Resource Allocation for Ultra-Reliable Low-Latency Communication (URLLC) in Future IoT Networks

Niall Donohoe
Student ID: 17340571

The thesis is submitted to University College Dublin in part fulfillment of the requirements for the degree of Master of

Electronic and Computer Engineering

Supervisor: Professor Mark Flanagan

UCD School of Electrical and Electronic Engineering

University College Dublin

April 2022

GRADUATE MINOR THESIS SUBMISSION FORM

At least two copies of this form should be completed (typed) by the student and signed and submitted together with the requisite number of copies of the thesis to the relevant School Office - students should check with their School Office to confirm the precise number of copies required. One copy of this form is to be retained by the School and one copy returned to the student, signed and dated by the relevant administrator in the School Office to confirm that the thesis has been received.

Student Name: Niall Donohoe

Student Number: 17340571

Programme: ME Electronic and Computer

Module Code and Title: EEEN40240

Supervisor: Professor Mark Flanagan

Personal Information

TITLE OF THESIS: Resource Allocation for Ultra-Reliable Low-Latency Communication (URLLC) in Future IoT Networks

---

To be completed by the Student:

I hereby certify that the submitted work is my own work, was completed while registered as a candidate for the degree stated above, and I have not obtained a degree elsewhere on the basis of the research presented in this submitted work. Where the submitted work is based on work done by myself jointly with others, I certify that a substantial part is my own original work, the extent of which is indicated in the collaborations page of the submitted work.

Student Name (printed): Niall Donohoe

Student Signature: Niall Donohoe

Date: 14/04/2022

To be completed by the School Office:

Number of copies received:

Received by (print name):

Date:

Appendix D: Hazard Identification / Risk Analysis

SCHOOL OF ELECTRICAL & ELECTRONIC ENGINEERING

ME ELECTRICAL POWER ENGINEERING
or ME ELECTRONIC & COMPUTER ENGINEERING

Please submit this signed form through Brightspace within **2 weeks** of your project being assigned.

At the end of the project, you must include a copy of this form in your final report. Note that **you may need to update the form** if new circumstances arise relating to safety during the course of the project.

Scan and email the original form to yourself with the Subject line "Hazard identification Risk Analysis Stamped", for easy retrieval at the end of the Spring trimester

Personal Information

Student Name:	Niall Donohoe
Student Number:	17340571
Supervisor:	Prof. Mark Flanagan
Project Title:	Resource Allocation for Secure URLLC in Future IoT Networks
Programme:	ME Electrical Power or Electronic & Computer Engineering

Identify any potential safety hazards for your project, e.g. dangerous voltages, rotating machinery, radiation, soldering fumes, sitting in front of a computer for prolonged periods

The ergonomic and comfort risks associated with prolonged computer work, and eyestrain.

Covid-19 is also a major risk, especially when conducting in person meetings or while working in the projects laboratory.

Indicate how all identified safety hazards will be managed to provide a safe working environment

To address prolonged periods working on the computer, and eyestrain, regular breaks will be taken
to ensure that discomfort is eliminated.

In relation to Covid-19 in person meetings will require facemasks, and social distancing, further aided
by good ventilation by opening doors and windows. Should any member of the team feel unwell the
meeting should be conducted online.

Indicate procedures in place for on-campus activities relating to the project, particularly in relation to unsupervised access to the project laboratory
When on-campus all UCD rules and regulations will be observed. Masks will be worn in the Project Laboratory. Care will be taken to minimise disruption to other students in the project Laboratory, especially when coming and leaving the Lab. If there are any issues with equipment I am to alert the Project Laboratory facilitators. There will be no eating, drinking or playing games in the Project Laboratory.
Indicate how project activities and final objectives will be revised, if appropriate, if there is an extended lockdown period on the UCD campus
In the case of an extended lockdown, meetings will be held online. Depending on access rules to the project lab work may have to retrieved such that project activities can be continued from home. In the case of a extended lockdown, in person project related milestones may also have to be conducted online, this includes the week 11 presentation and the project oral examination.

Office Use Only

Date and Time Received:	Received by:
[School Stamp]	Document Tracking
	Lecturer:

Table of Contents

Abstract

Ultra-reliable, low-latency communication (URLLC), is one of three service categories introduced by 3GPP for 5G and beyond. In order to meet the strict latency requirements of 1ms or less, packet length must be short. This presents a major challenge as fundamental information theory has shown that in order to achieve high reliability (low error probability) infinitely long blocklength is required.

As a consequence Shannon's capacity expression no longer holds true for short packet, finite blocklength communications. In order to gain insight and develop methods to tackle these challenges facing URLLC communication, a factory automation scenario is considered. A base station must deliver a packet each to two actuators. Under this model expressions are derived to aid investigation of blocklength and power allocation between multiple users in an orthogonal multiple access (OMA) downlink system.

Non-orthogonal multiple access (NOMA) has received much attention of late for its characteristics which are suggested to be suitable for URLLC applications. Following this NOMA expressions are derived and under the same system model OMA and NOMA are compared.

Algorithms are then proposed to allocate resources such that sum throughput is maximised. To conclude numerical simulation results show the performance of the algorithms proposed and in which base station (BS) transmit power regions they have significant impact.

The principal results of this thesis show that NOMA has desirable characteristics for URLLC applications, including higher sum throughput and also lower latency, when compared to OMA. The algorithms proposed are compared with equal allocation and exhaustive search allocation optimization methods. The simulation results show that the algorithms proposed achieve performance very close to global optimum.

Based off the simulations some conclusions are made as regards when there is a performance gain to be had by optimization and when optimization yields little performance gain. Some areas for future work are also suggested as URLLC and finite blocklength communication is still in its infancy and requires significant work before it will be practically applicable.

Lay Abstract

Future generations of technologies will undoubtedly have higher data rate requirements as they aim to provide more rich and immersive experiences for users. Furthermore IoT is revolutionising the way manufacturing and industry is monitored and controlled. Even in our cities we are moving to smart, connected cities which intelligently manage traffic, healthcare innovation brings remote patient monitoring, activity tracking wearables and remote surgeries. All of these technological advances share a limited resource. That resource is the network access that will enable all these functions and experiences to occur. In this thesis we investigate how we can ensure high-reliability and also low-latency for these future IoT networks.

The fundamental challenge for ultra-reliable low-latency communication (URLLC) is that the two requirements of low-latency and high reliability are competing objectives. Within this thesis novel and efficient techniques to achieve high reliability and low latency communications are analysed and proposed. Simulations show how these algorithms perform in a practical IoT scenario and from this we provide some concluding guidelines on the practical and real world application of the proposed algorithms.

Chapter 1: Project Background

1.1 Introduction

In years past the developments in information theory and communications engineering has led to significant advances in wireless technology, especially in the area of mobile broadband (MBB), in terms of higher throughput and capacity [1]. Today the majority of data is human-generated, as such it consists of high resolution images, audio and video [2]. However future networks (5G and beyond) must be capable of supporting not only human-generated traffic but also machine type communication (MTC) [3].

As a consequence of the shift to MTC the nature of the packets carrying this data is also changing, with MTC there is a greater emphasis on short packets [1], this type of communication may also have a strict latency requirement. In order to meet the demands posed by future network use cases, 3GPP have outlined three areas which define the 5G and beyond approach. The three areas are enhanced mobile broadband (EMBB), massive machine type communication (MMTC) and the principal focus of this work ultra-reliable low latency communication (URLLC).

URLLC will be the framework which critical infrastructure will be developed around, this is due to the high reliability 99.999% and low-latency (less than 1ms)[4]. Applications include traffic management, smart-cities, factories and hospitals equipped to perform remote surgeries. Given the mission-critical nature of the applications listed it is clear there is a need for the low-latency and high-reliability provided by URLLC.

Security is typically provided with traditional cryptographic methods, using key management. This will likely be problematic in the future as IoT devices will likely have limited computational power and memory, furthermore they are often constrained by power restrictions due to being battery powered [5]. As a result physical layer security (PLS) shows good promise as a solution for security in URLLC [6]. PLS leverages the inherent randomness of the wireless channel to secure the sensitive information.[7]

URLLC will enable the next generation of cities, healthcare and industry and for it to be implemented successfully research and investigation is required to ensure it meets the demands of its proposed applications. This serves as motivation for the following work.

1.2 Objective of Project

The primary objective for this work is to investigate resource allocation in URLLC systems. Following this the work should propose suitable methods to efficiently allocate resources, such as power and blocklength, to ensure the latency and reliability requirements of URLLC are met.

Simulations comparing the proposed resource allocation methods should show the performance and trade-offs of the methods. Following this some analysis and guidelines for practical URLLC systems should be discussed and presented.

1.3 Structure of Thesis

- Chapter 2 explores existing literature and their relevance to this thesis. From the literature review expressions are taken which will serve as the starting point for simulations and solutions presented later. The literature review also provides insights as to the different optimization methods used by different works.
- Chapter 3 shows preliminary work, where the expressions from the literature review were investigated and different parameters varied. This provides insight into the relationship between the expressions that govern URLLC and real word performance characteristics.
- Chapter 4 details the solutions developed by the author to address the resource allocation problem. Different solutions are proposed and depending on the specific system requirements one may be more suitable than the other. NOMA and OMA are compared and NOMA shows itself to outperform OMA in the analysis presented.
- Chapter 5 provides some concluding remarks on the solutions presented but also on the project as a whole.
- Chapter 6 discusses the impact of the research and explains how the proposed solutions may be leveraged to provide enchanced communications enabling future technological developments.
- Chapter 7 uses the experiences of the author gained by this project to highlight some potentially worthwhile areas for future work and research.
- Acknowledgements, bibliography and appendices conclude this thesis.

Chapter 2: **Literature Review**

2.1 Finite Blocklength

URLLC will feature short packets, in the region of 1000 bits or less [5], this poses a fundamental challenge as wireless networks typically operate with long packets (greater than 1000 bytes) this is due to the inherent nature of the types of communication the networks carry, such as video and voice. These packets are large enough to approach the Shannon secrecy capacity [1] and the law of large numbers can be leveraged. As outlined in [8] the effect of these short packets means the infinite blocklength assumption no longer holds for practical URLLC IoT applications.

The achievable secrecy rate under the finite blocklength regime is analysed in [1, 9–12]. Secrecy rate is the maximum rate at which information can be reliably and securely sent along a given channel for the finite blocklength case. [1, 9–12] define the secrecy rate (R_s) by the following approximation (2.1).

$$R_s = C_s - \sqrt{\frac{V_A}{N}}\frac{Q^{-1}(\epsilon)}{\ln(2)} - \sqrt{\frac{V_E}{N}}\frac{Q^{-1}(\delta)}{\ln(2)} \tag{2.1}$$

N denotes the blocklength, C_s is the secrecy capacity, $C_s = \log_2(1+\gamma_A) - \log_2(1+\gamma_E)$, where γ_A denotes the signal to noise ratio (SNR) at the intended user and γ_E denotes the SNR at the eavesdropper. $V_A = 1-(1+\gamma_A)^{-2}$ is the channel dispersion for the intended user, $V_E = 1-(1+\gamma_E)^{-2}$ is the channel dispersion for the eavesdropper. ϵ is the error rate. δ is the information leakage rate and Q^{-1} is the inverse of the Q-function $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$.

An important point regarding (2.1) is highlighted in the papers [1, 9, 10], is that the SNR at the intended user must be greater than that at the eavesdropper, this condition expressed mathematically is $\gamma_A > \gamma_E$ otherwise the secrecy rate is zero. As shown in [13], as blocklength increases we approach the Shannon secrecy capacity.

This forms the starting point for the analysis presented in related works, and from this expression metrics are derived to assess the performance for various blocklengths and other parameters. Simplifications of (2.1) exist, such as those which do not include a secrecy penalty [14, 15]. The secrecy penalty is in place when there is a secrecy requirement in the system. In later sections we will present simulations both with and without the secrecy penalty.

Further exploration of (2.1) is undertaken in [13, 16, 17] where different parameter values are investigated. Quality of experience (QoE) and its relation to expression (2.1) are considered in [18].

2.2 Secrecy Throughput

The next step in the related works considered it to isolate a performance metric which will subsequently be used for optimization and then for evaluation of system performance. In [1] (2.1) is re-arranged to obtain the error rate in terms of blocklength, information leakage and number of information bits (B). This gives the error rate expression (2.2), as shown below.

$$\epsilon = Q\Bigg(\sqrt{\frac{N}{V_A}}\bigg(\ln\frac{1+\gamma_A}{1+\gamma_E} - \sqrt{\frac{V_E}{N}}Q^{-1}(\delta) - \frac{B}{N}\ln 2\bigg)\Bigg) \tag{2.2}$$

Expression (2.2) is used to calculate the error rate for each sample of a channel realization, this is then averaged to get the average error rate denoted by $\bar{\epsilon}$. From this it is clear that $1-\bar{\epsilon}$ is the remaining portion without errors, following this the secrecy throughput, denoted T_s, can be easily expressed by expression (2.3) below, where B is the number of bits, N is the blocklength and $\bar{\epsilon}$ is the average error rate.

$$T_s = \frac{B}{N}(1-\bar{\epsilon}) \tag{2.3}$$

A similar process is performed in [7, 10], these works derive a secrecy throughput expression from the secrecy rate expression they define previous. In [9] a slightly different method is used to find the secrecy throughput given that in the system model it has multiple legitimate users. Serving multiple users will be explored later in this thesis as it is more realistic than serving one user and has better applicability to industry.

Secrecy throughput as defined by expression (2.3) is the principal metric used in the evaluation of numerical results in related works.

2.3 Optimization

The related works all seek to maximise either the secrecy rate or the secrecy throughput expression they derive respectively. However they differ greatly in their optimization methods and system models. Some works propose search based algorithms other works propose analytical expressions. Some system models consider multiple antennas and many users other consider 2 users with single antennas.

2.3.1 Orthogonal Multiple Access - OMA

Joint optimization of power and blocklength for OMA transmission is explored extensively in [8]. The system model presented does not have an eavesdropper so as a result the rate expression in this paper is simplified by dropping the secrecy penalty. The optimization problem presented uses the energy constraint and channel state information (CSI) to find the search bounds of blocklength. This neatly reduces the optimization complexity. An iterative algorithm is then presented to allocate power and blocklength. The introduction of the search bounds for blocklength allocation ensures that the algorithm converges rapidly. Using CSI to reduce the optimization complexity shows good promise and following this is explored in solutions proposed in chapter 4. Further simplification of the algorithm and solution is done by leveraging the high SNR regime.

In [1] a framework for optimizing secrecy throughput for the single and multi antenna case by varying blocklength is presented. For the multi-antenna case the effects of artificial noise (AN) and are examined. AN is used to degrade the eavesdroppers channel. The most significant results and conclusions of this paper assert that adding more antennas to the base station (BS) can reduce the performance drop due to short packet communications. The numerical results show that the an improved secrecy throughput can be achieved with reduced blocklength when multiple antennas are used. Most importantly in relation to this work is that for the single antenna case an analytical expression for the optimal blocklength is obtained. This differs greatly from many of the other works in this area where typically an algorithm is used to compute optimal blocklength. Numerical results show good performance of the proposed solutions and most significantly the analytical expression allows efficient computation of the optimal blocklength.

The first optimization problem considered in [9], is the weighted sum throughput, this is because in this system model there are multiple users. Weighted sum throughput is used to ensure fairness among users. The optimization performed in this work is done through jointly optimizing bandwidth and power allocation this is done whilst ensuring that the minimum security capacity of each device is maintained, from this the minimum blocklength can be found. This paper goes into detail about the specific algorithms used to solve the non-convex optimization problem, namely block coordinate descent and successive convex approximation. Block coordinate descent is also used for effective throughput maximization and transmission time minimization in [19]. The second optimization problem considered by [9] is the minimization of the total transmit power , whilst ensuring minimum security capacity to each user. Total transmit power is minimized to ensure energy efficiency. Energy efficiency is one of the key areas of focus by 3GPP and ensuring efficient use of energy is an area of focus of 5G and beyond. This paper [9] is novel in its contribution in the sense that it provides analysis and an algorithm based solution for the minimization of total transmit power. The performance metrics used to show algorithm operation focus on power savings, as a result this makes the contributions of this paper hard to balance against other papers where more conventional metrics are presented.

