Communication Resource Allocation in Wireless Networked Control Systems

Vom Fachbereich 18
Elektrotechnik und Informationstechnik
der Technischen Universität Darmstadt
zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Dissertation

von
M.Sc. Kilian Kiekenap

Referentin:	Prof. Dr.-Ing. Anja Klein
Korreferent:	Prof. Dr.-Ing. habil. Tobias Weber
Tag der Einreichung:	10. Mai 2022
Tag der mündlichen Prüfung:	12. Oktober 2022

D 17
Darmstädter Dissertation

The work of Kilian Kiekenap was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Projektnummer 210487104 - SFB 1053 (`https://www.maki.tu-darmstadt.de`) and the LOEWE Center emergenCITY (`https://www.emergencity.de/`).

Kiekenap, Kilian: Communication Resource Allocation in Wireless Networked Control Systems
Darmstadt, Technische Universität Darmstadt,
Jahr der Veröffentlichung der Dissertation auf TUprints: 2023
URN: urn:nbn:de:tuda-tuprints-229324
Tag der mündlichen Prüfung: 12.10.2022

Veröffentlicht unter CC BY-NC-SA 4.0 International
https://creativecommons.org/licenses/

Kurzfassung

Heutzutage konzentriert sich die mobile Kommunikation hauptsächlich auf die menschliche Kommunikation, wie Textnachrichten, Video- und Sprachanrufe und die Übertragung großer Datenmengen für z. B. Audio- und Video-Streaming-Anwendungen. Audio- und Videokommunikation erfordern mäßige Latenzzeiten und niedrige oder mittlere Datenraten, während die Übertragung großer Dateien im Allgemeinen hohe Datenraten erfordert, aber auch mit hohen Latenzzeiten zurechtkommen könnte. Neben Modi mit noch höheren Datenraten und mehr Geräten pro einzelner Zelle im Vergleich zu früheren Generationen wird der aktuelle 5G-Mobilfunkstandard mehr Anwendungen aus der Regelungstechnik ermöglichen, da niedrige Latenzen und garantierte maximale Fehlerraten für sogenannte Wireless Networked Control Systems (WNCS) erforderlich sind. Der neue 5G-Standard berücksichtigt WNCS in seinem Ultra-Reliable Low Latency Communication (URLLC)-Szenario, das eine niedrigratige Kommunikation mit minimaler Latenz und verbesserter Fehlerkorrektur für diese spezielle Art der Kommunikation mit viel kleineren Datenmengen im Vergleich zu anderen Szenarien bietet. Die genauen Anforderungen an Latenzzeit, maximale Fehlerwahrscheinlichkeit und Datenrate werden durch die Dynamik der jeweiligen Anlagen bestimmt. Ein geschlossenes Regelsystem besteht aus drei Hauptkomponenten: Regler, Regelstrecke und Sensor. Der Regler sendet auf der Grundlage des geschätzten Streckenzustands Steuerbefehle an das Stellglied in der Regelstrecke. Das Stellglied setzt die Befehle um und ändert so den Zustand der Regelstrecke. Der Sensor sendet Messungen des Streckenzustands zurück an den Regler, um die Rückkopplungsschleife zu schließen. Die drahtlose Übertragung kann dann entweder zur Übermittlung von Sensorwerten an den Regler, zur Übermittlung von Steuerbefehlen vom Regler an das Stellglied in der Strecke oder sogar für beides verwendet werden. Wenn mehrere Teilsysteme auf denselben drahtlosen Kommunikationsressourcen betrieben werden, muss ein Mehrfachzugriffsschema zur Vermeidung von Störungen implementiert werden, das auch den Zustand und die Anforderungen des Regelungssystems berücksichtigt. Dies kann entweder zentral erfolgen, wobei eine zentrale Instanz die Ressourcen den einzelnen Teilsystemen zuweist, oder dezentral, wobei die einzelnen Teilsysteme kooperativ handeln.

Ein wesentliches Element der Regelungstheorie ist die möglichst genaue Kenntnis des Systemzustands durch den Sensor und die anschließende Übertragung an den Regler, damit dieser die optimale Stellgröße zur Minimierung einer von der jeweiligen Anwendung abhängigen Kostenfunktion erzeugen kann. Für die Übertragung über einen digitalen Kommunikationskanal wird der Anlagenzustand, der als Vektor kontinuierlicher

Werte modelliert werden kann, zunächst erfasst und muss anschließend in digitale Symbole übersetzt werden. Bei diesem Prozess gibt es zwei Hauptfehlerquellen. Erstens werden die Messungen selbst durch das Messrauschen beeinträchtigt. Der aus diesem Rauschen resultierende Fehler kann verringert werden, wenn mehrere unabhängige Messungen desselben Wertes vorgenommen und kombiniert werden. Zweitens ist die Anzahl der verfügbaren Symbole, die in einem festen Zeitrahmen mit einer festen Datenrate übertragen werden können, begrenzt, so dass die Werte quantisiert werden müssen. In dieser Arbeit wird gezeigt, dass es ein Optimum im Kompromiss zwischen der Anzahl der verrauschten Messungen und der Anzahl der verfügbaren Sendesymbole mit einer jeweils gleichen Anzahl von Bits gibt, wenn die insgesamt verfügbare Zeit oder Energie begrenzt ist und durch den Mess- und den Sendeprozess nacheinander genutzt wird. Da Gebiete des kontinuierlichen Zustandsraums vor der Übertragung auf ein einziges Symbol abgebildet werden, gehen durch die Quantisierung Informationen über den Systemzustand verloren. Generell gilt: Je mehr unterscheidbare Symbole in der Kommunikationsverbindung vom Sensor zum Controller zur Verfügung stehen, desto kleiner ist der resultierende Quantisierungsfehler. Das zur Quantifizierung des Fehlers verwendete Maß ist das Bayes-Risiko. Um die Informationen über den von jedem Symbol übertragenen Systemzustand zu erhöhen, wird ein Schema nicht-äquidistanter Quantisierungsintervallgrenzen abgeleitet, das den Quantisierungsfehler für eine gegebene Anzahl von übertragenen Datensymbolen und eine bekannte Verteilung möglicher Sensorwerte minimiert. Da das Optimierungsproblem für dieses Schema rechenintensiv ist, wird ein zweites Schema implementiert, das lediglich die Wahrscheinlichkeit aller möglichen Übertragungsdatensymbole ausgleicht. Schließlich wird eine äquidistante Abtastung des Sensorwerteraums für eine gegebene Anzahl von Sendedatensymbolen mit den beiden vorhergehenden Schemata verglichen, um eine Basislinie zu erhalten. Durch Anwendung der drei Verfahren auf drei verschiedene Verteilungen eines skalaren Sensorwerts kann gezeigt werden, dass das informationsoptimale Verfahren im Vergleich zum linearen Verfahren bei gleichem Bayes-Risiko bis zu 20 % der erforderlichen Bits einsparen kann. In den meisten Fällen erreicht das äquidistante Schema etwa die Hälfte der Reduktion des Bayes-Risikos, die das informationsbasierte Schema im Vergleich zum äquidistanten Schema erreicht.

Wenn mehrere Teilsysteme mit einzelnen Sensoren und Anlagen um drahtlose Kommunikationsressourcen in WNCS konkurrieren, reichen die Ressourcen in manchen Situationen nicht aus, um immer Sensormesswerte von allen Sensoren zu einem zentralen Regler zu übertragen. In diesem Fall muss eine Teilmenge von Sensoren ausgewählt und die verfügbaren Ressourcen müssen auf diese verteilt werden, um Interferenzen zwischen den Übertragungen zu vermeiden. Bei einem zentral geplanten, zeitdiskreten System kann eine zentrale Stelle die Sensoren auswählen, von denen in den jeweiligen Zeitschlitzen Messwerte angefordert werden sollen. Für die jeweils nicht ausgewählten Teilsysteme muss eine Vorhersage des aktuellen Zustands gemacht werden. Die Regelungsleistung hängt von der Minimierung der Unsicherheit über den aktuellen Systemzustand ab, der als weißes Gaußsches Systemrauschen modelliert wird, wobei die Varianz des Rauschens ein Maß für die Unsicherheit ist. Zu diesem Zweck wird die Optimalität eines regelmäßigen Aktualisierungsschemas für lineare Teilsysteme mit additivem weißem Gaußschen Rauschen gezeigt. Danach wird der optimale Anteil der Kommunikationsressourcen für jedes Teilsystem bei einer gegebenen Gesamtzahl von Kommunikationsressourcen abgeleitet. Die berechneten Ressourcenanteile für die Subsysteme aus dieser Optimierung werden dann einem Algorithmus zur Planung der tatsächlichen Übertragungen zugeführt. Dieser zweistufige Ansatz ermöglicht eine Offline-Berechnung der Ressourcenanteile, während zur Laufzeit nur die tatsächliche Planung auf der Grundlage der vorberechneten Anteile erfolgen muss. Die durchschnittliche Unsicherheit über die Subsystemzustände wird im Vergleich zu bestehenden Planungsalgorithmen um bis zu 20 % reduziert. Darüber hinaus verringert sich die Schwankung der Unsicherheit über die Teilsystemzustände im Zeitverlauf um bis zu 60 %.

Schließlich wird die Verringerung des Energieverbrauchs der drahtlosen Übertragung von Steuerbefehlen an die Stellglieder in den Regelstrecken untersucht. In Regelungsanwendungen müssen die berechneten Regelgrößen, die vom Regler an die Stellglieder gesendet werden, innerhalb einer systemabhängigen Frist korrekt geliefert werden. Diese drei Anforderungen stehen in Konkurrenz zueinander, so dass ein Kompromiss gefunden werden muss. Da die Datenpakete, die die Befehle enthalten, klein sind, wird statt der bekannten Shannon-Kapazitätsformel für unendliche Paketlängen eine angepasste Formel für kurze Pakete angewendet, um die erforderliche Energie zu bestimmen. Die angepasste Formel kann dann verwendet werden, um die optimale Anzahl von Zeit-Frequenz-Ressourcen für einen minimalen Gesamtenergieverbrauch zu finden, die für eine einzelne Übertragung zugewiesen werden müssen. Für das resultierende Optimierungsproblem, das die individuellen Fristen, Befehlspaketgrößen und Kanaleigenschaften jedes Agenten berücksichtigt, wird die Konvexität gezeigt. Die berechnete optimale Verteilung der begrenzten Zeit- und Bandbreitenressourcen auf die einzelnen Subsysteme bei minimalem Energieverbrauch wird dann unter Zuhilfenahme eines Orthogonal Frequency Division Multiplex (OFDM)-Schemas angewendet. Da OFDM keine kontinuierliche, sondern nur eine auf Ressourcenblöcken basierende Aufteilung der Ressourcen erlaubt, wird ein Algorithmus für die Zuteilung von Zeit- und Frequenzblöcken aus dem OFDM-Schema entwickelt und gezeigt, dass er nahe an die theoretischen Grenzen der kontinuierlichen Lösung herankommt. Im Vergleich zu einem Schema, das nur die Gesamtzahl der jedem Agenten zugewiesenen Zeit-Frequenz-

Ressourcen ausgleicht, wird bei Anwendung des vorgeschlagenen Schemas die erforderliche Gesamtenergie zur Einhaltung der Fehlerraten- und Übertragungszeitgrenzen um bis zu 50 % reduziert.

Abstract

Today, mobile communication is mainly focused on human communication, like text messaging, video and voice calls, and the transmission of large data volumes for e.g. audio and video streaming applications. Audio and video communication require moderate latencies and low or medium data rates, while the transmissions of large files generally require high data rates, but could also cope with high latencies. In addition to modes with even higher data rates and more devices per single cell compared to previous generations, the current 5G mobile radio standard will allow for more applications from the control domain, since low latencies and guaranteed maximum error rates are required for so-called Wireless Networked Control Systems (WNCS)s. The new 5G standard considers WNCS in its Ultra-Reliable Low Latency Communication (URLLC) scenario, which provides a low-rate communication with minimal latency and improved error correction for this special type of communication with much smaller amounts of data compared to other scenarios. The exact requirements on latency, maximum error probability and data rate are determined by the dynamics of the respective plants. A closed-loop control system consists of three main components, controller, plant and sensor. The controller sends control commands to the plant, based on the estimated plant state. The plant then applies the commands, thus changing its state. The sensor transmits measurements of the plant state back to the controller to close the feedback loop. Wireless transmission can then be used to transmit either sensor values to the controller, transmit control commands from the controller to the actuator at the plant, or even for both. If multiple subsystems are operated on the same wireless communication resources, a multiple access scheme to prevent interference, which considers also the state and demands of the control system, has to be implemented. This can be done either in a centralized fashion, where a central entity allocates the resources to the individual subsystems, or in a decentralized fashion, where the individual subsystems act cooperatively.

A crucial element in control theory is the acquisition of as accurate as possible knowledge of the system state by the sensor and the subsequent transmission to the controller to enable it to generate the optimum control input to minimize a cost function, which depends on the respective application. For the transmission over a digital communication channel, the plant state, which can be modeled as a vector of continuous values, is first sensed and has to be translated to digital symbols afterwards. There are two main sources of error in this process. First, the measurements themselves are impaired by the measurement noise. The error resulting from this noise can be reduced, if multiple independent measurements of the same value are taken and combined. Second, the number of available symbols, which can be transmitted in a fixed time frame with

a fixed data rate, is limited, so the values have to be quantized. In this thesis, the existence of an optimum in the tradeoff of the number of noisy measurements and the number of available transmit data symbols with an equal number of bits is shown, if the available time or energy is limited and shared by the measurement and the transmission process. Since domains of the continuous state value space are mapped to a single symbol before transmission, information about the system state is lost due to the quantization. Generally, the more distinct symbols are available in the communication link from the sensor to the controller, the smaller is the resulting quantization error. The measure applied to quantify the error is the Bayes risk. To increase the information about the system state carried by each symbol, a scheme of non-equidistant quantization interval bounds, minimizing the quantization error for a given number of transmit data symbols, and a known distribution of possible sensor values is derived. Since the optimization problem for this scheme is computationally demanding, a second scheme purely equalizing the probability of all possible transmit data symbols is implemented. Finally, as a baseline, an equidistant sampling of the sensor value space for a given number of transmit data symbols is compared to the previous two schemes. By applying the three schemes to three different distributions of a scalar sensor value, it can be shown that the information-optimal scheme can save up to 20 % of the required bits for the same Bayes risk, when compared to the linear scheme. In most of the cases, the equidistant scheme achieves about half of the reduction of Bayes risk the information based scheme achieves, when compared to the equidistant scheme.

If multiple subsystems with individual sensors and plants are competing for wireless communication resources in WNCS, in some situations the resources are insufficient to always transmit sensor readings from all sensors to the central controller. In this case, a subset of sensors has to be selected and the available resources have to be distributed to them to prevent interference between the transmissions. With a centrally scheduled, discrete time system, a central entity can select the sensors to request readings from in each time slot. For the non-selected subsystems, a prediction of the current state has to be made. The control performance depends on minimizing the uncertainty about the current system state, which is modeled as a white Gaussian system noise, where the variance of the noise is a measure for the uncertainty. For this purpose, the optimality of a regular update scheme for linear subsystems with additive white Gaussian noise is shown. After that, the optimum communication resource share for each subsystem for a given number of communication resources is derived. The calculated resource shares for the subsystems from this optimization are then fed to an algorithm to schedule the actual transmissions. This two-step approach allows for an offline calculation of the resource shares, while during runtime only the actual scheduling based on the precalculated shares has to be done. The average uncertainty about the subsystem states is reduced by up to 20 % compared to existing scheduling algorithms. Furthermore, the variation over time of the uncertainty about the subsystem states is reduced by up to 60 %.

Finally, the reduction of the energy consumption of the wireless transmission of control commands to the actuators at the plants is investigated. In control applications, the calculated control inputs sent from the controller to the actuators must be delivered correctly before a system-dependent deadline. These three requirements are competing, so a tradeoff has to be found. Since the data packets containing the commands are small, instead of the well-known Shannon capacity formula for infinite packet length, an adapted formula for short packets is applied to determine the required energy. The adapted formula can then be used to find the optimum number of time-frequency resources for minimal total energy consumption to be allocated for a single transmission. The resulting optimization problem considering the individual deadlines, command packet sizes and channel characteristics of each agent is shown to be convex. The derived optimal distribution of the limited time and bandwidth resources to the individual subsystems for minimal energy consumption is then applied using an Orthogonal Frequency Division Multiplex (OFDM) scheme. Since OFDM does not allow for a continuous, but only for a resource-block based splitting of resources, an algorithm for the allocation of time-frequency blocks from the OFDM scheme is developed and shown to perform close to the theoretical bounds from the continuous solution. Compared to a scheme only balancing the time-frequency resources allocated to each agent, the total required energy to fulfill the error rate and transmission time limits is reduced by up to 50 % when applying the proposed scheme.

Contents

Previously Published Material

This thesis contains material that has been previously published in scientific conferences. Table 1 summarizes the papers relevant for the content of this thesis. This is done in order to make correct use of the gathered data and previous results as well as the reused text passages. A comprehensive list of all scientific publications of the author of this thesis is available in Chapter 5.2 at the end of the thesis.

The pronoun "I" will be used exclusively in this chapter to describe the specific contributions of the author of this thesis. For the remainder of the thesis, the pronoun "we" will be used to refer to the contribution of all co-authors of the respective publication.

In Chapter 2, "Sensor Value Quantization and Transmission", the influence of the number of measurements and quantization accuracy of a quantizing sensor on the error between the true value, which is measured by a sensor and then transmitted over a wireless link, and the value reconstructed from the received quantized measurements was investigated. The model and part of the mathematical derivations and numerical results were already presented in [KASK19]. In [KASK19], I derived the stochastic properties of each step for an arbitrary distribution of the value of interest in the processing chain. Based on these results, I suggested a quantization scheme to maximize the mutual information between the measured and the estimated value of interest. In this thesis, the quantization scheme is improved further to minimize the actual estimation error. This new scheme is then compared to the scheme from [KASK19] and a non-optimized scheme.

In Chapter 3, "Scheduling of Sensor Readings with Constrained Communication", I optimized the scheduling of sensor value transmissions for minimum state estimation error at a central controller to achieve optimal control performance in linear control

Table 1: List of publications related to this thesis

Chapter	Publication
Chapter 2: Sensor Value Quantization and Transmission	[KASK19]
Chapter 3: Scheduling of Sensor Readings with Constrained Communication	[KK20]
Chapter 4: Deadline-Aware Control Command Transmission	[KOK21]

subsystems. The model, optimization problem and scheduling algorithm as well as part of the results were already shown in [KK20], which is in turn based on the model from [AVK+19]. This thesis extends the results for a single set of subsystem parameters from [KK20] by investigating the performance of the scheduling algorithm for different system parameters for the linear subsystems. Therefore, in addition to [KK20], different ratios of higher and lower dynamic systems as well as completely random parameter sets are investigated in this thesis.

Chapter 4, "Deadline-Aware Control Command Transmission", is about the allocation of time-frequency resources for the timely reception of control commands sent from a central controller to multiple agents. The agents have individual deadlines, which determine the maximum allowable time after command generation to receive the command correctly. For the short-packet nature of control commands the approximation from [PPV10] is used. The convexity of the required energy was shown in [SSY+19], which allowed me to rewrite the problem as a convex optimization problem. While the model, problem and the algorithm are already shown in [KOK21], in this thesis I extend the numerical results to include also different total available bandwidth configurations, which result in a higher number of available resources to illustrate the effects of a less constrained environment compared to the scenario shown in the paper.

Chapter 1

Introduction

1.1 Communication and Control

The current roll-out of 5G mobile networks will greatly influence the types of services used over mobile networks [KK20]. Nowadays, the Internet of Things (IoT) receives growing attention from many different research fields, e.g. industrial communication [WSJ17] or connected cars [PFL$^+$16], [KASK19]. While the growth of the number of personal devices is already slowing down, the fraction of traffic of autonomous devices communicating with each other is rapidly growing [Cis19], [KK20]. Currently, most IoT devices serve as home appliances, building control or environmental sensors [Cis20], [KOK21]. While previous generations of mobile networks focused on human users and high data rates, 5G networks will also provide communication channels particularly suited for industrial communication and control applications [WSJ17], as well as connected cars and autonomous driving [PFL$^+$16], [KK20]. The enormous amount of additional smart devices deployed will greatly increase the number of devices per area [GRC$^+$14], [KASK19]. Despite the ever-growing number of devices per cell and increased data rates, also latency and reliability guarantees can be established [SYQ17]. This allows for new services, especially from the control domain, which heavily depend on latency guarantees [TC03], [KK20]. Many devices will act as autonomous agents and not only do sensing, but also cooperate to fulfill tasks [YSL13]. Most of them will use a wireless connection for communication, which results in increasing competition for the available communication resources like frequency bands. The sensors will be used for sensing many different types of values like temperature, humidity, air pressure, filling levels of tanks, positions, or velocities [KASK19]. One important application from the control domain is autonomous driving, where information from sensors in the vehicles and along the road has to be transmitted between vehicles and central entities to manage traffic. Obviously, the amount of data to be exchanged varies heavily, depending on the number of vehicles per area, the weather conditions etc [KK20].

For Industry 4.0, a shift to wireless instead of the current wired connections is desired. Wireless systems are rapidly reconfigurable and can be easily adapted to current production requirements [KK20]. Recent developments in industrial automation introduce wireless communication to production facilities for e.g. real-time monitoring or process control [KRZ+20]. In IoT for industrial manufacturing, called Industrial Internet of Things (IIoT), devices are part of industrial production processes and, therefore, directly embedded into control loops, which impose different demands on wireless communication systems compared to previous applications like voice, web browsing and video streaming [KOK21]. The real-time requirements of industrial control systems imply challenges on the communication system design very different from previous generations of communication systems. The control counterpart also has to be adapted to the specific characteristics of wireless connections. Current control systems rely on high data rates, low latency, and low error probabilities [KK20]. In control, the data amounts are small, in the order of tenths or a few hundred bytes, while the constraints on latency and packet error rate are even tighter than in other fields. Packet error rates as low as 10^{-9} and latencies of less than 0.25 to 5 ms are required [SMK+17]. The new 5G mobile radio standard is the first to define requirement profiles for these use cases [3GP19], [KOK21].

The combination of control and wireless communication systems, called Wireless Networked Control Systems (WNCS), is an important area of current research, especially the joint optimization of both, to adapt either part to possibly varying conditions of each other. If, for example, an autonomous vehicle drives along a straight and empty road, the communication can be reduced. In a crowded city scenario with a lot of intersections, the communication effort is much higher. At the same time, the speed of the vehicle can be adapted to the available communication resources [KK20].

WNCS are composed of subsystems which communicate over a shared wireless communication channel to exchange sensor values and control commands. The subsystems either work cooperatively to achieve a common task, or they compete for communication and other resources [KK20]. In the aforementioned autonomous driving example, the subsystems are, for example, a lane keeping subsystem, a subsystem for keeping the distance to the preceding vehicle and so on. In the connected cars application, multiple vehicles exchange data to form platoons or reduce distances in intersections. In all cases, the sensors need to transmit data like the position of the road marking or the distance of the preceding vehicle to the corresponding controller. Inside a vehicle this can either be done over a bus system, like the widely used CAN-Bus [ISO15], or also a wireless network of all components. Between different vehicles or elements of the infrastructure, as discussed in [3GP15], a wireless transmission is required.

In the industrial domain, there are many different types of plants [BHCW18], [KRZ+20]. Like in the autonomous driving example, the requirements on the control loop depend on the type of tasks to be fulfilled. While there are tasks like heating and stirring large volumes of liquid, which is an inherently slow process [Lun16a], there are other tasks, which might involve moving machines, which have more demanding requirements on the reaction time and precision of the underlying control system. In either case, the available transmission resources are limited and shared between the subsystems. In a digital control system, the data rate transmitted by a sensor depends on two factors, the resolution of the sensed value and the update rate. Both of these parameters influence the ability of the controller to know the current subsystem state, which is crucial for an optimal control action. For the second link from the controller to the actor, the relation is similar, the finer the resolution of the control input and the more often a new control input is sent, the higher is the required data rate. Also for this link, the finer the resolution and the higher the control input update rate, the closer is the applied control input to the calculated optimum, which improves the control performance. The control performance describes the difference between the true subsystem state and the desired state. For the lane keeping example above the control performance could be the distance of the center of the vehicle from the center of the lane, for the distance keeping example the difference between the desired and the true distance.

WNCS can generally be divided into two types, centralized and decentralized, as shown in [GYH17]. In both types, the subsystems of decentralized WNCS compete for communication resources in a shared communication medium such as a frequency band. For the decentralized type, a multiple-access scheme has to be implemented, which allows the individual subsystems to prevent collisions on the medium without central coordination. The centralized type has an additional central scheduler, which coordinates the multiple access and centrally allocates the available resources to the subsystems. This central allocation reduces the complexity of the scheduling process and leads, in general, to a better exploitation of the available resources, e.g. transmission time, bandwidth for wireless communication e.t.c., especially if these resources are scarce. This thesis will therefore only consider the centralized type of WNCS. The wireless

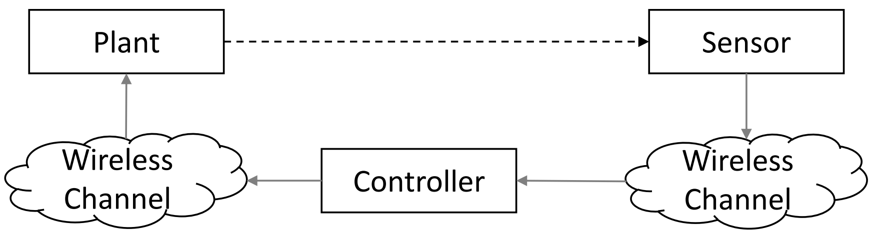

Figure 1.1: The general layout of a wireless networked control system

link can be either on the connection from the sensor to the controller or from the controller to the actuator at the plant. Fig. 1.1 illustrates the possibilities, either of the links or even both can be wireless. Examples for sensors remote from the plants might be cameras, e.g. for scanning the road marking in autonomous driving, infrared thermometers or ultrasonic distance sensors, e.g. for measuring the distance to the preceding vehicle.

1.2 Open Issues

Based on the general problems identified in the previous section, we have identified several research questions not yet covered by existing literature. Since the topics in the following chapters are related, but the focus is individual, the relevant literature will be discussed in the beginning of each chapter.

In the first step, a sensor takes measurements, which are required by the controller to assess the subsystem state and generate control inputs accordingly. If this wireless link is limited in capacity, the following questions arise:

1. How to adapt the transmission data rate between sensor and controller to the current state of the respective subsystem? Which reduction in data rate is possible, if only a certain accuracy of the quantity of interest is required by the current subsystem state?
2. What is the influence of the initial measurement error of the sensor and the error introduced by quantizing the measured quantity on the final subsystem state estimation error at the controller?

If there are multiple subsystems competing for transmission resources, it might be not possible to transmit sensor values constantly for all subsystems, if the available communication resources are insufficient. Instead, only part of the subsystems can transmit a new sensor reading to the controller at a time, while the others will have to wait. This impairs the ability of the controller to observe the subsystem states, which ultimately leads to a suboptimal control input. Facing this situation, we identified the following questions:

3. How can we improve on the knowledge of the controller about the subsystem states, if there are not enough resources to transmit an update of each subsystem sensor to each controller permanently? Is there a simple scheme to follow, which takes the different characteristics of subsystems into account, which does not require exhaustively checking all possible subsets of subsystems for transmission?

4. How to select subsystems for state transmission, if the age of the previously sent information and the subsystem characteristics are equal?

Finally, the controller has decided for a control input to the plant. Now, if this input has to be transmitted over a shared communication medium, the real-time properties of this medium are influenced by all subsystems jointly. We investigated, how this problem can be tackled in an Orthogonal Frequency Division Multiplex (OFDM) system, which results in these questions:

5. What are the implications of the very small amounts of data such commands are comprised of?
6. How should time-frequency resources be allocated to reduce the energy required to transmit commands to multiple agents at a time, while each agent has a certain deadline to receive the command?

1.3 Contributions and thesis overview

This thesis will focus on centralized WNCS, where the allocation of communication resources is carried out by a central scheduler. While the exact system models in the following chapters will differ, a common property is the central controller working with one or more agents. Generally, each agent comprises a control loop, which consists of a sensor, a controller, and a plant, like shown in Fig. 1.1.

First, only the link from the sensor to the controller will be considered in Chapter 2. The available communication resources on this link are assumed to be limited in the available transmit time or energy. To assess the influence of these reduced resources on the plant state estimation at the controller, in a first step, the acquisition of sensor data from the plant by the sensor, the subsequent wireless transmission to the controller and, finally, the plant state estimation at the controller is investigated and adapted to the available resources. Therefore, the quantization and mapping of the continuous plant state to a data word is adjusted to reduce the state estimation error at the central controller addressing issue No. 1 from Section 1.2. There are two sources of the estimation error at the controller, the measurement noise at the sensor itself and the quantization error caused by the translation to data words. While the measurement noise can be tackled by combining multiple independent measurements, the quantization error can be reduced by increasing the number of different data words and, hence, quantizations steps. In a second step, we show that there is a tradeoff between the number of measurements to average over at the sensor, which reduces the measurement error at the sensor, and number of quantization steps, which allows for a finer resolution of the transmitted values to reduce the quantization error, if the total time or energy available for taking measurements and transmit them is limited. The tradeoff is shown to have a pareto-optimal solution for minimal estimation error at the central controller, addressing issue No. 2.

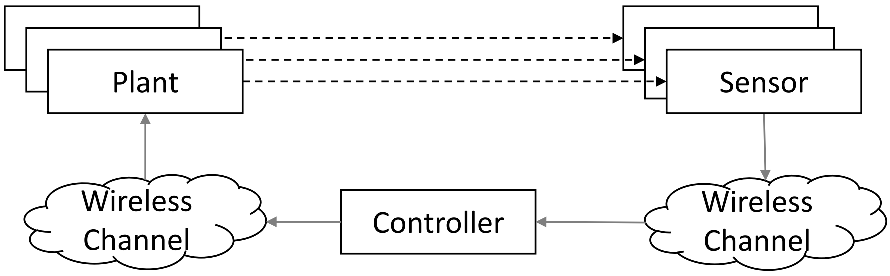

Figure 1.2: The layout of a wireless networked control system with multiple subsystems and wireless command and sensor value transmission

In Chapter 3, multiple subsystems, each consisting of a plant and a sensor, which are controlled by a central entity, share wireless transmission resources for sensor value and control command transmission, as depicted in Fig. 1.2. The subsystems are assumed to be scalar and linear. The focus in this chapter is on the central scheduler at the controller itself. The scheduler now has to coordinate the sensor value and control command transmissions of all the subsystems, each consisting of a linear plant with Gaussian system noise and a sensor. The control goal in Chapter 3 is to steer each plant to an equilibrium state. To calculate optimal control inputs, which minimize the deviation from the equilibrium state, the controller must have information about each of the plant states, which is gathered by the corresponding sensors [Lun16b]. If multiple subsystems are to be monitored, the wireless communication resources have to be split and, thus, a scheduling of the updates sent from the sensor to the controller is needed to prevent collisions of the transmissions. If the available communication resources are insufficient to always transmit updates from all sensors, a prioritization has to be implemented. Based on the deviation from the control goal as the optimality measure, an adapted update scheme for the sensor readings is derived, considering the individual system constant of each subsystem and the system noise levels, like in question No. 3 from Section 1.2. We show that the optimum update scheme for the considered scalar linear subsystems is a regular update scheme, where the update rate is calculated based on the subsystem characteristics. To apply this optimum rate, we present a scheduling algorithm, which distributes the inevitable deviation from this calculated optimum rate, which is induced by integer effects to all subsystems, equally, considering the time passed since previous transmissions for the same subsystem and, thus, tackling question No. 4.

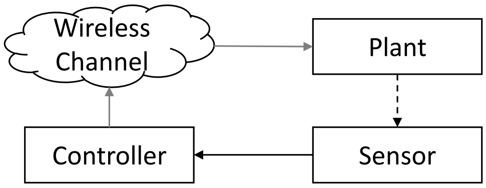

Figure 1.3: The layout of a wireless networked control system with wireless command transmission

The transmission of commands to the individual agents is then further investigated in Chapter 4. The considered system layout is shown in Fig. 1.3. The mentioned requirements of control systems, like minimizing the amount of data transmitted while maintaining required low latency, are considered by applying a correction term derived in [PPV10] to the well-known Shannon capacity formula. The Shannon capacity is only valid for infinitely long code lengths, which is approximately true for systems transmitting larger amounts of data or over longer times. The correction term from [PPV10] introduces an allowable probability of error to cope with the short packages. Based on this correction, an optimum time-frequency resource allocation for an OFDM system, considering individual command deadlines for each agent, is found.

Chapter 2

Sensor Value Quantization and Transmission

Several parts of the content of this section have been originally published by the author of this thesis in [KASK19]. This paragraph shall illustrate the previous work from [KASK19], the relation to and the additional work presented in this chapter. Especially the model and the derivations of the probability density functions, as well as the results for the entropy based quantization scheme have already been published. Extending [KASK19], another adaptive nonlinear quantization scheme is introduced for even better estimation results, compared to [KASK19]. As a benchmark for both nonlinear schemes, a fixed linear quantization scheme is employed and shown to be inferior to the proposed nonlinear schemes. Furthermore, results for different distributions of the parameter of interest are shown.

