AFRL-AFOSR-UK-TR-2012-0009

Application of Financial Risk-reward Theory to Link and Network Optimization

Aarne O. Mämmelä

VTT Technical Research Centre of Finland
Telecommunications, Communication Platforms
Kaitoväylä 1
Oulu, Finland FI-90571

EOARD SPC 08-4002

Report Date: October 2011

Final Report for 01 October 2008 to 01 October 2011

Air Force Research Laboratory
Air Force Office of Scientific Research
European Office of Aerospace Research and Development
Unit 4515 Box 14, APO AE 09421

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
11-10-2011	Final Report	01 October 2008 – 01 October 2011

4. TITLE AND SUBTITLE

Application of Financial Risk-reward Theory to Link and Network Optimization

5a. CONTRACT NUMBER

FA8655-08-C-4002

5b. GRANT NUMBER

SPC 08-4002

5c. PROGRAM ELEMENT NUMBER

61102F

6. AUTHOR(S)

Professor Aarne O. Mämmelä

5d. PROJECT NUMBER

5d. TASK NUMBER

5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

VTT Technical Research Centre of Finland
Telecommunications, Communication Platforms
Kaitovayla 1
Oulu, Finland FI-90571

8. PERFORMING ORGANIZATION REPORT NUMBER

N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

EOARD
Unit 4515 BOX 14
APO AE 09421

10. SPONSOR/MONITOR'S ACRONYM(S)

AFRL/AFOSR/RSW (EOARD)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

AFRL-AFOSR-UK-TR-2012-0009

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In this work, we have introduced a general framework for analysis and optimization of adaptive transmission systems in information-unstable channels. In information-unstable channels, the information density does not converge to a one-point measure and the maximum achievable transmission rate is seen as a random variable because it depends on the actual channel state. For that reason, instead of conventional channel capacity, we propose to use new performance indicators such as expected utility and riskiness that are commonly used to order probability density functions in axiomatic decision theory.We show the analogies between decision theoretic problems and information-theoretic problems in adaptive transmission systems. We present different single-user, multi-user, and multi-terminal communication scenarios and map them to various rationality concepts and uncertainty models in decision theory. To the author's best knowledge, adaptive transmission systems and networks have not been yet analyzed within the framework of rational decision theory.

15. SUBJECT TERMS

EOARD, Communications, Cognition, Complex Systems, Software Agents

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18, NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			JAMES LAWTON Ph. D.
UNCLAS	UNCLAS	UNCLAS	SAR	77	19b. TELEPHONE NUMBER *(Include area code)* +44 (0)1895 616187

Application of financial risk-reward theory to link and network optimization

Project's full name:	Application of financial risk-reward theory to link and network optimization
Project's short name:	REWARD
Project number:	30896
Customer reference:	European Office of Aerospace Research and Development (EOARD)
Drawn up by:	Adrian Kotelba

Version:	0.8	
Status:	[ ]	Draft
	[x]	Proposal/change proposal
	[ ]	Approved

DISTRIBUTION
James Lawton (EOARD)
Aarne Mämmelä (VTT)
Pertti Järvensivu (VTT)
Adrian Kotelba (VTT)

Change history:

Version	Date	Made by	Comments
0.1	29.01.2009	Kotelba	First draft
0.2	29.05.2009	Kotelba	Added section on St. Petersburg paradox
0.3	29.09.2009	Kotelba	Added section on rational decision-making
0.4	11.11.2009	Kotelba	Requirements and specifications for demonstrator
0.5	31.03.2010	Kotelba	Added section on adaptive power control
0.6	30.09.2010	Kotelba	Simulation results added
0.7	30.03.2011	Kotelba	Simulation results of Telser's safety-approach added
0.8	30.09.2011	Kotelba	Simulation results of performance in interference channels

Declaration of Technical Data Conformity

The Contractor, VTT Technical Research Centre of Finland, hereby declares that, to the best of its knowledge and belief, the technical data delivered herewith under Contract No. FA8655-08-C-4002 is complete, accurate, and complies with all requirements of the contract.

Date: 30 September, 2011

Name and Title of Authorized Official: Mämmelä, Aarne O., Research Professor

I certify that there were no subject inventions to declare as defined in FAR 52.227-13, during the performance of this contract.

Date: 30 September, 2011

Name and Title of Authorized Official: Mämmelä, Aarne O., Research Professor

Table of Contents

Abstract

In this work, we introduce a general framework for analysis and optimization of adaptive transmission systems in information-unstable channels. In information-unstable channels, the information density does not converge to a one-point measure and the maximum achievable transmission rate is seen as a random variable because it depends on the actual channel state. For that reason, instead of conventional channel capacity, we propose to use new performance indicators such as expected utility and riskiness that are commonly used to order probability density functions in axiomatic decision theory. We show the analogies between decision-theoretic problems and information-theoretic problems in adaptive transmission systems. We present different single-user, multi-user, and multi-terminal communication scenarios and map them to various rationality concepts and uncertainty models in decision theory. To the author's best knowledge, adaptive transmission systems and networks have not been yet analyzed within the framework of rational decision theory.

1 Introduction

The fundamental problem of telecommunication is that of efficiently encoding an information message, transmitting it over a communication channel, and decoding it at destination point with respect to a predefined fidelity criterion [35]. To process and transmit a given amount of information the communication system uses up available system resources: energy, frequency, time, and space. Furthermore, the transmission of messages over communication channels is subject to various practical constraints, for example, average or peak limitation of transmission power, available frequency bandwidth, available computing power, maximum allowable delays in encoding and decoding, and complexity of encoding-decoding scheme. Considerations of delays and complexity of encoding-decoding scheme specify the maximum length of codewords to be used in a given communication system, which in turn determines a trade-off between transmission power, transmission rate, and probability of decoding error. The communication system accomplishes efficient and reliable transmission if it uses as little system resources as possible to reproduce information at destination point within the prescribed fidelity criterion.

Ad hoc communication systems without any infrastructure have been interesting in many military applications. In commercial applications there is presently a high interest also in wireless mesh and relay networks, which are an intermediate between conventional cellular networks and ad hoc networks. Mesh and relay networks are more cost-effective and they have better performance than ad hoc networks. In addition, there are new generalized adaptive radios called cognitive radios that are based on ambient intelligence using different sensors to estimate the status of the environment and making intelligent decisions, usually to improve the performance of the network. The design of an efficient and reliable wireless communication network, which additionally operates under stringent delay constraints, poses significant engineering challenges. The main difficulties are the result of an underlying wireless radio channel which can be characterized by path loss, random composite shadowing and multipath fading, and interference. The wireless channel makes the design of wireless communication systems extremely difficult because the aforementioned channel impediments change over time in an unpredictable way. Adaptive transmission is currently considered a very promising method to cope with time-varying effects and improve transmission reliability. In adaptive transmission systems, transmitter adjusts the transmission parameters, for example transmission power, transmission rate, coding scheme, or any combination of those, to the actual state of communication channel in a way that the fidelity criterion is eventually satisfied. The adaptive control rule that governs the selection of transmission parameters with respect to various transmission constraints is usually called adaptive transmission strategy.

Information theory, by assigning a numerical measure to the information content of messages and relating various system resources through the expression of transmission rate, provides an excellent starting point for discovering methods to design efficient and reliable communication systems as well as optimal adaptive transmission strategies. However, variability of random fading process during transmission of a finite-length codeword leads to different information-theoretic models of the wireless communication channel and consequently different notions of the reliable transmission rate [5, 24]. For example, if the stochastic process which represents random fading is stationary and ergodic and the length of the codeword is sufficiently large as to reveal ergodic properties of the channel, the channel is information-stable. Consequently, the normalized information density converges in distribution to a constant which equals the normalized mutual information [38]. The supremum of the normalized mutual information is the Shannon capacity, that is, the maximum transmission rate for which error-free transmission is still possible. However, wireless communication channels are usually not information-stable channels because either the random fading process is not stationary or ergodic or the length of codeword is not sufficiently large as to reveal ergodic properties of the channel. As a consequence, the normalized information density does not converge in distribution to a single point and the Shannon capacity equals the infimum of the support set of the distribution of normalized information density [38]. In a multi-user environment

the channel is no longer characterized by a single parameter but by a set of parameters being the achievable transmission rates of all users. The set of achievable transmission rates is called capacity region. Depending on the transmission scenario certain capacity notions arise: multiple-access channel in uplink transmission, broadcast channel in downlink transmission, and interference channel in a simultaneous multi-terminal transmission.

Strict adherence to classical Shannon-sense capacity may lead to overpessimistic conclusions. In many practical systems, transmission of useful information does not require error-free communication because these systems can easily incorporate and cope with errors, outages, delays, high level distortions coming not in a stationary fashion, and the like. These practical considerations give rise to other information-theoretic notions of channel capacity such as distribution of capacity and zero-outage capacity where the instantaneous rate of reliable transmission is considered a random variable which depends on the actual realization of the fading process [5]. The probabilistic description of the rate of reliable transmission that can be supported by the communication channel represents an uncertainty on what its value could be during transmission. In adaptive systems, the distribution of reliable transmission rates depends not only on the statistical properties of the channel but also on the selection of adaptive transmission strategy. Consequently, a fair comparison of various transmission strategies as well as identification of the best strategy is problematic. It is because the subject of comparison is no longer a set of real numbers, as in the case of Shannon capacity, but the space of probability distributions. Numerous studies on the adaptive transmission in information-unstable channels, for example block-fading channels, have been conducted using expected value of mutual information or information outage probability as relevant performance indicators [5]. The information outage probability is defined as the probability that the instantaneous mutual information of the channel is below the transmitted code rate [5]. However, it is unknown whether these performance indicators, although intuitively pleasing, are indeed optimal ones because they are usually introduced without rigorous justification. The question we address in this study is how to compare different adaptive transmission strategies which can be used for transmission in information-unstable channels. In particular, what are appropriate performance indicators which objectively and meaningfully describe the performance of adaptive systems in information-unstable channels?

2 Methods, Assumptions, and Procedures

The purpose of this study is to establish a unified framework for analysis and optimization of adaptive transmission systems in information-unstable channels; a framework which takes into account an inherent uncertainty about the channel state during transmission and, consequently, the statistical nature of instantaneous rate of reliable transmission. The determination of the best choice among several alternatives, each of which leads to certain or uncertain consequences, is the subject of theory of rational decision-making. Therefore, we fit various information-theoretic models of wireless communication systems to a classification of decision-making problems in decision theory. Furthermore, we apply methods and mathematical tools of rational decision theory to construct the relevant performance indicators and order the distributions of reliable transmission rate together with the corresponding adaptive transmission strategies. Construction of the relevant performance indicators inevitably leads to a definition of some measure associated with the probability distribution, axiomatization of measurement procedure, and derivation of representation and uniqueness theorems which define homomorphisms between space of probability distributions and the set of real numbers. Consequently, with the help of representation theorem, probability distributions can be ordered in a way similar to real numbers. The set of axioms to be used in measurement provides a logically sound justification for selection of a given measure as the performance indicator and ensures that determination of the best adaptive transmission strategy is done rationally.

We define and solve the problem of finding appropriate performance indicators in information-

unstable channels within the framework of theory of rational decision-making. We focus exclusively on normative theories of rational decision-making which describe how agent should behave in order to be rational. Descriptive theories, on the other hand, analyze how human beings actually make decisions which sometimes could be considered irrational. Normative theories of rational decision-making can be roughly divided into three categories depending on the number of rational agents present in a considered scenario. We study various models of individual rationality (single-agent behavior), strategic rationality (small-group behavior), and competitive rationality (large-group behavior). Furthermore, we review and analyze different models of rational decision-making. In particular, we present the models of rational decision-making under certainty, risk, uncertainty, and ignorance [11, 22]. The classification of these models is based on the presumed level of knowledge about distribution of states of nature. For example, under complete probabilistic knowledge about distribution of states of nature, the decision is made under risk. Similarly, decision-making under uncertainty and ignorance is performed under assumptions of partial probabilistic knowledge and complete lack of probabilistic knowledge about distribution of states of nature, respectively. Decision-making under certainty is the most trivial decision-making problem where each action is known to lead invariably to a specific consequence. Consequently, a classical Shannon's theory can be used.

The models of individual rationality include ordinal utility (for choice under conditions of certainty), expected utility (for choice under conditions of risk, where probabilities for events are objectively determined), subjective expected utility and state-preference theory (for choice under conditions of uncertainty, where probabilities for events are subjectively determined or perhaps undetermined), and several kinds of nonexpected utility theory (in which total utility is not representable as a sum of utilities associated with disjoint events). Models of strategic rationality include a wide spectrum of different equilibrium concepts of game theory. Finally, models of competitive rationality include the general equilibrium model and particular forms of it that are used in capital asset pricing theory in finance. Figure 1 shows a rough taxonomy of these theories.

To order different probability distributions objectively, we define suitable fundamental measurement procedures [21]. For every possible class of decision-making problems, we will present a set of measurement axioms and propose corresponding representation and uniqueness theorems. Representation theorems define homomorphism between space of probability distributions and the set of real numbers [21, p. 12]. Uniqueness theorems, on the other hand, set forth permissible transformations that also yield homomorphisms into the same set of real numbers [21, p. 12]. Consequently, the probability distributions can be compared as easily as real numbers. In particular, we use two complementary measurement theories which are widely used in economics: risk theory [31] and utility theory [8, 30].

In utility theory, the order of probability distributions is established by comparing the expected utility associated with respective probability distributions. Utility functions are typically either linear or concave continuously differentiable functions [11, 22]. The use of linear and concave utility functions in telecommunication scenarios can be justified using results from rate-distortion theory. Furthermore, there are simplified decision-making models, for example, mean-risk [3], stochastic dominance [25], and safety-first models [16, 32], that provide orderings of probability distributions compatible with those obtained by comparing expected utilities. We emphasize mean-risk models because they are commonly used in practical applications of expected utility theory, e.g. portfolio theory [27].

The notion of risk is applied in economics as a property of uncertain options, or lotteries, which affects decision making. In risk theory, the order of probability distributions is established by comparing the riskiness associated with respective probability distributions. In this case, the relevant performance indicator could be any of the coherent risk measures, that is, risk measures which satisfy axioms of risk measurement [2].

In decision theory, and especially in mean-risk models, the term "efficiency" receives a new meaning. Namely, efficiency refers to the optimal trade-off between mean performance and risk associated with a

given mean performance. The optimal combinations of mean performance and risk are called efficient combinations [27, pp. 22–26]. The efficient frontier is the fundamental limit of the mean-risk performance because no other scheme can be constructed that achieves the performance above the efficient frontier. We study how to construct adaptive transmission schemes that achieve the optimal trade-off between average transmission rate and the risk associated with that particular transmission rate. In other words, we study how to determine efficient frontier of mean-risk transmission rate and how to design an adaptive system that achieves mean-risk performance on the efficient frontier. The mean-risk performance of various state-of-the-art adaptive transmission schemes will be studied as well. Our initial results are summarized in [19, 20]. The results in [19, 20] were derived for modulation schemes with Gaussian inputs, that is, modulation with infinite granularity. We plan to take a more practical approach, where we assume that transmitter uses a finite QAM modulation schemes.

One of the fundamental assumptions in game theory and asset pricing theory is rational behavior of agents. Therefore, our initial focus in this study is determination of rational behavior in all possible single-user scenarios. We start from idealized situation of choice under risk to find the fundamental performance limits and then turn to more practical situation when agents act under uncertainty. Next, we extend the results to include multi-user and multi-terminal scenarios. We will use game-theoretic approach to predict the performance of multi-user communication and various capital asset pricing models (CAPM) to assess the performance of possibly large multi-terminal networks. We plan to implement and verify the performance of different adaptive strategies developed for single-user scenarios using numerous game-theoretic and capital asset pricing models in our software demonstrator.

3 Channel Model

We consider a discrete-time linear vector channel with t inputs and r outputs. Let $\boldsymbol{y} \in \mathbb{C}^t$ denote a vector of complex input symbols and $\boldsymbol{z} \in \mathbb{C}^r$ be a vector of complex output symbols. The output $\boldsymbol{z}$ and input $\boldsymbol{y}$ symbols are related by the matrix equation

$$\boldsymbol{z}_k = \boldsymbol{H}_k \boldsymbol{y}_k + \boldsymbol{n}_k, \quad k = 1, \ldots, K \tag{1}$$

where k denotes a discrete-time index, K denotes the length of the codewrod, and $\boldsymbol{n} \in \mathbb{C}^r$ is a vector of complex white Gaussian noise samples with zero mean and variance σ_n^2. Complex channel coefficients h_{ij} that describe radio propagation conditions between ith input and jth output are assembled into the $r \times t$ channel matrix $\boldsymbol{H} = [h_{ij}]_{i,j=1}^{r,t}$.

The channel model given by (1) is very general and applies to several situation arising in wireless communications, depending on the definition of the channel matrix $\boldsymbol{H}$ [18]. For example,

1. Frequency-nonselective fading channel — $\boldsymbol{H}_k$ is a random scalar
2. Frequency-selective time-invariant channel — $\boldsymbol{H}_k$ is a fixed Toeplitz matrix
3. Frequency-selective block-fading channel — $\boldsymbol{H}_k$ is a random Toeplitz matrix
4. Frequency-nonselective fading channel with multiple antennas — $\boldsymbol{H}_k$ is a random matrix
5. Frequency-selective block-fading channel with multiple antennas — $\boldsymbol{H}_k$ is a random block-Toeplitz matrix

A frequency-selective fading can be modeled by (1) provided that the channel memory is finite. The channel matrix $\boldsymbol{H}_k$ is then a convolution matrix with Toeplitz or block-Toeplitz structure.

In practice, wireless systems are designed under very coarse assumptions on the channel model. For example, only the maximum values of Doppler frequency spread $F_{\max}$ and delay spread $\tau_{\max}$ are assumed to be known. Furthermore, the distribution of the entries of $\boldsymbol{H}_k$ is assumed to belong to some class of possible statistics, for example, the entries of $\boldsymbol{H}_k$ are jointly Gaussian. Another common assumption is that the entries of the channel matrix $\boldsymbol{H}$ evolve in time as wide-sense stationary complex processes,

and that coefficients at different delays in frequency-selective channels are uncorrelated. This model is commonly referred as to wide-sense stationary uncorrelated scattering (WSSUS) fading channel. The number of complex dimensions for which the fading channel can be considered as locally time-invariant is approximately given by the product of the channel coherence bandwidth W_c and the channel coherence time T_c. These are inversely proportional to the channel delay spread $\tau_{\max}$ and the channel Doppler bandwidth $F_{\max}$, respectively. In modern wireless/mobile communications the product $W_c T_c$ ranges between 100 and several thousands. We refer to such a fading channel as block-fading channel. The block-fading approximation is accurate at least in the range of spectral efficiencies of practical interest, typically ranging between 1/2 and 6 bit/s/Hz [23].

The block-fading regime enables channel estimation schemes based on training sequences. For each block of symbols over which the channel is locally time-invariant, the transmitter sends a pilot signal that enables the receiver to produce an estimate of the channel response. Explicit training is used by virtually all modern wireless/mobile communication systems, and it is ubiquitous in the next generation of wireless standards. The main factor motivating explicit training is simplicity: indeed, schemes that avoid explicit training are generally either suboptimal (e.g., schemes based on noncoherent/differential symbol by symbol detection) or require very high complexity, both in terms of receiver algorithms and in terms of coding and modulation design. Hence, the assumption of perfect channel state information at the receiver is well justified.

Most of wireless/mobile communication systems carry communications in two directions: from the base station to the user terminals and from the user terminals to the base station. This bidirectional communication can be used to provide channel state information also to the transmitter, either implicitly ("open-loop") or explicitly ("closed-loop"). Open loop schemes are based on measuring the channel parameters in one direction based on training symbols in the reverse direction. If transmission directions are multiplexed in time, and provided that the terminal radio frequency (RF) up and down conversion chains are calibrated so that their gain are known, the radio channel reciprocity can be exploited to achieve very accurate amplitude information on the channel coefficients at the transmitter. When reciprocity cannot be exploited, open-loop schemes can only provide coarse information such as average received signal strength. Then, closed-loop schemes must be used, where the channel at the receiver end is estimated, and some information is sent back via a feedback channel. Generally speaking, we may assume that the channel state information at the transmitter is represented by the signal $\{\boldsymbol{G}_k : k = 1, \ldots, K\}$ where $\{\boldsymbol{H}_k, \boldsymbol{G}_k\}$ are jointly distributed according to some known probability law and the $\{\boldsymbol{G}_k\}$ signal is revealed to the transmitter in a causal way, i.e., when producing the input $\boldsymbol{y}_k$ the transmitter knows the sequence $\{\boldsymbol{G}_1, \ldots, \boldsymbol{G}_k\}$.

Let us assume that both the transmitter and the receiver include additional linear transformations to diagonalize the channel matrix $\boldsymbol{H}_k$. In particular, let $\boldsymbol{H}_k = \boldsymbol{U}_k \boldsymbol{D}_k \boldsymbol{V}_k^*$ where $\boldsymbol{U}_k$ is an $r \times r$ unitary matrix, the matrix $\boldsymbol{D}_k = [d_{ij}]$ is $r \times t$ diagonal matrix with nonnegative real numbers on the diagonal, and $\boldsymbol{V}^*$ denotes the conjugate transpose of $\boldsymbol{V}_k$, a $t \times t$ unitary matrix. For example, in OFDM systems the matrices $\boldsymbol{V}_k$ and $\boldsymbol{U}_k$ are Fourier matrices which diagonalize a circulant or block-circulant matrix $\boldsymbol{H}_k$ [18]. In multi-antenna systems, theses unitary matrices are beamforming matrices that send the signal through the best channel eigenmodes [18]. We assume, therefore, that the channel input symbols $\boldsymbol{y}$ are the result of the linear transformation

$$\boldsymbol{y}_k = \boldsymbol{V}_k \boldsymbol{Q}_k \boldsymbol{x}_k \tag{2}$$

where $\boldsymbol{Q}_k$ is a $t \times t$ power control matrix and $\boldsymbol{x} \in \mathbb{C}^t$ is a vector of complex channel encoder symbols. Similarly, the receiver applies a linear transformation

$$\tilde{\boldsymbol{z}}_k = \boldsymbol{U}_k^* \boldsymbol{z}_k \tag{3}$$

to the channel output symbols $\boldsymbol{z}_k$. By combining (1), (2), and (3), we obtain

$$\tilde{\boldsymbol{z}}_k = \boldsymbol{D}_k \tilde{\boldsymbol{x}}_k + \tilde{\boldsymbol{n}}_k \tag{4}$$

where $\tilde{\boldsymbol{x}}_k = \boldsymbol{Q}_k \boldsymbol{x}_k$ and $\tilde{\boldsymbol{n}}_k = \boldsymbol{U}_k^* \boldsymbol{n}_k$. From (4), we conclude that with additional unitary transformation in the transmitter and the receiver, the linear vector channel is transformed into a set of $m = \min(r,t)$ parallel Gaussian subchannels.

Let $\lambda_i = d_{ii}^2$, where $1 \leq i \leq m$, denote the instantaneous energy gain of the ith subchannel and $q_i = |\tilde{x}_i|^2$ be the corresponding instantaneous transmitted energy. In most of the theoretical fading models, including Rayleigh, Nakagami, and log-normal fading models, the subchannel energy gains λ_i are unbounded [5]. Thus, we assume that λ_i are random variables with an arbitrary joint probability density function $p(\lambda_1, \lambda_2, \ldots, \lambda_m)$ defined on the nonnegative orthant of m-dimensional Euclidean space $\mathbb{R}^m$. Furthermore, let both the representative energy gain of the channel [18]

$$G_0 = \frac{1}{t} E\left[\sum\nolimits_{i=1}^{m} \lambda_i\right] = \frac{1}{t} \sum\nolimits_{i=1}^{m} \int_0^\infty \lambda_i p(\lambda_i)\, d\lambda_i \tag{5}$$

and the average transmitted energy

$$E_{\text{av}} = E\left[\sum\nolimits_{i=1}^{m} q_i\right] = \sum\nolimits_{i=1}^{m} \int_0^\infty q_i p(q_i)\, dq_i \tag{6}$$

be finite. The marginal probability density functions of λ_i and q_i are denoted by $p(\lambda_i)$ and $p(q_i)$ in (5) and (6), respectively.

The instantaneous link spectral efficiency of the transmission over a multi-antenna fading channel is a random variable which can be expressed as the function of $m = \min(t,r)$ positive eigenvalues of the Wishart matrix $\boldsymbol{W}_k = \boldsymbol{H}_k \boldsymbol{H}_k^*$ [36]. The value of the ith eigenvalue λ_i represents the energy gain of the ith subchannel associated with the ith eigenvalue. We denote the transmission power allocated to the kth channel by q_i. Then, the instantaneous link spectral efficiency η_r is [36]

$$\eta_r(\boldsymbol{\lambda}, \boldsymbol{q}) = \sum\nolimits_{i=1}^{m} \log_2\left(1 + q_i \lambda_i / \sigma_n^2\right) \tag{7}$$

where we have used vector notation $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$ and $\boldsymbol{q} = (q_1, \ldots, q_m)$ for simplicity. The average power constraint, on the other hand, becomes [36]

$$\mu_q(\boldsymbol{\lambda}, \boldsymbol{q}) = E\left[\sum\nolimits_{i=1}^{m} q_i\right] \leq P_{\text{av}}. \tag{8}$$

Let the region $Q \subset \mathbb{R}_+^m$ be defined by

$$Q = \left\{\boldsymbol{\lambda} \in \mathbb{R}_+^m : \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m\right\}. \tag{9}$$

We can assume, without loss of generality, that $\boldsymbol{\lambda} \in Q$, i.e., the subchannel gains $\{\lambda_1, \ldots, \lambda_m\}$ in (7) and (8) are sorted in the nonincreasing order. The assumption $\boldsymbol{\lambda} \in Q$ does not restrict our analysis in any way because the solution, which is obtained for $\boldsymbol{\lambda} \in Q$, can be easily extended to include any $\boldsymbol{\lambda} \notin Q$ [6]. In particular, one can always sort the components of $\boldsymbol{\lambda} \to \pi(\boldsymbol{\lambda})$ in the nonincreasing order, find the optimal power allocation vector $\boldsymbol{q}$ for the sorted vector $\pi(\boldsymbol{\lambda})$, and apply the inverse permutation to the vector $\boldsymbol{q}$, i.e., $\boldsymbol{q} \to \pi^{-1}(\boldsymbol{q})$. Consequently, one can obtain the optimal solution for any $\boldsymbol{\lambda} \in \mathbb{R}_+^m$ by considering only a restricted set of $\boldsymbol{\lambda} \in Q$.

The evaluation of (8) requires averaging over the joint distribution of ordered positive eigenvalues of $\boldsymbol{W}$ which is given by [36]

$$p(\boldsymbol{\lambda}) = c^{-1} \prod\nolimits_{i=1}^{m} \lambda_i^{u-m} e^{-\lambda_i} \prod\nolimits_{j<i} (\lambda_i - \lambda_j)^2 \tag{10}$$

where $u = \max(t, r)$ and c is a normalization constant

$$c = \prod_{i=1}^{m} (u - i)! \, (m - i)!. \tag{11}$$

It is shown in [6] that a Gaussian-like, fixed-rate code achieves optimal link spectral efficiency performance where a state-dependent amplifier controls the power according to the optimal power control strategy. Therefore, we exclusively focus on adaptive power control schemes.

4 Results and Discussion

4.1 Connections Between Decision Theory and Information Theory

The performance of adaptive transmission schemes in information-unstable channels is commonly described by information outage probability or expected value of reliable transmission rate. However, within the framework of decision theory the use of both performance criteria is at least questionable [2,4]. From descriptive point of view, expected value is a relevant performance measure if the experiment is repeated infinitely many times under the same conditions. This is rather uncommon situation in practice because complexity considerations set upper limit on the length of the codeword. Furthermore, one can encounter paradoxes similar to the St. Petersburg paradox, which demonstrates that under some conditions the use of expected value as the performance indicator leads to irrational choices [4]. The use of information outage probability, although intuitively pleasing, has also many disadvantages. For example, outage probability only gives the probability that an undesirable event occurs but gives no clues as to how severe it could be. Many practical communication systems can easily cope with outages and intermittent connectivity. Furthermore, adaptive transmission systems can reduce the actual transmission rate during outage rather than suspend it completely. Thus, the use of information outage probability as the relevant performance indicator may lead to overpessimistic conclusions. Furthermore, commonly used performance indicators, that is, information outage probability and expected value of transmission rate, are usually introduced without rigorous justification. The use of information outage probability as the relevant performance indicator can be justified using the concept of ε-achievable capacity [38]. However, a serious limitation of this approach is the assumption of constant code rate. We will show that outage probability is not in general coherent measure of risk and, as such, it should not be used as the relevant performance indicator. Furthermore, we will demonstrate limitations of risk theory when applied to adaptive transmission in information-unstable channels. In particular, we will show that some decision-making problems cannot be resolved within the framework of risk theory.