A general method for finding optimum blocklength to maximise throughput is proposed in [10], furthermore the high-SNR regime is analysed in detail and analytical results are presented. A significant contribution of this paper is the outage-based metric which considers reliability and security. In this regard the paper is novel and provides a detailed analysis into outage probability and throughput. This paper also proposes a numerical technique for finding the optimal blocklength. Numerical results are also presented to support the methods proposed throughout the paper. Of key importance in this paper is the results and justification for the high-SNR approximation, the paper asserts that the high-SNR approximation is valid as the IoT nodes will have little signal processing capabilities and as such the SNR must be maintained high such that successful decoding can occur. This high-SNR approximation is also done in [20] and [21] for the same reason, this also makes the optimization problem less intensive to solve.

In [7] secrecy throughput is optimised by blocklength variation for the single antenna case. In the multi antenna case this work investigates AN-aided transmission and power allocation. AN is used to degrade the eavesdroppers channel and thus increase secrecy capacity, this is also done in [1]. An algorithm is presented to solve the throughput maximization. Results show that increasing the number of BS antennas allows for higher secrecy throughput as AN can be increased accordingly at the BS. A noteworthy result is that this paper asserts that in the on-off policy using the the maximum blocklength maximises secrecy throughput, this differs with observations in [22, 23].

[21] looks at optimization by building upon the secrecy rate obtained in [12], this paper presents methods to optimize resource allocation problems associated with URLLC in the high SNR regime. Furthermore this paper takes a look at energy-efficient resource allocation for URLLC. This paper presents numerical results suggesting that circuit power is more significant compared with transmit power in the case of the average time between packets is greater than the delay requirement. In this case, especially with the small packets in URLLC high transmit power may be used to increase secrecy rate. This contribution is of interest because few works in this area analyse the circuit power and transmit power comparison. The bandwidth allocation in particular is analysed in this work, differentiating it from others where blocklength was investigated and optimized.

Energy efficiency in considered in [24, 25]. In [25] a joint allocation algorithm for power and blocklength, with the overall objective of minimizing energy consumption. This minimization is done whilst ensuring that each user's error rate is above the system requirements, thus ensuring the reliability constraint, and that the total blocklength is below a certain threshold, this ensures the latency requirement. This paper provides detailed numerical results showing the performance of the algorithm with regards to energy consumption. [24] examines two sub-optimal but efficient algorithms, with energy efficiency being maximised.

2.3.2 Non-Orthogonal Multiple Access - NOMA

NOMA has been suggested as a potential solution for URLLC by numerous papers [26–33]. This is because of the way that it uses superposition coding (SC) to communicate with multiple receivers simultaneously. The key reason for its performance advantage is that successive interference cancellation (SIC), is performed at the receiver. This is where the signals are decoded in order of the highest received signal strength. As a result, the BS transmits to multiple users simultaneously, this greatly reduces latency and as a result NOMA is a solution gaining significant interest for URLLC applications. NOMA and SIC are explained in further detail and simulations are provided in section 3.4.3.

A basis for NOMA and SIC is presented in [26], the basic operation of NOMA is explicitly defined for the infinite blocklength case. This provides a starting point for our analysis as the infinite blocklength case is essentially the Shannon channel capacity. From the Shannon capacity we take away the reliability penalty and the secrecy penalty to arrive at the NOMA finite blocklength expression.

As outlined in [8, 27] NOMA in collaboration with successive interference cancellation (SIC) the throughput to users can often exceed that of an OMA system, further evidence of superior NOMA performance under different analysis and performance metrics is demonstrated in [28–30].

Detailed optimization of finite blocklength NOMA is done in [8], interestingly the secrecy constraint is dropped such that the only penalty is the reliability penalty. This is common place in works relating to URLLC NOMA, secrecy is often dropped as a requirement. As a result secrecy analysis of NOMA is an open opportunity.

The NOMA optimization problem presented in [8] is minimising an error rate of one user whilst keeping the other users error rate below a certain threshold. This optimization is done by varying the blocklength and power, as a result a joint optimization algorithm is proposed for NOMA resource allocation. Simulation results show the performance of the joint optimization algorithm. Another important contribution of this paper is the fact that analysis is presented for OMA optimization and NOMA optimization on a comparable system model. The result is that this paper provides insight for a detailled OMA and NOMA comparison. The general trend of the results is that for URLLC applications NOMA provides superior performance characteristics.

In [34] finite blocklength downlink NOMA transmissions are optimised. This is done by maximising throughput at the user with joint decoding. The error constraints are used to simplify the search region and a simple bisection search algorithm is proposed to solve this resource allocation problem. Complexity analysis is presented showing small number of iterations required for the algorithm to converge. Numerical results presented show that in general NOMA outperforms OMA in terms of throughput, but also that joint decoding NOMA outperforms NOMA with SIC which was analysed in [8, 27].

Joint power and blocklength allocation is investigated in [25], specifically in regards to energy-efficient URLLC. The algorithm jointly allocates power and blocklength such that total power is minimised, whilst ensuring that error rates are above a minimum threshold and one user is allocated more power than the other, a fundamental requirement for power domain NOMA. Again in this NOMA analysis secrecy rate is neglected instead considering only the reliability constraint. Numerical results show the performance of the NOMA resource allocation algorithm. In this paper OMA is also considered and a detailed comparison of OMA and NOMA shows that NOMA outperforms OMA in URLLC applications. However a key insight of this paper is that it shows that NOMA consumes more energy than the OMA equivalent and as a result is less energy efficient. This is an important finding for practical systems as 3GPP have stated the importance of energy efficiency in future networks.

Power allocation finite blocklength NOMA is investigated in [35]. Energy efficiency is maximised by the algorithm proposed and simulations show that NOMA is better than OMA in terms of energy efficiency. This differs to energy efficiency analysis in [25].

2.4 Literature Review Summary

- Literature combining NOMA, URLLC and security constraints are sparse and in many ways this represents an open opportunity. It appears that when analysing NOMA works often drop the secrecy constraint possibly in order to simplify analysis. This will be highlighted as an area for future work in chapter 7 of this thesis.
- It is presented that NOMA has good performance when the delay requirement is strict, this is the case in URLLC, this suggests NOMA may be a promising candidate for URLLC applications.
- Some works manage to derive analytical expressions to allocate resources. However the vast majority of works solve the resource allocation problem by an algorithm based approach.
- Using CSI to simplify the search region of the optimization reduces complexity and ensures a rapid convergence of algorithms.

With the insights gathered from the relevant related works we can investigate the effects of different system parameters with the fundamental equations defining URLLC. From these investigations we will later develop solutions for URLLC resource allocation.

Chapter 3: Preliminary Work

In order to provide a solid basis for building upon fundamentals the preliminary works involved simulations where the effect of parameters such as blocklength, error rate and information leakage rate are investigated for a single user. Following this multiple access techniques such as OMA and NOMA have been investigated and compared for the system model proposed.

3.1 Channel Models

In order to simulate channels effectively the simulations in this thesis used Rayleigh fading. Rayleigh fading models a channel where there is no line-of-sight (LOS) component. As a result the received signal arrives at the receiver by randomly reflections off objects in the environment. This model was chosen as it effectively replicates a realistic scenario where URLLC IoT might be deployed. Such as a factory with many obstacles between the base station and the receiver. Furthermore the analysis presented in this thesis looks at downlink only. That is when the base station must transmit to one or multiple distributed users. This again was chosen due to it similarity to real world industrial IoT applications. A simple illustration of Rayleigh channel is provided in figure 3.1.

3.2 Simulation Methods

In the preceding section it was outlined that Rayleigh fading is the channel model. Rayleigh fading is a random process, in order to achieve repeatable simulations it was necessary to repeat the simulations many times and average the results to account for the random nature of the channel model. This process of multiple simulations and averaging is called Monte Carlo simulation and is widely used in communication and information theory to produce accurate results. MATLAB was the specific tool used for the simulations, in order to maximise learning opportunity for the author, few inbuilt functions were used, simulations were done from first principles.

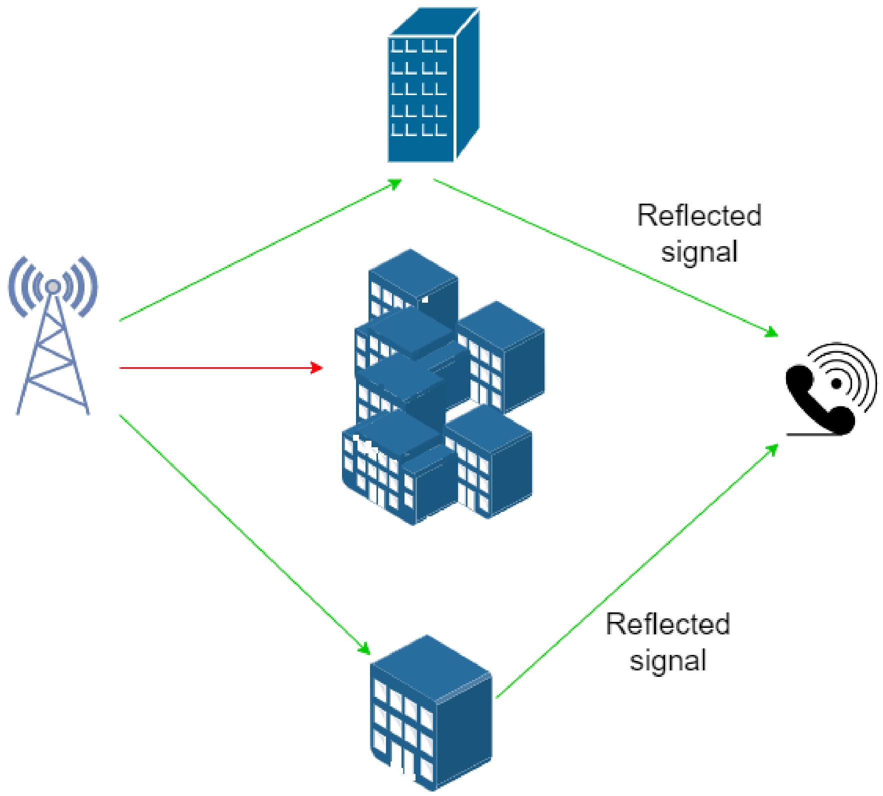

Figure 3.1: Rayleigh Channel

3.3 Single User Analysis

In order to gain insight as to the effect of varying different parameters, first we consider a one user downlink scenario, were the BS has to send some data to a single actuator A. The packet structure contains the data bits and also some redundancy bits.

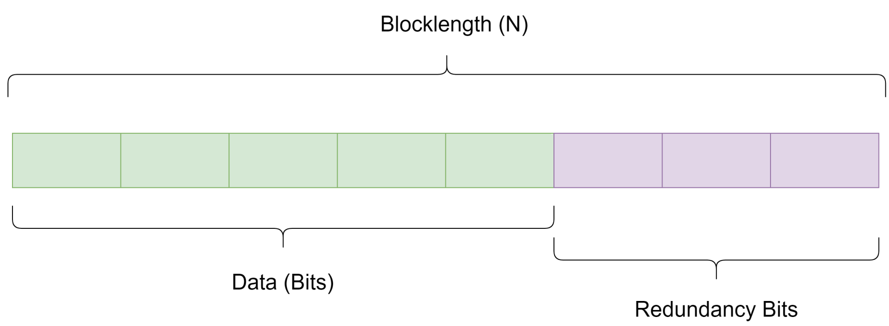

Figure 3.2: One user packet structure

3.3.1 Effect of Blocklength on Secrecy Rate

In the infinite blocklength regime, we know that the secrecy capacity, the maximal secrecy rate at which information can be reliably and securely received can be denoted as follows.

$$C_s = \log_2(1+\gamma_A) - \log_2(1+\gamma_E) \tag{3.1}$$

In equation (3.1), γ_A denotes the signal to noise ratio (SNR) at the actuator, γ_E denotes the SNR at the eavesdropper. Using the secrecy capacity expression we can express the finite blocklength secrecy rate as done in [1].

$$R_s = C_s - \sqrt{\frac{V_A}{N}}\frac{Q^{-1}(\epsilon)}{\ln(2)} - \sqrt{\frac{V_E}{N}}\frac{Q^{-1}(\delta)}{\ln(2)} \tag{3.2}$$

Expression (3.2) is the same as expression (2.1), for clarity it has been repeated here. The meaning of the variables has already been defined in section 2.1 above.

The reliability and secrecy requirements imposed serve to decrease the secrecy rate, we can think of the sub-expression $\sqrt{\frac{V_A}{N}}\frac{Q^{-1}(\epsilon)}{\ln(2)}$ as the penalty imposed by reliability requirement. Similarly we can think of $\sqrt{\frac{V_E}{N}}\frac{Q^{-1}(\delta)}{\ln(2)}$ as the penalty imposed by secrecy requirement.

We can now simulate the effect of blocklength on the secrecy rate, denoted as R_s. Using the Monte Carlo method we simulate 10^5 channel realisations. We choose ϵ and δ as 10^{-2}. These selected values for error rate and information leakage rate are arbitrary and the particular application will dictate their value, their effects will be evaluated in upcoming simulations. We consider blocklengths less than 1000 as this is the short-blocklength regime.

From figure 3.3 it is clear that as N increases the secrecy rate approaches the secrecy capacity of the channel. As N approaches infinity the secrecy rate becomes asymptotic to the secrecy capacity. This is clear from the graph but also the expression (3.2), as the denominator N tends to infinity the penalties associated with error and information leakage tend towards zero, and the secrecy rate approaches the secrecy capacity.

The insights from this simulation are as follows, for a given error rate and information leakage rate, we can choose the blocklength such that we achieve a desired secrecy rate. If a high secrecy rate is desired then a longer blocklength should be used as more redundancy is required to ensure the reliability and secrecy constraints. The converse is also true at a lower secrecy rate the blocklength can be shorter as we need less redundancy in the packet.

In essence this simulation captures the rate-reliability trade-off inherent to the finite blocklength regime. Upcoming simulations investigate the effects of parameters such as error rate and information leakage rate on the secrecy rate.

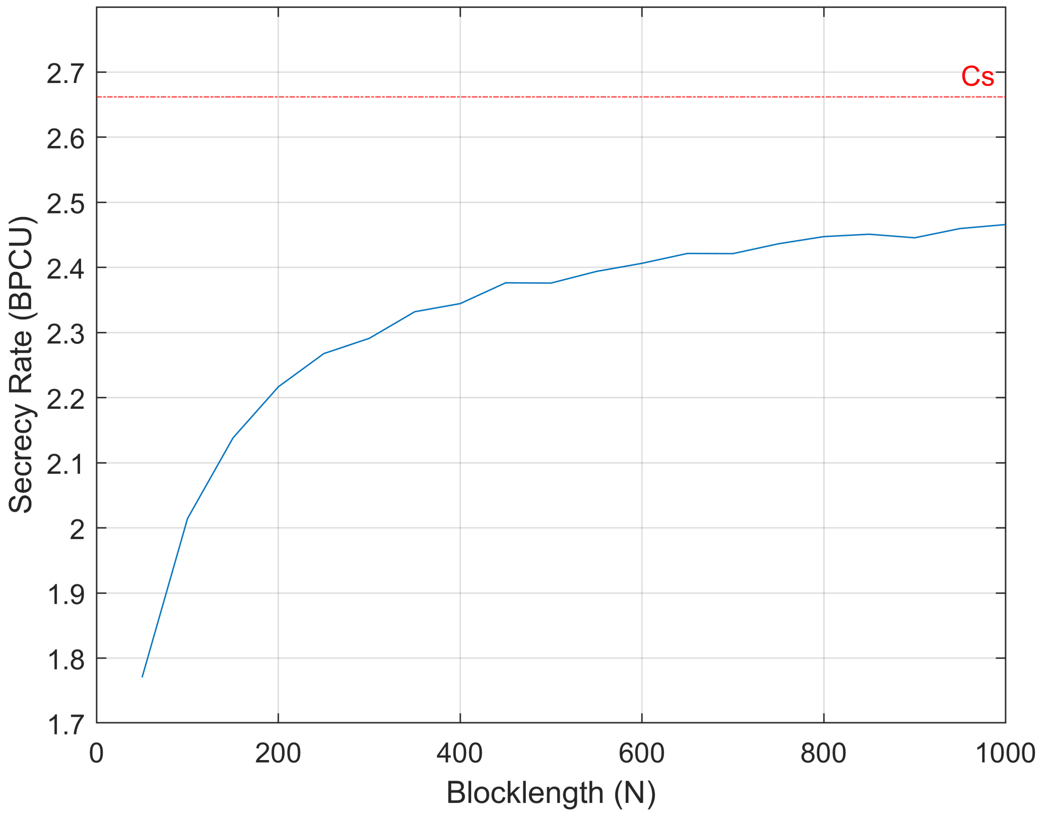

Figure 3.3: Effect of blocklength on secrecy rate

3.3.2 Effect of Error Rate on Secrecy Rate

The Monte Carlo method can be used to investigate the effect of error rate on the secrecy rate. In this case we fix the information leakage rate denoted by δ as 10^{-2} and repeat the simulation multiple times for different error rates.