2.1 Introduction

This chapter considers the wireless transmission of values from the sensor to the controller over a wireless channel as part of the control loop, as shown in Fig. 2.1. The sensor measures the state of the plant once or multiple times and quantizes the measurements to a single data word. This data word is then transmitted over the wireless

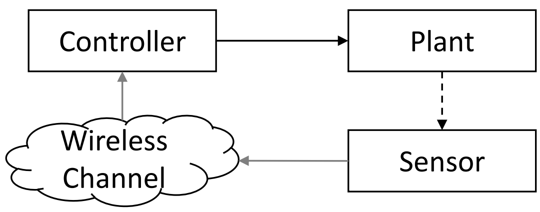

Figure 2.1: The networked control loop with wireless measurement data transmission from the sensor to the controller

communication link to a controller, which generates a control value to control the plant. The plant is either connected to the controller by wire or also uses a wireless connection to receive control values. This second connection is assumed to be ideal throughout the chapter, hence it is not considered in the problem formulation. Such a layout is often used in multi-agent control systems, where the sensors will take measurements of the plants' state parameters, which are not used directly at or close to the sensor, thus needing communication. This chapter focuses on the measurement, quantization and subsequent transmission of sensor values to the controller.

One example for such control systems is smart logistics [WHZ18], where many small vehicles act as the plants and distribute products. During operation, a huge amount of data, e.g. position, remaining fuel, or battery power, is collected by sensors on the vehicles and is then used at central controllers to calculate individual control actions for the devices, for example to schedule refueling of vehicles. Since the devices are moving most of the time, a wireless connection is required.

Another example are cognitive buildings, which have numerous sensors installed in all areas [PBB18]. Here, environmental data, like temperature, humidity, or air pressure is collected as well as the presence of humans is detected. From the collected data, the overall system behavior including the user preferences can be learned by a central controller. This controller then drives plants like blinds, lightning or heating. Even in this static application with fixed sensors, a wireless connection simplifies the installation, especially in existing buildings.

From those examples, it can be clearly seen that the sensors and corresponding controllers are often separated and a wired communication link is not desired for different reasons. In this case, the communication has to be wireless, and the increasing density of devices increases the competition for the limited wireless communication resources. Additionally, the devices often rely on battery power, which imposes additional constraints on the energy consumption of the systems.

According to the first two questions in Section 1.2, we want to find a scheme to translate these input values from the different sensors to data words. Examples for sensors are temperature sensors, humidity sensors and Global Navigation Satellite System (GNSS)s, providing the data as analog voltages or in digital form as high-resolution values. To reduce the amount of different data words, ranges of sensor output values can be aggregated. In the temperature case for example, ranges of 5 °C could be translated to a single data word and too high or too low temperatures, which do not occur in the considered system, could be neglected completely in the encoding. Similarly, for the GNSS, locations, which are never used by the device, e.g. in the middle of the ocean or on a different continent, could be left out in the quantization. In this chapter, quantization schemes with a certain number of steps are optimized to minimize the system state estimation error at the controller. Therefore, the distribution of possible sensor values, i.e. which actually occur in the considered system, has to be known.

In this work, we focus on a single sensor-receiver pair and use the properties of the underlying measurement model and communication model to jointly optimize the quality of the estimation at the receiver. Multiple noisy measurements of a parameter are taken and aggregated afterwards. The aggregated value is quantized and transmitted over a wireless channel. For both phases, measurement and transmission, only limited time and energy is available. The time limit results from the large amount of devices competing for transmission time. This limit can directly be translated to a limit of the data bits that can be transmitted in one time slot. On the other hand, a high transmit power to improve the signal-to-noise ratio (SNR) at the receiver might drain a prohibitively high amount of energy from the batteries of mobile or embedded devices. For this reason, a limitation in resolution of the quantization prior to the transmission is needed. As each individual measurement is also consuming time and power, the two phases of measuring and transmitting compete for the available time and energy resources. To find the best ratio of number of measurements and number of quantization intervals for a given time or energy limit, the Bayes risk is used as an estimation quality measure. This joint optimization of the number of measurements and the number of quantization steps allows for a minimum Bayes risk for given time or energy resource constraints.

Next, we will give an overview over the related work and then introduce the system model of this chapter in Section 2.3. To model the problem, first the aggregated probability-density function (pdf) is derived in Section 2.4.1. In the next two subsections, Section 2.4.2 and Section 2.4.3, the considered estimators are introduced. Afterwards, the two adapted quantization schemes are shown in Section 2.4.4 and Section 2.4.5. For comparison, a linear benchmark quantization scheme is shown in Section 2.4.6. Finally, the schemes are compared using numerical experiments in terms of conveyed mutual information in Section 2.5.2 and Bayes risk in Section 2.5.3. Furthermore, the tradeoff between the number of measurements and the number of quantization steps for limited time and energy is shown in Section 2.5.4.

2.2 Related Work

To give an overview over the current state of the art of this topic, we will now discuss other works, which investigated related problems. While our focus is on a centralized sensing scenario in the networked control systems domain, the aspect of sensor value transmission was already subject in other fields. Hence, first three papers with direct relation to WNCS are presented, after that three papers considering the transmission of sensor values over limited communication links in general are shown. A central sensing scenario is considered in [ZCWF18], where a discrete-time multi-agent scenario is assumed. Each agent consists of a controller and the plant, a central entity can take noisy measurements of the plant states and signals them over a wireless link to the controllers. The communication resources are limited, thus, not all controllers can receive updates in all time slots, so the central entity has to find a scheduling for the transmissions. The communication is scheduled based on the system state deviation from the equilibrium, but the control law and the quantization prior to transmission is not adapted to the communication channel state. Effects of the quantization on the plant state values received at the controllers are not considered.

A decentralized model is used in [VMKH16], where, similar to the previous example, autonomous agents in the field compete for limited communication resources used to transmit sensor values. In [VMKH16], however, there is no central entity sensing the plant states and scheduling the available communication resources. Instead, the allocation is done in an ALOHA-fashion. Again, the effects of state quantization are not considered, as well as there is no mechanism to prioritize plants with a higher state deviation for transmission.

In [CL16], a single control loop is considered. A wireless link is used to transmit sensor data to the controller, the controller in turn is directly attached to the plant, like in Fig. 2.1. The transmission is done in a multiple-input multiple-output (MIMO) fashion, the actual transmit signal is generated directly from the plant state variables by multiplying with a MIMO precoding vector as a preprocessing step. This analog transmission allows for a direct translation of communication channel noise to errors in the plant state estimation at the controller. The transmission is constrained by the available transmit energy, which is generated from a stochastic energy harvesting process. The objective is to guarantee the stability of the control loop in the Lyapunov sense and minimize the plant state estimation error at the controller for a given sensor battery capacity. While the sensor is assumed to be error-free, the noise during the wireless transmission and the system noise lead to uncertainty at the controller.

In all these works, the data from the sensor is not interpreted and processed prior to transmission, except for the MIMO precoding in [CL16], but rather the raw values are transmitted. The acquisition is not adapted to the communication system state, and similarly, the process of measuring and then transmitting the data is not adapted to the state of the underlying control system plant. Sensor outputs with a certain resolution are not compressed for transmission by reducing the resolution, even if the control system plant state would only require a coarse control action, which can also be generated from a coarse input. This missing preprocessing and adaptation leads to the research question No. 1 from Section 1.2 on how such a preprocessing could improve the tradeoff between estimation accuracy and data rate, which will be discussed in this chapter in Section 2.4.9.

To adapt also the communication itself and not only schedule the transmissions for the subsystems, there are multiple possibilities. In [SKMN15], the influence of very low resolution analog-to-digital conversion is considered. A receiver with only 1-bit quantization is used for estimation. The estimation performance is improved by exploiting information about the temporal evolution of the estimated parameter known a-priori at both, the transmitter and receiver. While not directly targeted at control systems, the crucial knowledge of the distribution and temporal evolution of the parameter of interest is common to our work.

In [LMZ+16], a digital transmission chain is used to transmit sensor data, which is already discretized by the sensor nodes, in a sensor network. The objective is to find the optimum power allocation in transmitting the sensor values reliably, while maximizing the battery lifetime of the sensor network. Tuning the power has direct influence on the packet loss probability for a given packet size.

A slightly different objective is considered in [KYY+17], which looks at the packet size of sensor networks and tries to find a trade-off between packet error probability and data integrity.

While those works consider the quantization and preprocessing, the influence and possible tradeoffs with the sensing and control task are not considered. Hence, we also elaborate on research question No. 2 in this chapter, which influence the initial measurement error of the sensor and the quantization for digital transmission, respectively, have on the state estimate at the controller.

2.3 System Model

This chapter focuses on the sensor value measurement and transmission, so only the part of the control loop in Fig. 2.1 with the sensor, wireless channel and controller will be modeled. The detailed chain of the measuring sensor with transmitter, receiver and estimator is shown in Fig. 2.2. The parameter of interest, denoted by w, is observed by a sensor and impaired by noise, denoted by m. A batch of N_{meas} noisy measurements of w is taken sequentially, denoted by $x_1, \ldots, x_{N_{\text{meas}}}$. Each individual measurement takes the time T_{meas} and the energy E_{meas}. The measurement values are then aggregated into a single value s, which is subsequently quantized into one of Q_{quant} data symbols, denoted by y. The symbol y is then transmitted over a wireless communication channel and distorted by receiver noise z. The number Q_{quant} of quantization steps determines the time and energy spent for transmission, denoted by T_{tx} and E_{tx}, respectively. The total energy and time for measuring and transmitting are limited by $E_{\max}$ and $T_{\max}$, respectively. The received symbol, denoted by y', is then decoded according to a codebook. The output v of the decoder is used by the estimator Ψ to generate the estimate $\hat{w}$ of w. In the next subsections, the individual steps are described in detail.

2.3.1 Measurement Model

The value of $w \in \mathbb{R}$ is assumed to lie between $w_{\min}$ and $w_{\max}$ and follow a known pdf $p_W(w)$. w is assumed to stay constant during the N_{meas} measurements, but with varying noise m. The noise is assumed to be Additive White Gaussian Noise (AWGN) with zero-mean and variance σ_{M}^2. The complete measurement phase takes the time $T_{\mathrm{acq}} = N_{\mathrm{meas}} T_{\mathrm{meas}}$. Likewise, the complete energy for measuring is $E_{\mathrm{acq}} = N_{\mathrm{meas}} E_{\mathrm{meas}}$. The pdf $p_W(w)$ as well as $w_{\min}$, $w_{\max}$ and σ_{M}^2 are assumed to be known at the transmitter and the receiver, since they all are properties of the sensor and the observed process.

To generate the aggregated value s, the measurement values

$$x_n = w + m_n, n = 1, \ldots, N_{\mathrm{meas}} \tag{2.1}$$

are summed up

$$s = \sum_{n=1}^{N_{\mathrm{meas}}} x_n, \tag{2.2}$$

which is assumed to take no additional time or energy. Instead of the sum the mean could also be chosen, since they are related by the number of measurements and the receiver can find the same optimum trade-off between the number N_{meas} of measurements and the number Q_{quant} of quantization steps as the sensor, since the properties of the random value w, the measurement noise m and the resource limits are known to the receiver.

2.3.2 Quantization and Transmission Model

The quantization of the sum value s to the data symbol y is carried out according to a Q_{quant}-step function $\phi_{\mathrm{Q}} : \mathbb{R} \mapsto \{1, 2, \ldots, Q_{\mathrm{quant}}\}$. To allow for a high accuracy of the state estimate, the amount of information about the parameter of interest w contained in a single data symbol y should be maximized. In Section 2.4, ϕ_{Q} is derived based on the pdfs $p_W(w)$ of the parameter of interest w, the measurement noise pdf $p_M(m)$ of m and the resulting aggregated value pdf $p_S(s)$ of s. The resulting transmit symbol y is transmitted over a wireless communication channel. This transmission is subject to receiver noise z, which is assumed to be AWGN with zero-mean and variance σ_{Z}^2. The channel has the constant channel coefficient h and the transmit power is given by P. For a capacity

$$C = \log_2\left(1 + \frac{hP}{\sigma_{\mathrm{Z}}^2}\right), \tag{2.3}$$

the channel is assumed allow for error-free communication [Sha48], thus, for the received symbol, $y' = y$ applies. (2.3) is an upper bound for the number of bits, which can be transmitted error-free in one channel use, i.e. per second and per Hertz. This bound will be attained for an infinite code length. Therefore, it is not a valid assumption for very small data packets. However, throughout this chapter, we will assume the sensor readings are transmitted as parts of larger packets, which allows for using the Shannon Capacity as an approximation of the achievable capacity, when spending a certain amount of power, or, looking at it the other way around, how much energy is required when transmitting a data packet of given size over a channel with given noise level.

All transmit symbols are encoded by the same number N_{bits} of bits. In [KASK19], the quantization is designed to make the transmit symbols equally probable and, thus, carry the same amount of information. N_{bits} is then determined by

$$N_{\mathrm{bits}} = \log_2(Q_{\mathrm{quant}}), \tag{2.4}$$

which is not necessarily an integer number, if Q_{quant} is not a power of 2. To transmit a non-integer number N_{bits} of bits, the communication systems symbol alphabet has to be designed with Q_{quant} different symbols. In this thesis, a quantization scheme with optimized estimation capabilities will be designed. Thus, the transmit symbols will not necessarily always carry the same amount of information, but to compare with the scheme from [KASK19], the equal size of the transmit symbols is kept.

The time T_{bit} consumed for transmitting one bit is determined by the channel capacity C as $T_{\mathrm{bit}} = \frac{1}{C}$. Since increased transmission power P increases the capacity C logarithmically, linearly increasing the energy E_{tx} for transmission logarithmically increases the possible number N_{bits} of bits. This results in a direct proportionality of Q_{quant} and E_{tx}, i.e. $E_{\mathrm{tx}} = Q_{\mathrm{quant}} E_{\mathrm{quant}}$.

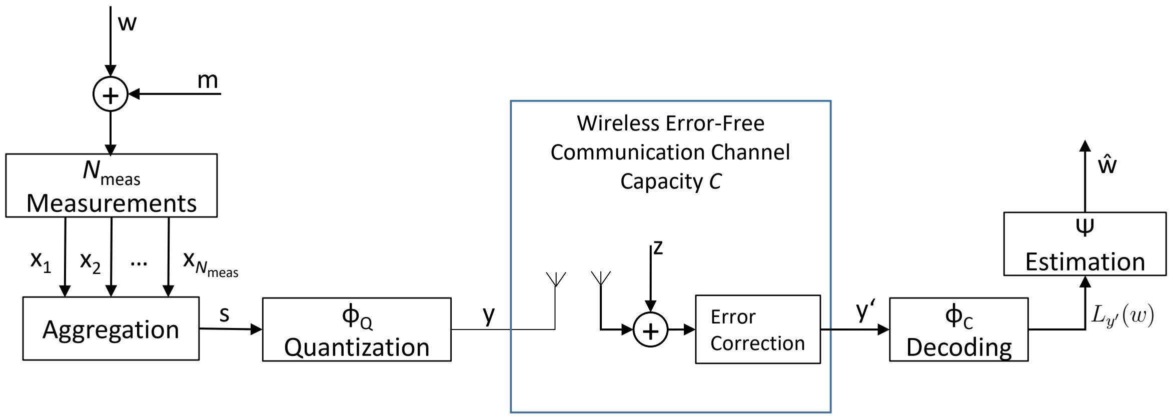

Figure 2.2: System model of the sensor value quantization and transmission

2.3.3 Estimation Model

The received data symbol y' is used to generate a likelihood function $L_{y'}(w)$ for a certain received y' for the parameter of interest w considering the knowledge about the aggregated value s. The likelihood function $L_{y'}(w)$ is subsequently used by the estimator Ψ to generate the estimate $\hat{w}$ of w. The optimization objective is to minimize the Bayes risk of the estimation, which is a metric for the estimation accuracy [Shy12]. It is calculated according to a distortion function l, i.e. $R_\text{B} = \text{E}\{l\,(W;\Psi(V))\}$. In this work, l is chosen to take the form $l(w,\Psi(v)) = |w-\Psi(v)|^p$ with $p \geq 1$, i.e.

$$R_\text{B} = \text{E}\{|w-\Psi(v)|^p\}. \tag{2.5}$$

For the well-known minimum mean-square error (MMSE) estimator, $p = 2$ applies. The MMSE estimation minimizes the squared error, which is suitable, if the cost increases quadratically with the parameter of interest, e.g. for an error in a voltage measurement, which results in increased power consumption. In other cases, where there is a linear dependency of the cost on the estimation error, the minimum absolute-value error (MAVE) estimator is used with $p = 1$. An application example would be a distance measurement of drones, where the error in distance linearly increases with the time to reach a certain point with constant velocity.

2.3.4 Constraint Model

The estimation process is constrained by limited time or energy resources, which are shared between the measuring and the transmission phase. The sum time or sum energy taken by the measurement of w and the transmission of y must not exceed a certain limit, T_{max} or E_{max}, respectively. In the time limited case, the number N_{meas} of measurements determines T_{acq}, the number N_{bits} of bits determines T_{tx}. Since the transmission of the aggregated value s cannot start till all measurements have been taken, the total time spent for measuring and transmitting is sum of these times. It must not exceed the available time, i.e.

$$T_{\text{max}} \geq T_{\text{tx}} + T_{\text{meas}}. \tag{2.6}$$

Likewise, in the energy limited case, the total energy is the sum of the energy E_{acq} consumed for measuring, determined by N_{meas}, and the energy E_{tx}, determined by Q_{quant}. The energy is constraint is then

$$E_{\text{max}} \geq E_{\text{acq}} + E_{\text{tx}}. \tag{2.7}$$

2.4 Problem Formulation

The estimation of the quantity of interest w at the receiver is based on a Bayes estimation scheme. First, the probability distributions for the non-quantized case are derived. Based on these distributions, the estimators for the MMSE and MAVE case are calculated. Then, the quantization and codebook-based reconstruction is introduced. Finally, the Bayes risk, which includes the influence of the quantization, is calculated.

2.4.1 Probability distribution $p_S(s)$

In this section the pdf $p_S(s)$ is derived. The pdf of the parameter w of interest is $p_W(w)$, which is only non-zero for $w_{\text{min}} \leq w \leq w_{\text{max}}$. Each measurement is impaired with the measurement noise m, which is i.i.d. Gaussian distributed, i.e.

$$p_M(m) = \frac{1}{\sqrt{2\pi\sigma_{\text{M}}^2}} \exp\left(-\frac{m^2}{2\sigma_{\text{M}}^2}\right). \tag{2.8}$$

N_{meas} measurement values are taken, which results in the vector

$$\boldsymbol{x} = w \cdot \mathbf{1}_{N_{\text{meas}}} + (m_1, m_2, \ldots, m_{N_{\text{meas}}})^{\text{T}} \tag{2.9}$$

of measurement values, where $\mathbf{1}_L$ is the all-ones vector with L elements. The sum of i.i.d. Gaussian random variables is again Gaussian, which results in

$$\sum_{n=1}^{N_{\text{meas}}} x_n = s \tag{2.10}$$

$$p_{S|W}(s|w) = \frac{1}{\sqrt{2\pi\sigma_{\mathrm{M}}^2}} \exp\left(\frac{(s - N_{\text{meas}} \cdot w)^2}{2\sigma_{\mathrm{M}}^2}\right). \tag{2.11}$$

This leads to the joint probability of s and w:

$$p_{S,W}(s,w) = p_{S|W}(s|w) \cdot p_W(w) \tag{2.12}$$

To get the unconditional pdf of s, the marginal probability w.r.t. w is calculated as

$$p_S(s) = \int_{w_{\min}}^{w_{\max}} p_{S,W}(s,w)\mathrm{d}w. \tag{2.13}$$

Applying the Bayesian theorem [PP02], the conditional pdf of the parameter of interest w for a given sum of measurements s can be calculated as

$$p_{W|S} = \frac{p_{S|W} p_W}{p_S} \tag{2.14}$$

2.4.2 Minimum Absolute Value Error (MAVE) Estimator

The MAVE estimator Ψ_{MAVE} based on the aggregated value s minimizes the mean absolute value error of the estimate, i.e. $p = 1$ in (2.5), and is defined in [Shy12] as the upper or lower bound, respectively, which splits the integral over the a-posteriori pdf $p_{W|S}(w,s)$ of the parameter of interest w given the aggregated value s into two equal parts, i.e.

$$\int_{-\infty}^{\Psi_{\text{MAVE}}(s)} p_{W|S}(w,s)\mathrm{d}w = \int_{\Psi_{\text{MAVE}}(s)}^{\infty} p_{W|S}(w,s)\mathrm{d}w = 0.5, \tag{2.15}$$

and after quantization, i.e. given only the bounds q_n and q_{n+1} of the quantization interval containing s, the a-posteriori pdf of w is given by

$$p_{W|Q}(w,n) = \frac{\int_{q_n}^{q_{n+1}} p_{W|S}(w,s)\mathrm{d}s}{\int_{q_n}^{q_{n+1}} p_S(s)\mathrm{d}s} = \frac{\int_{q_n}^{q_{n+1}} p_{W|S}(w,s)\mathrm{d}s}{\Pr(y' = n)} \tag{2.16}$$

and the MAVE estimator for quantization interval n by

$$\int_{-\infty}^{\Psi_{\text{MAVEq}}(n)} p_{W|Q}(w,n)\mathrm{d}w = \int_{\Psi_{\text{MAVEq}}(n)}^{\infty} p_{W|Q}(w,n)\mathrm{d}w = 0.5, \tag{2.17}$$

2.4.3 Minimum Mean-Square Error (MMSE) Estimator

The MMSE estimator $\Psi_{\mathrm{MMSE}}(s)$ based on the aggregated value s minimizes the mean squared error of the estimate, i.e. $p = 2$ in (2.5), and is defined as

$$\Psi_{\mathrm{MMSE}}(s) = \int_{w_{\min}}^{w_{\max}} w p_{W|S}(w, s)\mathrm{d}w, \tag{2.18}$$

which is the expected value of the a-posteriori pdf $p_{S,W}$. With the a-posteriori pdf of w after quantization $p_{W|Q}(w, n)$ we have

$$\Psi_{\mathrm{MMSEq}}(n) = \int_{w_{\min}}^{w_{\max}} w p_{W|Q}(w, n)\mathrm{d}w, \tag{2.19}$$

2.4.4 Mutual Information Based Quantizer Design

From information theory, it is known that the optimum communication channel usage is achieved, if the mutual information between the transmitted and received values is maximized [PS02]. In this scenario this corresponds to the mutual information $I(W; Y')$ between the parameter of interest w and the received data symbol y'. $I(W; Y')$ depends on the joint entropy $h(W, Y')$ of w and y', which in turn depends on the joint probability $p_{Y',W}(y', w)$ of y' and w

$$p_{Y',W}(y', w) = \int_{q_{y'}}^{q_{y'+1}} p_{S,W}(s, w)ds \text{ for } y' = 1, \dots, Q_{\mathrm{quant}} \tag{2.20}$$

$$h(W, Y') = \sum_{y'=1}^{Q_{\mathrm{quant}}} \int_{w_{\min}}^{w_{\max}} p_{Y',W}(y', w) \log_2 \left(p_{Y',W}(y', w)\right) dw, \tag{2.21}$$

the differential entropy h_W of the parameter of interest w

$$h_W = \int_{w_{\min}}^{w_{\max}} p_W(w) \log_2 \left(p_W(w)\right) dw, \tag{2.22}$$

and the entropy H_Y of the received symbol

$$H_{Y'} = \sum_{n=1}^{Q_{\mathrm{quant}}} \int_{q_n}^{q_{n+1}} p_S(s)ds \log_2 \left(\int_{q_n}^{q_{n+1}} p_S(s)ds \right). \tag{2.23}$$

Please note, that h_W and $h(W, Y')$ are differential entropies since the random variable is continuous and, thus, denoted by a small h, whereas the possible values for the received symbols y' are discrete. Therefore, its entropy $H_{Y'}$ is denoted by a capital H. The mutual information is then given as

$$I(W; Y') = H_{Y'} + h_W - h(W, Y') \tag{2.24}$$

The differential entropy of the parameter of interest w, h_W, only depends on its pdf $p_W(w)$. The joint entropy $h(W, Y')$ and the entropy of the received symbols $H_{Y'}$ depend on the number of measurements N_{meas}, the characteristics of the measurement noise m and the quantization intervals $q_1, \ldots, q_{Q_{\text{quant}}}$. This leads to the optimization problem

$$\max_{q_2,\ldots,q_{Q_{\text{quant}}}} \quad I(W; Y') \tag{2.25}$$

$$\text{s. t.}$$

$$q_n < q_{n+1}, \text{ for } n = 1, \ldots, Q_{\text{quant}} \tag{2.26}$$

$$q_1 = -\infty \tag{2.27}$$

$$q_{Q_{\text{quant}}+1} = \infty \tag{2.28}$$

for finding the optimum quantization interval bounds $q_1, \ldots, q_{Q_{\text{quant}}-1}$. Constraint (2.26) ensures that the bounds of the quantization intervals are in monotonically increasing order. Constraints (2.27) and (2.28) fix the bounds of the intervals with the largest and smallest values to make sure the quantization covers all real numbers.

2.4.5 Simplified Transmit Symbol Entropy Based Quantizer Design

For finite symbol alphabets, a uniform distribution of the symbols maximizes the mutual information [PS02]. To achieve this uniform distribution for the quantizer outputs, the quantization intervals are designed according to $p_S(s)$. In a first step, the cumulative distribution function (cdf) of S is calculated as

$$P_S(s) = \int_{-\infty}^{s} p_S(s')\mathrm{d}s'. \tag{2.29}$$

Then, the quantization interval bounds q_n are calculated, with $-\infty$ as left bound of the first interval q_1 and with $+\infty$ as right bound of the last interval $q_{Q_{\text{quant}}+1}$. The bounds $q_2, \ldots, q_{Q_{\text{quant}}}$ in between are calculated by solving the equation

$$\frac{n-1}{Q_{\text{quant}}} \stackrel{!}{=} P_S(q_n) = \int_{-\infty}^{q_n} p_S(s')\mathrm{d}s', \tag{2.30}$$

which provides equally probable output symbols. Making the transmitted symbols $y_1, \dots, y_{Q_{\text{quant}}}$ equally probable maximizes (2.23), but does not consider (2.21). Solving (2.30), however, is much less complex than solving (2.25), reducing computational effort.

2.4.6 Linear Quantization Based Design

For comparison, a simplistic linear quantizer with intervals of equal width is considered. It can also be used, if the exact pdf $p_W(w)$ is not known, but only the limits w_{min} and w_{max} of the parameter of interest w. For this approach, $-\infty$ is, like for the mutual information based and the transmit symbol entropy based quantizers from Section 2.4.4 and Section 2.4.5, chosen as left bound q_1 of the first interval and with $+\infty$ as right bound $q_{Q_{\text{quant}}+1}$ of the last interval. If $Q_{\text{quant}} = 2$, there are three bounds, of which the left is already fixed to $q_1 = -\infty$ and the right to $q_3 = \infty$. The central bound q_2 is set to

$$q_2 = N_{\text{meas}} w_{\text{min}} + N_{\text{meas}} \frac{w_{\text{max}} - w_{\text{min}}}{2}. \tag{2.31}$$

If $Q_{\text{quant}} > 2$, the right bound of the first interval is set to

$$q_2 = N_{\text{meas}} w_{\text{min}} \tag{2.32}$$

and the left bound of the last interval to

$$q_{Q_{\text{quant}}} = N_{\text{meas}} w_{\text{max}}. \tag{2.33}$$

The remaining interval bounds are set to

$$q_n = N_{\text{meas}} w_{\text{min}} + N_{\text{meas}} \frac{w_{\text{max}} - w_{\text{min}}}{Q_{\text{quant}} - 2} \tag{2.34}$$

2.4.7 Likelihood of the Quantity of Interest

Since the transmission errors from the noisy channel are assumed to be completely removed by the error correction, the received symbol y' is equal to the transmitted symbol y. At the receiver, the received symbol y' is used to look up the bounds $q_{y'}$ and $q_{y'+1}$ of the quantization interval containing the aggregated value s. Therefore, the likelihood of the parameter of interest w can be calculated from (2.14) and for the nth quantization interval it is given by

$$L_n(w) = \int_{q_n}^{q_{n+1}} p_{S,W}(w, s)\mathrm{d}s. \tag{2.35}$$

$L_n(w)$ is the pdf of w, given the received symbol was $y' = n$, weighted by the probability $\Pr(y' = n)$, i.e.

$$L_n(w) = p_{W|Q}(w, n) \Pr(y' = n) \tag{2.36}$$

Since the intervals and their bounds are known at the receiver in advance, it can create a codebook, which assigns an estimated value to each received symbol y'.

2.4.8 Calculation of the Bayes Risk

From (2.5), the Bayes risk of the MAVE estimator is given by

$$R_{\text{B, MAVE}} = \int_{-\infty}^{+\infty} \int_{w_{\min}}^{w_{\max}} |\Psi_{\text{MAVE}}(s) - w| \, p_{V,W}(s, w)\mathrm{d}w\mathrm{d}s, \tag{2.37}$$

and similarly for the MMSE estimator by

$$R_{\text{B, MMSE}} = \int_{-\infty}^{+\infty} \int_{w_{\min}}^{w_{\max}} (\Psi_{\text{MMSE}}(s) - w)^2 \, p_{V,W}(s, w)\mathrm{d}w\mathrm{d}s. \tag{2.38}$$

Since the quantization interval bounds $q_{y'}$ and $q_{y'+1}$ derived from the received symbol y', serving as the input value to the decoder, only take Q_{quant} different value combinations, there are only Q_{quant} possible likelihood functions $L_{y'}(w)$ and, thus, the estimates $\Psi_{\text{MAVEq}}(y')$ and $\Psi_{\text{MMSEq}}(y')$ will also take only Q_{quant} different values each. Then, the total Bayes risk for the estimators is given by

$$R_{\text{B, MAVEq}} = \sum_{n=1}^{Q_{\text{quant}}} \Pr(y' = n) \int_{w_{\min}}^{w_{\max}} |\Psi_{\text{MAVE}}(n) - w| \, p_{W|Q}(w, n)\mathrm{d}w \tag{2.39}$$

$$= \sum_{n=1}^{Q_{\text{quant}}} \int_{w_{\min}}^{w_{\max}} |\Psi_{\text{MAVE}}(n) - w| \, L_n(w)\mathrm{d}w \tag{2.40}$$

and

$$R_{\text{B, MMSEq}} = \sum_{n=1}^{Q_{\text{quant}}} \Pr(y' = n) \int_{w_{\min}}^{w_{\max}} (\Psi_{\text{MMSE}}(n) - w)^2 \, p_{W|Q}(w, n)\mathrm{d}w \tag{2.41}$$

$$= \sum_{n=1}^{Q_{\text{quant}}} \int_{w_{\min}}^{w_{\max}} (\Psi_{\text{MMSE}}(n) - w)^2 \, L_n(w)\mathrm{d}w. \tag{2.42}$$

2.4.9 Resource Constraints

The time and energy constraints, see section 2.3.4, limit the number N_{meas} of measurements as well as the number N_{bits} of bits which can be transmitted over the wireless channel in a given time or with a given amount of energy, respectively. Since it is assumed that the data words considered in this chapter are parts of large packets, Shannon capacity can serve as a valid approximation of the marginal increase of the required energy. For the transmission, a linear relation between the transmitted and received power expressed as the channel coefficient h is assumed. In the energy limited case, this leads to a linear relation between the available quantization steps in a given time interval T_{tx} and the transmission energy, as $Q_{\mathrm{quant}} = 2^{N_{\mathrm{bits}}}$ and $N_{\mathrm{bits}} = T_{\mathrm{tx}}C$. For this reason, a parameter γ_{E} is introduced to characterize the relation between the transmission energy E_{quant} needed for transmission of one additional quantization step and the energy E_{meas} consumed for each measurement, i.e.

$$E_{\mathrm{quant}} = \gamma_{\mathrm{E}} E_{\mathrm{meas}}. \tag{2.43}$$

For the time limited case, the time needed to transmit one bit and the time needed to take one measurement are related by a linear coefficient γ_{T}, i.e.

$$T_{\mathrm{bit}} = \gamma_{\mathrm{T}} T_{\mathrm{meas}}. \tag{2.44}$$

2.5 Numerical Results

2.5.1 Setup

All calculations are carried out with fixed measurement noise $\sigma_{\mathrm{M}}^2 = 9$. The range of w is set to $w_{\min} = 0$ and $w_{\max} = 100$. We will compare the quantization scheme with steps designed to maximize the mutual information $I(W; Y')$ from Section 2.4.4 to the simpler entropy based scheme from Section 2.4.5 and the linear quantization scheme described in Section 2.4.6. To show the influence of different pdfs of w, we will compare three different cases:

1. A triangular distribution, which was already shown in [KASK19]. One example for such a distribution is a system which senses the distance between a central point and an agent like a drone, which is located at a random spot around the central point with a uniform probability for each spot.

$$p1_W(w) = \begin{cases} \frac{2w}{w_{\max}^2 - w_{\min}^2} & \text{if } w_{\min} \leq w \leq w_{\max} \\ 0 & \text{otherwise.} \end{cases} \tag{2.45}$$