Numerous telecommunication scenarios can be easily fitted into a classification of decision theory. For example, single adaptive links, as shown in Figure 2, can be considered within framework of individual rationality. Similarly, traditional multi-user communication, as shown in Figure 3 for multiple-access channel, corresponds to strategic rationality. Finally, the performance of large multi-terminal networks can be analyzed using concepts of competitive rationality. We say that the transmitter or receiver has channel state information (CSI) and channel distribution information (CDI) available, when they have information about instantaneous state of the channel and all information about distribution of of channel states, respectively. Taking into account different assumptions about availability of channel distribution information and channel state information to the transmitter and receiver, we classify various information-theoretic models of single-user, multi-user, and general multi-terminal wireless scenarios within aforementioned models of decision-making problems. For example, the rational-decision making under risk corresponds to idealized situation where the receiver and transmitter have perfect channel state information and channel distribution information available. Rational-decision making under uncertainty or ignorance describes a more realistic scenario where cognitive radio learns the statistical properties of

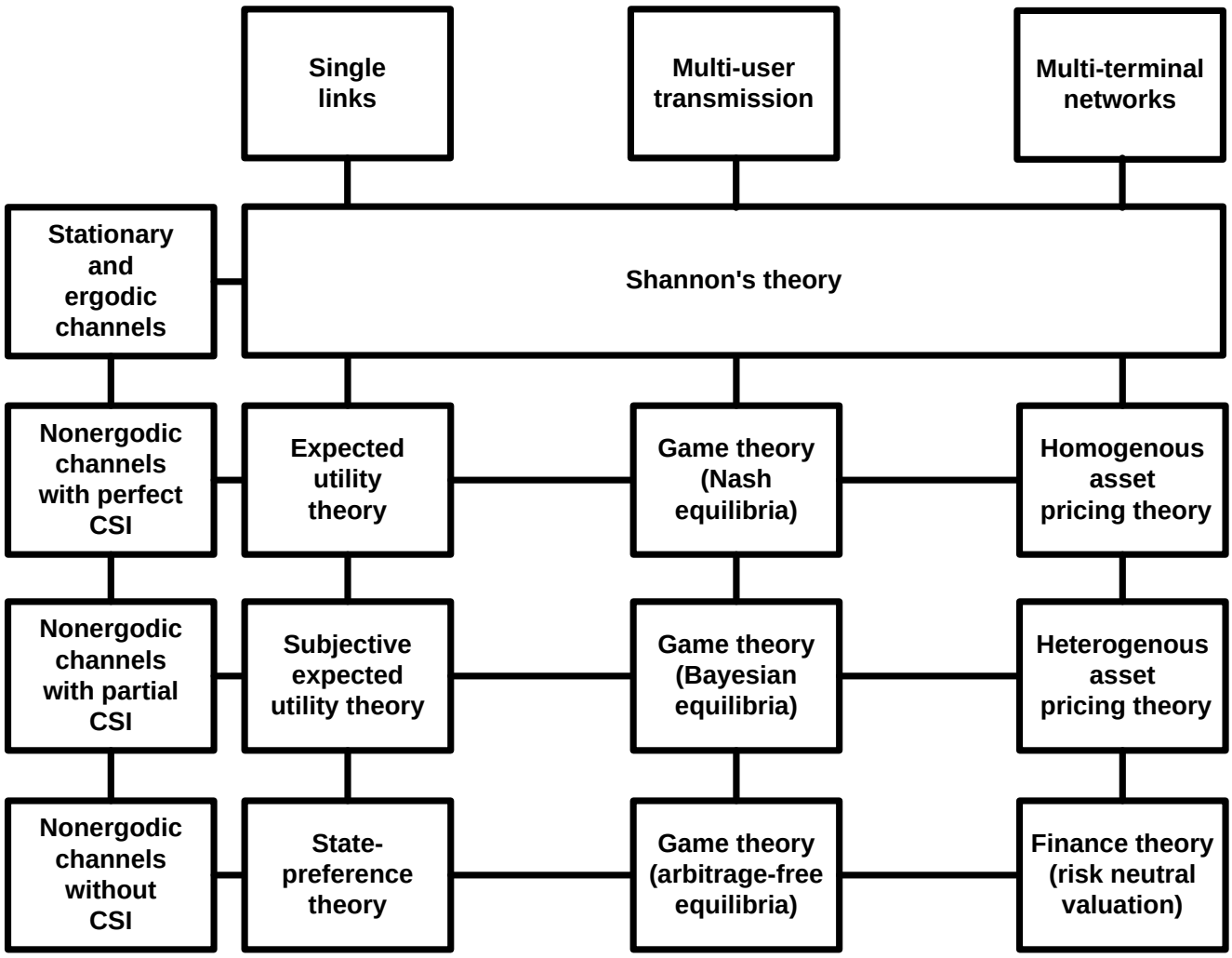

Figure 1: Information-theoretic models of communication channels and uncertainty concepts.

the environment in which it operates. In other words, transmitter and receiver have only partial channel state information available. Figure 4 shows connections between certain problems in decision theory and information-theoretic models of communication.

4.2 Rational Decision-Making Under Risk

4.2.1 Expected Utility Hypothesis

Let $\mathcal{C}$ be a set of all possible outcomes of an experiment, that is, a sample space. Let $\mathcal{P}$ be a set of all probability measures on $\mathcal{C}$. The elements of $\mathcal{P}$ are commonly referred to as objects of preference. Axioms of preference specify the behavior of a binary relation "is preferred to" $\succ$ on the set $\mathcal{P}$ of objects of preference. Before we proceed to state the axioms of preference and representation theorem, we need a few definitions.

Definition 1. *A sigma-algebra $\mathcal{F}$ for $\mathcal{C}$ is a collection of subsets of $\mathcal{C}$ that contains $\mathcal{C}$ ($\mathcal{C} \in \mathcal{F}$), is closed under complementation ($\mathcal{A} \in \mathcal{F} \Rightarrow \mathcal{C} \backslash \mathcal{A} \in \mathcal{F}$), and is closed under countable unions of its elements ($\mathcal{A}_k \in \mathcal{F}$ for $k = 1, 2, \ldots \Rightarrow \bigcup_{k=1}^{\infty} \mathcal{A}_k \in \mathcal{F}$).*

Definition 2. *Let $\mathcal{M}$ be the set of all open sets in $\mathcal{C}$. The sigma-algebra generated by $\mathcal{M}$, that is, the smallest sigma-algebra that includes $\mathcal{M}$, is called the Borel sigma-algebra for $\mathcal{C}$ and its elements $\mathcal{A}_k$*

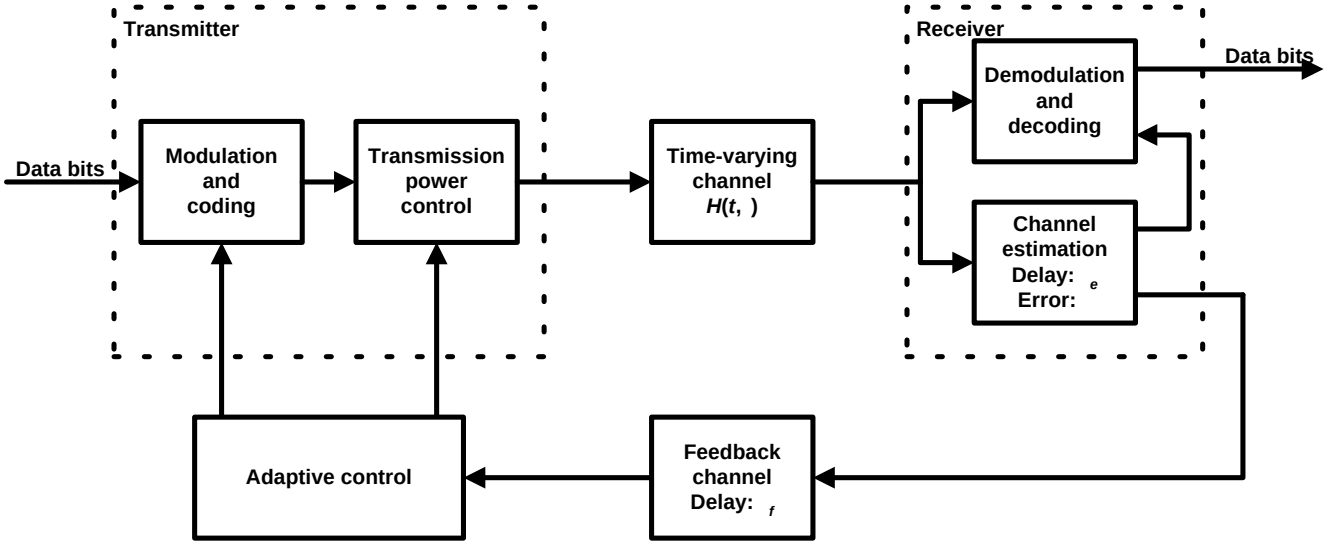

Figure 2: Model of a single-user communication.

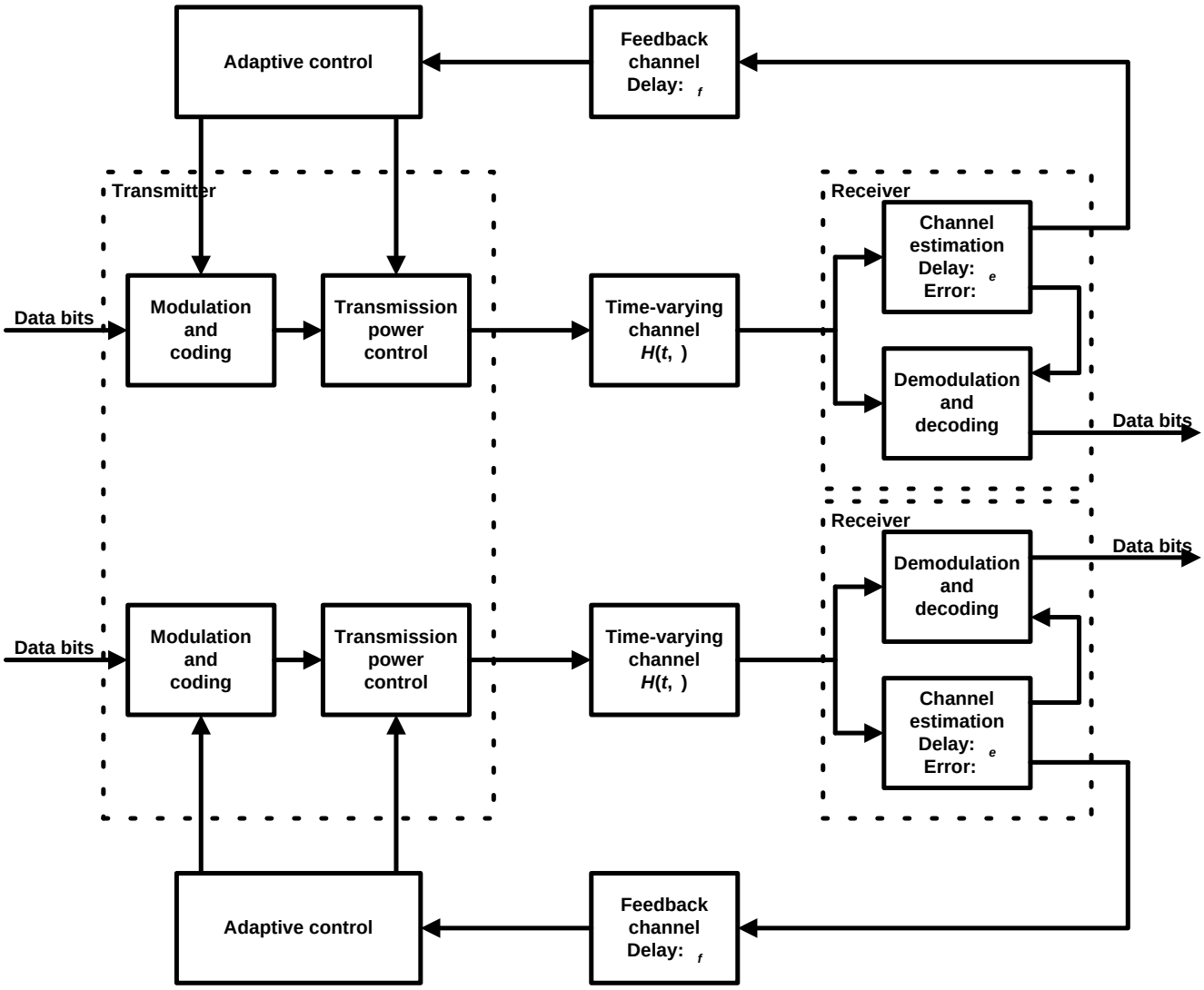

Figure 3: Model of a multi-user communication.

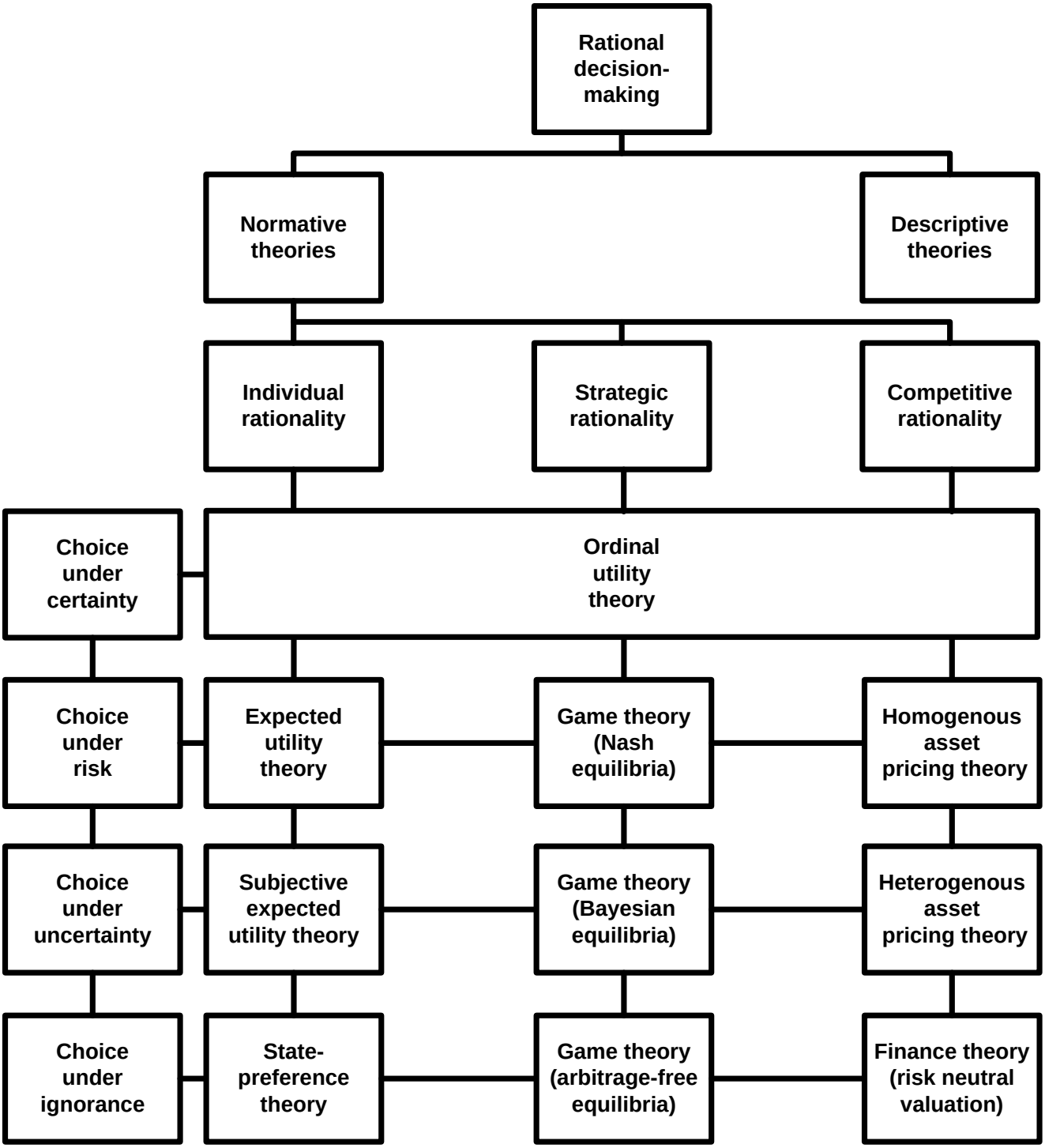

Figure 4: Rationality theories and associated uncertainty concepts.

are called Borel sets.

Definition 3. *A countably additive probability measure on an algebra $\mathcal{F}$ is a nonnegative real-valued function $P : \mathcal{F} \to [0,1]$ with $P(\varnothing) = 0$, $P(\mathcal{C}) = 1$, and*

$$P\left(\bigcup_{k=1}^{\infty} \mathcal{A}_k\right) = \sum_{k=1}^{\infty} P(\mathcal{A}_k) \tag{12}$$

whenever $\{\mathcal{A}_1, \mathcal{A}_2, \dots\}$ is a countable collection of pairwise disjoint elements in $\mathcal{F}$.

Definition 4. *A set $\mathcal{P}$ of probability measures on $\mathcal{F}$ is closed under finite convex combinations if and only if $\alpha P + (1-\alpha) Q \in \mathcal{P}$ whenever $P,\ Q \in \mathcal{P}$ and $0 < \alpha < 1$.*

In order to define utilities of consequences in an unambiguous way, we assume that each singleton subset $\{x\}$ from $\mathcal{C}$ is in $\mathcal{F}$ and the corresponding one-point measure $P(\{x\})$ is in $\mathcal{P}$. That is, if $x \in \mathcal{C}$ and $P(\{x\}) = 1$ then $P(\{x\}) \in \mathcal{P}$. Consequently, $x \succ y$ indicates that consequence x is preferred to consequence y when $P(\{x\}) = Q(\{y\}) = 1$. Similarly, $x \preccurlyeq y$ means that consequence x is not preferred to consequence y when $P(\{x\}) = Q(\{y\}) = 1$.

Definition 5. *A subset $\mathcal{Z}$ of $\mathcal{C}$ is a preference interval if $z \in \mathcal{Z}$ whenever $x, y \in \mathcal{Z}$, $x \preccurlyeq z$, and $z \preccurlyeq y$.*

$$\mathcal{Z} = \{z \in \mathcal{C} : x \preccurlyeq z \preccurlyeq y\} \tag{13}$$

Finally, to ensure that expected utilities are well defined, we require that $\mathcal{P}$ is also closed under conditional measures on preference intervals. When all preference intervals are in $\mathcal{F}$, the set $\mathcal{P}$ is closed under conditional measures on preference intervals if for all $\mathcal{Y} \in \mathcal{F}$ the conditional measure of p given $\mathcal{Z}$, $P_{\mathcal{Z}}(\mathcal{Y}) = P(\mathcal{Y} \cap \mathcal{Z}) / P(\mathcal{Z})$, is in $\mathcal{P}$ whenever $P \in \mathcal{P}$, $\mathcal{Z}$ is a preference interval from $\mathcal{C}$, and $P(\mathcal{Z}) > 0$.

We can now state a basic requirements that must be satisfied by our preference relation $\succ$ on the set $\mathcal{P}$ of probability measures.

Axiom 1 (Measurability)**.** *The set $\mathcal{F}$ is a Borel sigma-algebra of subsets of $\mathcal{C}$ which contains the singleton subset $\{x\}$ for each $x \in \mathcal{C}$ and every preference interval $\mathcal{Z}$ from $\mathcal{C}$. The set $\mathcal{P}$ is a set of countably additive probability measures defined on $\mathcal{F}$ that contains every one-point measure, is closed under finite convex combinations, and is closed under conditional measures on preference intervals.*

Axiom 2 (Ordering)**.** *The binary relation $\succ$ on $\mathcal{P}$ is asymmetric $P \succ Q \Rightarrow \neg(Q \succ P)$ and negatively transitive $P \succ Q \Rightarrow (R \succ Q \vee P \succ R)$ for all $P,\ Q,\ R \in \mathcal{P}$.*

Axiom 3 (Independence)**.** *If $P,\ Q,\ R \in \mathcal{P}$, $P \succ Q$, and $0 < \alpha < 1$, then $\alpha P + (1-\alpha) R \succ \alpha Q + (1-\alpha) R$.*

Axiom 4 (Continuity)**.** *If $P,\ Q,\ R \in \mathcal{P}$, $P \succ Q$ and $Q \succ R$, then $\alpha P + (1-\alpha) R \succ Q$ and $Q \succ \beta P + (1-\beta) R$ for some $0 < \alpha < 1$ and $0 < \beta < 1$.*

Axiom 5 (Dominance)**.** *If $P,\ Q \in \mathcal{P}$, $\mathcal{A} \in \mathcal{F}$, $P(\mathcal{A}) = 1$ and $y \in \mathcal{C}$, then $Q(\{y\}) \preccurlyeq P$ if $y \preccurlyeq x$ for all $x \in \mathcal{A}$, and $P \preccurlyeq Q(\{y\})$ if $x \preccurlyeq y$ for all $x \in \mathcal{A}$.*

Axiom 6 (Boundedness)**.** *If $P \in \mathcal{P}$ and Q is a simple probability measure, then $Q \preccurlyeq P_{\mathcal{Z}}$ for some $y \in \mathcal{C}$ and $\mathcal{Z} = \{z \in \mathcal{C} : z \preccurlyeq y\}$ if $P \succ Q$, and $P_{\mathcal{Z}} \preccurlyeq Q$ for some $y \in \mathcal{C}$ and $\mathcal{Z} = \{z \in \mathcal{C} : y \preccurlyeq z\}$ if $Q \succ P$.*

Axiom 1 is purely technical; it ensures the existence of measurable utility structure. Axiom 2, on the other hand, defines an ordering relation on the set of probability measures. Axiom 3 is an independence or linearity assumption which preserves preference under similar combinations. Axiom 4 is an Archimedean-type assumption which facilitates the derivation of continuous real-valued utilities. Axiom 5 is a dominance principle which says, for example, that if all random consequences are inferior to some certain consequence, then the certain consequence is preferred. Finally, Axiom 5 is again a technical axiom which ensures the existence of finite expected utility with unbounded utility functions.

Theorem 1 (Fishburn 1975)**.** *Suppose Axiom 1 holds. Then there exists a real-valued function u on $\mathcal{C}$ for which*

$$U(P) = \int_{\mathcal{C}} u(x)\, dP(x) \tag{14}$$

is well defined and finite for all $P \in \mathcal{P}$ and such that

$$P \succ Q \Longleftrightarrow U(P) > U(Q) \tag{15}$$

holds for all $P,\ Q \in \mathcal{P}$, if and only if Axioms 2 through 6 hold.

Moreover, $u(x)$ in (14) *is unique up to a positive linear transformation: that is, if $u(x)$ satisfies* (14) *then a real-valued function $v(x)$ on $\mathcal{C}$ satisfies* (15) *for all $P, Q \in \mathcal{P}$, if and only if there are numbers $a > 0$ and b such that*

$$v(x) = au(x) + b \quad \text{for all } \ x \in \mathcal{C}. \tag{16}$$

Proof. The proof of the theorem can be found in [9]. □

In words, Theorem 1 means that one probability distribution P is preferred to another probability distribution Q if and only if the expected utility associated with P is higher than the expected utility associated with Q. Consequently, one can order different probability measures by comparing the expected utilities.

4.2.2 Risk Theory

The risk theory is formulated in terms of a set $\mathcal{P} = \{P, Q, R, \ldots\}$ of probability distributions on the real line. Let $\circ$ denote convolution of probability distributions. The key concept of the risk theory is a binary relation of comparative risk denoted $\succcurlyeq$. Thus, $P \succcurlyeq Q$ states that P is at least as risky as Q, while $P \sim Q$ states that P and Q are equally risky. Let $P \succ Q$ whenever $P \succcurlyeq Q$ and not $Q \succcurlyeq P$. Furthermore, let β denote the degenerate distribution where the value zero is obtained with probability one and let nP denote the distribution obtained by multiplying all values of P by some real number n. For continuous distributions the density function of nP has to be properly normalized.

The axioms of the risk theory are incorporated into the definition of a risk system and a regular risk system. Namely,

Axiom 7 (Weak Ordering)**.** *The binary relation $\succcurlyeq$ is connected and transitive.*

Axiom 8 (Cancellation)**.** *$P \succcurlyeq Q$ if and only if $P \circ R \succcurlyeq Q \circ R$.*

Axiom 9 (Solvability)**.** *If $P \succcurlyeq \beta$ for all P in $\mathcal{P}$, then for any $P \succ Q$ there exists some R in $\mathcal{P}$ such that $P \sim Q \circ R$. If, on the other hand, $\beta \succ P$ for some P in $\mathcal{P}$, then for any P in $\mathcal{P}$ there exists some R in $\mathcal{P}$ such that $P \circ R \sim \beta$.*

Axiom 10 (Archimedean)**.** *If $P \succ Q \succ \beta$, then there exists some positive integer n such that $n \bullet Q \succcurlyeq P$; where $n \bullet Q$ is defined inductively $1 \bullet Q = Q$, $n \bullet Q = [(n-1) \bullet Q] * Q$.*

The first axiom is the usual ordering assumption which states that $\mathcal{P}$ can be weakly-ordered with respect to risk. The second axiom asserts that the risk ordering is compatible with the convolution operation, in the sense that order between any pair of distributions is preserved when a third distribution is convoluted with each of them. The solvability axioms states that when P is considered riskier than Q then the risk of P can always be matched by combining Q with some appropriate R. Finally, the Archimedean axiom is introduced to ensure that no risky option is infinitely riskier than any other one, provided that both options are riskier than β.

Definition 6. *The system $(\mathcal{P}, \circ, \succcurlyeq)$ is a risk system if it satisfies Axioms 7–10 for all P, Q, R in $\mathcal{P}$.*

Theorem 2 (Pollatsek & Tversky 1970)**.** *If $(\mathcal{P}, \circ, \succcurlyeq)$ is a risk system then there exists a real-valued function V defined on $\mathcal{P}$, such that for any P, Q in $\mathcal{P}$*

(i) $P \succcurlyeq Q$ *if and only if* $V(P) \geq V(Q)$
(ii) $V(P \circ Q) = V(P) + V(Q)$
(iii) *If W is another function satisfying (i) and (ii), then* $W(P) = \alpha V(P)$ *for some* $\alpha > 0$.

Proof. The proof of the theorem can be found in [31]. □

Definition 7. *A risk system $(\mathcal{P}, \circ, \succcurlyeq)$ is called a regular risk system if it also satisfies the following axioms for all P, Q, R in $\mathcal{P}$,*

Axiom 11 (Positivity)**.** *If S is a degenerate distribution with $s > 0$, then $P \succcurlyeq P \circ S$ for all P in $\mathcal{P}$.*

Axiom 12 (Monotonicity)**.** *For all P, Q in $\mathcal{P}$ with expected values $E[P] = E[Q] = 0$ and for any real $n > 1$, $nP \succ P$ and $P \succcurlyeq Q$ if and only if $nP \succcurlyeq nQ$.*

Axiom 13 (Continuity)**.** *If a sequence of distributions $\{P_n : n = 1, 2, \dots\}$ approaches a limiting distribution P, that is, whenever $\Pr(x \leq P_n \leq y)$ approaches $\Pr(x \leq P \leq y)$ as $n \to \infty$ for any real x, y, then $V(P_n)$ approaches $V(P)$, provided $E[P_n] = E[P]$ and $\mathrm{Var}[P_n] = \mathrm{Var}[P]$, where $\mathrm{Var}[P]$ denotes the variance of P.*

The first axiom states that addition of the positive sure-thing to an risky option cannot increase its risk. The second axiom asserts that, for distribution with zero expectation, risk increases with multiplication by any $n > 1$, and that the risk ordering is preserved upon multiplication by a positive real number. The last axiom is technical in nature. Namely, it establishes the continuity of the risk scale in the sense that if all the distributions in the sequence have the same mean and variance then $\lim V(P_n) = V(P)$ whenever $\lim P_n = P$.

Theorem 3 (Pollatsek & Tversky 1970)**.** *If $(\mathcal{P}, \circ, \succcurlyeq)$ is a regular risk system then there exists a unique $0 < \alpha \leq 1$ such that for all P, Q in $\mathcal{P}$ with finite expectations $E[P]$, $E[Q]$ and variances $\mathrm{Var}[P]$, $\mathrm{Var}[Q]$*

$$P \succcurlyeq Q \iff V(P) \geq V(Q) \tag{17}$$

where

$$V(P) = \alpha \mathrm{Var}[P] - (1 - \alpha) E[P]. \tag{18}$$

Proof. The proof of the theorem can be found in [31]. □

We can conclude that in a regular risk system, the risk ordering is generated by a linear combination of expectation and variance. In other words, the risk associated with any probability distribution can be readily computed, once a single parameter, α, is determined. Furthermore, α is attainable form a single judgment of risk-equality between two distinct distributions, and its value determines the relative contribution of the expectations and the variance to the riskiness of a given probability distribution.

4.2.3 Risk-Value Hypothesis

In the theory of finance, there is a long tradition of work in which individuals' preferences over asset portfolios are assumed to depend only on a few summary statistics. The aim is to develop statistics which describe the salient features of risky assets without making any reference to preferences, and to model the utility of a portfolio as some function of the values of those statistics. Individual-specific attitudes to risk are then modeled by the parameters of that portfolio utility functions. In the most widely used

model, assets are treated as one-dimensional prospects and for any asset, the two salient statistics are its expected return and some dispersion measure of its returns. In this section, we briefly review a few basic concepts of risk-value theory such as risk and reward measures, optimal risk-reward trade-offs, optimal as well as simplified decision models [29]. We present how these concepts can be applied in the context of adaptive transmission.

