_m \ , \quad (5.33)$$

which is illustrated in Figure 5.9. Note that in contrast to (5.11), the effective noise at antenna m, $m = 1, \ldots, M$, is of the form

$$\begin{aligned} \Delta_m = & \underbrace{\sum_{j=1}^{n} \sum_{i=1}^{N} \sum_{\substack{i'=1 \\ i' \neq i}}^{N} H^*_{mi}[j] H_{mi'}[j] X^*_i[j] X_{i'}[j]}_{=:\Delta_{1m}} \\ & + \underbrace{2 \sum_{j=1}^{n} \sum_{i=1}^{N} \mathrm{Re}\{H_{mi}[j] X_i[j] Z^*_m[j]\}}_{=:\Delta_{2m}} + \underbrace{\sum_{j=1}^{n} \big|Z_m[j]\big|^2}_{=:\Delta_{3m}} \ . \end{aligned} \quad (5.34)$$

Remark 5.13. The equivalent WMACs $W'_1, \ldots, W'_M$ could also be seen as a distributed

multiple-input multiple-output system represented by the linear equation

$$\begin{pmatrix} Y'_1 \\ \vdots \\ Y'_M \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} \left|H_{11}[j]\right|^2 & \dots & \sum_{j=1}^{n} \left|H_{1N}[j]\right|^2 \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{n} \left|H_{M1}[j]\right|^2 & \dots & \sum_{j=1}^{n} \left|H_{MN}[j]\right|^2 \end{pmatrix} \begin{pmatrix} P_1 \\ \vdots \\ P_N \end{pmatrix} + \begin{pmatrix} \Delta_1 \\ \vdots \\ \Delta_M \end{pmatrix} .$$

Now, based on the channel outputs $Y'_1, \dots, Y'_M$, a corresponding computation-receiver makes a guess on the desired function-value, that is,

$$\mathsf{Rx} : \mathbb{R}_+^M \to \mathbb{R} \ , \ (Y'_1, \dots, Y'_M) \mapsto \hat{f}(\mathbf{s}) \ . \tag{5.35}$$

In what follows, we choose (5.35) as

$$\mathsf{Rx}(Y'_1, \dots, Y'_M) = \Psi \left[\psi \left(h_\varphi \left(\sum_{m=1}^{M} Y'_m \right) \right) \right] \ , \tag{5.36}$$

where the antenna outputs are simply accumulated before the signal post-processing is applied (see Figure 5.3). Note that the continuous function $\Psi : \mathbb{R} \to \mathbb{R}$ represents the part of the computation-receiver that is responsible for certain estimation properties such as unbiasedness and consistency. In the case of the desired function "arithmetic mean", for instance, χ ist of the form (cf. Definition 5.1)

$$\Psi(\xi) = \xi - \mathbb{E}\left\{ \psi(\Delta_3/(\alpha_{\text{arit}} n)) \right\} \ .$$

5.4.2 How Much Channel Knowledge is Needed?

In this section, we analyze the impact of transmitter CSI on the computation performance. More precisely, we assume three different types of CSI at the sensor nodes, resulting in different transmit signals summarized in Table 5.1. Accordingly, "Full CSI" refers to the case considered in Sections 5.1–5.3 where each node perfectly knows its own complex-valued channel coefficients. In contrast, "Modulus CSI" corresponds to a scenario in which each node knows only the modulus of its coefficients whereas "No CSI" means that nodes have no channel state information at all. Notice that here the FC does not need any *instantaneous* channel knowledge but may has access to some statistical CSI.

Full CSI Versus Modulus CSI at Sensor Nodes

First, we assume that the FC has a single receive antenna (i.e., $M = 1$) and consider the "Modulus CSI" case, which stands in contrast to Sections 5.1–5.3 where "Full CSI" was assumed for sensor nodes to perfectly invert their channels. The goal is to show that with regard to (5.6) there is no performance loss compared to the "Full CSI" case. We conjecture this because, by the first term on the right-hand side of (5.33), fading impacts the computation over the channel only through the instantaneous *channel gains*.

Table 5.1: Transmit signals at node i, $i = 1, \ldots, N$, depending on the available CSI.

Full CSI	$X_i[j] = \frac{\sqrt{P_i}}{H_{1i}[j]} \mathrm{e}^{\mathrm{i}\Theta_i[j]}$
Modulus CSI	$X_i[j] = \frac{\sqrt{P_i}}{\lvert H_{1i}[j] \rvert} \mathrm{e}^{\mathrm{i}\Theta_i[j]}$
No CSI	$X_i[j] = \sqrt{P_i}\, \mathrm{e}^{\mathrm{i}\Theta_i[j]}$

With "Full CSI" at sensor nodes and $M = 1$, (5.33) reduces to

$$Y_1' = n \sum_{i=1}^{N} g_\varphi(\varphi_i(S_i)) + \Delta_1 \ , \tag{5.37}$$

where the statistical moments of $\Delta_1 = \Delta_{11} + \Delta_{21} + \Delta_{31}$ are independent of fading (see Section 5.2.1). Proposition 5.7 guarantees that this is also fulfilled in the case of "Modulus CSI".

Proposition 5.7. *Regardless of the distribution of fading coefficients, the first and second moment of (5.37) under "Modulus CSI" are identical to those under "Full CSI".*

Proof. The proof is deferred to Appendix 5.A.9. □

By Proposition 5.7, the full channel knowledge at sensor nodes in Sections 5.1–5.3 can be replaced without any difference[10] by the knowledge of instantaneous channel magnitudes. This will significantly reduce the channel estimation effort in practical systems as costly phase tracking is superfluous.

Note that "Full CSI" and "Modulus CSI" are considered for the single-antenna case only because for $M > 1$ there exists a *single-input multiple-output channel* between each sensor node, say node i, and the FC. This channel is for some fixed channel use j characterized by the M channel coefficients

$$\big(H_{1i}[j], H_{2i}[j], \ldots, H_{Mi}[j]\big) \qquad i = 1, \ldots, N \ . \tag{5.38}$$

Even if we assume that node i has "Full CSI" (i.e., node i is aware of all M channel coefficients (5.38)), it is not clear how to appropriately employ this information. Node i is only able to invert a single channel per transmission, which corresponds to the case $M = 1$. Of course, one can think of transmission schemes in which each node generates some statistical CSI out of (5.38). This, however, would significantly increase the channel estimation effort as each node would have to estimate M complex-valued channel coefficients prior to transmission instead of simply one. Since the focus of this section is to reduce the channel estimation effort rather than to increase it, we did not consider the cases "Full CSI" and "Modulus CSI" for $M > 1$.

[10] Due to Markov's inequality (see Theorem B.1), the behavior of the outage probability is mainly determined by the first and second moment.

No CSI at Sensor Nodes

Let us now consider a FC with $M \geq 1$ antennas in order to show that for a large class of fading distributions, there is *no need for instantaneous CSI* at the sensor nodes, provided that the FC has some a priori knowledge about the fading statistics. Towards this end, we model the channel gains in (5.33) as

$$\left|H_{mi}[j]\right|^2 \coloneqq r_i^{-\alpha_i}\left|F_{mi}[j]\right|^2 .$$

Here and hereafter, each $r_i^{-\alpha_i}$ describes the slow-fading part of the channel (i.e., the slowly varying mean path attenuation) where $\alpha_i \geq 2$ is the path-loss exponent and $r_i \geq 0$ is the distance between node i, $i = 1, \ldots, N$, and the FC. On the other hand, $F_{mi}[j] \in \mathbb{C}$ captures the random fast-fading part induced by multipath propagation. The notion of *performance loss* is defined in this subsection as follows.

Definition 5.3 (Performance Loss)**.** The loss of performance due to the lack of CSI at nodes is quantified by

$$\lambda(L) \coloneqq \left| \kappa_H \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{j=1}^{n} \left|H_{mi}[j]\right|^2 P_i - \frac{1}{N} \sum_{i=1}^{N} P_i \right| , \tag{5.39}$$

where $L \coloneqq nNM$ and $\kappa_H \in \mathbb{R}$ denotes some statistical CSI to be specified later. For given κ_H, we say that the performance loss vanishes if $\lambda(L) \to 0$ (in a probabilistic sense) as $L \to \infty$.

Remark 5.14. The intuition behind (5.39) is to have a reasonable measure of distance between $\sum_{m,i,j} |H_{mi}[j]|^2 P_i$ and $\sum_i P_i$ when employing statistical CSI at the FC, since $\sum_i P_i = \sum_i g(\varphi_i(S_i))$ contains all the relevant information about $f(\mathbf{s})$.

Block-Fading With Equal Mean Path Losses Suppose that the channel gains are constant for the duration of n channel uses so that $H_{mi}[1] = \cdots = H_{mi}[n]$ for all m, i. Furthermore, suppose that the coefficients are iid over nodes and antennas with

$$\mathbb{E}\left\{\left|H_{mi}[1]\right|^2\right\} = r_i^{-\alpha_i}\mathbb{E}\left\{\left|F_{mi}[1]\right|^2\right\} = r_1^{-\alpha_1}\left((\sigma_{11}^{(1)})^2 + |\mu_{11}^{(1)}|^2\right) < \infty .$$

Here, $\mu_{11}^{(1)} \coloneqq \mathbb{E}\{F_{11}[1]\} \in \mathbb{C}$ captures the line-of-sight components and $(\sigma_{11}^{(1)})^2 \coloneqq \mathbb{V}\mathrm{ar}\{F_{11}[1]\} > 0$ denotes some finite variance of fast-fading effects. Prominent examples among others are Rice fading as well as Rayleigh fading (i.e., if $\mu_{11}^{(1)} = 0$).

This model reflects a homogeneous propagation environment with nodes located at similar distances to the FC, which results in equal mean path losses. Then, under mild conditions, the averaging property of (5.33) allows a simple correction of fading effects at the FC, which further reduces the channel estimation effort in comparison to "Modulus CSI".

Proposition 5.8. *Suppose that the FC scales the sum of channel outputs, $\sum_m Y'_m$, by*

$$\kappa_H \coloneqq \left(L r_1^{-\alpha_1} \left((\sigma_{11}^{(1)})^2 + |\mu_{11}^{(1)}|^2 \right) \right)^{-1} .$$

Then, for some given $\varepsilon > 0$ and $N \in \mathbb{N}$ there exists $L_0 = L_0(\varepsilon, N)$ such that $\lambda(L) \leq \varepsilon$ with probability one for all $L' \coloneqq L/n = NM \geq L_0$.

Proof. The proof is deferred to Appendix 5.A.10. □

Proposition 5.8 requires that κ_H is a priori known to the FC. This information can be obtained from an unbiased estimation of $r_1^{-\alpha_1}((\sigma_{11}^{(1)})^2 + |\mu_{11}^{(1)}|^2)$: During an initialization phase, all nodes concurrently transmit with unit power (i.e., $p_i \equiv 1$ for all $i = 1, \ldots, N$) and block length n large enough so that the FC can obtain a sufficiently good estimate of the second moment directly from the sum of channel outputs, $\sum_{m=1}^{M} Y'_m$.

Corollary 5.2. *Assume "No CSI" at the sensor nodes (see Table 5.1) and let*

$$r_1^{-\alpha_1} \left((\sigma_{11}^{(1)})^2 + |\mu_{11}^{(1)}|^2 \right) = 1 . \tag{5.40}$$

Then, $\lambda(L) \to 0$ with probability one as $L' \coloneqq L/n = NM \to \infty$. In other words, if condition (5.40) is fulfilled, for sufficiently large L' there is no need for channel estimation, neither at the sensor nodes nor at the FC.

Remark 5.15. Although Definition 5.3 and Proposition 5.8 consider only the first term on the right-hand side of

$$\sum_{m=1}^{M} Y'_m = \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{j=1}^{n} \left| H_{mi}[j] \right|^2 P_i + \sum_{m=1}^{M} \Delta_m , \tag{5.41}$$

with the computation receiver of (5.36) modified to

$$\mathsf{Rx}(Y'_1, \ldots, Y'_M) = \hat{f}(\mathbf{s}) = \Psi \left[\psi \left(h_\varphi \left(\kappa_H \sum_{m=1}^{M} Y'_m \right) \right) \right] ,$$

the results remain valid if we would also incorporate the effective noise $\kappa_H \sum_m \Delta_m$.

The results above indicate that for sufficiently large values of NM (or, equivalently, $L = nNM$ for some given n), CSI at sensor nodes is not necessary, provided that some knowledge about the channel statistics is available at the FC. From the proof of Proposition 5.8 and with the law of the iterated logarithm (see Theorem B.15 in conjunction with Remark B.4), we can conclude that

$$\lambda \in \mathcal{O} \left(\sqrt{\frac{\log_{\mathrm{e}} \log_{\mathrm{e}} NM}{NM}} \right)$$

almost surely. The numerical examples in Section 5.4.3 will demonstrate that already $M = 2$ antennas at the FC lead to a noticeable performance gain compared with the "Full CSI" case with a single antenna at the FC.

Independent and Identically Distributed Fading If we have flat fading with a sufficiently short coherence time, then the first term on the right-hand side of (5.41) has $L = nNM$ iid channel gains. By Proposition 5.8, we therefore have

$$\lambda \in \mathcal{O}\left(\sqrt{\frac{\log_{\mathrm{e}} \log_{\mathrm{e}} nNM}{nNM}}\right) \tag{5.42}$$

almost surely. This implies that rapid changes of the fading environment can be beneficial when no CSI is available at nodes.

Since almost sure convergence implies convergence in probability, (5.42) leads us, along with Theorem 5.41, to the conclusion that at least $\lambda \in \mathcal{O}(1/(nNM))$ in probability. On the other hand, Theorems B.11 and B.12 provide conditions for a faster convergence. For instance, if and only if the fading distributions are such that for all $\varepsilon > 0$ there exist positive constants $C(\varepsilon)$ and $\tau(\varepsilon)$ such that

$$\forall t \in [-\tau(\varepsilon), \tau(\varepsilon)] : \mathbb{E}\left\{\exp_{\mathrm{e}}\left(t \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{j=1}^{n} |H_{mi}[j]|^2 P_i\right)\right\} \leq C(\varepsilon) \mathrm{e}^{\varepsilon |t| n} ,$$

then we even have that (5.39) vanishes *exponentially fast* in probability, that is,

$$\lambda \in \mathcal{O}\left(\frac{1}{\rho^{nNM}}\right)$$

in probability for some $1 < \rho < \infty$.

Independent but not Identically Distributed Fading Now, we consider a heterogeneous propagation environment in which the fading coefficients are independent but have different distributions. More precisely, we assume

$$\mathbb{E}\left\{\left|H_{mi}[j]\right|^2\right\} = r_i^{-\alpha_i}\left((\sigma_{mi}^{(j)})^2 + |\mu_{mi}^{(j)}|^2\right) < \infty$$
$$(\sigma_{mi}^{(j)})^2 \coloneqq \mathbb{V}\mathrm{ar}\{F_{mi}[j]\} > 0 ,$$

for $m = 1, \ldots, M$, $i = 1, \ldots, N$, and $j = 1, \ldots, n$.

Proposition 5.9. *Let $\forall m, i, j : U_{mi}[j] \coloneqq |H_{mi}[j]|^2 P_i$ with finite second moments and*

$$\sum_{m=1}^{\infty} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{\mathbb{V}\mathrm{ar}\{U_{mi}[j]\}}{(mij)^2} < \infty . \tag{5.43}$$

Suppose that the FC scales the sum of channel outputs, $\sum_m Y'_m$, by

$$\kappa_H \coloneqq \left(\sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{j=1}^{n} r_i^{-\alpha_i}\left((\sigma_{mi}^{(j)})^2 + |\mu_{mi}^{(j)}|^2\right)\right)^{-1} ,$$

which is known a priori. Then, for any $\varepsilon > 0$ there exists $L_0(\epsilon)$ such that $\lambda(L) \leq \varepsilon$ with probability one for all $L = nNM$ with $\min\{n, N, M\} \geq L_0$.

Proof. The proof is deferred to Appendix 5.A.11. □

Note that if $1/\kappa_H$ is too small, the noise amplification due to the scaling of (5.41) may be unacceptable. In such cases, the mitigation of fading effects can be divided into two parts:

(i) Each sensor node estimates its $r_i^{-\alpha_i}$, $i = 1, \ldots, N$, in order to appropriately adapt to the slowly varying path attenuation by power control.