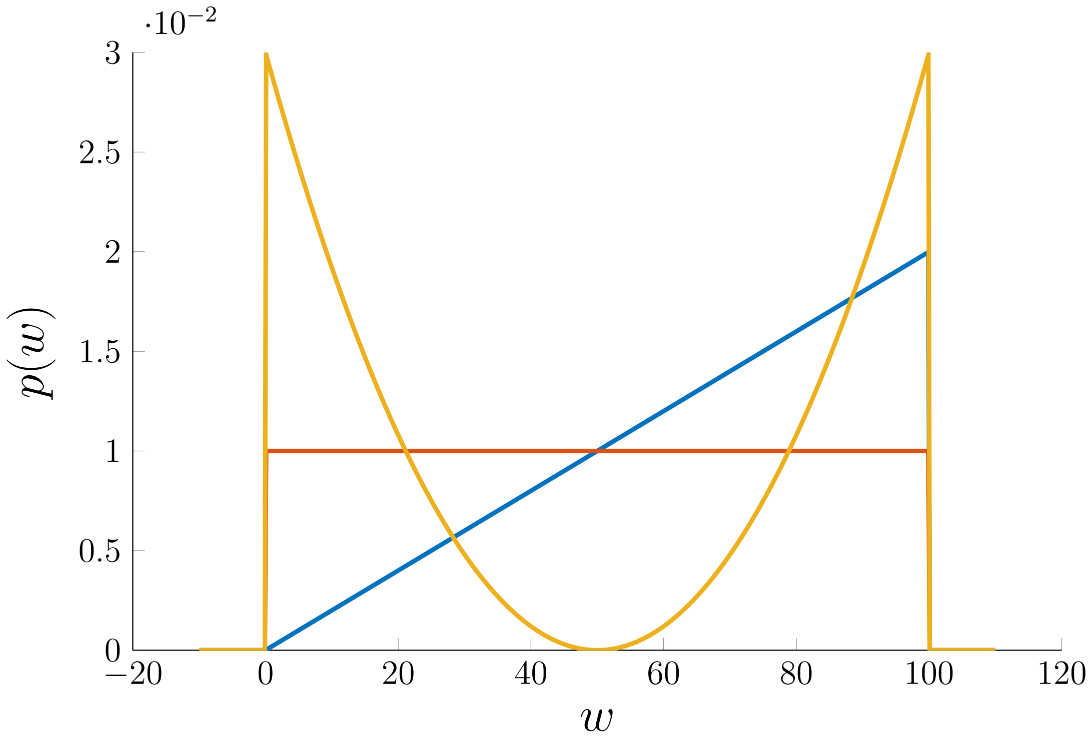

Figure 2.3: The applied distributions for the parameter w for $w_{\text{min}} = 0$ and $w_{\text{max}} = 100$

2. A uniform distribution of w between w_{min} and w_{max}. This distribution occurs, if all sensor values are equally likely. This distribution would also be used, if no a-priori information about the parameter of interest w and its distribution is available.

$$p2_W(w) = \begin{cases} \frac{1}{w_{\text{max}} - w_{\text{min}}} & \text{if } w_{\text{min}} \le w \le w_{\text{max}} \\ 0 & \text{otherwise.} \end{cases} \tag{2.46}$$

3. A bimodal distribution. In many processes the system states can be differentiated into two classes, which have a smooth transition region between them. An example is the weight of two different sizes of containers, which are filled with a liquid. While it is possible to get values in between for the not completely filled larger container, it is most likely to get values close to the two capacities. As a simple example for such a bimodal pdf we take the U-quadratic distribution:

$$p3_W(w) = \begin{cases} \frac{12}{(w_{\text{max}} - w_{\text{min}})^3} \left(w - \frac{w_{\text{min}} + w_{\text{max}}}{2}\right)^2 & \text{if } w_{\text{min}} \le w \le w_{\text{max}} \\ 0 & \text{otherwise.} \end{cases} \tag{2.47}$$

These pdfs (2.45)–(2.47) are shown in Fig. 2.3 for $w_{\text{min}} = 0$ and $w_{\text{max}} = 100$. In Fig. 2.4, the distribution of the measured parameter$s$ is shown for $N_{\text{meas}} = 1$. Since s includes

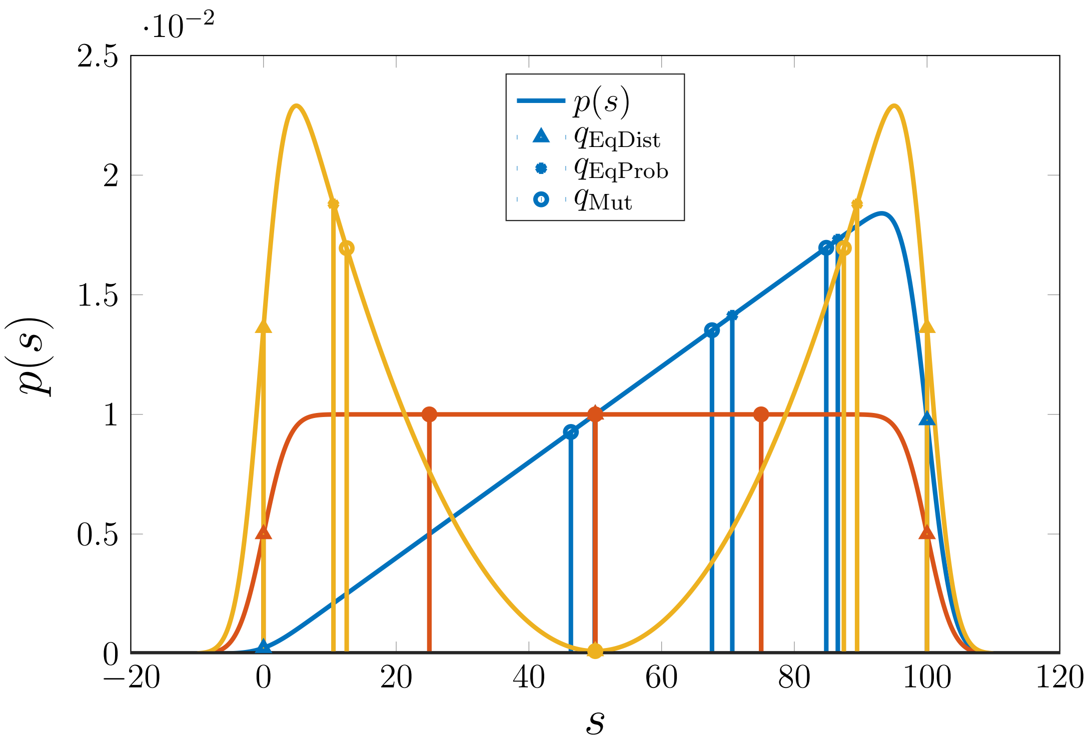

Figure 2.4: The resulting distributions for the measured parameter s for $N_{\mathrm{meas}} = 1$ and the quantization interval bounds for $Q_{\mathrm{quant}} = 4$

the measurement noise, the shape from Fig. 2.3 is convolved with the Gaussian pdf. For each pdf, the resulting quantization interval borders are marked by vertical lines. The quantization schemes are distinguished by the symbol at the top of each line, which are denoted as follows:

- Linear quantization scheme Section 2.4.6 $\rightarrow q_{\mathrm{EqDist}}$
- Entropy based quantization scheme Section 2.4.5 $\rightarrow q_{\mathrm{EqProb}}$
- Mutual information based quantization scheme Section 2.4.4 $\rightarrow q_{\mathrm{Mut}}$

The linear quantization scheme obviously has the same interval bounds for each pdf, since it only depends on w_{min} and w_{max}, but not on the actual shape of the pdf in between. The two other schemes are adapted to the pdfs, which leads to smaller intervals for larger values of the pdf, because the probability for w and, therefore, s to lie in this range is higher. A high probability for a data word corresponds to a low information of this symbol, hence the width of the intervals is reduced to equalize the information of all symbols, which optimizes the overall mutual information. It also shows that these two quantization schemes are generally not equivalent.

2.5.2 Mutual information

First, the mutual information conveyed using the different quantization schemes described in Section 2.4.4 – Section 2.4.6 shall be investigated and compared for different numbers Q_{quant} of quantization intervals. Fig. 2.5 shows, as the ordinates, the mutual

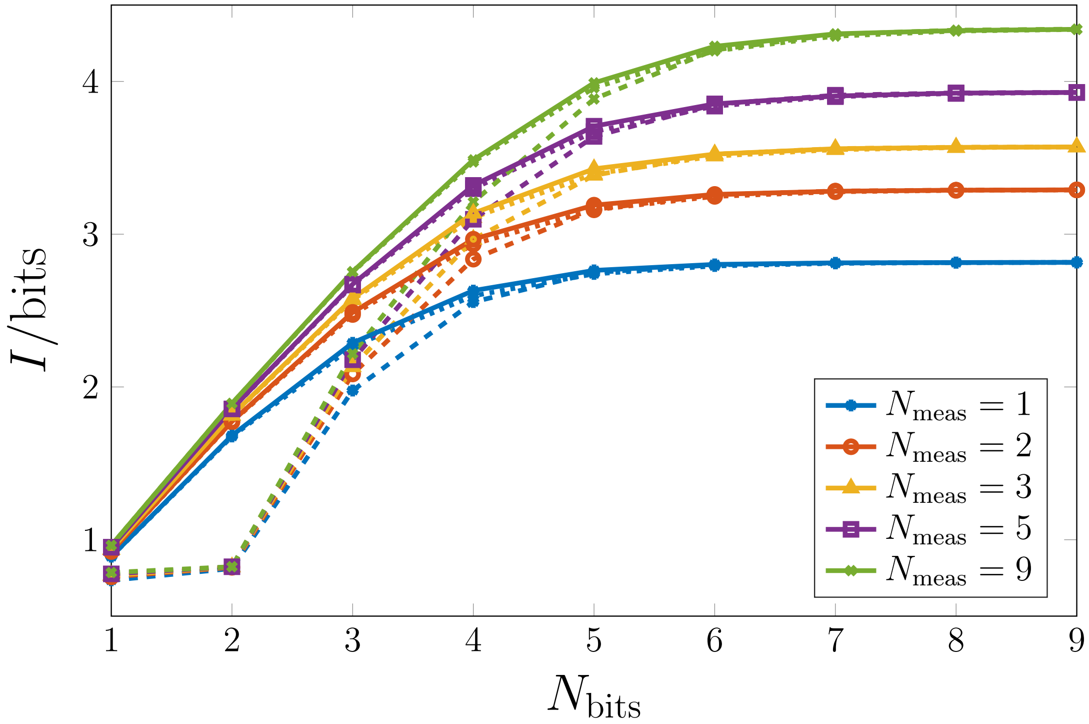

Figure 2.5: Mutual information for $p1_W$ (Linear quantizer: -- Entropy based: ··· Mutual info. based: —)

information between the parameter of interest w and the quantized value y, when w is distributed according to $p1_W$ with a triangular shape. As the abscissae, the number N_{bits} of bits is shown, which directly relates to the number Q_{quant} of quantization intervals, according to (2.4). Each color represents a different number N_{meas} of measurements. With increasing N_{bits}, the number Q_{quant} of quantization intervals increases accordingly, resulting in a finer quantization resolution. This leads to a higher mutual information I. For high N_{bits}, the mutual information I asymptotically reaches a bound, which is dictated by the measurement noise m.

The more measurements are taken, the better is the representation of w in the aggregated value s. Therefore, the influence of the measurement noise is reduced and the aforementioned bound is higher and, thus, the mutual information can also be higher for more measurements, if a sufficient number of quantization intervals is available. Especially for lower N_{bits}, the quantization based on linear quantization described in

Section 2.4.6 is largely outperformed by the quantization based on maximizing the mutual information from Section 2.4.4. The scheme described in Section 2.4.5 is better than the linear quantization, for low N_{bits}, approximately up to $N_{\text{bits}} = 5$, but not as good as the mutual information optimized scheme. The knee in the curves of the linear quantizer is due to the fact that for $N_{\text{bits}} = 2$, which translates to $Q_{\text{quant}} = 4$, two of the intervals are used for the range of $s \leq N_{\text{meas}} w_{\min}$ and $s \geq N_{\text{meas}} w_{\max}$. These ranges have a low probability to contain s, which is leading to a low information of the corresponding data symbols y.

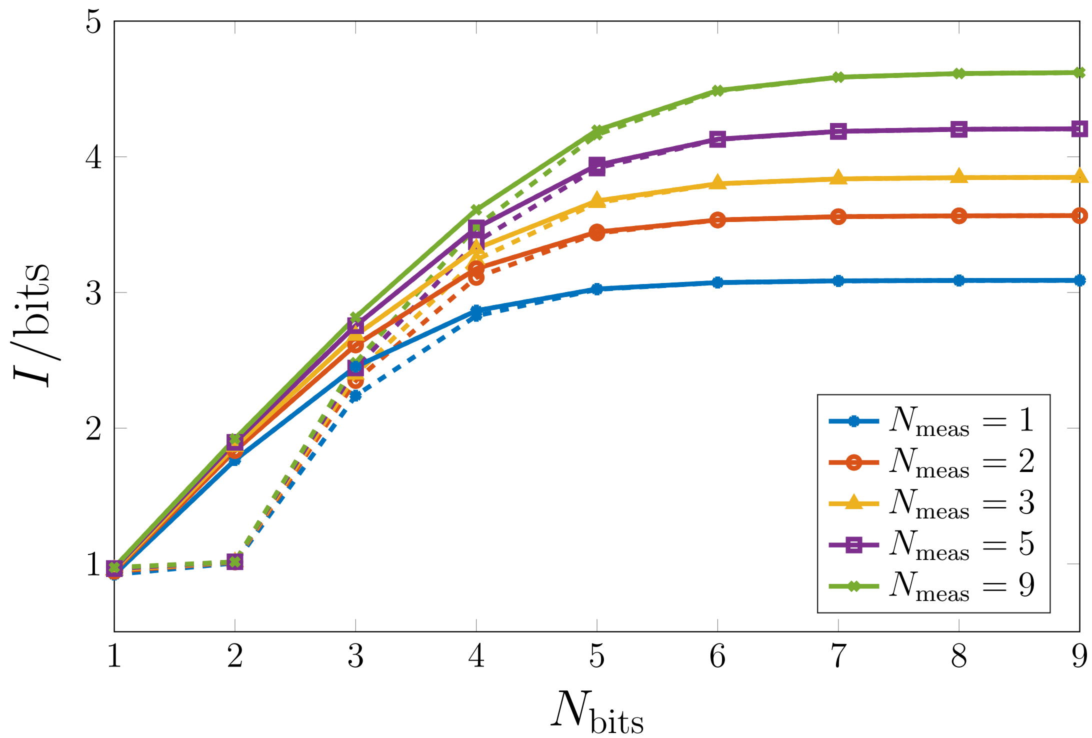

Figure 2.6: Mutual information for $p2_W$ (Linear quantizer: -- Entropy based: $\cdots$ Mutual info. based: —)

In Fig. 2.6 the mutual information for the three quantization schemes is, in the same way as in Fig. 2.5, shown for the distribution $p2_W$ from (2.46), which has a uniform shape. The entropy based quantization scheme shows the same results as the mutual information based one for this distribution. This is due to the fact, that the linear quantization provides the same quantization intervals for this pdf $p2_W$. The linear scheme can achieve results similar to the other schemes for $N_{\text{bits}} > 3$, because the optimum scheme is also a linear one, which can be seen in Fig. 2.4.

In Fig. 2.7, the mutual information for the U-qadratic distribution $p3_W$ is shown. Since the parameter of interest w tends to take values at the borders with a much

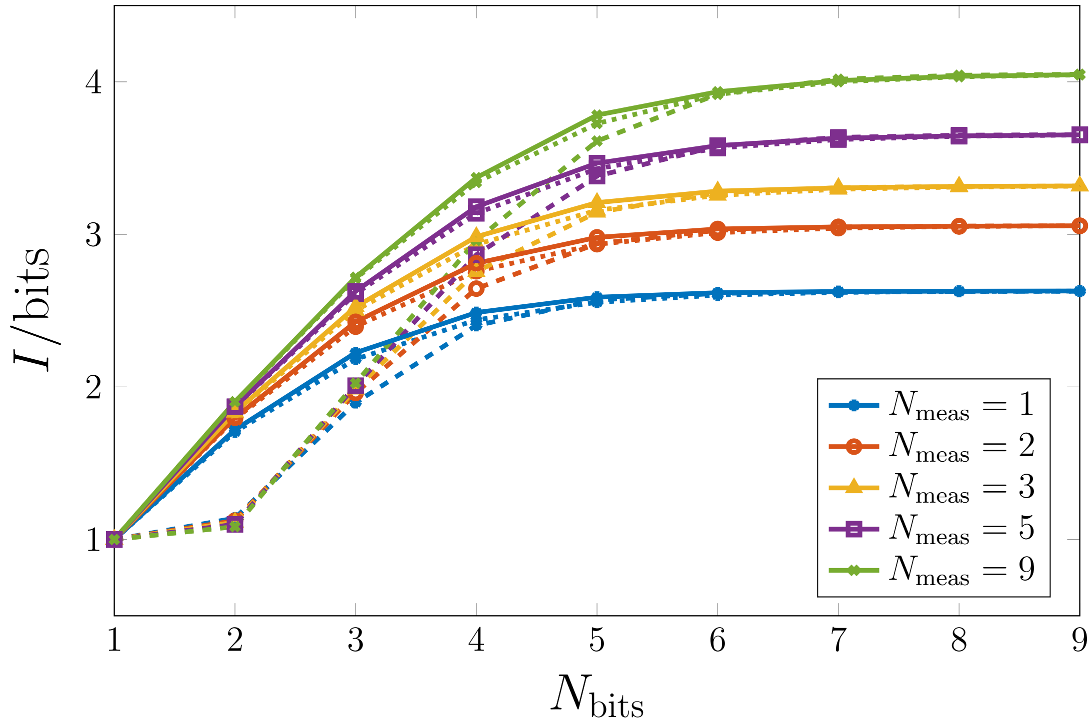

Figure 2.7: Mutual information for $p3_W$ (Linear quantizer: -- Entropy based: ··· Mutual info. based: —)

higher probability than values in the center, the performance hit of the linear scheme, compared to the entropy based and mutual information based, is the largest for this pdf $p3_W$. The linear scheme needs at least $N_{\text{bits}} \geq 5$ to reduce this performance gap to the other schemes.

2.5.3 Bayes risk

To get a general overview of the influence of N_{meas} and N_{bits} on the Bayes risk, serving as a measure for the error at the estimator, the Bayes risk R_{B} is calculated for up to $N_{\text{meas}} = 9$ measurements and $N_{\text{bits}} = 9$, resulting in up to $Q_{\text{quant}} = 512$ quantization steps with both estimators, MMSE and MAVE, for the three distributions of w. For each distribution, the mutual information based quantizer from Section 2.4.4 is compared to the entropy based quantizer from Section 2.4.5 and the linear quantizer from Section 2.4.6.

Fig. 2.8 shows $R_{\text{B, AVE}}$ as the ordinates for different values of N_{meas} as the abscissae for the MAVE estimator with all three quantizers for the triangular distribution $p1_W$. The more measurements are taken and the higher the quantization resolution is, the lower

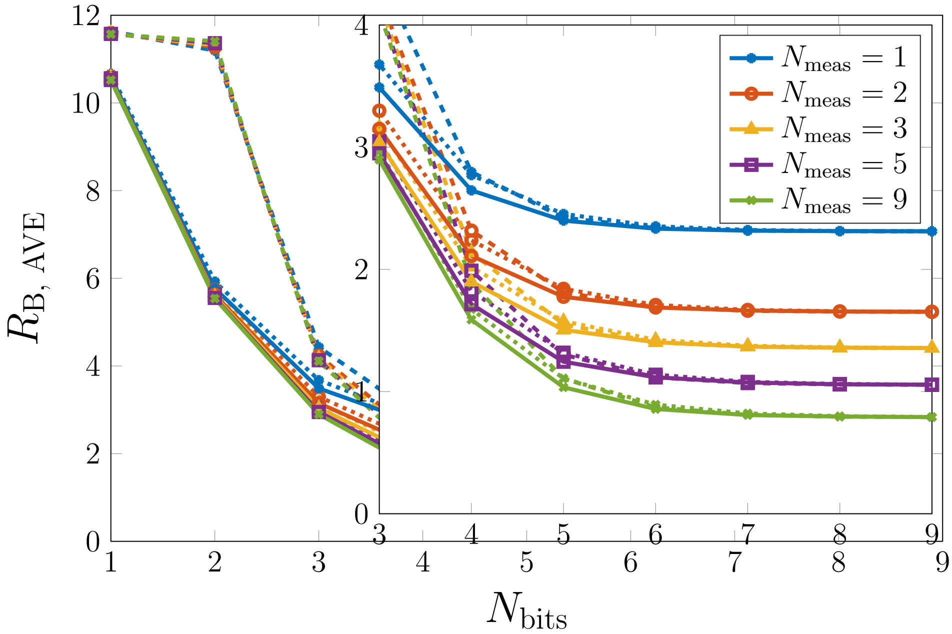

Figure 2.8: Bayes risk for the MAVE estimator with $p1_W$ (Linear quantizer: -- Entropy based: $\cdots$ Mutual info. based: —)

is the Bayes risk $R_{B,\ AVE}$. For an increasing N_{meas} and a fixed N_{bits}, $R_{B,\ AVE}$ decreases asymptotically, so that for a higher N_{meas} the improvement on $R_{B,\ AVE}$ decreases. This is also true for increasing N_{bits} with fixed N_{meas}. Since the Bayes risk R_B is directly related to the mutual information I, the asymptotic behaviour for larger N_{bits} can also be observed here.

The gain of the mutual information based quantizer to the linear quantizer is high, if the number of quantization steps Q_{quant} is low, but the results get closer, when Q_{quant} gets larger. This is because the quantization interval size shrinks for growing Q_{quant}, so that the increase of the Bayes risk of the suboptimal quantization scheme diminishes, just as the mutual information from the previous set of experiments suggests. The knee in the curves for the linear quantization scheme is also visible here, because the lower mutual information leads to a higher estimation error.

Fig. 2.9 shows a similar behaviour for $p2_W$, but the Bayes risks R_B of the mutual information based scheme and the linear scheme are closer, while the entropy based scheme achieves the same results as the mutual information based scheme. This is in accordance to Fig. 2.6, where both schemes are shown to result in similar mutual information I. The higher overall level of mutual information in Fig. 2.6 does not lead to a lower Bayes risk in Fig. 2.9, since the pdfs are different and have different differential entropies. Therefore, there is no direct connection between the mutual information and the resulting Bayes risk.

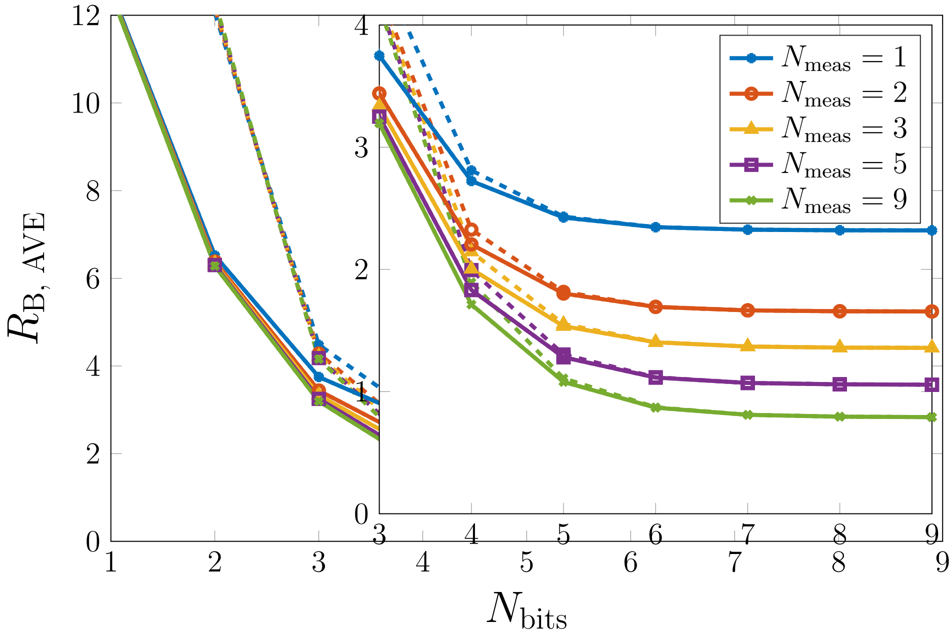

Figure 2.9: Bayes risk for the MAVE estimator with $p2_W$ (Linear quantizer: -- Entropy based: $\cdots$ Mutual info. based: —)

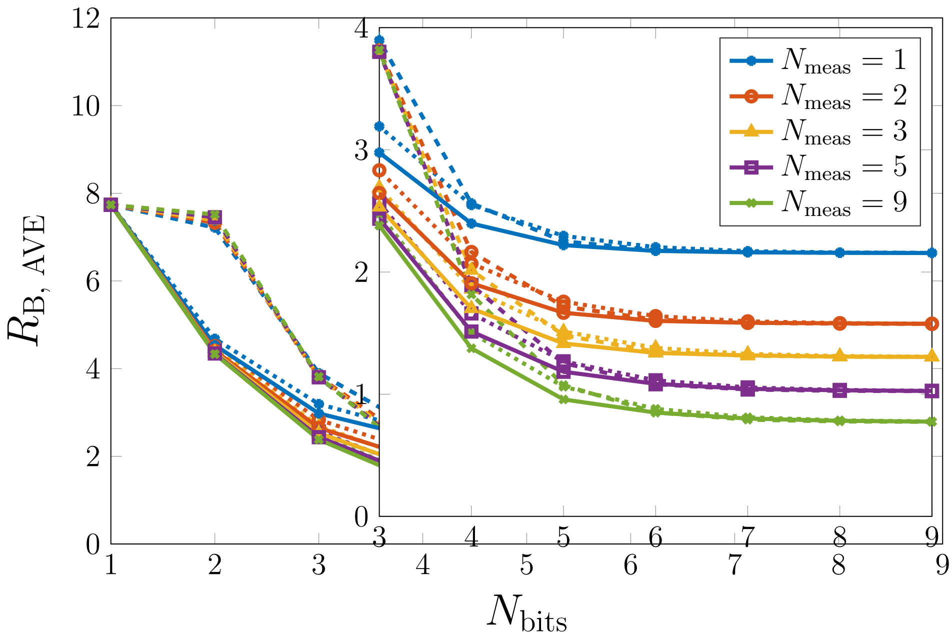

Figure 2.10: Bayes risk for the MAVE estimator with $p3_W$ (Linear quantizer: -- Entropy based: $\cdots$ Mutual info. based: —)

Fig. 2.10 with the distribution $p3_W$ in turn shows results similar to Fig. 2.8, but the difference between the adaptive entropy and mutual information based schemes to the non-adaptive linear quantizer is, like in Fig. 2.7, the largest of all three considered pdfs.

As an exception, for $4 \leq N_{\text{bits}} \leq 6$, the linear scheme slightly outperforms the entropy based one. Still, the mutual information based scheme yields the lowest Bayes risk for all N_{bits}.

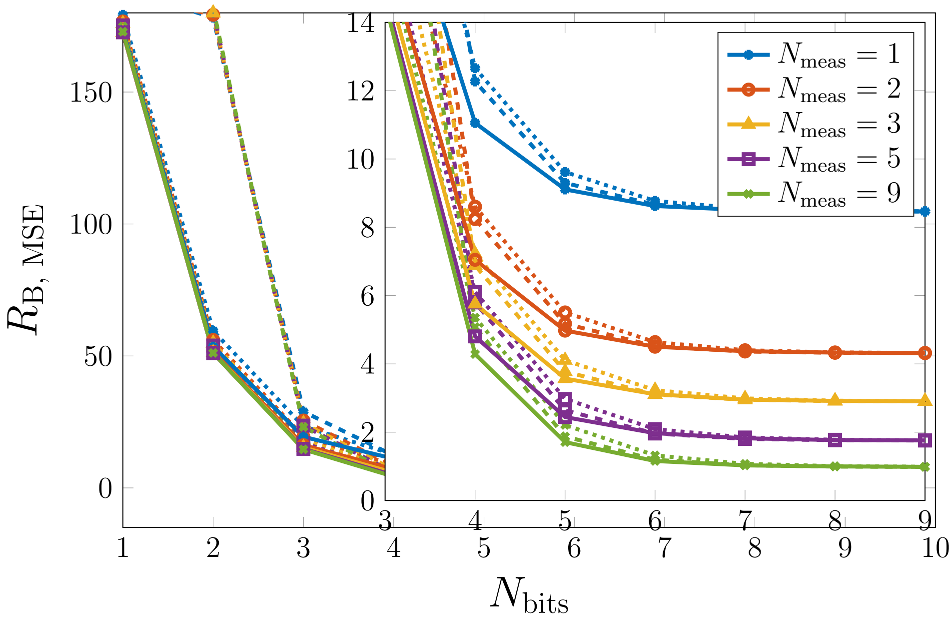

Figure 2.11: Bayes risk for the MMSE estimator with $p1_W$ (Linear quantizer: -- Entropy based: $\cdots$ Mutual info. based: —)

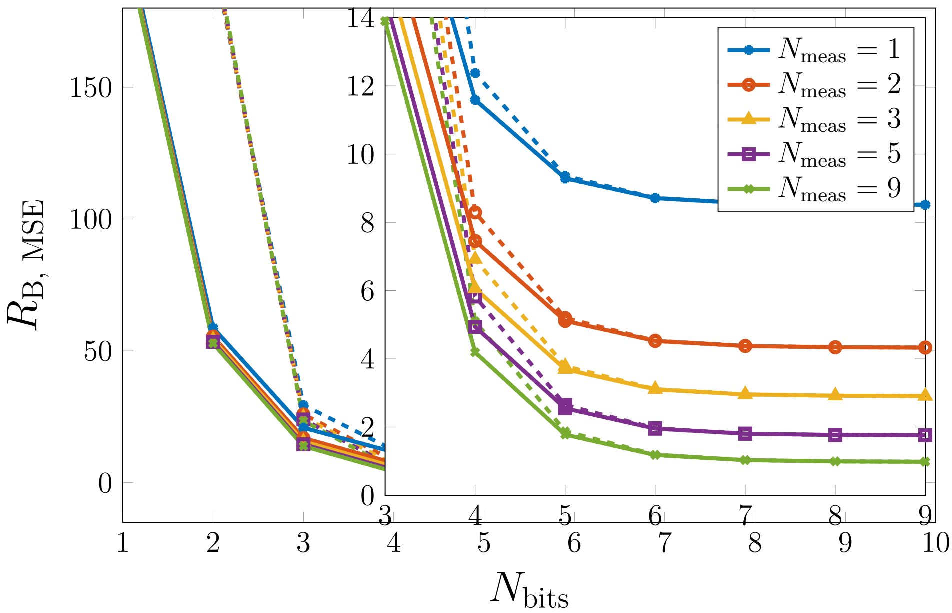

Figure 2.12: Bayes risk for the MMSE estimator with $p2_W$ (Linear quantizer: -- Entropy based: $\cdots$ Mutual info. based: —)

For the MMSE estimator in Fig. 2.11–Fig. 2.13, a behaviour similar to the MAVE

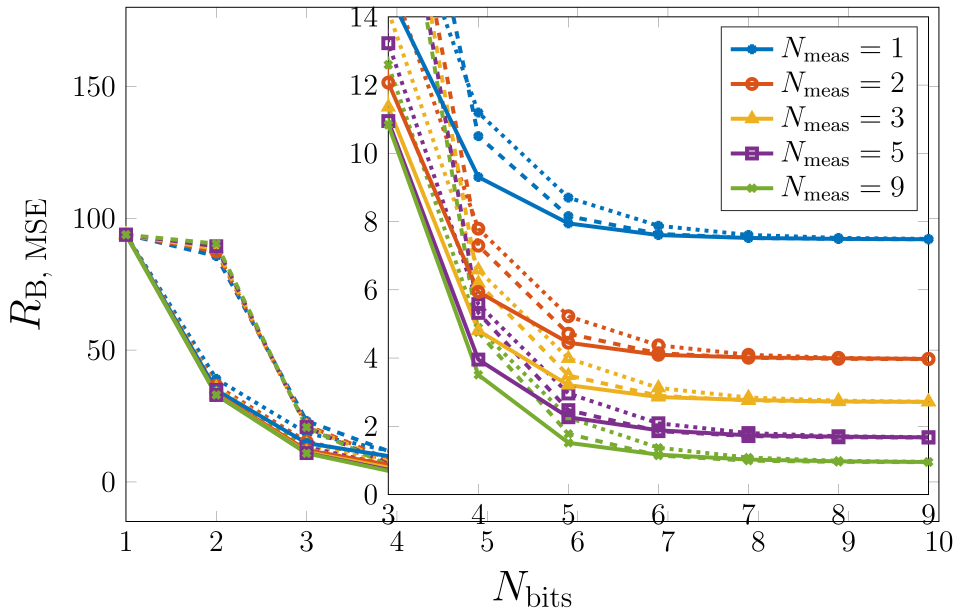

Figure 2.13: Bayes risk for the MMSE estimator with $p3_W$ (Linear quantizer: -- Entropy based: ··· Mutual info. based: —)

estimator is observed. Please note that $R_{\text{B, MSE}}$ is quadratic, since the MMSE estimator is optimal for the squared error, which leads to the larger magnitudes of Bayes risk in the plots for MMSE.

More measurements lead to a stronger reduction of $R_{\text{B, MSE}}$ than for the MAVE estimator, because the quadratic nature of the MMSE estimator penalizes the larger errors for low N_{bits} or N_{meas} more than the linear MAVE estimator. The knee in the linear quantizer curves, which could already be observed in the plots of the mutual information I in Fig. 2.5–Fig. 2.7 and the plots of the Bayes risk $R_{\text{B, AVE}}$ of the MAVE estimator in Fig. 2.8–Fig. 2.10, is also visible here. The effect of the different differential entropies of the three pdfs is apparent especially for $N_{\text{bits}} < 3$. A uniform distribution like $p2_W$ has the maximum differential entropy, since all possible values $w_{\min} \leq w \leq w_{\max}$ have equal probability. This results in the highest Bayes risk for all quantization schemes when using this pdf.

Fig. 2.14 shows the Bayes risk for a given maximum time of $T_{\max} = 13$ and $\gamma_{\text{T}} = 0.25$ as the ordinates. The abscissae are the used transmission times T_{tx}. The experiments are done for the distribution $p1_W$. The available time is completely used, i.e.

$$T_{\text{acq}} + T_{\text{tx}} = T_{\max}, \tag{2.48}$$

so increasing T_{tx} decreases T_{acq} and vice versa. The results show the trade-off between T_{acq} and T_{tx} with minimum R_{B} can be found. This result suggests, that there is an optimum ratio of time spent for measuring on the one hand and transmission on the other hand. A similar relation can be found for limited energy. In the next section the behaviour of this optimum for different constraints is investigated further.