From figure 3.4 we see that error rate has a significant effect on the secrecy rate, we note that if we are willing to tolerate more errors, i.e. a higher error rate, then we can achieve a higher secrecy rate. The converse is also true, if a low error rate is required a lower secrecy rate will be achieved.

Looking at the expression (3.2) this result may be counter intuitive, this is due to the fact that the inverse of the Q-function gets larger the smaller the argument passed into it is. So for a small error rate the result of the inverse of the Q-function is large and for a large error rate result of the inverse of the Q-function is small.

The insights from this simulation is that there is a penalty associated with reliability and more demanding reliability constraints reduce the secrecy rate. Less demanding reliability

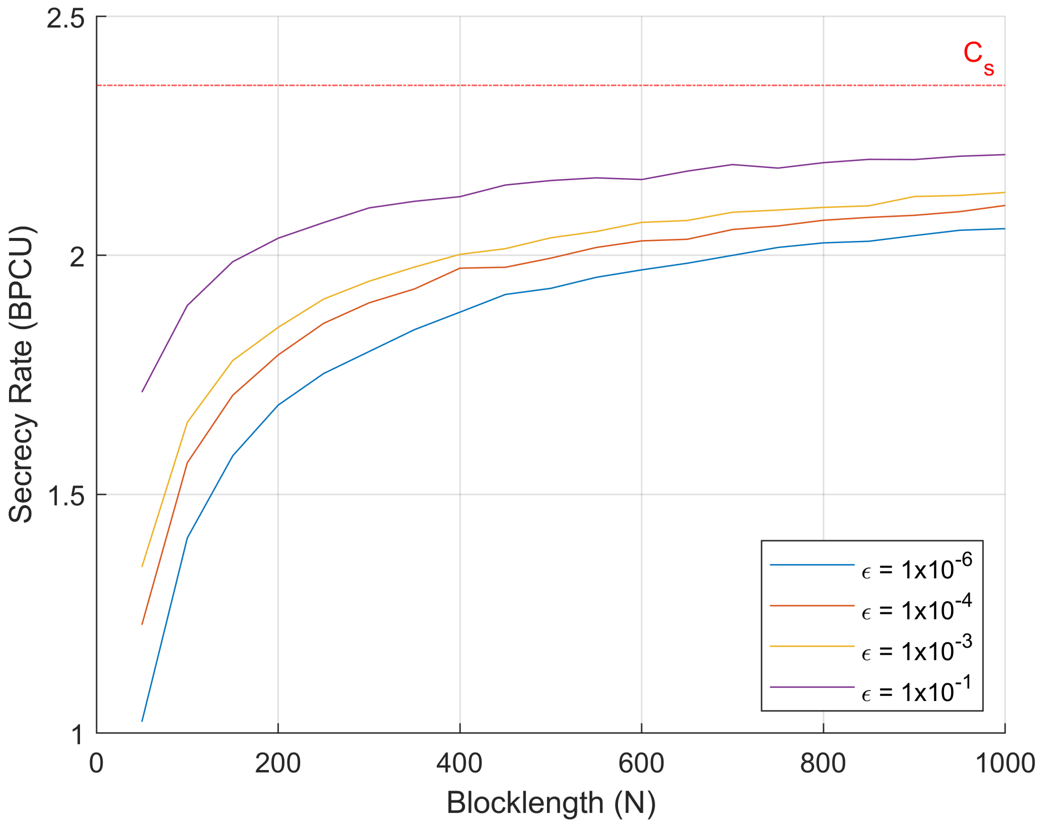

Figure 3.4: Effect of error rate on secrecy rate

constraints lead to a higher secrecy rate.

3.3.3 Effect of Information Leakage Rate on Secrecy Rate

Another simulation can be conducted, this time investigating the effect of information leakage rate on the secrecy rate. In this case we fix the error rate denoted by ϵ as 10^{-2} and repeat the simulation multiple times for different information leakage rates.

In figure 3.5 we see comparable results to that achieved in the previous section. A high tolerable information leakage rate results in a higher secrecy rate and a lower tolerable information leakage rate results in a lower secrecy rate.

This result is the case, again due to the inverse of the Q-function in the information leakage penalty expression.

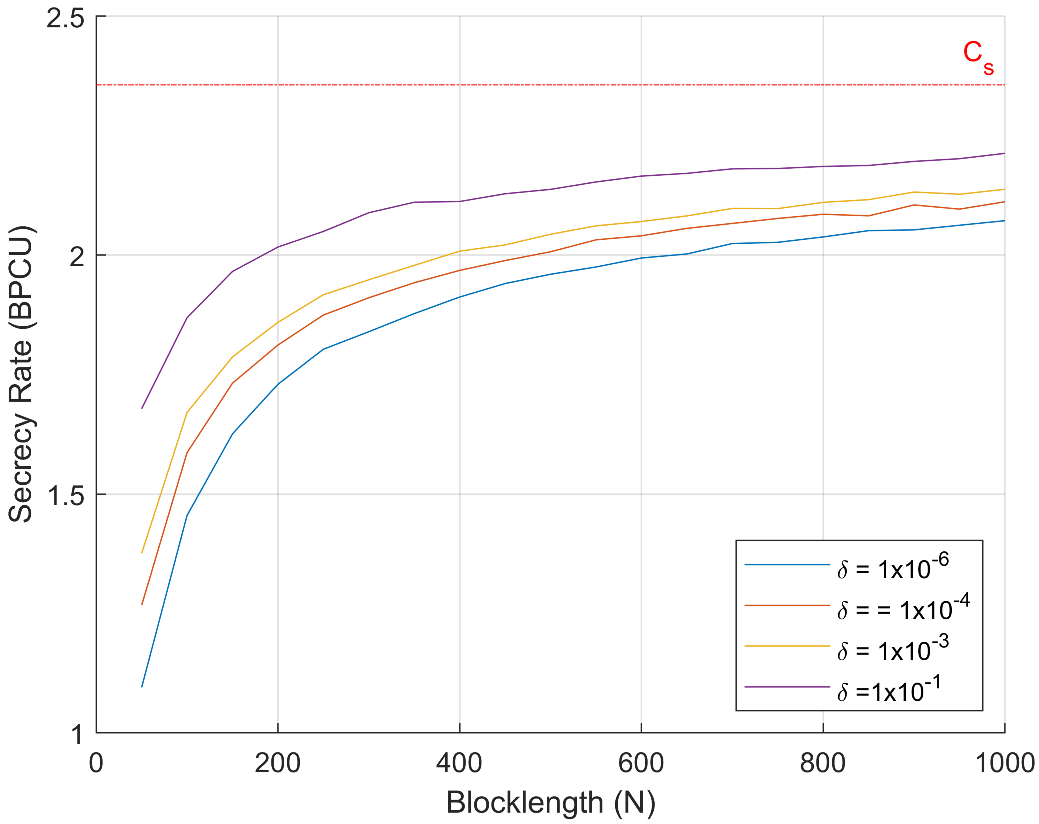

Figure 3.5: Effect of information leakage rate on secrecy rate

We see from the simulations presented in section 3.3.2 and 3.3.3 that there is a considerable penalty associated with the chosen error rate and information leakage rate, and that by carefully choosing information leakage rate and error rate we may affect secrecy rate for a given channel.

3.3.4 Secrecy Throughput

In many ways secrecy rate is not the complete picture, we must consider other metrics to evaluate performance. In section 3.3.1 we spoke about the rate-reliability trade-off. In order to achieve the higher secrecy rate, we must add redundancy to the packet, this is to say that we add extra bits to the data which serve as redundancy-bits and are not actually part of the information bits we wish to send. Assuming that there is some redundancy added means that the amount of useful data in the packet is less than the total bits in the packet, we call this rate of useful data throughput. As shown in [1], the secrecy throughput can be expressed as follows.

$$T_s = \frac{B}{N}(1 - \bar{\epsilon}) \tag{3.3}$$

In expression (3.3), B is the information bits in the packet in question, in URLLC this is expected to range from 50 to 1000 bits. N is the blocklength. $\bar{\epsilon}$ is the average error rate, this is computed by rearranging the rate expression in term of error, given by expression (2.2), and averaging over many channel realisations. We perform a Monte Carlo simulation using expression (3.3), varying blocklength and using the blocklength to compute error rate using expression (2.2).

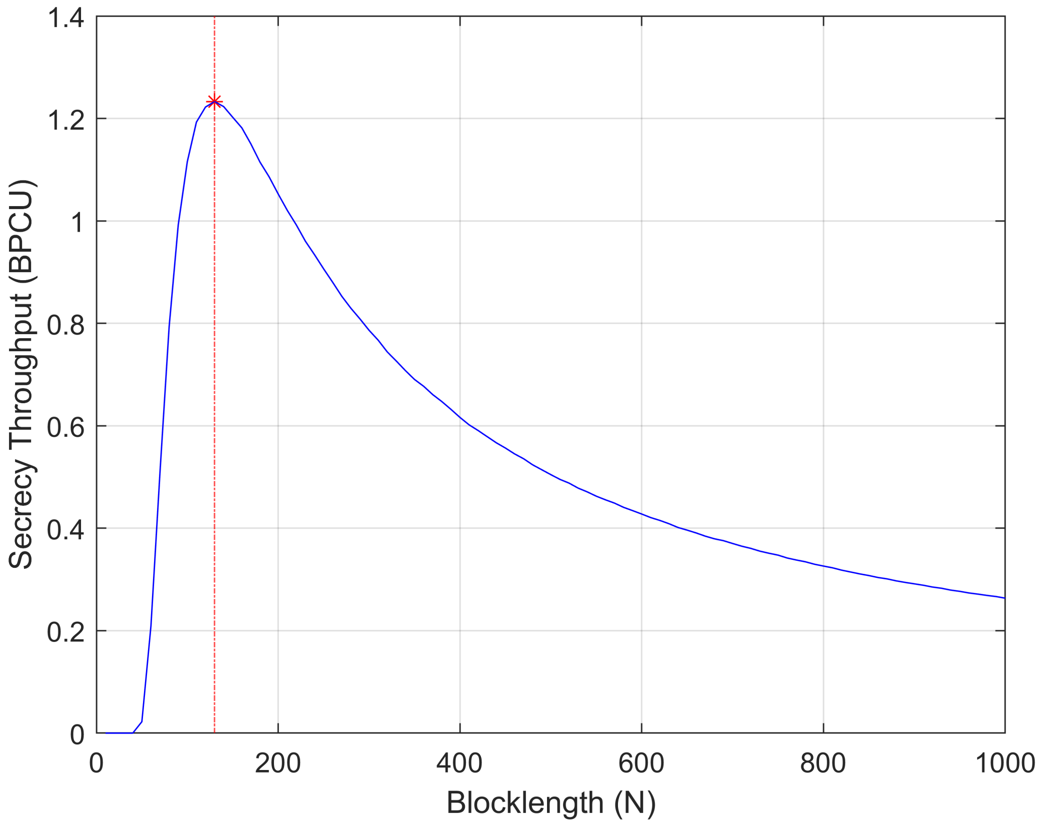

Figure 3.6: Secrecy throughput with $B = 300$

The key result from this simulation is that, for a given number of information bits, there exists an optimum blocklength which maximises throughput for a given channel. This is a significant result and following this we can conclude that in order to maximise secrecy throughput we must choose the optimum blocklength. The motivates the upcoming solutions in chapter 4, where efficient techniques for determining the optimum blocklength are developed.

3.4 Multiple Access Techniques

Thus far we have only examined the case of 1 actuator and 1 eavesdropper, the analysis of multiple users is a more real-world example as we expect in URLLC IoT there will be multiple devices. Consider an automated factory in which there is downlink communication from a base station (BS) to two different actuators, each having one antenna. These actuators (A_1 and A_2) have high reliability and low-latency requirements and as such fall under the URLLC requirements umbrella. The actuators each require a data packet, containing the same number of information bits, denoted as B. It is assumed that we know exactly the channel gain characteristics h_1 and h_2 of each of the paths respectively. The channel between BS and A_1 is h_1, and the between BS and A_2 is h_2. These channels are assumed to be independent and undergo Rayleigh fading and remains constant over the total blocklength N. This system model is summarised in figure 3.7. We also note that there is no eavesdropper in this system model, furthermore there is no secrecy requirement. The secrecy requirement is dropped at this point to simplify the expressions and as shown by the simulation presented in section 3.3.3 the secrecy requirement serves to reduce the overall rate. For the analysis and solutions beyond this point we can assume that if the secrecy requirement was reimposed it would simply reduce the overall rate or throughput.

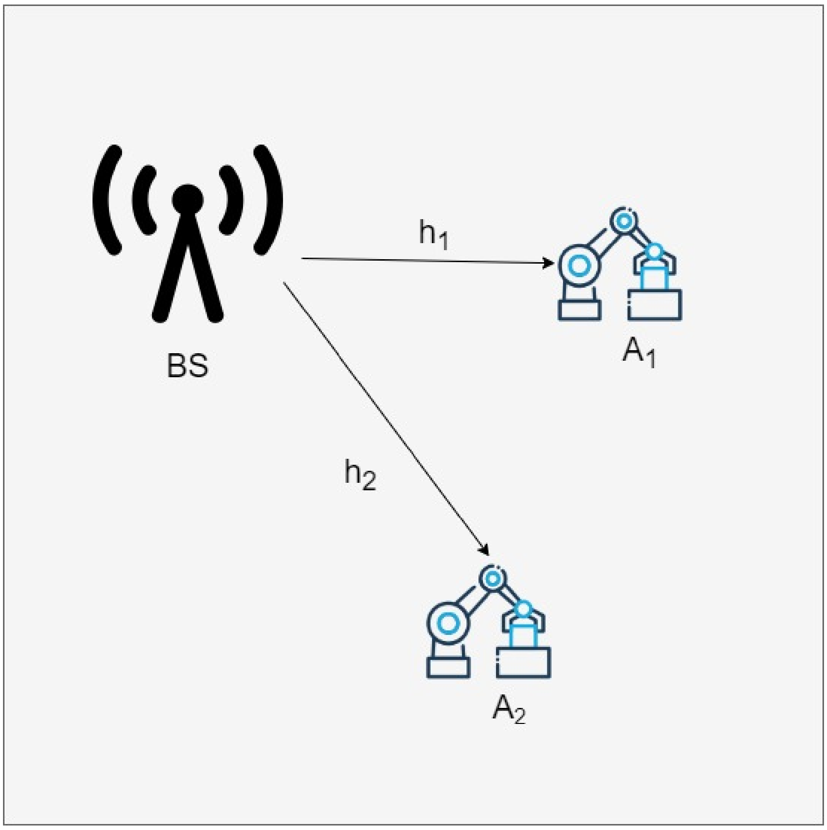

Figure 3.7: System Model

3.4.1 Orthogonal Multiple Access - OMA

3.4.2 Blocklength Allocation

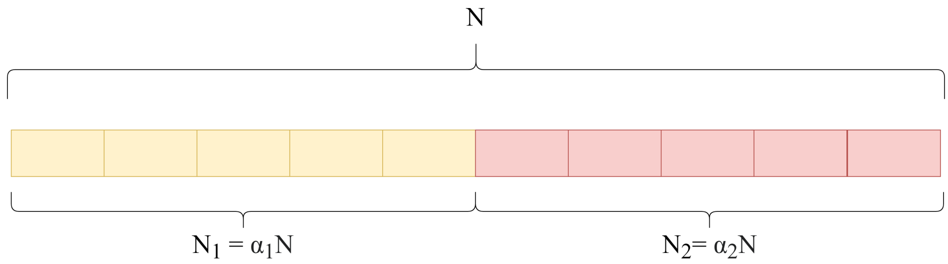

Figure 3.8: OMA two user packet structure

We will introduce a blocklength allocation factor α_i, representing the factor of blocklength allocated to the i-th actuator. This blocklength allocation factor is introduced in order to share the blocklength between multiple users.