In the financial theory the reward is conventionally measured as the expected return on investment in excess of some predefined threshold [29]. In adaptive transmission, where energy "investment" creates a transmission channel that supports a certain link spectral efficiency r, the return is the actual link spectral efficiency r. The reward μ_t is therefore the expected value of the difference between the actual η_r and the target η_t link spectral efficiency, i.e., $\mu_t = E[\eta_r - \eta_t]$.

The risk has not yet received any consensus as to what constitutes its proper measure [29]. For example, risk can be viewed as the standard deviation of η_r, outage probability $\varepsilon(\eta_t)$, or lower partial moments of η_r. The standard deviation as a measure of risk has an important limitation [29]. Namely, the standard deviation measures both positive and negative deviations from the mean value. This is in contradiction to a common perception of risk as a chance of getting performance worse than desired. For that reason, lower partial moments of link spectral efficiency distribution are usually used as the risk measure. The lower partial moments measure only deviations below a predefined threshold, their use as the risk measures is fully justified within the framework of stochastic dominance and expected utility models, and they represent a large number of common engineering risk measures as special cases [29]. The nth order lower partial moment of the random variable η_r with respect to a reference point η_t is [29]

$$l_n^-(\eta_t) = \int_{-\infty}^{\eta_t} (\eta_t - \eta_r)^n \, p_{\eta_r}(\eta_r)\, d\eta_r, \quad n \geq 0 \tag{19}$$

where pdf of η_r is denoted by $p_{\eta_r}(\eta_r)$.

The outage probability $\varepsilon(\eta_t) = \Pr(\eta_r < \eta_t)$, a commonly used performance measure in communication engineering, is simply the zeroth order lower partial moment, i.e., $\varepsilon(\eta_t) = l_0^-(\eta_t)$. The major shortcoming of the outage probability as the risk measure is that it gives the probability that an undesirable event might occur but gives no clues as to how severe it might be. Another risk measure, which is commonly used in nuclear engineering, is expected shortfall $l_1^-(\eta_t)$. Expected shortfall measure incorporates both the probability and magnitude of the potential shortfall if it does occur. However, large losses which occur infrequently represent the same risk as the small and frequent losses. In general, the lower partial moments of order $n \leq 1$ lack the ability to represent user's risk aversion [29]. The risk aversion means that a rational user, when offered several transmission schemes with the same expected link spectral efficiency, prefers the scheme with the lowest risk. For that reason, in practical considerations, one uses the second order lower partial moment $l_2^-(\eta_t)$ which is known as the below-target semivariance [29]

$$s_t(t) = l_2^-(t) = E\{[\max(0, t-r)]^2\}. \tag{20}$$

We define the maximum-return scheme as a transmission scheme which achieves the highest level of return. Similarly, a transmission scheme with the lowest risk is the minimum-risk scheme. Furthermore, a transmission scheme is considered efficient if it maximizes the return for a given amount of risk or minimizes the risk for a given level of return [26, pp. 22–26].

A trade-off between reward and risk is quantitatively measured by a reward-to-semivariability ratio. The most general reward-to-semivariability ratio is Kappa ratio $\kappa_n(\eta_t)$ which is defined as the ratio of the reward to the nth root of the nth order lower partial moment, or [15]

$$\kappa_n(\eta_t) = \frac{E[\eta_r - \eta_t]}{\sqrt[n]{l_n^-(\eta_t)}}, \quad n > 0. \tag{21}$$

The optimal risk-reward scheme is a transmission scheme with the highest reward-to-semivariability ratio $\kappa_n(\eta_t)$ because it maximizes the reward per unit of risk taken [15]. Furthermore, it can be easily verified that the transmission scheme which maximizes $\kappa_n(\eta_t)$ has the highest expected value of utility for a certain class of utility functions defined in [10, eq. (11)–(12)] with a given risk aversion rate and, as such, it is considered the optimal choice under uncertainty [26, Ch. 10].

The maximum-return curve and the minimum-risk curve are curves in the risk-return plane which graphically represent the set of transmission schemes with the highest level of return for a given amount of risk and the lowest risk for a given level of return, respectively. Finally, the efficient curve is a curve in the risk-return plane which graphically represents the set of efficient transmission schemes.

4.2.4 Safety-first schemes

The origin of safety-first approaches stems from the belief that the complexity of optimal transmission schemes, and especially optimal decision algorithms, is so high that they cannot be implemented in any practical system. One relies then on a simple decision algorithm that concentrates exclusively on undesirable events. Three basic safety-first schemes have been proposed in [16, 32, 37].

In the scheme proposed by Roy in [32] one chooses the transmission scheme which minimizes the outage probability $\Pr(\eta_r < \eta_t)$, i.e.,

$$\min \Pr(\eta_r < \eta_t). \tag{22}$$

On the other hand, according to Kataoka's safety-first scheme [16], one chooses the transmission scheme that maximizes the target link spectral efficiency t such that the outage probability does not exceed a predefined value ξ, i.e.,

$$\max \{\eta_t \,|\Pr(\eta_r < \eta_t) \leq \xi\}. \tag{23}$$

The third safety-first approach is proposed in [37] by Telser. By predicting the outage probability ξ and the minimum acceptable link spectral efficiency α, one should try to maximize the expected link spectral efficiency $\mu_r = E[\eta_r]$, i.e.,

$$\max \{\mu_r \,|\Pr(\eta_r \leq \alpha) \leq \xi\}. \tag{24}$$

Roy's and Kataoka's safety-first approaches have already been used, without reference to risk-reward theory, in [6] as objective criteria for designing adaptive transmission systems.

4.3 St. Petersburg Paradoxes in Performance Analysis of Adaptive Wireless Systems

The St. Petersburg game was a problem presented in 1713 by Nicholas Bernoulli [34]. It involves tossing a fair coin repeatedly until the first time it lands "heads." If this happens on the kth toss, the prize w_k is 2^k ducats. The expected winnings are

$$E[W] = \sum_{k=1}^{\infty} w_k \Pr(W = w_k) = \sum_{k=1}^{\infty} 2^k 2^{-k} = \infty \tag{25}$$

where $\Pr(\mathcal{A})$ denotes the probability of the event $\mathcal{A}$. A naive person, who only wants to maximize the expected difference between the winnings and the price of game ticket, would pay any finite sum of money to play this game. But, as Gabriel Cramer observed, no reasonable man would be willing to pay more than 20 ducats to play the game [4]. Thus, rational behavior does not imply maximization of winnings. In utility theory, this anomaly is usually called St. Petersburg paradox.

The expected utility hypothesis was formulated in Cramer's suggestion for resolving this St. Petersburg paradox. Namely, one game is preferred to another one if and only if its expected utility of the

winnings $E[U]$ is higher. Gabriel Cramer and Daniel Bernoulli suggested taking strictly concave functions $U = \sqrt{W}$ and $U = \ln W$, respectively [4]. It can be easily shown that for both of these utility functions $u(w)$, the expected utility of winnings $E[U]$ is finite, that is,

$$E[U] = \sum_{k=1}^{\infty} u(w_k) \Pr(W = w_k) < \infty. \tag{26}$$

The solutions of St. Petersburg paradox proposed by Cramer and Bernoulli are not yet completely satisfying. As pointed out by Menger [28], the game can easily be changed in such a way that the paradox reappears. More specifically, one needs to offer even larger winnings $w_k = \exp(2^k)$ for $U = \ln W$ or $w_k = 2^{2k}$ for $U = \sqrt{W}$. It can be easily verified that with those winnings, the expected utility $E[U]$ is unbounded. In utility theory, this modified variant of St. Petersburg paradox is commonly referred to as super St. Petersburg paradox.

In this section, we explore analogies between St. Petersburg game and adaptive transmission over fading channels to show, by *reductio ad absurdum*, the limitations of commonly used theoretical fading models and the limitations of performance analysis that is based on those theoretical models. In particular, we demonstrate that St. Petersburg paradoxes can occur in analysis of mathematical models of adaptive wireless systems. We construct an adaptive energy control scheme which, according to the standard fading models, leads to infinite average received energy although both the average transmitted energy and the representative energy gain of the channel are finite. Obviously, practical transmission systems cannot achieve infinite average received energy without violating the energy conservation law. We discuss possible reasons for the St. Petersburg paradoxes to occur in analysis of wireless systems and how they can be resolved using suitable performance metrics and fading models. To our best knowledge, no such analysis of St. Petersburg paradoxes, different fading channel models, various adaptive transmission schemes, and allowable class of performance metrics exists in communications literature.

In stationary and ergodic fading channels, a common performance indicator is the expected value of some additive performance metric [5]

$$f(\gamma_1, \gamma_2, \ldots, \gamma_m) = \sum\nolimits_{i=1}^{m} f_i(\gamma_i) \tag{27}$$

which is a function of the received signal-to-noise ratios

$$\gamma_i = \frac{s_i}{\sigma_n^2} = \frac{|\tilde{z}_i - \tilde{n}_i|^2}{\sigma_n^2} = \frac{\lambda_i q_i}{\sigma_n^2}. \tag{28}$$

The symbol s_i in (28) denotes the energy contained in the signal component of $\tilde{z}_i$. Examples of additive performance functions (27) include ergodic capacity of a linear vector channel, where $f_i(\gamma_i) = \log_2(1+\gamma_i)$ for all i, and the average received signal-to-noise ratio with $f_i(\gamma_i) = \gamma_i$ for all i. The optimal adaptive transmission system is identified by finding the extrema of performance indicator $E[f(\gamma_1, \gamma_2, \ldots, \gamma_m)]$ subject to the average energy constraint (6).

The form of (5), (6), and (27) suggests that no generality is lost if we restrict our analysis to single-channel transmission. For that reason, in the rest of the paper, we assume that $m = 1$ and omit the subchannel index i.

4.3.1 Ordinary St. Petersburg Paradox

Let us assume that the transmitter sets the instantaneous transmission energy level q according to the following rule

$$q(\lambda) = \begin{cases} 0, & \text{for } \lambda < \mu \\ \dfrac{\alpha}{\lambda^2 p(\lambda)}, & \text{for } \lambda \geq \mu \end{cases} \tag{29}$$

where μ and α are some positive real constants. Since the transformation (29) associates a unique value of a random variable q with each value of the random variable λ, (6) can be rewritten as [17, p. 615]

$$E[q] = \int_0^\infty q(\lambda)\, p(\lambda)\, d\lambda. \tag{30}$$

Similarly, the energy of the signal component of $\tilde{z}_i$ becomes

$$E[s] = \int_0^\infty gq(\lambda)\, p(\lambda)\, d\lambda. \tag{31}$$

With the energy control rule (29), the average transmitted energy

$$E[q] = \int_\mu^\infty \alpha\lambda^{-2} d\lambda = \alpha\mu^{-1} \tag{32}$$

is finite. The constants α and μ in (32) are selected such that the average energy constraint $E[q] \leq E_{\text{av}}$ is satisfied. However, the average received energy is unbounded because

$$E[s] = \lim_{b\to\infty} \int_\mu^b \frac{\alpha}{\lambda} d\lambda = \alpha \lim_{b\to\infty} \ln b - \alpha \ln \mu = \infty \tag{33}$$

does not converge on $[\mu, \infty)$. Consequently, the average received signal-to-noise ratio $E[\gamma] = E[s]/\sigma_n^2$ is also unbounded.

We can conclude that a naive maximization of the average received signal-to-noise ratio leads to the absurd. Namely, the average received signal-to-noise ratio, as predicted by the model, is unbounded for every possible value of average transmitted energy E_{av} and representative energy gain of the channel G_0. Thus, from mathematical point of view, it is not possible to discriminate different adaptive transmission systems because all of them appear to be equally good. Furthermore, from physical point of view, these adaptive transmission systems violate the energy conservation law.

4.3.2 Super St. Petersburg Paradox

The ordinary St. Petersburg paradox can be sometimes avoided if optimization of adaptive wireless systems is done with respect to some function $f(\gamma)$ of signal-to-noise ratio γ. The optimal system is determined by finding extrema of

$$E[f(\gamma)] = \int_0^\infty f(\gamma)\, p(\gamma)\, d\gamma \tag{34}$$

where $p(\gamma)$ is the probability density function of signal-to-noise ratio γ. However, as the following examples illustrate, the function $f(\gamma)$ cannot be completely arbitrary.

Let us assume that all probability density functions $p(\gamma)$, including those with infinite means, are allowed in comparison. Let $f(\gamma)$ be an increasing convex function. Then, by a well-known property of convex functions

$$f(\gamma) \geq f(\xi) + f'(\xi)(\gamma - \xi) \tag{35}$$

where $f'(\xi) > 0$ denotes the first-order derivative of $f(\gamma)$ at an arbitrary point $0 \leq \xi < \infty$. Since both sides of the inequality

$$f(\gamma) - f(0) \geq f'(\xi)(\gamma - \xi) + f(\xi) - f(0) \tag{36}$$

are nonnegative, we obtain

$$\int_0^\infty f(\gamma)\, p(\gamma)\, d\gamma - f(0) \geq f'(\xi) \int_0^\infty \gamma p(\gamma)\, d\gamma + C \tag{37}$$

where

$$C = f(\xi) - \xi f'(\xi) - f(0) < \infty. \tag{38}$$

Therefore, we can conclude that $E[f(\gamma)]$ is unbounded if $E[\gamma]$ is unbounded and $f(\gamma)$ is an increasing convex function of γ.

On the other hand, if $f(\gamma)$ is an increasing and concave function, then

$$f(\gamma) \leq f(\xi) + f'(\xi)(\gamma - \xi) \tag{39}$$

and

$$\int_0^\infty [f(\gamma) - f(0)]\, p(\gamma)\, d\gamma \leq f'(\xi) \int_0^\infty \gamma p(\gamma)\, d\gamma + C. \tag{40}$$

However, the integral on the left hand side of (40) can diverge when $E[\gamma]$ is unbounded. For example, it can be easily shown that both $E[\gamma]$ and $E[f(\gamma)]$ are unbounded when

$$f(\gamma) = \log_2(1+\gamma) \tag{41}$$

and

$$p(\gamma) = \begin{cases} 0, & \text{for } \gamma < \gamma_0 \\ \dfrac{\ln(1+\gamma_0)}{(1+\gamma)\ln^2(1+\gamma)}, & \text{for } \gamma \geq \gamma_0 \end{cases} \tag{42}$$

where $\gamma_0 > 0$ is an arbitrary positive constant.

4.3.3 Resolving Ordinary St. Petersburg Paradox

Fading Channel with Finite Peak Gain Let us assume that the channel gain λ is a random variable with bounded support, that is,

$$\operatorname{supp}(\lambda) = \overline{\{\lambda \geq 0 : p(\lambda) > 0\}} \subset [0, \overline{\lambda}] \tag{43}$$

where $\overline{\lambda} < \infty$ and $\overline{\mathcal{X}}$ denotes the closure of a set $\mathcal{X}$. In other words, we assume that the energy gain of the channel is bounded above by some finite $\overline{\lambda}$, that is, $0 \leq \lambda \leq \overline{\lambda}$. Thus, from (30) and (31), we obtain

$$E[s] = \int_0^\infty g q(\lambda)\, p(\lambda)\, d\lambda \leq \overline{\lambda} \int_0^\infty q(\lambda)\, p(\lambda)\, d\lambda = \overline{\lambda} E[q]. \tag{44}$$

Since the right hand side of (44) is finite by our assumption $E[q] \leq E_{\text{av}} < \infty$, we conclude that the average received signal-to-noise ratio $E[\gamma]$ must be finite. Consequently, it is always possible to discriminate different transmission system using the value of $E[\gamma]$ as a ranking criterion.

The introduction of a fading model with a bounded energy gain can be fully justified using the energy conservation law. Namely, the energy at the output of the channel $E_{\text{out}} = s + \sigma_n^2$ cannot exceed the energy at its inputs $E_{\text{in}} = q + \sigma_n^2$ or, equivalently, $s \leq q$ for any value of λ. It can be easily seen that the condition $s \leq q$ implies that $\overline{\lambda} \leq 1$.

Energy Control Rule with Finite Peak Gain Let us now assume that the transmitted energy q is a random variable with bounded support, that is,

$$\operatorname{supp}(q) = \overline{\{q \geq 0 : p(q) > 0\}} \subset [0, \overline{q}] \tag{45}$$

where $\overline{q} < \infty$. Consequently, from (30) and (31), we obtain

$$E[s] = \int_0^\infty g q(\lambda)\, p(\lambda)\, d\lambda \leq \overline{q} \int_0^\infty \lambda p(\lambda)\, d\lambda = \overline{q} E[\lambda]. \tag{46}$$

Since the right hand side of (46) is finite by our assumption $G_0 = t^{-1}E[\lambda] < \infty$, we conclude that the average received signal-to-noise ratio $E[\gamma]$ must be finite.

Examples of energy control rules with finite peak gain include water-filling [5, p. 2627]

$$q = \begin{cases} 0, & \text{for } \lambda < \mu_{\text{wf}} \\ \mu_{\text{wf}}^{-1} - \lambda^{-1}, & \text{for } \lambda \geq \mu_{\text{wf}} \end{cases} \tag{47}$$

and truncated channel inversion [5, p. 2629]

$$q = \begin{cases} 0, & \text{for } \lambda < \mu_{\text{tci}} \\ \beta\lambda^{-1}, & \text{for } \lambda \geq \mu_{\text{tci}} \end{cases}. \tag{48}$$

The positive constants μ_{wf}, μ_{tci}, and β in (47) and (48) are selected such that the average energy constraint $E[q] \leq E_{\text{av}}$ is satisfied. It can be easily verified that the peak transmission energy is $\sup q = \mu_{\text{wf}}^{-1} < \infty$ for water-filling and $\sup q = \beta\mu_{\text{tci}}^{-1} < \infty$ for truncated channel inversion. Consequently, these two adaptive transmission schemes ensure existence of finite $E[\gamma]$ in all fading channels, including those with unbounded gains.

Transmission Energy Level Uncorrelated with Channel Gain Ordinary St. Petersburg paradox does not occur when q and λ are uncorrelated random variables. A special case of this transmission strategy is transmission with constant energy, that is, q is a degenerate random variable. The average received signal-to-noise ratio

$$E[\gamma] = \frac{E[s]}{\sigma_n^2} = \frac{E[\lambda \cdot q]}{\sigma_n^2} = \frac{E[\lambda] \cdot E[q]}{\sigma_n^2} \tag{49}$$

is finite because $E[\lambda]$ and $E[q]$ are finite in our system model. Hence, we conclude that $E[q] < \infty$ and $E[\lambda] < \infty$ are sufficient conditions to ensure that $E[\gamma]$ is finite if instantaneous transmission energy level q and channel energy gain λ are uncorrelated.

Energy Control Rule with Finite Second-Order Moment Another way to ensure that $E[\gamma]$ is finite is to limit the higher-order moments of channel energy gain λ and instantaneous transmitted energy q. In particular, by applying Hölder's inequality [17, p. 118] to $E[s] = E[\lambda \cdot q]$, we obtain

$$E[s] = E[\lambda \cdot q] \leq \sqrt{E[\lambda^2]\, E[q^2]}. \tag{50}$$

Therefore, we can conclude that $E[\gamma]$ is finite when second-order moments of random variables q and λ are finite.

Adaptive control of transmission energy level clearly requires stronger assumption about the moments of channel energy gain λ and transmission energy level q than traditionally used first-order conditions $E[\lambda] = tG_0$ and $E[q] < E_{\text{av}}$.

4.3.4 Resolving Super St. Petersburg Paradox

Unbounded Performance Metrics Let us assume that $f(\gamma)$ is a continuous and unbounded function. Furthermore, let the received signal-to-noise ratio γ be a random variable with bounded support, that is,

$$\text{supp}(\gamma) = \overline{\{\gamma \geq 0 : p(\gamma) > 0\}} \subset [a, b] \tag{51}$$

where $0 \leq a < b < \infty$. In other words, we assume that the amount of received energy is bounded below and above by some finite a and b. Then, a continuous function $f(\gamma)$ is necessarily bounded

$$c \leq f(\gamma) \leq d, \quad \text{for } a \leq \gamma \leq b \tag{52}$$

because it is defined on a bounded and closed interval $[a,b]$ [17, p. 93]. The constants c and d are, respectively, the finite minimum

$$c = \min\{f(\gamma) : a \leq \gamma \leq b\} \tag{53}$$

and the finite maximum

$$d = \max\{f(\gamma) : a \leq \gamma \leq b\} \tag{54}$$

of $f(\gamma)$ on $[a,b]$. By the mean value theorem [17, p. 119], there is at least one point $a < \xi < b$ such that

$$E[f(\gamma)] = \int_a^b f(\gamma)\, p(\gamma)\, d\gamma = f(\xi)\,. \tag{55}$$

Since $f(\gamma)$ is continuous and bounded on $[a,b]$, we can conclude that $E[f(\gamma)]$ is finite.

In communications, the peak received signal-to-noise ratio $\sup\gamma$ is of course finite when both the peak energy gain of the channel $\sup\lambda$ and peak transmission energy $\sup q$ are finite. However, these conditions are sufficient but not necessary ones. For example, the peak received signal-to-noise ratio is finite when the received signal-to-noise ratio γ has a so-called simple probability distribution, that is, a probability distribution which assigns probability one to a finite subset of $[0,\infty)$. With truncated channel inversion (48), γ is a discrete random variable with probability density function

$$p(\gamma) = \Pr(\lambda < \mu_{\text{tci}})\,\delta(0) + \Pr(\lambda \geq \mu_{\text{tci}})\,\delta\left(\beta/\sigma_n^2\right) \tag{56}$$

where $\delta(x)$ denotes Dirac's delta function. Thus, the peak received signal-to-noise ratio $\sup\gamma = \beta/\sigma_n^2$ is finite without requiring $\sup\lambda$ or $\sup q$ to be finite.

Let us now assume that γ is a random variable with unbounded support, that is, $p(\gamma) > 0$ for all $a \leq \gamma < \infty$. Furthermore, let $f(\gamma)$ be an lth degree polynomial

$$f(\gamma) = c_l\gamma^l + c_{l-1}\gamma^{l-1} + \ldots + c_1\gamma + c_0 \tag{57}$$

where $c_0, c_1, \ldots, c_l$ are constant real coefficients. Then,

$$E[f(\gamma)] = \sum\nolimits_{j=0}^{l} c_j \int_a^\infty \gamma^j p(\gamma)\, d\gamma < \infty \tag{58}$$

if and only if $E\left[\gamma^l\right] < \infty$ because, by Hölder's inequality [17, p. 118],

$$\int_a^\infty \gamma^j p(\gamma)\, d\gamma \leq \int_a^\infty \gamma^l p(\gamma)\, d\gamma \tag{59}$$

for all $1 \leq j \leq l$. Most of the practically used probability distributions have finite moments with, for example, Cauchy and Pareto distributions being exceptions [33].

Unbounded Concave Performance Metrics Let $f(\gamma)$ be a continuous and monotone increasing function of γ that is concave for some $\gamma \geq \xi$. Furthermore, let γ has unbounded support, that is, $p(\gamma) > 0$ for all $a \leq \gamma < \infty$. Then,

$$E[f(\gamma)] = \int_a^\xi f(\gamma)\, p(\gamma)\, d\gamma + \int_\xi^\infty f(\gamma)\, p(\gamma)\, d\gamma. \tag{60}$$

The first integral on the right hand side of (60) is finite because $f(\gamma)$ is continuous and bounded on $[a,\xi]$ and [17, p. 119]

$$\int_a^\xi f(\gamma)\, p(\gamma)\, d\gamma = f(\zeta) \int_a^\xi p(\gamma)\, d\gamma < \infty \tag{61}$$

where $a < \zeta < \xi$. Therefore, it is sufficient to consider only the second integral on the right hand side of (60) to verify whether $E[f(\gamma)]$ is finite. By a well-known property of concave functions, we obtain

$$f(\gamma) - f(\xi) \leq f'(\xi)(\gamma - \xi). \tag{62}$$

Since both sides of (62) are nonnegative, we obtain

$$\int_{\xi}^{\infty} f(\gamma)\, p(\gamma)\, d\gamma - f(\xi) \leq f'(\xi) \int_{\xi}^{\infty} \gamma p(\gamma)\, d\gamma - \xi f'(\xi). \tag{63}$$

Furthermore,

$$\int_{\xi}^{\infty} \gamma p(\gamma)\, d\gamma \leq \int_{a}^{\infty} \gamma p(\gamma)\, d\gamma = E[\gamma] \tag{64}$$

because $\gamma p(\gamma)$ is nonnegative and $\xi \geq a$. Finally, by combining (60), (61), (63), and (64), we obtain

$$E[f(\gamma)] \leq f'(\xi)\, E[\gamma] + C' \tag{65}$$

where

$$C' = f(\zeta) \int_{a}^{\xi} p(\gamma)\, d\gamma + f(\xi) - \xi f'(\xi) < \infty. \tag{66}$$

Therefore, we can conclude that $E[f(\gamma)]$ is finite provided that $E[\gamma]$ is finite and $f(\gamma)$ is a continuous and monotone increasing function of γ that is concave on some interval $[\xi, \infty)$. Consequently, those functions only permit discrimination among probability distributions of γ which have finite means (cf. [1]).

In communications, an example of unbounded and concave performance metric is channel capacity because $f(\gamma) = \log_2(1+\gamma)$ is concave function of γ but it tends to infinity when γ increases.

Bounded Performance Metrics Consider a continuous and bounded function $f(\gamma)$, that is,

$$c \leq f(\gamma) \leq d, \quad \text{for } 0 \leq \gamma < \infty \tag{67}$$

where

$$c = \inf\{f(\gamma) : 0 \leq \gamma < \infty\} > -\infty \tag{68}$$

and

$$d = \sup\{f(\gamma) : 0 \leq \gamma < \infty\} < +\infty. \tag{69}$$

Finite constants c and d are the infimum and the supremum of $f(\gamma)$ on $[0, \infty)$, respectively. From (55) we can conclude that there is at least one point $0 < \xi < \infty$ such that $E[f(\gamma)] = f(\xi) < \infty$ because $f(\gamma)$ is continuous and bounded on $[0, \infty)$. Consequently, one can discriminate among all possible adaptive transmission systems corresponding to all possible probability distributions of γ, including those with infinite means (cf. [28]).

In communications, an example of continuous and bounded performance metric is symbol error rate because probability of symbol error as a function of the received signal-to-noise ratio γ is always a real number between zero and one.

4.4 Improving Link Budget Analysis of Adaptive Wireless Systems with Probabilistic Inequalities

The ratio of the desired received power P_d to thermal noise power σ_n^2 before detection is commonly called the desired signal-to-noise ratio γ_d. The equation that relates desired signal-to-noise ratio to the link parameters, such as transmitted power P_t, path loss L_p, receiver and transmitter antenna gains G_r and G_t, etc., is called the link budget [40].

In this section, we focus on the link budget analysis in adaptive transmission systems. We demonstrate fundamental differences in the link budget analysis of adaptive and nonadaptive systems. The novelty of our approach arises from the fact that we consider the effects of possible adaptation of transmission power at the transmitter, a fact that is usually ignored in the conventional link budget analysis. We propose to complement the conventional link budget analysis with the correction term that we call *adaptation gain*. We derive the upper and lower bounds of the adaptation gain. Furthermore, we recognize that the calculation of the required fade margin also needs a refinement because adaptation of the transmission power changes not only the average received power $E[P_r]$ but also the distribution of the received power P_r. We derive the lower and upper bounds of the fade margin.

4.4.1 Propagation Modeling and Link Budget Analysis

Empirical propagation models are usually based on measurements. The transmitter sends a constant-power sounding signal, usually a bandlimited white noise-like signal. At the receiver, a local mean is obtained by measuring squared-envelope of the received signal and averaging it over a spatial distance of several tens of wavelengths. Usually, the local mean experiences slow variations due to the presence of large objects such as buildings or trees. This phenomenon is commonly called shadowing. If the local mean is averaged over sufficiently large spatial distances to average over the shadows, the area mean is obtained [40]. The area mean μ_r is the average signal strength that is received to or from a mobile station over a large area that lies at approximately the same distance from the base station [40], i.e.,

$$\mu_r = E[G]\, P_t. \tag{70}$$

The inverse of $E[G]$ is usually referred to as empirical path loss. The empirical path loss describes all the propagation effects between the transmitter and receiver such as antenna loses, shadowing, etc. In nonadaptive systems, all the parameters such as transmission power, antenna gains, receiver sensitivity, required minimum received signal-to-noise ratio are fixed. For that reason, designers usually determine the maximum allowable path loss that would still yield acceptable link performance. In a nonadaptive transmission system, the maximum allowable path loss is [40, p. 28, eq. 1.27])

$$L_{\max(\mathrm{dB})} = P_{t(\mathrm{dBm})} + G_{t(\mathrm{dB})} + G_{r(\mathrm{dB})} - S_{r(\mathrm{dBm})} - M_{(\mathrm{dB})} - L_{I(\mathrm{dB})} + G_{h(\mathrm{dB})} \tag{71}$$

where G_t and G_r are, respectively, transmitter and receiver antenna gains, S_r is the receiver sensitivity, M is the fade margin, L_I is the interference margin, and G_h is the handoff gain.