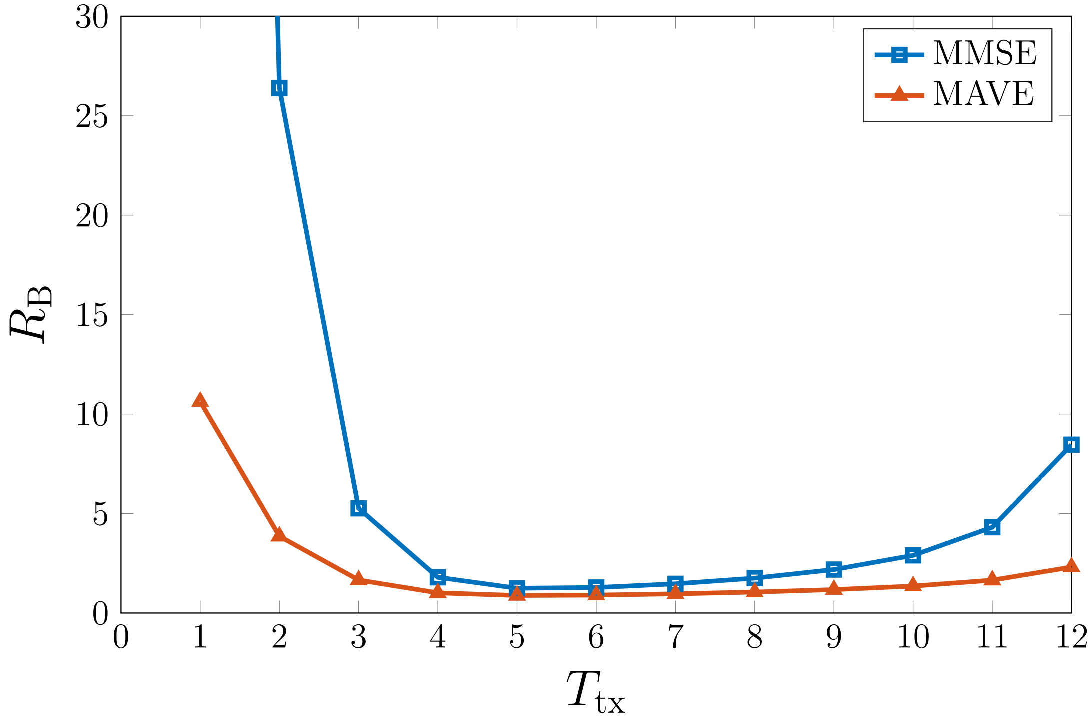

Figure 2.14: Trade-off T_{tx} vs. T_{acq} with $T_{max} = 13$ and $\gamma_T = 0.25$ [KASK19]

2.5.4 Optimum ratio of number of measurements and quantization steps

To investigate the influence of time constraints on the optimal selection of N_{meas} and N_{bits} or, likewise, energy constraints on N_{meas} and Q_{quant}, the minimum $R_{B,\,MSE}$ and $R_{B,\,AVE}$ for different constraint sets is investigated. The experiments are done for the distribution $p1_W$.

First, the influence of changing constraints T_{max} and E_{max} is considered. For a fixed ratio γ_T and γ_E, respectively, the optimum allocation of time or energy, respectively, was found. From the previous results for the Bayes risk, it is clearly visible that a finer quantization and more measurements will always lead to a lower Bayes risk. This result suggests that the optimum resource allocation should always use all available time or energy resources.

To find the optimum allocation for a given limit, all possible combinations of N_{bits} and N_{meas} which fully use T_{max} in the time limited case, and all possible combinations of Q_{quant} and N_{meas} which fully use E_{max} in the energy limited case, are considered. The resulting optimal values for T_{tx} and T_{acq} as the ordinates for a given T_{max} as the respective abscissae are shown in Fig. 2.15 and for E_{tx} and E_{acq} as the ordinates for a given T_{max} as the respective abscissae in Fig. 2.16.

In both cases, for $T_{\mathrm{max}} = 1$ or $E_{\mathrm{max}} = 1$, it is only possible to carry out one measurement, but not to transmit, so the estimate is based solely on $p_W(w)$. For the time constrained case, $\gamma_{\mathrm{T}} = 2$ and $T_{\mathrm{max}} = 1$ results in only 0.5 bits available to quantize the value, so only a single value can be transmitted and the estimation is again carried out solely based on the knowledge of $p_W(w)$.

Both, in the time limited and in the energy limited case, the time or energy spent for measuring, T_{acq} or E_{acq}, and transmitting, T_{tx} or E_{tx}, increase. In the time limited case, the time spent for transmission increases faster with increasing T_{max}, because one additional bit always doubles the number of available quantization steps, while the corresponding two additional measurements, which could be done in the same time, give less and less improvement for higher T_{max}.

In the energy limited case, the growth is almost proportional, because now there is just a linear instead of an exponential relation between E_{quant} and Q_{quant}. Since the improvement for each additional quantization step and measurement reduces, they increase alternatingly. The two estimators only show minor differences in the optimal resource allocation.

Now, the influence of changing ratio γ_{T} is investigated. The available time is set to $T_{\mathrm{max}} = 13$, the ratio γ_{T} is varied between 0.75 and 4. For rising γ_{T}, this makes the transmission relatively more time-consuming. The results are shown in Fig. 2.17. The longer time per bit results in more time spent for transmitting data than measuring. This leads to a growing T_{tx}. As shown in the previous results, the overall R_{B} increases, since the duration of a single measurement and the available time T_{max} is held constant, while less bits can be transmitted. Decreasing N_{bits} and increasing N_{meas} is generally not an option, because decreasing N_{bits} by one halves Q_{quant}.

For the energy constrained case, a similar result is shown in Fig. 2.18. Here, the linear relation between Q_{quant} and N_{meas} results in a more constant ratio of E_{tx} and E_{acq}, because, in contrast to the previous case, an additional measurement can often compensate for a smaller Q_{quant}.

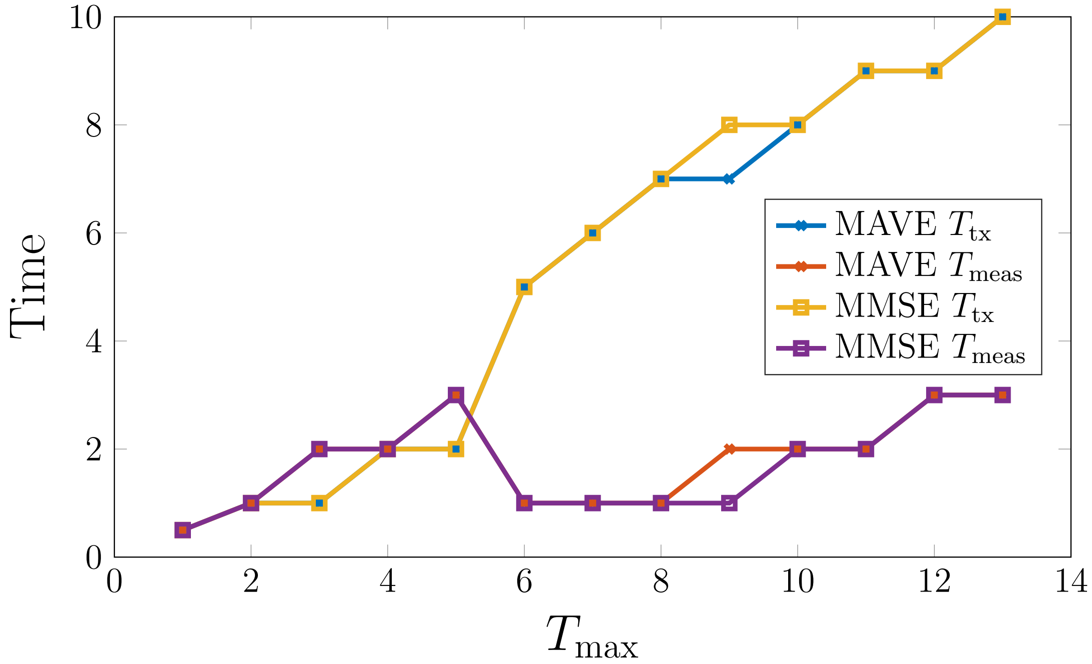

Figure 2.15: Constrained time usage, $\gamma_{\mathrm{T}} = 2$ [KASK19]

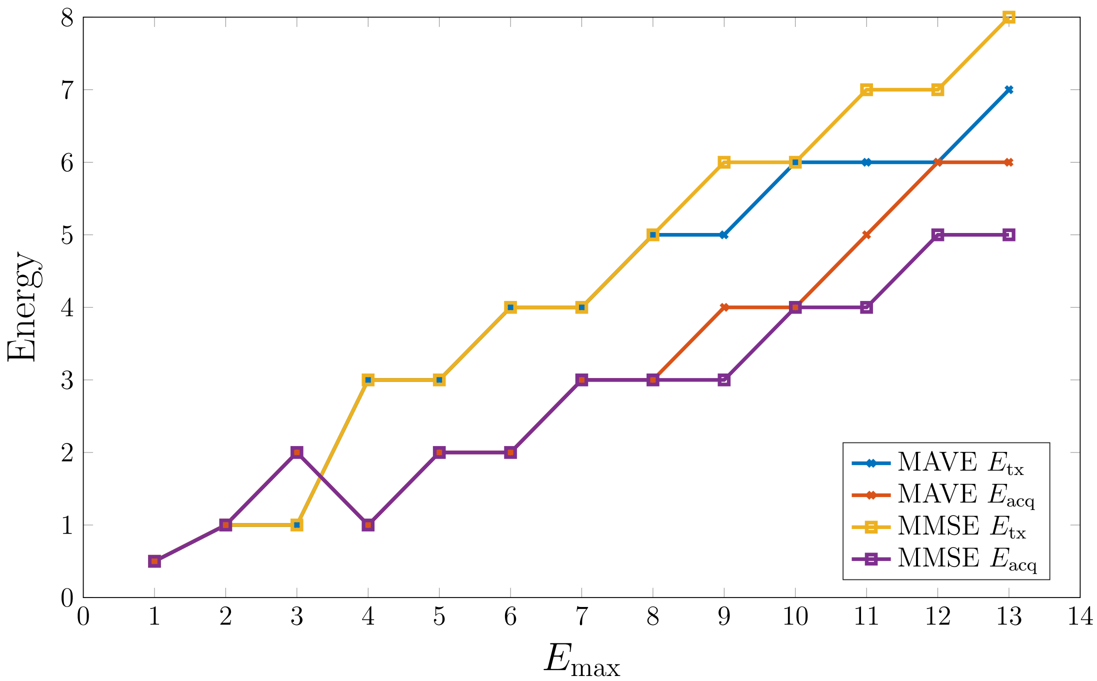

Figure 2.16: Constrained time usage, $\gamma_{\mathrm{E}} = 0.25$ [KASK19]

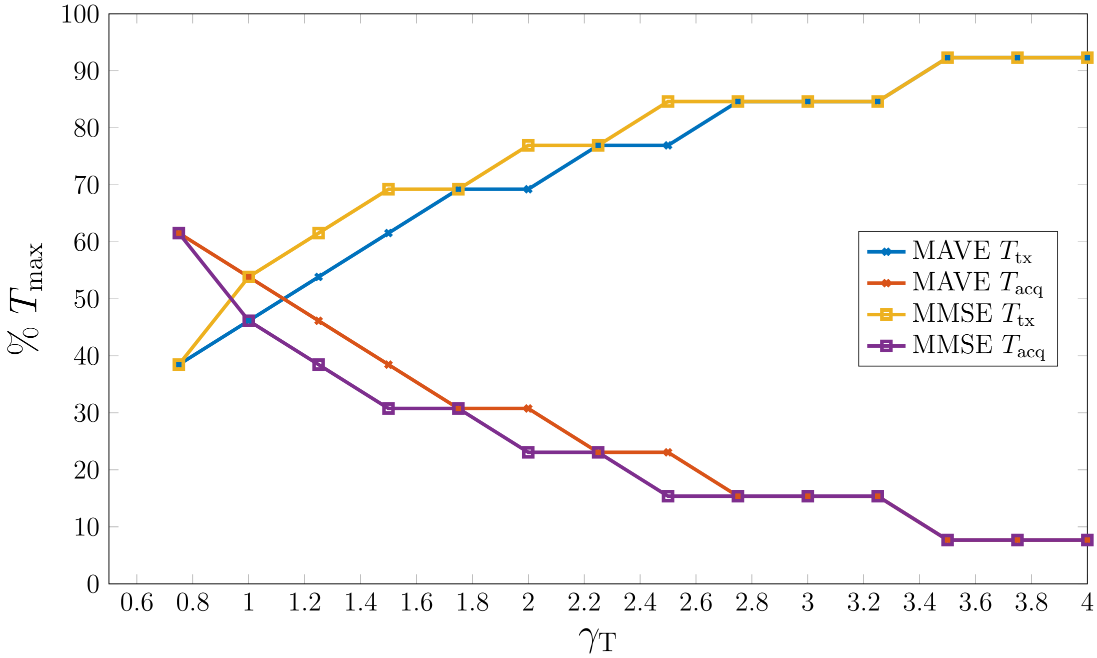

Figure 2.17: Variable time usage ratio, $T_{\max} = 13$ [KASK19]

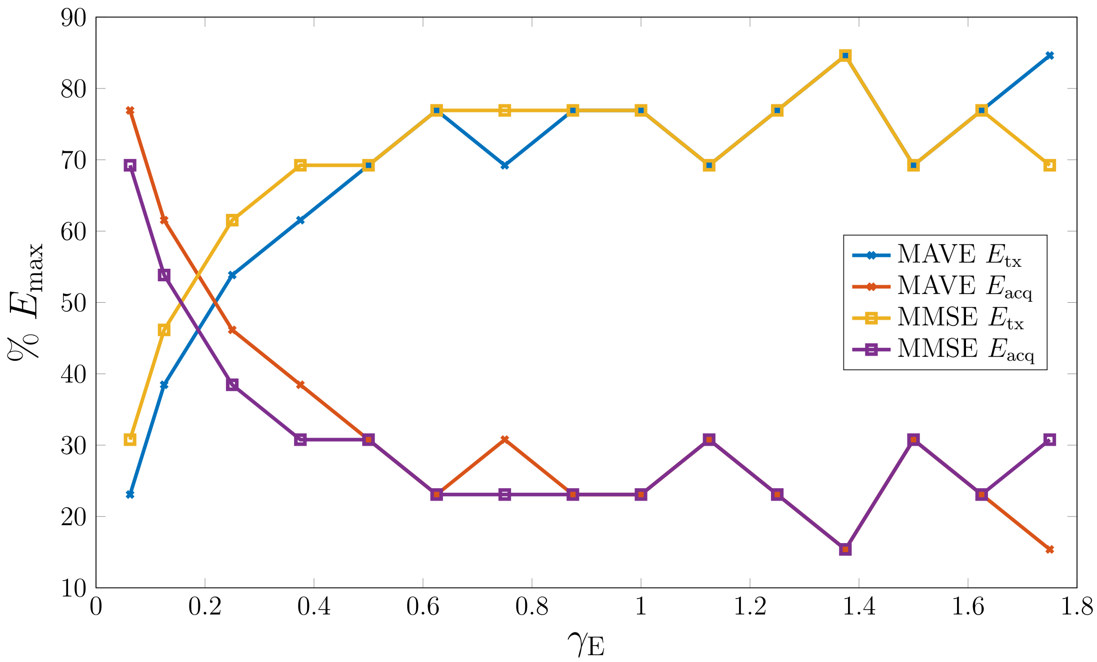

Figure 2.18: Variable energy usage ratio, $E_{\max} = 13$ [KASK19]

2.6 Conclusion

In this chapter, the estimation of a parameter following a given pdf using quantized measurements transmitted over a wireless channel was considered. Multiple noisy measurements of a parameter can be taken sequentially and are aggregated afterwards. The aggregated value is quantized and transmitted to the receiver, which executes the estimation of the measured parameter.

Three different quantization schemes were developed and compared with respect to the conveyed mutual information between the quantity of interest and the resulting data symbols, as well as the resulting Bayes risk after estimation based in these data symbols. We have shown, that the lowest Bayes risk is achieved by the scheme which is created by maximizing the mutual information. A scheme, which aims at making all data symbols equally likely, often delivers similar results, but reduces the complexity of calculating the quantization steps.

Besides the quantization scheme, the two main parameters to tune are the accuracy of the measurement and the precision of the quantization. In the used model, multiple measurements can improve the measurement accuracy, while the precision of the quantization is determined by the number of quantization steps. If the required time and energy resources for measuring and transmitting are comparable and limited, a trade-off between measurement accuracy and quantization precision can be made to reduce the overall estimation error. This trade-off is highly dependent on the resource constraints and the ratio of the energy or time resources used for measurement and quantization improvement. For an optimum quantization the distribution of the sensor value and of the sensor noise must be known, otherwise the linear quantization scheme with equidistant intervals has to be used.

Chapter 3

Scheduling of Sensor Readings with Constrained Communication

Several parts of the content of this section have been originally published by the author of this thesis in [KK20]. This paragraph shall illustrate the previous work from [KK20], the relation to and the additional work presented in this chapter. The model, problem formulation and the scheduling algorithm were already published. Results for a single configuration of subsystem coefficients were also shown previously in [KK20]. The simulations in the numerical part are extended to show results for different system characteristics and also examine the influence of the ratio of different subsystem characteristics on the overall system.

3.1 Introduction

The performance of control processes is not only determined by the accuracy of the plant state measurement, which was investigated in the previous chapter. Another important aspect is also the timely reception of the sensor values. Depending on the system, which is to be controlled, "timely" can result in vastly different requirements.

There are highly dynamic processes like keeping the lane in autonomous driving, which requires fast reactions to getting away from the center. The reaction times should be as fast as 0.1 s, in such lane keeping systems, as shown in [BDH+20]. Then there are processes, which require only a medium response time, i.e. keeping the distance in autonomous driving or correcting the altitude of a drone. The vehicle distance control is needed for platooning scenarios as described in [PYZ+20]. The response time are in the range of about 5 s seconds here. Similarly, when talking about multicopter drones, the correction times are in the range of a 2–3 s seconds, as shown in [XMH19]. These reactions result in a constant stream of input commands to the plant to keep it in a safe operating region. On the other hand, there are processes with low dynamics, which even return to an equilibrium, if no external input is given. An example for such a process is a crane, where the load attached will swing, but eventually remain static in the center. By a sophisticated control scheme, however, the amplitude of the oscillations and/or the time until it is virtually stationary can be reduced.

The dynamics of the process determine the control law in the controller, but also the uncertainties about the plant state evolution have to be considered. Such uncertainties are caused on the one hand by random influences on the process and measurement inaccuracies, but on the other hand also properties of the process not considered in the control model contribute to this uncertainty. If no uncertainty was there, the initial state of the plant is perfectly known and all inputs for the whole future are known, the plant state could be precalculated for the whole future. Since this is not possible, the plant state has to be sensed from time to time to correct the plant state estimate at the controller.

In this chapter, we will now deal with the questions No. 3 and No. 4 from Section 1.2. We consider multiple discrete time linear subsystems with different dynamics and a central scheduler, which can request sensor readings from each subsystem. The communication resources are limited, so it is not possible to get a sensor reading from every sensor in every time slot, the central scheduler rather has to select a subset of subsystems to request sensor readings from in every time slot. In this chapter we will assume fine quantization and, thus, treat the errors introduced by quantization as negligible. Hence, only the system noise is assumed to lead to uncertainty about the system state.

In this chapter, the focus is on the deterministic case without packet loss. There are multiple control loops and a central scheduler, which is aware of the subsystems characteristics and schedules the transmissions from sensors and to controllers. The optimality of a fixed update frequency scheme for a minimum mean-square estimation error at the controller is shown. For derivation of the frequencies and the actual scheduling, we propose a two-step approach. In the first step, the individual update frequency for each subsystem is determined based on the system noise power and the subsystem dynamics. We show in Section 3.4, that this resource allocation problem is in fact a convex problem, which can be solved existing optimization frameworks like [GB14]. For the second step, we developed an algorithm in Section 3.5, which schedules the available communication resources in each time step fairly to the subsystems according to the derived update frequencies from the first step. The main advantage of this two-step approach is the reduced effort during runtime, because only the second step has to be carried out during runtime. Finally, we show the advantage of our approach with numerical results in Section 3.6.

3.2 Related Work

This section discusses related work and will give an overview of the state of the art. The basic requirement is to keep the plant state information fresh at the controller, as well as transmit the calculated control commands to the plants. Since these requirements are strongly related, papers for each of the problems as well as works considering both are discussed.

A very general solution to minimize the overall time passed since the last update of the sensor value at the controller is investigated in [HYE16]. Multiple sensors, acting as sources of information, generate data packets for their respective destinations and store them at queues at the transmitters. Transmitters can serve one or multiple sources. The actual transmission of the data packets by the transmitters is done over a wireless broadcast channel to the respective receivers. The scheduling of these transmissions is now optimized centrally for interference free reception while minimizing the ages of the last received update for all sensor – receiver pairs. This ensures a timely update of the state information for all subsystems, but does not consider the individual subsystem dynamics.

If the system has a central scheduler to schedule the updates, which is aware of the dynamics, the scheduling decision can be further optimized. This system layout is investigated in [ZCWF18], which was already mentioned in Section 2.1. Here, the central scheduler transmits the sensor values to the subsystems according to a precalculated schedule, which is based on the uncertainty about the current subsystem states. In each time step, a fixed number of transmission slots is available. The required optimization is a mixed-integer problem, which only gives a schedule for a fixed time horizon and is considerably hard because of its nonconvex structure.

In [MGW+19], a model similar to [ZCWF18] with multiple independent subsystems is used. Unlike [ZCWF18], the wireless connection is not between sensor and controller, but rather between controller and plant. The available resources elements are defined as the time slots in the IEEE 802.15.4 standard, [IEE20]. The scheduling also considers packet loss and varying link qualities by a prediction model. The loss probability can be reduced by using multiple resource elements to transmit the same data. Due to the lossy links, the optimization has to be redone for each time step, taking the predicted link qualities into account.

The restriction of the time limited schedule is relieved in [AVK+19], which also uses a central entity to schedule the updates. Similar to [ZCWF18], the subsystems all consist of a sensor, a plant and a controller. The sensor cannot directly transmit values to the controller, but rather the central entity can request sensor readings and transmit them to the respective controllers. The communication resources are limited, so it is not possible to always update all the controllers. For the scheduling decision the time passed since the last transmission of the sensor value for a subsystem is used to find the most outdated values. This approach is compared to an improved scheduling based on the error covariance of the current state estimate at the controller, Value of Information (VoI). The variance-based approach considers different dynamics of the subsystems.

In [AVK20], the same model as in [AVK+19] is used. Based on the age of the last successful transmission, the problem is modeled as a Markov Decision Process (MDP) with the times passed since the last transmission for each subsystem as the multidimensional state and the scheduling decision for the current time step as action. To limit the infinite state space, the number of possible states is capped by limiting the maximum considered ages. The model also incorporates packet-loss probabilities for the communication links, which results then in a deterministic scheduling policy.

The major drawback of [ZCWF18] and [MGW+19] is the time limited schedule, so only a limited time horizon can be used for scheduling optimization. In [AVK20] the time horizon is infinite, but the state space is artificially limited to render the problem tractable. This leads to research question No. 3 from Section 1.2, which is asks for a general rule for an update rate based on the subsystem characteristics. During research, the problem of integer effects on the performance became apparent, since the general rule found for state update transmissions can give ambiguous results, if more subsystems are equally eligible for transmission. Therefore, an algorithm to equalize these effects on all subsystems and reduce the overall fluctuation of error covariances was developed, tackling question No. 4.

3.3 System Model

3.3.1 Control System Model

The overall control system consists of N_{sys} independent subsystems as shown in Fig. 3.1. Each subsystem i consists of a plant $\mathcal{P}_i$, a controller $\mathcal{C}_i$ and a sensor $\mathcal{S}_i$. A central scheduler polls measurements y_i from the sensor $\mathcal{S}_i$ (uplink) and forwards them to the controller $\mathcal{S}_i$ (downlink). Each subsystem $i \in 1, \dots, N_{\text{sys}}$ is modeled as a discrete time linear system with a scalar state $x_i(k)$ at time instant k, a system coefficient a_i, measurement noise $w_i(k)$ and the control variable is $u_i(k)$. Each subsystem i follows a linear system equation

$$x_i(k+1) = a_i x_i(k) + u_i(k) + w_i(k). \tag{3.1}$$

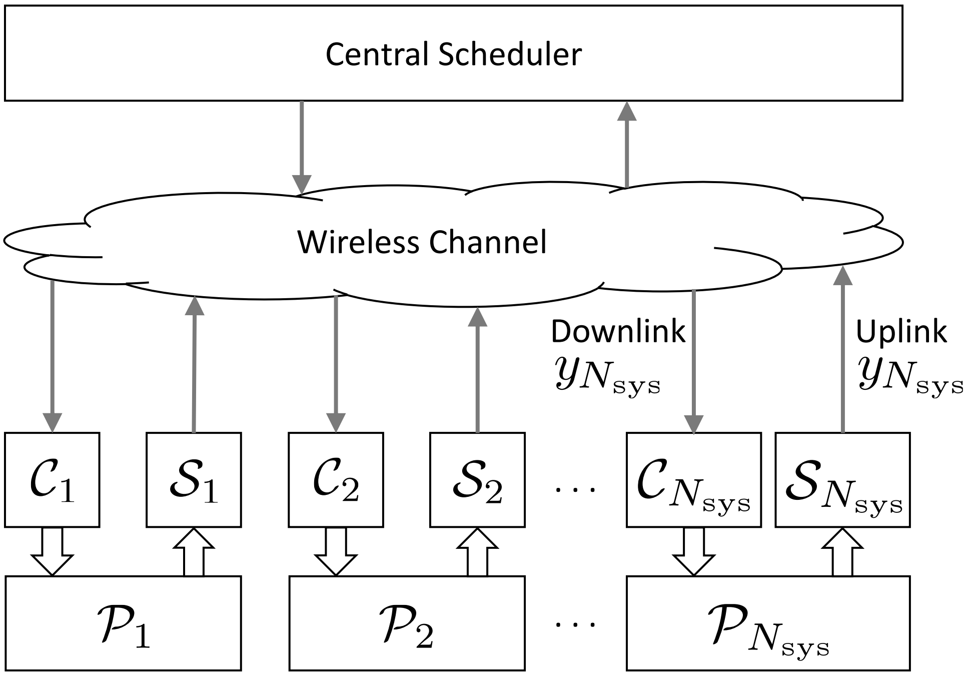

Figure 3.1: System model of the WNCS with wireless sensor readings

The system noise $w_i(k)$ is assumed to be zero-mean Gaussian i.i.d. for all times k and all systems i with variance W_i. The subsystems can be observed according to the observation equation

$$y_i(t) = x_i(k) + v_i(t) \tag{3.2}$$

with the measurement noise $v_i(k) \sim \mathcal{N}(0, V_i)$. Since the system is linear with Gaussian noise, the Kalman filter [Kal60] gives the MMSE estimate $\hat{x}$ of the system state x based on the observation y. The estimation error is $e_i(k) = x_i(k) - \hat{x}_i(k)$. The control variable u_i is then calculated according to a deadbeat law, i.e. $u_i(k) = -a_i\hat{x}_i(k)$, to achieve the control goal $x_i = 0$ for all subsystems. A cost function

$$J(k) := |\boldsymbol{x}(k)| \tag{3.3}$$

for deviating from this goal is assumed. The limited communication resources only allow for sensor readings of scheduled time slots. The Kalman filter is modified to predict the intermediate values, if no sensor reading is available for a subsystem in a given time slot. The availability of a new value is described by a binary scheduling decision variable $\pi_i(k)$, which is set to 1, if the system i is scheduled in time slot k and to 0 otherwise. The expression $x(a|b)$ is used to denote quantity x at time instant a with the knowledge from time instant b with $b \leq a$. The modified Kalman filter with the estimation error covariance P_i of e_i, and the Kalman gain g_i is then given as

$$\hat{x}_i(k|k-1) = a_i\hat{x}_i(k-1|k-1) + u_i(k-1) \tag{3.4}$$

$$P_i(k|k-1) = a_i^2 P_i(k-1) + W_i \tag{3.5}$$

$$g_i(k) = \pi_i(k)\frac{P_i(k|k-1)}{P_i(k|k-1) + V_i} \tag{3.6}$$

$$\hat{x}_i(k|k) = \hat{x}_i(k|k-1) + g_i(k)\left(y_i(k) - \hat{x}_i(k)\right) \tag{3.7}$$

$$P_i(k|k) = \left(1 - g_i(k)\right) P_i(k|k-1) \tag{3.8}$$

Because the instantaneous value $x_i(k)$ is not known at the central scheduler, π_i can only be based on P_i. The scheduling should minimize the uncertainty about the system state $x_i(k)$ at the controller, which is induced by the system noise $w_i(k)$ and results in the estimation error $e_i(k)$, which is also Gaussian distributed with zero-mean and variance P_i, since (3.4) – (3.8) resemble a linear time-invariant (LTI) system. Gaussian noise filtered with a LTI filter is still Gaussian [PP02]. The uncertainty about the subsystem state $x_i(k)$ results in the system error variance P_i. So minimizing the uncertainty about the subsystem states corresponds to minimizing $\sum_{i=1}^{N_{\text{sys}}} P_i(k) \quad \forall k$. The transmission of the current observation $y_i(k)$ reduces the uncertainty. The variances P_i do not depend directly on the unknown subsystem states x_i and, thus, it is possible to precalculate them based on (3.4) - (3.8). Henceforth, the measurement noise is assumed to be negligible, i.e. $V_i = 0$. This results in $g_i(t)$ being either 1 or 0 and the system state $x_i(k)$ as well as the error covariance P_i is set to zero every time a transmission to the controller is scheduled. After this, the system noise $w_i(k)$ adds uncertainty in every system time step k. If the system has been scheduled last in time step $l \leq k$, the variance P_i is

$$P_i = \text{Var}(e_{i,k}) = \begin{cases} W_i(k-l) & \text{if } |a_i| = 1 \\ W_i\frac{1-a_i^{2(k-l)}}{1-a_i^2} & \text{else.} \end{cases} \tag{3.9}$$

Since only the time duration $d = k - l$ between consecutive transmissions at time steps k and l determines the error covariance P_i, a function $f_i(d)$ to calculate the sum of variances P_i after d timeslots without transmission can be written, as

$$f_i(d) := \sum_{m=l}^{k} P_i(m) = \begin{cases} W_i\frac{d(d+1)}{2} & \text{if } a_i = 1 \\ W_i\frac{1}{1-a_i^2}\left(1 - \frac{1-a^{2d}}{1-a_i^2}\right) & \text{else.} \end{cases} \tag{3.10}$$

The control cost function (3.3) can be minimized for each time step individually to minimize the overall cost. Because of the deadbeat control, the state x_i is always zero-mean Gaussian distributed. The remaining deviation from the control goal $x_i = 0$ after application of the control variable u_i derived from $\hat{x}_i$ is almost equal to the estimation error e_i, i.e. $\text{Var}(x_i(k)) \approx \text{Var}(e_i(k)) = P_i(k)$. The approximation comes from the fact that the measurement is received with delay due to latency induced by the communication. It is possible that more recent information is available at the central scheduler, reducing the estimation error $e_i(k)$ there, but it could not be used for better control, since it has not been transmitted to the controller yet. The expected value of the cost function J can then be calculated from the standard normal distribution function as

$$\mathbb{E}\{|x_i(k)|\} = \sqrt{\frac{2}{\pi}}\sqrt{\mathrm{Var}(x_i(k))}. \tag{3.11}$$

The overall control objective with respect to the scheduling is

$$\arg\min_{\pi} \textstyle\sum_{i=1}^{N_{\text{sys}}} \sum_{k=1}^{T_{\text{sim}}} \mathrm{Var}(x_i(k)). \tag{3.12}$$

3.3.2 Communication System Model

The communication time slots, denoted by t, are shorter than the control time slots, denoted by k. In each of the control system time slots k for subsystem i, T_i^s communication time slots t take place. Additionally, a sampling offset T_i^0 between the subsystems is used. The relation between t and k is then given like in [AVK+19] as $k_i(t) = \lfloor \frac{t-T_i^0}{T_i^s} \rfloor$. The communication system is based on the scheduling decisions π_i made by a central scheduler and is used to transmit measurements from sensors to controllers, which are directly attached to the plants. Transmission takes place in a packet based fashion with equally sized packets. In one communication time step t, R_{UL} packets can be transmitted in the uplink and R_{DL} packets in the downlink. The system is assumed to have no packet loss. The scheduling decision is modeled by the variable $\boldsymbol{\pi}_{\text{UL}(t)} \in \mathbb{R}^{N_{\text{sys}}\times 1}$ and $\boldsymbol{\pi}_{\text{DL}(t)} \in \mathbb{R}^{N_{\text{sys}}\times 1}$ for uplink and downlink, respectively.

3.4 Problem Formulation

3.4.1 Influence of Long-Term evaluation of subsystem variances

In [AVK+19], the error covariance P_i is used as VoI to derive the scheduling π_i in a greedy fashion. This minimizes the uncertainty about subsystem i in the current time step, but (3.12) rather asks for the minimization of the overall sum of uncertainty. This means, a greedy scheduling decision might be suboptimal.