By removing the secrecy requirement from equation (2.1) as shown in [15], the achievable channel coding rate R_i for the actuator A_i in the finite blocklength case is given by

$$R_i = C_i - \sqrt{\frac{V_i}{\alpha_i N}} \frac{Q^{-1}(\epsilon_i)}{\ln(2)} \tag{3.4}$$

N denotes the total blocklength, α_i is the portion of N allocated to the respective actuator. C_i is the channel Shannon capacity, or the infinite blocklength capacity, $C_i = \alpha_i \log_2(1+\gamma_i)$, where γ_i denotes the signal to noise ratio (SNR) at the $i-th$ actuator, $\gamma_i = \frac{P|h_i|^2}{\alpha_i N_0}$. N_0 is the noise PSD at the actuator. $V_i = 1-(1+\gamma_i)^{-2}$ is the channel dispersion for the actuator. ϵ_i is the error rate and Q^{-1} is the inverse of the Q-function $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$.

It is important to realise that for the two actuator case $i \in 1, 2$, following this it is clear that $\alpha_2 = 1 - \alpha_1$. To illustrate the blocklength allocation trade-offs we run simple simulations and plot the results as figures 3.9 and 3.10. As the blocklength allocation α_1 increases and the blocklength allocation α_2 decreases, as a result the rate of A_1 increases and the rate of A_2 decreases. It is clear that there in an inherent trade off in rate and blocklength allocation between the 2 users. We can also examine the throughput for the 2 users as shown in figure 3.11. From this we see that the there is a distinct optimum blocklength for both users. This further highlights the importance of allocating blocklength sensibly between multiple users such that throughput is maximised.

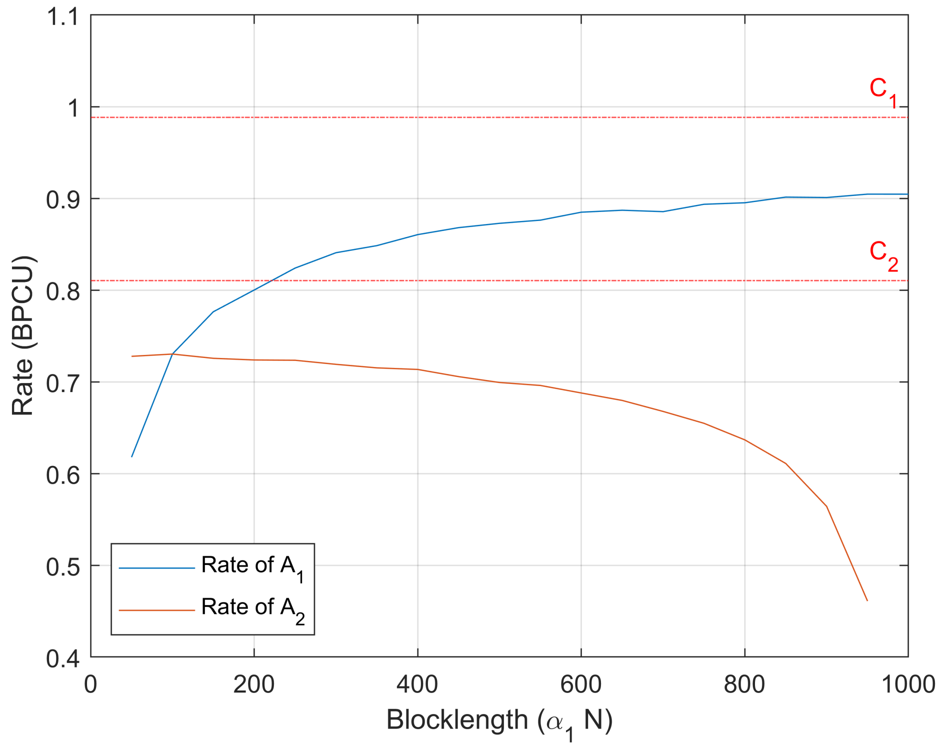

Figure 3.9: Effect of variation of blocklengths on rates of multiple users

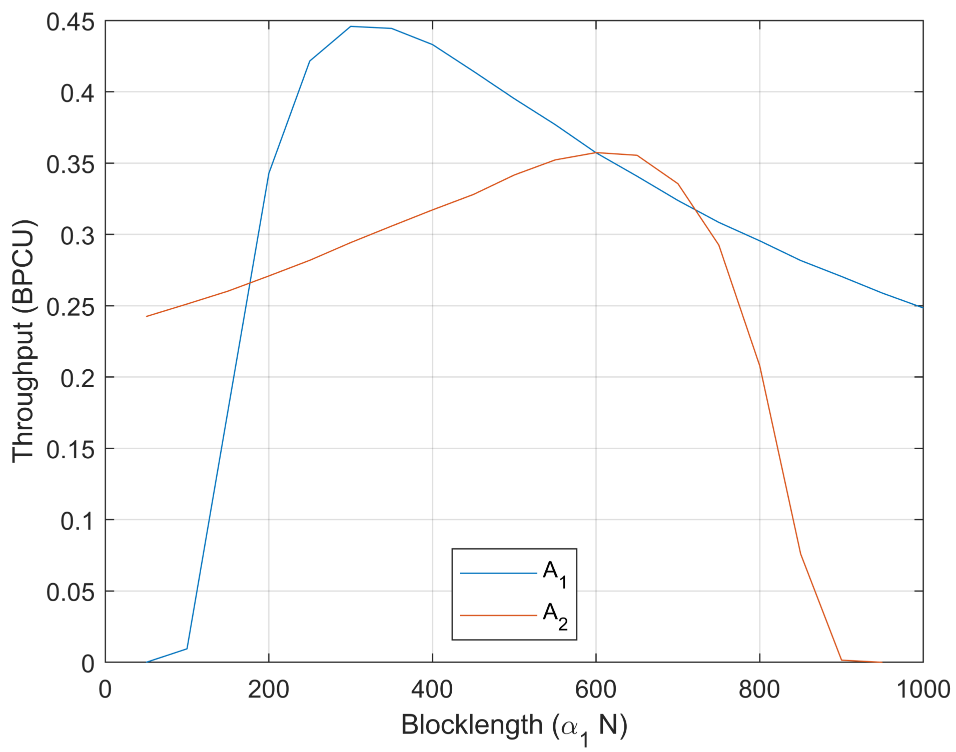

Figure 3.10: Effect of variation of blocklength on throughputs for multiple users

3.4.3 Non-Orthogonal Multiple Access - NOMA

NOMA has been shown in [26–33] to outperform OMA in certain applications. The basis for power domain NOMA is that multiple signals are superimposed on one transmitted signal. Successive interference cancellation (SIC) is then used to decode the signals in order of strongest received signal first. In this thesis, where NOMA is mentioned SIC is implied as the decoding method unless stated otherwise. There are other NOMA decoding methods however they remain an area for future work in this field.

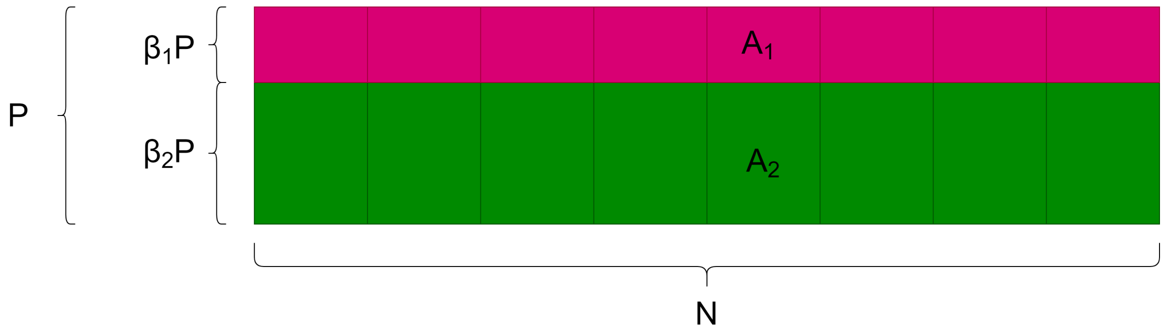

Figure 3.11: Power domain NOMA packet structure for 2 Users

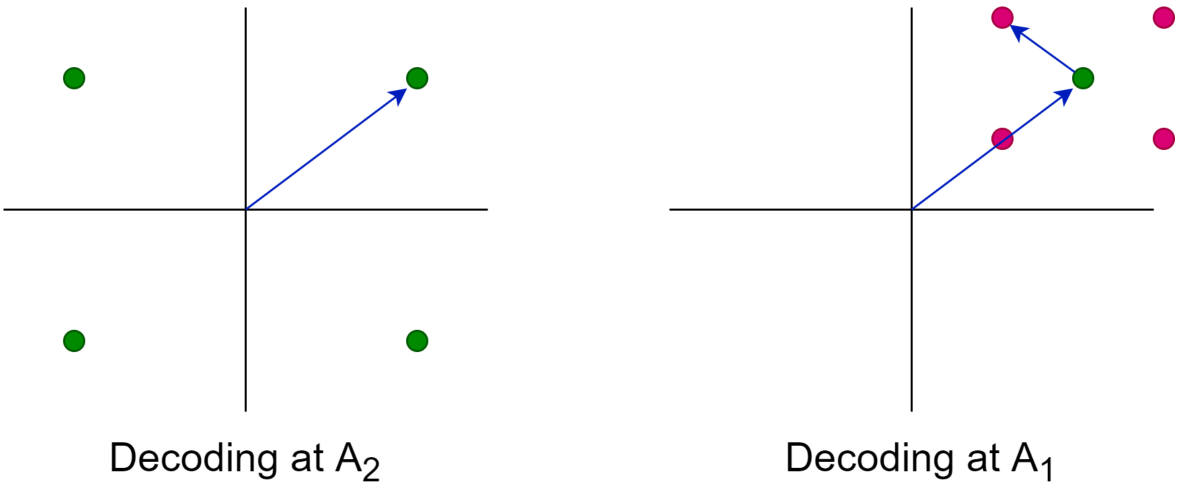

Figure 3.12: SIC for 2 Users

For the 2 user case, we assume A_1 has a better channel than A_2. The decoding process can then be analysed with the aid of figures 3.11 and 3.12. A_2 is allocated more power as it has a worse channel, this means that at A_2 the first signal in order of strongest signal is the signal intended for A_2. As a result A_2 decodes the signal. At A_1, who has a better channel, the signal inteded for A_2 is first in order of signal strength. As a result A_1 first decodes the signal for A_2, and removes it from the original received signal, A_1 can then decode the signal intended for itself.

With the basic concept outlined we can explore expressions for NOMA transmissions. In order to allocate power between the 2 users will will introduce a power allocation factor β_i. This represents the factor of total power P allocated to the i-th actuator.

The capacity of a channel depends on the SNR at the actuator, however for NOMA we don't just have noise degrading the channel, we also have interference due to the superposition of multiple signals. As a result we must redefine the capacity in the NOMA case to take into account the signal to interference and noise ratio (SINR). Leveraging equations derived in [26, 27] we can define SINR for NOMA with successive interference cancellation (SIC). Using these NOMA expressions from [26, 27] we will later compare OMA and NOMA performance using expression (3.4) and upcoming expressions.

SINR for NOMA can be defined as

$$\gamma_i = \frac{\beta_i P |h_i|^2}{P|h_i|^2 \sum_{k=i+1} \beta_k + N_0} \tag{3.5}$$

Expression (3.5) states that the SINR for NOMA is the signal strength for the i-th user divided by the sum of all of the signal strengths of users with better channels than the i-th user plus noise. In other words the SINR for a particular user is the signal strength divided by all the other users signal strengths it doesn't have to decode plus noise.

Assuming that user 1 has a stronger channel, i.e $\frac{|h_1|^2}{N_0} > \frac{|h_2|^2}{N_0}$, then we can explicitly write the our 2 user NOMA SINR expressions as

$$\gamma_1 = \frac{\beta_1 P |h_1|^2}{N_0} \tag{3.6}$$

A_1 does not suffer interference from the A_2 signal as first the signal for A_2 is decoded as it is a stronger received signal, this is because more power is allocated to A_2. The decoded signal from A_2 can then be subtracted from the received signal, thus eliminating interference.

$$\gamma_2 = \frac{\beta_2 P |h_2|^2}{\beta_1 P |h_2|^2 + N_0} \tag{3.7}$$

A_2 does suffer interference from A_1 as it comes first in the decoding order and can not subtract the signal for A_1 from the received signal.

The finite blocklength NOMA rate expressions are obtained by applying the reliability penalty to the NOMA capacity expressions outlined previous.

$$R_i = C_i - \sqrt{\frac{V_i}{N}} \frac{Q^{-1}(\epsilon_i)}{\ln(2)} \tag{3.8}$$

Expression (3.8) differs from expression (3.4) as it does not have a blocklength allocation term, as in NOMA the whole blocklength is used by all the users. This time C_i is the channel Shannon capacity, or the infinite blocklength capacity $C_i = \log_2(1 + \gamma_i)$, note the lack of a blocklength allocation factor. Where γ_i denotes SINR at the i-th actuator, as outlined above. The other parameters are the same as previous expressions.

Using the expression (3.4) for the OMA rate and the expression (3.8) for the NOMA rate, we can visualise the rate region for two users.

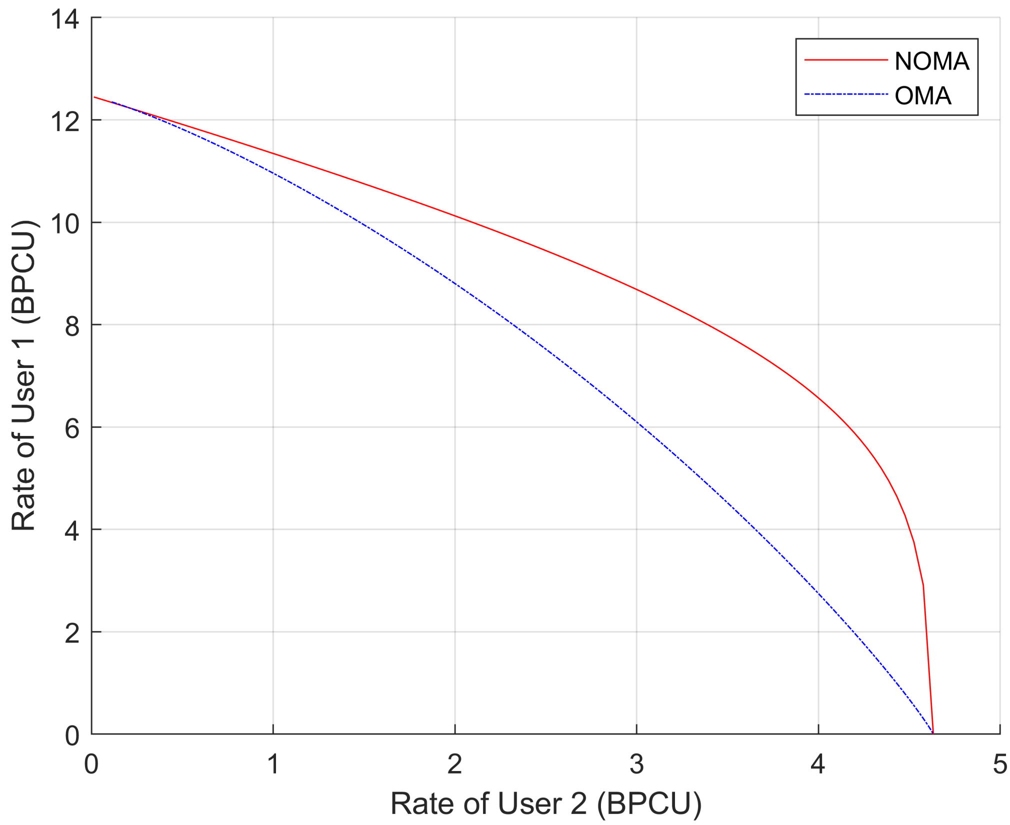

Figure 3.13: NOMA with SIC vs. OMA

From figure 3.13 it is clear that the rate region for NOMA with SIC dominates that for OMA. This means that under the same conditions NOMA can provide a greater rate for the actuators.

The results from figure 3.13 in combination with results in [26–30] presents a great opportunity for the application of NOMA to URLLC IoT applications. This is because in NOMA multiple users are served over one blocklength, thus reducing latency.

3.5 Problem Statement

The problem we will address in the upcoming proposed solutions is for a 2 user URLLC scenario, how should we efficiently allocate power and blocklength such that sum throughput is maximised.

In order to solve this problem we will develop algorithms which allocate resources efficiently.

Furthermore we will investigate the 2 different transmission models OMA and NOMA. For OMA we will investigate blocklegnth and power allocation, whereas for NOMA we will investigate only power allocation. Algorithms for OMA and NOMA will be developed to achieve sum throughput maximisation.