4.4.2 Adaptation Gain

In adaptive transmission systems which employ adaptive transmission power control, the instantaneous transmission power P_t and instantaneous channel gain G are correlated. Thus, the average received power is [41]

$$E[P_r] = E[G \cdot P_t] = E[G]\, E[P_t] \left(1 + \frac{\mathrm{Cov}[G, P_t]}{E[G]\, E[P_t]}\right) \tag{72}$$

where $\mathrm{Cov}\left[G, P_t\right]$ denotes the covariance between random variables G and P_t. In a system without adaptive power control $E\left[P_t\right] = P_t$ and $\mathrm{Cov}\left[G, P_t\right] = 0$ and thus (72) reduces to (70). We refer to the term

$$A = \frac{E\left[P_r\right]}{E\left[G\right] E\left[P_t\right]} = 1 + \frac{\mathrm{Cov}\left[G, P_t\right]}{E\left[G\right] E\left[P_t\right]} \tag{73}$$

as *adaptation gain*. In words, the adaptation gain describes how the area mean changes when adaptive power control is used at the transmitter. Consequently, in an adaptive transmission system, the maximum allowable path loss is

$$L_{\max(\mathrm{dB})} = P_{t(\mathrm{dBm})} + G_{t(\mathrm{dB})} + G_{r(\mathrm{dB})} + A_{(\mathrm{dB})} - S_{r(\mathrm{dBm})} - M_{(\mathrm{dB})} - L_{I(\mathrm{dB})} + G_{h(\mathrm{dB})} \tag{74}$$

where the adaptation gain $A_{(\mathrm{dB})}$ is an additional correction term that describes additional effects due to adaptation of transmission power at the transmitter.

Proposition 1. *In any transmission system, the average received power satisfies the condition*

$$E\left[P_r\right] \geq E^2\left[\sqrt{P_t}\right] \left(E\left[G^{-1}\right]\right)^{-1} \tag{75}$$

with equality if and only if $P_t = E\left[P_t\right] \left(E\left[G^{-2}\right]\right)^{-1} G^{-2}$*. Consequently, the lower bound on adaptation gain is*

$$A_{(\mathrm{dB})} \geq 10 \log_{10} \frac{E^2\left[\sqrt{P_t}\right]}{E\left[P_t\right] E\left[G\right] E\left[G^{-1}\right]}. \tag{76}$$

Proof. We prove the proposition using a modified version of the Hölder inequality [17, p. 118]. We introduce the modification to the original form of the Hölder inequality to account for an arbitrary weighting function $\phi\left(\lambda\right)$. For arbitrary functions $f\left(\lambda\right)$, $g\left(\lambda\right)$, and $\phi\left(\lambda\right)$

$$\begin{aligned} &\left| \int_a^b f^{\frac{1}{p}}\left(\lambda\right) \phi\left(\lambda\right) d\lambda \right| \\ &= \left| \int_a^b \left[f\left(\lambda\right) g\left(\lambda\right) \phi\left(\lambda\right)\right]^{\frac{1}{p}} \cdot g^{-\frac{1}{p}}\left(\lambda\right) \phi^{\frac{1}{q}}\left(\lambda\right) d\lambda \right| \\ &\leq \left[\int_a^b \left| f\left(\lambda\right) g\left(\lambda\right) \phi\left(\lambda\right) \right| d\lambda \right]^{\frac{1}{p}} \left[\int_a^b \left| g^{-\frac{q}{p}}\left(\lambda\right) \phi\left(\lambda\right) \right| d\lambda \right]^{\frac{1}{q}} \end{aligned} \tag{77}$$

where $p^{-1} + q^{-1} = 1$ and $p > 1$. By rearranging the terms in (77), we obtain

$$\begin{aligned} &\int_a^b \left| f\left(\lambda\right) g\left(\lambda\right) \phi\left(\lambda\right) \right| d\lambda \\ &\geq \left| \int_a^b f^{\frac{1}{p}}\left(\lambda\right) \phi\left(\lambda\right) d\lambda \right|^p \left[\int_a^b \left| g^{-\frac{q}{p}}\left(\lambda\right) \phi\left(\lambda\right) \right| d\lambda \right]^{-\frac{p}{q}}. \end{aligned} \tag{78}$$

The equality in (77) and in (78) is obtained if and only if

$$f\left(\lambda\right) g\left(\lambda\right) = c g^{-\frac{q}{p}}\left(\lambda\right) \tag{79}$$

where $c > 0$ is an arbitrary positive constant. By using the substitutions $P_t = f\left(\lambda\right)$, $G = g\left(\lambda\right)$, and $p = q = 2$ in (78), we obtain

$$E\left[P_r\right] \geq E^2\left[\sqrt{P_t}\right] \left(E\left[G^{-1}\right]\right)^{-1} \tag{80}$$

with equality if and only if

$$P_t G = c G^{-1}. \tag{81}$$

Since $E[P_t] = cE[G^{-2}]$, we conclude that

$$c = E[P_t]\left(E\left[G^{-2}\right]\right)^{-1} \tag{82}$$

and equality in (80) is obtained if and only if

$$P_t = E[P_t]\left(E\left[G^{-2}\right]\right)^{-1} G^{-2} \tag{83}$$

which was to demonstrate. □

Proposition 2. *In any transmission system, the average received power satisfies the condition*

$$E[P_r] \leq \sqrt{E[P_t^2]\,E[G^2]} \tag{84}$$

with equality if and only if $P_t = E[P_t](E[G])^{-1} G$*. Consequently, the upper bound on adaptation gain is*

$$A_{(\mathrm{dB})} \leq 10\log_{10}\frac{\sqrt{E[P_t^2]\,E[G^2]}}{E[P_t]\,E[G]}. \tag{85}$$

Proof. We prove the proposition using a modified version of the Hölder inequality [17, p. 118]. We introduce the modification to the original form of the Hölder inequality to account for an arbitrary weighting function $\phi(\lambda)$. For arbitrary functions $f(\lambda)$, $g(\lambda)$, and $\phi(\lambda)$

$$\begin{aligned}\left|\int_a^b f(\lambda)\,g(\lambda)\,\phi(\lambda)\,d\lambda\right| &= \left|\int_a^b f(\lambda)\,\phi^{\frac{1}{p}}(\lambda)\cdot g(\lambda)\,\phi^{\frac{1}{q}}(\lambda)\,d\lambda\right| \\ &\leq \left[\int_a^b |f^p(\lambda)\,\phi(\lambda)|\,d\lambda\right]^{\frac{1}{p}}\left[\int_a^b |g^q(\lambda)\,\phi(\lambda)|\,d\lambda\right]^{\frac{1}{q}}\end{aligned} \tag{86}$$

where $p^{-1} + q^{-1} = 1$ and $p > 1$. The equality in (86) is obtained if and only if

$$f^p(\lambda) = cg^q(\lambda) = c[g(\lambda)]^{\frac{p}{p-1}} \tag{87}$$

where $c > 0$ is an arbitrary positive constant. By using the substitutions $P_t = f(\lambda)$, $G = g(\lambda)$, and $p = q = 2$ in (86), we obtain

$$E[P_r] \leq \sqrt{E[P_t^2]\,E[G^2]} \tag{88}$$

with equality if and only if

$$P_t = \sqrt{c}G. \tag{89}$$

Since $E[P_t] = \sqrt{c}E[G]$, we conclude that

$$\sqrt{c} = E[P_t](E[G])^{-1} \tag{90}$$

and equality in (88) is obtained if and only if

$$P_t = E[P_t](E[G])^{-1} G \tag{91}$$

which was to demonstrate. □

4.4.3 Fade Margin

We define the fade margin as $M = E[P_r]/P_d$ and denote probability of the outage event by $\alpha = \Pr(P_r < P_d)$ and probability of exceeding the desired received power by $\beta = \Pr(P_r > P_d)$.

Proposition 3. *For any distribution of P_r and any value of P_d, the required fade margin*

$$M \geq 1 - \alpha \tag{92}$$

with equality if and only if P_r has the two-point distribution with $\Pr(P_r = 0) = \alpha$ and $\Pr(P_r = P_d) = 1 - \alpha$.

Proof. By Markov's inequality [42, p. 44], for any non-negative random variable P_r with expected value $E[P_r]$ and $P_d > 0$, the probability

$$\Pr(P_r \geq P_d) \leq \frac{E[P_r]}{P_d} = M \tag{93}$$

with equality if and only if the random variable P_r has the two-point distribution such that

$$\Pr(P_r = P_d) = 1 - \Pr(P_r = 0) = \frac{E[P_r]}{P_d}. \tag{94}$$

From (93) and (94) we can conclude that

$$M \geq \Pr(P_r \geq P_d) = 1 - \Pr(P_r < P_d) = 1 - \alpha \tag{95}$$

with equality if and only if P_r has the two-point distribution with $\Pr(P_r = 0) = \alpha$ and $\Pr(P_r = P_d) = 1 - \alpha$. □

Proposition 4. *For any distribution of P_r and any $P_d < E[P_r]$, the required fade margin*

$$M \leq 1 + P_d^{-1}\sqrt{\beta(1-\beta)^{-1} V[P_r]} \tag{96}$$

with equality if and only if P_r has the two-point distribution with

$$\Pr(P_r = P_d) = 1 - \beta \tag{97}$$

$$\Pr\left(P_r = \beta^{-1}E[P_r] - \beta^{-1}(1-\beta)P_d\right) = \beta \tag{98}$$

where $V[P_r]$ denotes the variance of P_r.

Proof. We prove the proposition using the one-sided Chebyshev inequality [44, p. 152]. First, we note that

$$\Pr(P_r \leq P_d) = \Pr(P_r - E[P_r] \leq P_d - E[P_r]) = \Pr(E[P_r] - P_r \geq E[P_r] - P_d). \tag{99}$$

To apply the one-sided Chebyshev inequality, the term $E[P_r] - P_d$ in (99) must be positive, that is, $E[P_r] - P_d > 0$. The condition $E[P_r] - P_d > 0$ implies that $P_d < E[P_r]$, or equivalently, $M > 1$ because $E[P_r] = MP_d$. By the one-sided Chebyshev inequality [44, p. 152],

$$\Pr(P_r \leq P_d) = \Pr(E[P_r] - P_r \geq E[P_r] - P_d) \leq \frac{V[P_r]}{V[P_r] + (E[P_r] - P_d)^2}. \tag{100}$$

After some simple algebra in (100), we obtain

$$\frac{\left(E\left[P_r\right]-P_d\right)^2}{V\left[P_r\right]}=\frac{P_d^2\left(M-1\right)^2}{V\left[P_r\right]}\leq\frac{1}{\Pr\left(P_r\leq P_d\right)}-1. \tag{101}$$

Since both sides of inequality (101) are nonnegative, we can rewrite (101) as

$$\frac{P_d\left(M-1\right)}{\sqrt{V\left[P_r\right]}}\leq\sqrt{\frac{1}{\Pr\left(P_r\leq P_d\right)}-1} \tag{102}$$

After some simple algebra in (102), we obtain

$$M\leq 1+P_d^{-1}\sqrt{\beta\left(1-\beta\right)^{-1}V\left[P_r\right]} \tag{103}$$

because

$$\Pr\left(P_r\leq P_d\right)=1-\Pr\left(P_r>P_d\right)=1-\beta. \tag{104}$$

The equality in one-sided Chebyshev inequality of the form

$$\Pr\left(X-E\left[X\right]\geq kD\left[X\right]\right)\leq\frac{1}{1+k^2}, \tag{105}$$

where $k>0$ and $D\left[X\right]$ denotes the standard deviation of a random variable X, is obtained when the random variable X has a two-point distribution with probabilities (cf. [45, 46])

$$\Pr\left(X=E\left[X\right]+kD\left[X\right]\right)=\frac{1}{1+k^2} \tag{106}$$

$$\Pr\left(X=E\left[X\right]-k^{-1}D\left[X\right]\right)=\frac{k^2}{1+k^2}. \tag{107}$$

By comparing (100) and (105) we conclude that $X=-P_r$, $E\left[X\right]=-E\left[P_r\right]$, $D\left[X\right]=D\left[P_r\right]$, and

$$kD\left[X\right]=E\left[P_r\right]-P_d>0. \tag{108}$$

By substituting (108) into (106) we obtain (97). Similarly, by substituting (108) into (107) and using

$$\Pr\left(P_r=P_d\right)=\Pr\left(X=E\left[X\right]+kD\left[X\right]\right)=\frac{1}{1+k^2}=1-\Pr\left(P_r>P_d\right)=1-\beta \tag{109}$$

we obtain (98). □

Remark 1. *We demonstrate the relationship between the adaptation gain A and the maximum fade margin $\max M$ using the results from [43]. In particular, the variance of the received signal-to-noise ratio in* (96) *is (cf. [43, p. 712])*

$$V\left[P_r\right]\approx E^2\left[P_t\right]V\left[G\right]+E^2\left[G\right]V\left[P_t\right]+2E^2\left[G\right]E^2\left[P_t\right]\left(A-1\right). \tag{110}$$

In general, Propositions 3 and 4 together with Remark 1 imply that the adaptation of transmission power at the transmitter has a significant effect on the value of the required fade margin. In general, the fade margin achieves its minimum and maximum values when the P_r has a two-point distribution. Furthermore, as (110) suggests, the maximum value of fade margin M increases when the adaptation gain A increases.

The adaptation gains $A_{(\mathrm{dB})}$ and fade margins $M_{(\mathrm{dB})}$ achieved by water-filling (WF) [5, p. 2627] and truncated channel inversion (CI) [5, p. 2629] in a composite gamma/lognormal fading channel are shown

in Fig. 5. In a composite gamma/lognormal fading channel, the probability density function of G has the form [40, p. 102]

$$p(G) = \frac{\xi m^m G^{m-1}}{\sqrt{2\pi}\sigma_s \Gamma(m)} \int_0^\infty \frac{1}{w^{m+1}} \exp\left[-\frac{(\xi \ln w - \mu_s)^2}{2\sigma_s^2} - \frac{mG}{w} \right] dw \tag{111}$$

where $\xi = 10/\ln 10$. We set the multipath parameter m to 4, the mean μ_s and the standard deviation σ_s of $10\log_{10} G$ to 0 dB and 4 dB, respectively. The transmitted signal-to-noise ratio was set to $10\log_{10} E[P_t/\sigma_n^2] = 5$ dB and the target outage probability α was set to 0.1.

The results in Fig. 5 show that the adaptation gain $A_{(\mathrm{dB})}$ of the water-filling is positive whereas the adaptation gain $A_{(\mathrm{dB})}$ of truncated channel inversion is negative. Furthermore, the water-filling power control scheme requires much larger fade margin $M_{(\mathrm{dB})}$ than the nonadaptive system to meet the same requirement of outage probability. The truncated channel inversion, on the other hand, does not require any fade margin to meet the outage probability requirement. In fact, the fade margin $M_{(\mathrm{dB})}$ of truncated channel inversion is negative.

In Fig. 6 we plot the adaptation gains $A_{(\mathrm{dB})}$ of water-filling and truncated channel inversion power control rules as the function of the average transmitted signal-to-noise ratio $E[P_t/\sigma_n^2]$. The upper and lower bounds of the adaptation gain $A_{(\mathrm{dB})}$ are shown for comparison. The numerical results suggest that with the water-filling power control rule one achieves almost the highest possible value of adaptation gain. However, the adaptation gain of water-filling power control rule vanishes when the transmitted signal-to-noise ratio increases. The adaptation gain $A_{(\mathrm{dB})}$ of truncated channel inversion is independent of the average transmitted signal-to-noise ratio and it is almost the lowest possible adaptation gain.

In Fig. 7 we plot the fade margin $M_{(\mathrm{dB})}$ required to meet the outage probability requirement $\alpha = 0.1$ both with water-filling and truncated channel inversion. Fade margins are plotted as functions of the average transmitted signal-to-noise ratio $E[P_t/\sigma_n^2]$. The numerical results suggest that the truncated channel inversion does not need any fade margin to meet the outage requirement $\alpha = 0.1$ and the fade margin does not depend on $E[P_t/\sigma_n^2]$. On the other hand, the required fade margin with water-filling power control significantly depends on $E[P_t/\sigma_n^2]$. The fade margin of water-filling reduces to the fade margin of a nonadaptive systems when average transmitted signal-to-noise ratio increases. However, the fade margin of water-filling increases without any bound when the transmitted signal-to-noise ratio decreases. In fact, for the value of the average transmitted signal-to-noise ratio below approximately 4.5 dB meeting the outage probability requirement $\alpha = 0.1$ with water-filling power control is not possible and the value of the fade margin is infinite.

4.5 Application of Rational Decision Theory to Adaptive Power Control

4.5.1 Safety-first schemes

Roy's (22) and Kataoka's (23) safety-first approaches have already been used, without reference to risk-reward theory, in [6] as objective criteria for designing adaptive transmission systems. Therefore, we consider only the Telser's safety first approach defined in (24).

According to Telser's safety-first approach, one needs to maximize the expected link spectral efficiency $E[\eta_r]$ subject to the outage probability constraint $\Pr(\eta_r \le \alpha) \le \zeta$ and the minimum acceptable link spectral efficiency η_t

$$\max\left\{ E[\eta_r] \,|\Pr(\eta_r \le \eta_t) \le \zeta \right\}. \tag{112}$$

The optimization problem (112) can be formulated as the variational problem with the expected link spectral efficiency $E[\eta_r]$ as the variational integral and the constraints

$$\Pr\left(\sum_{k=1}^{m} \log_2\left(1 + \frac{\rho_k^2 \lambda_k}{\sigma_n^2} \right) \le \eta_t \right) \le \zeta \tag{113}$$

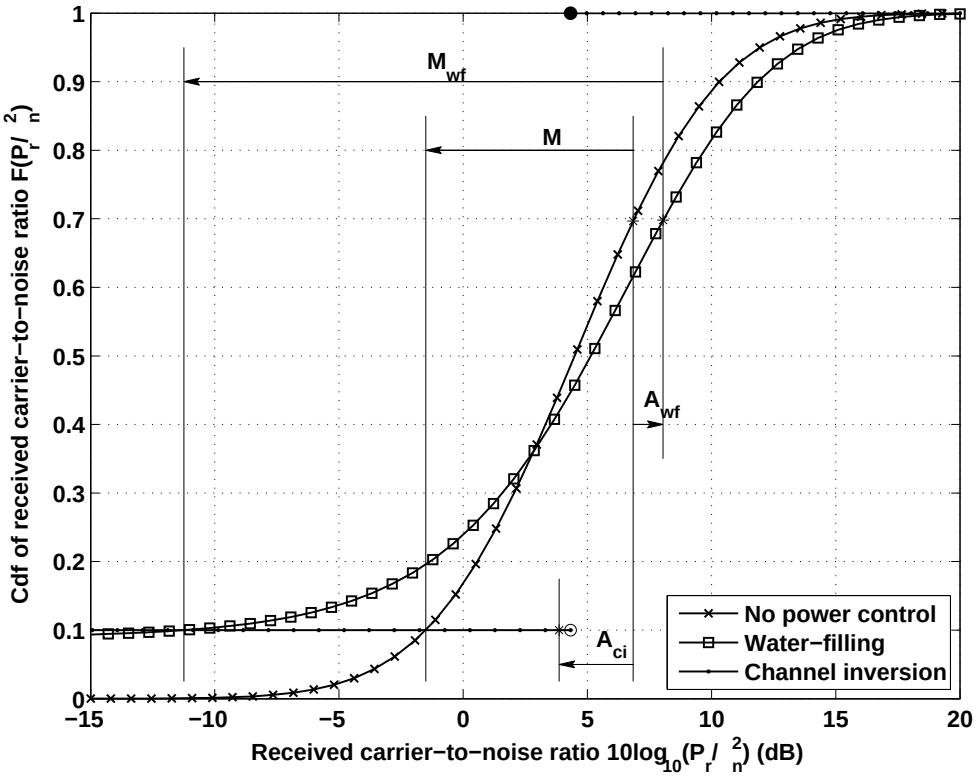

Figure 5: Adaptation gain and fade margin for two adaptive power control rules in a composite gamma/lognormal fading channel.

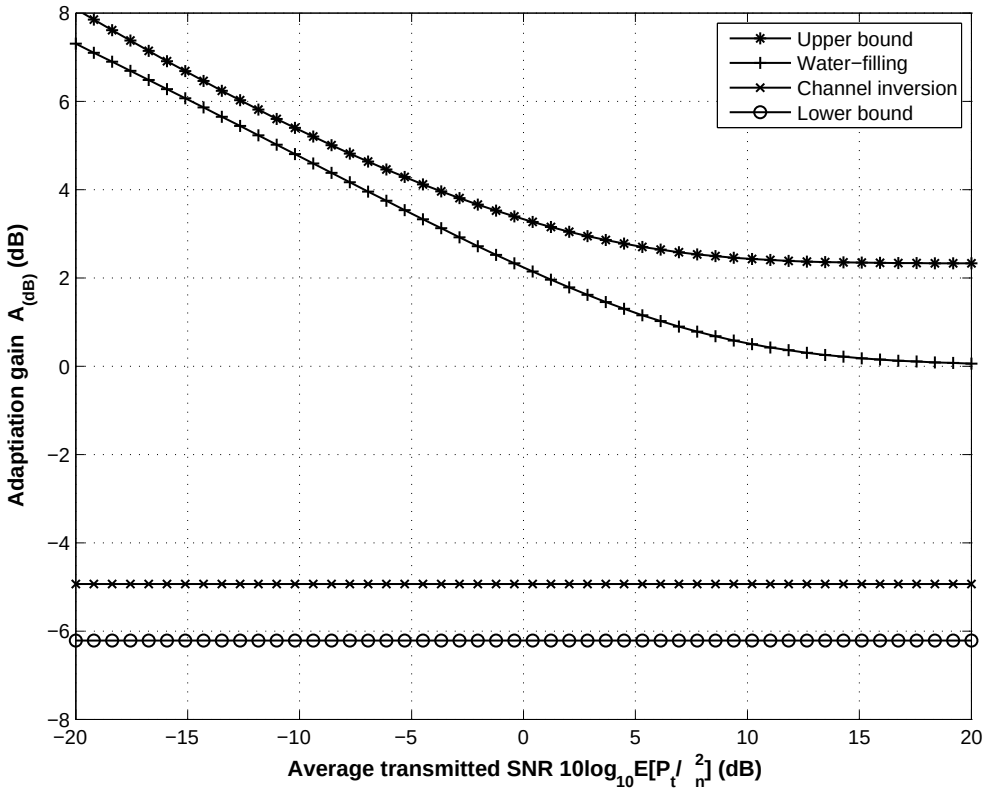

Figure 6: Adaptation gain of two adaptive power control rules in a composite gamma/lognormal fading channel.

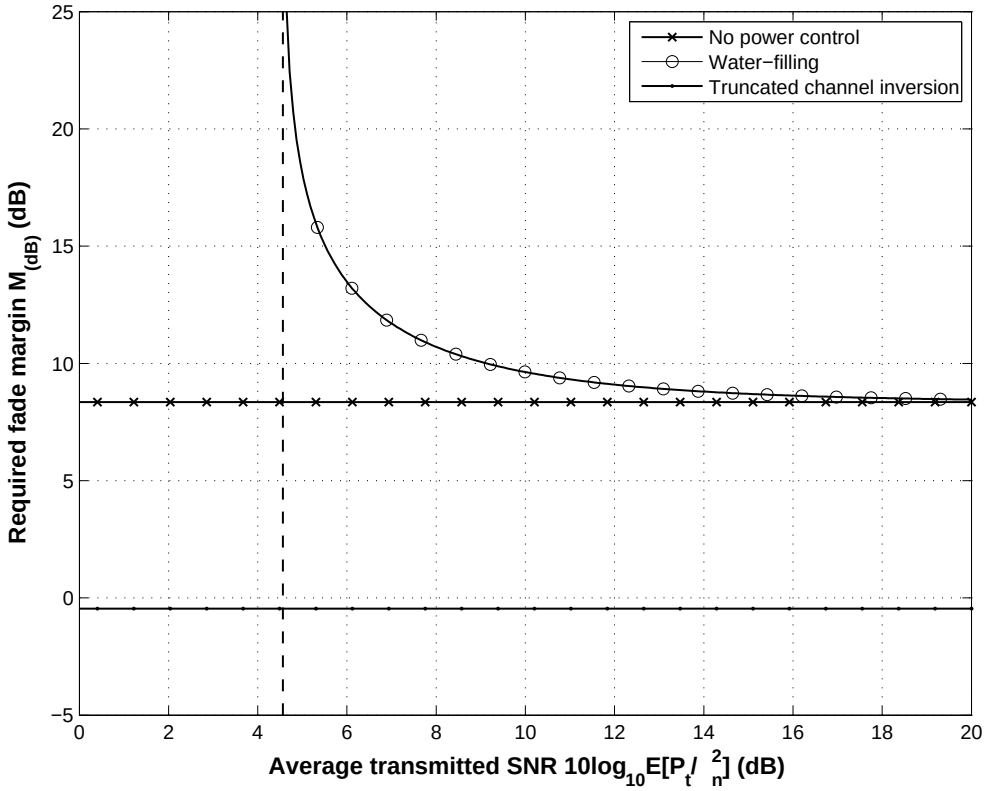

Figure 7: Fade margin for two adaptive power control rules in a composite gamma/lognormal fading channel.

$$\int_R \sum_{k=1}^{m} \rho_k^2 \, p(\boldsymbol{\lambda}) \, d\boldsymbol{\lambda} = (1-\beta) \, P_{\text{av}} \tag{114}$$

$$\int_T \sum_{k=1}^{m} \rho_k^2 \, p(\boldsymbol{\lambda}) \, d\boldsymbol{\lambda} = \beta P_{\text{av}} \tag{115}$$

as subsidiary conditions [12, p. 45]. The integration regions R and T are defined as follows

$$T = \left\{ (\lambda_1, \lambda_2, \ldots, \lambda_n) : \sum_{k=1}^{m} \log_2 \left(1 + \frac{\rho_k^2 \lambda_k}{\sigma_n^2} \right) \geq \eta_t \right\} \tag{116}$$

and

$$R = \left\{ (\lambda_1, \lambda_2, \ldots, \lambda_n) : \sum_{k=1}^{m} \log_2 \left(1 + \frac{\rho_k^2 \lambda_k}{\sigma_n^2} \right) \leq \eta_t \right\}. \tag{117}$$

The constraint (113) is an ordinary outage probability constraint. Constraints (114) and (115) together constitute the average power constraint. The parameter $0 < \beta \leq 1$ controls the division of transmission power below and above the target link spectral efficiency η_t. Furthermore, we have additional inequality constraints, i.e., $q_k \geq 0$ for all $k = 1, 2, \ldots, m$. Since the inequality constraints cannot be easily added to variational problems, we introduce a new vector-valued function $\boldsymbol{\rho} : \mathbb{R}_+^m \to \mathbb{R}_+^m$ and reformulate the optimization problem accordingly. Namely, we assume that $q_k = \rho_k^2$ which ensures that $\boldsymbol{q} \in \mathbb{R}_+^m$. The optimization problem (112) can therefore be stated as

$$\min \left\{ J[\boldsymbol{\rho}] = -E\left[\eta_r\right] | K_1[\boldsymbol{\rho}] = \zeta, K_2[\boldsymbol{\rho}] = \beta P_{\text{av}}, K_3[\boldsymbol{\rho}] = (1-\beta) \, P_{\text{av}} \right\} \tag{118}$$

where

$$J\left[\boldsymbol{\rho}\right] = \int_0^{\infty} \int_0^{\lambda_1} \cdots \int_0^{\lambda_{m-1}} F(\boldsymbol{\lambda}, \boldsymbol{\rho}) \, d\lambda_m \cdots d\lambda_2 d\lambda_1 \tag{119}$$

$$K_1\left[\boldsymbol{\rho}\right] = \int_R G_1(\boldsymbol{\lambda}, \boldsymbol{\rho}) \, d\lambda_m \cdots d\lambda_2 d\lambda_1 \tag{120}$$

$$K_2\left[\boldsymbol{\rho}\right] = \int_R G_2(\boldsymbol{\lambda}, \boldsymbol{\rho}) \, d\lambda_m \cdots d\lambda_2 d\lambda_1 \tag{121}$$

$$K_3\left[\boldsymbol{\rho}\right] = \int_T G_3(\boldsymbol{\lambda}, \boldsymbol{\rho}) \, d\lambda_m \cdots d\lambda_2 d\lambda_1 \tag{122}$$

are integrals over the fixed region Q. The respective variational integrands in (119), (120), (121), and (122) are given by

$$F(\boldsymbol{\lambda}, \boldsymbol{\rho}) = -\left[\sum_{k=1}^{m} \log_2 \left(1 + \frac{\rho_k^2 \lambda_k}{\sigma_n^2} \right) \right]_+ p(\boldsymbol{\lambda}) \tag{123}$$

$$G_1(\boldsymbol{\lambda}, \boldsymbol{\rho}) = p(\boldsymbol{\lambda}) \tag{124}$$

$$G_2(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \sum_{k=1}^{m} \rho_k^2 p(\boldsymbol{\lambda}) \tag{125}$$

$$G_3(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \sum_{k=1}^{m} \rho_k^2 p(\boldsymbol{\lambda}) \tag{126}$$

A necessary condition for the vector-valued function $\boldsymbol{\rho}$ to be an extremal of $J\left[\boldsymbol{\rho}\right]$ is that the function components $\rho_k = \rho_k(\lambda_k)$ satisfy Euler-Lagrange equations, i.e., [12, p. 152]

$$\frac{\partial}{\partial \rho_k} L(\boldsymbol{\lambda}, \boldsymbol{\rho}) = 0, \quad k = 1, 2, \ldots, m \tag{127}$$

where

$$L(\boldsymbol{\lambda}, \boldsymbol{\rho}) = F(\boldsymbol{\lambda}, \boldsymbol{\rho}) + \xi_1 G_1(\boldsymbol{\lambda}, \boldsymbol{\rho}) + \xi_2 G_2(\boldsymbol{\lambda}, \boldsymbol{\rho}) + \xi_3 G_3(\boldsymbol{\lambda}, \boldsymbol{\rho}) \tag{128}$$

and where $\xi_1 \neq 0$, $\xi_2 \neq 0$, and $\xi_3 \neq 0$ are three constants, also known as Lagrange multipliers, which are independent of $\boldsymbol{\lambda}$ [12, p. 45].