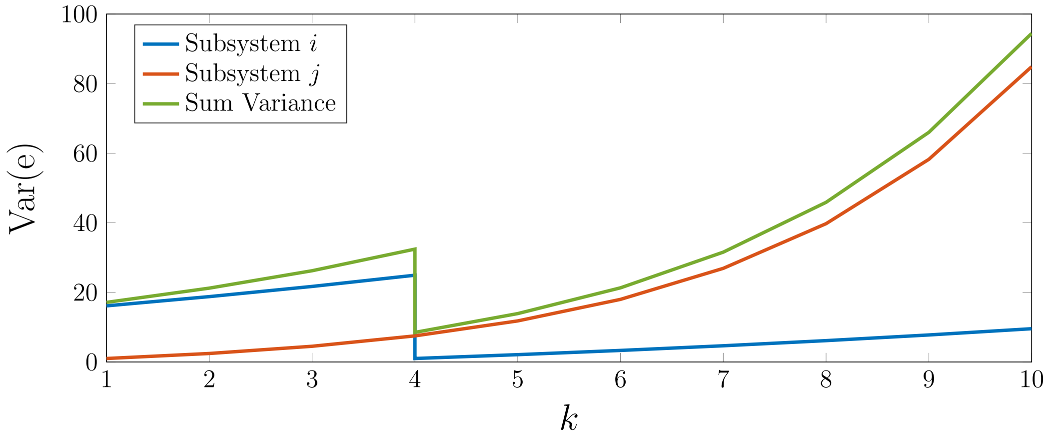

Figure 3.2: Greedy Scheduling from [AVK+19]

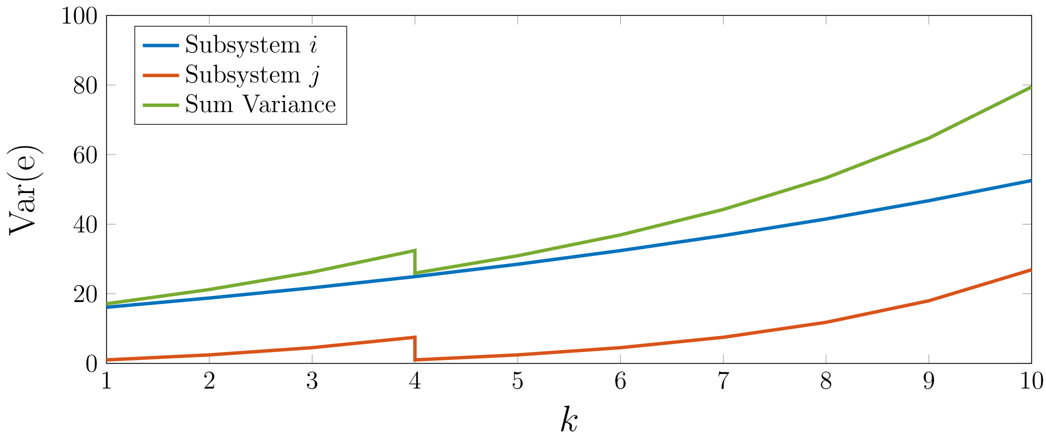

Figure 3.3: Scheduling according to (3.12)

Figure 3.4: Scheduling examples

Fig. 3.4 illustrates the problem: For two subsystems i and j, it is assumed that in time step $k = 4$ one communication resource is available and $T_i^s = T_j^s = 1$. For subsystem i with small $a_i = 1.05$, a high P_i can accumulate over time, so this system will be scheduled in time step $k = 4$ as shown in Fig. 3.2, whereas a second system j with system coefficient $a_j = 1.2$ is not scheduled. According to a_i and a_j, the variances P_i and P_j grow in the next time steps. In Fig. 3.3, another scheduling possibility is shown. The sum of P_i and P_j does not grow as fast as in Fig. 3.2, so the scheduling in Fig. 3.3 outperforms the greedy one over time.

The functions f_i are used to consider this in the scheduling. Their exponential shape makes a regular scheduling of each subsystem desirable. The individual slope of f_i suggests an update rate depending on the system constant a_i, the system noise covariance W_i, and the available communication resources. Stable systems with $|a_i| < 1$ get a low or zero rate, while system with large a_i are scheduled more often. When looking at the scheduling based on the current variance, a regular scheme for the scheduling of the subsystems becomes apparent. The following proof shows the optimality of such a scheme.

3.4.2 Optimality of a Regular Update Scheme

To show the optimality of a regular update scheme, we consider a single subsystem i with system constant $a_i \neq 1$, a system noise variance of W_i, and a finite operation time horizon T_{sim}, i.e. $k = 1, \dots, T_{\mathrm{sim}}$. During the operation time, $N+1$ sensor values are transmitted, where the first and last transmission take place at $k = 0$ and $k = T_{\mathrm{sim}}$, respectively. The time durations between two consecutive transmissions are denoted by $d_{i,1}, \dots, d_{i,N}$ with $d_{i,n} \geq 0, n = 1, \dots, N$. Now, the scheduling minimizing the sum of error variances over k up to T_{sim} is to be found. Using (3.10), this leads to the optimization problem

$$\arg\min_{\boldsymbol{d}_i} \textstyle\sum_{n=1}^{N} f_i(d_{i,n}) \tag{3.13}$$

$$\text{s.t.} \quad \textstyle\sum_{n=1}^{N} d_{i,n} = T_{\mathrm{sim}} \tag{3.13a}$$

with the Langragian

$$L(\boldsymbol{d}, \mu) = \inf_{\mu} \sum_{n=1}^{N} \frac{W_i}{1 - a_i^2} \left(1 - \frac{1 - a_i^{2d_{i,n}}}{1 - a_i^2} \right) + \mu \left(d_{i,n} - T_{\mathrm{sim}} \right) \tag{3.14}$$

and its partial derivatives with respect to the d_i:

$$\frac{\partial L}{\partial d_{i,n}} = \frac{W_i}{(1 - a_i^2)^2} \left(-2 \log{(a_i)} a_i^{2d_{i,n}} \right) + \mu. \tag{3.15}$$

Since μ, W_i and a_i do not depend on n, they are equal for all partial derivatives of $L(\mu, \boldsymbol{d}_i)$. This results in equal $d_{i,n}$ to bring all components of the gradient to zero, i.e., equal durations between the transmissions for a single subsystem are optimal for reducing the sum of error variances over time, but are not necessarily integer multiples of the time slot duration. This is also applicable in case of multiple subsystems. Henceforth, scalar d_i will be used for the durations between consecutive transmissions of subsystem i.

3.4.3 The rate optimization problem

To find the optimum duration d_i for each subsystem, the optimization problem (3.13) is modified to use an update rate

$$r_i = \frac{1}{d_i}. \tag{3.16}$$

The results from [AVK+19] suggest a resource allocation, which is defined by $\min\{R_{\mathrm{DL}}, R_{\mathrm{UL}}\}$, since there is no gain in receiving information at the central scheduler, which cannot be forwarded to the systems. On the other hand, if no data was received from the sensors, the downlink capacity R_{DL} cannot be fully used. Since the data itself does not influence the scheduling decision, but rather the calculated error variance P_i, the scheduling for uplink and downlink, $\boldsymbol{\pi}_{\mathrm{UL}}$ and $\boldsymbol{\pi}_{\mathrm{DL}}$, is always the same. Then, the sum of the average variances per timeslot k is minimized:

$$\begin{aligned}
&\arg\min_{\boldsymbol{r}} \textstyle\sum_{i=1}^{N_{\mathrm{sys}}} r_i \max\left(W_i, f_i(1/r_i)\right) && (3.17)\\
&\text{s.t.} \quad 0 \leq r_i, i = 1, \ldots, N_{\mathrm{sys}} && (3.17a)\\
&\qquad\quad R = T^{\mathrm{s}} \min\{R_{\mathrm{UL}}, R_{\mathrm{DL}}\} && (3.17b)\\
&\qquad\quad R \geq \textstyle\sum_{i=1}^{N_{\mathrm{sys}}} r_i. && (3.17c)
\end{aligned}$$

The maximum in the objective (3.17) sets the lower bound of uncertainty to the system noise covariance W_i. Constraint (3.17a) ensures positive rates, while (3.17c) limits the sum rate to the available communication resources given by (3.17b).

3.5 Scheduling Algorithm

After calculating the rates, the actual scheduling is derived. As discussed in Section 3.4.3, the uplink and downlink scheduling is equal. The primary goal of the scheduling algorithm for the possibly non-integer duration values d_i is to bring the individual durations as close to the desired values as possible. As shown in Section 3.4.2, the duration between consecutive transmissions is more important than the average, so Algorithm 1 only considers the time since the last scheduling. The first transmission for all subsystems is assumed to take place at $t = 0$. Then, in each communication time slot t, Algorithm 1 is run to find the R_{UL} or R_{DL} subsystems, which have the longest time passed since their respective last transmission, relative to their desired duration d_i. The resulting vector $\boldsymbol{\pi}$ has elements for every subsystem.

Algorithm 1 Transmission scheduling

$R_{\text{remain}} \leftarrow \min\{R_{\text{DL}}, R_{\text{UL}}\}$ {Remaining resources for this time slot}
$\boldsymbol{\pi}(t) \leftarrow \mathbf{0}$
while $R_{\text{remain}} \neq 0$ **do**
  $\boldsymbol{\delta_n} \leftarrow \frac{1}{\delta_r}(t\mathbf{1} - \boldsymbol{t_l})$ {Normalized time since last TX}
  $n_i \leftarrow \arg\max_i \boldsymbol{\delta_n}(i)$
  $\pi_{n_i}(t) \leftarrow 1$ {Schedule subsystem with longest duration since last transmission}
  $\boldsymbol{t}_{\text{l},n_i} \leftarrow t$ {Save current transmission time step}
  $R_{\text{remain}} \leftarrow R_{\text{remain}} - 1$
end while
return $\boldsymbol{\pi}(t)${Return the scheduling for the current time slot}

3.6 Numerical Results

In this section, the performance of the proposed scheduling algorithm from Section 3.5 is evaluated. For this purpose, the influence of the number N_{sys} of subsystems, the difference of the subsystem coefficients a_i between each other, and the ratio of subsystems with high dynamics, i.e. larger a_i, to subsystems with lower dynamics, i.e. smaller a_i, shall be investigated. All experiment are conducted for $T_{\text{sim}} = 4000$ simulation time steps.

The influence of the number of subsystems with constant communication resources R_{UL} and R_{DL} was derived for a set of four system classes, each of them containing one quarter of the subsystems, with

$$a_i \in \{0.75, 1, 1.25, 1.5\}. \tag{3.18}$$

The noise variance for all subsystems i is set to $W_i = 1$. The number T_{i}^{s} of communication time slots per system time slot is set to $T_{\text{i}}^{\text{s}} = 10$.

First, we want to investigate, how this constraining environment influences the time duration d_i between two transmissions of the different subsystem classes. By increasing N_{sys} while keeping R_{UL} and R_{DL} constant, the resources per subsystem are reduced, which leads to increasing durations d_i between consecutive transmissions. Additionally, the share of resources changes between the subsystem classes; subsystems with smaller a_i get a smaller share. Fig. 3.5 shows the average duration between transmissions as the ordinates for different numbers of subsystems N_{sys} as the abscissae with resources $R_{\text{UL}} = R_{\text{DL}} = 1$. The optimization shown in (3.17) reduces the resources assigned to the systems with a low system constant, when N_{sys} increases, resulting in longer times between two subsequent transmissions. For the stable subsystems with $a_i = 0.75$, almost no transmissions are scheduled for $N_{\text{sys}} \geq 60$. For small r_i, due to (3.16), even

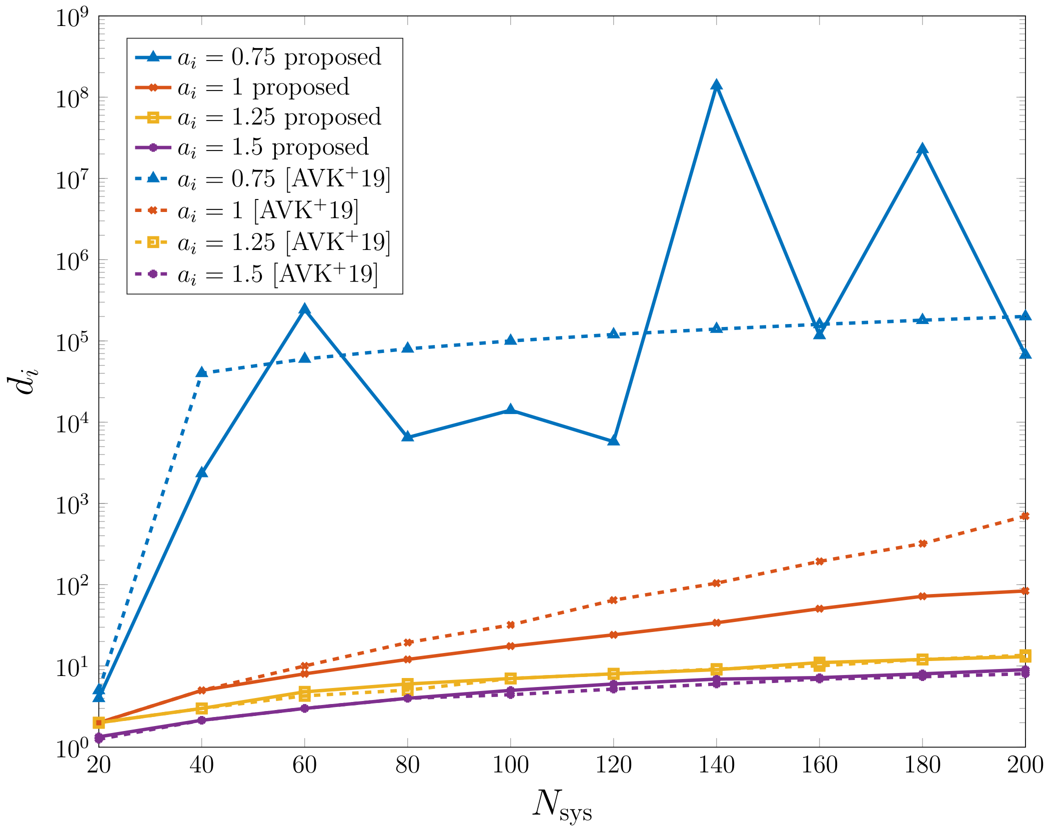

Figure 3.5: Duration between transmissions for $R_{\mathrm{UL}} = 1$

small changes in r_i result in great changes of the corresponding d_i and lead to the fluctuations visible in Fig. 3.5 for $a_i = 0.75$. These small variations result from the numerical solution of (3.17). Since these subsystems are stable, they do not contribute much to the overall uncertainty optimized in (3.17) and the fluctuations do not impact the result. For comparison, the results from [AVK+19] are shown. The differences between the two schemes are significant, especially for the subsystems with $a_i = 1$, which are scheduled more often with the results from (3.17).

Next, the control performance of the different scheduling algorithms is compared in terms of the mean absolute estimation error $\Sigma_{\mathrm{e}}(k) = \frac{1}{N_{\mathrm{sys}}} \sum_{i=1}^{N_{\mathrm{sys}}} \alpha \sqrt{P_i(k)}$. Fig. 3.6 shows the results for $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 1$ in the upper part and $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 3$ in the lower part. As the ordinates, the sum of the mean absolute estimation error $\Sigma_{\mathrm{e}}(k)$ over all subsystems k is shown versus the total number N_{sys} of subsystem as the abscissae. With $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 3$, both algorithms achieve the same Σ_{e} except for many subsystems, i.e. $N_{\mathrm{sys}} > 180$. In the more constrained case ($R_{\mathrm{UL}} = R_{\mathrm{DL}} = 1$), the reduction of Σ_{e} of the proposed method compared to [AVK+19] becomes apparent. Furthermore, as explained already in the previous paragraph, the fluctuations from Fig. 3.5 are not influencing in the total uncertainty.

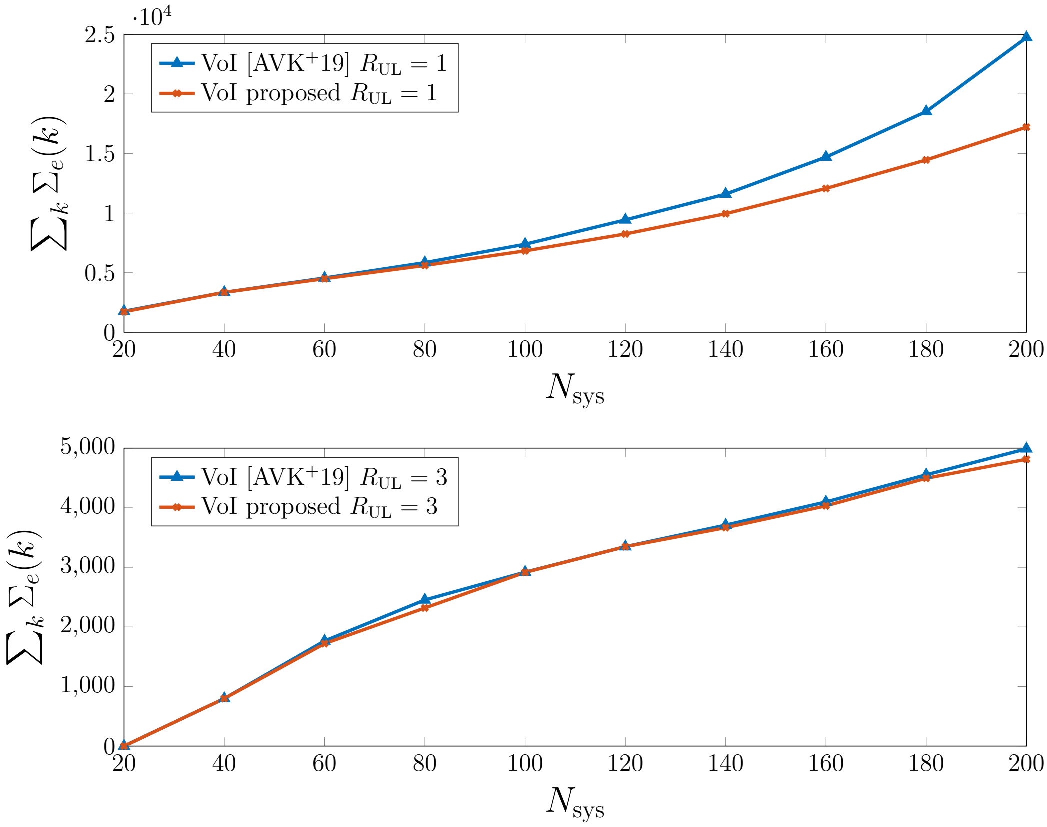

Figure 3.6: $\sum_k \Sigma_{\mathrm{e}}$ for $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 1$ and $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 3$

To illustrate the advantage of the proposed scheme from Section 3.5 for different sets of dynamics of the subsystems, i.e. different sets of coefficients a_i, with respect to the mean absolute estimation error Σ_{e} at the central controller, the next experiments were conducted with changing sets of dynamics of the subsystems. First, we will show, how a large spread in system coefficients a_i influences the results. This spread is modeled by the spreading factor

$$\alpha \in \mathbb{R}^+ \tag{3.19}$$

The four classes of subsystems are now not fixed, but given by

$$a_i \in 1.05 + \{0, 0.3, 0.45, 0.75\}\alpha. \tag{3.20}$$

As a consequence, all subsystems will always be unstable, but with different dynamics. The number of subsystems was fixed to $N_{\mathrm{sys}} = 160$, with 40 subsystems per each class as described in (3.20). First, the influence of the spreading factor α on the durations d_i between two consecutive transmissions for subsystem i was investigated. The results are shown in Fig. 3.7. The available communication resources were set to $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 1$.

Similar to the results shown in Fig. 3.5, the proposed algorithm selects smaller d_i for

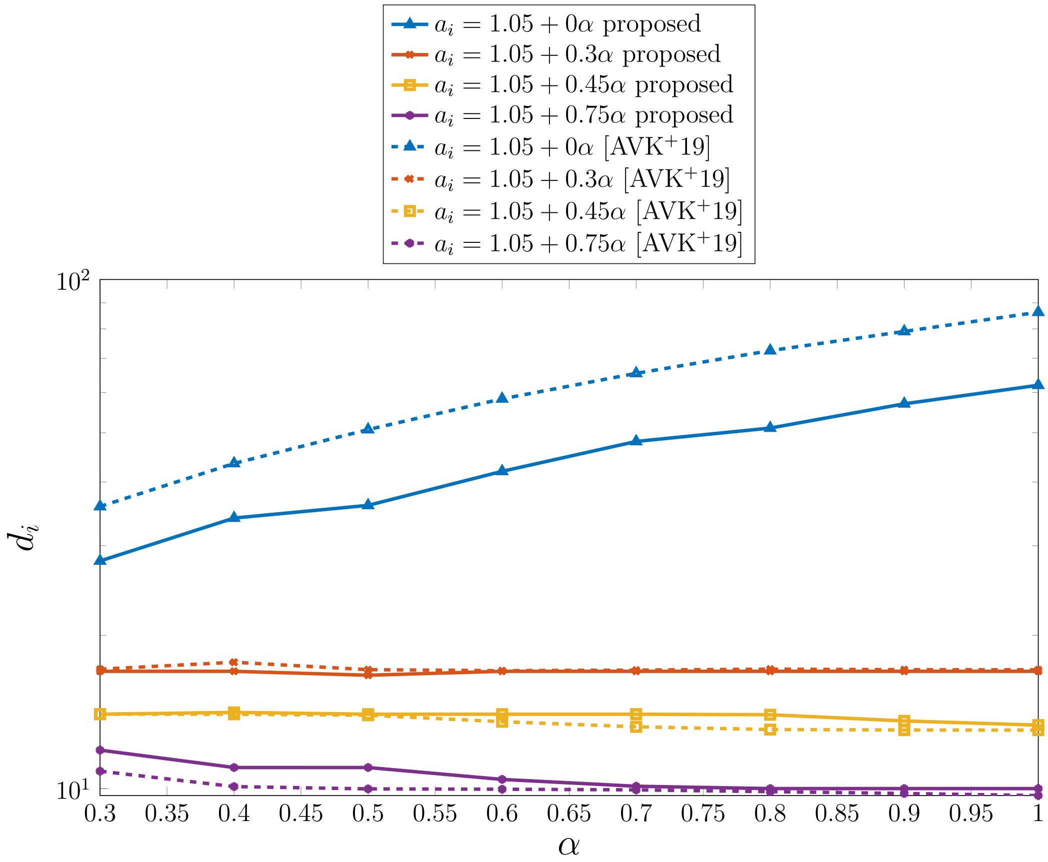

Figure 3.7: Duration between transmissions for $R_{\mathrm{UL}} = 1$ and different spreads α between the subsystem coefficients

smaller system coefficients a_i compared to the reference scheme from [AVK+19], i.e. scheduling them more often, while the update rate for subsystems with large a_i, and, therefore, higher dynamics, is reduced. For a low spreading factor α, the difference in the durations d_i is also low, because even the subsystems in the highest class with $a_i = 1.05 + 0.75\alpha$ have low dynamics for small α compared to the setups with larger α.

The higher the spreading factor α, the more demanding the update task gets, because especially those subsystems in the class with the largest coefficients a_i result in high dynamics for subsystems of this class, while the subsystems from the lowest class are kept constant at $a_i = 1.05$. This results in higher dynamics of the overall system, making the estimation task more demanding and resulting in higher estimation errors.

This is shown in Fig. 3.8, where the influence of the spreading factor α on the sum of the mean absolute estimation error Σ_{e} is depicted. For an increasing spread between the largest a_i and the smallest a_i, the proposed algorithm from Section 3.5 reduces the estimation error at the central controller compared to the VoI based greedy scheme from [AVK+19].

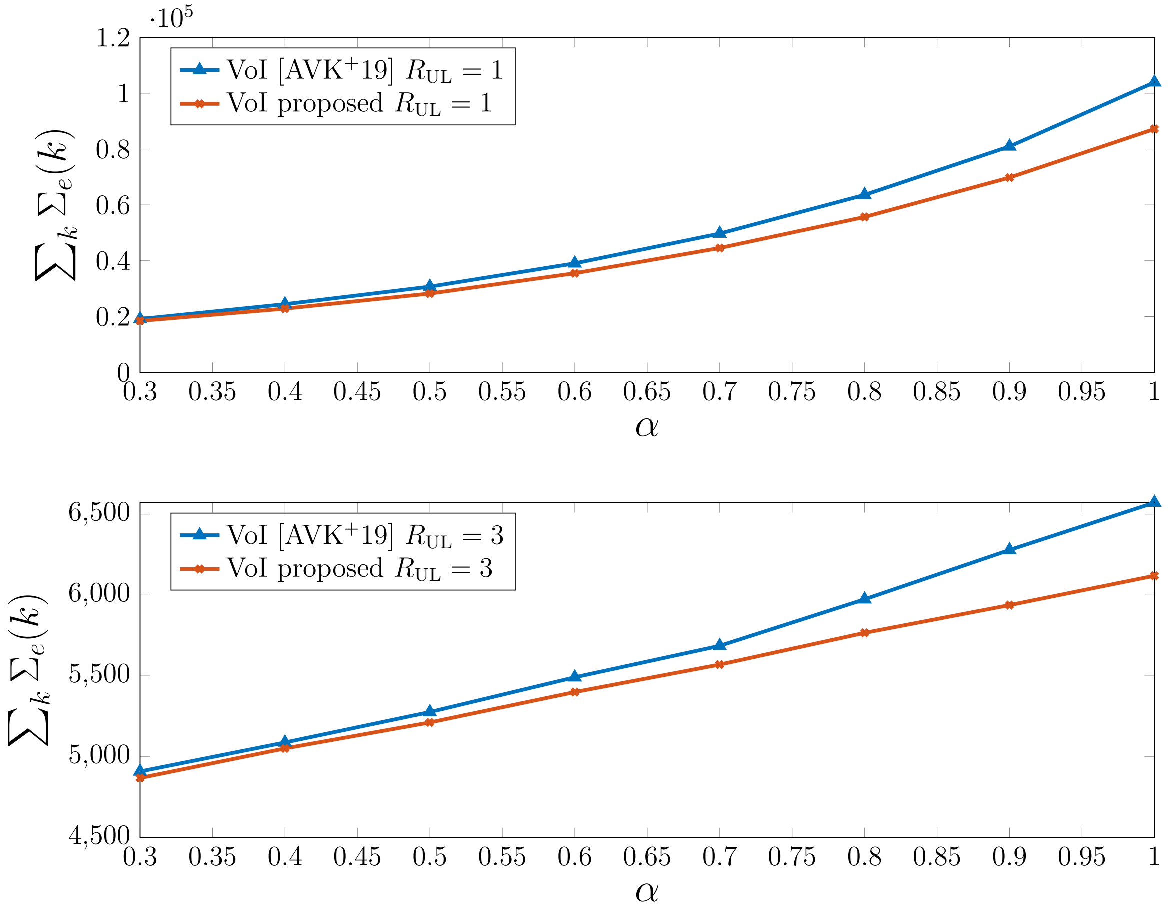

Figure 3.8: $\sum_k \Sigma_{\mathrm{e}}$ for $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 1$ and $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 3$ with varying spread α of a_i and $N_{\mathrm{sys}} = 160$

Now, the influence of the ratio of the number of subsystems with high a_i to the number of subsystems with low a_i shall be investigated. For this purpose, only two classes of subsystems are used, $a_i \in \{1.1, 1.7\}$. The factor β determines the ratio of the two types, with

$$a_i = \begin{cases} 1.7 & \text{for } i < \beta N_{\mathrm{sys}} \\ 1.1 & \text{else.} \end{cases} \tag{3.21}$$

Defining β as in (3.21) lets β directly resemble the ratio of the number of subsystems with $a_i = 1.7$ the total number N_{sys} of subsystems.

Fig. 3.9 shows results similar to Fig. 3.5 and Fig. 3.7. The more subsystems with high dynamics with $a_i = 1.7$ are in the system, the more demanding the estimation task gets, because the uncertainty about the subsystem states grows faster for higher a_i, so that more subsystems are scheduled more often, resulting in a lower d_i for these subsystems. Since the numbers R_{UL} and R_{DL} of communication resources are not changed, no

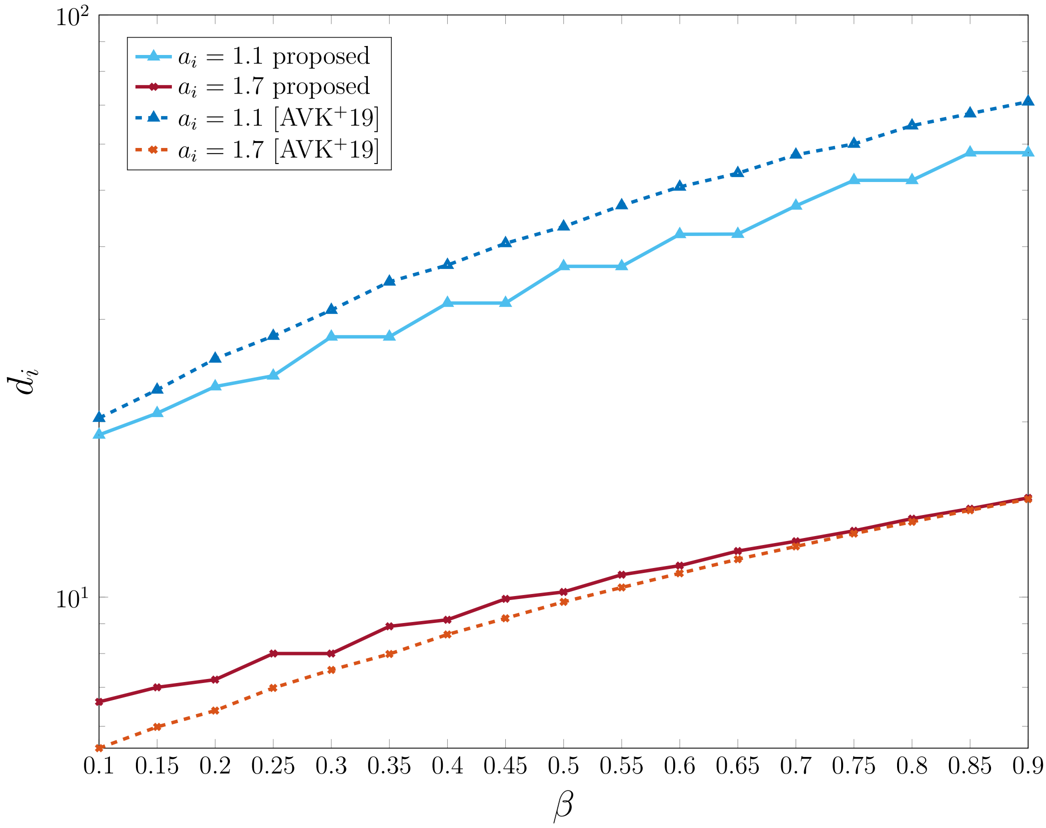

Figure 3.9: Duration between transmissions for $R_{\mathrm{UL}} = 1$ and different shares β between the subsystem coefficients

additional resources are added to the system, all d_i increase with an increasing share of the subsystems with $a_i = 1.7$, because the number of subsystems N_{sys} is kept constant. Hence, decreasing d_i for one subsystem i results in (slightly) increasing d_j for all other subsystems j. Thus, β is also influencing the total performance, even if the number N_{sys} of subsystems and the communication resources R_{UL} and R_{DL} are kept constant as in the previous experiment with a varying α, changing the spread of the coefficients a_i. The algorithm from [AVK+19] generally spends less resources on the lower dynamic subsystems than the algorithm proposed in Section 3.5. This was already visible in Fig. 3.5 and Fig. 3.7. Our proposed algorithm from Section 3.5, in comparison, gives more resources to the subsystems with $a_i = 1.1$, especially if β is low and, thus the number of subsystems with $a_i = 1.1$ is low, while the other subsystems with $a_i = 1.7$ are scheduled more rarely, if β increases and the number of subsystems with $a_i = 1.7$ therefore also increases.¸

Fig. 3.11, similar to Fig. 3.8 for the spread α of the subsystem coefficients a_i, shows the influence of β, the ratio of the number of subsystems with $a_i = 1.7$ to the total number of subsystems N_{sys}, as the abscissae on the mean absolute estimation error Σ_{e} as the

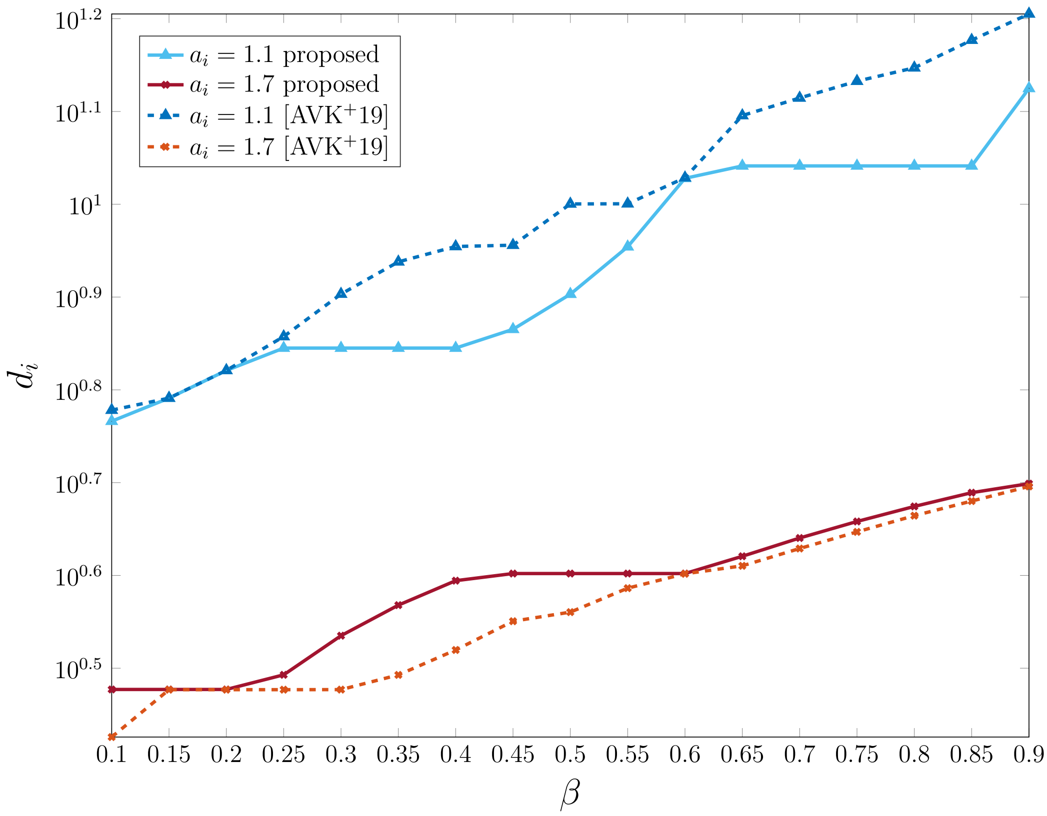

Figure 3.10: Duration between transmissions for $R_{\mathrm{UL}} = 3$ and different shares β between the subsystem coefficients

ordinates. In the upper half, the communication resources are set to $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 1$, in the lower half to $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 3$. In both cases, the increase of $\sum_k \Sigma_{\mathrm{e}}$ for higher β, which could already be anticipated by the shape of the curves in Fig. 3.9 and Fig. 3.10, is clearly visible. For most of the ratios β, our proposed algorithm from Section 3.5 achieves slightly lower sums of mean absolute errors $\sum_k \Sigma_{\mathrm{e}}$. The only exception happens for $\beta = 0.6$ and $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 3$: The gap between curves for our algorithm from Section 3.5 and the algorithm from [AVK+19] is zero here, mainly because the latter apparently can handle this situation much better compared to the other scenarios with different β. The reason for the same $\sum_k \Sigma_{\mathrm{e}}$ is, that here the resource allocation of both algorithms, the one from [AVK+19] and the one proposed in Section 3.5, result in the same update rates for the subsystem types, as shown in Fig. 3.10.