Chapter 4: Proposed Solutions

In order to solve the problem outlined in section 3.5, we will leverage the simulations done in the preliminary work sections and the knowledge gained from existing relevant works.

In section 4.1 we consider sum throughput maximization. This is is the maximization of the sum throughput only for the time that the BS is serving that actuator. Sum throughput is maximised in 4.1 under OMA transmission, where blocklength and power allocation are optimised.

The solutions presented in 4.1 have some drawbacks, for example as a consequence of only optimizing for the time that the BS is serving that actuator, the overall or average sum throughput is not considered, and may be quite poor.

This motivates section 4.2 where we optimise the average sum throughput, this is the sum throughput over the whole blocklength. Average sum throughput is a more significant metric and as a result, solutions in section 4.2 are the principal contribution of this thesis.

4.1 Sum Throughput Maximization

In order to maximise sum throughput, we first consider the throughput only for the time that the BS is serving that actuator. In 4.1.1 sum throughput is maximised by blocklength allocation. In 4.1.3 sum throughput is maximised by power allocation.

4.1.1 Blocklength Allocation

For A_i using (3.4) we can express the OMA case of finite blocklength rate with the blocklength allocation factor α_i.

For A_i we define rate as

$$\frac{B}{\alpha_i N} = C_i - \sqrt{\frac{V_i}{\alpha_i N}} \frac{Q^{-1}(\epsilon_i)}{\ln(2)} \tag{4.1}$$

Where $i \in 1, 2$, B is the number of information bits. The other parameters have already been identified in section 3.4.2.

It is important to recall that that in the 2 actuator case $\alpha_2 = 1 - \alpha_1$.

Rearranging (4.1) in terms of error rate yields

$$\epsilon_i = Q\left(\ln 2 \sqrt{\frac{\alpha_i N}{V_i}} \left(C_i - \frac{B}{\alpha_i N} \right) \right) \tag{4.2}$$

Following (4.2) instantaneous throughput for the i-th user is given by

$$T_i = \frac{B}{\alpha_i N}(1 - \epsilon_i) \tag{4.3}$$

From the relationship (4.3), and substituting in the error rate expression (4.2) we obtain the instantaneous throughput equation for the i-th user.

$$T_i = \frac{B}{\alpha_i N} \left[1 - Q\left(\ln 2 \sqrt{\frac{\alpha_i N}{V_i}} \left(C_i - \frac{B}{\alpha_i N} \right) \right) \right] \tag{4.4}$$

We can thus phrase our optimization problem as follows.

$$\begin{aligned} \max_{\alpha_i} \quad & T_{sum} \\ \text{s.t.} \quad & \alpha_1 + \alpha_2 = 1 \end{aligned} \tag{4.5}$$

Where $T_{sum} = T_1 + T_2$. In order to successfully solve this optimization problem, we employ a simple variation on the popular gradient ascent (GA) algorithm. GA and other similar algorithms rely on the derivatives. Examining (4.4), this expression is already unwieldy and differentiation makes it intractable. Therefore we make use of the high-SNR assumption, where $V_1 \approx 1$ and $V_2 \approx 1$, this greatly simplifies differentiation. The high-SNR approximation holds true for an SNR greater than around 3dB.

We can rewrite our instantaneous throughput expression after the high-SNR approximation.

$$T_i \approx \frac{B}{\alpha_i N} \left[1 - Q\left(\ln 2 \sqrt{\alpha_i N} \left(C_i - \frac{B}{\alpha_i N} \right) \right) \right] \tag{4.6}$$

From here we differentiate (4.6). Despite the high-SNR approximation the differentiated expression is unwieldy and long and as a result has been included in the appendices to maintain a neat format of this thesis. For simplicity the derivatives will be denoted $dT_1/d\alpha_1$ and $dT_2/d\alpha_1$.

An novel approach to finding the optimum blocklength allocation is presented in Algorithm 1. L is the step size taken by the proposed algorithm.

Algorithm 1 - Blocklength Allocation Algorithm

```
  α1 ← 0.5                                        ▷ initial estimate
  L ← 0.001                                       ▷ set step size
  repeat
      if |h1|^2 ≥ |h2|^2 then
          α1 ← α1 + L × dT1/dα1
      else
          α1 ← α1 + L × dT2/dα1
      end if
  until convergence
Output: α1*
```

Convergence of this algorithm can be determined by the value of the derivative, and using a tolerance value. For example when the value of the derivative is $0\pm$ the chosen tolerance value, then this is the optimum blocklength $\alpha_1^\star$.

The efficiency of this algorithm relies on the fact that instead of searching a function for the optimum we are able to simplify our optimization by the observation that depending on the channel gain characteristic shape of the sum throughput as shown in figures 4.1 and 4.2 changes. As shown in figure 4.1 a multi-modal sum throughput is observed, other channel gain coefficients such as in figure 4.2 result in uni-modal sum throughput plots.

Crucially however, and this is the key simplification of the algorithm, we note that the peak of the sum throughput always lies within the region above the peak of the throughput of the user with the greatest channel gain. This is highlighted by the vertical lines in figures 4.1 and 4.2. We see that the allocated blocklength may not result in the global optimum however the efficient algorithm manages to find a near-optimal sum throughput value by exploiting CSI.

The result of this is that, assuming we know the channel gains, we can use the derivative of the single user throughput to find the optimal blocklength allocation factor. This simplifies the problem reduces optimization complexity and ensures GA converges to the an optimum.

Results showing performance of algorithm 1 are presented in section 4.1.2.

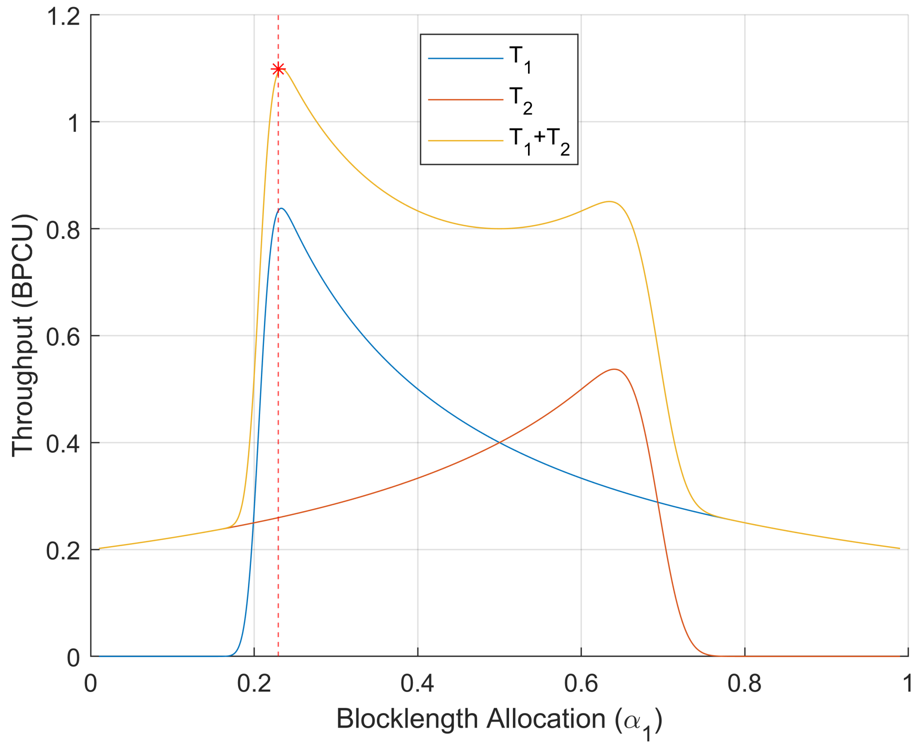

Figure 4.1: Multi-modal sum throughput case - $B = 200, |h_1|^2 = 5, |h_2|^2 = 1, P = 0dB$

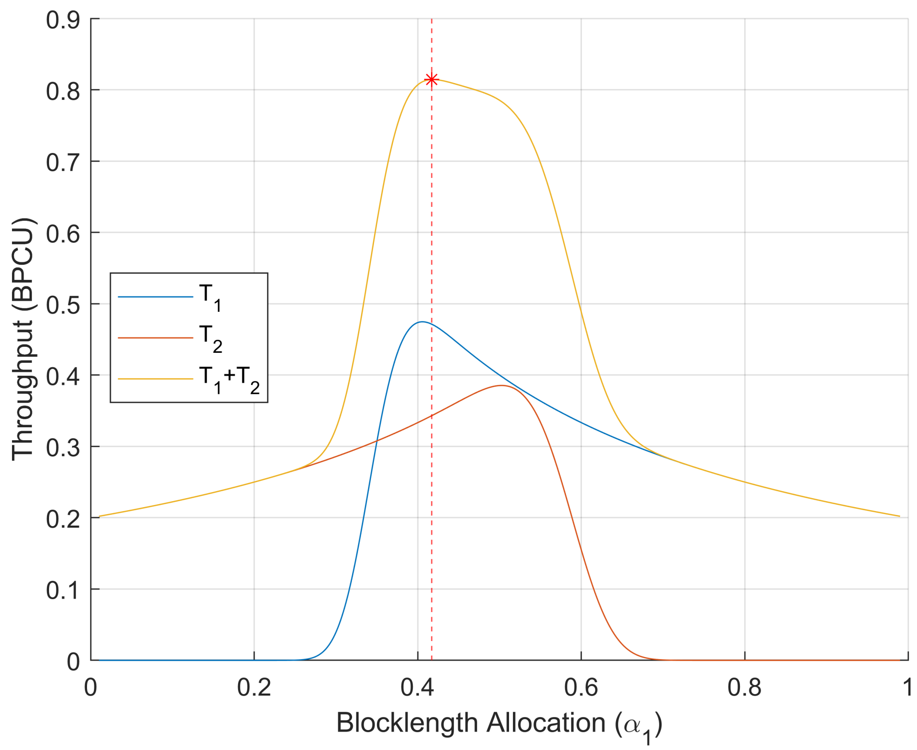

Figure 4.2: Uni-modal sum throughput case - $B = 200, |h_1|^2 = 0.75, |h_2|^2 = 0.5, P = 0dB$

4.1.2 Algorithm 1 Simulation Results

The first simulation investigated is the performance of the blocklength allocation algorithm 1 proposed on section 4.1.1. In order to produce the simulation and accompanying plots, the power of the BS was incremented, for each power level, 1000 channels realisations were generated. A exhaustive search finds the average optimum sum throughput, but is computationally expensive. Our efficient proposed algorithm also finds the average optimum sum throughput. For comparison purposes to show the benefits of optimization of blocklength, an equal blocklength allocation between actuators is shown ($\alpha_1 = \alpha_2 = 0.5$). This serves to contrast the benefit of optimisation and equal allocation.

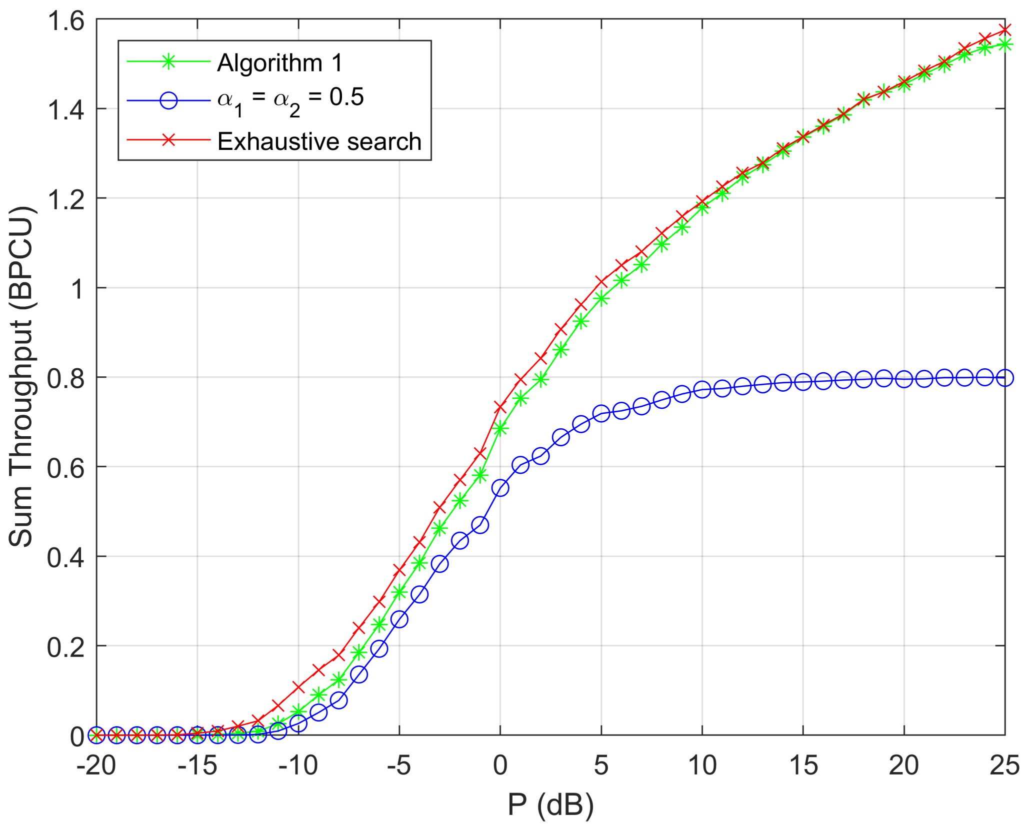

Figure 4.3: Performance of Algorithm 1 vs. BS Transmit Power

Viewing the performance of algorithm 1 we note a almost perfect match between the exhaustive search average optimised throughput and the proposed algorithm. Also worth noting is that low BS power (less than about 0dB) there is little difference between the optimised cases and the equal allocation case where $\alpha_1 = \alpha_2 = 0.5$. This would indicate that at low BS transmit power there is marginal gains optimally allocating resources, and that a simple allocation such as an even split is not too far away from optimal.

The opposite is the case at high transmit power and we note that as the transmit power increases the throughput of the optimised systems increases accordingly. The benefit of optimizing is higher the greater the BS transmit power. For example at 25dB BS transmit power the optimised sum throughput is twice the equal allocation sum throughput.

The equal allocation eventually plateaus due to the fact that this simulation used $B = 200$, using the equal allocation of blocklength meant each actuator was assigned a blocklength of 500. When the BS transmit power was sufficiently high, past about 10dB, the error computed by equation (4.2) was close to zero, as a result our sum throughput was $B/\alpha_1 N + B/\alpha_2 N \approx 200/500 + 200/500 = 0.8$

4.1.3 Optimization of Power Allocation

Power allocation can also be optimised. Introducing β_i as the power allocation factor and using equal blocklength allocation we can state the OMA power allocation throughput as expression (4.4) with $\alpha_1 = \alpha_2 = 0.5$.

However we must make one change to the SNR expression. Taking into account the power allocation OMA SNR is now

$$\gamma_i = \frac{\beta_i P |h_i|^2}{0.5 N_0} \tag{4.7}$$

The optimization problem can therefore be expressed as

$$\begin{aligned} \max_{\beta_i} \quad & T_{sum} \\ \text{s.t.} \quad & \beta_1 + \beta_2 = 1 \end{aligned}$$

Unlike the blocklength allocation system we can't employ the high-SNR approximation. We must deal with equation (4.4) without simplification.

In the case of equal blocklength allocation, we investigate the effects of power allocation. The sum throughput in the power allocation case is heavily dependent on the channel gain coefficients and depending on their value the optimization can be reasonably trivial, with a large flat peak. Other channel gain coefficients show a more complicated multi-modal function for sum throughput as shown in figure 4.4. In order to find the optimum power allocation irrespective of channel gains, a simple search algorithm is proposed below.