(ii) The FC scales (5.41) by

$$\kappa_H := \left(\sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{j=1}^{n} \mathbb{E}\left\{ \left| F_{mi}[j] \right|^2 \right\} \right)^{-1} .$$

Remark 5.16. Note that such a two-step procedure is necessary if the channel gains vary too fast, since then reliably tracking and applying CSI at sensor nodes may not be possible.

5.4.3 Numerical Examples

Now, we provide two numerical examples in order to validate the results of Section 5.4.2. Similar to Section 5.3, we consider the scenario in which the sensor nodes measure temperature-values that are uniformly iid in the interval $[5\,°\mathrm{C}, 25°\mathrm{C}]$. The performance measure is again chosen to be the outage probability (5.6) along with the distortion measure (5.5), that is, for $\varepsilon > 0$

$$\mathbb{P}\big(|D_n| \geq \varepsilon\big) = \mathbb{P}\left(\left| \frac{\mathsf{Rx}(Y_1', \ldots, Y_M') - f(\mathbf{s})}{f_{\max} - f_{\min}} \right| \geq \varepsilon \right) . \tag{5.44}$$

Example 5.6. Consider the homogeneous case of iid Rician fading with unit mean path attenuations: $r_i^{-\alpha_i} = 1$ and $\mathrm{Re}\{F_{mi}[j]\}, \mathrm{Im}\{F_{mi}[j]\} \sim \mathcal{N}_{\mathbb{R}}(\sqrt{.125}, .375)$ for all m, i, j. The network consists of $N = 25$ nodes, the block length is $n = 15$, the desired function is chosen to be the "arithmetic mean", and $P = \sigma_Z^2 = 1$. A comparison of the block-fading case with the iid fading case for a FC with $M = 1, 2, 4$ antennas is depicted, with regard to (5.44), in Figure 5.10. △

Figure 5.10 confirms that there is no benefit from having "Full CSI" instead of "Modulus CSI", as stated in Proposition 5.7. Moreover, the plots show that already for $M = 2$ antenna elements and relatively *small* values of n, N, "No CSI" outperforms the "Full CSI" case (with a single antenna). Since the assumptions on the fading coefficients fulfill the conditions of Corollary 5.2, there was no need for channel estimation neither at the sensor nodes nor at the FC.

Example 5.7. Let $N = 250$, the block length $n = 100$, and "geometric mean" be the desired function. Consider a heterogeneous Rician fading environment in which

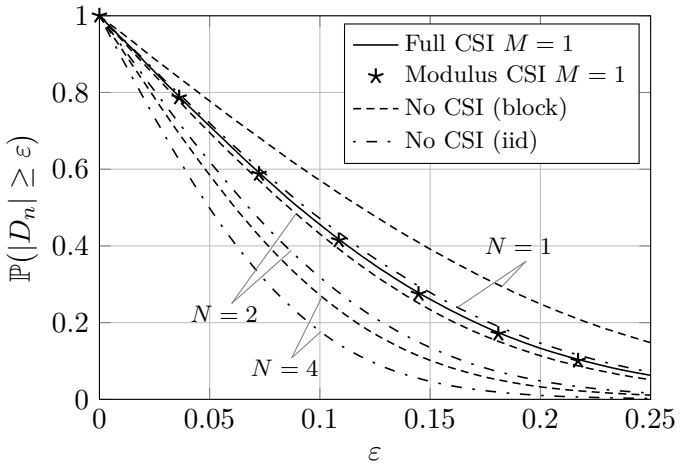


Figure 5.10: Full CSI and Modulus CSI ($M = 1$) vs. No CSI ($M = 1, 2, 4$) for the desired function "arithmetic mean" in a network of $N = 25$ sensor nodes.

$\mathrm{Re}\{\mu_{mi}^{(j)}\}$, $\mathrm{Im}\{\mu_{mi}^{(j)}\}$, and $(\sigma_{mi}^{(j)})^2$ are uniformly drawn from $[0.1, 1]$, for all m, i, j. A comparison of "Full CSI" with "No CSI" for different numbers of antennas at the FC is depicted, with regard to (5.44), in Figure 5.11. △

The plots in Figure 5.11 demonstrate that if n, N are appropriately chosen, CSI at the nodes provides no advantage already for $M = 1$ antenna elements at the FC. This fact can be used to significantly reduce the complexity of sensor nodes.

5.5 Summary and Conclusions

In this chapter, we proposed a simple *analog* transmission scheme for reliably and efficiently computing nomographic functions of measurement data over a wireless multiple-access channel. In order to relax the need for perfect synchronization as it was assumed in the previous chapters, the nodes transmit some random sequences of symbols at a transmit power that is proportional to the individual pre-processed sensor readings. As a consequence, only a coarse frame synchronization is required so that the scheme is robust against synchronization errors on the symbol and phase level. The second essential part of the scheme consists of an analog computation-receiver that is designed to appropriately estimate desired function values from the post-processed received sum of transmit energies. Since the estimator has to be matched to the desired function, we considered two canonical function examples and proposed corresponding receivers with good statistical properties. We are strongly convinced that other functions of practical relevance can also be computed in a beneficial manner using the proposed technique.

Numerical comparisons with a TDMA- and a CDMA-based approach have shown that in terms of computation accuracy, the proposed analog computation scheme has

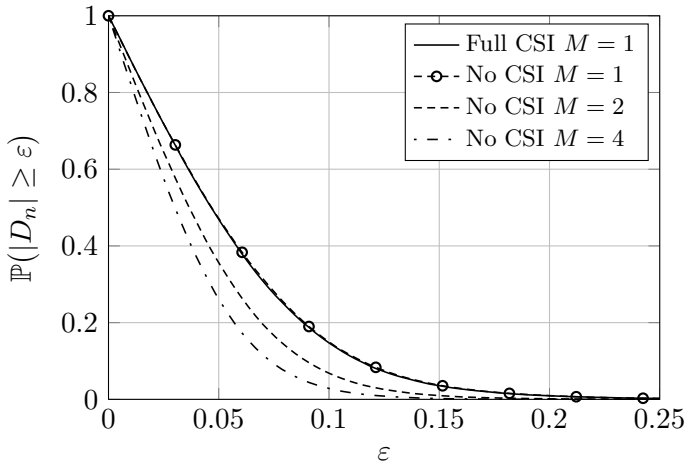


Figure 5.11: Full CSI and Modulus CSI ($M = 1$) vs. No CSI ($M = 1, 2, 4$) for the desired function "geometric mean" in a network of $N = 250$ sensor nodes.

the potential to achieve huge performance gains over separation-based strategies. In addition to the weaker requirements regarding the synchronization of sequences, the scheme needs no explicit protocol structure, which significantly reduces overhead. Computation schemes following the described design rule are therefore energy and complexity efficient and can be easily implemented in practice. Finally, the hardware-effort is reduced as well since energy consuming digital components (e.g., analog-to-digital converters, micro-controllers, registers) are not necessarily needed.

For the main part of the chapter, it was assumed that the sensor nodes have full access to complex-valued channel state information prior to transmissions in order to invert their channels. Since accurately estimating complex-valued channel state information at sensor nodes is generally a difficult and costly task, we devoted the remaining part of the chapter to the central question of how much channel knowledge is actually needed and how the channel estimation effort can significantly be reduced. In this context, we first have shown that knowing the channel magnitude at the sensor nodes is sufficient to achieve the same performance as with full (complex-valued) channel state information. It was further shown that for a wide range of fading distributions, no channel state information is needed at the transmitters provided that the FC has access to some statistical channel knowledge and is equipped with multiple antennas.

Instead of computation rates, the main figure of merit in this chapter was the achievable outage performance with regard to some predefined distortion. Even though this constitutes a reasonable measure of robustness, an open problem for future work could be to determine the computation rates that are achievable with the proposed analog scheme. Comparing these rates with those provided in Chapter 4 would then help to better understand the trade-off between synchronization effort and achievable computation rate performance.

Appendix 5.A Proofs

5.A.1 Proof of Proposition 5.1

Let $g'_\varphi \coloneqq n g_\varphi$ and $h'_\varphi \coloneqq h_\varphi / n$ so that we have to show that

$$h'_\varphi \left(\sum_i g'_\varphi(\xi_i) \right) = \sum_i \xi_i \,,$$

with $\xi_i \in [\varphi_{\min}, \varphi_{\max}]$, $i = 1, \ldots, N$, holds if and only if g_φ and h_φ are affine functions. The "$\Leftarrow$" direction is trivial whereas the other direction is shown by contradiction. Suppose g'_φ is bijective and continuous but not affine. Then, there exist at least two points $(\xi_1, \ldots, \xi_N)$ and $(\tilde{\xi}_1, \ldots, \tilde{\xi}_N)$ in $[\varphi_{\min}, \varphi_{\max}]^N$ with

$$\sum_i \xi_i \neq \sum_i \tilde{\xi}_i \qquad \text{but} \qquad \sum_i g'_\varphi(\xi_i) = \sum_i g'_\varphi(\tilde{\xi}_i) \,.$$

By the last equation, we have

$$\sum_i \xi_i = h'_\varphi \left(\sum_i g'_\varphi(\xi_i) \right) = h'_\varphi \left(\sum_i g'_\varphi(\tilde{\xi}_i) \right) = \sum_i \tilde{\xi}_i \,,$$

which however contradicts $\sum_i \xi_i \neq \sum_i \tilde{\xi}_i$. Hence, g'_φ is affine and so is g_φ. Moreover, we have $h_\varphi(\sum_i g'_\varphi(\xi_i)) = h_\varphi(n g_\varphi(\sum_i \xi_i) + \tilde{c})$ for some $\tilde{c} \in \mathbb{R}$, from which we conclude that h_φ is an affine function as well, with $h_\varphi \equiv g_\varphi^{-1} - c$ and some constant $c \in \mathbb{R}$ that depends on g_φ.

5.A.2 Proof of Proposition 5.2

Since the sum terms of $\Delta|\boldsymbol{s}$ are neither identically distributed nor independent, the convergence to a normal distribution is not clear. Therefore, let us, for some fixed $\boldsymbol{s} \in \mathcal{S}^N$, rearrange the sum in order to obtain:

$$\begin{aligned}
\Delta|\boldsymbol{s} &= \Delta_1|\boldsymbol{s} + \Delta_2|\boldsymbol{s} + \Delta_3 \\
&= \sum_{k=1}^{K} \sum_{j=1}^{n} \sqrt{p'_k} C_k[j] + 2 \sum_{i=1}^{N} \sum_{j=1}^{2n} \sqrt{p_i} U_{ij} Z'_j + \sum_{j=1}^{n} \big| Z[j] \big|^2 \\
&= \sum_{j=1}^{n} \Bigg[\sum_{k=1}^{K} \sqrt{p'_k} \cos(\Theta'_k[j]) + \sum_{i=1}^{N} \sqrt{p_i} \Big(\mathrm{Re}\{Z[j]\} \cos(\Theta_i[j]) \\
&\qquad + \mathrm{Im}\{Z[j]\} \sin(\Theta_i[j]) \Big) + \big| Z[j] \big|^2 \Bigg] \\
&= \sum_{j=1}^{n} \Lambda_j \,.
\end{aligned}$$

This makes clear that $\Lambda_1, \ldots, \Lambda_n$ are iid nondegenerate (see Definition B.1) random variables. Moreover, for every $N, P, \sigma_Z^2 < \infty$ and every compact set $\mathcal{S}$ it follows that

$$\mathbb{E}\{\Lambda_1^2 \,|\, \mathbf{s} = \boldsymbol{s}\} = 2\left(\sum_{k=1}^{K} p_k' + \sigma_Z^2 \sum_{i=1}^{N} p_i + \sigma_Z^4\right)$$

is finite. Hence, the assertion follows from Theorem B.7 with (5.19) and $\mathbb{E}\{\Delta \,|\, \mathbf{s} = \boldsymbol{s}\} = n\sigma_Z^2$.

5.A.3 Proof of Proposition 5.3

With the notions introduced in Definition 5.1 in mind, we can write (5.20), for some fixed $n \in \mathbb{N}$, as

$$\hat{f}_n(\mathbf{s}) = f(\mathbf{s}) + \frac{1}{\alpha_{\text{arit}} N n}\left(\Delta - n\sigma_Z^2\right) .$$

From this, we immediately conclude

$$\mathbb{E}\left\{\hat{f}_n(\mathbf{s}) \,|\, \mathbf{s} = \boldsymbol{s}\right\} = f(\boldsymbol{s}) + \frac{1}{\alpha_{\text{arit}} N n}\left(\mathbb{E}\{\Delta \,|\, \mathbf{s} = \boldsymbol{s}\} - n\sigma_Z^2\right)$$

so that the proposition follows since $\forall \boldsymbol{s} \in \mathcal{S}^N : \mathbb{E}\{\Delta \,|\, \mathbf{s} = \boldsymbol{s}\} = \mathbb{E}\{\Delta_3\} = n\sigma_Z^2$.

5.A.4 Proof of Proposition 5.4

According to Definition B.3, we have to prove that

$$\forall \varepsilon > 0 \, \forall \boldsymbol{s} \in \mathcal{S}^N : \lim_{n\to\infty} \mathbb{P}\left(\left|\hat{f}_n(\boldsymbol{s}) - f(\boldsymbol{s})\right| \geq \varepsilon\right) = 0 .$$

To this end, let $\boldsymbol{s} \in \mathcal{S}^N$, $c \coloneqq 1/(\alpha_{\text{arit}} N) > 0$ and $\varepsilon > 0$ be arbitrary but fixed. By the proof of Proposition 5.3, we know that

$$E_n(\boldsymbol{s}) \coloneqq \hat{f}_n(\boldsymbol{s}) - f(\boldsymbol{s}) = \frac{c}{n}\Big(\Delta|\boldsymbol{s} - \mathbb{E}\{\Delta_3\}\Big) .$$

Hence, as $\mathbb{E}\{\Delta\} = \mathbb{E}\{\Delta_3\}$, we obtain

$$\begin{aligned}
\mathbb{P}\left(|E_n(\boldsymbol{s})| \geq \varepsilon\right) &= \mathbb{P}\left(E_n(\boldsymbol{s})^2 \geq \varepsilon^2\right) \\
&= \mathbb{P}\left(\frac{c^2}{n^2}\left(\Delta|\boldsymbol{s} - \mathbb{E}\{\Delta_3\}\right)^2 \geq \varepsilon^2\right) \\
&\leq \frac{c^2}{(n\varepsilon)^2}\mathbb{E}\left\{\left(\Delta - \mathbb{E}\{\Delta\}\right)^2 \,|\, \mathbf{s} = \boldsymbol{s}\right\} \qquad (5.45) \\
&= \frac{c^2}{(n\varepsilon)^2}\sigma_{\Delta|\boldsymbol{s}}^2 , \qquad (5.46)
\end{aligned}$$

where (5.45) follows from Markov's inequality (see Theorem B.1). By (5.19), we have for $N, P, \sigma_Z^2 < \infty$ that $\sigma_{\Delta|\boldsymbol{s}}^2 \in \mathcal{O}(n)$ so that the right-hand side of the above inequality goes to zero as the block length n tends to infinity. Since $\varepsilon > 0$ and $\boldsymbol{s} \in \mathcal{S}^N$ are both arbitrary, this completes the proof.