The next set of experiments shall illustrate the properties of the scheduling resulting from the scheduling algorithm described in Section 3.5 during runtime of the system. The evolution of the absolute estimation error over time behaves differently for the two schemes. The same experiment as in Fig. 3.5 and Fig. 3.6 is now conducted and the

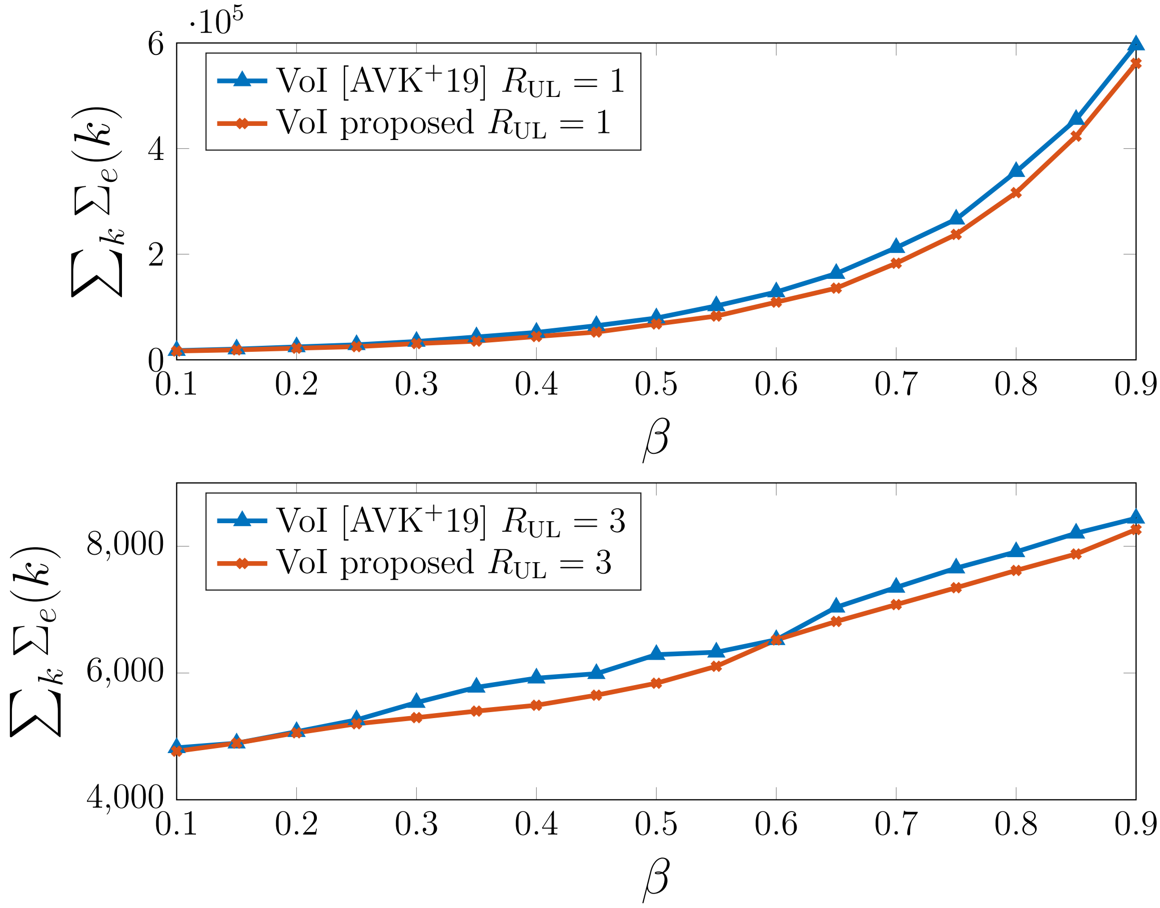

Figure 3.11: $\sum_k \Sigma_{\mathrm{e}}$ for $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 1$ and $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 3$ with varying share β of $a_i = 1.1$ and $a_i = 1.7$, $N_{\mathrm{sys}} = 160$

behaviour of the uncertainty Σ_{e} at each time step k as the ordinates over the whole simulation time $k = 0, \ldots, T_{\mathrm{sim}}$ as the abscissae is investigated. Exemplarily, Σ_{e} for $N_{\mathrm{sys}} = 160$ and $R_{\mathrm{UL}} = R_{\mathrm{DL}} = 1$ is shown in Fig. 3.12 for each control system time step k. Since the subsystems are assumed to have been reset at $k = 0$, multiple systems have to be scheduled in the first time slots, which results in a transient phase for both algorithms. After this initial phase, the uncertainty about the system state, in terms of the mean absolute error Σ_{e}, when applying our algorithm, is almost constant, while the one from [AVK+19], besides having about 20% higher mean absolute error, shows a regular oscillating pattern as described in Section 3.4.1.

In Fig. 3.13 the same experiment is conducted with $R_{\mathrm{DL}} = R_{\mathrm{UL}} = 3$. While in Fig. 3.6 both, the proposed and the method from [AVK+19], seem almost equivalent, Fig. 3.13 reveals the slightly lower average of the mean absolute error Σ_{e} of the proposed method during the whole simulation time. The small spike around $k = 3900$ of the proposed method comes from the fact, that at this time the subsystems with $a_i = 0.75$ are scheduled the first time, resulting in a small disturbance, which can be viewed as a very long transient phase. Like in Fig. 3.12, the method from [AVK+19] results in a strong oscillation of Σ_e, which does not converge to a limit like our algorithm.

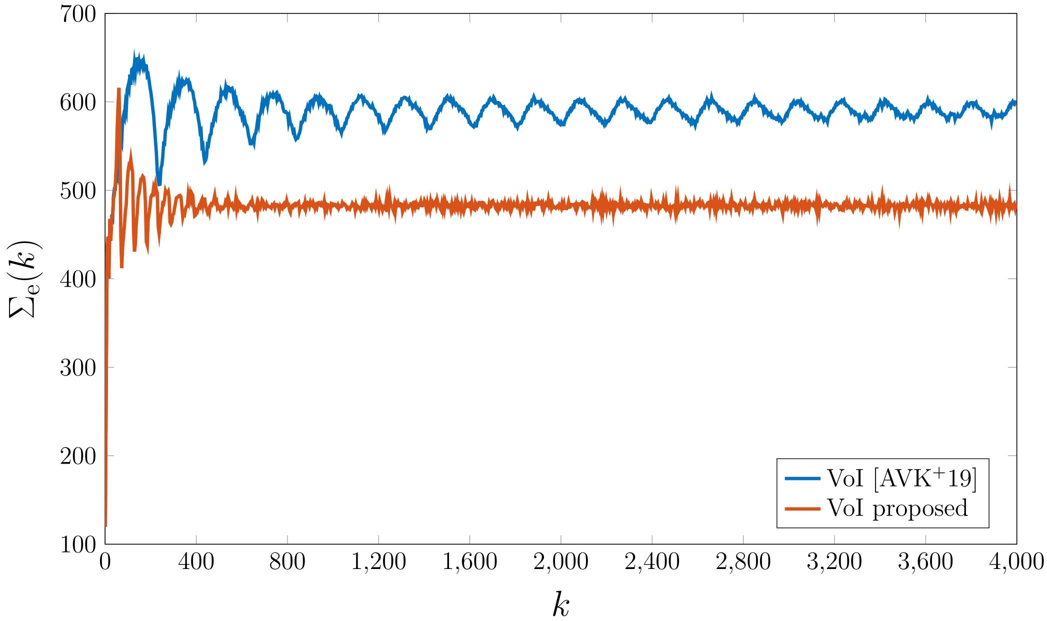

Figure 3.12: Σ_e for each k for $R_{UL} = 1$ and $N_{sys} = 160$

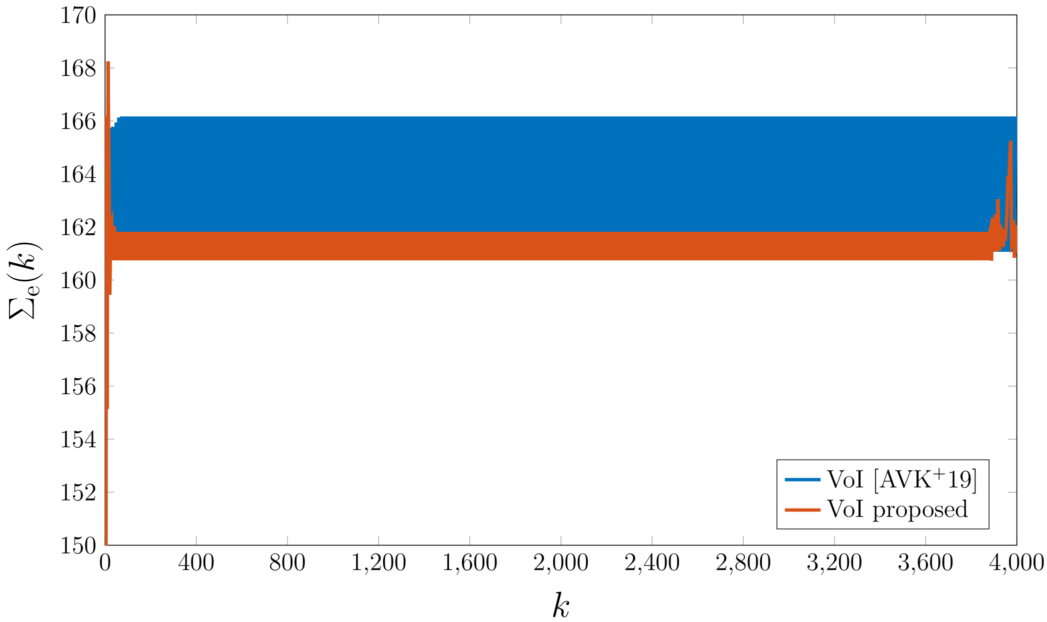

Figure 3.13: Σ_e for each k for $R_{UL} = 3$ and $N_{sys} = 160$

3.7 Conclusion

In this chapter, an optimized scheduling of sensor value transmission for discrete time LTI subsystems was developed. First we have shown, that the optimum control performance is achieved by a regular scheduling of sensor value transmission. Then, the communication resource allocation optimization problem was stated and an algorithm to generate the scheduling was developed. The results could be evaluated by the derived analytic expressions for the state estimation error variance.

The estimation error variance Σ_e is, compared to existing methods, not only reduced in mean, but is also almost constant during the system operation time.

The precalculation of transmission frequencies for each subsystem greatly reduces the runtime computational complexity, because the optimization problem is only solved once, while during runtime only the time duration between subsequent transmissions must be adjusted to approximate the precalculated update rates. The derived analytic expressions for evaluation make Monte-Carlo simulations obsolete. While the current system model of scalar linear systems with a deadbeat control is very simplistic, extensions to other linear-quadratic regulation models can easily be made. Another extension to multidimensional subsystem states requires only minor changes to the cost function (3.10).

Chapter 4

Deadline-Aware Control Command Transmission

Several parts of the content of this section have been originally published by the author of this thesis in [KOK21]. This paragraph shall illustrate the previous work from [KOK21], the relation to and the additional work presented in this chapter. The model, problem and scheduling algorithm were already published, also results for one bandwidth configuration were shown. The simulations are extended in this chapter to also display results for different amounts of communication resources. Furthermore, the actual exploitation of the available resources by the different allocation algorithms is investigated.

4.1 Introduction

In this chapter, the transmission of control commands from the central controller to the agents over a wireless link is investigated. The sensor is assumed to be attached directly to the controller, so no additional communication for state estimation at the controller is required for deriving the control commands. Hence, the focus in this chapter is now on the connection between the central scheduler, which also acts as controller in this case, and the actor at the plant, like shown in Fig. 4.1.

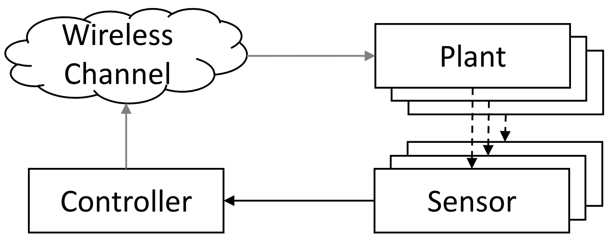

Figure 4.1: The networked control loop with wireless command transmission from the controller to multiple plants

Like in the previous chapter, a timely reception of the command at the plant is crucial for optimal control [Lun16b]. While there were no actual limits assumed on the system states in the previous chapter, in real world applications such limits always exist. In the autonomous driving example from [BDH+20], where the current lane has to be kept, the limits might be the road markings, which must not be crossed by any part of the vehicle. In the distance keeping example from [PYZ+20], the distance must not fall below a certain security distance, because otherwise the braking distance might lead to fatal accidents.

The control process dictates, depending on its characteristics, a maximum latency, after which the control commands have to be successfully delivered to the plant and the maximum percentage of failed transmissions. To achieve energy-optimal communication while maintaining the error rate and latency requirements, the allocation of time-frequency resources has to be adapted to the individual wireless channel conditions and transmission deadlines resulting from the control perspective. The capacity of a communication channel for infinite time-frequency resources according to Shannon is determined by the SNR [Sha48]. This is a valid assumption for transmissions of large amounts of data over infinite time-frequency resources, but since IIoT is especially about short packets and low latencies, this estimation is way too optimistic. The scenario considered in this chapter therefore requires a different estimation of achievable rate. Error free transmission in limited time and bandwidth is not possible, so a relation of the SNR at the receiver, latency, allowable packet error rate and time-frequency resources is required. Hence, the investigated energy minimization is formulated using a more realistic capacity formula for short packets derived by Polyanskiy et al. in their seminal paper [PPV10]. The characteristics of the short packet formula as well as its implications on the energy minimization problem considered in this chapter are investigated, according to 5 from Section 1.2.

Since the allowable packet error rate and number of time-frequency resources are limited, the SNR at the receiver, which is determined by the transmit power and the channel gain, has to be tuned to meet the requirements. To increase the SNR, the transmit power has to be increased, which results in an increase of total energy consumption.

The controller can sense the states of the agents, generates control commands according to the states and transmits them to the agents via a wireless link. The control commands are assumed to be short data packets of up to a few hundred bits in size. For each agent, we consider the different dynamics of the various types of machines by means of the definition of agent-specific deadlines. Additionally, the maximum allowable packet error rate is constrained to a low constant value to account for e.g. the safety requirements of industrial production plants. The available bandwidth for transmission is limited. Under these constraints, we find the optimal time-frequency resource allocation to the agents minimizing the required energy for transmission. This energy minimization problem relates the time-frequency resource allocation for the transmission of the control commands to each agent to the required transmit power and, therefore, energy consumption. Question 6 from Section 1.2 asks for this relation between deadlines, time-frequency resource allocation and energy consumption.

For this purpose, we first formulate a problem with a continuous amount of resources for each agent in Section 4.4. The given agent-specific maximum latencies lead to deadlines, when the transmission has to be finished the latest. The channel conditions are also given, as well as the common maximum packet error rate. This problem is shown to be convex. Then, we propose a gradient-based algorithm in Section 4.5 to allocate the time-frequency resources in an OFDM scheme in a quantized fashion. For comparison, an allocation balancing the number of resources for each agent, as far as the deadlines allow, is calculated. The three approaches are compared in Section 4.6 and the gradient based allocation is shown to be close to the continuous lower bound. Moreover, the balancing allocation of resources to all agents is shown to perform worse than the gradient-based algorithm.

4.2 Related Work

In this section, an overview of existing work and the state of the art for wireless control command transmission is given.

In [LNL+21], a system with a single controller and agent is investigated. The wireless link is situated between the controller and the plant, similar to the setup shown in Fig. 4.1 for multiple agents. The focus of this work is to find the requirements on the communication channel to enable the stabilization of the control loop. It is shown, that only the SNR of the communication channel determines the stabilizability of the control loop, not the latency of the channel. The effects of quantization to discrete commands and packet loss due to the short packet effect are considered. Further control requirements, like bounds on the state or the control input, leading to latency restrictions, are not considered.

In [dIS+20], the transmission of control commands from the controller to the plant over a wireless link is considered. The effects on the communication channel characteristics due to the short packets utilized to transmit the commands are also regarded. The main goal is to minimize the required energy for the transmission, which is achieved by minimizing the power spectral density of the time and bandwidth limited transmit signal. The limited number of time-frequency resources is split between an initial transmit and a potential retransmit, potentially saving energy, if no retransmission is needed. The receiver incorporates a buffer for control commands for the next few time steps and is assumed to send an acknowledgment for each correctly received packet. The acknowledgment is assumed to require a fixed portion of the available time and bandwidth resources. Each transmitted packet contains commands to refill the buffer completely. A maximum probability of a buffer underrun at the receiver has to be reached, thus including maximum transmission latency requirements. In this scenario, all available time-frequency resources are used for the single agent, while the multi-agent scenario is not considered.

A multi-agent bandwidth minimization problem is studied in [WQQ20]. While not directly considering a control scenario, the transmission from a single transmitter to multiple agents with strict reliability and latency requirements can also be used in the scenario in Fig. 4.1 for command transmission. The channel is assumed to be frequency selective, but only known to the receiver. The transmit power per resource element is fixed. To minimize the energy required for transmissions while fulfilling the requirements on latency, error rate and minimum throughput, it is therefore sufficient to minimize the total allocated bandwidth.

[SSY$^+$19] maximizes the energy efficiency in a scenario with multiple sensors transmitting data to agents. Sensors and agents are assumed to be in different mobile radio cells, the transmissions were done in Ultra-Reliable Low Latency Communication (URLLC) style. This results in a multi-hop scenario from the sensor as transmitter to the first base station, from the first base station via the backhaul link to a second base station and from the second base station to the agent. The deadlines for these transmissions are assumed identical, which is unrealistic for an industrial plant with agents belonging to different classes of machines, such as heaters with lower dynamics and transportation devices keeping a lane or driving in a platoon, resulting in higher dynamics, as discussed in Section 3.1.

The small amounts of data required for transmitting control commands to each agent individually require a different assessment of the communication channel capacity, which is given by Polyanskiy et. al. in [PPV10]. The deadlines induced by the control requirements are already considered in [WQQ20] and [dIS$^+$20], but the combination of different latency requirements for multiple agents and minimization of transmit energy, as formulated as question No. 5 in Section 1.2, is not considered. In [SSY$^+$19], it is shown that increased time-frequency resources do not always result in reduced energy consumption, especially for short data packets, which leads to question No. 6 on how to distribute the available resources for minimal transmit energy consumption.

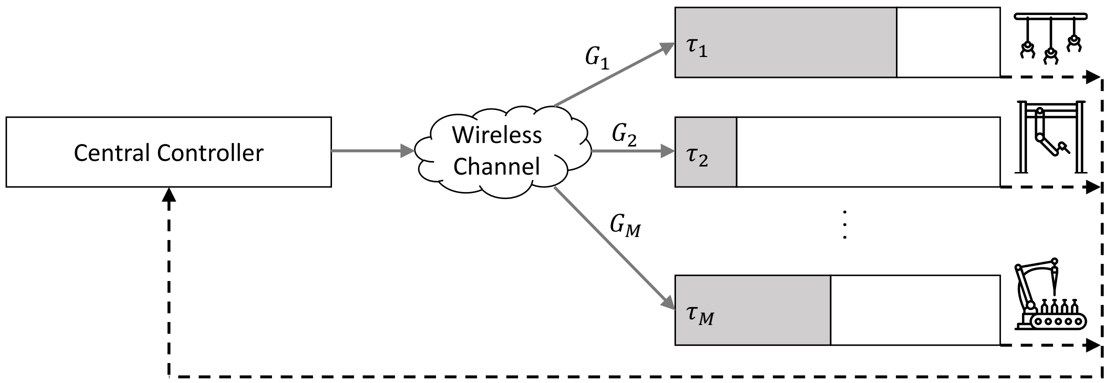

Figure 4.2: System model of the deadline-aware control command transmission

4.3 System Model

The system consists of a single central controller and M agents randomly distributed around the controller, as shown in Fig. 4.2. The central controller senses the control system states of all agents and generates control commands accordingly, which are then transmitted to the agents. The control system is assumed to be discrete-time with a time slot duration T. For each time slot a new control command is generated for every agent, the commands are all available at the beginning of the time slot. The time elapsed since the beginning of the time slot is denoted by $t, 0 \leq t \leq T$, $t = 0$ indicates where transmission starts.

The performance of the control system is determined by the latency of the control commands, so each agent m has an individual deadline $\tau_m, 0 \leq \tau_m \leq T$ for the successful reception of its command after the beginning of the time slot. The value of τ_m depends on the dynamics of agent m, where higher dynamics generally lead to shorter deadlines. Allocating resources to agent m after its deadline τ_m has passed would not contribute to a timely reception, so we assume no resources after the deadline are allocated. Additionally, the probability of a lost control command must not exceed p_c to keep the agents in a safe operation region.

Throughout this paper, a continuous quantity x will be denoted by $x'(t)$, while its piecewise continuous counterpart will be denoted by x_t. The total bandwidth available for transmission is denoted by B. The time-variant bandwidth assigned to agent m at time t is $b'_m(t) \leq B$. $b'_m(t)$ is assumed to fulfill the uncertainty principle, i.e. it does not change arbitrarily fast. Moreover, the sum of all assignments must not exceed the total bandwidth, i.e.

$$\sum_{m=1}^{M} b'_m(t) \leq B \text{ for } 0 \leq t \leq T. \tag{4.1}$$

The commands for each agent, consisting of N bits, are transmitted over a wireless channel, which is perfectly known at the central controller and the receiving agents. The agents are assumed to be stationary. The channel between the controller and every agent is modelled as line-of-sight (LOS). Thus, the channel is assumed to stay constant over T and B. The power gain of the channel from the central controller to agent m is denoted by the scalar channel gain G_m.

The transmission is performed interference free by using frequency division multiple access on the available bandwidth B and time T for each agent. The integral of $b'_m(t)$ with respect to t corresponds to the time-frequency resources of agent m, denoted by n'_m

$$n'_m = \int_0^T b'_m(t)dt. \tag{4.2}$$

The Power-Spectral-Density (PSD) of the transmit power for agent m is denoted by q_m. It is assumed to stay constant for the whole transmission. The total energy E_m spent for the transmission to agent m is then given by

$$E_m = q_m \int_0^T b'_m(t)dt. \tag{4.3}$$

To account for the deadlines τ_m in (4.3), the assigned bandwidth for agent m, $b'_m(t)$, must be set to zero for $t > \tau_m$. The receiver noise is assumed to be AWGN, whose power $\sigma_m^2(t)$ depends only on the noise PSD N_0 and the bandwidth $b'_m(t) \leq B$ assigned to agent m at t, i.e. $\sigma_m^2(t) = N_0 b'_m(t)$. The SNR at agent m is then

$$\gamma_m = \frac{q_m b'_m(t)}{\sigma_m^2} = \frac{q_m b'_m(t)}{N_0 b'_m(t)} = \frac{q_m}{N_0}. \tag{4.4}$$

Since the commands are short and transmit time and bandwidth is limited, the well-known Shannon capacity formula

$$C_m = \log_2\left(1 + \gamma_m\right) \tag{4.5}$$

for error-free transmission is too optimistic to determine the minimum SNR and has to be extended for short packets. Therefore, also the channel dispersion for agent m,

$$V_m = \gamma_m \frac{2 + \gamma_m}{\left(1 + \gamma_m\right)^2} \log_2^2\left(e\right), \tag{4.6}$$

has to be considered. In [SSY+19] the approximation

$$V_m \approx \log_2^2(e) \tag{4.7}$$

is given, which is valid for $\gamma_m \geq 5\text{dB}$. For the strict demands on p_c, the short packet sizes N and limited resources n_m, generally $\gamma_m \geq 5\text{dB}$ is required. For an AWGN channel, the normal approximation from [PPV10] gives the short packet formula for a packet error rate p_c, given a certain packet size N, the number of time-frequency resources n_m, the channel dispersion V_m and the SNR at the receiver γ_m. The packet error probability $p_{c,m}$ for agent m can then be approximated by

$$p_{c,m} \approx Q\left(\frac{n_m C_m - N + \frac{\log_2 n_m}{2}}{\sqrt{n_m V_m}}\right), \tag{4.8}$$

where $Q(\cdot)$ is the Gaussian Q-function.

The minimal Shannon capacity corrected for short packets $C_{\text{corr},m}$ and therefore E_m required to fulfill the latency and error rate requirements $p_{c,m}$ for each agent m can be calculated using a reformulated version of (4.8):

$$C_{\text{corr},m} \approx \frac{1}{n'_m}\left(\sqrt{n'_m V_m} Q^{-1}(p_c) + N - \frac{\log_2(n'_m)}{2}\right) \tag{4.9}$$

$$E_m\left(n'_m, G_m, N, p_c\right) = (2^{C_{\text{corr},m}} - 1)\frac{N_0 n'_m}{G_m} \tag{4.10}$$

In the next section, the minimum total energy for continuous n_m will be derived. In practice, however, a continuous allocation of time-frequency resources is not possible. Therefore, we follow the approach used in mobile radio standards like 5G New Radio (NR) to implement the OFDM scheme, dividing the time-frequency plane into a grid of rectangles, called resource elements. The total available bandwidth B is split into N_{sc} subcarriers. The subcarrier bandwidth is b_{sc}, such that $B = b_{\text{sc}} N_{\text{sc}}$. The number of OFDM symbols per time slot is N_{sym}, such that $T = N_{\text{sym}} t_{\text{sym}}$. Considering the available bandwidth, we define a time-frequency resource element as $t_{\text{sym}} b_{\text{sc}}$. Each resource element is identified by its time index $t, t = 1, \ldots, N_{\text{sym}}$ and subcarrier index $s, s = 1, \ldots, N_{\text{sc}}$. The number of resource elements for agent m is denoted by n_m. The complete distribution of resource elements is collected in the vector $\boldsymbol{n} = [n_1, \ldots, n_M]^{\mathrm{T}}$. Due to the nature of OFDM, the area of one resource element is always $t_{\text{sym}} b_{\text{sc}} = 1\text{s} \cdot \text{Hz}$. In Section 4.5, two algorithms to derive distributions of resource elements $\boldsymbol{n}$ are proposed.

4.4 Problem Formulation

4.4.1 General formulation

The overall goal is to minimize the total energy $E = \sum_{m=1}^{M} E_m\left(n'_m, G_m, N, p_c\right)$ used for the command transmission. Since B as well as the available time, due to the deadlines, is limited, n'_m is also limited. The energy minimization problem for the continuous resource case is then

$$\min_{n'_1,\ldots,n'_M} \sum_{m=1}^{M} E_m\left(n'_m, G_m, N, p_c\right), \tag{4.11a}$$

s.t.

$$\sum_{m=1}^{M} b'_m(t) \leq B \text{ for } 0 \leq t \leq T, \tag{4.11b}$$

$$b'_m(t) = 0 \text{ for } \tau_m < t \leq T, m = 1, \ldots, M, \tag{4.11c}$$

where (4.11b) enforces the bandwidth limitation and (4.11c) effectively restricts the transmission to $0 \leq t \leq \tau_m$.

4.4.2 Convex reformulation

Problem (4.11a)–(4.11c) is hard to tackle, because the solution space is non-convex. We will now implement constraints on n_m and $b'_m(t)$ to get a convex subset of the original solution space, still containing the optimal solution.

First, we restrict on n_m to make (4.11a) convex. In [SSY+19], the partial convexity of (4.10) in n_m up to an inflection point $n_{m,\text{thr}}$, i.e. for $n_m \leq n_{m,\text{thr}}$, is shown. For illustrative purposes of this partial convexity of (4.10), three examples for different packet sizes N are shown in Fig. 4.3.

Furthermore, the number n_m of resources achieving the global minimum of (4.10), $n_{m,\text{min}}$, is shown to be $0 \leq n_{m,\text{min}} \leq n_{m,\text{thr}}$. Thus, adding the constraint $n_m \leq n_{m,\text{thr}}$ will turn (4.10) and (4.11a) into convex functions in n_m. To illustrate these properties, Fig. 4.4 shows graphs for (4.10) with $p_c = 10^{-9}$, $G_m = -70\,\text{dB}$ and $N = 512$.

Since (4.11a) is only based on n_m and not on $b'_m(t)$ directly, $b'_m(t)$ can be restricted to be piecewise constant, without further restrictions on n_m. The values of the constant pieces are then collected in a vector $\boldsymbol{b}_m$.

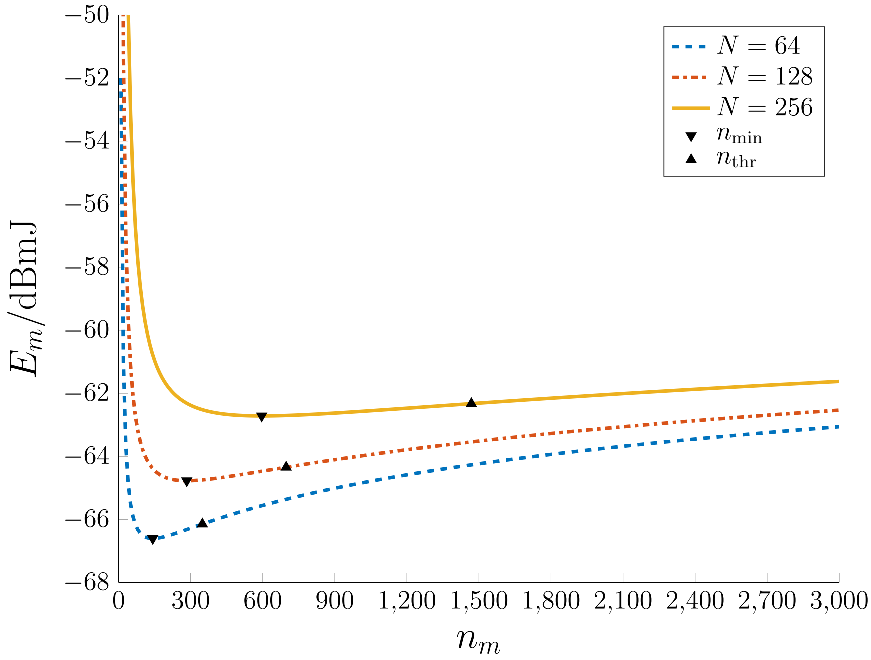

Figure 4.3: Required energy E_m

Furthermore, only the integral over $b'_m(t)$ determines n_m in (4.2). and therefore E_m in (4.10). Hence, $b'_m(t)$ can be assumed to be piecewise constant. As a consequence, (4.2) becomes a sum of rectangular areas. The width of the rectangles is selected as the distance between two consecutive deadlines.

An example is given in Fig. 4.4. A piecewise constant allocation is shown for three agents with deadlines τ_1, τ_2, τ_3 in increasing order. The vectors are then $\boldsymbol{b}_1 = (2, 0, 0)^{\mathrm{T}}$, $\boldsymbol{b}_2 = (1, 4, 0)^{\mathrm{T}}$ and $\boldsymbol{b}_3 = (5, 3, 1)^{\mathrm{T}}$. The first constant starts at $t = 0$ and ends at τ_1, the second ranges from τ_1 to τ_2 and so on, up to τ_M.