We observe that some first order partial derivatives $\partial F/\partial \rho_k$ may not exist when $\eta_r = \eta_t$ because the function $F(\boldsymbol{\lambda}, \boldsymbol{\rho})$ may have a jump discontinuity on the ridge between regions R and T.Hence, we assume that the general solution to (118) is given by discontinuous functions ρ_k [12, p. 61]. Consequently, the functions ρ_k are defined separately in the region $T = \{\boldsymbol{\lambda} \in Q : \eta_r(\boldsymbol{\lambda}, \boldsymbol{\rho}) \geq \eta_t\}$ and its complement $R = \{\boldsymbol{\lambda} \in Q : \eta_r(\boldsymbol{\lambda}, \boldsymbol{\rho}) < \eta_t\}$. Furthermore, on the boundary surface $S = \{\boldsymbol{\lambda} \in Q : \eta_r(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \eta_t\} \subset T$ the functions ρ_k must satisfy the Weierstrass-Erdmann conditions [12, p. 63]

$$\lim_{\eta_r \to \eta_t^-} \frac{\partial L}{\partial \rho_k'} = \lim_{\eta_r \to \eta_t^+} \frac{\partial L}{\partial \rho_k'} \tag{129}$$

$$\lim_{\eta_r \to \eta_t^-} \left(L - \rho_k' \frac{\partial L}{\partial \rho_k'} \right) = \lim_{\eta_r \to \eta_t^+} \left(L - \rho_k' \frac{\partial L}{\partial \rho_k'} \right). \tag{130}$$

The symbol ρ_k' in (129) and (130) denotes the first order derivative of ρ_k with respect to λ_k, i.e., $\rho_k' = \partial \rho_k / \partial \lambda_k$.

Proposition 5. *If $\boldsymbol{\lambda} \in T$, then the kth component of $\boldsymbol{q}$ is*

$$q_k = \sigma_n^2 \left(\theta_T^{-1} - \lambda_k^{-1} \right)_+, \quad k = 1, 2, \ldots, m \tag{131}$$

where

$$\theta_T = \sigma_n^2 \xi_3 \ln 2 > 0. \tag{132}$$

On the other hand, if $\boldsymbol{\lambda} \in R$, then the kth component of $\boldsymbol{q}$ is

$$q_k = \sigma_n^2 \left(\theta_R^{-1} - \lambda_k^{-1} \right)_+, \quad k = 1, 2, \ldots, m \tag{133}$$

where

$$\theta_R = \sigma_n^2 \xi_2 \ln 2 > 0. \tag{134}$$

Proof. Deferred to Appendix. □

Remark 2. *The optimization problem* (118) *involves four constants or variables: $\beta, \xi_1, \xi_2, \xi_3$. However, we have only three necessary conditions* (127)*,* (129)*, and* (130)*. It means that the solution of the optimization problem cannot be uniquely determined. In fact, the solution to the optimization problem* (118) *is a function of one of those parameters.*

The value of the third Lagrange multiplier ξ_3 can be determined from the second Weierstrass-Erdmann condition (130) provided that ξ_1 and ξ_2 are found from Euler-Lagrange equations (127) and β is defined by the system designer.

We illustrate the performance of the adaptive power control algorithm in a single-input single-output (SISO) Rayleigh fading channel. The adaptive power control rule algorithm is designed based on the Telser's safety-first approach. The average transmitted signal-to-noise ratio is set to 10 dB and target link spectral efficiency is 2 bit/(s·Hz), that is, $P_{\text{av}}/\sigma_n^2 = 10$ dB and $\eta_t = 2$. The relationship between all Lagrange multipliers is presented in Figure 8. The first Lagrange multiplier ξ_1 is negative whereas the second ξ_2 and third ξ_3 Lagrange multipliers are positive. Since

$$\xi_1 \approx -\frac{\Delta J}{\Delta K_1} = -\frac{\Delta \mathrm{E}\left[\eta_r\right]}{\Delta \zeta} < 0, \tag{135}$$

$$\xi_2 \approx -\frac{\Delta J}{\Delta K_2} = -\frac{1}{\beta}\frac{\Delta \mathrm{E}\left[\eta_r\right]}{\Delta P_{\mathrm{av}}} > 0, \tag{136}$$

and

$$\xi_3 \approx -\frac{\Delta J}{\Delta K_3} = -\frac{1}{1-\beta}\frac{\Delta \mathrm{E}\left[\eta_r\right]}{\Delta P_{\mathrm{av}}} > 0, \tag{137}$$

we conclude that the average link spectral efficiency decreases when we increase the power in regions R and T and increases when we increase the allowable outage probability ζ. The symbols ΔJ, ΔK_1, ΔK_2, and ΔK_3 denote the arbitrarily small changes of objective function and the first, second, and third constraint, respectively. The decrease of mean link spectral efficiency when we increase the transmission power in the region T is shown in Figure 9. In Figure 10 we plot the relationship between the second ξ_2, the third ξ_3 Lagrange multiplier, and mean link spectral efficiency $\mathrm{E}\left[\eta_r\right]$. It can be seen that the mean link spectral efficiency increases when the value of the first Lagrange multiplier ξ_1 is increased towards zero. This is expected behavior, because the mean link spectral efficiency achieves its maximum value when the outage probability constraint is relaxed, that is when $\xi_1 = 0$. Then, $\xi_2 = \xi_3$ and the power control rule becomes the conventional water-filling.

In a general case, the Lagrange multipliers ξ_2 and ξ_3 are different, that is $\xi_2 \neq \xi_3$. Consequently, the power control rule as well as achievable link spectral efficiency is not continuous on the boundary of regions R and T. In a SISO channel, the boundary of regions R and T is a point. This effect is illustrated in Figure 11 for a power control rule that needs to satisfy the requirement of minimum target link spectral efficiency of $\eta_t = 2$ bit/(s·Hz) with outage probability $\zeta = 0.25$. It can be easily shown that the difference between integrand values on the boundary of R and T must be equal to ξ_1. By comparing the results in Figures 8 and 11, we see that the difference is approximately 0.6 as expected. Furthermore, the power control rule becomes continuous if and only if $\xi_1 = 0$, that is, when we relax the outage probability constraint K_1.

There is a fundamental trade-off between target link spectral efficiency η_t and outage probability ζ. This trade-off is illustrated in Figures 12, 13, 14, and 15. The results suggest that only some combination of target link spectral efficiency η_t and outage probability ζ are possible. In fact, those combinations lie below the red line in Figures 14 and 15. From practical point of view, all combinations of target link spectral efficiency η_t and outage probability ζ below water-filling line are inferior to water-filling. It is because, the water-filling power control rule offers either higher target link spectral efficiency η_t for a given outage probability ζ or a lower outage probability ζ for a given target link spectral efficiency η_t. From practical point of view, a useful operating regions for pairs of η_t and ζ lies between water-filling boundary and our proposed transmission scheme boundary. The useful operating region lies between blue and red lines in Figures 16 and 17.

4.5.2 Risk Theory

In risk theory, one constructs a risk measure system where different random variables are ordered with respect to their risk measure V. In this work, we prefer the least risky transmission scheme, so our goal is to minimize the risk measure

$$V\left(\eta_r\right) = \alpha \mathrm{Var}\left[\eta_r\right] - \left(1-\alpha\right) E\left[\eta_r\right] \tag{138}$$

where $0 < \alpha \leq 1$ is some arbitrary constant. Specifically, we look for the adaptive power control scheme which implies the lowest possible risk.

To minimize the risk measure V (138), we need to obtain a power control rule which minimizes the variance $\mathrm{Var}\left[\eta_r\right]$ subject to the average link spectral efficiency constraint

$$E\left[\eta_r\right] = \mu_r(\boldsymbol{\lambda}, \boldsymbol{q}) = r_{\mathrm{av}}, \tag{139}$$

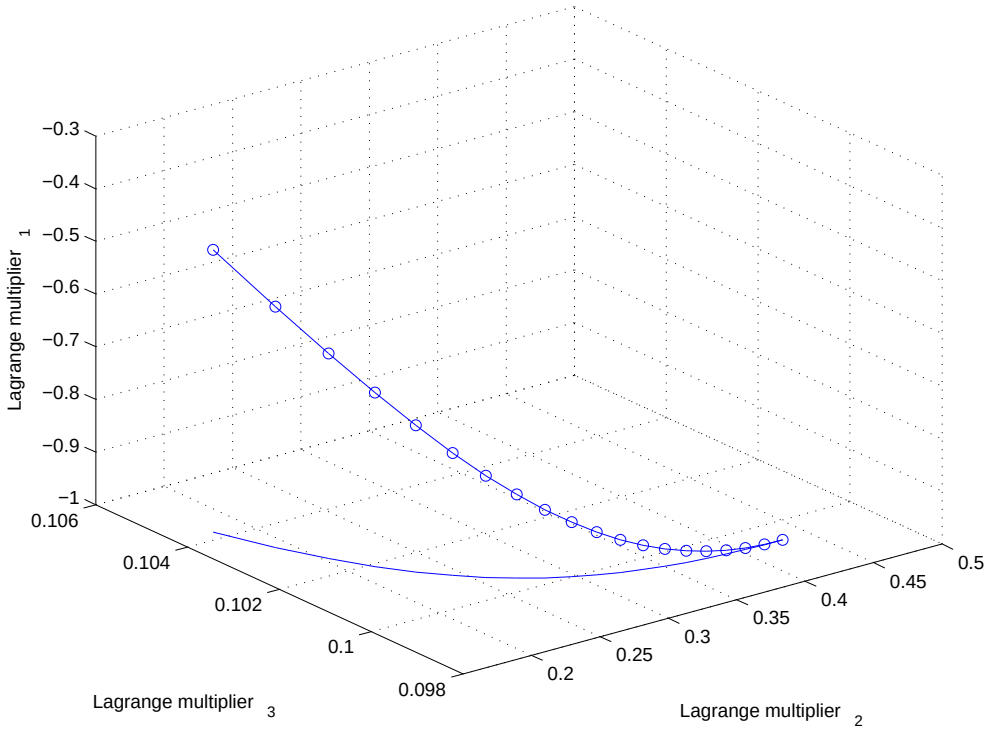

Figure 8: Relationship between Lagrange multipliers with Telser's safety-first rule in a SISO Rayleigh fading channel.

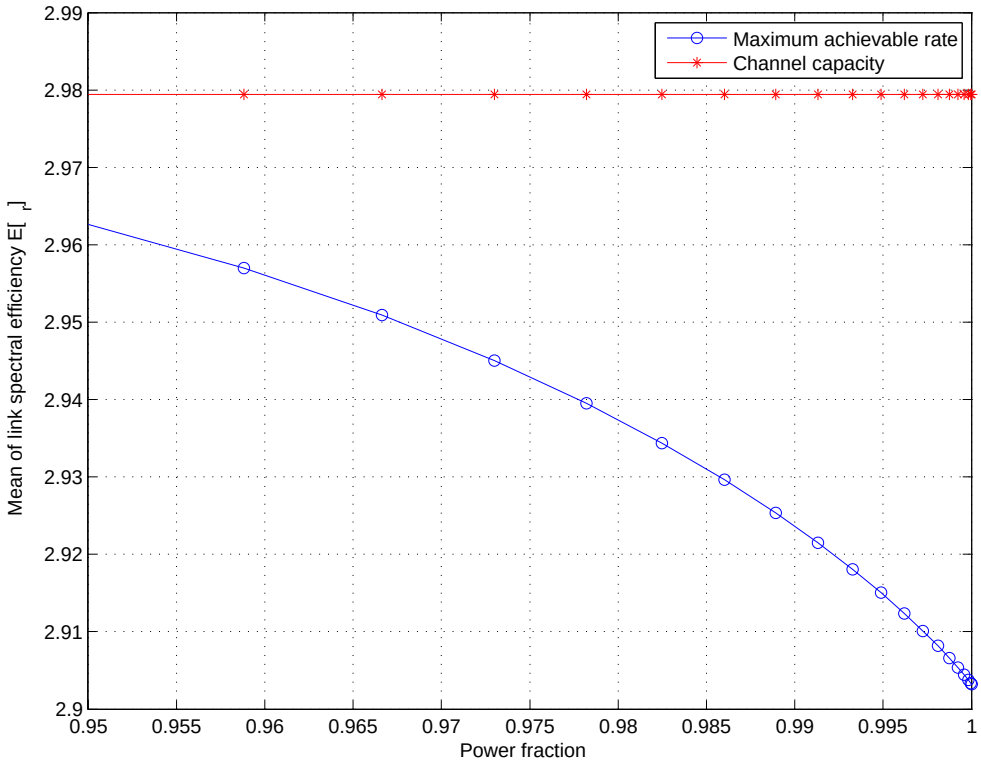

Figure 9: Illustration of the effect of power division between regions R and T on the mean link spectral efficiency.

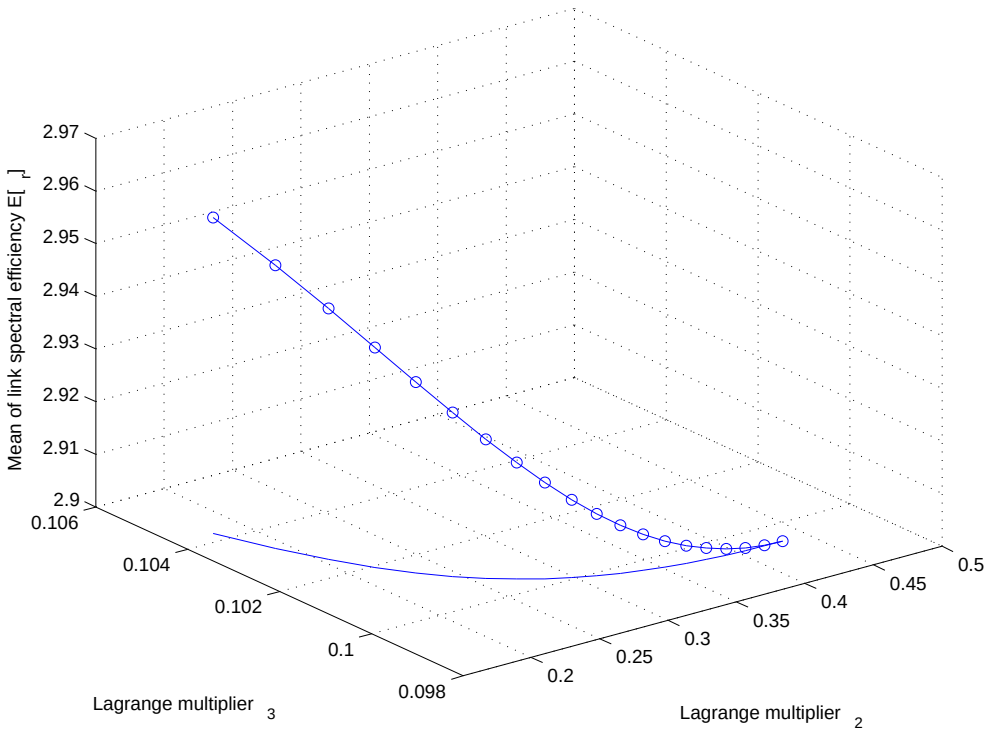

Figure 10: Relationship between the second and third Lagrange multipliers and mean link spectral efficiency with Telser's safety-first rule in a SISO Rayleigh fading channel.

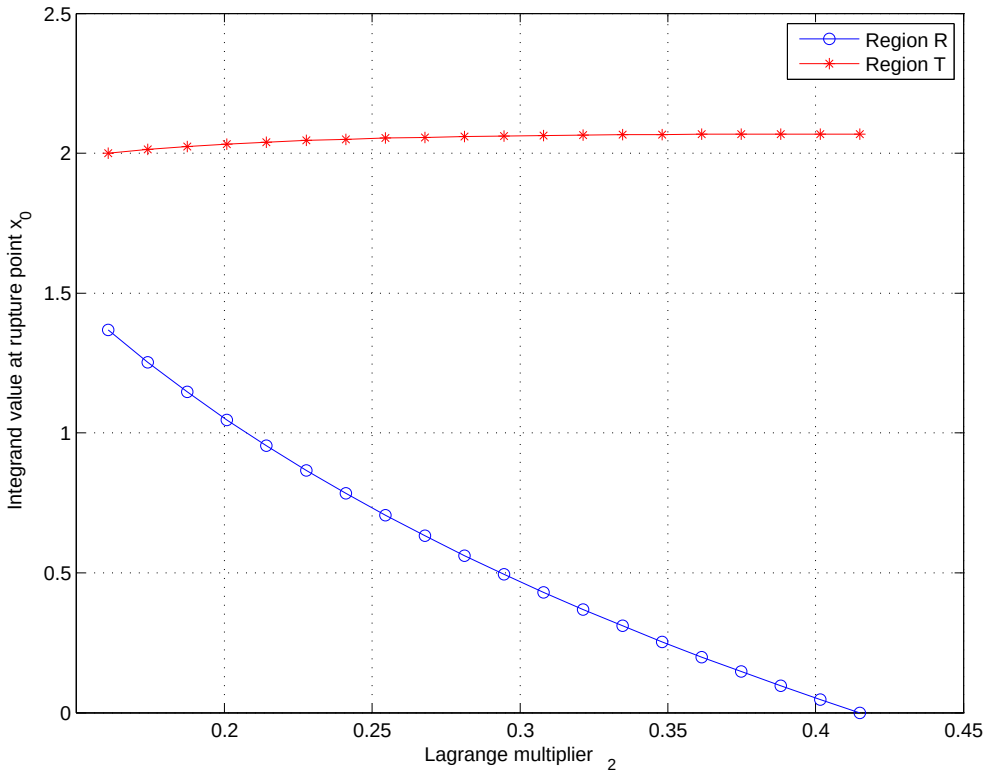

Figure 11: Illustration of the discontinuity of transmission link spectral efficiency on the boundary of regions R and T ($\eta_t = 2, \zeta = 0.25$).

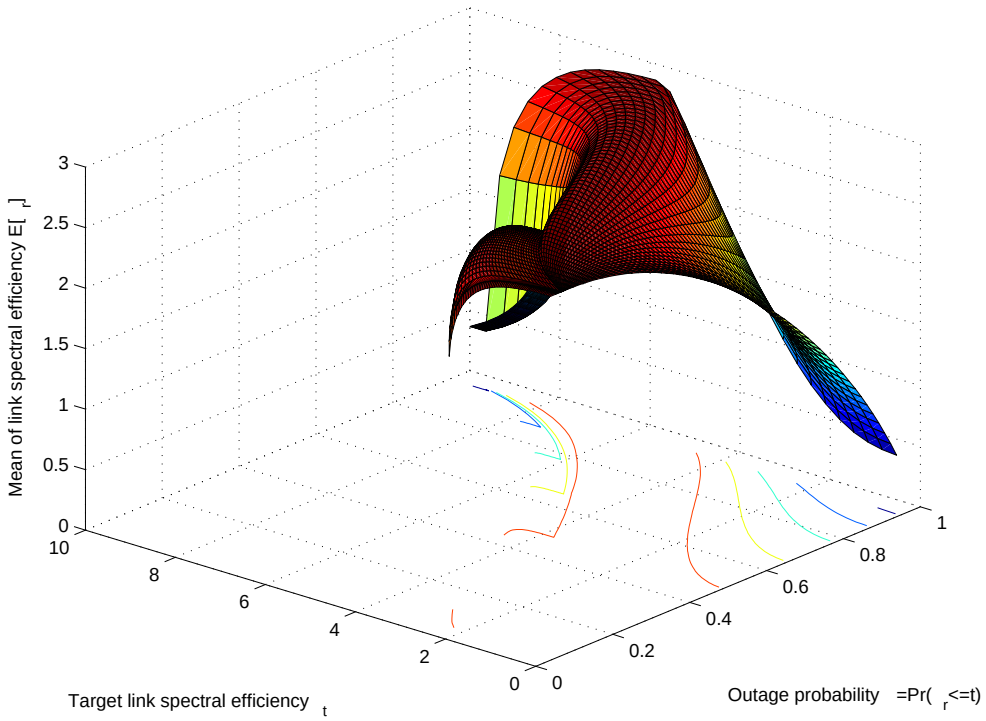

Figure 12: Mean link spectral efficiency as a function of target link spectral efficiency η_t and outage probability ζ in a SISO channel.

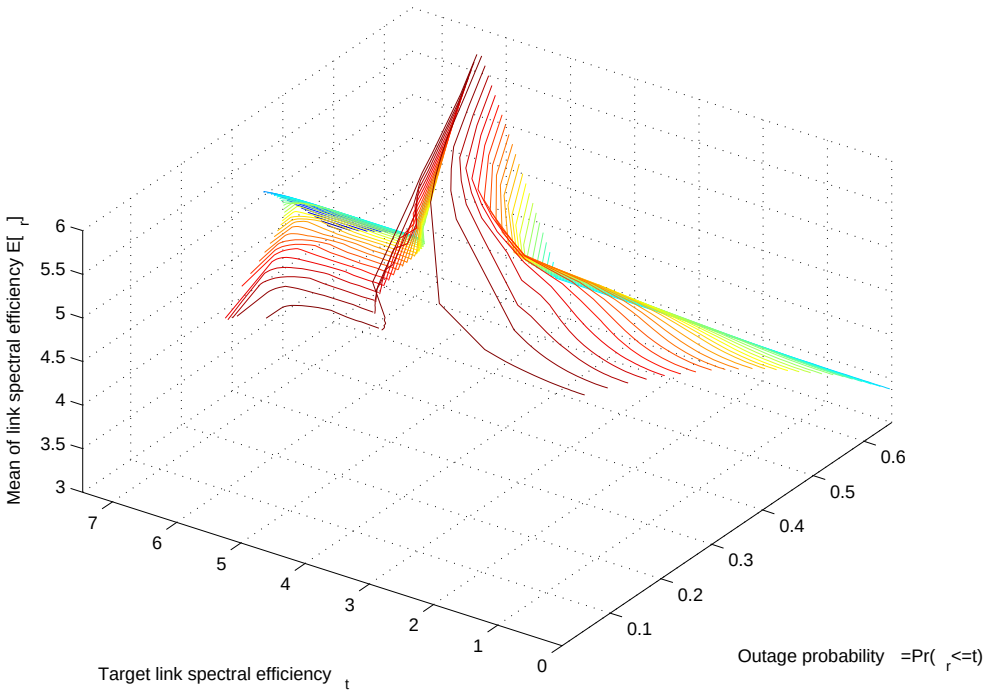

Figure 13: Mean link spectral efficiency as a function of target link spectral efficiency η_t and outage probability ζ in a MIMO channel.

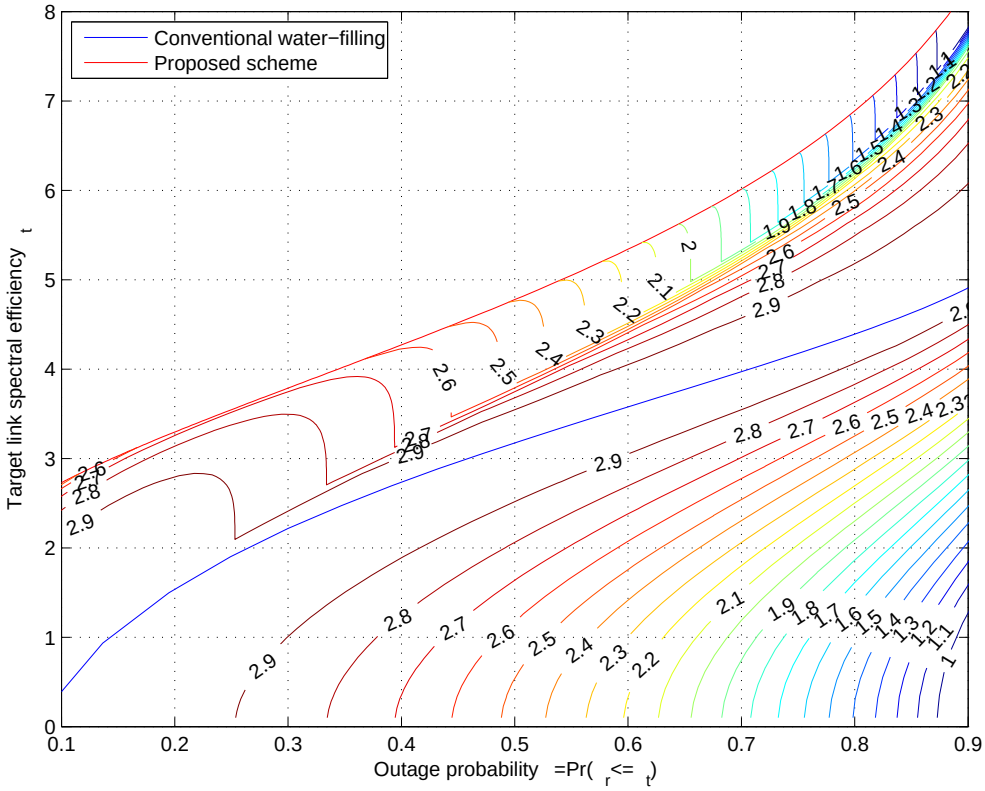

Figure 14: Contour plot of achievable mean link spectral efficiency as a function of target link spectral efficiency η_t and outage probability ζ in a SISO channel.

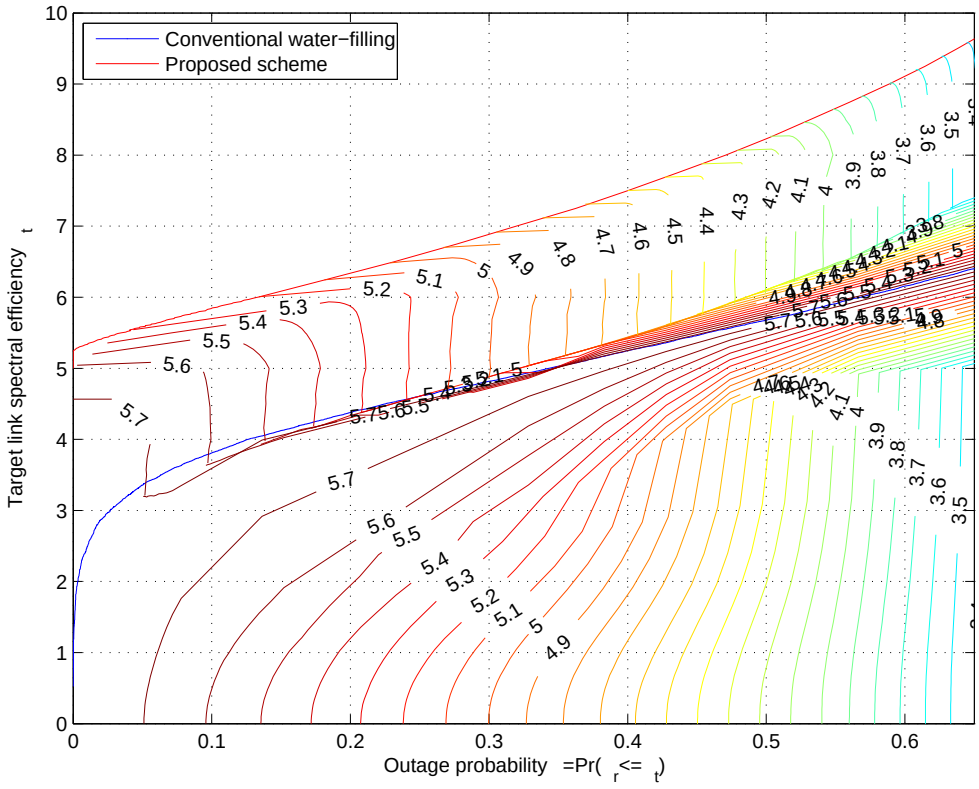

Figure 15: Contour plot of achievable mean link spectral efficiency as a function of target link spectral efficiency η_t and outage probability ζ in a MIMO channel.