Algorithm 2 - Power Allocation Algorithm

Require: P, N
  $step \leftarrow 0.01$ ▷ set step size
  $\beta_1 \leftarrow 0$ ▷ initial value
  while $\beta_1 \leq 1$ **do**
    $\beta_2 \leftarrow 1 - \beta_1$
    $T_{sum} \leftarrow T_1(\beta_1) + T_2(\beta_2)$ ▷ compute sum throughput
    if $T_{sum} > T^{\star}_{sum}$ **then**
      $T^{\star}_{sum} \leftarrow T_{sum}$
      $\beta^{\star}_1 \leftarrow \beta_1$
    end if
    $\beta_1 \leftarrow \beta_1 + step$
  end while
Output: $\beta^{\star}_1, T^{\star}_{sum}$

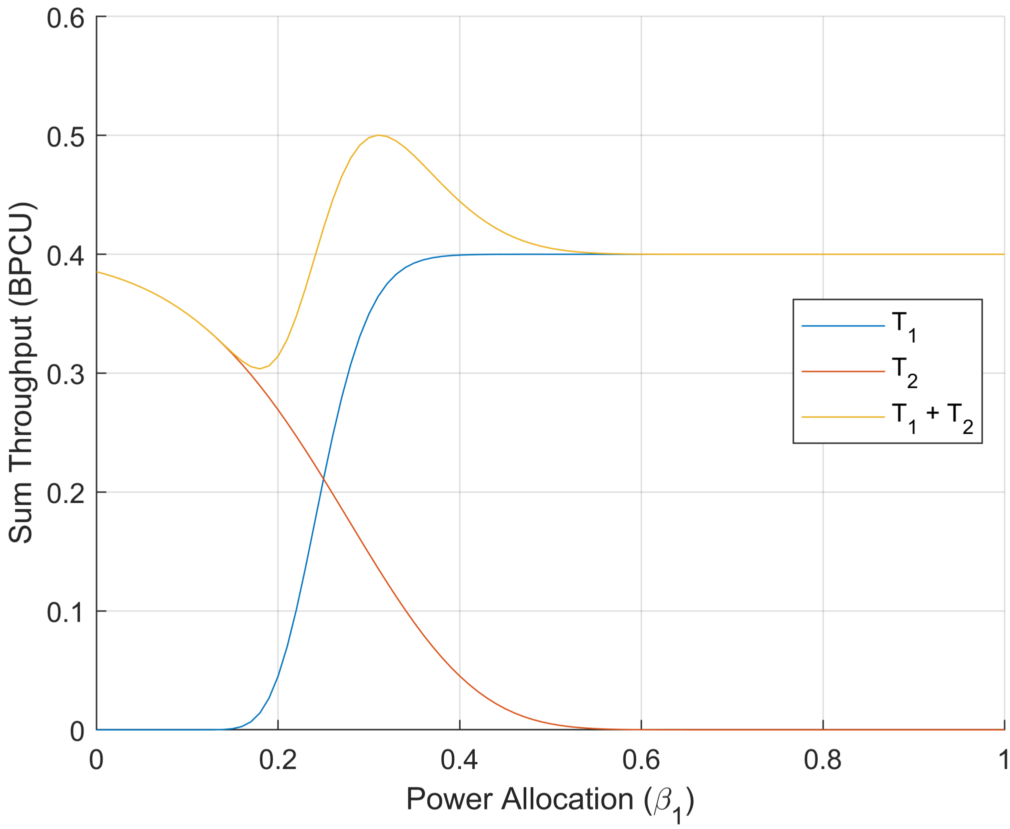

Figure 4.4: Throughput vs. Power Allocation - $B = 200, |h_1|^2 = 1.5, |h_2|^2 = 0.5, P = 0dB, N = 1000$

4.1.4 Algorithm 2 Simulation Results

In order to sufficiently investigate the performance of the power allocation algorithm, the transmit power of the BS was varied and the maximum sum throughput was computed, this was averaged over 1000 channel realisations. The simulation also computed the sum throughput for the case where the power allocation was the same for each user $\beta_1 = \beta_2 = 0.5$. For the simulation equal blocklength allocation between users was assumed.

From the figure 4.5 it is clear to see that there is a significant increase in average sum throughput for allocating power optimally between the 2 actuators. Of significant importance is the observation that the most significant increase in performance versus arbitrarily allocating power to each user equally is in the region of -10dB to 10dB. Outside of this region the transmit power is either too small to have an appreciably large throughput or the transmit power is high enough that the equal allocation results in low error probability communication.

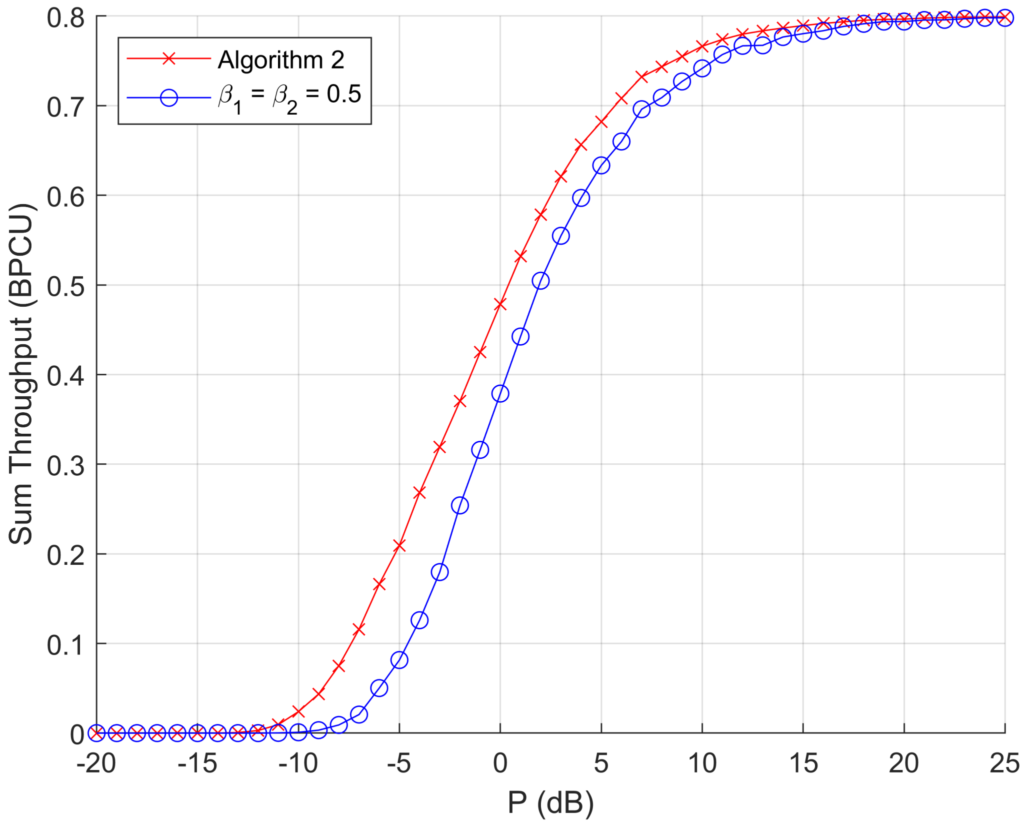

Figure 4.5: Performance of Algorithm 2 vs. BS Transmit Power

4.1.5 Comments on Sum Throughput Maximization

An efficient algorithm for allocation of blocklength and supporting evidence has been presented. It is clear the the allocation of blocklength is a significant factor in maximising sum throughput at high power. For a certain range of transmit powers (between -20dB and 0dB) there is not a significant difference between the optimised case and the equal allocation case. Following this it may be worthwhile optimizing allocation only when the transmit power from the BS is above 0dB. At high power there is a clear increase in average sum throughput and optimizing via the proposed algorithm achieves sum throughput very close to the global optimum.

A power allocation algorithm has been shown to achieve greater average sum throughput than equal allocation to each user, however this is only within a certain region of BS power. Outside of this BS power optimization yields little increase in average sum throughput and in certain applications may not be worthwhile undertaking.

As outlined previously however, the algorithms proposed to maximise sum throughput don't guarantee good performance when looking at the overall blocklength. Following this fact in section 4.2 we will resolve this with algorithms and analysis that takes into account the average sum throughput.

4.2 Average Sum Throughput Maximization

In the previous section we analysed the sum throughput for the time that the BS is serving each user and for a fixed number of information bits. This means that the optimization methods presented in 4.1 don't take into account the overall average sum throughput over the whole blocklength. As a consequence of this a more complete optimization process is proposed in this section. This is the principal contribution of this thesis.

We now consider average sum throughput, that is the sum throughput over the whole blocklength. Furthermore we don't consider a fixed number of information bits.

Referring back to the figure 3.8 we see that averaging over the whole blocklength gives the throughput for the actuator over the whole blocklength N, as opposed to over their specific blocklength allocation $\alpha_i N$. This subtle change will have a substantial effect on our analysis and results.

In 4.2.1 we maximise average sum throughput by blocklength allocation, a simple global optimum algorithm is proposed. The algorithm proposed in 4.2.1 does not ensure that each user gets an average minimum throughput, this is resolved in 4.2.3 where a proposed algorithm ensures fairness by allocating blocklength such that an average minimum throughput constraint for each user is met.

Lastly in 4.2.5 an algorithm is proposed for power allocation for OMA and NOMA. The algorithm ensures an average minimum throughput constraint is met.

4.2.1 Blocklength Allocation - Global Optimum

The OMA rate over the whole blocklength, i.e. the averaged rate $\bar{R}_i$ is expressed as,

$$\bar{R}_i = \alpha_i R_i = \alpha_i \left[C_i - \sqrt{\frac{V_i}{\alpha_i N}} \frac{Q^{-1}(\epsilon_i)}{\ln(2)} \right] \tag{4.8}$$

Expressing throughput through the averaged rate expression above

$$\bar{T}_i = \bar{R}_i(1 - \epsilon_i) \tag{4.9}$$

Where ϵ_i is the error rate at the i-th actuator, this is a design parameter for this simulation and as such it is chosen instead of computed.

From (4.9) we can define average sum throughput for 2 users as

$$\bar{T}_{sum} = \bar{T}_1 + \bar{T}_2 \tag{4.10}$$

In this case we will will define our optimization problem slightly differently. As opposed to computing the error rate ϵ_i, will will take a target value and design around that. ϵ_i must be less than the allowable error rate of the system ϵ_{MAX}

$$\begin{aligned} \max_{\alpha_i} \quad & \bar{T}_{sum} \\ \text{s.t.} \quad & \alpha_1 + \alpha_2 = 1 \\ & \epsilon_i \leq \epsilon_{MAX} \end{aligned}$$

We now propose an algorithm to allocate blocklength. It is clear from figure 4.6 that in order to achieve maximum sum rate, all of the blocklength should be allocated to the user with the better channel. Due to the all or nothing allocation achieved by this algorithm it achieves the global optimum, however it does not guarantee a minimum throughput to users.

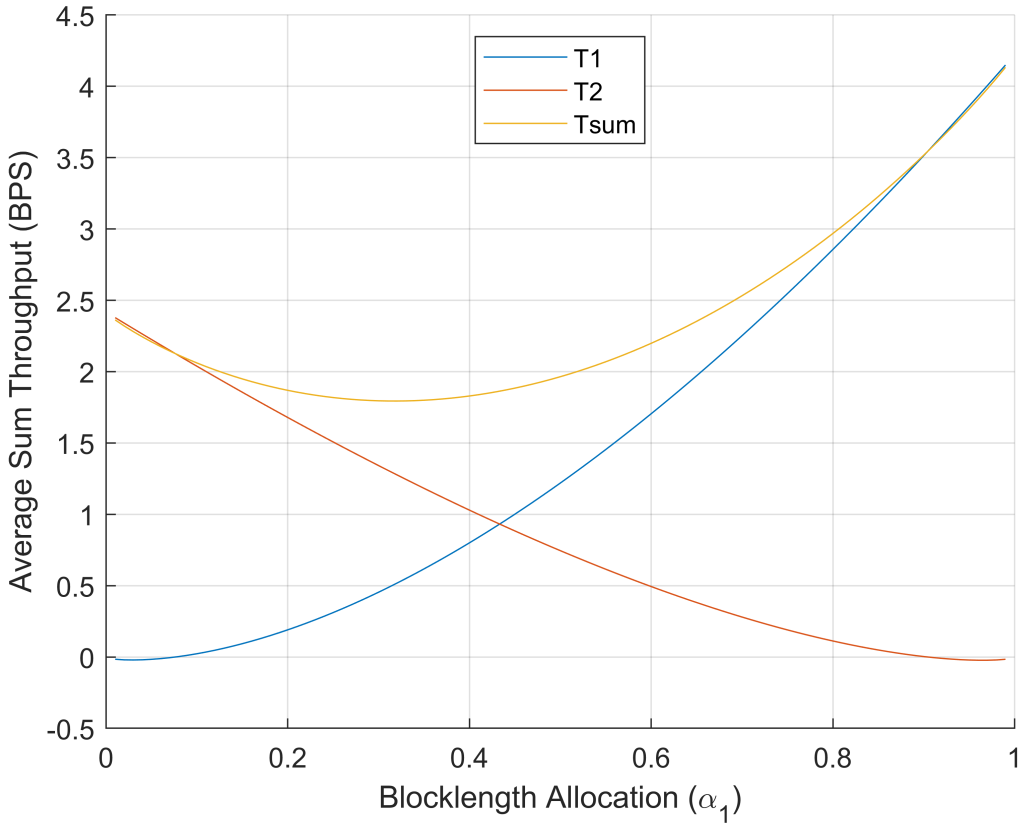

Figure 4.6: Average sum throughput case - $|h_1|^2 = 2, |h_2|^2 = 0.5, P = 10dB$

Algorithm 3 - Unconstrained Blocklength Allocation Algorithm

Require: P, N, h_1, h_2
  $\alpha_1 \leftarrow 0.5$ ▷ allocation will remain equal if channels equivalent
  $\alpha_2 \leftarrow 0.5$
  if $|h_1|^2 > |h_2|^2$ **then**
    $\alpha_1 \leftarrow 1$
    $\alpha_2 \leftarrow 0$
  else
    $\alpha_2 \leftarrow 1$
    $\alpha_1 \leftarrow 0$
  end if
Output: $\alpha_1^\star$

4.2.2 Algorithm 3 Simulation Results

Investigating the average sum throughput for a range of BS power levels delivers insight to the performance of the algorithm outlined above.

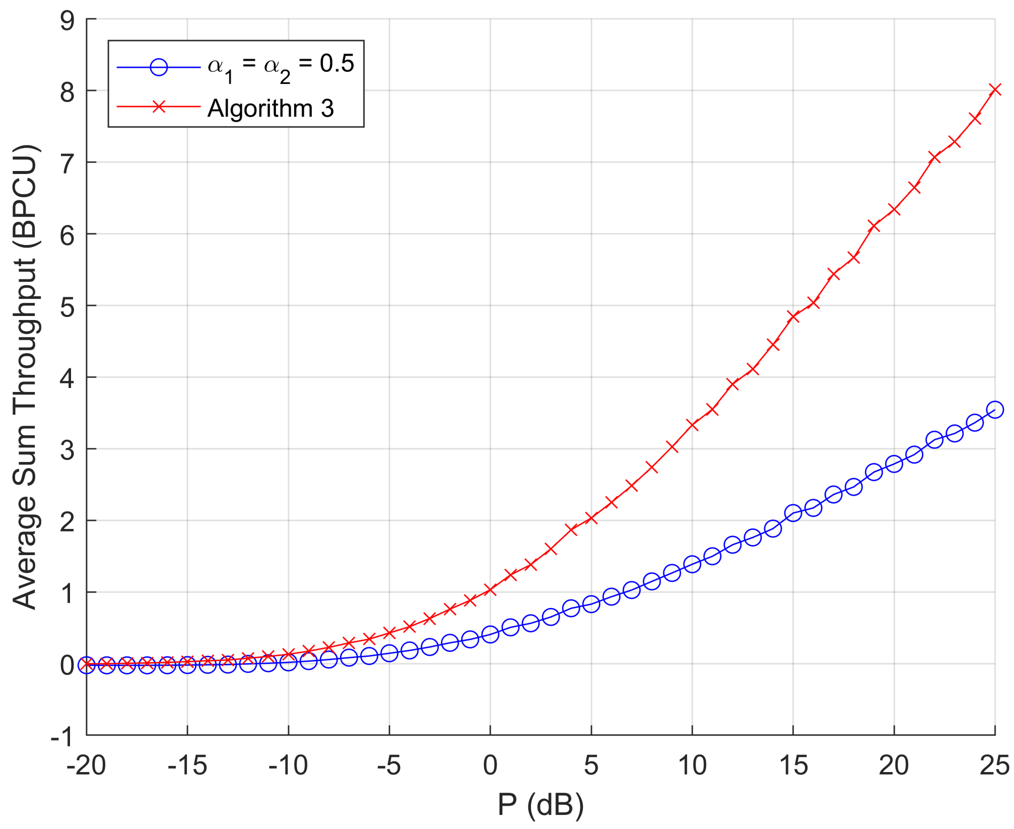

Figure 4.7: Performance of Algorithm 3 vs BS Transmit Power

It is clear that allocation using the proposed solution results in a large performance gain over equal allocation. Furthermore the simple algorithm finds the global optimum with very low computational complexity.