5.A.5 Proof of Lemma 5.2

Since $\Delta_3 \sim \chi^2_{2n}$, the probability density of Δ_3 is given by

$$p_{\Delta_3} : \mathbb{R} \to \mathbb{R}_+ \, , \; \xi \mapsto p_{\Delta_3}(\xi) = \frac{1}{\sigma_Z^{2n}\Gamma(n)} \xi^{n-1} \mathrm{e}^{-\xi/\sigma_Z^2} \mathbb{1}_{[0,\infty)}(\xi) \, ,$$

where $\Gamma(z)$, with $\mathrm{Re}\{z\} > 0$, denotes the Gamma function [AS64, p. 255]. Hence, one obtains

$$\mathbb{E}\left\{\psi(\Delta_3/\alpha_{\mathrm{geo}}n)\right\} = \frac{1}{\sigma_Z^{2n}\Gamma(n)} \int_0^\infty \xi^{n-1} \exp_{\mathrm{e}}\left(-\left(\tfrac{\alpha_{\mathrm{geo}}Nn-\sigma_Z^2}{\sigma_Z^2\alpha_{\mathrm{geo}}Nn}\right)\xi\right) \mathrm{d}\xi \, . \tag{5.47}$$

Now, assume $\sigma_Z^2 < \alpha_{\mathrm{geo}}Nn$ and note that

$$\Gamma(z) = \int_0^\infty \xi^{z-1}\mathrm{e}^{-\xi}\mathrm{d}\xi = k^z \int_0^\infty \xi^{z-1}\mathrm{e}^{-k\xi}\mathrm{d}\xi \, ,$$

$\mathrm{Re}\{z\} > 0$, which holds for every $k \in \mathbb{C}$ with $\mathrm{Re}\{k\} > 0$ [AS64, p. 255]. Substituting this into (5.47) with an appropriately chosen k proves (i). As for (ii), if $\sigma_Z^2 < \alpha_{\mathrm{geo}}Nn$, then it follows from (i) that

$$\lim_{n\to\infty} \left(\frac{\alpha_{\mathrm{geo}}Nn}{\alpha_{\mathrm{geo}}Nn - \sigma_Z^2}\right)^n = \lim_{n\to\infty}\left(1+\frac{u}{n}\right)^{-n} = \mathrm{e}^{-u} \, ,$$

where $u \coloneqq -\frac{\sigma_Z^2}{\alpha_{\mathrm{geo}}N}$. This proves the lemma.

5.A.6 Proof of Proposition 5.5

Let $N, P, \sigma_Z^2 < \infty$ and $\varepsilon > 0$ be arbitrary but fixed and let $\{\hat{f}_n\}_{n\in\mathbb{N}}$ be the sequence of receivers given by (5.24). We show that for every $\boldsymbol{s} \in \mathcal{S}^N$, the outage probability $\mathbb{P}(|D_n| \geq \varepsilon \,|\, \mathbf{s} = \boldsymbol{s})$ goes to zero as the block length n tends to infinity. Towards this end, note that $f(\boldsymbol{s}) > 0$, $\beta(\boldsymbol{s}) > 0$, $\lambda_n > 0$, and $\Xi|\boldsymbol{s} > 0$ for all $\boldsymbol{s} \in \mathcal{S}^N$. By (5.25), we have

$$\mathbb{P}\Big(|D_n| \geq \varepsilon \,\Big|\, \mathbf{s} = \boldsymbol{s}\Big) = \mathbb{P}\left(\frac{\Xi}{\lambda_n} \geq 1 + \frac{\varepsilon}{\beta(\boldsymbol{s})} \,\Big|\, \mathbf{s} = \boldsymbol{s}\right) + \mathbb{P}\left(1 - \frac{\Xi}{\lambda_n} \geq \frac{\varepsilon}{\beta(\boldsymbol{s})} \,\Big|\, \mathbf{s} = \boldsymbol{s}\right) \, .$$

An application of Theorem B.1 yields an upper bound on the first sum term:

$$\begin{aligned}\mathbb{P}\left(\frac{\Xi}{\lambda_n} \geq 1 + \frac{\varepsilon}{\beta(\boldsymbol{s})} \,\Big|\, \mathbf{s} = \boldsymbol{s}\right) &= \mathbb{P}\left(\log_{\mathrm{e}}\left(\frac{\Xi}{\lambda_n}\right) \geq \log_{\mathrm{e}}\left(1 + \frac{\varepsilon}{\beta(\boldsymbol{s})}\right) \Big|\, \mathbf{s} = \boldsymbol{s}\right) \\ &\leq \frac{\mathbb{E}\{\log_{\mathrm{e}}(\Xi|\boldsymbol{s})\} - \log_{\mathrm{e}}(\lambda_n)}{\log_{\mathrm{e}}(1+\varepsilon/\beta(\boldsymbol{s}))} \, . \end{aligned} \tag{5.48}$$

By part (ii) of Lemma 5.2, we have

$$\lim_{n\to\infty} \log_{\mathrm{e}}(\lambda_n) = \log_{\mathrm{e}}\left(\lim_{n\to\infty}\lambda_n\right) = \frac{\sigma_Z^2}{\alpha_{\mathrm{geo}}N} \, .$$

Due to the results on the distribution functions of random variables that are functions of other random variables [Shi96, pp. 239–240], we obtain

$$\mathbb{E}\{\log_{\mathrm{e}}(\Xi|\boldsymbol{s})\} = \frac{\mathbb{E}\left\{\Delta|\boldsymbol{s}/(\alpha_{\mathrm{geo}}n)\right\}}{N} = \frac{\sigma_Z^2}{\alpha_{\mathrm{geo}}N} \; ,$$

where we used $\mathbb{E}\{\Delta \,|\, \mathbf{s} = \boldsymbol{s}\} = n\sigma_Z^2$ in the last step. Combining the results shows that the upper bound in (5.48) tends to zero as $n \to \infty$.

As for $\mathbb{P}(1 - \Xi/\lambda_n \geq \varepsilon/\beta(\boldsymbol{s}) \,|\, \mathbf{s} = \boldsymbol{s})$, note that we can focus on $\varepsilon/\beta(\boldsymbol{s}) < 1$ since $\Xi|\boldsymbol{s}/\lambda_n > 0$. With this in mind, we have

$$\mathbb{P}\left(1 - \frac{\Xi}{\lambda_n} \geq \frac{\varepsilon}{\beta(\boldsymbol{s})} \,\Big|\, \mathbf{s} = \boldsymbol{s}\right) = \mathbb{P}\left(\frac{\lambda_n}{\Xi} \geq \frac{1}{1 - \varepsilon/\beta(\boldsymbol{s})} \,\Big|\, \mathbf{s} = \boldsymbol{s}\right) \; .$$

Proceeding essentially along the same lines as above shows that this probability goes to zero with $n \to \infty$ and therefore

$$\forall \varepsilon > 0 \, \forall \boldsymbol{s} \in \mathcal{S}^N : \lim_{n\to\infty} \mathbb{P}\Big(|D_n| \geq \varepsilon \,|\, \mathbf{s} = \boldsymbol{s}\Big) = 0 \; .$$

As $\varepsilon > 0$ was chosen arbitrarily, this completes the proof.

5.A.7 Proof of Lemma 5.3

Let $\mathcal{S}$ be an arbitrary compact set and $N < \infty$ any fixed natural number. In accordance with Definition 5.2, we have that

$$\Xi|\boldsymbol{s} = \psi\left(\frac{\Delta|\boldsymbol{s}}{\alpha_{\mathrm{geo}}n}\right) = \exp_{\mathrm{e}}\left(\frac{\Delta|\boldsymbol{s}}{\alpha_{\mathrm{geo}}Nn}\right) \; ,$$

where $N, \alpha_{\mathrm{geo}} > 0$. Since the exponential function is continuous and strictly increasing, it follows for the conditional distribution function $P_{\Xi|\boldsymbol{s}} : \mathbb{R}_{++} \to [0,1]$,

$$P_{\Xi|\boldsymbol{s}}(\xi) = \mathbb{P}(\Xi \leq \xi \,|\, \mathbf{s} = \boldsymbol{s}) = \mathbb{P}(\Delta \leq \alpha_{\mathrm{geo}}Nn\log_{\mathrm{e}}(\xi) \,|\, \mathbf{s} = \boldsymbol{s}) = P_{\Delta|\boldsymbol{s}}(\alpha_{\mathrm{geo}}Nn\log_{\mathrm{e}}(\xi)) \; .$$

With Corollary 5.1, we conclude that for n sufficiently large, $\Delta|\boldsymbol{s}$ can be approximated (in distribution) by a random variable $\tilde{\Delta}|\boldsymbol{s} \sim \mathcal{N}_{\mathbb{R}}(N\sigma_Z^2, \sigma_{\Delta|\boldsymbol{s}}^2)$. An immediate consequence of this is that for sufficiently large values of n, the conditional distribution function of $\Delta|\boldsymbol{s}$ can be approximated by

$$P_{\tilde{\Delta}|\boldsymbol{s}} : \mathbb{R} \to [0,1] \; , \; \xi \mapsto P_{\tilde{\Delta}|\boldsymbol{s}}(\xi) = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{\xi - n\sigma_Z^2}{\sigma_{\Delta|\boldsymbol{s}}\sqrt{2}}\right) \; ,$$

that is,

$$\forall \boldsymbol{s} \in \mathcal{S}^N \, \forall \delta > 0 \, \exists n_0(\boldsymbol{s}, \delta) \, \forall n \geq n_0 : \left\| P_{\Delta|\boldsymbol{s}} - P_{\tilde{\Delta}|\boldsymbol{s}} \right\|_\infty < \delta \; .$$

Writing this shortly as $P_{\Delta|\boldsymbol{s}} \approx P_{\tilde{\Delta}|\boldsymbol{s}}$, we conclude that for n large enough, along with Theorem B.5, that

$$P_{\Xi|\boldsymbol{s}}(\xi) \approx P_{\tilde{\Xi}|\boldsymbol{s}}(\xi) = P_{\tilde{\Delta}|\boldsymbol{s}}(\alpha_{\mathrm{geo}} N n \log_{\mathrm{e}}(\xi)) = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{\alpha_{\mathrm{geo}} N n \log_{\mathrm{e}}(\xi) - n\sigma_Z^2}{\sigma_{\Delta|\boldsymbol{s}}\sqrt{2}}\right). \tag{5.49}$$

Observe that (5.49) describes the distribution function of a log-normally distributed random variable with parameters $\frac{\sigma_Z^2}{\alpha_{\mathrm{geo}} N} =: \mu_\Xi$ and $\left(\sigma_{\Delta|\boldsymbol{s}}/(\alpha_{\mathrm{geo}} N n)\right)^2 =: \sigma_{\Xi|\boldsymbol{s}}^2$. Thus, for some $\boldsymbol{s} \in \mathcal{S}^N$, $\Xi|\boldsymbol{s}$ can be approximated (in distribution) by $\tilde{\Xi}|\boldsymbol{s} \sim \mathcal{LN}(\mu_\Xi, \sigma_{\Xi|\boldsymbol{s}}^2)$.

5.A.8 Proof of Proposition 5.6

Note that it is sufficient to show (5.28) and (5.29). Let $\boldsymbol{s} \in \mathcal{S}^N$ be arbitrary but fixed. Because $|D_n|\boldsymbol{s}| = |\gamma(\boldsymbol{s})^{-1}\Xi|\boldsymbol{s} - \beta(\boldsymbol{s})|$ is continuous in $\Xi|\boldsymbol{s}$, Lemma 5.3 in combination with Theorem B.5 allows for the approximation of $|D_n|\boldsymbol{s}|$ (in distribution) by

$$|\tilde{D}_n|\boldsymbol{s}| = \left|\frac{1}{\gamma(\boldsymbol{s})}\tilde{\Xi}|\boldsymbol{s} - \beta(\boldsymbol{s})\right|,$$

where the probability distribution function of $\tilde{\Xi}|\boldsymbol{s} \sim \mathcal{LN}(\mu_\Xi, \sigma_{\Xi|\boldsymbol{s}}^2)$ is given by the right-hand side of (5.49). Since $0 < \beta(\boldsymbol{s}), \gamma(\boldsymbol{s}) < \infty$, we have

$$\begin{aligned}
\mathbb{P}\Big(|D_n| \geq \varepsilon \,\Big|\, \mathbf{s} = \boldsymbol{s}\Big) &\approx \mathbb{P}\Big(|\tilde{D}_n| \geq \varepsilon \,\Big|\, \mathbf{s} = \boldsymbol{s}\Big) \\
&= 1 - \mathbb{P}\Big(-\varepsilon < \tilde{D}_n < \varepsilon \,\Big|\, \mathbf{s} = \boldsymbol{s}\Big) \\
&= 1 - \mathbb{P}\Big(-\varepsilon < \gamma(\mathbf{s})^{-1}\tilde{\Xi} - \beta(\mathbf{s}) < \varepsilon \,\Big|\, \mathbf{s} = \boldsymbol{s}\Big).
\end{aligned}$$

This immediately leads to

$$\mathbb{P}\Big(|\tilde{D}_n| \geq \varepsilon \,\Big|\, \mathbf{s} = \boldsymbol{s}\Big) = \begin{cases} 1 - P_{\tilde{\Xi}|\boldsymbol{s}}(\rho^+(\boldsymbol{s},\varepsilon)) + P_{\tilde{\Xi}|\boldsymbol{s}}(\rho^-(\boldsymbol{s},\varepsilon)) & 0 < \varepsilon < \beta(\boldsymbol{s}) \\ 1 - P_{\tilde{\Xi}|\boldsymbol{s}}(\rho^+(\boldsymbol{s},\varepsilon)) & \beta(\boldsymbol{s}) \leq \varepsilon < \infty \end{cases} \tag{5.50}$$

with

$$\begin{aligned}
\rho^+(\boldsymbol{s},\varepsilon) &\coloneqq \gamma(\boldsymbol{s})\big(\beta(\boldsymbol{s}) + \varepsilon\big) \\
\rho^-(\boldsymbol{s},\varepsilon) &\coloneqq \gamma(\boldsymbol{s})\big(\beta(\boldsymbol{s}) - \varepsilon\big).
\end{aligned}$$

Inserting the right-hand side of (5.49) into the expression (5.50) and using the fact that $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$, for all $x \in \mathbb{R}$, shows (5.28) and (5.29) and thus completes the proof.

5.A.9 Proof of Proposition 5.7

For simplicity, we consider here only the first moment. The second moment can be treated in a similar manner. Towards this end, we show that the expected value

$$\mathbb{E}\{\Delta_1\} = \mathbb{E}\{\Delta_{11}\} + \mathbb{E}\{\Delta_{21}\} + \mathbb{E}\{\Delta_{31}\} \tag{5.51}$$

is independent of fading so that we have to analyze the error terms Δ_{11} and Δ_{21} in (5.34), which depend on the channel coefficients.

Obviously, $\mathbb{E}\{\Delta_{21}\} \equiv 0$ always holds due to the independence of transmit signals, fading and noise. As far as Δ_{11} is concerned, the following lemma from [Sch83] is helpful.

Lemma 5.4 (Schatte)**.** *Let A, B be real independent random variables. If A or B is uniformly distributed in $[0, 2\pi)$, then the reduced sum $C = (A + B) \bmod 2\pi$ is uniformly distributed in $[0, 2\pi)$ as well.*

We write the random fading coefficient between sensor node i, $i = 1, \ldots, N$, and the FC at channel use j, $j = 1, \ldots, n$, in polar form as

$$H_{1i}[j] = |H_{1i}[j]| \mathrm{e}^{\mathrm{i}\Lambda_{1i}[j]} ,$$

where $\Lambda_{1i}[j]$ is used to denote the corresponding random phase. Then, in the case of "Modulus CSI", Δ_{11} becomes

$$\Delta_{11} = 2 \sum_{j=1}^{n} \sum_{i=2}^{N} \sum_{i'=1}^{i-1} \sqrt{P_i P_{i'}} \cos(\Delta\Lambda_{ii'}[j] + \Delta\Theta_{ii'}[j])$$

with $\Delta\Lambda_{ii'}[j] \coloneqq \Lambda_{1i}[j] - \Lambda_{1i'}[j]$ and $\Delta\Theta_{ii'}[j] \coloneqq \Theta_i[j] - \Theta_{i'}[j]$. Note that the moduli of the channel coefficients are removed but the phases are still present. Let $D_{ii'}[j] \coloneqq \Delta\Lambda_{ii'}[j] + \Delta\Theta_{ii'}[j]$ and $C_{ii'}[j] \coloneqq \cos(D_{ii'}[j])$.

A sufficient condition for $\mathbb{E}\{\Delta_{11}\}$ to be zero is that for all i, i', j and any distribution of $\Delta\Lambda_{ii'}[j]$, $\mathbb{E}\{C_{ii'}[j]\} = 0$ holds. Since $\Theta_i[j]$ and $\Theta_{i'}[j]$ are uniformly iid in $[0, 2\pi)$ for all $i \neq i'$, Lemma 5.4 implies that the differences $\Delta\Theta_{ii'}[j]$ are uniformly iid in $[0, 2\pi)$ as well. Moreover, since $\Delta\Theta_{ii'}[j]$ and $\Delta\Lambda_{ii'}[j]$ are independent for all $m, i', i \neq i'$, we conclude from Lemma 5.4 that all $D_{ii'}[j]$ are uniformly distributed in $[0, 2\pi)$. Since $\mathbb{E}\{\cos(X)\} = \int_0^{2\pi} \frac{1}{2\pi} \cos(x) \mathrm{d}x = 0$ with X being uniformly distributed in $[0, 2\pi)$, it follows $\mathbb{E}\{C_{ii'}[j]\} = 0$, for all i' and $i \neq i'$. Finally, considering the linearity of the expectation operator, we conclude $\mathbb{E}\{\Delta_{11}\} \equiv 0$ so that (5.51) behaves as in the case of "Full CSI", regardless of the fading distribution.