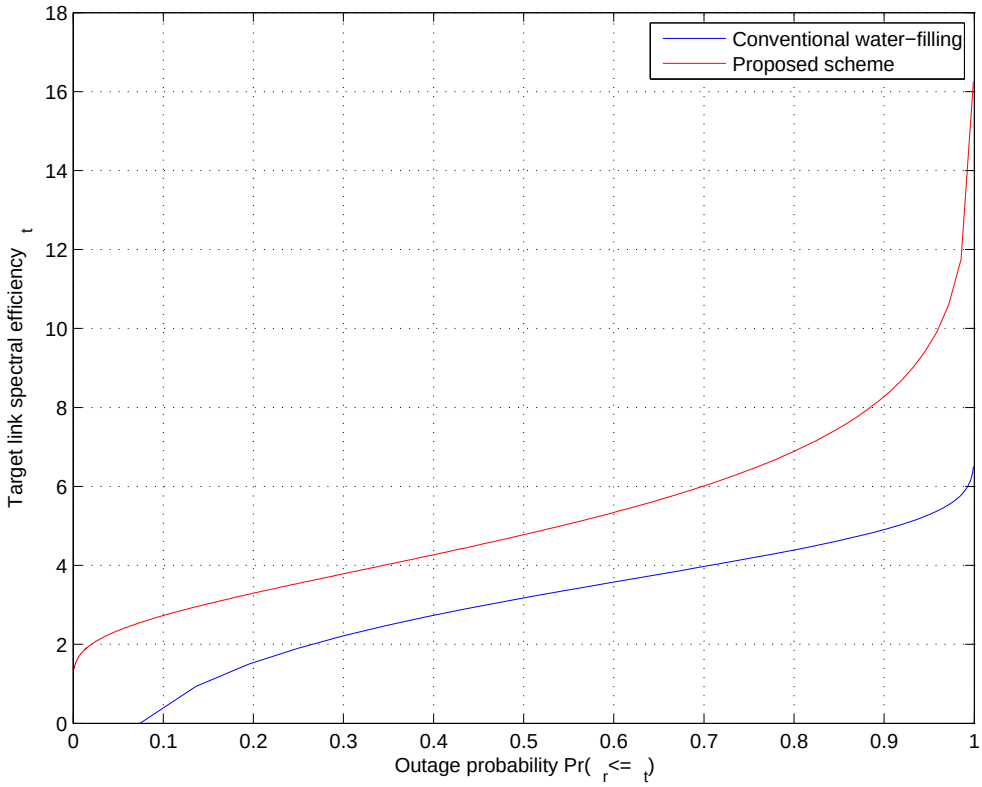

Figure 16: Practical operating region of adaptive power control rule based on Telser's safety-first approach in a SISO channel.

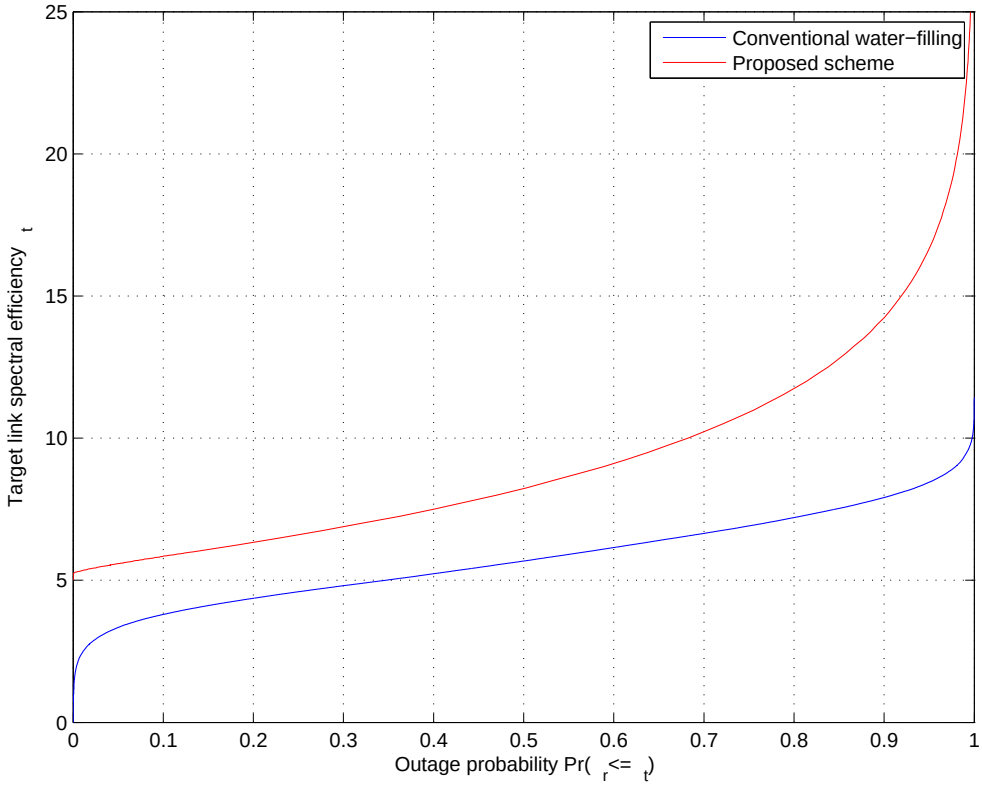

Figure 17: Practical operating region of adaptive power control rule based on Telser's safety-first approach in a MIMO channel.

the average power constraint (8), and the physical constraint that the transmission power cannot be negative, i.e.,

$$\min\left\{\mathrm{Var}\left[\eta_r\right]|\mu_r = r_{\mathrm{av}}, \mu_q = P_{\mathrm{av}}, \boldsymbol{q} \in \mathbb{R}_+^m\right\}. \tag{140}$$

The optimization problem (140) can be formulated as the variational problem with the variance as the variational integral and the constraints (8) and (139) as subsidiary conditions [12, p. 45]. However, we have additional inequality constraints, i.e., $q_k \geq 0$ for all $k = 1, 2, \ldots, m$. Since the inequality constraints cannot be easily added to variational problems, we introduce a new vector-valued function $\boldsymbol{\rho} : \mathbb{R}_+^m \to \mathbb{R}_+^m$ and reformulate the optimization problem accordingly. Namely, we assume that $q_k = \rho_k^2$ which ensures that $\boldsymbol{q} \in \mathbb{R}_+^m$. The optimization problem (157) can therefore be stated as

$$\min\left\{J[\boldsymbol{\rho}] = \mathrm{Var}\left[\eta_r\right]|K_1[\boldsymbol{\rho}] = r_{\mathrm{av}}, K_2[\boldsymbol{\rho}] = P_{\mathrm{av}}\right\} \tag{141}$$

where

$$J\left[\boldsymbol{\rho}\right] = \int_0^{\infty}\int_0^{\lambda_1}\cdots\int_0^{\lambda_{m-1}} F(\boldsymbol{\lambda}, \boldsymbol{\rho})\, d\lambda_m \cdots d\lambda_2 d\lambda_1 \tag{142}$$

$$K_1\left[\boldsymbol{\rho}\right] = \int_0^{\infty}\int_0^{\lambda_1}\cdots\int_0^{\lambda_{m-1}} G_1(\boldsymbol{\lambda}, \boldsymbol{\rho})\, d\lambda_m \cdots d\lambda_2 d\lambda_1 \tag{143}$$

$$K_2\left[\boldsymbol{\rho}\right] = \int_0^{\infty}\int_0^{\lambda_1}\cdots\int_0^{\lambda_{m-1}} G_2(\boldsymbol{\lambda}, \boldsymbol{\rho})\, d\lambda_m \cdots d\lambda_2 d\lambda_1 \tag{144}$$

are integrals over the fixed region Q. The respective variational integrands in (142), (143), and (144) are given by

$$F(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \left[\sum_{k=1}^{m} \log_2\left(1 + \frac{\rho_k^2 \lambda_k}{\sigma_n^2}\right) - r_{\mathrm{av}}\right]^2 p(\boldsymbol{\lambda}) \tag{145}$$

$$G_1(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \sum_{k=1}^{m} \log_2\left(1 + \frac{\rho_k^2 \lambda_k}{\sigma_n^2}\right) p(\boldsymbol{\lambda}) \tag{146}$$

$$G_2(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \sum_{k=1}^{m} \rho_k^2 p(\boldsymbol{\lambda}) \tag{147}$$

A necessary condition for the vector-valued function $\boldsymbol{\rho}$ to be an extremal of $J\left[\boldsymbol{\rho}\right]$ is that the function components $\rho_k = \rho_k(\lambda_k)$ satisfy Euler-Lagrange equations, i.e., [12, p. 152]

$$\frac{\partial}{\partial \rho_k} L(\boldsymbol{\lambda}, \boldsymbol{\rho}) = 0, \quad k = 1, 2, \ldots, m \tag{148}$$

where

$$L(\boldsymbol{\lambda}, \boldsymbol{\rho}) = F(\boldsymbol{\lambda}, \boldsymbol{\rho}) + \xi_1 G_1(\boldsymbol{\lambda}, \boldsymbol{\rho}) + \xi_2 G_2(\boldsymbol{\lambda}, \boldsymbol{\rho}) \tag{149}$$

and where $\xi_1 \neq 0$ and $\xi_2 \neq 0$ are two constants, also known as Lagrange multipliers, which are independent of $\boldsymbol{\lambda}$ [12, p. 45].

Proposition 6. *The transmission power level in the kth subchannel, that is, the kth component of $\boldsymbol{q}$ is*

$$q_k = \sigma_n^2\left[\theta_R^{-1}(w, \boldsymbol{\lambda}) - \lambda_k^{-1}\right]_+, \quad k = 1, 2, \ldots, m \tag{150}$$

where

$$\theta_R(w, \boldsymbol{\lambda}) = \frac{\sigma_n^2 \left(\ln 2\right)^2}{2w\xi_2^{-1}} \left\{W\left[\frac{\sigma_n^2 \left(\ln 2\right)^2}{2w\xi_2^{-1}} \sqrt[w]{\frac{2^{(2r_{\mathrm{av}} - \xi_1)/2}}{\prod_{k=1}^{w} \lambda_k}}\right]\right\}^{-1} \tag{151}$$

with $\xi_2 > 0$, and where w is a unique integer in $\{1, 2, \ldots, m\}$ such that $\lambda_w > \theta_R(w, \boldsymbol{\lambda}) \geq \lambda_{w+1}$. The symbol $W(z)$ in (151) denotes the principal branch of the Lambert W function [7].

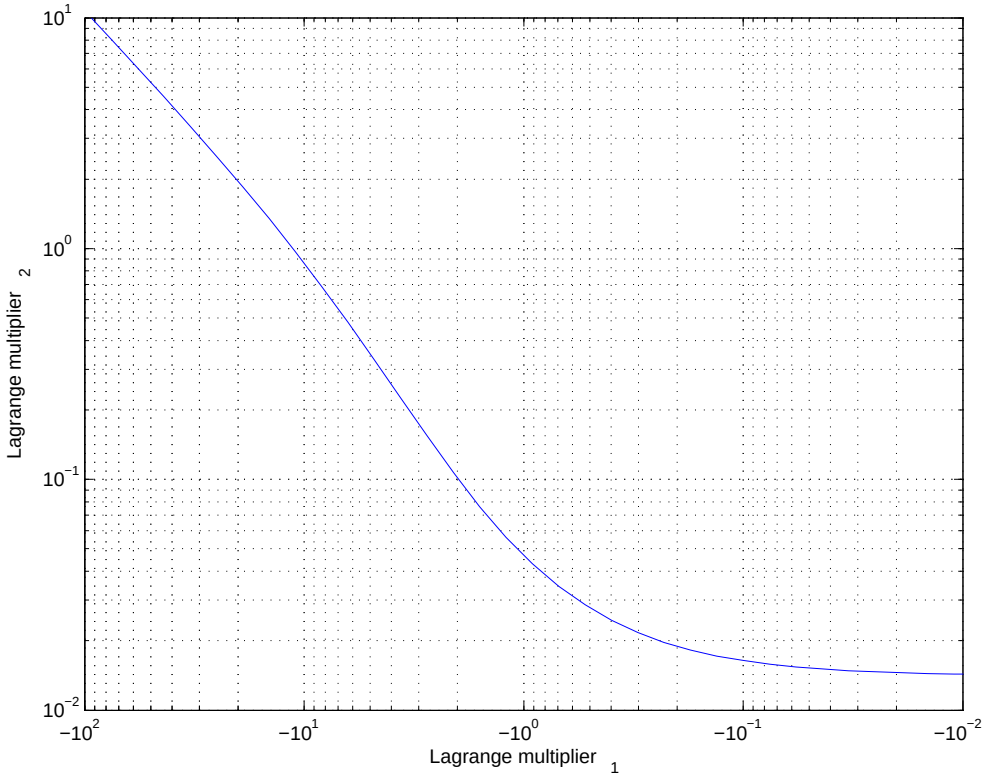

Figure 18: The optimal pairs of Lagrange multipliers in a SISO channel.

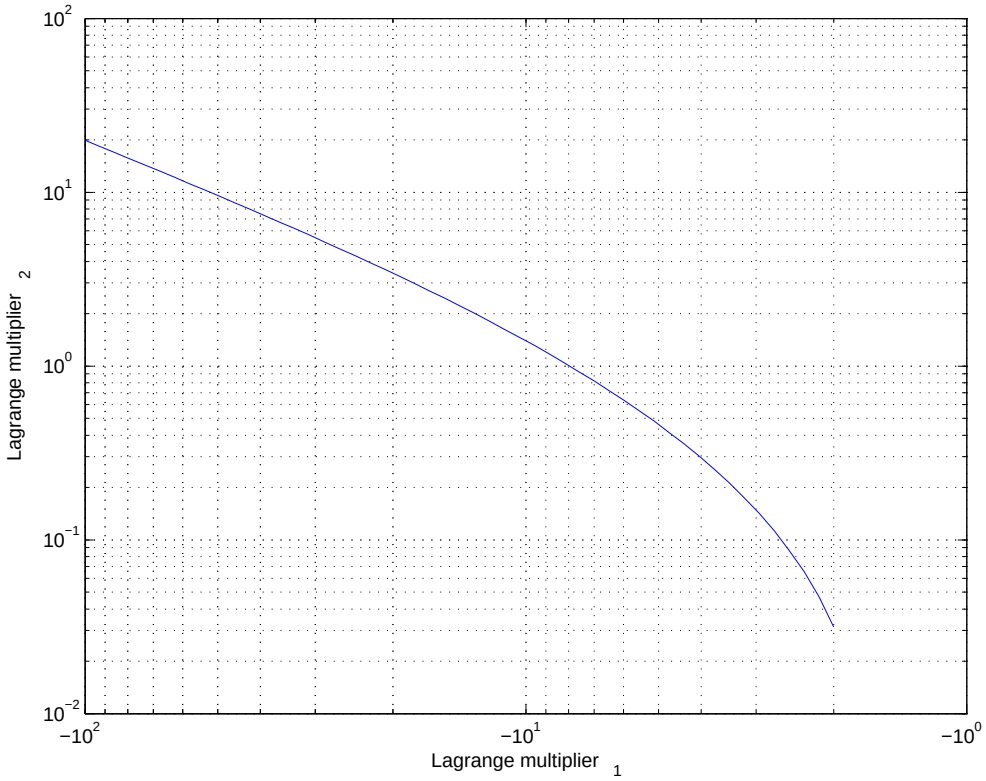

Figure 19: The optimal pairs of Lagrange multipliers in a MIMO channel.

Proof. Deferred to Appendix. □

We illustrate the performance of the adaptive power control algorithm in a single-input single-output (SISO) Rayleigh fading channel as well as multiple-input multiple-output (MIMO) Rayleigh fading channel with two transmitter and two receiver antennas. The adaptive power control algorithm is designed based on the theory of risk. In both cases the average transmitted signal-to-noise ratio is set to 10 dB, that is, $P_{\text{av}}/\sigma_n^2 = 10$ dB. The relationship between Lagrange multipliers is presented in Figures 18 and 19 for a SISO and MIMO channel, respectively. In both cases, the first Lagrange multiplier ξ_1 is negative wheras the second Lagrange multiplier ξ_2 is positive. Since

$$\xi_1 \approx -\frac{\Delta J}{\Delta K_1} = -\frac{\Delta \text{Var}\left[\eta_r\right]}{\Delta r_{\text{av}}} < 0 \tag{152}$$

and

$$\xi_2 \approx -\frac{\Delta J}{\Delta K_2} = -\frac{\Delta \text{Var}\left[\eta_r\right]}{\Delta P_{\text{av}}} > 0, \tag{153}$$

we can conclude that the variance of link spectral efficiency $\text{Var}\left[\eta_r\right]$ will increase when we decrease the average transmission power or increase the average transmission rate requirement. On the other hand, the variance will decrease when the average transmission power is increased or the average rate requirement is decreased. The symbols $\Delta\text{Var}\left[\eta_r\right]$, Δr_{av}, and ΔP_{av} denote arbitrarily small changes in variance of link spectral efficiency, average transmission rate, and average power constraint, respectively. These changes of link spectral efficiency variance with respect to average transmission rate are shown in Figures 20 and 21. The optimal combinations of mean link spectral efficiency $\text{E}\left[\eta_r\right]$ and variance of link spectral efficiency $\text{Var}\left[\eta_r\right]$, shown in Figures 20 and 21, form an efficient frontier. The efficient frontier is a fundamental limit in the sense that it is not possible to construct a transmission scheme with combinations of mean $\text{E}\left[\eta_r\right]$ and variance of link spectral efficiency $\text{Var}\left[\eta_r\right]$ which lie above the frontier in Figures 20 and 21. All combinations of mean and variance of link spectral efficiency which lie below the frontier are suboptimal. It is because there is other combination of mean $\text{E}\left[\eta_r\right]$ and variance $\text{Var}\left[\eta_r\right]$ which offers lower variance for the same mean or higher mean for the same variance. In particular, this combination lies on the efficient frontier.

In risk theory, we are primarily interested in a risk index (138) and we want to minimize it subject to some preselected value of α. We plot the value of risk index $V\left[\eta_r\right]$ for different values of α in Figures 22 and 23. The value of risk $V\left[\eta_r\right]$ index is normalized such that its minimum value equals zero. It was achieved by adding a constant $c_V = \text{E}\left[\eta_r\right]$ to $V\left[\eta_r\right]$. Note that addition of constant value does not change the ordering of power control schemes given by (17) in any way. The minimum values of risk index $V\left[\eta_r\right]$ for respective values of α are denoted by asterisk. The results suggest that for every value of α a different combination of mean and variance from the efficient frontier is the optimal one. As expected, when α approaches zero, schemes with higher mean value are preferred. On the other hand, when α approaches one, schemes with lower variance are preferred. The selection of α offers flexibility in selecting the optimal transmission scheme. In practice, the selection of a specific combination of mean $\text{E}\left[\eta_r\right]$ and variance of link spectral efficiency $\text{Var}\left[\eta_r\right]$ is achieved by properly selecting one of Lagrange multipliers. We illustrate this in Figures 24 and 25 where we plot the mean of link spectral efficiency and the corresponding variance of link spectral efficiency for a given value of the second Lagrange multiplier ξ_2. By changing the value of ξ_2 one moves along the efficient frontier.

4.5.3 Expected Utility Hypothesis

In this work, we propose to use the following utility function

$$u\left(\eta_r\right) = \begin{cases} \eta_r, & \text{for } \eta_r \geq \eta_t \\ \eta_r - \beta\left(\eta_t - \eta_r\right)^2, & \text{for } \eta_r < \eta_t \end{cases} \tag{154}$$

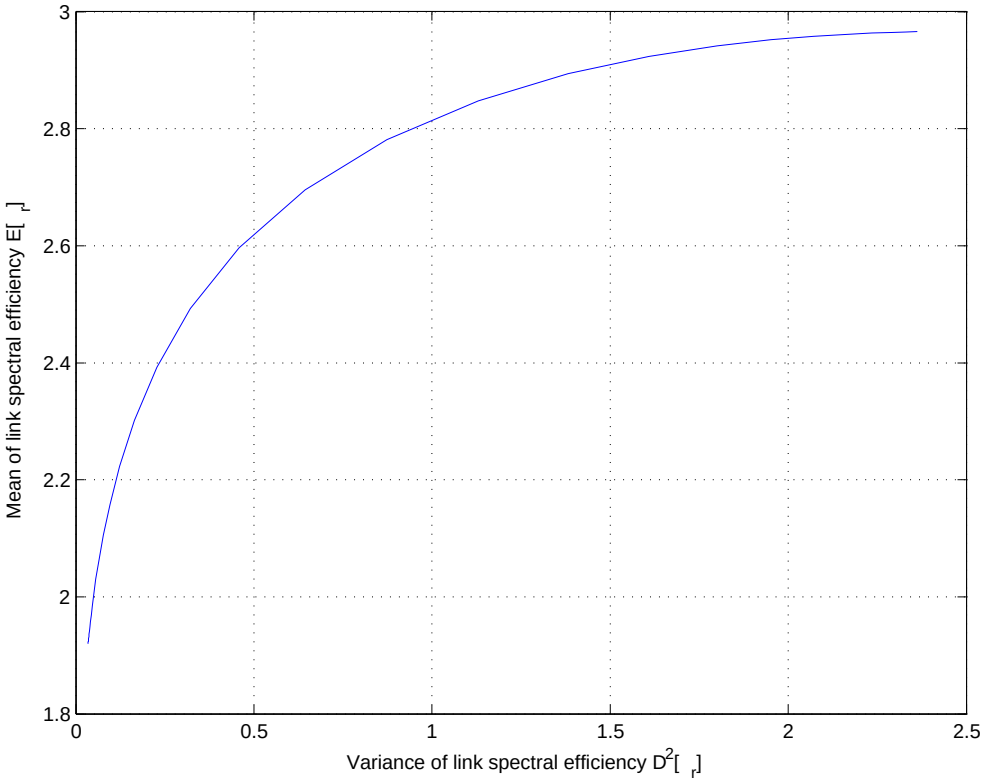

Figure 20: Efficient frontier of power control rule inspired by Pollatsek and Tversky theory of risk (SISO channel).

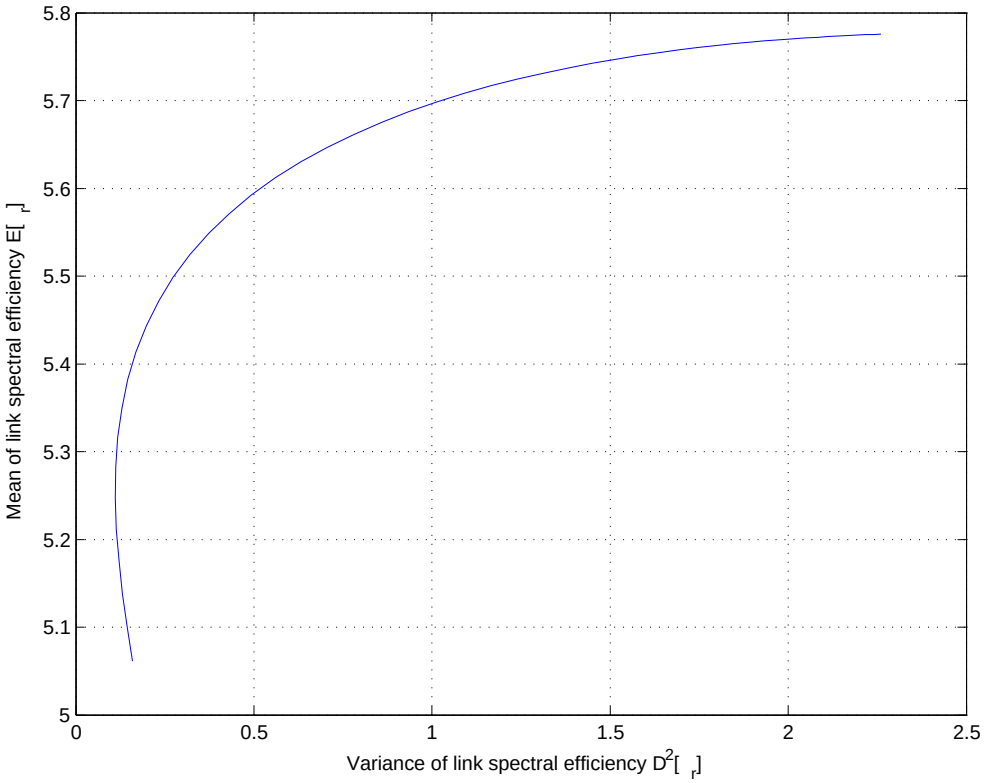

Figure 21: Efficient frontier of power control rule inspired by Pollatsek and Tversky theory of risk (MIMO channel).

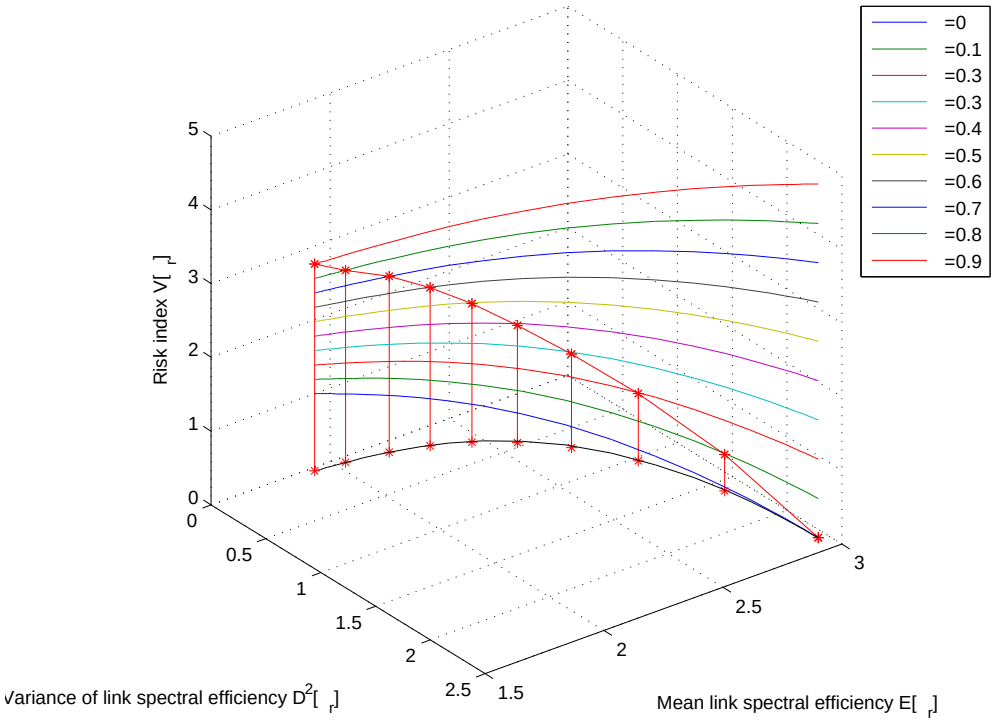

Figure 22: Risk index V as a function of parameter α in a SISO channel.

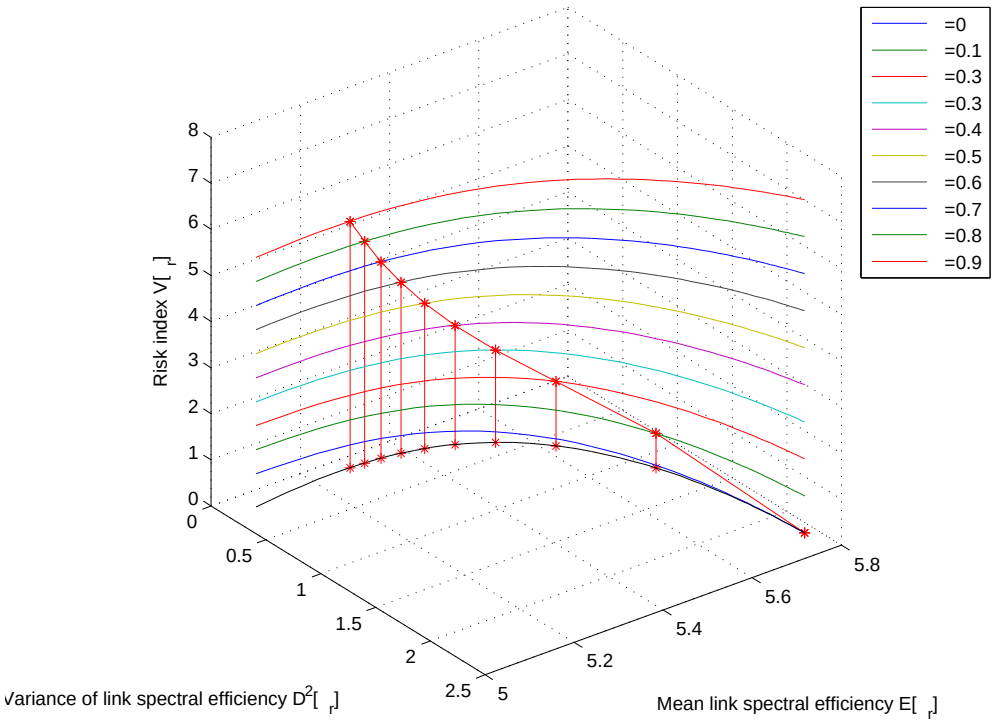

Figure 23: Risk index V as a function of parameter α in a MIMO channel.