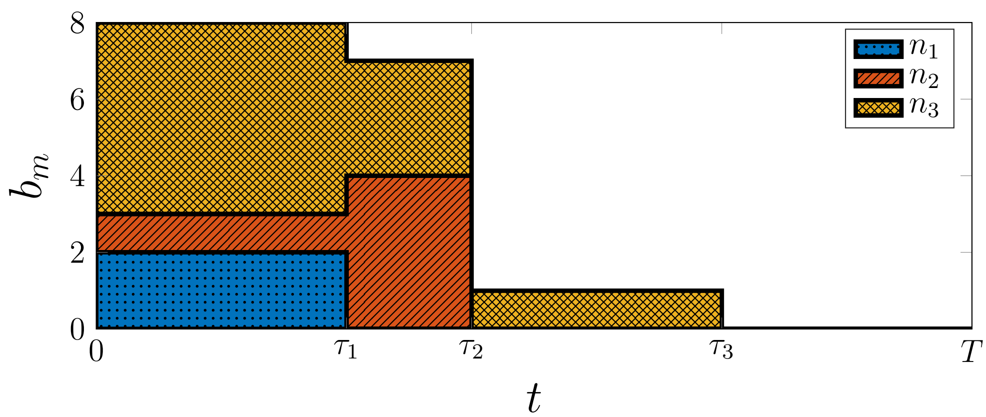

Figure 4.4: Example for constant bandwidth assignments

With the auxiliary variable

$$\tau_0 = 0, \tag{4.12}$$

we have the convex problem

$$\min_{n_1,\dots,n_M} \sum_{m=1}^{M} E_m\left(n'_m, G_m, N, p_c\right) \tag{4.13a}$$

$$n_m := \sum_{k=1}^{M} (\tau_k - \tau_{k-1}) b_{m,k} \tag{4.13b}$$

s.t.

$$\sum_{m=1}^{M} b_{m,k} \leq B \text{ for } k = 1, \dots, M \tag{4.13c}$$

$$b_{m,k} = 0 \text{ for } m = 1, \dots, M, k = m+1, \dots, M \tag{4.13d}$$

$$n_m \leq n_{m,\text{thr}} \text{ for } m = 1, \dots, M \tag{4.13e}$$

4.5 Resource Scheduling Algorithms

4.5.1 Gradient-Based Resource Scheduling Algorithm

The continuous allocation of time-frequency resources is not possible in an OFDM scheme, which splits the time-frequency plane into a grid of small rectangles. Therefore, (4.13a)–(4.13e) can only be used as a lower bound on E. To find a solution for the discrete-time and discrete-bandwidth problem, two scheduling algorithms are developed. The first algorithm is based on the fact that (4.13a) is convex in n_m for all m up to $n_{m,\text{thr}}$. All resource elements are iteratively allocated to the agents. In each iteration, the resource elements $\boldsymbol{n} = [n_1, \dots, n_M]^{\mathrm{T}}$, which were allocated in the previous iteration, determine the possible reduction of E_m for each agent m, if an additional resource element is allocated to it. Therefore, the gradient of E, $\frac{\partial E}{\partial \boldsymbol{n}}(\boldsymbol{n})$ is used as the decision criterion to select the agent for the resource element in the current iteration. The resource elements can be distributed to the agents according to (4.13d) and (4.13e). However, not all resource elements are beneficial to all agents because of (4.13d). For resources at t, only agents with $\tau_m \geq t$ can benefit. The larger t, the more deadlines τ_m have passed, hence less agents will benefit from these resources. Before allocating the resource elements, the level of competition, i.e. how many agents can actually benefit from a certain resource element, must be calculated for each resource element.

Therefore, the algorithm consists of two phases. First, the level of competition for each resource element is determined. Second, the resource elements are allocated to the agents, starting with the resource elements with the lowest level of competition. If multiple agents can use a resource element, the agent who achieves a greater energy reduction with this additional resource element gets it.

Algorithm 2 Algorithm Phase 1: Determine level of competition

Input: $\tau_1, \dots, \tau_M$
Output: $n_1, \dots, n_M$
PHASE 1: Calculate levels of competition
for $t = 1 : N_{\text{sym}}$ **do**
  for $s = 1 : N_{\text{sc}}$ **do**
    for $m = 1 : M$ **do**
      if $t \leq \tau_m$ **then** ▷ check, if resource element at s and t is before deadline of agent m
        $r_{s,t,m} = 1$
      else
        $r_{s,t,m} = 0$
      end if
    end for
  end for
end for
$c_{s,t} = \sum_{m=1}^{M} r_{s,t,m}$ ▷ calculate levels of competition

In the first phase, the level of competition is stored in matrix $\boldsymbol{C} \in \mathbb{N}^{N_{\text{sc}} \times N_{\text{sym}}}$. The element $c_{s,t}$ of matrix $\boldsymbol{C}$ contains the number of agents, which can use the resource element at subcarrier s and time instant t. To calculate $\boldsymbol{C}$, first, the three-dimensional array $\boldsymbol{R} \in \{0; 1\}^{N_{\text{sc}} \times N_{\text{sym}} \times M}$ is generated. The element $r_{s,t,m}$ is set to 1, if agent m can use the resource element at subcarrier s and time instant t, and to 0 otherwise. Finally, the array $\boldsymbol{R}$ is summed up along the third dimension to get $\boldsymbol{C}$, i.e. $c_{s,t} = \sum_{m=1}^{M} r_{s,t,m}$.

In the second phase, the resource elements are allocated to the agents in increasing level of competition, starting with elements with $c_{s,t} = 1$ up to $c_{s,t} = M$. Elements with $c_{s,t} = 0$ are neglected, because no agent benefits from them. Now, all resource elements with the current level of competition are determined and their subcarrier and time indices s and t are stored in the vectors $\boldsymbol{s}$ and $\boldsymbol{t}$, respectively. In each iteration, one resource element identified by corresponding s and t from $\boldsymbol{s}$ and $\boldsymbol{t}$ is considered. First, the agents competing for this element are stored in the vector $\boldsymbol{m}$. Then, the current total number n_m of elements allocated to agent m is calculated. The agent m from $\boldsymbol{m}$ with the smallest derivative $g_m = \frac{\partial E_m}{\partial n_m}(n_m)$ is assigned the resource element, because this results in the greatest reduction of E. If $g_m \geq 0$, $n_{m,\min}$ is achieved, agent m has

Algorithm 3 Phase 2 of gradient-based Scheduling algorithm

PHASE 2: Allocate resource elements to agents

for $l = 1 : M$ **do**
    $(\boldsymbol{s}, \boldsymbol{t}) = \text{findindex}(c_{s,t} == l)$ ▷ find all resource elements for current level of competition
    for $(s, t) \in (\boldsymbol{s}, \boldsymbol{t})$ **do**
        $n_m = \sum_{m=1}^{M} \sum_{s=1}^{N_{sc}} a_{s,t,m}$ ▷ calculate current numbers of resource elements
        $\boldsymbol{m} = \text{findindex}(r_{s,t} == 1)$ ▷ find all agents competing for this element
        $g_m = \frac{\partial E_m}{\partial n_m}(n_m)$ ▷ calculate derivative for $m \in \boldsymbol{m}$
        $o = \text{sort}(\boldsymbol{g})$
        for $m \in o$ **do**
            if $m \in \boldsymbol{m}$ **then** ▷ check if element usable by agent m
                if $g_m < 0$ **then** ▷ check if not yet larger than $n_{m,\min}$
                    $a_{s,t,m} = 1$
                    break for
                else
                    $a_{s,t,m} = 0$
                end if
            end if
        end for
    end for
end for
$n_{m,\text{opt}} = \sum_{m=1}^{M} \sum_{s=1}^{N_{sc}} a_{s,t,m}$ ▷ calculate final numbers of resource elements

no benefit from any more resources. The allocation is stored in the three-dimensional array $\boldsymbol{A} \in \{0; 1\}^{N_{\text{sc}} \times N_{\text{sym}} \times M}$.

Finally, all allocations from $\boldsymbol{A}$ are combined to get the total numbers of resources n_m. If the resource elements can only be assigned in groups like the physical resource blocks in 5G NR [3GP21], the N_{sc} for the algorithm has to be reduced accordingly. For the calculation of g_m and the final counting to get n_m, the size of one resource element group has to be adapted. A summary of the code is presented in Algorithm 2 and Algorithm 3.

4.5.2 Deadline-Aware Balancing Scheduling Algorithm

For comparison, a simpler resource element balancing algorithm is developed. The gradient of E is not considered, but rather the number of resource elements already allocated to agent m is the decision criterion for the current iteration. The resource element is allocated to the agent with the least number of resources, i.e. with the lowest n_m whose deadline has not yet passed and, thus, can use the resource element of the current iteration. This is done by replacing line 6 in Algorithm 3 by $g_m = n_m$. Furthermore, the check in line 10 must be removed, since the number n_m of allocated resources is always positive. A summary of this modified second phase is shown in Algorithm 4.

Algorithm 4 Phase 2 of balancing scheduling algorithm

PHASE 2 of balancing allocation: Allocate resource elements to agents

for $l = 1 : M$ **do**
  $(\boldsymbol{s}, \boldsymbol{t}) = \text{findindex}(c_{s,t} == l)$ ▷ find all resource elements for current level of competition
  for $(s, t) \in (\boldsymbol{s}, \boldsymbol{t})$ **do**
    $n_m = \sum_{m=1}^{M} \sum_{s=1}^{N_{sc}} a_{s,t,m}$ ▷ calculate current numbers of resource elements
    $\boldsymbol{m} = \text{findindex}(r_{s,t} == 1)$ ▷ find all agents competing for this element
    $g_m = n_m$ ▷ take number of already allocated resources
    $o = \min(\boldsymbol{g})$
    $a_{s,t,o} = 1$
  end for
end for
$n_{m,\text{opt}} = \sum_{m=1}^{M} \sum_{s=1}^{N_{sc}} a_{s,t,m}$ ▷ calculate final numbers of resource elements

As a consequence, an equal allocation, as far as the deadlines allow, is achieved. Moreover, the channel gain G_m is not considered. Hence, the number $n_{m,\text{min}}$, which achieves the minimum energy E_m as shown in Fig. 4.3 can also be exceeded if sufficient resources are available, i.e. n_{tot} is large, leading to an increase of the required total energy E. This case is investigated in detail in the next section.

4.6 Numerical Results

The numerical results are generated for parameters based on the 5G NR standard. In particular, we consider the frame structure [3GP21] and the possibility to make shorter time allocations instead of assigning a whole frame to an agent, so-called minislots [3GP20]. The carrier frequency f_c is chosen to be 6 GHz, corresponding to unlicensed band n96 of 5G NR. The channel is assumed to be pure LOS, so G_m only depends on the distance of agent m to the central controller, but not on the subcarrier frequency or the time t. Random deadlines $\tau_1, \ldots, \tau_M$ are used and the results are derived from a Monte-Carlo simulation. Each Monte-Carlo run uses a new set of deadlines, drawn from the uniform distribution $\mathcal{U}\left(14t_{\text{sym}}; 70t_{\text{sym}}\right)$ for each τ_m to ensure $\tau_m \leq T$ on the one hand and make the problem feasible on the other hand. Parameters common to

Carrier frequency f_c	6 GHz
Noise power spectral density N_0	$-174\,\frac{\text{dBm}}{\text{Hz}}$
Number of subcarriers N_{sc}	4
Subcarrier bandwidth B_{sc}	15 kHz
Symbol duration t_{sym}	66.666 μs
Deadlines τ_m	$\tau_m \sim \mathcal{U}(14t_{\text{sym}}; 70t_{\text{sym}})$
Maximum packet error probability p_c	10^{-9}
Number of OFDM symbols N_{sym}	70
Monte-Carlo runs per experiment R	10,000

Table 4.1: Simulation parameters

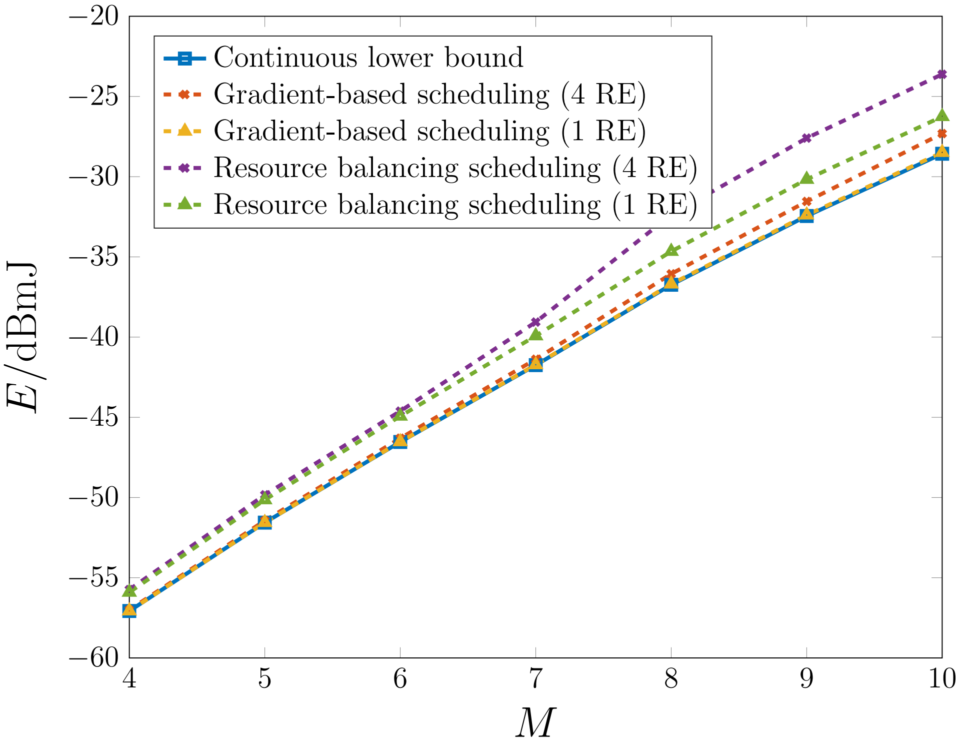

Figure 4.5: Energy for different number M of agents

all simulations are given in Table 4.1. The allocations generated by the gradient-based scheduling algorithm described in Section 4.5.1 and the balancing scheduling algorithm described in Section 4.5.2 are compared to the continuous lower bound derived in Section 4.4.2. Both, the gradient-based scheduling algorithm and the balancing scheduling algorithm, are used to either assign a single resource element or four resource elements from a single OFDM symbol per iteration.

In Fig. 4.5 the required energy E for different numbers M of agents is shown. The available resources are kept constant as in Table 4.1. On the axis of the ordinate, the

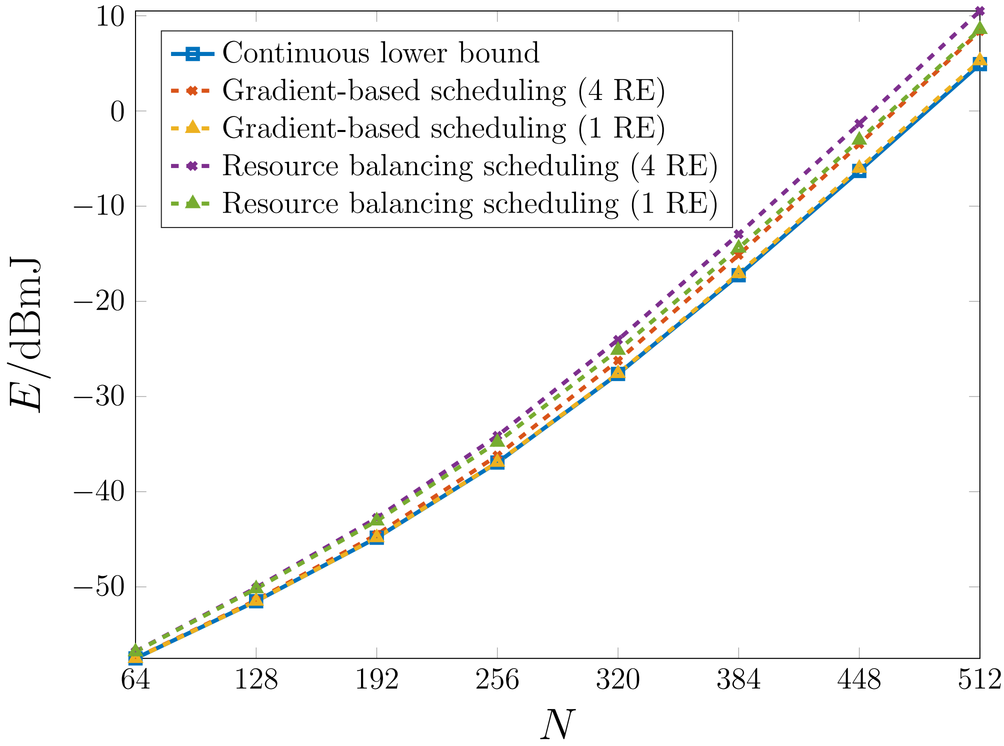

Figure 4.6: Energy for different packet sizes N

required total energy E is shown in a logarithmic scale. The axis of abscissa is used to depict the number M of agents competing for resources. The agents are spaced equidistant on a straight line starting at the central controller. The agent $m = 1$ is at a distance $d_{\text{min}} = 5\,\text{m}$ from the central controller, the agent $m = M$ at $d_{\text{max}} = 100\,\text{m}$. Consequently, the larger the number M of agents is, the closer the individual agents are spaced. The packet size is $N = 256\,\text{bits}$, which is in the center of the range for N, where (4.8) is valid, according to [PPV10]. The more agents are in the scenario, the less resources per agent are available, therefore the required energy E increases. In a highly constrained scenario, i.e. $n_m \ll n_{m,\text{min}}$, changes in n_m have greater influence on E_m, because, as shown in Fig. 4.3, (4.10) is strictly convex in n_m for $n_m < n_{m,\text{min}}$. The continuous lower bound gives the minimum E, if there were no quantization effects of n_m on E. The influence of the coarse grid with the allocation in blocks of four resource elements on the performance becomes apparent especially for $M \geq 8$. For $M \geq 8$, the gradient based scheduler is about $0.1\,\text{dB}$ worse than the optimum in single resource element case and about $1.25\,\text{dB}$ in the 4-resource element case. Meanwhile, the balancing algorithm needs $5\,\text{dB}$ and $2.3\,\text{dB}$ more than the lower bound, respectively.

In Fig. 4.6, the influence of different packet sizes N as the abscissae on the required energy E as the ordinates is investigated. The number of agents is $M = 7$, the agents

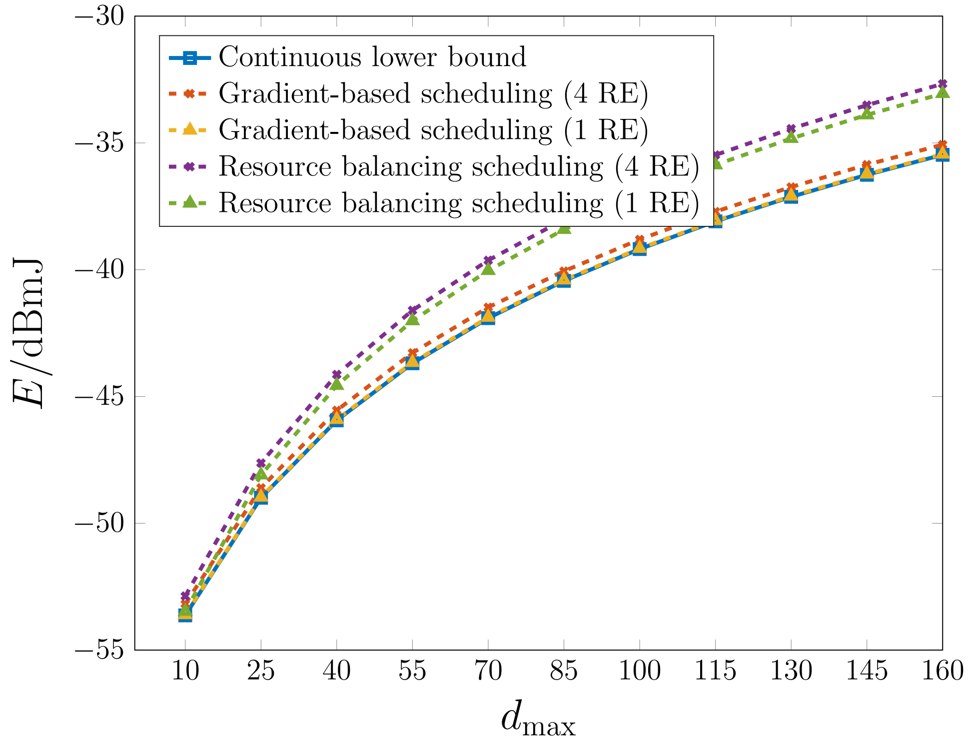

Figure 4.7: Energy for varying maximum agent distances $d_{\max}$, resulting in different G_m

are spaced equidistant from $d_{\min} = 5\text{m}$ to $d_{\max} = 100\text{m}$, like in the previous experiment. Since the number of resources and agents is fixed, the scenario becomes more constrained when the packet size increases. This is because the more bits are transmitted, the larger is the number of bits per resource element. The effect on E is similar to the previous result, due to the curvature of (4.10) for small n_m in constrained scenarios. The gradient-based scheduling algorithm for a single resource element gets results about 0.3 dB worse than the optimum derived by solving (4.11a), even for high N. The balancing benchmark scheduler always needs about 3 dB more energy than the gradient-based algorithm, even with the fine grid of only one resource element, because it does not consider the different gradients of E_m caused by the different channel gains G_m and packet sizes N.

In Fig. 4.7, the influence of different channel gains G_m on the required energy E is shown. The $M = 7$ agents are again placed equidistantly, agent $m = 1$ is at $d_{\min} = 5\,\text{m}$, but agent $m = M$ is varied from $d_{\max} = 10\,\text{m}$ to $d_{\max} = 160\,\text{m}$. The agents in between are placed accordingly to keep the equidistant positioning. Since a pure LOS channel is assumed, the channel gains G_m in dB-scale are directly proportional to the distances d_m of the agents to the central controller, which are shown on the axis of abscissae.

The higher $d_{\max}$, the greater is the distance between neighboring agents and, thus, their difference in G_m. The optimal resource allocation has to account for this difference. Since the balancing scheduler only takes the number of resources into account, the energy requirement is up to 3 dB higher than for the gradient based scheduler. This is an interesting result especially for scenarios, where non-line-of-sight propagation leads to largely different channel gains.

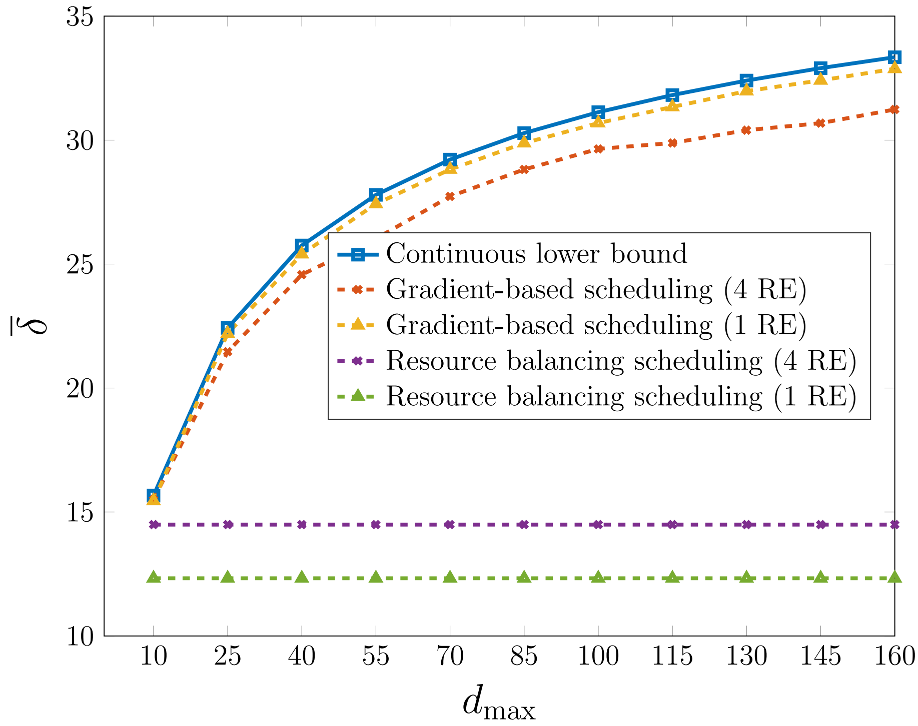

Figure 4.8: Avg. difference of min. and max. n_m for varying max. agent distances $d_{\max}$, resulting in different G_m

The benefit of assigning resources n_m based on the gradient of E, compared to the balancing scheduling for different channel gains G_m, is investigated in Fig. 4.8. The setup is the same as in Fig. 4.7, the maximum distance $d_{\max}$ is again shown as the abscissae.

The difference δ_m, as defined in (4.14), between the largest and the smallest $n_m(r)$ for all agents in a single run r of the experiment, averaged over the total number R of runs with the same $d_{\max}$, is shown as the ordinates.

$$\delta(r) = \max_m n_m(r) - \min_m n_m(r) \tag{4.14}$$

$$\bar{\delta} = \frac{1}{R} \sum_{r=1}^{R} \delta(r) \tag{4.15}$$

The continuous lower bound suggests a larger difference in the assigned n_m is beneficial in terms of energy consumption by assigning more resources to agents with low G_m. For the gradient based scheduler, almost the same $\overline{\delta}$ as the continuous lower bound is attained. Especially for larger $d_{\max}$, the 4-resource element case cannot achieve the results of the finer resource grid, since the adaptation is worse due to the coarser grid. The difference in the distances r_m and, thus, the channel gains G_m is not considered by the balancing scheduler, resulting in the constant average differences.

Now, the number of subcarriers is increased to $N_{\text{sc}} = 12$. First, the required energy E for different packet sizes N is shown in Fig. 4.9, similar to Fig. 4.6. Due to the

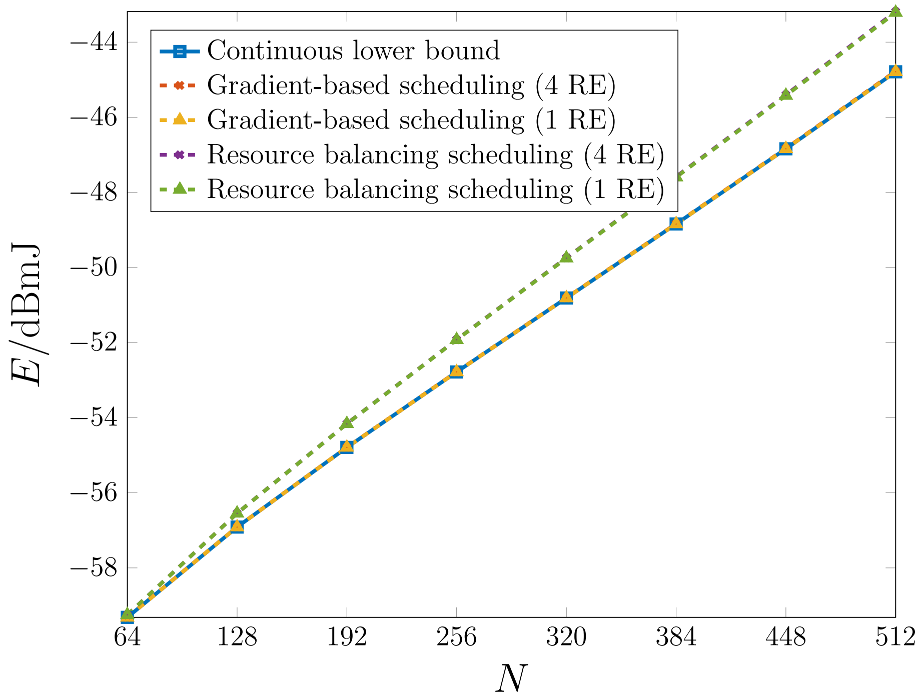

Figure 4.9: Energy for different packet sizes N with $N_{\text{sc}} = 12$

higher number of available resources, the transmit power can be reduced, resulting in a lower total energy consumption. On the other hand, the results of the gradient based scheduler are now even closer to the continuous lower bound, while the balancing scheduler uses about 1.5 dB more energy. The gap between the algorithms shrinks, because the absolute value of the gradient of E_m decreases for n_m close to $n_{\min}$, cf. Fig. 4.3.

Next, the required energy for a varying number of agents M like in Fig. 4.5 was investigated for $N_{\text{sc}} = 12$. The number of bits per packet was set to $N = 64$. The plot Fig. 4.10 shows the results. The granularity of the resource allocation has, like in Fig. 4.9, almost no influence on the required energy E. Compared to Fig. 4.5, the required energy E is lower, while the slightly concave shape of the curves is still visible. The reason for this shape lies partly in (4.9) and (4.10), but also the better exploitation of available resources contributes to it. To illustrate this, an additional curve is displayed in Fig. 4.10, showing the energy E required, if all time-frequency resources

$$n_{\text{tot}} = TB \tag{4.16}$$

are allocated to the agents, i.e.

$$n_{m,\text{fair}} = \frac{n_{\text{tot}}}{M}. \tag{4.17}$$

This allocation is not possible under the constraints, because (4.17) does not consider the individual deadlines τ_m, but rather assumes $\tau_m = T$. With a growing number M of agents, the benefit, i.e. the reduction of the sum energy E, of ignoring the deadlines shrinks. This is due to the better exploitation of the available time-frequency resources, even if the deadlines are considered.

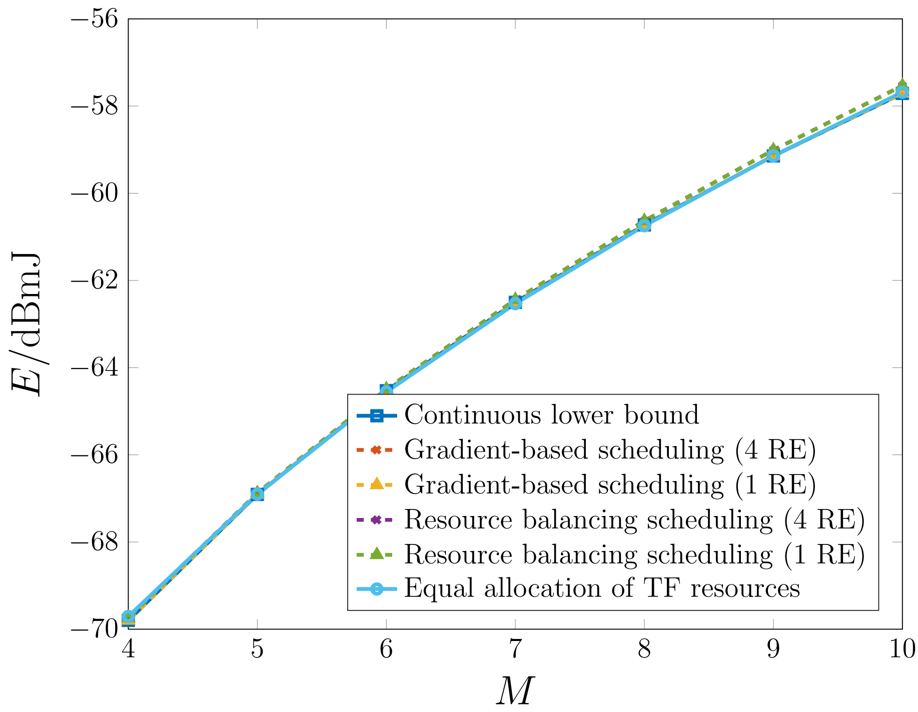

Figure 4.10: Energy for different number M of agents with $N_{\text{sc}} = 12$

The percentage of used resources of the available total resources n_{tot}

$$\%n_{\text{tot}} = 100 \cdot \frac{\sum_{M=1}^{M} n_m}{n_{\text{tot}}} \tag{4.18}$$

is shown as the abscissae in Fig. 4.11. If the number of agents M is increased, the probability to have at least one of them with a late deadline increases. Therefore, in mean, the exploitation of the available time-frequency resources gets better, leading to the increasing values for $\%n_{\mathrm{tot}}$ in Fig. 4.10 for increasing M.

The less agents are competing for the resources, i.e. the smaller M, the smaller gets the sum $\sum_{M=1}^{M} n_{m,\min}$ of resources required to achieve the minimum required energy. Since the balancing allocation scheme does not consider the gradient g_m, more resources than necessary to achieve $n_{m,\min}$ might be allocated to agent m. Hence, the exploitation of available resources is better, but the required energy E is increased due to the curvature of (4.10) shown in Fig. 4.3.

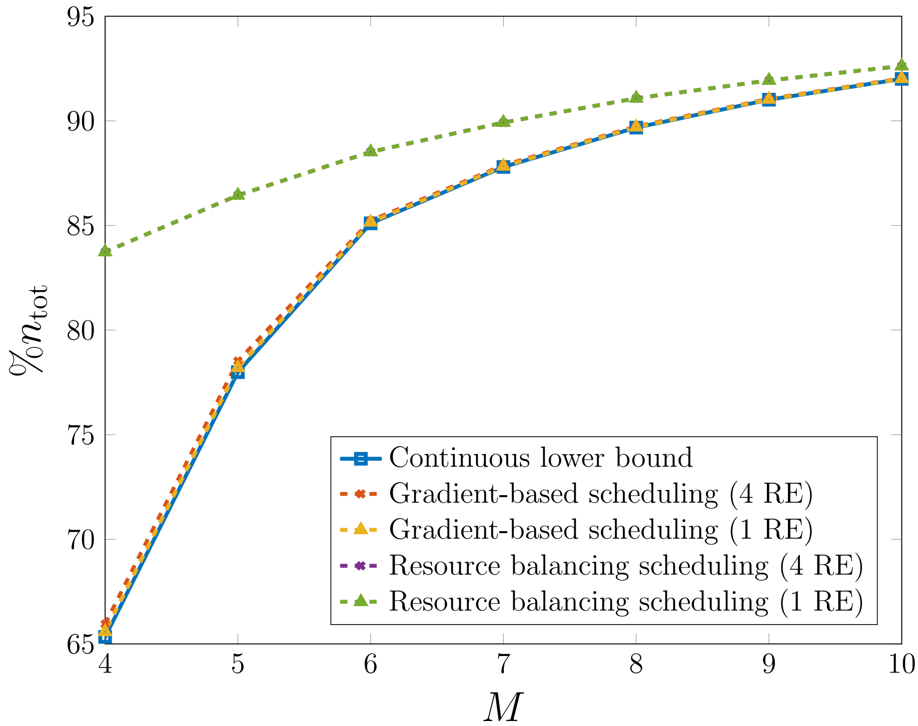

Figure 4.11: Share of resources allocated to agents for number M of agents with $N_{\mathrm{sc}} = 12$

To make this effect even more visible, the number of available resource is now increased to $N_{\mathrm{sc}} = 20$, while M is set to the values from the previous experiments. Now, for $M < 8$ the gap in energy consumption between the balancing and the gradient based algorithm is wider than for the more constrained scenarios with $M > 8$. This effect on the required energy E for $N_{\mathrm{sc}} = 20$ is visible in Fig. 4.12.