5.A.10 Proof of Proposition 5.8

In order to show that (5.39) vanishes with probability one as $L' = L/n = NM$ goes to infinity,

$$\kappa_H \sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{j=1}^{n} \big|H_{mi}[1]\big|^2 P_i = \frac{\sum_{m=1}^{M} \sum_{i=1}^{N} \sum_{j=1}^{n} \big|H_{mi}[1]\big|^2 P_i}{L r_1^{-\alpha_1} ((\sigma_{11}^{(1)})^2 + |\mu_{11}^{(1)}|^2)} = \frac{\sum_{m=1}^{M} \sum_{i=1}^{N} \big|H_{mi}[1]\big|^2 P_i}{L' r_1^{-\alpha_1} ((\sigma_{11}^{(1)})^2 + |\mu_{11}^{(1)}|^2)}$$

has to tend to $\frac{1}{N}\sum_{i=1}^{N} P_i$ with probability one. Let $U_{mi}[1] \coloneqq |H_{mi}[1]|^2 P_i$ and note that

$$\forall m,i : \mathbb{E}\left\{|U_{mi}[1]|\right\} = r_i^{-\alpha_i}\mathbb{E}\left\{\left|F_{mi}[1]\right|^2\right\} P_i = r_1^{-\alpha_1}\left((\sigma_{11}^{(1)})^2 + |\mu_{11}^{(1)}|^2\right) P_i < \infty$$

because P_i is finite for all $i = 1, \ldots, N$. Since this is a necessary and sufficient condition of Kolmogorov's strong law of large numbers for iid variables (see Theorem B.13 along with Remark B.3), it follows that

$$\frac{\sum_{m=1}^{M}\sum_{i=1}^{N} U_{mi}[1]}{L' r_1^{-\alpha_1}\left((\sigma_{11}^{(1)})^2 + |\mu_{11}^{(1)}|^2\right)} \xrightarrow{\text{a.s.}} \frac{\mathbb{E}\left\{\left|H_{11}[1]\right|^2\right\}\xi}{r_1^{-\alpha_1}\left((\sigma_{11}^{(1)})^2 + |\mu_{11}^{(1)}|^2\right)} = \xi \,,$$

for $L' \to \infty$, where ξ is either $\frac{1}{N}\sum_i P_i$ or $\lim_{N\to\infty}\frac{1}{N}\sum_i P_i$ (the limit exists and is finite with probability one). As a consequence, $\lambda(L)$ vanishes almost surely so that for every fixed $\varepsilon > 0$ and N, there exists an $L_0(\varepsilon, N)$ such that $\lambda(L) \leq \varepsilon$ with probability one for all $L' \geq L_0$, which proves the proposition.

5.A.11 Proof of Proposition 5.9

Let $\forall m,i,j : U_{mi}[j] \coloneqq |H_{mi}[j]|^2 P_i$ so that

$$\kappa_H \sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{j=1}^{n}\left|H_{mi}[j]\right|^2 P_i = \frac{\sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{j=1}^{n} U_{mi}[j]}{\sum_{m=1}^{M}\sum_{i=1}^{N}\sum_{j=1}^{n} r_i^{-\alpha_i}\left((\sigma_{mi}^{(j)})^2 + |\mu_{mi}^{(j)}|^2\right)} \,.$$

If condition (5.43) is fulfilled, Kolmogorov's strong law of large numbers for independent but not identically distributed variables (see Theorem B.14) yields

$$\frac{\sum_m\sum_i\sum_j U_{mi}[j]}{\sum_m\sum_i\sum_j r_i^{-\alpha_i}\left((\sigma_{mi}^{(j)})^2 + |\mu_{mi}^{(j)}|^2\right)} \xrightarrow{\text{a.s.}} \frac{\sum_m\sum_i\sum_j \mathbb{E}\{|H_{mi}[j]|^2\}\xi}{\sum_m\sum_i\sum_j r_i^{-\alpha_i}\left((\sigma_{mi}^{(j)})^2 + |\mu_{mi}^{(j)}|^2\right)} = \xi$$

as $\min\{n, N, M\} \to \infty$, where $\xi = \lim_{N\to\infty}\frac{1}{N}\sum_i P_i$. As a consequence, (5.39) vanishes almost surely so that for every fixed $\varepsilon > 0$, there exists an $L_0(\varepsilon)$ such that $\lambda(L) \leq \varepsilon$ with probability one for all L with $\min\{n, N, M\} \geq L_0$. This completes the proof.

6

Conclusion

> *Wir wissen nichts – das ist das Erste. Deshalb sollen wir sehr bescheiden sein – das ist das Zweite. Dass wir nicht behaupten zu wissen, wenn wir nicht wissen – das ist das Dritte.*

Karl R. Popper, *Alles Leben ist Problemlösen*, 1994

In this thesis, we dealt with the problem of reliably and efficiently computing real-valued linear and nonlinear functions of the measurements in wireless sensor networks. Instead of avoiding the interference that occurs when distinct sensor nodes concurrently access the same frequency spectrum, we considered it as a freely available *computational resource*. In this context, we first asked the fundamental question which real-valued functions of several variables can essentially be computed directly *in the air*. To our own surprise, this question can shortly be answered with *every*. The reason is that by using an appropriate signal pre- and post-processing strategy, in fact every function of the sensor readings can be matched to the linear structure of the wireless channel. In other words, the wireless channel is capable of performing all basic arithmetic operations over the reals and can therefore be regarded as a *computer*.

Of course, this insight alone is not sufficient in order to evaluate if harnessing interference is a superior paradigm because separation-based approaches also allow for computing every function of the sensor readings: transmit all the data to the fusion center by employing some form of interference avoidance/cancellation and compute the desired function-value afterwards. Towards this end, we proposed two novel transmission schemes (digital and analog) and studied their properties as well as their achievable performance. It turns out that both schemes can be used to compute a variety of linear and nonlinear real-valued functions at computation rates that are in a wide range of

network operating points not achievable with separation-based methods. As a conclusion, we therefore advise harnessing interference for efficiently and reliably computing real-valued functions in wireless sensor networks.

Outlook

In Sections 3.4, 4.4, and 5.5, which concluded Chapters 3–5, we already outlined some open problems and future research directions. Instead of repeating them here, we provide some further topics that can be of interest for future work.

- **Proof of Concept:** As mentioned above, we proposed in this thesis a digital and an analog transmission scheme for efficiently computing real-valued functions over the wireless channel. A next natural step would therefore be to provide a first proof of concept in order to demonstrate that the shown performance gains can be realized in practice. Towards this end, the computation schemes have to be implemented on, for instance, software defined radio devices that establish an appropriate testbed. It is expected that the corresponding system design fundamentally differs from that for standard message transfer. In this context, one has also to think about appropriate wake-up strategies, which are necessary to initiate function computations.

- **Secure Function Computation:** As the wireless communication channel is an open and freely accessible medium, it is inherently vulnerable to eavesdropping. Currently used security methods heavily rely on higher-layer implementations of cryptographic concepts, which typically require that users share a private key. If the key is not yet available, a dedicated secure channel between prospective communication partners has to be established for private key exchange. In contrast, physical layer security methods, which are gaining more and more importance [SCW+11, BB11], do not need such an additional channel. They provide security (in some sense) by explicitly taking into account the properties of the wireless channel. This paradigm can also be relevant for *secure function computation* since many wireless sensor network applications not only require to efficiently compute function-values at fusion centers but also to keep them secret from non-legitimate receivers. Some very recent considerations in this direction can be found in [TNG11, LA14, DDMP14].

- **The Reverse Computation Problem:** Consider a fusion center that already knows all the raw sensor readings $s_1, \ldots, s_N$ but does not have enough computational capabilities in order to compute the desired function-value $f(s_1, \ldots, s_N)$. An approach that is related to the studies in this thesis is to distribute the information into the neighborhood, which then helps to compute the function by simultaneously transmitting the (processed) information back. One of the corresponding challenges is to determine the exact trade-off between the required computation and communication resources.

- **Theory of Implementation:** The results of Chapters 3 and 5 demonstrate that *analog systems* can be well suited for efficiently solving computation problems in wireless sensor networks. Indeed, Boche and Mönich show in [BM11, BM14] that with ordinary sampling, purely analog linear time-invariant systems are not always stably representable in discrete time domain. Moreover, sampling is usually followed by quantization, which generates additional instabilities that are not always controllable by oversampling [BM10, BM12]. Thus, digital signal processing has some fundamental limits and analog systems are gaining more and more attention in the sensor network community. Under these circumstances, some results of this thesis can also be seen in a much broader context as they provide a step towards a general *theory of implementation*. The development of such a theory is certainly a big challenge but it would be invaluable for practice as it would give a comprehensive answer to the question: With which kind of technology (i.e., analog, digital, hybrid) communication systems should be implemented and what are the corresponding costs in terms of energy and hardware [BM14].

Note that the results of this thesis are general in the sense that they are not limited to wireless sensor networks. In particular, the proposed computation schemes can be used as a building block for complex in-network processing schemes and wireless networking protocols. Furthermore, they can be well suited for establishing efficient machine type communication in next-generation cellular networks, which are expected to handle the massive access of machines, objects, and devices of any kind by a very dense infrastructure (i.e., cell densification).

A

Real-Valued Functions of Several Variables

As real-valued functions of several variables are the main point of interest in this thesis, we recap on the following pages some useful definitions and relationships in order to make the thesis more self-contained. A comprehensive treatment of the topic can be found for instance in [Wol66, Lan73, Fle77] as well as in the standard analysis textbook [Rud76].

To this end, let $\|\cdot\|_2$ be the ordinary Euclidean norm, which induces a metric on the Euclidean N-space $\mathbb{R}^N$. Then, the notion of a convergent sequence in $\mathbb{R}^N$ is as follows.

Definition A.1 (Convergent Sequence)**.** A sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ of elements $\boldsymbol{x}_k \in \mathbb{R}^N$, $k \in \mathbb{N}$, is called *convergent*, if there exists some $\boldsymbol{x} \in \mathbb{R}^N$ such that

$$\forall \varepsilon > 0 \, \exists k_0(\varepsilon) \in \mathbb{N} \, \forall k \geq k_0(\varepsilon) : \|\boldsymbol{x}_k - \boldsymbol{x}\|_2 < \varepsilon \, . \tag{A.1}$$

If (A.1) applies, we call $\boldsymbol{x}$ the *limit* of the sequence and write $\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}$ or $\boldsymbol{x}_k \to \boldsymbol{x}$ for $k \to \infty$, respectively.

The following theorem can be used to examine properties of sequences in $\mathbb{R}^N$ by examining properties of sequences on the real line.

Theorem A.1. *The sequence $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ of elements $\boldsymbol{x}_k \coloneqq (x_1^{(k)}, \ldots, x_N^{(k)}) \in \mathbb{R}^N$ converges to the limit $\boldsymbol{x} \coloneqq (x_1, \ldots, x_N) \in \mathbb{R}^N$ if and only if for each fixed i, $i = 1, \ldots, N$, $\lim_{k\to\infty} x_i^{(k)} = x_i$.*

Proof. The statement follows immediately from the inequalities [Wol66, p. 4]

$$\max\left\{|x_1^{(k)} - x_1|, \ldots, |x_N^{(k)} - x_N|\right\} \leq \|\boldsymbol{x}_k - \boldsymbol{x}\|_2 \leq N \max\left\{|x_1^{(k)} - x_1|, \ldots, |x_N^{(k)} - x_N|\right\}.$$

□

Now, we provide the essential definition of a real-valued function of N variables.

Definition A.2 (Real-Valued Function of N Variables)**.** Let $N \in \mathbb{N}$, $N \geq 2$, and $\mathbb{D}$ be some subset of $\mathbb{R}^N$. An instruction f that assigns every $\boldsymbol{x} \coloneqq (x_1, \ldots, x_N) \in \mathbb{D}$ to exactly one element

$$y = f(\boldsymbol{x}) \coloneqq f(x_1, \ldots, x_N) \in \mathbb{R}$$

is called *real-valued function of N variables*.

Remark A.1. We denote real-valued functions of N variables often shortly as *multivariate functions* or as *real-valued functions of several variables*.

Special cases that are frequently mentioned throughout the thesis are linear functions of several variable variables.

Definition A.3 (Linear Function)**.** A multivariate function $f : \mathbb{D} \subseteq \mathbb{R}^N \to \mathbb{R}$ is said to be *linear* (or $\mathbb{R}$-*linear*) if and only if

$$f(\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2) = \alpha_1 f(\boldsymbol{x}_1) + \alpha_2 f(\boldsymbol{x}_2)$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{D}$ and all $\alpha_1, \alpha_2 \in \mathbb{R}$.

Example A.1. Let $f : \mathbb{R}^N \to \mathbb{R}$ with $f(x_1, \ldots, x_N) = f(\boldsymbol{x}) = \sum_{i=1}^N x_i$. Then, f is linear since for all $\boldsymbol{x}_1 = (x_1^{(1)}, \ldots, x_N^{(1)}) \in \mathbb{R}^N$, all $\boldsymbol{x}_2 = (x_1^{(2)}, \ldots, x_N^{(2)}) \in \mathbb{R}^N$ and all $\alpha_1, \alpha_2 \in \mathbb{R}$, we have

$$\begin{aligned} f(\alpha_1 \boldsymbol{x}_1 + \alpha_2 \boldsymbol{x}_2) &= \sum_{i=1}^N \left(\alpha_1 x_i^{(1)} + \alpha_2 x_i^{(2)} \right) \\ &= \alpha_1 \sum_{i=1}^N x_i^{(1)} + \alpha_2 \sum_{i=1}^N x_i^{(2)} \\ &= \alpha_1 f(\boldsymbol{x}_1) + \alpha_2 f(\boldsymbol{x}_2) \, . \end{aligned}$$

△

In contrast to the notion of the limit of a sequence in $\mathbb{R}^N$ such as provided in Definition A.1, the limit of a multivariate functions is defined as follows.

Definition A.4 (Limit of a Multivariate Function)**.** Let $f : \mathbb{D} \subseteq \mathbb{R}^N \to \mathbb{R}$ be a multivariate function and $\boldsymbol{x}_0 \in \mathbb{R}^N$ some accumulation point of $\mathbb{D}$. If there exists $c \in \mathbb{R}$ such that for every sequence $\{\boldsymbol{x}_k\}_{k \in \mathbb{N}}$, $\boldsymbol{x}_k \in \mathbb{D}$, $\boldsymbol{x}_k \neq \boldsymbol{x}_0$, the implication

$$\lim_{k \to \infty} \boldsymbol{x}_k = \boldsymbol{x}_0 \Rightarrow \lim_{k \to \infty} f(\boldsymbol{x}_k) = c$$

holds true, then c is called *the limit* of f at $\boldsymbol{x}_0$.

Definition A.5 (Continuity)**.** A multivariate function $f : \mathbb{D} \subseteq \mathbb{R}^N \to \mathbb{R}$ is said to be *(pointwise) continuous* if

$$\forall \boldsymbol{x}_0 \in \mathbb{D}\, \forall \varepsilon > 0\, \exists \delta > 0\, \forall \boldsymbol{x} \in \mathbb{D} : \|\boldsymbol{x} - \boldsymbol{x}_0\|_2 < \delta \Rightarrow \big|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)\big| < \varepsilon \; .$$

This so called (ε, δ)-definition goes back to Weierstrass. An equivalent definition of (pointwise) continuity, which is more related to the limit of functions given in Definition A.4, is the following.