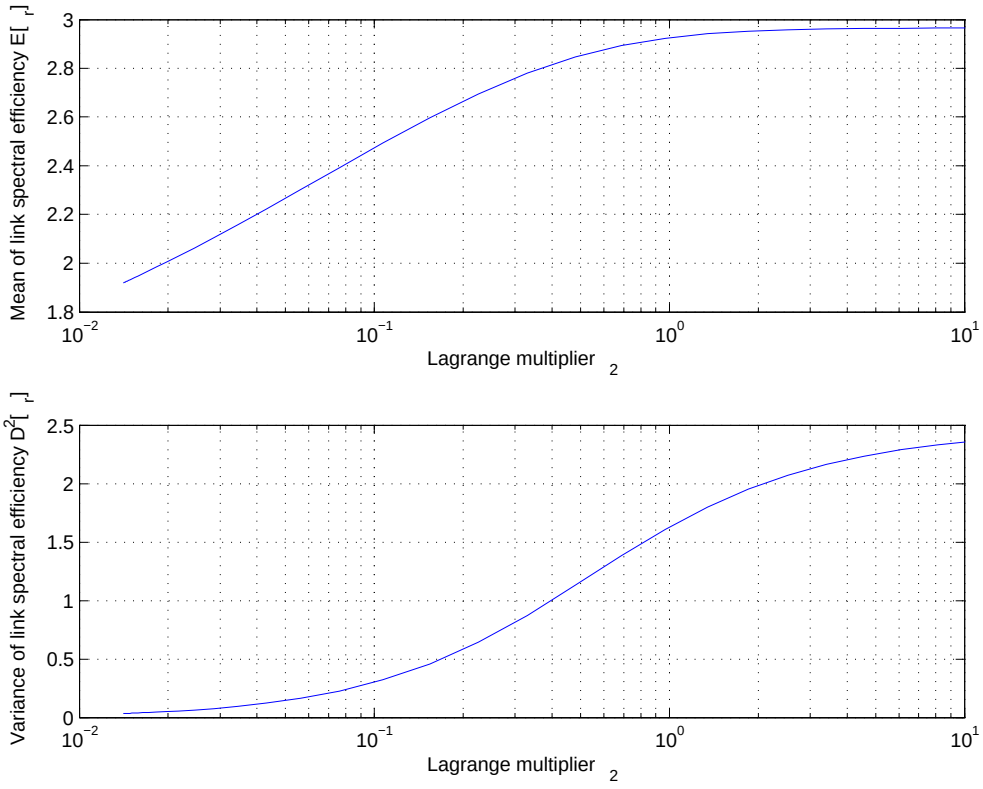

Figure 24: Mean and variance of link spectral efficiency as functions of ξ_2 in a SISO channel.

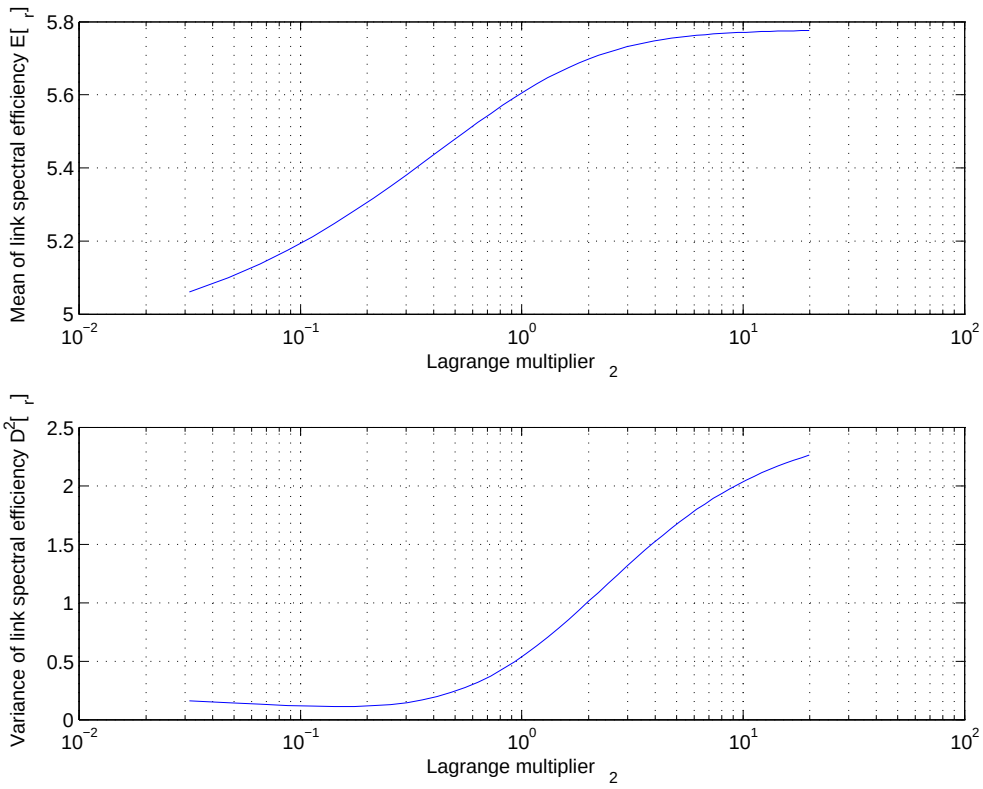

Figure 25: Mean and variance of link spectral efficiency as functions of ξ_2 in a SISO channel.

to rank the random distributions of achievable link spectral efficiency η_r. The constant β in (154) is an arbitrary positive constant $\beta > 0$. It can be easily verified that the ordering function U is

$$U = \int_0^\infty u\left(\eta_r\right) p\left(\eta_r\right) d\eta_r = \underbrace{\int_0^\infty \eta_r p\left(\eta_r\right) d\eta_r}_{E[\eta_r]} - \underbrace{\int_0^{\eta_t} \left(\eta_t - \eta_r\right)^2 p\left(\eta_r\right) d\eta_r}_{l_2^-(\eta_t)}. \tag{155}$$

Consequently, to maximize the expected utility U (155), we need to obtain a power control rule which minimizes the below-target semivariance $l_2^-(\eta_t)$ subject to the average link spectral efficiency constraint

$$E\left[\eta_r\right] = \mu_r(\boldsymbol{\lambda}, \boldsymbol{q}) = r_{\text{av}}, \tag{156}$$

the average power constraint (8), and the physical constraint that the transmission power cannot be negative, i.e.,

$$\min\left\{l_2^-\left(\eta_t\right) | \mu_r = r_{\text{av}}, \mu_q = P_{\text{av}}, \boldsymbol{q} \in \mathbb{R}_+^m\right\} \tag{157}$$

for some target link spectral efficiency η_t.

The optimization problem (157) can be formulated as the variational problem with the below-target semivariance (20) as the variational integral and the constraints (8) and (156) as subsidiary conditions [12, p. 45]. However, we have additional inequality constraints, i.e., $q_k \geq 0$ for all $k = 1, 2, \ldots, m$. Since the inequality constraints cannot be easily added to variational problems, we introduce a new vector-valued function $\boldsymbol{\rho} : \mathbb{R}_+^m \to \mathbb{R}_+^m$ and reformulate the optimization problem accordingly. Namely, we assume that $q_k = \rho_k^2$ which ensures that $\boldsymbol{q} \in \mathbb{R}_+^m$. The optimization problem (157) can therefore be stated as

$$\min\left\{J[\boldsymbol{\rho}] = l_2^-\left(\eta_t\right) | K_1[\boldsymbol{\rho}] = r_{\text{av}}, K_2[\boldsymbol{\rho}] = P_{\text{av}}\right\} \tag{158}$$

where

$$J\left[\boldsymbol{\rho}\right] = \int_0^\infty \int_0^{\lambda_1} \cdots \int_0^{\lambda_{m-1}} F(\boldsymbol{\lambda}, \boldsymbol{\rho})\, d\lambda_m \cdots d\lambda_2 d\lambda_1 \tag{159}$$

$$K_1\left[\boldsymbol{\rho}\right] = \int_0^\infty \int_0^{\lambda_1} \cdots \int_0^{\lambda_{m-1}} G_1(\boldsymbol{\lambda}, \boldsymbol{\rho})\, d\lambda_m \cdots d\lambda_2 d\lambda_1 \tag{160}$$

$$K_2\left[\boldsymbol{\rho}\right] = \int_0^\infty \int_0^{\lambda_1} \cdots \int_0^{\lambda_{m-1}} G_2(\boldsymbol{\lambda}, \boldsymbol{\rho})\, d\lambda_m \cdots d\lambda_2 d\lambda_1 \tag{161}$$

are integrals over the fixed region Q. The respective variational integrands in (159), (160), and (161) are given by

$$F(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \left\{\left[t - \sum_{k=1}^m \log_2\left(1 + \frac{\rho_k^2 \lambda_k}{\sigma_n^2}\right)\right]_+\right\}^2 p(\boldsymbol{\lambda}) \tag{162}$$

$$G_1(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \sum_{k=1}^m \log_2\left(1 + \frac{\rho_k^2 \lambda_k}{\sigma_n^2}\right) p(\boldsymbol{\lambda}) \tag{163}$$

$$G_2(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \sum_{k=1}^m \rho_k^2 p(\boldsymbol{\lambda}) \tag{164}$$

A necessary condition for the vector-valued function $\boldsymbol{\rho}$ to be an extremal of $J\left[\boldsymbol{\rho}\right]$ is that the function components $\rho_k = \rho_k(\lambda_k)$ satisfy Euler-Lagrange equations, i.e., [12, p. 152]

$$\frac{\partial}{\partial \rho_k} L(\boldsymbol{\lambda}, \boldsymbol{\rho}) = 0, \quad k = 1, 2, \ldots, m \tag{165}$$

where

$$L(\boldsymbol{\lambda},\boldsymbol{\rho}) = F(\boldsymbol{\lambda},\boldsymbol{\rho}) + \xi_1 G_1(\boldsymbol{\lambda},\boldsymbol{\rho}) + \xi_2 G_2(\boldsymbol{\lambda},\boldsymbol{\rho}) \tag{166}$$

and where $\xi_1 \neq 0$ and $\xi_2 \neq 0$ are two constants, also known as Lagrange multipliers, which are independent of $\boldsymbol{\lambda}$ [12, p. 45].

We observe that some first order partial derivatives $\partial F/\partial\rho_k$ may not exist when $\eta_r = \eta_t$ because the function $f(z) = (z)_+$ is not differentiable at $z = 0$. Hence, we assume that the solution to (158) is given by continuous piecewise smooth functions ρ_k [12, p. 61]. Consequently, the functions ρ_k are defined separately in the region $T = \{\boldsymbol{\lambda} \in Q : \eta_r(\boldsymbol{\lambda},\boldsymbol{\rho}) \geq \eta_t\}$ and its complement $R = \{\boldsymbol{\lambda} \in Q : \eta_r(\boldsymbol{\lambda},\boldsymbol{\rho}) < \eta_t\}$. Furthermore, on the boundary surface $S = \{\boldsymbol{\lambda} \in Q : \eta_r(\boldsymbol{\lambda},\boldsymbol{\rho}) = \eta_t\} \subset T$ the functions ρ_k must satisfy the Weierstrass-Erdmann conditions [12, p. 63]

$$\lim_{\eta_r \to \eta_t^-} \frac{\partial L}{\partial \rho_k'} = \lim_{\eta_r \to \eta_t^+} \frac{\partial L}{\partial \rho_k'} \tag{167}$$

$$\lim_{\eta_r \to \eta_t^-} \left(L - \rho_k' \frac{\partial L}{\partial \rho_k'} \right) = \lim_{\eta_r \to \eta_t^+} \left(L - \rho_k' \frac{\partial L}{\partial \rho_k'} \right). \tag{168}$$

The symbol ρ_k' in (167) and (168) denotes the first order derivative of ρ_k with respect to λ_k, i.e., $\rho_k' = \partial\rho_k/\partial\lambda_k$.

Proposition 7. *If $\boldsymbol{\lambda} \in T$, then the kth component of $\boldsymbol{q}$ is*

$$q_k = \sigma_n^2 \left(\theta_T^{-1} - \lambda_k^{-1}\right)_+, \quad k = 1, 2, \ldots, m \tag{169}$$

where

$$\theta_T = -\sigma_n^2 \xi_1^{-1} \xi_2 \ln 2 > 0. \tag{170}$$

Proof. Deferred to Appendix. □

Proposition 8. *If $\boldsymbol{\lambda} \in R$, then the kth component of $\boldsymbol{q}$ is*

$$q_k = \sigma_n^2 \left[\theta_R^{-1}(w,\boldsymbol{\lambda}) - \lambda_k^{-1}\right]_+, \quad k = 1, 2, \ldots, m \tag{171}$$

where

$$\theta_R(w,\boldsymbol{\lambda}) = \frac{\sigma_n^2 (\ln 2)^2}{2w\xi_2^{-1}} \left\{ W\left[\frac{\sigma_n^2 (\ln 2)^2}{2w\xi_2^{-1}} \sqrt[w]{\frac{2^{(2\eta_t - \xi_1)/2}}{\prod_{k=1}^w \lambda_k}} \right] \right\}^{-1} \tag{172}$$

with $\xi_2 > 0$, and where w is a unique integer in $\{1, 2, \ldots, m\}$ such that $\lambda_w > \theta_R(w,\boldsymbol{\lambda}) \geq \lambda_{w+1}$. The symbol $W(z)$ in (172) denotes the principal branch of the Lambert W function [7].

Proof. Deferred to Appendix. □

Remark 3. *By Proposition 5, Lagrange multipliers ξ_1 and ξ_2 have different signs, i.e., $\xi_1\xi_2 < 0$. Since $\xi_2 > 0$ by Proposition 6, we conclude that $\xi_1 < 0$ and $\xi_2 > 0$.*

Remark 4. *The achievable link spectral efficiency $\eta_r(\boldsymbol{\lambda},\boldsymbol{q})$ in the region T depends on $\boldsymbol{\lambda}$ only through a constant θ_T. This allows us to describe the regions T and R in a simple way. Namely, let us assume that $\eta_r(\boldsymbol{\lambda},\boldsymbol{q}) \geq \eta_t$ is satisfied with transmission in exactly w subchannels, i.e., in the region*

$$V_w = \left\{ \boldsymbol{\lambda} \in Q : \sum\nolimits_{k=1}^{w} \log_2\left(1 + q_k\lambda_k/\sigma_n^2\right) \geq \eta_t \right\} \tag{173}$$

where $w = 1, 2, \ldots, m$. Substituting (169) into (173) yields

$$V_w = \left\{ \boldsymbol{\lambda} \in Q : \prod\nolimits_{k=1}^{w} \lambda_k \geq 2^{\eta_t} \left(\theta_T\right)^w \right\}. \tag{174}$$

Since $\eta_r(\boldsymbol{\lambda},\boldsymbol{q}) \geq \eta_t$ can be satisfied with any value of w, the region T must be a union of regions V_w, i.e., $T = \bigcup_{w=1}^m V_w$. Consequently, the region $R = Q\backslash T$, must be an intersection of the regions $U_w = Q\backslash V_w$, i.e., $R = \bigcap_{w=1}^m U_w$.

Remark 5. *The regions T and R are divided into, respectively, v and $v+1$ subregions T_w and R_w given by*

$$T_w = \{\boldsymbol{\lambda} \in T : \lambda_w > \theta_T \geq \lambda_{w+1}\} \tag{175}$$
$$R_w = \{\boldsymbol{\lambda} \in R : \lambda_w > \theta_R(w, \boldsymbol{\lambda}) \geq \lambda_{w+1}\} \tag{176}$$

where $w = 1, 2, \ldots, m$. The additional region R_0 is given by

$$R_0 = \{\boldsymbol{\lambda} \in R : \lambda_1 \leq \theta_R(1, \boldsymbol{\lambda})\}. \tag{177}$$

Since the thresholds θ_T and θ_R exist and are unique, $\{T_1, \ldots, T_m\}$ and $\{R_0, \ldots, R_{v+1}\}$ are partitions of the respective regions T and R. If $\boldsymbol{\lambda} \in (R_w \cup T_w)$, then exactly w subchannels are used for transmission. On the other hand, if $\boldsymbol{\lambda} \in R_0$, then the transmission is suspended.

The boundary surface of T_i and T_{i+1} is given by $P_i = \{\boldsymbol{\lambda} \in T_i : \lambda_{i+1} = \theta_T\}$ where $i = 1, 2, \ldots, m-1$. On the other hand, the boundary surface of R_{i-1} and R_i is $O_i = \{\boldsymbol{\lambda} \in R_{i-1} : \lambda_i = \theta_R(i, \boldsymbol{\lambda})\}$ where $i = 1, 2, \ldots, m$. If $\boldsymbol{\lambda} \in O_i$, then solving the equality $\theta_R(i, \boldsymbol{\lambda}) = \lambda_i$ for λ_i yields

$$\theta_R(i, \boldsymbol{\lambda}) = \lambda_i = \begin{cases} \sigma_n^2 \xi_2 \left(2\eta_t - \xi_1\right)^{-1} \ln 2, & i = 1 \\ \theta_R(i-1, \boldsymbol{\lambda}), & i \geq 2 \end{cases} \tag{178}$$

which implies that $\theta_R(i, \boldsymbol{\lambda})$, and thus $\mathbf{q}$, is continuous on all boundary surfaces O_i. The set of equalities (178) is obtained by substituting λ_i into (172) and rearranging the terms such that the resulting equality has the form $\tilde{a} = W(\tilde{b})$. Then, using the identity $\tilde{a}e^{\tilde{a}} = \tilde{b}$ [7] and grouping the terms which contain λ_i on one side lead to the equation $a^z e^a = b$, whose solution is $a = \ln b$ if $z = 0$ and $a = zW(\sqrt[z]{b}/z)$ otherwise.

The regions R_w and T_w can be defined over the whole $\mathbb{R}_+^m$ by permuting the coordinate axes, e.g. $R_w \to R_w^{(\pi)}$. The respective regions R_w, $R_w^{(\pi)}$, T_w, and $T_w^{(\pi)}$ in a generic two-input two-output channel are shown in Figure 26.

Proposition 9. *The extremal $\mathbf{q}$ defined by (169) and (171) satisfies Weierstrass-Erdmann conditions (167) and (168) on the boundary surface S and is an absolute minimizer of $l_2^-(\eta_t)$.*

Proof. Deferred to Appendix. □

By using (175) and (176) and substituting the optimal power control rules (169) and (171) into (159), (160), and (161), we obtain

$$\begin{aligned} l_2^-(\eta_t) = &\sum_{w=1}^{m} \int_{R_w} \left\{ \eta_t - \sum_{k=1}^{w} \log_2 \left[\frac{\lambda_k}{\theta_R(w, \boldsymbol{\lambda})} \right] \right\}^2 p(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} \\ &+ \int_{R_0} \eta_t^2 p(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} \end{aligned} \tag{179}$$

$$\begin{aligned} \mu_r = &\sum_{w=1}^{m} \sum_{k=1}^{w} \left\{ \int_{R_w} \log_2 \left[\frac{\lambda_k}{\theta_R(w, \boldsymbol{\lambda})} \right] p(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} \right. \\ &\left. + \int_{T_w} \log_2 \left(\frac{\lambda_k}{\theta_T} \right) p(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} \right\} = r_{\text{av}} \end{aligned} \tag{180}$$

$$\begin{aligned} \mu_q = \sigma_n^2 &\sum_{w=1}^{m} \sum_{k=1}^{w} \left\{ \int_{R_w} \left[\frac{1}{\theta_R(w, \boldsymbol{\lambda})} - \frac{1}{\lambda_k} \right] p(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} \right. \\ &\left. + \int_{T_w} \left(\frac{1}{\theta_T} - \frac{1}{\lambda_k} \right) p(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} \right\} = P_{\text{av}} \end{aligned} \tag{181}$$

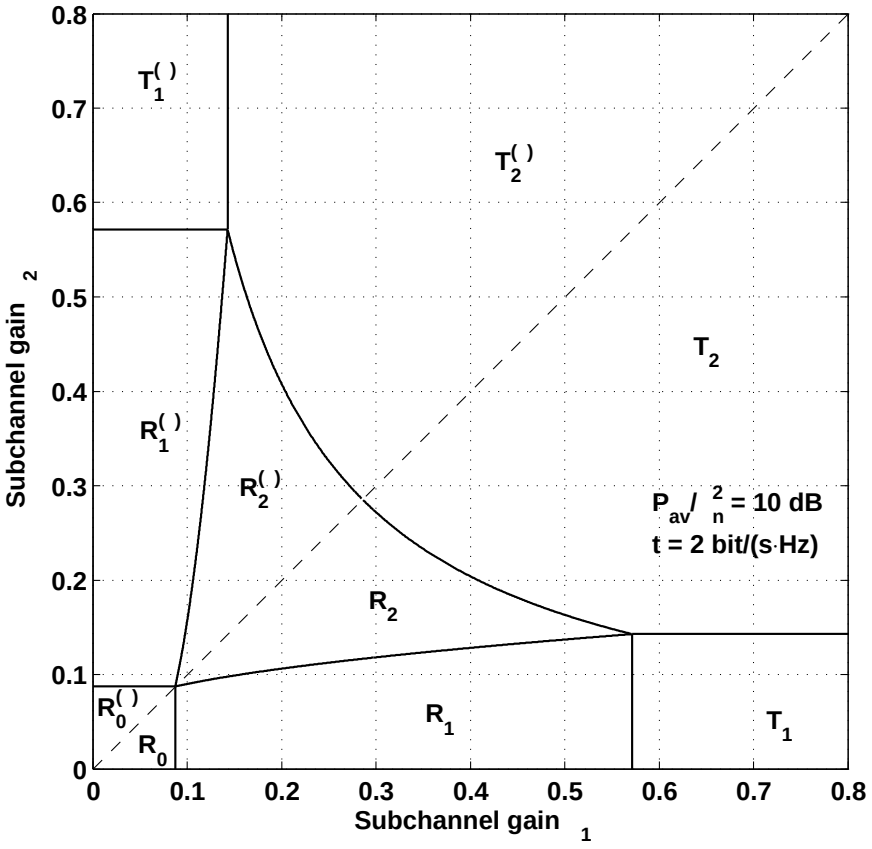

Figure 26: Illustration of respective transmission regions R and T.

where $d\boldsymbol{\lambda} = d\lambda_1 \cdots d\lambda_m$. In general, the multiple integrals in (179), (180), and (181) do not have analytical closed-form solutions and one has to resort to numerical Monte Carlo methods to evaluate them. The valid pair (ξ_1, ξ_2) for a given pair $(r_{\text{av}}, P_{\text{av}})$ can be found by solving set of nonlinear equations formed by (180) and (181) provided that the solution exists, i.e., a pair $(r_{\text{av}}, P_{\text{av}})$ is feasible.

We illustrate the performance of the adaptive power control algorithm in a single-input single-output (SISO) Rayleigh fading channel as well as multiple-input multiple-output (MIMO) Rayleigh fading channel with two transmitter and two receiver antennas. In both cases the average transmitted signal-to-noise ratio is set to 10 dB, that is, $P_{\text{av}}/\sigma_n^2 = 10$ dB. The relationship between Lagrange multipliers is presented in Figures 27 and 28 for a SISO and MIMO channel, respectively. In both cases, the first Lagrange multiplier ξ_1 is negative whereas the second Lagrange multiplier ξ_2 is positive. Since

$$\xi_1 \approx -\frac{\Delta J}{\Delta K_1} = -\frac{\Delta l_2^- (\eta_t)}{\Delta r_{\text{av}}} < 0 \tag{182}$$

and

$$\xi_2 \approx -\frac{\Delta J}{\Delta K_2} = -\frac{\Delta l_2^- (\eta_t)}{\Delta P_{\text{av}}} > 0, \tag{183}$$

we can conclude that the below-target semideviation of link spectral efficiency $l_2^- (\eta_t)$ will increase when we decrease the average transmission power or increase the average transmission rate requirement. On the other hand, the variance will decrease when the average transmission power is increased or the average rate requirement is decreased. The symbols $\Delta l_2^- (\eta_t)$, Δr_{av}, and ΔP_{av} denote the arbitrarily small changes of the below-target semideviation, average transmission rate, and average transmission power, respectively.

The changes of the below-target semideviation with respect to the average transmission rate and average transission power are shown in Figures 29 and 30. The optimal combinations of mean link spectral efficiency $\mathrm{E}[\eta_r]$ and below-target semideviation of link spectral efficiency $l_2^-(\eta_t)$, shown in Figures 29 and 30, form an efficient frontier. The efficient frontier is a fundamental limit in the sense that it is not possible to construct a transmission scheme with combinations of mean and below-target semideviation of link spectral efficiency which lie above the frontier in Figures 29 and 30. All combinations of mean and below-target semideviation of link spectral efficiency which lie below the frontier are suboptimal. It is because there is other combination of mean and below-target semideviation which offers lower variance for the same mean or higher mean for the same variance. In particular, this combination lies on the efficient frontier.

In Figures 29 and 30 we also plot the risk-reward performance of practical adaptive transmission scheme — truncated channel inversion [19]. Clearly, its performance is suboptimal because it lies below the efficient frontier. The optimal power control rules (169) and (171) are continuous power control rules. As a consequence, there is some maximum value of below-target semideviation which cannot be exceeded without violating average power constraint. Similarly, there is some minimum value of average link spectral efficiency constraint which cannot be lowered without violating the average power constraint. At these two extrema, the Lagrange multipliers achieve their maximum allowable values. Then, the maximum rate boundary and minimum risk boundary for continuous power control rules are found by relaxing the average power constraint. Relaxing the average power constraint means that actual average transmission power is smaller or equal to our target value P_{av}. The boundaries are shown with black and violet lines in Figures 29 and 30. We observe that the below-target semideviation increases as we reduce the average transmission power. This trend is the expected behavior, because the second Lagrange multiplier ξ_2 is positive. We postulate that the minimum risk boundary below the minimum risk scheme and maximum rate boundary beyond the maximum rate scheme are achieved by discontinuous power control rules. For the minimum risk boundary, we set $\xi_2 = 0$. It can be easily shown that it leads to truncated channel inversion and explains why the curves for truncated channel inversion approach the minimum risk boundary. It can be shown that the maximum rate boundary is actually achieved by the power control rule based on Telser's safety-first approach. We have derived the maximum rate boundary for a general MIMO channel. The maximum rate boundaries are shown in Figures 29 and 30. In general, the ridge along which the variational integrand is discontinuous can be located either on the boundary of regions R and T or inside a region R. This fact is reflected in Figure 30 where the maximum rate boundary is marked with a pink and green line for a ridge located on the boundary of regions R and T and ridge located inside a region R, respectively.

In practice, the selection of a specific combination of the mean and the below-target semideviation of link spectral efficiency is achieved by properly selecting one of Lagrange multipliers. We illustrate this in Figures 31 and 32 where we plot the mean of link spectral efficiency and the corresponding below-target semideviation of link spectral efficiency for a given value of the second Lagrange multiplier ξ_2. By changing the value of ξ_2 one moves along the efficient frontier.

4.5.4 Risk-Value Hypothesis

The risk-value model considered in Section 4.2.3 and expected-utility model presented in Section 4.5.3 are compatible models. Specifically, both models lead to the same form of adaptive power control rule. However, these models will definitely select a different power control rule from the set of efficient power control rules. Specifically, the expected utility model will select the scheme with the highest expected utility as the optimal power control rule. On the other hand, the risk-value model will select the scheme with the highest Kappa ratio as the optimal power control rule.

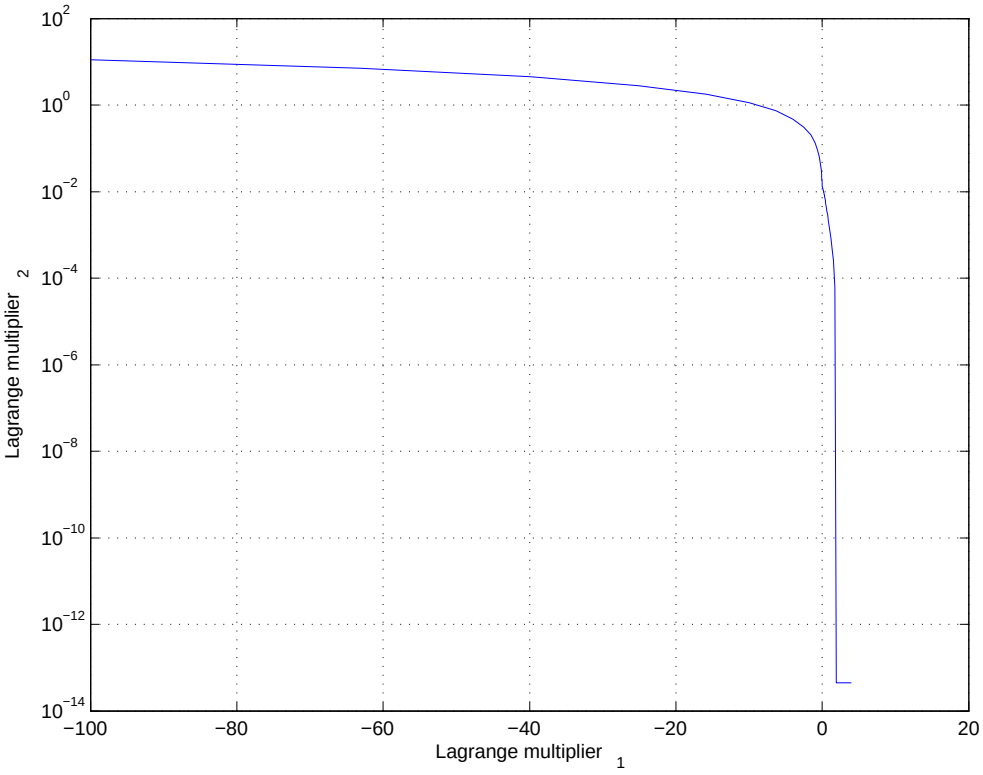

Figure 27: Relationship between Lagrange multipliers in a SISO Rayleigh fading channel.