The low complexity of this algorithm is about the only advantage it has, ensuring a minimum throughput to each user is an important requirement for URLLC systems and this motivates section 4.2.3 where an minimum throughput constraint is imposed.

4.2.3 Blocklength Allocation - With Minimum Throughput Constraint

Examining figure 4.7 and figure 4.6, it is clear to see that in order to maximise average sum throughput, the algorithm proposed should allocate all of the blocklength to the user with the better channel. Depending on the specific application this may not be a desirable system characteristic, as it may result in high latency for users with bad channels, or users faraway from the BS not being served. In order to impose fairness to ensure that a minimum throughput is achieved we propose an alternative to algorithm 3.

The key difference is that the algorithm proposed below seeks to find the maximum average sum throughput whilst ensuring that both user still achieve a minimum throughput, T_{MIN}.

$$\begin{aligned} \max_{\alpha_i} \quad & \bar{T}_{sum} \\ \text{s.t.} \quad & \alpha_1 + \alpha_2 = 1 \\ & \epsilon_i \leq \epsilon_{MAX} \\ & \bar{T}_i \geq T_{MIN} \end{aligned}$$

Algorithm 4 - Constrained Blocklength Allocation Algorithm

Require: P, N, h_1, h_2, T_{MIN}
  $\alpha_1 \leftarrow 0.5$ ▷ Initially equal allocation
  $\alpha_2 \leftarrow 0.5$
  $step \leftarrow 0.01$ ▷ Set step size
  while $T \geq T_{MIN}$ **do**
    $\alpha_2 \leftarrow 1 - \alpha_1$
    if $|h_1|^2 > |h_2|^2$ **then**
      $\alpha_1 \leftarrow \alpha_1 + step$
      $T \leftarrow R_2(\alpha_2) * (1 - \epsilon_2)$ ▷ Compute smaller throughput
    else
      $\alpha_1 \leftarrow \alpha_1 - step$
      $T \leftarrow R_1(\alpha_1) * (1 - \epsilon_1)$ ▷ Compute smaller throughput
    end if
  end while
Output: $\alpha_1^\star$

This simple algorithm works by allocating more blocklength to the user with the better channel until the minimum throughput constraint for the user with the worse channel is reached.

4.2.4 Algorithm 4 Simulation Results

By the very nature of the fact that additional constraints have been imposed it is clear that the maximum average sum throughput for the overall system will have been reduced. However, the algorithm ensures a degree of fairness by ensuring that each user gets at least their minimum throughput requirement.

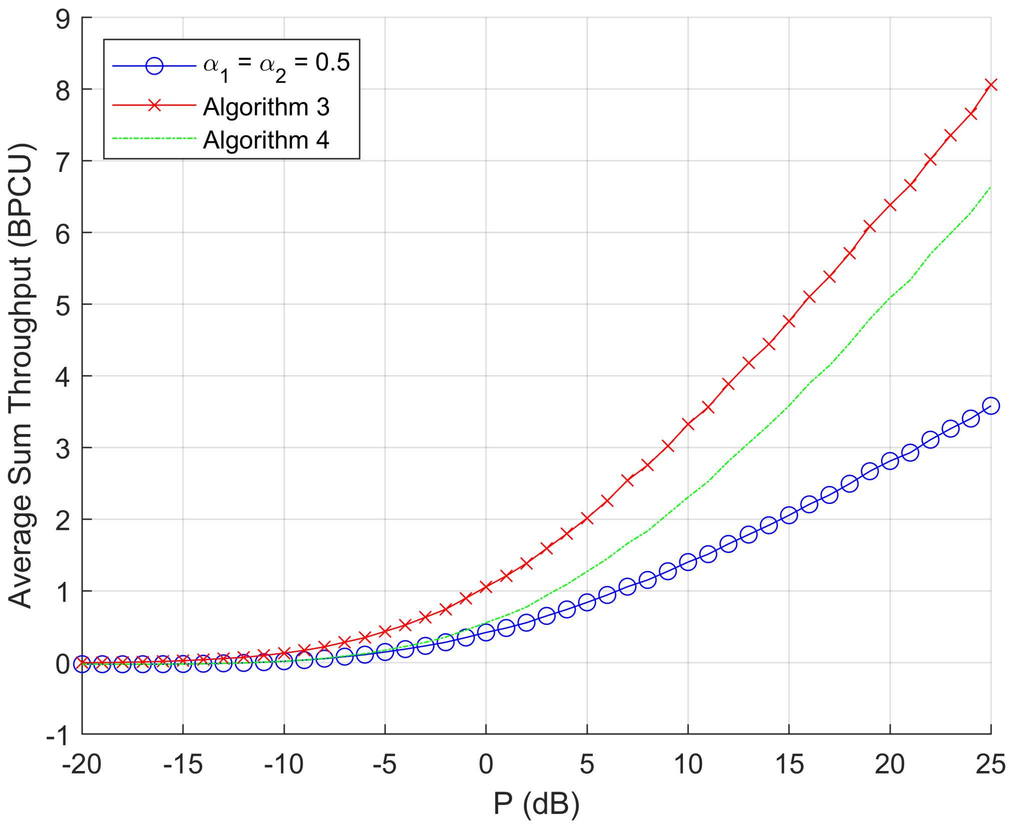

Figure 4.8: Performance of Algorithm 3 and 4 vs BS Transmit Power

It us clear from figure 4.8 that there is a reduction in the average sum throughput when imposing a minimum throughput constraint. However, depending on the application there may be merit in allocating resources this way as it ensures that each user has a guaranteed average minimum throughput.

4.2.5 Power Allocation - OMA and NOMA

In order to allocate power to each user in OMA and OMA we will develop a simple algorithm to optimise average sum throughput. For the OMA case equal blocklength allocation was assumed $\alpha_1 = \alpha_2 = 0.5$. In the NOMA case as outlined in section 3.4.3, communication to the 2 users occurs over the whole blocklength.

Recall that in section 3.4.3 in figure 3.13 it was shown that NOMA yields a greater rate when compared with OMA for a given channel. It follows that throughput would follow a similar trend as it is merely a fraction of rate.

In order to effectively allocate power we will propose a search based algorithm to solve NOMA and OMA power allocation. CSI will be used for some slight efficiency gain, by reducing the search complexity. The algorithm will also ensure a minimum throughput to each user, by using a minimum average sum throughput constraint.

For OMA we use the power allocation throughput expression (4.9) where the SNR is given by (4.7).

For NOMA we use expression (4.9) with the important change $\bar{R}_i$ is given by expression (3.8). Crucially we must also recall that γ_i in the NOMA case is the SINR at the actuator as defined by in section 3.4.3 by expression (3.5).

With the important relevant expressions defined we can use the optimization algorithm 5 to compare the average sum throughput for a range of BS transmit powers for both OMA and NOMA.

$$\begin{aligned} \max_{\beta_i} \quad & \bar{T}_{sum} \\ \text{s.t.} \quad & \beta_1 + \beta_2 = 1 \\ & \epsilon_i \leq \epsilon_{MAX} \\ & \bar{T}_i \geq T_{MIN} \end{aligned}$$

The basic principle behind this algorithm is that in the OMA case, in order to maximise average sum throughput, the power allocation to the user with the better channel is increased until the user with the worse channel reaches their minimum throughput constraint.

The NOMA case is slightly different. In order to maximise average sum throughput in the NOMA case we increase power allocation until the user with the better channel reaches their throughput constraint.

Algorithm 5 - Constrained Power Allocation Algorithm

Require: P, N, h_1, h_2, T_{MIN}
  $\beta_1 \leftarrow 0.5$ ▷ Start with equal allocation
  $\beta_2 \leftarrow 0.5$
  $step \leftarrow 0.01$ ▷ Set step size
  while $T \geq T_{MIN}$ **do**
    if NOMA **then**
      $\beta_1 \leftarrow \beta_1 - step$ ▷ NOMA - Allocate less power to user with better channel
      $\beta_2 \leftarrow 1 - \beta_1$
      $T \leftarrow R_1(\beta_1) * (1 - \epsilon_1)$ ▷ Compute smaller throughput
    else
      $\beta_1 \leftarrow \beta_1 + step$ ▷ OMA - Allocate more power to user with better channel
      $\beta_2 \leftarrow 1 - \beta_1$
      $T \leftarrow R_2(\beta_2) * (1 - \epsilon_2)$ ▷ Compute smaller throughput
    end if
  end while
Output: $\beta_1^\star$

4.2.6 Algorithm 5 Simulation Results

In order to show the performance of algorithm 5, we will vary the BS power level, take a large amount of channel realisations, apply the algorithm and average the resulting sum throughput. A significant point to recall is that as described in section 3.4.3, the NOMA expressions derived are for the case where A_1 has a better channel than A_2. As a result for simulation, channels were randomly generated and the better channel was assigned to A_1. This ensures that the expressions used in the simulation hold true such that SIC can occur.

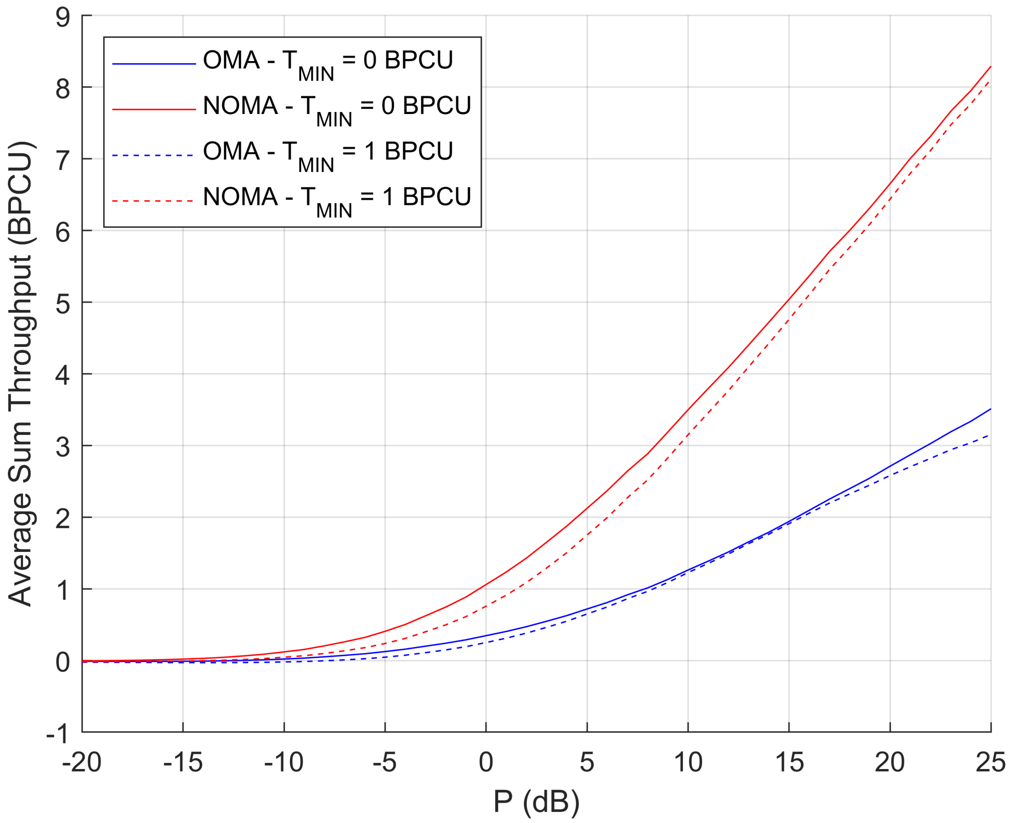

Figure 4.9: Performance of Algorithm 5 vs. BS Transmit Power

Figure 4.9 shows some interesting and desirable characteristics. The first point to note is that as NOMA dominates OMA in terms of average sum throughput. As regards the performance of algorithm 5, we note that as algorithm 5 applies a minimum throughput constraint it will have a lower average sum throughput than the unconstrained case. This is clear from figure 4.9, where the value of $T_{MIN} = 0$, this is essentially the case where there is no minimum throughput constraint and is shown in figure 4.9 by the solid lines on the plot. Adding the constraint clearly reduces the average sum throughput, however it ensures a minimum throughput.

We note from the figure that as BS power increases, the average sum throughput increases linearly beyond about 5dB. Another point to note is that the linear increase for NOMA is greater than that of OMA, the slope of the linear NOMA curve is greater than that of OMA. This means that an increasing the BS power by a fixed amount will have a greater return on average sum throughput if NOMA is used.

The algorithm proposed has some limitations and depending on the combination of channel gains may not perform well, for example the OMA case of the algorithm ensures that the optimum average sum throughput is reached. However in the NOMA case the algorithm does not ensure an optimum sum throughput is reached. Despite this we note from figure 4.9 that on average the algorithm does a reasonable job at allocation of power between users.

An improvement of the NOMA case of algorithm would use the channel gain coefficients, to compute the bounds of the power allocation factors. We recall that the signal intended for actuator A_2, must have greater SINR than that intended for A_1, this is to ensure successful SIC. This would help reduce the search region and reduce complexity.

Also worth noting is that even with the drawbacks of algorithm 5 outlined above, the NOMA performance of the algorithm still dominates the optimal OMA performance of the algorithm.

4.2.7 Comments on Average Sum Throughput Optimization Results

From the results above we see that we have proposed 3 algorithms for solving the problem of maximising average sum throughput.

If there is a no minimum throughput required by the system then the simple algorithm 3 proposed in section 4 provides extremely efficient blocklength allocation.

In the case that the system has a minimum throughput constrain then algorithm 4 allocates blocklength by maximising average sum throughput, whilst ensuring that the minimum throughput requirements are observed.

Algorithm 5 allocates power for OMA and NOMA transmission and shows better overall performance in terms of average sum throughput for the NOMA transmission method. Furthermore this algorithm ensures that an minimum average throughput is imposed.

Algorithms 4 and 5 are highly applicable to real world URLLC IoT applications where multiple users require a certain guaranteed throughput.

Chapter 5: **Discussion and Conclusions**

The optimization problems explored in chapter 4 provide some interesting discussion points. The first being the reason for analysing sum throughput and also average sum throughput.

Sum throughput, as analysed in section 4.1, is the maximisation of the throughput for the time that the actuator is being served, i.e. the busy period. As outlined this does not guarantee overall performance, despite this the optimization algorithms proposed are novel, in particular algorithm 1 which uses channel gains to accurately converge to an optimal value close to global optimum.

The optimization considered in section 4.2 is more generally applicable as it seeks to maximise average sum throughput, that is the throughput over the whole blocklength N. This is of greater interest for practical IoT applications. Because of the quasi-static assumption that was made regarding Rayleigh channels, that the channel gain coefficients remain constant over the whole blocklength, we have an interesting optimization scenario. With the constant channel gain coefficients it is clear that to maximise average sum throughput we should allocate the whole blocklength to the user with the better channel, this is shown in figure 4.7. This all or nothing allocation is the approach taken by algorithm 3. In this case however there is no guaranteed rate for the user with the worse channel, hence algorithm 4 solves this problem. Imposing this minimum throughput constraint however, as expected with optimization problems only serves to reduce the overall average sum throughput as shown in figure 4.9.

The comparison of OMA and NOMA in section 4.2.6 shows that for a given BS power level the average sum throughput will be higher in the NOMA system. This is a significant result as it shows that NOMA with SIC is a good candidate for URLLC applications. Furthermore algorithm 5 proposes a method to solve the power allocation problem for both OMA and NOMA.

The solutions and analysis presented in this thesis have provided great learning opportunity for the author, advancing research skills, technical writing, MATLAB code and presentation skills. The author will undoubtedly take these learnings with him in the next phase of his engineering career.

Chapter 6: **Impact of Research**

The solutions presented in this thesis are applicable to the global problem of the ever increasing demand for higher and higher data speeds. These high data speeds will be critical to enable the next generation of technologies to be layered upon. Driver-less cars, smart factories and the haptic internet will all be achievable provided that URLLC is deployed leveraging solutions such as those outlined in this thesis. The availability of URLLC in 5G and beyond will lead to a wave of new innovations, with this comes economic stimulation, technical jobs and new challenges for future engineers.

An environment or city that has a functioning URLLC network will likely reap the rewards of such a networks as it will enable an innovative and enterprising economy, enabling industry to leverage cutting edge technology to increase output, reduce emissions and streamline industrial processes.

The result of many of the innovations that will be layered over the next generation of networks, is that the quality of life will be improved. The innovations such as remote patient monitoring will undoubtedly improve peoples lives, democratising access to quality healthcare. Drone delivery will enable the speedy delivery of donated organs and blood to save lives. Smart cities will intelligently manage traffic and driver-less cars to reduce carbon emissions and eliminate traffic and accidents.