Definition A.6 (Continuity)**.** A multivariate function $f : \mathbb{D} \subseteq \mathbb{R}^N \to \mathbb{R}$ is said to be *continuous* at $\boldsymbol{x}_0 \in \mathbb{D}$ if

$$\forall \{\boldsymbol{x}_k\}_{k \in \mathbb{N}} \subset \mathbb{D} : \lim_{k \to \infty} \boldsymbol{x}_k = \boldsymbol{x}_0 \Rightarrow \lim_{k \to \infty} f(\boldsymbol{x}_k) = f(\boldsymbol{x}_0) \; .$$

If f is continuous at every point of $\mathbb{D}$, then f said to be (pointwise) *continuous*.

Remark A.2. Note that continuity requires in contrast to Definition A.4 that f is defined at $\boldsymbol{x}_0$, which means that $\boldsymbol{x}_0$ has to be an element of $\mathbb{D}$. If f is continuous at $\boldsymbol{x}_0$, then we are allowed to exchange the order of the limits, that is,

$$\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} f(\boldsymbol{x}) = f\Big(\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} \boldsymbol{x}\Big) \; .$$

If some function $f : \mathbb{D} \subseteq \mathbb{R}^N \to \mathbb{R}$ is continuous, then it is said to belong to the space $\mathscr{C}^0(\mathbb{D})$.

A definition of continuity that is stronger than pointwise continuity is that of uniform continuity.

Definition A.7 (Uniform Continuity)**.** A multivariate function $f : \mathbb{D} \subseteq \mathbb{R}^N \to \mathbb{R}$ is said to be *uniformly continuous* if

$$\forall \varepsilon > 0\, \exists \delta > 0\, \forall \boldsymbol{x}, \boldsymbol{x}_0 \in \mathbb{D} : \|\boldsymbol{x} - \boldsymbol{x}_0\|_2 < \delta \Rightarrow \big|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)\big| < \varepsilon \; .$$

Remark A.3. The obvious difference to (pointwise) continuity (cf. Definition A.5) is that δ depends only on the choice of ε but not on the choice of $\boldsymbol{x}_0$.

The following theorem relates pointwise and uniform continuity by imposing some restriction on the domain $\mathbb{D}$.

Theorem A.2 (Heine·Cantor)**.** *Let $f : \mathbb{D} \subset \mathbb{R}^N \to \mathbb{R}$ be some multivariate continuous function and $\mathbb{D}$ a compact set. Then, f is* uniformly continuous*.*

Proof. A proof of the theorem, which is a corollary of the more general fact that every continuous function between two metric spaces is uniformly continuous if the domain is compact, can be found for instance in [Rud76, p. 91]. □

Another property of continuous functions that are defined on some compact subset of the Euclidean N-space provides the following theorem, which is also known as the *extreme value theorem*.

Theorem A.3 (Weierstrass)**.** *Let $f : \mathbb{D} \subset \mathbb{R}^N \to \mathbb{R}$ for some $N \in \mathbb{N}$ and $\mathbb{D}$ a compact set. If f is continuous, then it attains a minimum and a maximum. That is,*

$$\exists \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{D}\, \forall \boldsymbol{x} \in \mathbb{D} : f(\boldsymbol{x}_1) \leq f(\boldsymbol{x}) \leq f(\boldsymbol{x}_2) \; .$$

Proof. The proof of the theorem can be found for instance in [Rud76, pp. 89]. □

Besides pointwise and uniform continuity, there exist some other notions of continuity such as, for instance, Hölder continuity.

Definition A.8 (Hölder Continuity)**.** A multivariate function $f : \mathbb{D} \subset \mathbb{R}^N \to \mathbb{R}$ is called (L, α)-*Hölder continuous* if and only if there exist positive real constants L and α, where $0 < \alpha \leq 1$, such that

$$\forall \boldsymbol{x}, \boldsymbol{x}_0 \in \mathbb{D} : \left| f(\boldsymbol{x}) - f(\boldsymbol{x}_0) \right| \leq L \|\boldsymbol{x} - \boldsymbol{x}_0\|_2^{\alpha} \; .$$

Example A.2. Let $f : \mathbb{R}_+ \to \mathbb{R}$, $f(x) = \sqrt{x}$. Then, f is $(1, 1/2)$-Hölder continuous. To see this, let $x, x_0 \in \mathbb{R}_+$ be chosen without loss of generality such that $x \geq x_0 \geq 0$. It follows

$$|x - x_0| = x - x_0 = \left(\sqrt{x} - \sqrt{x_0}\right)\left(\sqrt{x} - \sqrt{x_0}\right) \geq \left(\sqrt{x} - \sqrt{x_0}\right)^2 = \left|\sqrt{x} - \sqrt{x_0}\right|^2$$

and thus

$$\left|\sqrt{x} - \sqrt{x_0}\right| \leq \sqrt{|x - x_0|} \; .$$

△

Remark A.4. It can be easily verified that Hölder continuity implies uniform continuity. To this end, let $f : \mathbb{D} \subset \mathbb{R}^N \to \mathbb{R}$ be (L, α)-Hölder continuous and set $\delta \coloneqq (\varepsilon L^{-1})^{1/\alpha}$ for $\varepsilon > 0$ arbitrary but fixed. Then, $\|\boldsymbol{x} - \boldsymbol{x}_0\|_2 < \delta$ implies $|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)| < \varepsilon$ for all $\boldsymbol{x}, \boldsymbol{x}_0 \in \mathbb{D}$. The converse, however, is not necessarily true.

Whereas the above given definitions of limits and continuity for functions of several variables are virtually simple extensions of the univariate case, the notion of differentiability fundamentally differs. Therefore, in order to introduce the meaning of differentiability of a function of several variables, we have to first provide the concept of a partial derivative.

Definition A.9 (Partial Derivative)**.** Let $f : \mathbb{D} \subset \mathbb{R}^N \to \mathbb{R}$ and $\mathbb{D}$ an open set. Then, f is said to be *partial differentiable* at $\boldsymbol{x} \coloneqq (x_1, \ldots, x_N) \in \mathbb{D}$ with respect to x_i, $i = 1, \ldots, N$, if the limit

$$\lim_{h \to 0} \frac{f(x_1, \ldots, x_{i-1}, x_i + h, x_{i+1}, \ldots, x_N) - f(x_1, \ldots, x_N)}{h}$$

exists. The limit is called the *partial derivative* of f at $\boldsymbol{x}$ with respect to x_i and is written as $\frac{\partial f}{\partial x_i}\Big|_{\boldsymbol{x}}$.

The partial derivative (if existent) reflects the behavior of a multivariate function in a certain direction and therefore does not provide information about the behavior in the neighborhood of some given point. This fact leads us to the following definition.

Definition A.10 (Differentiability)**.** Let $f : \mathbb{D} \subset \mathbb{R}^N \to \mathbb{R}$ and $\mathbb{D}$ an open set. The multivariate function f is said to be *differentiable at* $\boldsymbol{x}_0 \in \mathbb{D}$ if there exists a point $\boldsymbol{a} \in \mathbb{R}^N$ and a function $g : \mathbb{R}^N \to \mathbb{R}$ that is defined on a neighborhood $\mathcal{U}(\boldsymbol{x}_0)$ of $\boldsymbol{x}_0$ such that[1]

(i) $\forall \boldsymbol{x} \in \mathcal{U}(\boldsymbol{x}_0) : f(\boldsymbol{x}) = f(\boldsymbol{x}_0) + \boldsymbol{a}^\mathsf{T}(\boldsymbol{x} - \boldsymbol{x}_0) + g(\boldsymbol{x})\|\boldsymbol{x} - \boldsymbol{x}_0\|_2$

(ii) $\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} g(\boldsymbol{x}) = 0$.

If f is differentiable at every $\boldsymbol{x}_0 \in \mathbb{D}$, then f is said to be *differentiable*.

Remark A.5. If a multivariate function f is differentiable at some point $\boldsymbol{x}_0$ of its open domain, then $\boldsymbol{a}$ in Definition A.10 coincides with the vector of partial derivatives, that is, the gradient

$$\boldsymbol{a} = (a_1, \ldots, a_N) = \mathrm{grad} f(\boldsymbol{x}_0) \coloneqq \left(\frac{\partial f}{\partial x_1}\bigg|_{\boldsymbol{x}_0}, \ldots, \frac{\partial f}{\partial x_N}\bigg|_{\boldsymbol{x}_0} \right) .$$

Based on the partial derivatives, the following theorem provides a simple way to check whether a multivariate function is differentiable or not.

Theorem A.4. *Let $f : \mathbb{D} \subset \mathbb{R}^N \to \mathbb{R}$ and $\mathbb{D}$ an open set. Assume that all N partial derivatives of f exist at every $\boldsymbol{x} \in \mathbb{D}$ and that they are continuous. Then, f is differentiable.*

Proof. A proof of the theorem can be found for instance in [Lan73, p. 68]. □

Remark A.6. It has to be emphasized that the pure existence of partial derivatives does not necessarily imply differentiability.

Throughout the thesis, the notion of a *nomographic function* in the sense of Buck [Buc79] plays an important role. Under the same designation, however, Warmus gave an alternative definition already in 1959 [War59] (see also [Eve82, Sec. 4.4]). In order to make the differences clear, we provide Warmus's idea of a nomographic function below. As a preliminary, we have to first introduce the concept of the rank of a function of two or three variables.

Definition A.11. Let $\mathbb{D}_1$ and $\mathbb{D}_2$ be subsets of the real line and $f : \mathbb{D}_1 \times \mathbb{D}_2 \to \mathbb{R}$. Then, f is said to be of *rank* $n > 1$ if and only if there exist n functions $\varphi_i : \mathbb{D}_1 \to \mathbb{R}$ as well as n functions $\psi_i : \mathbb{D}_2 \to \mathbb{R}$ such that

$$f(x_1, x_2) = \varphi_1(x_1)\psi_1(x_2) + \varphi_2(x_1)\psi_2(x_2) + \cdots + \varphi_n(x_1)\psi_n(x_2) ,$$

[1] Recall that $\mathcal{U}(\boldsymbol{x}_0)$ is said to be a *neighborhood* of $\boldsymbol{x}_0 \in \mathbb{R}^N$ if it contains an *open* neighborhood $\mathcal{U}_\varepsilon(\boldsymbol{x}_0) \coloneqq \{\boldsymbol{x} \in \mathbb{R}^N \,|\, \|\boldsymbol{x} - \boldsymbol{x}_0\|_2 < \varepsilon\}$ for some $\varepsilon > 0$.

where n cannot be reduced. That is, there do not exist functions $\varphi_i' : \mathbb{D}_1 \to \mathbb{R}$ and $\psi_i' : \mathbb{D}_2 \to \mathbb{R}$, $i = 1, \ldots, n-1$, such that

$$f(x_1, x_2) = \varphi_1'(x_1)\psi_1'(x_2) + \varphi_2'(x_1)\psi_2'(x_2) + \cdots + \varphi_{n-1}'(x_1)\psi_{n-1}'(x_2) \; .$$

Definition A.12. Let $\mathbb{D}_1, \mathbb{D}_2, \mathbb{D}_3$ be subsets of the real line and $i \in \{1, 2, 3\}$ be fixed. A function $f : \mathbb{D}_1 \times \mathbb{D}_2 \times \mathbb{D}_3 \to \mathbb{R}$ is said to be of *rank n with respect to x_i* if and only if it is of rank n when considered as a function of the two variables x_i and (x_j, x_k), $j \neq k \neq i$.

Definition A.13 (Warmus's Nomographic Function)**.** Let $f : \mathbb{D}_1 \times \mathbb{D}_2 \times \mathbb{D}_3 \to \mathbb{R}$ with $\mathbb{D}_i \subset \mathbb{R}$, $i = 1, 2, 3$, some compact intervals. Then, f is said to be a *nomographic function (in the sense of Warmus)* if and only if there exist functions $\varphi_{ij} : \mathbb{D}_i \to \mathbb{R}$, $x_i \mapsto \varphi_{ij}(x_i)$, such that the following two conditions are satisfied:

(i) $f(x_1, x_2, x_3) = \det \begin{pmatrix} \varphi_{11}(x_1) & \varphi_{12}(x_1) & \varphi_{13}(x_1) \\ \varphi_{21}(x_2) & \varphi_{22}(x_2) & \varphi_{23}(x_2) \\ \varphi_{31}(x_3) & \varphi_{32}(x_3) & \varphi_{33}(x_3) \end{pmatrix}$

(ii) With respect to each of the variables x_1, x_2, x_3, f is of rank greater than 1.

Comparing the determinant representation above, written out in full as

$$\begin{aligned} f(x_1, x_2, x_3) = {} & \varphi_{11}(x_1)\varphi_{22}(x_2)\varphi_{33}(x_3) + \varphi_{12}(x_1)\varphi_{23}(x_2)\varphi_{31}(x_3) \\ & + \varphi_{13}(x_1)\varphi_{21}(x_2)\varphi_{32}(x_3) - \varphi_{11}(x_1)\varphi_{23}(x_2)\varphi_{32}(x_3) \\ & - \varphi_{12}(x_1)\varphi_{21}(x_2)\varphi_{33}(x_3) - \varphi_{13}(x_1)\varphi_{22}(x_2)\varphi_{31}(x_3) \; , \end{aligned}$$

with

$$f(x_1, x_2, x_3) = \psi(\varphi_1(x_1) + \varphi_2(x_2) + \varphi_3(x_3))$$

(see (3.6)) reveals the difference between the two concepts, even if both are closely related to nomography, which is the almost forgotten field dealing with nomographs [Eve82].

B

Some Standard Results of Probability Theory

For quick reference, we summarize in this appendix some standard definitions and results from probability theory. To this end, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space that is rich enough in the sense that it supports all occurring random elements.

Definition B.1. A random variable X is said to be *degenerate* if and only if there exists some fixed $c \in \mathbb{R}$ such that $\mathbb{P}(X = c) = 1$. Otherwise, X is called nondegenerate.

According to the definition, a random variable is degenerate if and only if it is with probability one concentrated on a set of measure zero, which implies that a degenerate random variable is with probability one equal to its expected value and thus of zero variance. The notion can immediately be extended to the multivariate case.

Definition B.2. A multivariate random variable is said to be degenerate if and only if it is with probability one concentrated on some set of measure zero.

Obviously, a multivariate random variable is degenerate if it lies with probability one on a hyperplane and thus if its covariance matrix is *not* positive definite [Mui05, pp. 3–4].[1]

One of the most frequently used inequality in probability theory is the following.

Theorem B.1 (Markov's Inequality)**.** *Let X be a nonnegative random variable (i.e., $\mathbb{P}(X \geq 0) = 1$). Then,*

$$\forall \varepsilon > 0 : \mathbb{P}(X \geq \varepsilon) \leq \frac{\mathbb{E}\{X\}}{\varepsilon} \ .$$

[1] Note that the converse is not true because hyperplanes are not the only sets of measure zero in higher dimensions.

Proof. Let $\mathbb{1}_{\mathbb{A}}$ denote the indicator function on set $\mathbb{A}$ and note that

$$X = X\mathbb{1}_{[\varepsilon,\infty)}(X) + X\mathbb{1}_{(-\infty,\varepsilon)}(X) \geq X\mathbb{1}_{[\varepsilon,\infty)}(X) \geq \varepsilon\mathbb{1}_{[\varepsilon,\infty)}(X) \,,$$

for all $\varepsilon > 0$, since X is nonnegative. Then,

$$\mathbb{E}\{X\} \geq \varepsilon\mathbb{E}\{\mathbb{1}_{[\varepsilon,\infty)}(X)\} = \varepsilon\mathbb{P}(X \geq \varepsilon) \,,$$

for all $\varepsilon > 0$, which proves the theorem. □

An immediate consequence of Markov's inequality is the not less important Chebyshev's inequality.

Corollary B.1 (Chebyshev's Inequality)**.** *Let X be a nondegenerate random variable with finite expected value and finite variance. Then,*

$$\forall\varepsilon > 0 : \mathbb{P}\left(|X - \mathbb{E}\{X\}| \geq \varepsilon\right) \leq \frac{\mathbb{V}\mathrm{ar}\{X\}}{\varepsilon^2} \,.$$

Another fruitful inequality is the so-called union bound (or Boole's inequality)

Theorem B.2 (Union Bound)**.** *Let $A_1, A_2, \ldots$ be a countable set of events for which $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$, and $\bigcup_i A_i \in \mathcal{F}$. Then,*

$$\mathbb{P}\left(\bigcup_i A_i\right) \leq \sum_i \mathbb{P}(A_i) \,.$$

Proof. The bound follows from the σ-subadditivity of probability measure $\mathbb{P}$. See [Shi96, p. 134] for a complete proof. □

Before dealing with sequences of random variables, we first recap some of the various kinds of the convergence of random variables.