The increasing gap between the balancing and the gradient-based schemes already observed in Fig. 4.11 is even more pronounced in Fig. 4.13. While the continuous lower

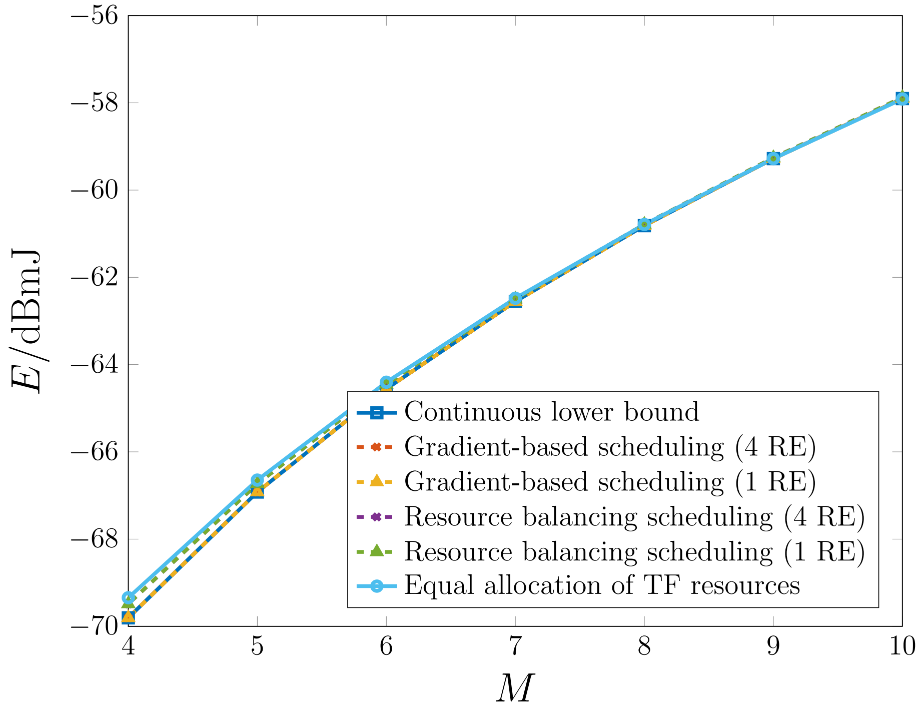

Figure 4.12: Energy for different number M of agents with $N_{\mathrm{sc}} = 20$

bound and the gradient based algorithm reduce the number of allocated resources for a $M < 8$, the balancing algorithm still allocates all resources usable by the agents, therefore, for $M < 8$, the sum $n_{\mathrm{tot}} > \sum_{M=1}^{M} n_{m,\mathrm{min}}$ of allocated time-frequency resources is larger than the sum of the minimal energy achieving allocations $n_{m,\mathrm{min}}$, which results in the increased energy consumption E shown in Fig. 4.12.

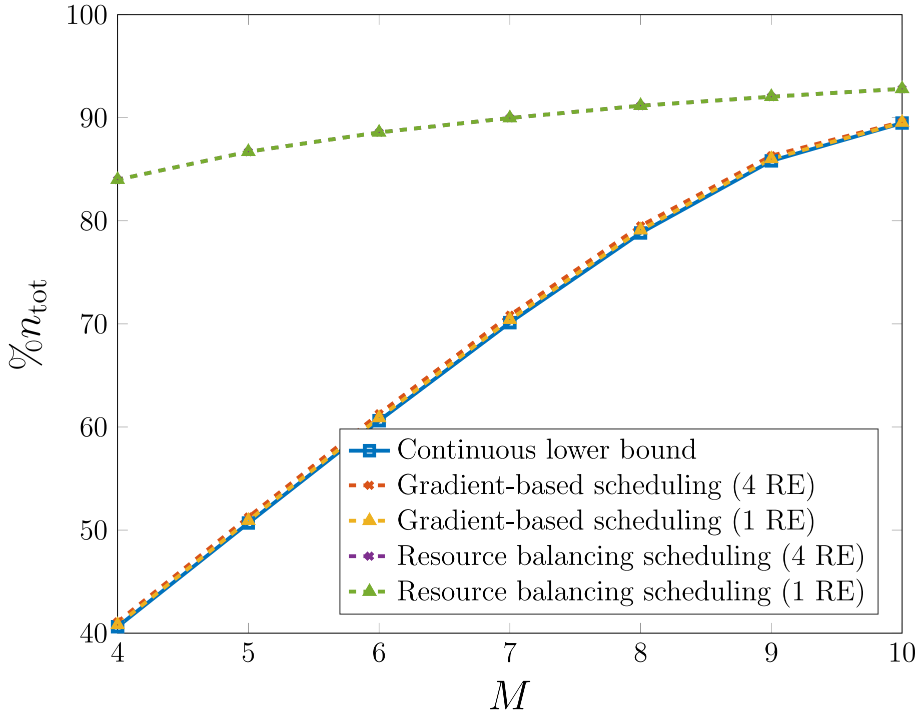

Figure 4.13: Share of resources allocated to agents for number M of agents with $N_{\text{sc}} = 20$

4.7 Conclusion

In this chapter, the time-frequency resource allocation for a single central controller transmitting control commands to multiple agents, was optimized for minimum energy consumption. The agents needed to receive the control commands before an individual deadline. Since control commands generally are of small size, the problem was stated based on an adapted version of the Shannon capacity formula for short packets. The resulting continuous minimization problem was shown to be convex.

For application in mobile radio systems like 5G, due to the OFDM scheme applied, the resource allocation has to be done based on fixed size resource elements, turning the problem into a mixed integer problem. An algorithm to find a scheduling of these resource elements based on the gradient of the required transmit energy was proposed and compared to a simple resource balancing algorithm only considering the deadlines in terms of the required total transmit energy and the amount of time-frequency resources allocated to the individual agents.

The resource blocks were configured to occupy either 1 Hz·s or 4 Hz·s in the time-frequency plane. The gradient-based algorithm was shown to perform only about

0.3 dB worse in terms of required energy than the continuous optimum and showed improvements of more than 50% to the balancing algorithm, when blocks of 1 Hz·s were allocated at once, especially if the channel gains among the agents are very different. If 4 Hz·s are allocated at once, the gradient-based algorithm performed about 1.25 dB worse than the lower bound, while the balancing algorithm again required about double the energy compared to the gradient-based.

Furthermore, the ability of the gradient-based scheduling algorithm to adapt to different requirements due to different deadlines was demonstrated by investigating the average difference between the largest and the smallest resource allocation to the agents. Especially for a high difference in channel gains the difference in resource allocations is large. Since the balancing scheduling algorithm does not consider channel gains, the difference stays constant for all channel gain configurations.

Another important aspect shown is the ability of the proposed algorithm to reduce the total number of allocated resource elements, compared to the balancing scheme, especially for the larger resource block configurations of 4 Hz·s. This larger allocation of the balancing scheme leads also to a higher noise energy at the receivers at the agents, which has to be compensated by a higher transmit energy for this agent, resulting in the observed increase of the total transmit energy.

Chapter 5

Conclusions

5.1 Summary

In this thesis, three important aspects of WNCS were investigated and existing methods were improved. First, the measurement and quantization of sensor values to measure the state of the control system was investigated. Then, the scheduling of the transmission of the sensor values to the controller was considered. Finally, the resource allocation for transmitting control commands from a central controller to distributed agents was optimized. A closed loop control system generally consists of one or more loops, each of them including sensors, controllers, and plants. Each of these components can be located remotely from its predecessor and/or successor, which imposes the requirement to transmit information. Since the resources for wireless communication are limited, an optimum utilization of this scarce resource is pursued.

In Chapter 2, the acquisition of sensor values and their transmission to the controller were optimized in terms of needed energy and time. There are two parameters available for adjustment, which were shown to have a tradeoff. First, there is the number of quantization intervals, determining the resolution of the transmitted values. Second, the number of measurements taken by a sensor impaired by measurement noise before quantization. A higher number of measurements allows for, under the assumption the noise of each measurement is uncorrelated, an averaging and therefore reduction of the noise. Each of those tasks consumes time and energy, which are generally limited. Hence, a pareto-optimal solution for minimum Bayes risk was shown to exist. For the known properties of the quantity measured by the sensor, determined by the a-priori distribution of the parameter and the number of measurements, a quantization scheme tailored to the random distribution of possible sensor values was developed to improve the utility of each transmitted bit. Therefore, the mutual information between the source, the sensed value of interest in this case, and the sink, the estimator at the controller, was maximized. The resulting estimation error at the controller was evaluated in terms of the Bayes risk for the mean-absolute and the mean-squared error. Finally, it was shown that this tradeoff has a pareto-optimum with a certain combination of the number of quantization steps and the number of measurements, if either the available total time or total energy for the acquisition and transmission are limited. Using the presented quantization scheme, a reduction of required bits of up to 20% for the same Bayes risk, which was used as a measure for the quality of the estimate, could be achieved on the considered a-priori distributions.

In Chapter 3, the scheduling of sensor value transmissions of a control system with multiple discrete time linear subsystems, each having a scalar system state and a dead-beat control scheme was investigated. The subsystems were assumed to accumulate Gaussian system noise. To achieve the control goal of steering each subsystem to the equilibrium, a central controller calculates a control input to each subsystem. To derive the correct control input, updates about the current subsystem states have to be sent to the central controller. The resources for transmission of these updates were limited, so it was not possible to send an update for each subsystem in each time slot. We have shown, that a periodic updating scheme is optimal to minimize the deviation of the subsystems from the equilibrium over time. Furthermore, we stated an optimization problem to find the optimum update rate for each subsystem based on its system constant and the available transmission resources. We also considered the influence of the different system constants on the behaviour of the uncertainty about each subsystem state in the future. This optimization problem was shown to be convex. Finally, we developed an algorithm to apply the calculated rates. Since the update rates do not necessarily translate to an integer number of time steps between consecutive updates, the resulting error is distributed equally by the scheduling algorithm among the subsystems. This scheme also reduces fluctuations of the estimation error covariance over time. Combining the results from the continuous optimization and the algorithm for the application of the calculated update rates, we see a reduction of up to 20% in the mean absolute error over all subsystems, compared to the reference scheme from prior work, which did only work on the current uncertainty about the subsystem states and did not include the future development. We could also reduce the variation of this mean absolute error of the runtime of the system by about 60% compared to the reference scheme.

In Chapter 4 the timely transmission of control commands from the controller to the actuator was investigated. The small amounts of data to be transmitted, as well as the required low latency between control input generation at the controller and the application of this input at the actuator, a modified version of the well-known Shannon channel capacity formula, which is adapted to data packets of short finite length, was used. A command of a certain number of bits had to be transmitted to each agent before an individual deadline. To achieve a minimal overall energy consumption for the transmission, an optimization problem to allocate time-frequency resources to the agents was derived. This problem was shown to be convex, which allows for an efficient solution using available optimization frameworks. Additionally, a energy-gradient-based algorithm to assign blocks of time-frequency resources in an

OFDM fashion was developed to tackle the mixed-integer problem resulting from the resource block structure in OFDM. The results of this algorithm were compared to the results acquired by solving the continuous optimization problem in terms of required total transmit energy. For further comparison, a simplistic algorithm, which only balances the time-frequency resources allocated to the agents was developed. The gradient-based algorithm was shown to achieve results close to the continuous solution derived from the convex optimization problem, while the resource-balancing algorithm required a much higher energy for transmission. The reduction of required energy for the balancing scheme was shown to be up to 3 dB or halving the required amount of energy. The gradient-based algorithm also considers the optimum amount of time-frequency resources, which is especially required for short packages. This is due to the fact that in short package communication more time-frequency resource also increase the noise power, which results in an optimum number of time frequency resource for a given packet size. In contrast to the classic Shannon formula, the required energy for transmission rises, if too many time-frequency resources are allocated.

5.2 Outlook

Finally, we want to give some further research directions. For the state estimation discussed in Chapter 2, independent realizations of the random variable were considered. However, in control systems, possible trajectories are at least partly known, so including this a priori knowledge can further help in reducing the required quantization intervals or improving the estimation performance. Furthermore, the communication energy model can be improved by applying the short packet formula from [PPV10] to get a tighter approximation of the required transmission energy, but also introduce possible transmission errors.

The simple linear subsystems in Chapter 3 do not include any measurement noise, which would make the scheduling more challenging. The value of information cannot be calculated in closed form anymore, because it is not reset to zero after a single transmission. The subsystems themselves can be extended to feature multidimensional states. This also allows for a model considering only partial observability. Furthermore, transmission error probabilities can make the model even more realistic.

The channel model in Chapter 4 does not consider channel fading or frequency selectivity. This would increase the complexity of the allocation algorithm, since the channel coefficient is an additional parameter for the allocation of resource elements besides the timeliness, which guarantees the transmission before the deadline, and the gradient of the transmission energy. Even more energy could be saved, if a retransmission scheme was implemented, which can first try with a lower transmit energy and only use a higher energy for a possible retransmission.

List of Acronyms

AWGN	Additive White Gaussian Noise
cdf	cumulative distribution function
GNSS	Global Navigation Satellite System
IIoT	Industrial Internet of Things
IoT	Internet of Things
LOS	line-of-sight
LTI	linear time-invariant
MAVE	minimum absolute-value error
MDP	Markov Decision Process
MIMO	multiple-input multiple-output
MMSE	minimum mean-square error
NR	New Radio
OFDM	Orthogonal Frequency Division Multiplex
pdf	probability-density function
PSD	Power-Spectral-Density
SNR	signal-to-noise ratio
URLLC	Ultra-Reliable Low Latency Communication
VoI	Value of Information
WNCS	Wireless Networked Control Systems

List of Mathematical Symbols

$\boldsymbol{A}^{\mathrm{T}}$	Transpose of matrix $\boldsymbol{A}$
$\boldsymbol{A}^{\mathrm{H}}$	Hermitian transpose of matrix $\boldsymbol{A}$
$\lvert\cdot\rvert$	$l2$-norm
$\mathrm{Var}\{\cdot\}$	Variance of a random variable
$\mathbb{E}\{\cdot\}$	Expected Value of a random variable
$I\left(X;Y\right)$	Mutual information of the random variables X and Y
$L_{y'}(x)$	Likelihood function of the parameter x for the observation y'
$L((x),\boldsymbol{\mu})$	Lagrangian function for argument $\boldsymbol{x}$ with Lagrange multipliers $\boldsymbol{\mu}$
$p_X(x)$	pdf of the random variable X
$p_{X,Y}(x,y)$	Joint pdf of the random variables X and Y
$p_{Y\lvert X}(y,x)$	Conditional density function of the random variable Y for a given $X=x$
$\mathbf{1}$	Vector of ones
$Q\left(\cdot\right)$	Gaussian Q-function
$Q^{-1}\left(\cdot\right)$	Inverse of the Gaussian Q-function
$\mathbb{R}$	Set of real numbers
$\mathbb{R}^{+}$	Set of positive real numbers
$\mathbb{R}^{m\times n}$	Set of matrices with real-numbered elements with m rows and n columns
$x(a\lvert b)$	Estimate of quantity x at time instant a with the knowledge from time instant b, $b\leq a$
$\lfloor x\rfloor$	Greatest integer less than or equal to x

List of Variables from Chapter 2

C	Communication channel capacity
E_{acq}	Total energy required for acquisition
E_{max}	Acquisition and transmission energy limit
E_{meas}	Energy required for a single measurement
E_{quant}	Energy required per quantization step
E_{tx}	Energy required for transmission
h	Communication channel coefficient
m_n	Measurement noise of the nth measurement of w
N_{meas}	Number of measurements
N_{bits}	Number of quantization bits
P	Transmit power
Q_{quant}	Number of quantization intervals
R_{B}	Bayes risk
$R_{\mathrm{B,\ AVE}}$	Bayes risk of the MAVE estimator
$R_{\mathrm{B,\ MSE}}$	Bayes risk of the MMSE estimator
s	Aggregated value derived from the measurements x_n
T_{max}	Acquisition and transmission time limit
T_{bit}	Time required to transmit a single bit
T_{tx}	Time required for transmission
T_{meas}	Time required for a single measurement
T_{acq}	Total time reqired for acquisition
v	Decoder output generated from y'
w	Value of interest
$\hat{w}$	Estimated value of w generated by Ψ
w_{max}	Upper bound of the value w of interest
w_{min}	Lower bound of the value w of interest
x_n	nth noisy measurement of w
y	Index of the quantization interval selected for s, transmitted over the wireless communication channel
y'	The symbol at receiver of the wireless communication channel
z	Receiver noise at the receiver of the wireless communication channel
γ_{E}	Ratio of E_{quant} to E_{meas}
γ_{T}	Ratio of T_{bit} to T_{meas}

σ_{M}	Measurement noise power
σ_{Z}	Communication noise power
ϕ_{Q}	Step function with Q_{quant} steps for quantization of s
Ψ	Estimator to estimate $\hat{w}$ based on v

List of Variables from Chapter 3

a_i	System coefficient of the ith subsystem
$\mathcal{C}_i$	Controller of the ith subsystem
$e_i(k)$	Estimation error of the ith subsystem at time instant k
$f_i(d)$	Function to calculate the sum of variances P_i of the ith subsystem after d timeslots without transmission
$g_i(k)$	Kalman gain of the ith subsystem at time instant k
$J(k)$	Cost function
N_{sys}	Number of subsystems
$\mathcal{P}_i$	Plant of the ith subsystem
$P_i(k)$	Covariance matrix of the estimation error $e_i(k)$
R_{DL}	Number of downlink resources
R_{UL}	Number of uplink resources
R_{remain}	Remaining resources for this time slot(in algorithm)
$\mathcal{S}_i$	Sensor of the ith subsystem
T_{sim}	Simulation total time
T_{i}^{s}	Number of communication time slots per system time slot
$u_i(k)$	Control variable of the ith subsystem at time instant k
V_i	Measurement noise covariance of the ith subsystem
$v_i(k)$	Measurement noise of the ith subsystem at time instant k
W_i	System noise covariance of the ith subsystem
$w_i(k)$	System noise of the ith subsystem at time instant k
$x_i(k)$	Scalar system state of the ith subsystem at time instant k
$\hat{x}_i(k)$	MMSE estimate of $x_i(k)$ based on $y_i(k)$
y_i	Measurement from the ith subsystem
α	Spreading factor of the subsystem coefficient classes
β	Ratio of the number of subsystems with higher dynamics to the subsystems with lower dynamics
$\boldsymbol{\mu}$	Lagrange multipliers
$\boldsymbol{\pi}_{\text{DL}(t)}$	Downlink scheduling in time slot t
$\boldsymbol{\pi}_{\text{UL}(t)}$	Uplink scheduling in time slot t
$\boldsymbol{\pi}_{\text{DL}}$	Downlink scheduling
$\boldsymbol{\pi}_{\text{UL}}$	Uplink scheduling
Σ_{e}	Mean absolute estimation error

List of Variables from Chapter 4

$\boldsymbol{A}$	Three-dimensional array of time-frequency resource allocations
B	Total available transmission bandwidth
b_{sc}	Bandwidth of one subcarrier
$b'_m(t)$	Bandwidth assigned to agent m at time instant t
$\boldsymbol{C}$	Matrix for storing the level of competition for each time-frequency resource element
C_m	Shannon channel capacity for agent m
C_{corr}	Channel capacity corrected for short packets
$C_{\mathrm{corr},m}$	Channel capacity of agent m corrected for short packets
$d_{\max}$	Maximum distance of agents from central controller
$d_{\min}$	Minimum distance of agents from central controller
E_m	Total transmission energy for agent m
G_m	Channel gain of agent m
g_m	Derivative of the required transmit energy w.r.t. n_m for agent m
M	Number of agents
N	Number of bits in a single command
N_0	Receiver noise PSD
N_{sc}	Number of subcarriers
N_{sym}	Number of OFDM symbols
n_m	Number of time-frequency resources of agent m
$n_{\min}$	Optimum numbers of time-frequency resource elements without resource limits
$n_{m,\min}$	Optimum number of time-frequency resource elements for agent m without resource limits
$n_{m,\mathrm{opt}}$	Optimum number of time-frequency resource elements for agent m
n_{thr}	Inflection point, where energy function changes from convex to concave
$n_{m,\mathrm{thr}}$	Inflection point of agent m, where energy function changes from convex to concave
n_{tot}	Total available time-frequency resource elements
q_m	PSD of the transmit power for agent m
p_c	Maximum allowable probability of a lost control command
q_m	Transmit PSD of agent m
s	Subcarrier index
T	Control time slot duration

t	Time elapsed since the beginning of the control time slot
t_{sym}	OFDM symbol duration
V_m	Channel dispersion for agent m
γ_m	Receiver SNR for agent m
γ_m	SNR at agent m
σ_m^2	Receiver noise power for agent m
τ_m	Deadline for agent m

Bibliography

[3GP15] 3GPP Technical Report 22.885, "Study on LTE support for Vehicle to Everything (V2X) services (Release14)," ETSI, Tech. Rep., 2015, v14.0.0.

[3GP19] 3GPP Technical Report 38.824, "Study on physical layer enhancements for NR ultra-reliable and low latency case (URLLC) (Release16)," ETSI, Tech. Rep., 2019, v16.0.0.

[3GP20] 3GPP Technical Report 38.912, "Study on New Radio (NR) access technology (Release16)," ETSI, Tech. Rep., 2020, v16.0.0.

[3GP21] 3GPP Technical Report 38.211, "Physical channels and modulation (Release16)," ETSI, Tech. Rep., 2021, v16.0.0.

[AVK+19] O. Ayan, M. Vilgelm, M. Klügel, S. Hirche, and W. Kellerer, "Age-of-information vs. value-of-information scheduling for cellular networked control systems," in *Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems*, ser. ICCPS '19. ACM, 2019, pp. 109–117.

[AVK20] O. Ayan, M. Vilgelm, and W. Kellerer, "Optimal scheduling for discounted age penalty minimization in multi-loop networked control," in *2020 IEEE 17th Annual Consumer Communications Networking Conference (CCNC)*, 2020, pp. 1–7.

[BDH+20] Y. Bian, J. Ding, M. Hu, Q. Xu, J. Wang, and K. Li, "An advanced lane-keeping assistance system with switchable assistance modes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, pp. 385–396, 01 2020.

[BHCW18] H. Boyes, B. Hallaq, J. Cunningham, and T. Watson, "The industrial internet of things (iiot): An analysis framework," *Computers in Industry*, vol. 101, p. 1–12, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0166361517307285

[Cis19] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2017–2022," Tech. Rep. C11-738429-01, 2019.

[Cis20] ——, "Cisco Annual Internet Report (2018-2023) White Paper," Tech. Rep. C11-741490, March 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[CL16] S. Cai and V. K. N. Lau, "MIMO Precoding for Networked Control Systems with Energy Harvesting Sensors," *IEEE Transactions on Signal Processing*, vol. 64, no. 17, pp. 4469–4478, Sep 2016.

[dIS+20] F. E. da Silva, A. L. V. Iaremczuk, R. D. Souza, G. Brante, G. L. Moritz, and S. Hussain, "Hybrid ARQ in Wireless Packetized Predictive Control," *IEEE Sensors Letters*, vol. 4, no. 5, pp. 1–4, 2020.

[GB14] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar 2014.

[GRC+14] L. A. Grieco, A. Rizzo, S. Colucci, S. Sicari, G. Piro, D. Di Paola, and G. Boggia, "IoT-aided robotics applications: Technological implications, target domains and open issues," *Computer Communications*, vol. 54, pp. 32–47, 2014.

[GYH17] X. Ge, F. Yang, and Q.-L. Han, "Distributed networked control systems: A brief overview," *Information Sciences*, vol. 380, pp. 117 – 131, 2017.

[HYE16] Q. He, D. Yuan, and A. Ephremides, "Optimizing freshness of information: On minimum age link scheduling in wireless systems," in *2016 14th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, May 2016, pp. 1–8.

[IEE20] *IEEE Standard for Low-Rate Wireless Networks*, IEEE Std. 802.15.4, 2020.

[ISO15] "Road vehicles — Controller area network (CAN) — Part 1: Data link layer and physical signalling," International Organization for Standardization, Geneva, CH, Standard, Mar. 2015.

[Kal60] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[KRZ+20] W. Khan, M. Rehman, H. Zangoti, M. Afzal, N. Armi, and K. Salah, "Industrial Internet of Things: Recent advances, enabling technologies and open challenges," *Computers & Electrical Engineering*, vol. 81, p. 106522, 2020.

[KYY+17] S. Kurt, H. U. Yildiz, M. Yigit, B. Tavli, and V. C. Gungor, "Packet Size Optimization in Wireless Sensor Networks for Smart Grid Applications," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 3, pp. 2392–2401, Mar 2017.

[LMZ+16] S. Lin, F. Miao, J. Zhang, G. Zhou, L. Gu, T. He, J. A. Stankovic, S. Son, and G. J. Pappas, "ATPC: Adaptive transmission power control for wireless sensor networks," *ACM Trans. Sen. Netw.*, vol. 12, no. 1, pp. 6:1–6:31, Mar 2016. [Online]. Available: http://doi.acm.org/10.1145/2746342

[LNL+21] W. Liu, G. Nair, Y. Li, D. Nesic, B. Vucetic, and H. V. Poor, "On the Latency, Rate, and Reliability Tradeoff in Wireless Networked Control Systems for IIoT," *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 723–733, 2021.

[Lun16a] J. Lunze, *Regelungstechnik 2: Mehrgrößensysteme, Digitale Regelung*, 9th ed. Berlin: Springer Vieweg, 2016, ch. 2.

[Lun16b] ——, *Regelungstechnik 2: Mehrgrößensysteme, Digitale Regelung*, 9th ed. Berlin: Springer Vieweg, 2016, ch. 7.

[MGW+19] Y. Ma, J. Guo, Y. Wang, A. Chakrabarty, H. Ahn, P. Orlik, and C. Lu, “Optimal dynamic scheduling of wireless networked control systems,” in *Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems*, ser. ICCPS ’19. ACM, 2019, pp. 77–86.

[PBB18] J. Ploennigs, A. Ba, and M. Barry, “Materializing the Promises of Cognitive IoT: How Cognitive Buildings Are Shaping the Way,” *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2367–2374, Aug 2018.

[PFL+16] S. Pandi, F. H. P. Fitzek, C. Lehmann, D. Nophut, D. Kiss, V. Kovacs, A. Nagy, G. Csorvasi, M. Toth, T. Rajacsis, H. Charaf, and R. Liebhart, “Joint Design of Communication and Control for Connected Cars in 5G Communication Systems,” in *Proc. IEEE Globecom Workshops*, Dec 2016, pp. 1–7.

[PP02] A. Papoulis and S. U. Pillai, “Probability, random variables, and stochastic processes,” Boston [u.a.], 2002. [Online]. Available: http://scans.hebis.de/HEBCGI/show.pl?10203526_toc.pdf

[PPV10] Y. Polyanskiy, H. V. Poor, and S. Verdu, “Channel Coding Rate in the Finite Blocklength Regime,” *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010.

[PS02] J. G. Proakis and M. Salehi, “Communication Systems Engineering,” Upper Saddle River, NJ, 2002.

[PYZ+20] B. Peng, D. Yu, H. Zhou, X. Xiao, and Y. Fang, “A platoon control strategy for autonomous vehicles based on sliding-mode control theory,” *IEEE Access*, vol. 8, pp. 81 776–81 788, 2020.

[Sha48] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, pp. 379–423, 1948. [Online]. Available: http://plan9.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf

[Shy12] J. Shynk, *Probability, Random Variables, and Random Processes: Theory and Signal Processing Applications.* Wiley, 2012.

[SKMN15] M. Stein, A. Kürzl, A. Mezghani, and J. Nossek, “Asymptotic Parameter Tracking Performance With Measurement Data of 1-Bit Resolution,” *IEEE Transactions on Signal Processing*, vol. 63, no. 22, pp. 6086–6095, Nov 2015.

[SMK+17] P. Schulz, M. Matthe, H. Klessig, M. Simsek, G. Fettweis, J. Ansari, S. A. Ashraf, B. Almeroth, J. Voigt, I. Riedel, A. Puschmann, A. Mitschele-Thiel, M. Muller, T. Elste, and M. Windisch, “Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture,” *IEEE Communications Magazine*, vol. 55, no. 2, pp. 70–78, February 2017.

[SSY+19] C. Sun, C. She, C. Yang, T. Q. S. Quek, Y. Li, and B. Vucetic, "Optimizing Resource Allocation in the Short Blocklength Regime for Ultra-Reliable and Low-Latency Communications," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 402–415, 2019.

[SYQ17] C. She, C. Yang, and T. Q. S. Quek, "Radio Resource Management for Ultra-Reliable and Low-Latency Communications," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 72–78, Jun 2017.

[TC03] Y. Tipsuwan and M.-Y. Chow, "Control methodologies in networked control systems," *Control Engineering Practice*, vol. 11, no. 10, pp. 1099 – 1111, 2003, special Section on Control Methods for Telecommunication.

[VMKH16] M. Vilgelm, M. H. Mamduhi, W. Kellerer, and S. Hirche, "Adaptive decentralized MAC for event-triggered networked control systems," in *Proc. of the 19th International Conference on Hybrid Systems: Computation and Control*. ACM, 2016, pp. 165–174.

[WHZ18] J. Wen, L. He, and F. Zhu, "Swarm Robotics Control and Communications: Imminent Challenges for Next Generation Smart Logistics," *IEEE Communications Magazine*, vol. 56, no. 7, pp. 102–107, Jul 2018.

[WQQ20] Y. Wu, D. Qiao, and H. Qian, "Efficient Bandwidth Allocation for URLLC in Frequency-Selective Fading Channels," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[WSJ17] M. Wollschlaeger, T. Sauter, and J. Jasperneite, "The Future of Industrial Communication: Automation Networks in the Era of the Internet of Things and Industry 4.0," *IEEE Industrial Electronics Magazine*, vol. 11, no. 1, pp. 17–27, Mar 2017.

[XMH19] N. Xuan Mung and S.-K. Hong, "Improved altitude control algorithm for quadcopter unmanned aerial vehicles," *Applied Sciences*, vol. 9, 05 2019.

[YSL13] H. Yu, Z. Shen, and C. Leung, "From Internet of Things to Internet of Agents," in *Proc. IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, Aug 2013, pp. 1054–1057.

[ZCWF18] M. Zanon, T. Charalambous, H. Wymeersch, and P. Falcone, "Optimal Scheduling of Downlink Communication for a Multi-Agent System With a Central Observation Post," *IEEE Control Systems Letters*, vol. 2, no. 1, pp. 37–42, Jan 2018.

Author's Publications

[KASK19] K. Kiekenap, H. Al-Shatri, and A. Klein, "Trade-Off Between Measurement Accuracy and Quantization Precision for Minimum Bayes Risk in Wireless Networked Control Systems," in *SCC 2019; 12th International ITG Conference on Systems, Communications and Coding*, 2019, pp. 1–6.

[KK20] K. Kiekenap and A. Klein, "Optimum Sensor Value Transmission Scheduling for Linear Wireless Networked Control Systems," in *2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall)*, 2020, pp. 1–5, © 2020 IEEE.

[KOK21] K. Kiekenap, A. Ortiz, and A. Klein, "Energy-Optimal Short Packet Transmission for Time-Critical Control," in *2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall)*, 2021, pp. 01–06, © 2021 IEEE.

[TASKK19] A. Tahir, H. Al-Shatri, K. Kiekenap, and A. Klein, "Ultra-Reliable Low Latency Communication for Consensus Control in Multi-Agent Systems," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, 2019, pp. 1–7, © 2019 IEEE.

Supervised Student Theses

Name	Title of the thesis	Thesis type	Date
Aslam, Muhammad Uzair	Behavioral Analysis of Adaptive Event-Triggered MAC in Networked Control Systems using the Slotted ALOHA Protocol and employing a Linear Control Model	Master thesis	04/2019
Hott, Ruben	Sensor Localization for UAV Data Collection using Artificial Neural Networks in Large-Scale Wireless Sensor Networks	Bachelor Thesis	05/2019
Hirsch, Daniel	Weighted Factor Based Iterative MMSE Filter-Design in Partially Connected Two-Way Relaying Networks	Bachelor Thesis	08/2019
Brand, Martin	Optimizing Sensor Value Transmission in Wireless Networked Control System	Bachelor Thesis	09/2020

Erklärungen laut Promotionsordnung

§8 Abs. 1 lit. c PromO
Ich versichere hiermit, dass die elektronische Version meiner Dissertation mit der schriftlichen Version übereinstimmt.

§8 Abs. 1 lit. d PromO
Ich versichere hiermit, dass zu einem vorherigen Zeitpunkt noch keine Promotion versucht wurde. In diesem Fall sind nähere Angaben über Zeitpunkt, Hochschule, Dissertationsthema und Ergebnis dieses Versuchs mitzuteilen.

§9 Abs. 1 PromO
Ich versichere hiermit, dass die vorliegende Dissertation selbstständig und nur unter Verwendung der angegebenen Quellen verfasst wurde.

§9 Abs. 2 PromO
Die Arbeit hat bisher noch nicht zu Prüfungszwecken gedient.

Datum und Unterschrift