The greatest impact of URLLC is not in the network innovations themselves, but rather the virtually unlimited possibilities such a network will unleash in future technologies.

There is also some potential abuses of the proposed solutions. The same benefits of low-latency and ultra-reliable communication will also be available to individuals who seek to cause harm to others. For example URLLC may be abused in order to remotely attack civilians using drones, furthermore robust URLLC security is needed to ensure data privacy but also to mitigate against cyber-attacks and cyber-terrorism which will likely be a feature of next generation warfare and terrorism. This is a typical concern with emerging technologies in the communications field, their misuse enables bad actors the access the same high reliability low latency performance, thus improving their ability to cause harm and destruction.

As engineers there is a ethical responsibility to consider the impacts of proposed innovations. On balance the overall impact of this research is that it will help guide future network designers to ensure high-reliability and low latency communications within their networks.

Assuming abuse of the technology is limited and appropriate punishment is enforced for misuse, URLLC will enable delivery much needed solutions for the world. Smart factories will drive down production costs, smart cities will better manage resources such that carbon emissions are minimised and traffic eliminated.

Chapter 7: **Suggestions for Future Work**

Despite the significant contributions outlined in this thesis, there are still significant challenges in URLLC which require further investigation, some of these challenges include

- This thesis only examined downlink communication. A worthwhile extension to this work would be to examine how the solutions and optimization algorithms cope when uplink and downlink communication is a system requirement. Further analysis on uplink optimization alone would also provide interesting results which when combined with this thesis would provide clear guidelines for uplink and downlink optimization for URLLC.

- Algorithms 1 and 4 are extremely efficient in the sense that they advantageously use CSI to simplify the optimization problem by using some general observations about the sum throughput, which result in algorithms that deliver an optimum very close to the global optimum obtained by exhaustive search. It is likely that algorithms 2,3 and 5 could in a similar way benefit from some intelligent observations as to their general behaviour and use this to simplify the optimization complexity. In the authors opinion a good starting point for this would be using the tolerable error rate to set a search region for blocklength, this would reduce the search space. Similarly in the NOMA case of algorithm 5, the channel gain coefficients could be used to find the power allocation bounds to ensure that the SINRs enable successful SIC.

- Secrecy and information leakage rate is investigated early on in this thesis, but when it comes to the proposed solutions secrecy is not considered. A worthwhile extension to this work would be to re-evaluate the results, this time with a secrecy requirement. It is clear that adding secrecy would reduce the throughput, but the exact performance trade-off would provide insight into the overhead imposed by secrecy. Furthermore the algorithms proposed would need slight alterations to take into account the the secrecy requirement, this would further provide interesting results when compared to the results in this thesis. Secrecy is clearly a very important aspect of future communication networks for the reasons outlined in chapter 6. Investigation of PLS in URLLC would likely yield interesting insights.

- The thesis investigates blocklength allocation and power allocation but provides no insight into the joint optimization of blocklength and power allocation. It is likely that optimization of both blocklength and power allocation would yield greater sum throughput results. A small extension to the thesis investigating joint optimization would provide greater insight and would effectively allocate blocklength and power in one algorithm as opposed to 2 separate algorithms.

- Algorithm 1 relies on a fixed number of information bits for each actuator, this might be a valid assumption for many systems, however more crucially - the number of information bits affects the shape of the throughput curve. A key step of the proposed algorithm is taking the best channel to locate the optimum. With a different number of information bits between the 2 users the algorithm wouldn't be guaranteed to converge by just taking the best channel gain coefficient as proposed. Instead the number of information bits and the channel gain coefficient would need to be combined in some way such that the correct peak identified. This would make the optimization algorithm more robust and would converge for a variety of different channel gain coefficients and differing numbers of information bits in the packets.

- For algorithm 1, at very high BS transmit power beyond 30dB, the peak sum throughput becomes incredibly sharp or sensitive over blocklength allocation factor α_i. As a result the algorithm does not converge with the proposed step size L. A smaller step-size is required to ensure that the algorithm converges to the optimum blocklength allocation. In order to remedy this, a better designed method of computing L depending on base station transmit power would ensure that the algorithm converges rapidly for transmit power level.

- All of the results only show the performance for 2 users. It is hard to imagine how the proposed solutions might work with an increased number of users. Analysis for more than 2 users would provide greater insight. Extending analysis and optimization to more users would highlight the benefits of the solutions proposed but also their downfalls, thus pointing to further future work.

- In the case of NOMA, only power domain NOMA with SIC was considered in this work. A detailed analysis of the different NOMA decoding methods such as joint decoding, C-NOMA (cooperative) etc. are compared. As shown in this thesis NOMA dominates OMA for URLLC applications, it would be worthwhile exploring all the different NOMA methods. From this analysing the different methods and providing overarching guidelines on their specific use cases (MMTC, EMBB, URLLC) would aid engineers as NOMA becomes increasingly discussed as a solution for next generation networks.

- A further extension of the NOMA decoding methods investigation above, would be to investigate the capabilities of simple IoT nodes to effectively perform such signal processing such as SIC and joint decoding. It is a concern of the author that in order to produce numerous cheap battery-powered IoT nodes, they will likely have limited computational power. From this an investigation of the power requirements of the IoT node battery and processor in order to achieve such decoding would provide significant contribution for practical IoT node designers.

- The energy efficiency of the proposed algorithms has not been considered. A worthwhile are for future work would be examining URLLC resource allocation with metrics focused on energy efficiency. As highlighted in the literature review there are conflicting results as regards if NOMA or OMA is more efficient for communications. A detailed analysis taking into account uplink and downlink transmissions and results showing which transmission methods and decoding methods could be used to reduce energy usage.

Acknowledgements

This thesis and previous technical work has been guided by the expertise of Professor Mark Flanagan and Dr. Chinmoy Kundu, their experience, knowledge and teaching has been immensely valuable to the author throughout the project.

This thesis is the final piece in what has been an enjoyable (at times even stressful) 5 years at UCD. And there are some individuals whose contribution the author would like to formally acknowledge.

Thanks to the fellow engineers in the project lab, who managed to find creative ways of maintaining morale when the going got tough. In particular Louis, Patrick, Adam and Shea who kindly accepted me into their clique when all my other friends dropped out.

Thanks also to my parents and 2 brothers who have kindly covered for me at family gatherings, for the past 5 years, from which I abstained due to "college commitments".

Thanks also to Elaine, who provides endless support and encouragement. Elaine now knows more about URLLC than any civil engineer ought to.

Bibliography

1. Wang, H.-M., Yang, Q., Ding, Z. & Poor, H. V. Secure Short-Packet Communications for Mission-Critical IoT Applications. *IEEE Transactions on Wireless Communications* **18,** 2565–2578 (2019).

2. Poor, H. V., Goldenbaum, M. & Yang, W. *Fundamentals for IoT networks: secure and low-latency communications* in (Jan. 2019), 362–364. ISBN: 9781450360944.

3. Popovski, P. *Ultra-reliable communication in 5G wireless systems* in *1st International Conference on 5G for Ubiquitous Connectivity* (2014), 146–151.

4. Ji, H. *et al.* Ultra-reliable and low-latency communications in 5G downlink: Physical layer aspects. *IEEE Wireless Communications* **25,** 124–130 (2018).

5. Schulz, P. *et al.* Latency Critical IoT Applications in 5G: Perspective on the Design of Radio Interface and Network Architecture. *IEEE Communications Magazine* **55,** 70–78 (2017).

6. Poor, H. V. Information and inference in the wireless physical layer. *IEEE Wireless Communications* **19,** 40–47 (2012).

7. Zheng, T.-X., Wang, H.-M., Ng, D. W. K. & Yuan, J. Physical-Layer Security in the Finite Blocklength Regime Over Fading Channels. *IEEE Transactions on Wireless Communications* **19,** 3405–3420 (2020).

8. Ren, H., Pan, C., Deng, Y., Elkashlan, M. & Nallanathan, A. Joint Power and Blocklength Optimization for URLLC in a Factory Automation Scenario. *IEEE Transactions on Wireless Communications* **19,** 1786–1801 (2020).

9. Ren, H., Pan, C., Deng, Y., Elkashlan, M. & Nallanathan, A. Resource Allocation for Secure URLLC in Mission-Critical IoT Scenarios. *IEEE Transactions on Communications* **68,** 5793–5807 (2020).

10. Feng, C., Wang, H.-M. & Vincent Poor, H. Reliable and Secure Short-Packet Communications. *IEEE Transactions on Wireless Communications,* 1–1 (2021).

11. Yang, W., Schaefer, R. F. & Poor, H. V. Wiretap Channels: Nonasymptotic Fundamental Limits. *IEEE Transactions on Information Theory* **65,** 4069–4093 (2019).

12. Yang, W., Durisi, G., Koch, T. & Polyanskiy, Y. Quasi-Static Multiple-Antenna Fading Channels at Finite Blocklength. *IEEE Transactions on Information Theory* **60,** 4232–4265 (2014).

13. Durisi, G., Koch, T. & Popovski, P. Toward Massive, Ultrareliable, and Low-Latency Wireless Communication With Short Packets. *Proceedings of the IEEE* **104,** 1711–1726 (2016).

14. Ren, H., Pan, C., Deng, Y., Elkashlan, M. & Nallanathan, A. *Resource Allocation for URLLC in 5G Mission-Critical IoT Networks* in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)* (2019), 1–6.

15. Polyanskiy, Y., Poor, H. V. & Verdu, S. Channel Coding Rate in the Finite Blocklength Regime. *IEEE Transactions on Information Theory* **56,** 2307–2359 (2010).

16. Yang, W., Schaefer, R. F. & Poor, H. V. *Secrecy-reliability tradeoff for semi-deterministic wiretap channels at finite Blocklength* in *2017 IEEE International Symposium on Information Theory (ISIT)* (2017), 2133–2137.

17. Sabuj, S. R., Khan, M. R. H. & Ziad, M. O. *An Analytical Investigation of Finite Blocklength Information Theory* in *2020 IEEE Region 10 Symposium (TENSYMP)* (2020), 568–571.

18. Dai, H., Zhang, C., Luo, J., Li, C. & Wang, B. *QoE- Driven Resource Allocation for Secure URLLC in 6G-Enabled IoT Networks* in *2021 13th International Conference on Wireless Communications and Signal Processing (WCSP)* (2021), 1–5.

19. Chen, J., Zhang, L., Liang, Y.-C., Kang, X. & Zhang, R. Resource Allocation for Wireless-Powered IoT Networks With Short Packet Communication. *IEEE Transactions on Wireless Communications* **18,** 1447–1461 (2019).

20. Ren, H., Pan, C., Deng, Y., Elkashlan, M. & Nallanathan, A. *Transmit Power Minimization for Secure Short-packet Transmission in a Mission-Critical IoT Scenario* in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference* (2020), 1–6.

21. Sun, C. *et al.* Optimizing Resource Allocation in the Short Blocklength Regime for Ultra-Reliable and Low-Latency Communications. *IEEE Transactions on Wireless Communications* **18,** 402–415 (2019).

22. Farhat, J., Brante, G. & Souza, R. D. *Secure Throughput Optimization of Selective Decode-and-Forward with Finite Blocklength* in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)* (2018), 1–5.

23. Hu, Y., Li, Y., Gursoy, M. C., Velipasalar, S. & Schmeink, A. Throughput Analysis of Low-Latency IoT Systems With QoS Constraints and Finite Blocklength Codes. *IEEE Transactions on Vehicular Technology* **69,** 3093–3104 (2020).

24. Xie, Y., Ren, P., Wang, Y. & Li, J. *Power Consumption-Oriented Resource Allocation Strategy for Ultra-Reliable Low-Latency Communication* in *2019 IEEE Wireless Communications and Networking Conference (WCNC)* (2019), 1–6.

25. Mugisha, R., Mahmood, A., Abedin, S. F., Beltramelli, L. & Gidlund, M. *Joint Power and Blocklength Allocation for Energy-Efficient Ultra- Reliable and Low- Latency Communications* in *2021 17th International Symposium on Wireless Communication Systems (ISWCS)* (2021), 1–6.

26. Islam, S. M. R., Avazov, N., Dobre, O. A. & Kwak, K.-s. Power-Domain Non-Orthogonal Multiple Access (NOMA) in 5G Systems: Potentials and Challenges. *IEEE Communications Surveys Tutorials* **19,** 721–742 (2017).

27. Saito, Y. *et al. Non-Orthogonal Multiple Access (NOMA) for Cellular Future Radio Access* in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)* (2013), 1–5.

28. Kumar, V., Ding, Z. & Flanagan, M. F. On the Effective Rate of NOMA in Underlay Spectrum Sharing. *IEEE Transactions on Vehicular Technology* **70,** 12220–12225 (2021).

29. Kumar, V., Cardiff, B. & Flanagan, M. F. Fundamental Limits of Spectrum Sharing for NOMA-Based Cooperative Relaying Under a Peak Interference Constraint. *IEEE Transactions on Communications* **67,** 8233–8246 (2019).

30. Kumar, V., Flanagan, M. F., Da Costa, D. B. & Tran, L.-N. *On the Secrecy Rate of Downlink NOMA in Underlay Spectrum Sharing with Imperfect CSI: Invited Paper* in *2021 28th International Conference on Telecommunications (ICT)* (2021), 167–173.

31. Yu, Y., Chen, H., Li, Y., Ding, Z. & Vucetic, B. On the Performance of Non-Orthogonal Multiple Access in Short-Packet Communications. *IEEE Communications Letters* **22,** 590–593 (2018).

32. Sun, X. *et al.* Short-Packet Downlink Transmission With Non-Orthogonal Multiple Access. *IEEE Transactions on Wireless Communications* **17,** 4550–4564 (2018).

33. Chen, J., Zhang, L., Liang, Y.-C. & Ma, S. Optimal Resource Allocation for Multicarrier NOMA in Short Packet Communications. *IEEE Transactions on Vehicular Technology* **69,** 2141–2156 (2020).

34. Yao, J., Zhang, Q. & Qin, J. Joint Decoding in Downlink NOMA Systems with Finite Blocklength Transmissions for Ultra-Reliable Low-Latency Tasks. *IEEE Internet of Things Journal,* 1–1 (2022).

35. Ryu, W. J. & Shin, S. Y. *Power Allocation for URLLC Using Finite Blocklength Regime in Downlink NOMA Systems* in *2019 International Conference on Information and Communication Technology Convergence (ICTC)* (2019), 770–773.

Appendices

Appendix A

Derivative of T_i w.r.t α_i

$$\frac{dT_i}{d\alpha_i} \approx \frac{d}{d\alpha_i}\left(\frac{B}{\alpha_i N}\left(1 - Q\left(\ln 2\sqrt{\alpha_i N}\left(C_i - \frac{B}{\alpha_i N}\right)\right)\right)\right)$$

$$\frac{\mathrm{dT}_i}{d\alpha_i} \approx \frac{-B}{\alpha_i^2 N} - Q\left(\ln 2\sqrt{\alpha_i N}\ \left(\alpha_i \log_2\left(1 + \frac{P|h_i|^2}{\alpha_i N_0}\right) - \frac{B}{\alpha_i N}\right)\right)\left(-\frac{B}{\alpha_i^2 N}\right) +$$
$$\left(\frac{B}{\alpha_i N}\left(Q'\left(\ln 2\sqrt{\alpha_i N}\ \left(\alpha_i \log_2\left(1 + \frac{P|h_i|^2}{\alpha_i N_0}\right) - \frac{B}{\alpha_i N}\right)\right)\right.\right.$$

$$\left(\frac{3N\ N_0 \ln\left(1 + \frac{P|h_i|^2}{\alpha_i N_0}\right)\alpha_i^3 + \left(3P|h_i|^2 \mathrm{N}\ln\left(1 + \frac{P|h_i|^2}{\alpha_i N_0}\right) - 2P|h_i|^2 N\right)\alpha_i^2 + \ln 2BN_0\alpha_i + \ln 2BP|h_i|^2}{2\alpha_i\sqrt{N\alpha_i}\,(N_0\alpha_i + P|h_i|^2)}\right)$$

Appendix B

Differentiation of Q function

Recall that $Q(x) = 1 - \Phi(x)$, where $\Phi(x)$ is the CDF of the standard Gaussian distribution. Hence $Q'(x) = -\Phi'(x)$. The differential of the CDF is the PDF, hence

$$Q'(x) = \frac{-1}{\sqrt{2\pi}} \exp -\left(\frac{x^2}{2}\right)$$