Definition B.3. The sequence $X_1, X_2, \ldots$ of random variables converges *in probability* to the random variable X, denoted as $X_n \xrightarrow{\mathbb{P}} X$, if

$$\forall\varepsilon > 0 : \lim_{n\to\infty} \mathbb{P}\left(|X_n - X| \geq \varepsilon\right) = 0 \,.$$

Definition B.4. The sequence $X_1, X_2, \ldots$ of random variables converges *almost surely (or with probability one)* to the random variable X, denoted as $X_n \xrightarrow{\text{a.s.}} X$, if

$$\mathbb{P}\left(\lim_{n\to\infty} X_n \neq X\right) = 0 \,.$$

A necessary and sufficient condition for a sequence to converge almost surely is given by the following theorem, which therefore provides an equivalent definition.

Theorem B.3. *Let $X_1, X_2, \ldots$ be a sequence of random variables. Then, $X_n \xrightarrow{\text{a.s.}} X$, for some random variable X, if and only if*

$$\forall \varepsilon > 0 : \lim_{n\to\infty} \mathbb{P}\left(\sup_{n' \geq n} \left| X_{n'} - X \right| \geq \varepsilon \right) = 0 \, .$$

Proof. For a proof, we refer to [Shi96, pp. 253–254]. □

Definition B.5. The sequence $X_1, X_2, \ldots$ of random variables converges *in distribution* to the random variable X, denoted as $X_n \xrightarrow{\text{d}} X$, if

$$\forall x \in C(F_X) : \lim_{n\to\infty} F_{X_n}(x) = F_X(x) \, .$$

Here, F_{X_n} and F_X are the cumulative distribution functions of X_n and X, respectively, and $C(F_X)$ denotes the continuity set of F_X (i.e., the set of points at which F_X is continuous).

Theorem B.4. *Let $X_1, X_2, \ldots$ be a sequence of random variables that converges to the random variable X almost surely. Then, we have the following implications:*

$$X_n \xrightarrow{\text{a.s.}} X \Rightarrow X_n \xrightarrow{\mathbb{P}} X \Rightarrow X_n \xrightarrow{\text{d}} X \, .$$

Proof. The proof can be found for instance in [Shi96, p. 256]. □

In applied probability theory, it is often of particular interest to infer from the convergence behavior of a sequence of random variables on the convergence of the sequence processed by some function.

Theorem B.5 (Mann-Wald)**.** *Let X and $\{X_n\}_{n\in\mathbb{N}}$ be random variables that take on values in a separable metric space $(\Theta, \mathcal{E}, P_X)$. Furthermore, let $g = g(x)$, $x \in \Theta$, be some measurable function that maps $(\Theta, \mathcal{E}, P_X)$ into another separable metric space $(\Theta', \mathcal{E}', P'_X)$ with $\mathbb{P}(X \in D(g)) = 0$, where $D(g)$ denotes the discontinuity set of g (i.e., the set of points at which g is discontinuous). Then,*

a) $X_n \xrightarrow{\text{a.s.}} X \Rightarrow g(X_n) \xrightarrow{\text{a.s.}} g(X)$

b) $X_n \xrightarrow{\mathbb{P}} X \Rightarrow g(X_n) \xrightarrow{\mathbb{P}} g(X)$

c) $X_n \xrightarrow{\text{d}} X \Rightarrow g(X_n) \xrightarrow{\text{d}} g(X)$.

Proof. The proof of the theorem, which is also known as the *continuous mapping theorem*, can be found for instance in [vdV98, pp. 7–8], [Shi96, p. 357], [Bil99, p. 21]. □

Theorem B.6 (Dominated Convergence)**.** *Let $X_1, X_2, \ldots$ be a sequence of random variables such that $|X_n| \leq Y$ almost surely for all $n \in \mathbb{N}$ and some random variable Y with finite expected value. Then, if $X_n \xrightarrow{\text{a.s.}} X$,*

$$\lim_{n\to\infty} \mathbb{E}\{X_n\} = \mathbb{E}\{X\} \, .$$

Proof. The theorem, which is a consequence of Lebesgue's theorem on dominated convergence, is proven for instance in [Shi96, p. 187] and [Bil95, p. 77]. □

As one of the most important results in probability theory, the central limit theorem for iid variables states that the corresponding (normalized) sum converges under rather mild conditions to a standard normal random variable.

Theorem B.7 (Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be a sequence of nondegenerate iid random variables with finite second moment (i.e., $\mathbb{E}\{X_1^2\} < \infty$). If $S_n \coloneqq \sum_{i=1}^n X_i$, then*

$$\frac{S_n - \mathbb{E}\{S_n\}}{\sqrt{\mathbb{V}\mathrm{ar}\{S_n\}}} \xrightarrow{\mathrm{d}} X \sim \mathcal{N}_{\mathbb{R}}(0,1) \ .$$

Proof. For a proof, we refer to [Shi96, p. 326] and [Bil95, Sec. 27]. □

Remark B.1. The cumulative distribution function $F_X : \mathbb{R} \to [0,1]$ of a random variable $X \sim \mathcal{N}_{\mathbb{R}}(0,1)$ is of the form

$$F_X(x) = \mathbb{P}(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{1}{2}t^2} \, \mathrm{d}t \ . \tag{B.1}$$

Since (B.1) is continuous over $\mathbb{R}$, the convergence in Theorem B.7 (see Definition B.5) is uniform, that is,

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_{T_n}(x) - F_X(x)| = 0 \ ,$$

with $F_{T_n}(x) = \mathbb{P}(T_n \leq x)$, $x \in \mathbb{R}$, the cumulative distribution function of

$$T_n \coloneqq \frac{S_n - \mathbb{E}\{S_n\}}{\sqrt{\mathbb{V}\mathrm{ar}\{S_n\}}} \ . \tag{B.2}$$

The uniform convergence of the cumulative distribution functions of (B.2) to (B.1) gives rise to the question how rapidly this occurs. The following theorem provides an answer for those sequences of iid random variables that have finite second and finite third absolute moments.

Theorem B.8 (Berry-Esseen)**.** *Let $X_1, X_2, \ldots$ be a sequence of iid nondegenerate random variables with finite $\mathbb{E}\{X_1^2\}$ and $\mathbb{E}\{|X_1|^3\}$. Then, for $n \in \mathbb{N}$*

$$\sup_{x\in\mathbb{R}} |F_{T_n}(x) - F_X(x)| \leq C_0 \frac{\beta}{\sqrt{n}} \ , \tag{B.3}$$

where $\beta \coloneqq \mathbb{E}\left\{|X_1|^3\right\} / \mathbb{V}\mathrm{ar}\{X_1\}^{3/2}$ and C_0 a finite absolute constant.

Proof. See [Shi96, pp. 374-375] for a proof in which $1/\sqrt{2\pi} \leq C_0 < 0.8$. A recent proof of the inequality that improves the upper bound on the absolute constant to $C_0 < 0.4784$ is given by Korolev and Shevtsova in [KS12]. □

Another family of limit theorems whose importance cannot be overestimated are the *laws of large numbers*. In this family a distinction is made between weak and strong laws.

Theorem B.9 (Weak Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with finite expected values and $S_n \coloneqq \sum_{i=1}^{n} X_i$. Then,*

$$\frac{S_n}{n} \xrightarrow{\mathbb{P}} \mathbb{E}\{X_1\}$$

as n tends to infinity.

Proof. The proof can be found for instance in [Shi96, p. 325]. When additionally assuming $\mathbb{V}\mathrm{ar}\{X_1\} < \infty$, then the theorem is a trivial consequence of Definition B.3 together with Chebyshev's inequality, that is,

$$\forall \varepsilon > 0 : \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\{X_1\}\right| \geq \varepsilon\right) \leq \frac{\mathbb{V}\mathrm{ar}\{S_n\}}{n^2\varepsilon^2} = \frac{\mathbb{V}\mathrm{ar}\{X_1\}}{n\varepsilon^2} \to 0 \tag{B.4}$$

as n tends to infinity. □

As for the central limit theorem, we are also interested in estimating the rate of convergence in the weak law of large numbers, which precisely means that we ask how rapidly $\mathbb{P}\left(\left|S_n/n - \mathbb{E}\{X_1\}\right| \geq \varepsilon\right)$ tends to zero with growing n. One of the simplest results is the following.

Theorem B.10. *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with finite variance and let $S_n \coloneqq \sum_{i=1}^{n} X_n$. Then,*

$$\forall \varepsilon > 0 : \mathbb{P}\left(\left|\frac{S_n}{n} - \mathbb{E}\{X_1\}\right| \geq \varepsilon\right) \in \mathcal{O}\left(\frac{1}{n}\right) .$$

Proof. As can be seen from (B.4), the assertion follows immediately from Chebyshev's inequality. □

A similar theorem can be proven if higher moments are assumed to exist.

Theorem B.11. *Let $X_1, X_2, \ldots$ be a sequence of independent random variables for which the $(t+1)^{th}$ moments exist, $t = 2, 3, \ldots$, and*

$$\exists C > 0 \, \forall n \in \mathbb{N} : \mathbb{E}\{|X_n|^{t+1}\} \leq C .$$

Then,

$$\forall \varepsilon > 0 : \mathbb{P}\left(\left|\frac{S_n - \mathbb{E}\{S_n\}}{n}\right| \geq \varepsilon\right) \in \mathcal{O}\left(\frac{1}{n^t}\right) ,$$

where $S_n \coloneqq \sum_{i=1}^{n} X_n$.

Proof. The proof can be found for instance in [Rév68, p. 54]. □

A necessary and sufficient condition to even achieve *exponential* convergence is given by the following.

Theorem B.12 (Baum-Katz-Read)**.** *Let $X_1, X_2, \ldots$ be a sequence of independent random variables and let $S_n \coloneqq \sum_{i=1}^{n} X_i$. Then,*

$$\forall \varepsilon > 0 : \mathbb{P}\left(\left|\frac{S_n - \mathbb{E}\{S_n\}}{n}\right| \geq \varepsilon\right) \in \mathcal{O}\left(\frac{1}{\rho^n}\right) , \quad 1 < \rho < \infty ,$$

if and only if for all $\varepsilon > 0$ there exist positive constants $C(\varepsilon)$ and $\tau(\varepsilon)$ such that

$$\forall t \in [-\tau(\varepsilon), \tau(\varepsilon)] : \prod_{i=1}^{n} \mathbb{E}\{\mathrm{e}^{tX_i}\} = \mathbb{E}\{\mathrm{e}^{tS_n}\} \leq C(\varepsilon)\mathrm{e}^{\varepsilon|t|n} . \tag{B.5}$$

Proof. For a proof, the reader is referred to [BKR62]. □

The following corollary, also given by Baum, Katz, and Read in [BKR62], demonstrates how condition (B.5) simplifies when a sequence of independent Gaussian random variables is considered.

Corollary B.2. *Let $X_1, X_2, \ldots$ be a sequence of independent Gaussian random variables with bounded expected values. Then, for all $\varepsilon > 0$, $\mathbb{P}(|S_n - \mathbb{E}\{S_n\}|/n \geq \varepsilon) \to 0$ exponentially fast with growing n, if and only if there exists $C \in \mathbb{R}$ such that*

$$\forall n \in \mathbb{N} : \frac{1}{n}\sum_{i=1}^{n} \mathbb{V}\mathrm{ar}\{X_i\} \leq C ,$$

that is, the sequence of averaged variances is bounded.

Remark B.2. Note that in general, $\mathbb{P}(|S_n - \mathbb{E}\{S_n\}|/n \geq \varepsilon)$ cannot vanish more rapidly than with the order $\mathcal{O}(1/\rho^n)$, for some $p \in (1, \infty)$ [Rév68, p. 56].

The *strong laws of large numbers* are a family of propositions in which the convergence in probability occurring in the weak laws is replaced by almost sure convergence. A well-known example is the following, proposed by Kolmogorov.

Theorem B.13 (Kolmogorov's Strong Law of Large Numbers for iid Variables)**.** *Let $X_1, X_2, \ldots$ be a sequence of iid random variables with finite first absolute moment (i.e., $\mathbb{E}\{|X_1|\} < \infty$). Then,*

$$\frac{S_n - \mathbb{E}\{X_1\}}{n} \xrightarrow{\text{a.s.}} 0$$

as n tends to infinity, where $S_n \coloneqq \sum_{i=1}^{n} X_i$.

Proof. For a proof see for instance [Shi96, p. 391]. □

Remark B.3. It can be shown that the condition $\mathbb{E}\{|X_1|\} < \infty$ is also necessary for $(S_n - \mathbb{E}\{X_1\})/n$ to converge to 0 with probability one [Shi96, p. 393].

Provided that the second moments exist as well, the sequence of random variables need not be identically distributed, which was also proven by Kolmogorov.

Theorem B.14 (Kolmogorov's Strong Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be a sequence of independent random variables with finite second moments and*

$$\sum_{n=1}^{\infty} \frac{\mathbb{V}\mathrm{ar}\{X_n\}}{n^2} < \infty \,.$$

Then,

$$\frac{S_n - \mathbb{E}\{S_n\}}{n} \xrightarrow{\text{a.s.}} 0$$

as n tends to infinity, where $S_n \coloneqq \sum_{i=1}^{n} X_i$.

Proof. The proof can be found for instance in [Rév68, p. 63] and [Shi96, p. 389]. □

For the special case of iid random variables, an answer to the question at which rate $(S_n - \mathbb{E}\{S_n\})/n$ converges to 0 with probability one can be inferred from Hartman's and Wintner's version of the *law of the iterated logarithm*.

Theorem B.15 (Hartman-Wintner)**.** *Let $X_1, X_2, \ldots$ be a sequence of nondegenerate iid random variables with finite second moment. Then,*

$$\mathbb{P}\left(\limsup_{n\to\infty} \frac{S_n - \mathbb{E}\{S_n\}}{\sqrt{2\,\mathbb{V}\mathrm{ar}\{X_1\} n \log_{\mathrm{e}} \log_{\mathrm{e}} n}} = 1\right) = 1$$

Proof. For a proof, the reader is referred to for instance [Bil95, p. 154]. □

Remark B.4. Since the Theorem states that

$$\limsup_{n\to\infty} \frac{S_n - \mathbb{E}\{S_n\}}{n} \sqrt{\frac{n}{2\,\mathbb{V}\mathrm{ar}\{X_1\} \log_{\mathrm{e}} \log_{\mathrm{e}} n}} = 1$$

almost surely, we can immediately specify the corresponding convergence rate to

$$\forall \varepsilon > 0 : \mathbb{P}\left(\sup_{n' \geq n} \left| \frac{S_{n'} - \mathbb{E}\{S_{n'}\}}{n'} \right| \geq \varepsilon\right) \in \mathcal{O}\left(\sqrt{\frac{\log_{\mathrm{e}} \log_{\mathrm{e}} n}{n}}\right) \,.$$

Publication List

Journal, Conference, and Workshop Papers

[1] Mario Goldenbaum, Sławomir Stańczak, and Michał Kaliszan, "On function computation via wireless sensor multiple-access channels," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, Budapest, Hungary, April 2009, pp. 1–6.

[2] Mario Goldenbaum and Sławomir Stańczak, "Computing functions via SIMO multiple-access channels: How much channel knowledge is needed?" in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, TX, USA, March 2010, pp. 3394–3397.

[3] Mario Goldenbaum and Sławomir Stańczak, "Computing the geometric mean over multiple-access channels: Error analysis and comparisons," in *Proceedings of the 44th Asilomar Conference on Signals, Systems and Computers (Asilomar)*, Pacific Grove, CA, USA, November 2010, pp. 2172–2178.

[4] Mario Goldenbaum, Rudi Abi Akl, Stefan Valentin, and Sławomir Stańczak, "On the effect of feedback delay in the downlink of multiuser OFDM systems," in *Proceedings of the 45th Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, USA, March 2011, pp. 1–6.

[5] Sławomir Stańczak, Michał Kaliszan, and Mario Goldenbaum, "Max-min fair rate control based on a saddle-point characterization of some Perron roots," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3112–3115.

[6] Sławomir Stańczak, Michał Kaliszan, and Mario Goldenbaum, "On distributed power control and transceiver optimization in wireless networks," in *Proceedings of 5th International ICST Conference on Performance Evaluation Methodologies and Tools (ValueTools)*, Cachan, France, May 2011, pp. 244–253, invited.

[7] Mario Goldenbaum, Holger Boche, and Sławomir Stańczak, "Analyzing the space of functions analog-computable via wireless multiple-access channels," in *Proceedings of the 8th International Symposium on Wireless Communication Systems (ISWCS)*, Aachen, Germany, November 2011, pp. 779–783.

[8] Mario Goldenbaum and Sławomir Stańczak, "On multiantenna sensor networks with interference: Energy consumption vs. robustness," in *Proceedings of the 16th International ITG Workshop on Smart Antennas (WSA)*, Dresden, Germany, March 2012, pp. 125–132.

[9] Mario Goldenbaum, Holger Boche, and Sławomir Stańczak, "On analog computation of vector-valued functions in clustered wireless sensor networks," in *Proceedings of the 46th Annual Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, USA, March 2012, pp. 1–6.

[10] Mario Goldenbaum, Holger Boche, and Sławomir Stańczak, "Analog computation via wireless multiple-access channels: Universality and robustness," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, April 2012, pp. 2921–2924.

[11] Meng Zheng, Mario Goldenbaum, Sławomir Stańczak, and Haibin Yu, "Fast average consensus in clustered wireless sensor networks by superposition gossiping," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, Paris, France, April 2012, pp. 1982–1987.

[12] Mario Goldenbaum, Holger Boche, and Sławomir Stańczak, "Nomographic gossiping for f-consensus," in *Proceedings of the 10th International Symposium on Modeling and Optimization in Mobile, Ad-Hoc, and Wireless Networks (WiOpt)*, Paderborn, Germany, May 2012, pp. 130–137.

[13] Sławomir Stańczak, Mario Goldenbaum, Renato L. G. Cavalcante, and Federico Penna, "On in-network computation via wireless multiple-access channels with applications," in *Proceedings of the 9th International Symposium on Wireless Communication Systems (ISWCS)*, Paris, France, August 2012, pp. 276–280, invited.

[14] Mario Goldenbaum, Holger Boche, and Sławomir Stańczak, "Reliable computation of nomographic functions over Gaussian multiple-access channels," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, May 2013, pp. 4814–4818.

[15] Mario Goldenbaum and Sławomir Stańczak, "Robust analog function computation via wireless multiple-access channels," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3863–3877, September 2013.

[16] Mario Goldenbaum, Holger Boche, and Sławomir Stańczak, "Harnessing interference for analog function computation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 61, no. 20, pp. 4893–4906, October 2013.

[17] Steffen Limmer, Sławomir Stańczak, Mario Goldenbaum, and Renato L. G. Cavalcante, "Exploiting interference for efficient distributed computation in cluster-based wireless sensor networks," in *Proceedings of the IEEE Global Conference*

on Signal and Information Processing (GlobalSIP), Austin, TX, USA, December 2013, pp. 933–936, invited.

[18] Mario Goldenbaum and Sławomir Stańczak, "On the channel estimation effort for analog computation over wireless multiple-access channels," *IEEE Wireless Communications Letters*, vol. 3, no. 3, pp. 261–264, June 2014.

[19] Andreas Kortke, Mario Goldenbaum, and Sławomir Stańczak, "Analog computation over the wireless channel: A proof of concept," in *Proceedings of the IEEE Sensors Conference*, Valencia, Spain, November 2014, pp. 1224–1227.

[20] Mario Goldenbaum, Holger Boche, and Sławomir Stańczak, "Nomographic functions: Efficient computation in clustered Gaussian sensor networks," *IEEE Transactions on Wireless Communications*, December 2014, in print.

Patents

[21] Holger Boche, Mario Goldenbaum, and Sławomir Stańczak, "Sensorfunksystem," DE Patent 10 2011 085 804, *Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V., Technische Universität München, Technische Universität Berlin*, August, 2013.

[22] Mario Goldenbaum and Sławomir Stańczak, "Funktionsvariablenwertsender, Funktionsempfänger und System," DE Patent 10 2010 000 735, *Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V.*, July, 2014.

References

[ACLY00] Rudolf Ahlswede, Ning Cai, Shuo-Yen R. Li, and Raymond W. Yeung, "Network information flow," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1204–1216, July 2000.

[AFKZ11] Rathinakumar Appuswamy, Massimo Franceschetti, Nikhil Karamchandani, and Kenneth Zeger, "Network coding for computing: Cut-set bounds," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 1015–1030, February 2011.

[ÅGB11] Johan Åkerberg, Mikael Gidlund, and Mats Björkman, "Future research challenges in wireless sensor and actuator networks targeting industrial automation," in *Proceedings of the 9th IEEE International Conference on Industrial Informatics (INDIN)*, Caparica Lisbon, Portugal, July 2011, pp. 410–415.

[Ahl71] Rudolf Ahlswede, "Multi-way communication channels," in *Proceedings of the 2nd International Symposium on Information Theory*, Tsaghkadzor, Armenia, September 1971, pp. 23–52.

[Arn57] Vladimir I. Arnol'd, "On the representation of functions of two variables in the form $\chi[\varphi(x) + \psi(y)]$," *Uspekhi Matematicheskikh Nauk*, vol. 12, no. 2, pp. 119–121, 1957, English translation: *Vladimir I. Arnold Collected Works Volume 1 – Representations of Functions, Celestial Mechanics, and KAM Theory, 1957-1965*, Berlin·Heidelberg: Springer, 2009.

[AS64] Milton Abramowitz and Irene A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. New York: Dover Publications, 1964.

[ASSC02] Ian F. Akyildiz, Weilian Su, Yogesh Sankarasubramaniam, and Erdal Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–114, August 2002.

[BB11] Matthieu Bloch and João Barros, *Physical-Layer Security – From Information Theory to Security Engineering*. Cambridge·New York: Cambridge University Press, 2011.

[Ber71] Toby Berger, *Rate Distortion Theory – A Mathematical Basis for Data Compression*, ser. Prentice-Hall Series in Information and System Sciences. Englewood Cliffs: Prentice-Hall, 1971.

[BG09] Jürgen Braun and Michael Griebel, "On a constructive proof of Kolmogorov's superposition theorem," *Constructive Approximation*, vol. 30, no. 3, pp. 653–675, December 2009.

[BGPS06] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, June 2006.

[BHSN07] Waheed U. Bajwa, Jarvis D. Haupt, Akbar M. Sayeed, and Robert D. Nowak, "Joint source-channel communication for distributed estimation in sensor networks," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3629–3653, October 2007.

[Bie32] Ludwig Bieberbach, "Operationsbereiche von Funktionen," in *Proceedings of the International Congress of Mathematicians (ICM)*, vol. I, Zurich, Switzerland, September 1932, pp. 162–172.

[Bil95] Patrick Billingsley, *Probability and Measure*, 3rd ed., ser. Wiley Series in Probability and Statistics. New York: John Wiley & Sons, 1995.

[Bil99] Patrick Billingsley, *Convergence of Probability Measures*, 2nd ed., ser. Wiley Series in Probability and Statistics. Hoboken: John Wiley & Sons, 1999.

[BKR62] Leonard E. Baum, Melvin Katz, and Robert R. Read, "Exponential convergence rates for the law of large numbers," *Transactions of the American Mathematical Society*, vol. 102, no. 2, pp. 187–199, February 1962.

[BM10] Holger Boche and Ullrich J. Mönich, "Behavior of the quantization operator for bandlimited, nonoversampled signals," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2433–2440, May 2010.

[BM11] Holger Boche and Ullrich J. Mönich, "Sampling of deterministic signals and systems," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2101–2111, May 2011.

[BM12] Holger Boche and Ullrich J. Mönich, "Unboundedness of thresholding and quantization for bandlimited signals," *Signal Processing*, vol. 92, no. 12, pp. 2821–2829, December 2012.

[BM14] Holger Boche and Ullrich J. Mönich, "Signal and system approximation from general measurements," in *New Perspectives on Approximation and Sampling Theory - Festschrift in honor of Paul Butzer's 85th birthday*, ser. Applied and Numerical Harmonic Analysis. New York: Birkhäuser / Springer, 2014, in press.

[BSTS12] Mahesh K. Banavar, Anthony D. Smith, Cihan Tepedelenlioğlu, and Andreas Spanias, "On the effectiveness of multiple antennas in distributed detection over fading MACs," *IEEE Transactions on Wireless Communications*, vol. 11, no. 5, pp. 1744–1752, May 2012.

[BT89] Dimitri P. Bertsekas and John N. Tsitsiklis, *Parallel and Distributed Computation – Numerical Methods*, ser. Prentice-Hall Series in Information and System Sciences. Englewood Cliffs: Prentice-Hall, 1989.

[BTS10] Mahesh K. Banavar, Cihan Tepedelenlioğlu, and Andreas Spanias, "Estimation over fading channels with limited feedback using distributed sensing," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 414–425, January 2010.

[BTS12] Mahesh K. Banavar, Cihan Tepedelenlioğlu, and Andreas Spanias, "Distributed SNR estimation with power constrained signaling over Gaussian multiple-access channels," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 3289–3294, June 2012.

[Buc79] R. Creighton Buck, "Approximate complexity and functional representation," *Journal of Mathematical Analysis and Applications*, vol. 70, no. 1, pp. 280–298, July 1979.

[Buc82] R. Creighton Buck, "Nomographic functions are nowhere dense," *Proceedings of the American Mathematical Society*, vol. 85, no. 2, pp. 195–199, June 1982.

[CS10] John H. Conway and Neil J. A. Sloane, *Sphere Packings, Lattices and Groups*, 3rd ed., ser. Comprehensive Studies in Mathematics. New York: Springer, 2010, vol. 290.

[CT06] Thomas M. Cover and Joy A. Thomas, *Elements of Information Theory*, 2nd ed. New York: John Wiley & Sons, 2006.

[CWL14] Xuanyu Cao, Xinbing Wang, and Songwu Lu, "Function computation over heterogeneous wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, 2014, in press.

[DDMP14] Deepesh Data, Bikash K. Dey, Manoj Mishra, and Vinod M. Prabhakaran. (2014) How to securely compute the modulo-two sum of binary sources. [Online]. Available: http://arxiv.org/abs/1405.2555v1

[DKM+10] Alexandros G. Dimakis, Soummya Kar, José M. F. Moura, Michael G. Rabbat, and Anna Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, November 2010.

[Doe09] Ron Doerfler, "The lost art of nomography," *Journal of Undergraduate Mathematics and its Applications*, vol. 30, no. 4, pp. 457–493, 2009.

[DS98] Tolga M. Duman and Masoud Salehi, "Decentralized detection over multiple-access channels," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 34, no. 2, pp. 469–476, April 1998.

[DSME10] Vishal Doshi, Devavrat Shah, Muriel Médard, and Michelle Effros, "Functional compression through graph coloring," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3901–3917, August 2010.

[Dud01] Wiesław A. Dudek, "On some old and new problems in n-ary groups," *Quasigroups and Related Systems*, vol. 8, pp. 15–36, 2001.

[ELZ05] Uri Erez, Simon Litsyn, and Ram Zamir, "Lattices which are good for (almost) everything," *IEEE Transactions on Information Theory*, vol. 51, no. 10, pp. 3401–3416, October 2005.

[Eps58] L. Ivan Epstein, *Nomography.* New York·London: Interscience Publishers, 1958.

[Eve82] Harold A. Evesham, *The History and Development of Nomography.* Boston: Docent Press, 1982.

[EZ04] Uri Erez and Ram Zamir, "Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2293–2314, October 2004.

[FA14] Gerhard Fettweis and Siavash Alamouti, "5G: Personal mobile internet beyond what cellular did to telephony," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 140–145, February 2014.

[Fle77] Wendell Fleming, *Functions of Several Variables*, 2nd ed., ser. Undergraduate Texts in Mathematics. New York: Springer, 1977.

[For03] G. David Forney, Jr., "On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," in *Proceedings of the 41st Annual Allerton Conference on Communication, Control and Computing (Allerton)*, Monticello, IL, USA, October 2003, pp. 430–439.

[FR13] Simon Foucart and Holger Rauhut, *A Mathematical Introduction to Compressive Sensing*, ser. Applied and Numerical Harmonic Analysis. New York: Birkhäuser / Springer, 2013.

[Fra10] Jacob Fraden, *Handbook of Modern Sensors – Physics, Designs, and Applications*, 4th ed. New York: Springer, 2010.

[Gas08] Michael Gastpar, "Uncoded transmission is exactly optimal for a simple Gaussian "sensor" network," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5247–5251, November 2008.

[GK05] Arvind Giridhar and Panganamala R. Kumar, "Computing and communicating functions over sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 755–764, April 2005.

[GK06] Arvind Giridhar and Panganamala R. Kumar, "Toward a theory of in-network computation in wireless sensor networks," *IEEE Communications Magazine*, vol. 44, no. 4, pp. 98–107, April 2006.

[GK11] Abbas El Gamal and Young-Han Kim, *Network Information Theory*. Cambridge·New York: Cambridge University Press, 2011.

[GV03] Michael Gastpar and Martin Vetterli, "Source-channel communication in sensor networks," in *Proceedings of the 2nd International Workshop on Information Processing in Sensor Networks (IPSN)*, Palo Alto, CA, USA, April 2003, pp. 162–177.

[GVD06] Michael Gastpar, Martin Vetterli, and Luigi Dragotti, "Sensing reality and communicating bits: A dangerous liaison," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 70–83, July 2006.

[HCB02] Wendi B. Heinzelman, Anantha P. Chandrakasan, and Hari Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660–670, October 2002.

[Hil00] David Hilbert, "Mathematische Probleme," *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse*, no. 3, pp. 253–297, 1900, English translation: *Bulletin of the American Mathematical Society*, vol. 8, no. 10, pp. 437–479, 1902.

[Hil02] David Hilbert, "Mathematical problems," *Bulletin of the American Mathematical Society*, vol. 8, no. 10, pp. 437–479, 1902.

[HS12] Sheng Huang and Mikael Skoglund, "Polynomials and computing functions of correlated sources," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, USA, July 2012, pp. 771–775.

[HW48] Witold Hurewicz and Henry Wallman, *Dimension Theory*, ser. Princeton Mathematical Series. Princeton: Princeton University Press, 1948, vol. 4.

[Ins07] Texas Instruments, "2.4 GHz IEEE 802.15.4/ZigBee-ready RF transceiver," CC2420 datasheet, March 2007.

[JCS14] Feng Jiang, Jie Chen, and A. Lee Swindlehurst. (2014) Detection in analog sensor networks with a large scale antenna fusion center. To appear in *Proceedings of the 8th IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*. [Online]. Available: http://arxiv.org/abs/1404.7586

[JSS+12] Predrag Jakimovski, Hedda R. Schmidtke, Stephan Sigg, Leonardo Weiss Ferreira Chaves, and Michael Beigl, "Collective communication for dense sensing environments," *Journal of Ambient Intelligence and Smart Environments*, vol. 4, no. 2, pp. 123–134, March 2012.

[JWG13] Sang-Woon Jeon, Chien-Yi Wang, and Michael Gastpar, "Computation over Gaussian networks with orthogonal components," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, July 2013, pp. 2139–2143.

[KGK07] Sachin Katti, Shyamnath Gollakota, and Dina Katabi, "Embracing wireless interference: Analog network coding," in *Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication (ACM SIGCOMM)*, Kyoto, Japan, August 2007, pp. 397–408.

[Kha97] Semyon Y. Khavinson, *Best Approximation by Linear Superpositions (Approximate Nomography)*, ser. Translations of Mathematical Monographs. Providence: American Mathematical Society, 1997, vol. 159.

[KK09] Hemant Kowshik and Panganamala R. Kumar, "Zero-error function computation in sensor networks," in *Proceedings of the 48th Conference on Decision and Control (CDC)*, Shanghai, China, December 2009, pp. 3787–3792.

[KKF10] Lorenzo Keller, Nikhil Karamchandani, and Christina Fragouli, "Function computation over linear channels," in *Proceedings of the IEEE International Symposium on Network Coding (NetCod)*, Toronto, Canada, June 2010, pp. 1–6.

[KKFF13] Nikhil Karamchandani, Lorenzo Keller, Christina Fragouli, and Massimo Franceschetti, "Function computation via subspace coding," *Physical Communication*, vol. 6, pp. 134–141, March 2013.

[KKK08] Nilesh Khude, Anurag Kumar, and Aditya Karnik, "Time and energy complexity of distributed computation of a class of functions in wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 5, pp. 617–632, May 2008.

[KM79] János Körner and Katalin Marton, "How to encode the modulo-two sum of binary sources," *IEEE Transactions on Information Theory*, vol. 25, no. 2, pp. 219–221, March 1979.

[KM03] Ralf Koetter and Muriel Médard, "An algebraic approach to network coding," *IEEE/ACM Transactions on Networking*, vol. 11, no. 5, pp. 782–795, October 2003.

[KND13] Nikhil Karamchandani, Urs Niesen, and Suhas Diggavi, "Computation over mismatched channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 4, pp. 666–677, April 2013.

[Kol57] Aleksandrov N. Kolmogorov, "On the representation of continuous functions of several variables by superposition of continuous functions of one variable and addition," *Doklady Akademii Nauk SSSR*, no. 114, pp. 953–956, 1957, English translation: *American Mathematical Society Translations*, vol. 28, no. 2, pp. 55–59, 1963.

[KS12] Victor Korolev and Irina Shevtsova, "An improvement of the Berry-Esseen inequality with applications to Poisson and mixed Poisson random sums," *Scandinavian Actuarial Journal*, vol. 2012, no. 2, pp. 81–105, 2012.

[Kul08] Ulrich Kulisch, *Computer Arithmetic and Validity: Theory, Implementation, and Applications*, ser. de Gruyter Studies in Mathematics. Berlin·New York: De Gruyter, 2008, vol. 33.

[LA14] Eun J. Lee and Emmanuel Abbe. (2014) A Shannon approach to secure multi-party computations. [Online]. Available: http://arxiv.org/abs/1401.7360v3

[Lan73] Serge Lang, *Calculus of Several Variables*, ser. Addison-Wesley Series in Mathematics. Reading·Menlo Park·London: Addison-Wesley, 1973.

[Lap09] Amos Lapidoth, *A Foundation in Digital Communication.* Cambridge·New York: Cambridge University Press, 2009.

[LD07] Wenjun Li and Huaiyu Dai, "Distributed detection in wireless sensor networks using a multiple access channel," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 822–833, March 2007.

[LD13] Chengzhi Li and Huaiyu Dai, "Efficient in-network computing with noisy wireless channels," *IEEE Transactions on Mobile Computing*, vol. 12, no. 11, pp. 2167–2177, November 2013.

[LED12] Feng Li, Jamie S. Evans, and Subhrakanti Dey, "Design of distributed detection schemes for multiaccess channels," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1552–1569, April 2012.

[Lia72] Henry H. J. Liao, "Multiple access channels," Ph.D. dissertation, University of Hawaii, Department of Electrical Engineering, Honolulu, USA, 1972.

[Lük99] Hans D. Lüke, "The origins of the sampling theorem," *IEEE Communications Magazine*, vol. 37, no. 4, pp. 106–108, April 1999.

[Loe97] Hans-Andrea Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1767–1773, November 1997.

[Lor66] George G. Lorentz, *Approximation of Functions*. New York: Holt, Rinehart and Winston, 1966.

[LS07] Ke Liu and Akbar M. Sayeed, "Type-based decentralized detection in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1899–1910, May 2007.

[LYC03] Shuo-Yen R. Li, Raymond W. Yeung, and Ning Cai, "Linear network coding," *IEEE Transactions on Information Theory*, vol. 49, no. 2, pp. 371–381, February 2003.

[MGA12] Elaheh Mohammadi, Amin Gohari, and Hassan Aghaeinia, "Transmission of non-linear binary input functions over a CDMA system," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Cambridge, MA, USA, July 2012, pp. 1401–1405.

[MM93] James L. Massey and Thomas Mittelholzer, "Welch's bound and sequence sets for code-division multiple-access systems," in *Sequences II: Methods in Communication, Security and Computer Science*, R. Capocelli, A. De Santis, and U. Vaccaro, Eds. Heidelberg·New York: Springer, 1993, pp. 63–78.

[Mön11] Ullrich J. Mönich, "Reconstruction and processing of bandlimited signals based on their discrete values," Ph.D. dissertation, Technische Universität München, Fakultät für Elektrotechnik und Informationstechnik, Munich, Germany, 2011.

[MNT07] Gökhan Mergen, Vidyut Naware, and Lang Tong, "Asymptotic detection performance of type-based multiple access over multiaccess fading channels," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 1081–1092, March 2007.

[MS06] Damon Mosk-Aoyama and Devavrat Shah, "Computing separable functions via gossip," in *Proceedings of the 25th Annual ACM Symposium on Principles of Distributed Computing (PODC)*, Denver, USA, July 2006, pp. 113–122.

[MT06] Gökhan Mergen and Lang Tong, "Type based estimation over multiaccess channels," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 613–626, February 2006.

[Mui05] Robb J. Muirhead, *Aspects of Multivariate Statistical Theory*, ser. Wiley Series in Probability and Statistics. Hoboken: John Wiley & Sons, 2005.

[NA11] Matthew Nokleby and Behnaam Aazhang, "Unchaining from the channel: Cooperative computation over multiple-access channels," in *Proceedings of the IEEE Information Theory Workshop (ITW)*, Paraty, Brazil, October 2011, pp. 593–597.

[NA12] Matthew Nokleby and Behnaam Aazhang. (2012) Cooperative compute-and-forward. Submitted to *IEEE Transactions on Information Theory*. [Online]. Available: http://arxiv.org/abs/1203.0695v1

[NCL10] Wooseok Nam, Sae-Young Chung, and Yong H. Lee, "Capacity of the Gaussian two-way relay channel to within $\frac{1}{2}$ bit," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5488–5494, November 2010.

[NDG11] Bobak Nazer, Alexandros G. Dimakis, and Michael Gastpar, "Local interference can accelerate gossip algorithms," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp. 876–887, August 2011.

[NG07] Bobak Nazer and Michael Gastpar, "Computation over multiple-access channels," *IEEE Transactions on Information Theory*, vol. 53, no. 10, pp. 3498–3516, October 2007.

[NG08a] Bobak Nazer and Michael Gastpar, "The case for structured random codes in network capacity theorems," *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 455–474, June 2008.

[NG11a] Bobak Nazer and Michael Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6463–6486, October 2011.

[NG11b] Bobak Nazer and Michael Gastpar, "Reliable physical layer network coding," *Proceedings of the IEEE*, vol. 99, no. 3, pp. 438–460, March 2011.

[OR01] Alon Orlitsky and James R. Roche, "Coding for computing," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 903–917, March 2001.

[OSFM07] Reza Olfati-Saber, J. Alex Fax, and Richard M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, January 2007.

[Poz05] David M. Pozar, *Microwave Engineering*, 3rd ed. Hoboken: John Wiley & Sons, 2005.

[Pre14] Kathy Pretz, "Tech topic: Smart sensors – making the internet of things soar," *The Institute, IEEE*, pp. 6–7, March 2014.

[PS08] John G. Proakis and Masoud Salehi, *Digital Communications*, 5th ed. New York: McGraw-Hill, 2008.

[PY06] Petar Popovski and Hiroyuki Yomo, "The anti-packets can increase the achievable throughput of a wireless multi-hop network," in *Proceedings of the IEEE International Conference on Communications (ICC)*, Istanbul, Turkey, June 2006, pp. 3885–3890.

[Rap02] Theodore S. Rappaport, *Wireless Communications – Principles and Practice*, 2nd ed., ser. Prentice Hall Communications Engineering and Emerging Technologies. Upper Saddle River: Prentice-Hall, 2002.

[Rei96] Constance Reid, *Hilbert*. New York: Copernicus / Springer, 1996.

[Ros09] Leif Roschier. (2009) PyNomo – a program to create nomographs with python. Freeware. [Online]. Available: http://www.pynomo.org

[Rud76] Walter Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, 1976.

[Rév68] Pál Révész, *The Laws of Large Numbers*, ser. Probability and Mathematical Statistics. New York·London: Academic Press, 1968, vol. 4.

[SB14] Rafael F. Schaefer and Holger Boche, "Physical layer service integration in wireless networks – signal processing challenges," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 147–156, May 2014.

[SBK05] Bharath Sundararaman, Ugo Buy, and Ajay D. Kshemkalyani, "Clock synchronization for wireless sensor networks: A survey," *Ad Hoc Networks*, vol. 3, no. 3, pp. 281–323, May 2005.

[Sch83] Peter Schatte, "On sums modulo 2π of independent random variables," *Mathematische Nachrichten*, vol. 110, no. 1, pp. 243–262, 1983.

[SCW+11] Yi-Sheng Shiu, Shih Y. Chang, Hsiao-Chun Wu, Scott C.-H. Huang, and Hsiao-Hwa Chen, "Physical layer security in wireless networks: A tutorial," *IEEE Wireless Communications*, vol. 18, no. 2, pp. 66–74, April 2011.

[SGS07] Sundar Subramanian, Piyush Gupta, and Sanjay Shakkottai, "Scaling bounds for function computation over large networks," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, Nice, France, June 2007, pp. 136–140.

[Sha48] Claude E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3 and 4, pp. 379–423 and 623–656, July and October 1948.

[Shi96] Albert N. Shiryaev, *Probability*, 2nd ed., ser. Graduate Texts in Mathematics. New York: Springer, 1996, vol. 95.

[SJB12] Stephan Sigg, Predrag Jakimovski, and Michael Beigl, "Calculation of functions on the RF-channel for IoT," in *Proceedings of the 3rd International Conference on the Internet of Things*, Wuxi, China, October 2012, pp. 107–113.

[Spr65] David A. Sprecher, "A representation theorem for continuous functions of several variables," *Proceedings of the American Mathematical Society*, vol. 16, no. 2, pp. 200–203, April 1965.

[Stü11] Gordon L. Stüber, *Principles of Mobile Communication*, 3rd ed. New York·Heidelberg: Springer, 2011.

[Ste85] Yaki Sternfeld, "Dimension, superposition of functions and separation of points, in compact metric spaces," *Israel Journal of Mathematics*, vol. 50, no. 1-2, pp. 13–53, March 1985.

[STM09] STMicroelectronics, "Ultra-low current 2.4 V precision analog temperature sensor," STLM20 datasheet, September 2009.

[SV12] Rajiv Soundararajan and Sriram Vishwanath, "Communicating linear functions of correlated Gaussian sources over a MAC," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1853–1860, March 2012.

[SW73] David Slepian and Jack K. Wolf, "A coding theorem for multiple access channels with correlated sources," *The Bell System Technical Journal*, vol. 52, no. 7, pp. 1037–1076, September 1973.

[SWB05] Sławomir Stańczak, Marcin Wiczanowski, and Holger Boche, "Distributed power control for optimizing a weighted sum of link-layer QoS levels," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)*, St. Louis, MO, USA, November 2005, pp. 3539–3543.

[SWB06] Sławomir Stańczak, Gerhard Wunder, and Holger Boche, "On pilot-based multipath channel estimation for uplink CDMA systems: An overloaded case," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 512–519, February 2006.

[SWB07] Sławomir Stańczak, Marcin Wiczanowski, and Holger Boche, "Distributed utility-based power control: Objectives and algorithms," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 5058–5068, October 2007.

[SZHT07] Ananthram Swami, Qing Zhao, Yao-Win Hong, and Lang Tong, Eds., *Wireless Sensor Networks – Signal Processing and Communications Perspectives*. Hoboken: John Wiley & Sons, 2007.

[TD11] Cihan Tepedelenlioğlu and Sivaraman Dasarathan, "Distributed detection over Gaussian multiple access channels with constant modulus signaling," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2875–2886, June 2011.

[TH98] David N. C. Tse and Steven V. Hanly, "Multiaccess fading channels – Part I: Polymatroid structure, optimal resource allocation and throughput capacities," *IEEE Transactions on Information Theory*, vol. 44, no. 7, pp. 2796–2815, November 1998.

[TNG11] Himanshu Tyagi, Prakash Narayan, and Piyush Gupta, "When is a function securely computable?" *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6337–6350, October 2011.

[VAT99] Pramod Viswanath, Venkat Anantharam, and David N. C. Tse, "Optimal sequences, power control, and user capacity of synchronous CDMA systems with linear MMSE multiuser receivers," *IEEE Transactions on Information Theory*, vol. 45, no. 6, pp. 1968–1983, September 1999.

[vdV98] Aad W. van der Vaart, *Asymptotic Statistics*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. New York: Cambridge University Press, 1998.

[Ver98] Sergio Verdú, *Multiuser Detection*. New York: Cambridge University Press, 1998.

[Vit04] Anatoli G. Vitushkin, "On Hilbert's thirteenth problem and related questions," *Russian Mathematical Surveys*, vol. 59, no. 1, pp. 11–25, 2004.

[VK67] Anatoli G. Vitushkin and Gennadi M. Khenkin, "Linear superpositions of functions," *Russian Mathematical Surveys*, vol. 22, no. 1, pp. 77–125, 1967.

[vN28] John von Neumann, "Ein System algebraisch unabhängiger Zahlen," *Mathematische Annalen*, vol. 99, no. 1, pp. 134–141, 1928.

[War59] Mieczysław Warmus, *Nomographic Functions*, ser. Rozprawy Matematyczne. Warsaw: Państwowe Wydawnictwo Naukowe, 1959, vol. 16.

[Wil08] Andreas Willig, "Recent and emerging topics in wireless industrial communications: A selection," *IEEE Transactions on Industrial Informatics*, vol. 4, no. 2, pp. 102–124, May 2008.

[WJG13] Chien-Yi Wang, Sang-Woon Jeon, and Michael Gastpar. (2013) Interactive computation of type-threshold functions in collocated broadcast-superposition networks. *submitted to IEEE Transactions on Information Theory*. [Online]. Available: http://http://arxiv.org/abs/1310.2860

[WNPS10] Makesh P. Wilson, Krishna Narayanan, Henry D. Pfister, and Alex Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Transactions on Information Theory*, vol. 56, no. 11, pp. 5641–5654, November 2010.

[Wol66] John W. Woll, Jr., *Functions of Several Variables*, ser. Harbrace College Mathematics. New York·Chicago·Burlingame: Harcourt, Brace & World, 1966.

[XCLG08] Jin-Jun Xiao, Shuguang Cui, Zhi-Quan Luo, and Andrea J. Goldsmith, "Linear coherent decentralized estimation," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 757–770, February 2008.

[Yam82] Hirosuke Yamamoto, "Wyner-Ziv theory for a general function of the correlated sources," *IEEE Transactions on Information Theory*, vol. 28, no. 5, pp. 803–807, September 1982.

[YF04] Ossama Younis and Sonia Fahmy, "HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366–379, October–December 2004.

[YKR06] Ossama Younis, Marwan Krunz, and Srinivasan Ramasubramanian, "Node clustering in wireless sensor networks: Recent developments and deployment challenges," *IEEE Network*, vol. 20, no. 3, pp. 20–25, May–June 2006.

[YSD07] Lei Ying, Rayadurgam Srikant, and Geir E. Dullerud, "Distributed symmetric function computation in noisy wireless sensor networks," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4826–4833, December 2007.

[Zam11] Ron Zamir, "Can structure beat random? – The story of lattice codes," *IEEE Information Theory Society Newsletter*, vol. 61, no. 1, pp. 20–29, March 2011.

[ZG04] Feng Zhao and Leonidas Guibas, *Wireless Sensor Networks: An Information Processing Approach*. San Francisco: Elsevier / Morgan Kaufmann, 2004.

[ZLL06] Shengli Zhang, Soung-Chang Liew, and Patrick P. Lam, "Hot topic: Physical-layer network coding," in *Proceedings of the 12th Annual Interna-*

tional Conference on Mobile Computing and Networking (ACM MobiCom), Los Angeles, CA, USA, September 2006, pp. 358–365.

[ZPGS13] Jiening Zhan, Se Yong Park, Michael Gastpar, and Anant Sahai, "Linear function computation in networks: Duality and constant gap results," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 4, pp. 620–638, April 2013.