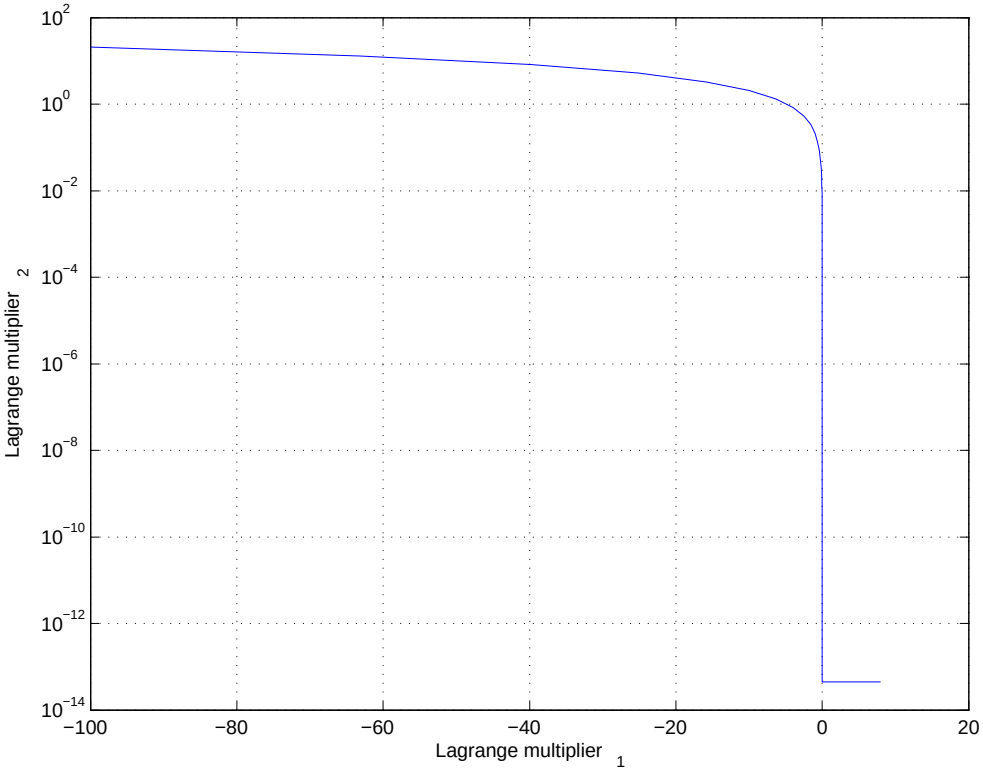

Figure 28: Relationship between Lagrange multipliers in a MIMO Rayleigh fading channel.

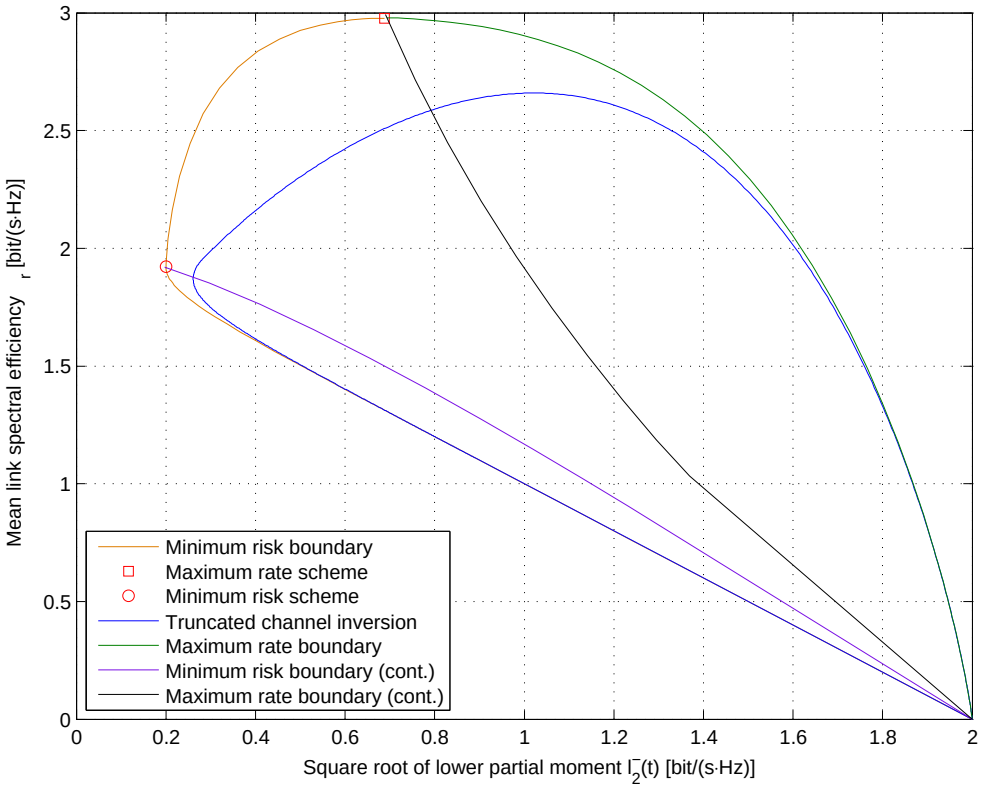

Figure 29: Risk-reward plane for a SISO Rayleigh fading channel.

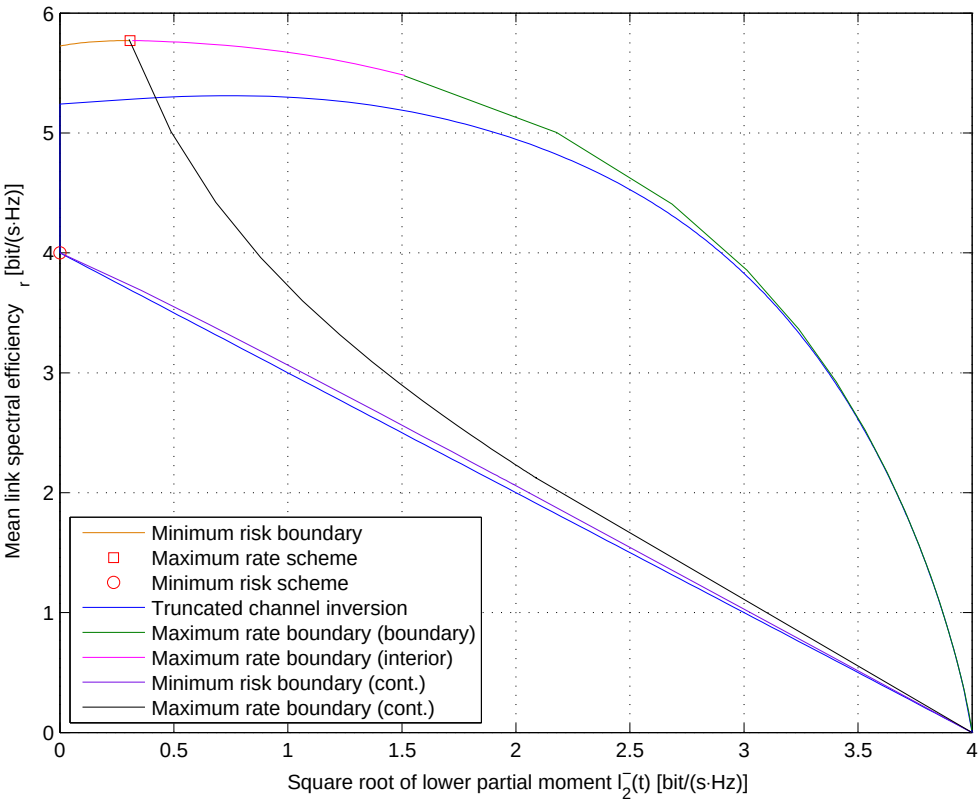

Figure 30: Risk-reward plane for a MIMO Rayleigh fading channel.

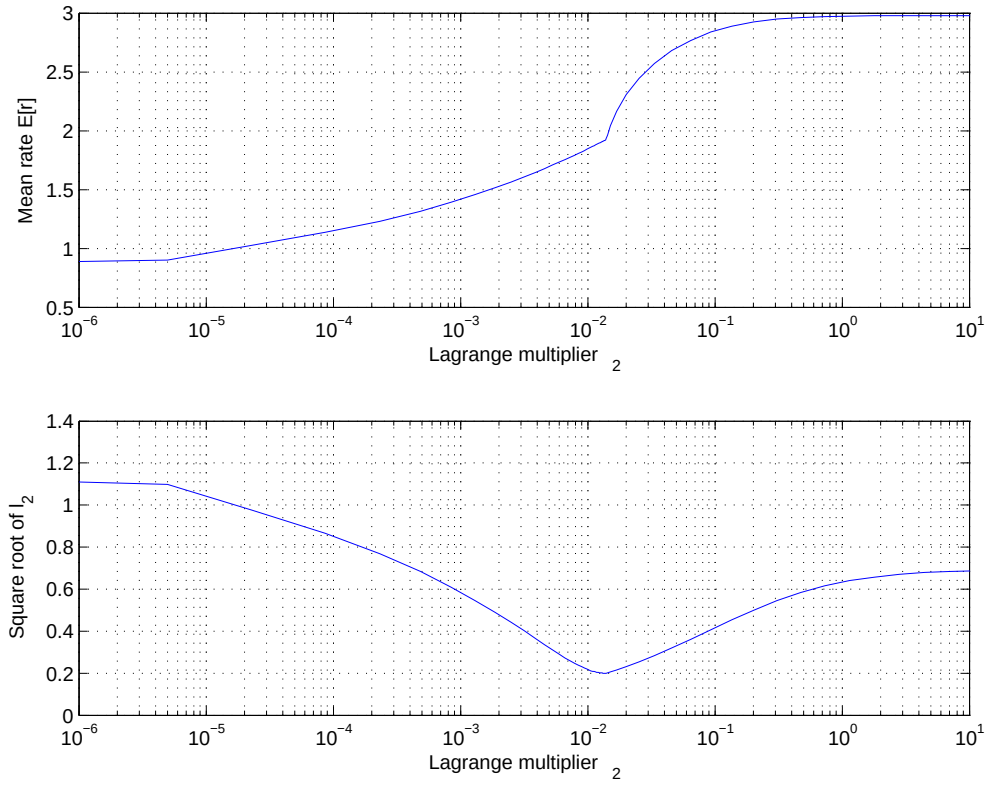

Figure 31: Mean of link spectral efficiency and below-target semideviation as the function of Lagrange multiplier ξ_2 in a SISO channel.

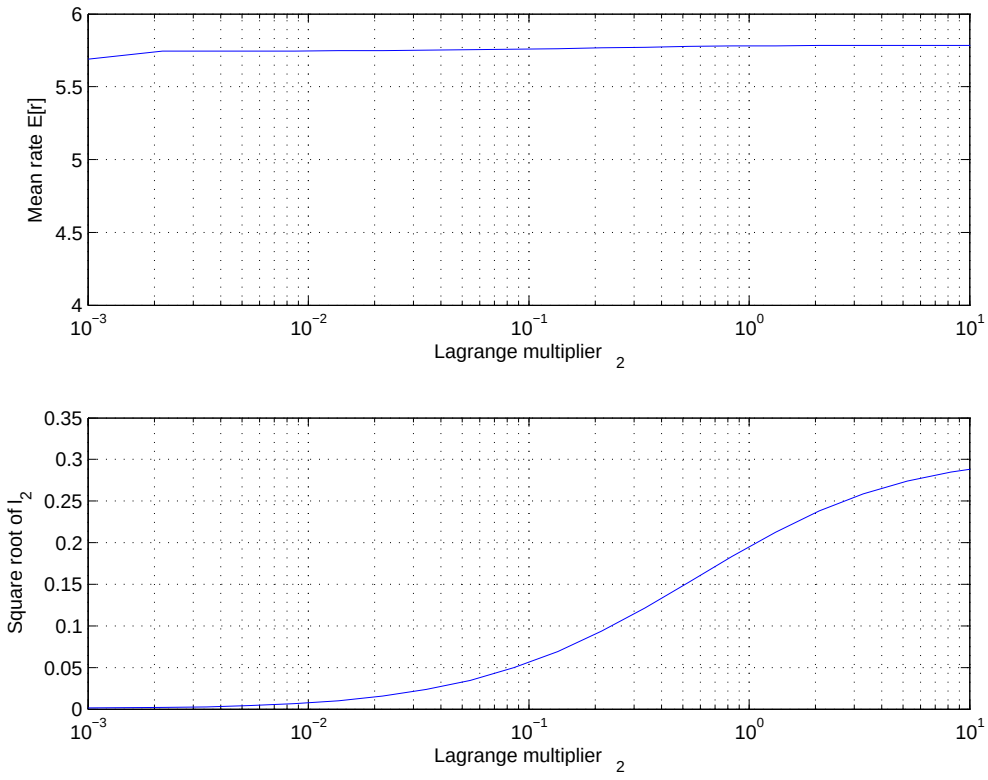

Figure 32: Mean of link spectral efficiency and below-target semideviation as the function of Lagrange multiplier ξ_2 in a MIMO channel.

4.6 Performance of Adaptive Power Control Schemes in Interference Channels

Ad-hoc networking has been of critical interest to the military for many years. In commercial applications there is presently a high interest also in wireless mesh networks, which are an intermediate between conventional cellular networks and ad-hoc networks. Typically, mesh networks are more cost-effective and have better performance than ad-hoc networks. Therefore, our focused software demonstrator will visualize operation of radio resource management block in the transceiver of a broadband mobile ad-hoc network (MANET) with mesh networking capability.

4.6.1 Requirements and specifications for the focused demonstrator

The main requirements of the focused software demonstrator can be divided into performance and complexity requirements. The main performance requirements are:

- R1 The demonstrator models wireless ad-hoc communications between at least three mobile nodes.
- R2 The demonstrator supports pedestrian and vehicular radio propagation environments in 2.5 GHz frequency band.
- R3 The demonstrator models multichannel transmission and reception, for example, simultaneous transmission using multiple antennas.
- R4 The demonstrator uses adaptive modulation and coding techniques that support flexible control of offered Quality of Service with special emphasis on precedence, error rate, overall delay, and throughput.
- R5 The demonstrator enables advanced radio resource management for efficient use of radio resources.

The main complexity requirement is:

- R6 The complexity of the networking equipment is kept at manageable level and excessive system complexity is avoided.

The performance requirements R1–R3 as well as complexity requirement R6 are satisfied by assuming that the demonstrator models a system that resembles an existing and commercially available state-of-the-art networking equipment. In particular, we propose that a reference system for our focused demonstrator is the system based on IEEE 802.16e-2005 standard, which is commonly referred to as “mobile WiMAX.” The major system parameters such as radio propagation models, minimum antenna requirements, and allowable transmission powers are taken from IEEE 802.16e-2005 specification [39].

Focused demonstrator models simultaneous bidirectional transmissions between several transceivers that are randomly spaced in a given geographical area as shown in Figure 33. A simplified block diagram of a single transceiver is shown in Figure 34. Shaded functional blocks in Figure 34 will not be implemented in focused software demonstrator and an appropriate physical abstraction layer will be used instead. In the transmitter part, information source generates a block of data bits to be transmitted. Data bits are encoded by the channel encoder and modulated to form data symbols. The channel encoder is convolutional turbo encoder with the following code rates: 1/2, 2/3, 3/4, and 5/6 [39]. The set of available modulation schemes includes quadrature amplitude modulation (QAM) with 2, 4, 16, and 64 constellation points [39]. In a spatial multiplexer, data symbols are converted from serial to parallel representation. Consequently, a separate stream of data symbols is obtained for every transmitter antenna. After spatial multiplexing, the data symbols are allocated to specific space/frequency subchannels by taking into account their Quality of Service requirements and the actual state of communication channel. Finally, the data symbols are subject to the orthogonal frequency-division multiplexing (OFDM) by inverse fast Fourier transform (IFFT) processors. In the receiver part, the received signal is first processed by fast Fourier transform (FFT) processors to obtain the data symbols. The original order of data symbols is restored in subchannel demapper. Next, the received symbols are converted from parallel to

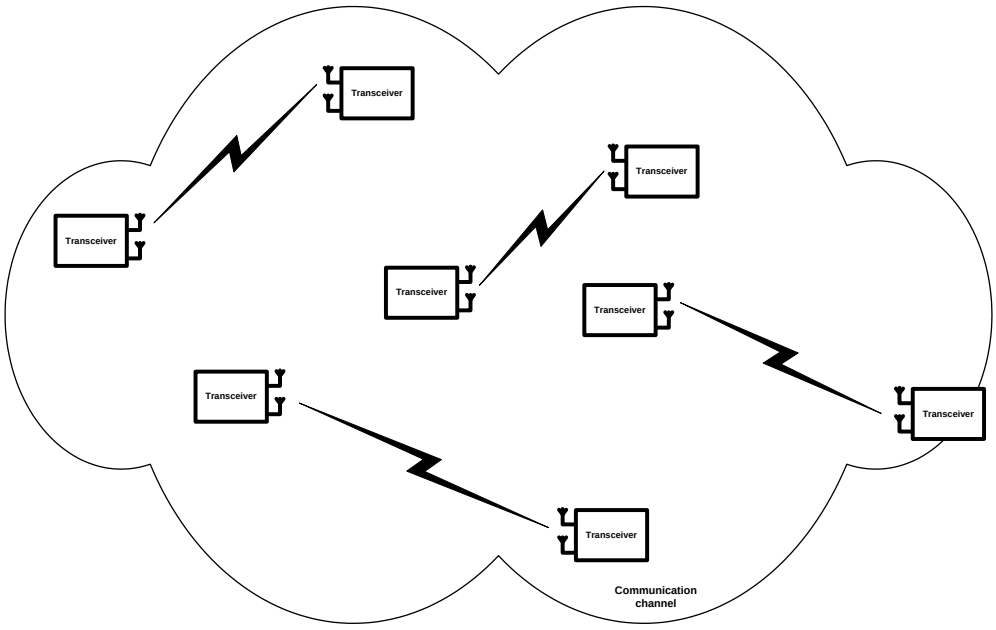

Figure 33: Communication scenario to be modeled by focused demonstrator.

serial representation in spatial demultiplexing block. A block of reconstructed data bits is obtained at the information destination after demodulation and decoding. The channel estimator provides the estimate of the actual channel state to demodulation and decoding blocks as well as radio resource management block. With time division duplex (TDD) the antenna switch swaps the antenna feeds between transmitter and receiver. The antenna switch is controlled by the radio resource management block to enable flexible allocation of radio resources between uplink and downlink transmissions.

Radio network system simulations typically use an abstraction of the physical layer to model link performance. Many simulation methodologies use a block fading channel where an instantaneous snapshot of the desired user's and the interferers' channels is used to determine the respective Signal-to-Interference-plus-Noise Ratios (SINR). Along with knowledge of the transmit power, pathlosses, shadowing and noise variances, various link performance parameters such as instantaneous throughput, bit error rate, and overall delay can be calculated from their dependency on the SINR value. In our focused software demonstrator, the input variables of the physical abstraction layer include modulation and coding scheme, number of transmitter and receiver antennas, Quality of Service requirements for a given block of data bits, number of OFDM subcarriers to be used for transmission, transmission power, and channel occupancy ratio for antenna switch. The outputs of the physical abstraction layer are throughput, bit error rate, and overall transmission delay. The focused software demonstrator will illustrate the principle of operation of advanced radio resource management and channel measurement blocks as well as link level performance as reported by the physical abstraction layer.

Efficient use of radio resources is achieved by appropriate channel measurement and reporting, interference management, and flexible resource allocation mechanisms. In radio resource management, our goal is to optimize the throughput rate between two nodes with respect to constraints on the overall delay, allowable transmission power, available transmission bandwidth, and allowable amount of interference to other users. In particular, we propose to use financial risk-reward theory for flexible radio resource management. The radio resource allocation algorithms are based on the expected utility hypothesis, risk measurement hypothesis, risk-reward hypothesis, and safety-first approach. A novel aspect of these algorithms is the ability to include additional constraint on the risk associated with data transmission. We believe that such a feature enables a very flexible control of Quality of Service parameters because reducing overall transmission delay and prioritizing a given packet can be directly translated

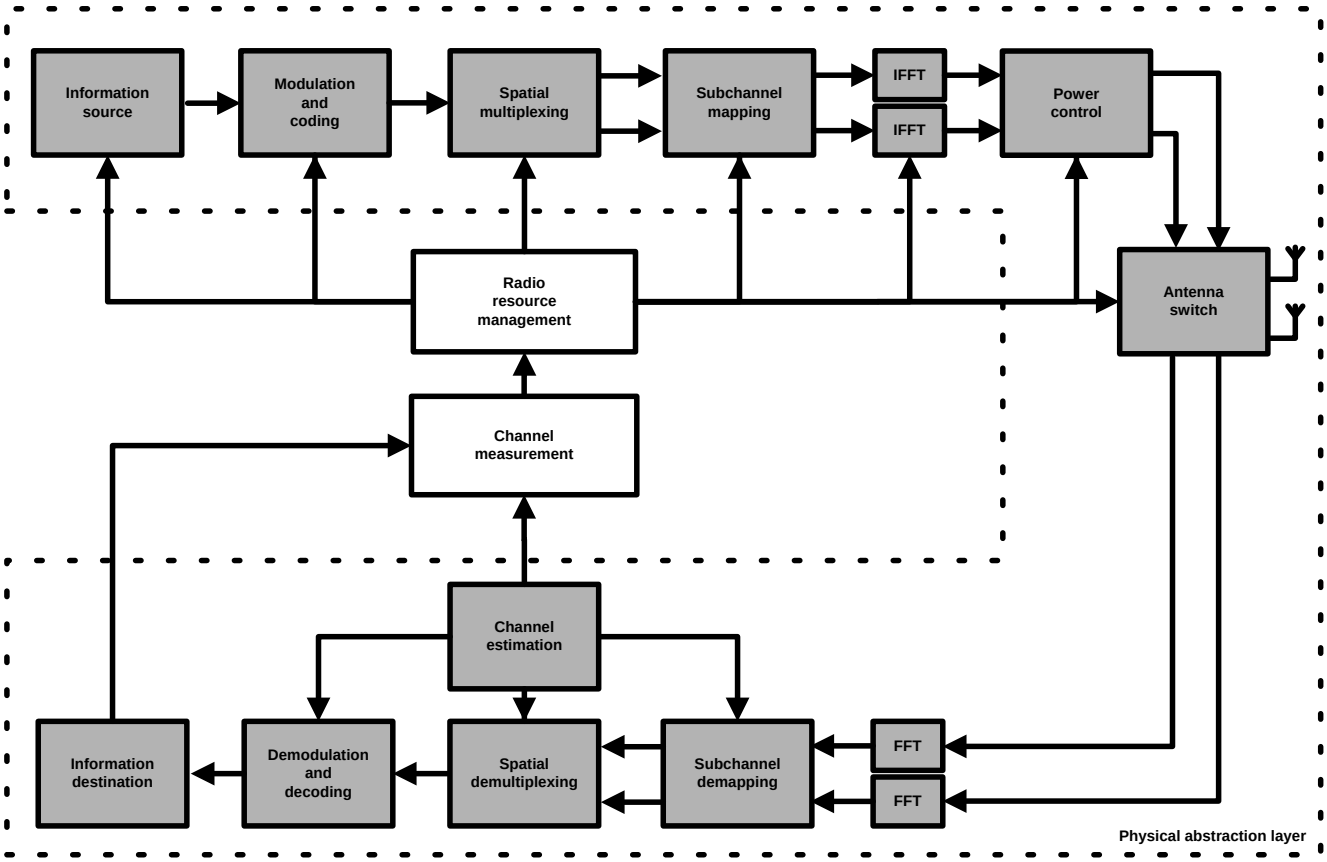

Figure 34: Simplified block diagram of the demonstrator transceiver.

into minimizing the risk in our risk-reward framework. Consequently, with those algorithms the various Quality of Service requirements can be flexibly supported already at the physical layer. As shown in Figure 34, the ultimate goal of radio resource allocation algorithms is to select the optimal combination of modulation scheme, error coding scheme, subchannel mapping permutation, transmission power, and partition of radio resources between uplink and downlink transmissions to maximize the throughput rate. The maximization of throughput rate is performed with respect to constraints on the overall delay, bit error rate, transmission power, transmission bandwidth, and interference level. Furthermore, the radio resource allocation algorithms need information about instantaneous radio propagation conditions as well as the collection of reliable link statistics over different timescales, for example, throughput rate statistics, received signal strength statistics, actual estimate of bit error rate, etc. These measurements are available locally at the transceiver and are reported to a remote transmitter via a dedicated fast feedback channel of the bidirectional radio link.

4.6.2 Numerical results

We study the performance of adaptive power control algorithms developed in Section 4.5.3 in interference channels. We simulate simultaneous transmissions between several transceivers that are randomly spaced in a given geographical are as shown in Figure 33. We assume that all transmitters and receivers are equipped with multiple antennas. Furthermore, we assume that a given receiver is decoding its intended message and treats other transmissions as interference. Consequently, signals received from other transmitters are treated as additional noise.

The link parameters of a given user are randomly selected from respective uniform distributions. The average power constraint P_{av} expressed in decibels is uniformly distributed between 0 and 10 dBm.

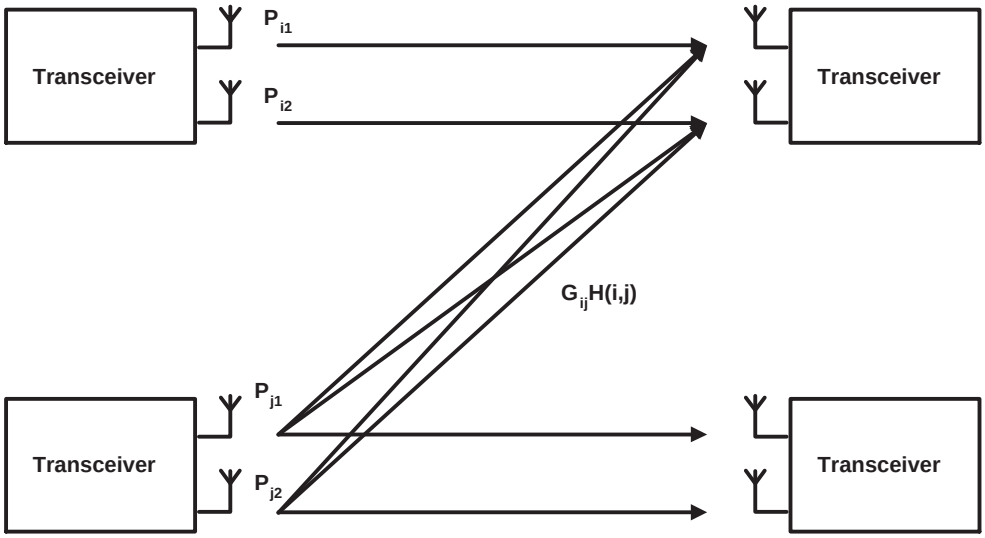

Figure 35: Communication scenario in interference channel.

Similarly, the target link spectral efficiency η_t is uniformly distributed between 0 and 5 bits per channel use, and the first Lagrange multiplier ξ_1 is uniformly distributed between -10^2 and -10^{-4}. The value of the second Lagrange multiplier ξ_2 is selected such that the average power constraint (8) is satisfied with equality.

The interference between users' links is modeled by treating transmissions of other links as interference. The principle of calculating the equivalent noise variance is illustrated in Figure 35. In words, we assume that the equivalent noise power experienced by the ith user at its lth receive antenna is

$$\sigma_{il}^2 = \sigma_n^2 + \sum_{\substack{j=1 \\ i \neq j}}^{N} G_{ij} \sum_{k=1}^{K} \left| H_{kl}\left(i,j\right) \right|^2 P_{jk} \tag{184}$$

where G_{ij} is the large-scale fading parameter that captures the effect of path loss and shadowing in radio wave propagation between ith and jth user, $H_{kl}\left(i,j\right)$ is the small-scale fading parameter that captures the effect of multipath fading between kth transmitter antenna of jth user and lth receiver antenna of ith user, and P_{jk} is the power of radio wave emitted from the kth transmitter antenna of the jth user. The parameters $G_{ij} = G_{ji}$ are randomly selected from a uniform distribution on the interval $(0, G_{\max})$ where $G_{\max}$ denotes the minimum link attenuation. We assume that the multipath fading is Rayleigh fading and consequently $H_{kl}\left(i,j\right)$ are circularly-symmetric complex Gaussian random variables with zero mean and unit variance.

We consider two interference scenarios: strong and mild interference. In the strong interference scenario $G_{\max} = 1$ which corresponds to minimum link attenuation of 0 dB. In other words, the power of the interfering signal is comparable with the power of the useful signal. In the mild interference scenario $G_{\max} = 0.1$ which corresponds to minimum link attenuation of 10 dB.

We study two aspects of adaptive power control algorithms developed in Section 4.5.3: stability and long-term performance. In the stability analysis, all transmission channels are fixed during iterations of adaptive power control algorithm and we observe the transient behavior of the adaptive power control algorithm. In particular, we analyze the transmission power level P_i and achievable link spectral efficiency η_r of every user. In the long-term performance analysis, all transmission channels are randomly selected for every iteration and we observe the behavior of the adaptive power control algorithm. We analyze the transmission power level and achievable link spectral efficiency of every user.

Table 1: Link attenuation between respective users in dB

G_{ij}	User 1	User 2	User 3	User 4	User 5	User 6
User 1	0	10.5268	14.1915	13.5718	12.5164	11.9060
User 2	10.5268	0	20.7023	11.0017	22.9864	13.4486
User 3	14.1915	20.7023	0	17.1914	10.4037	10.0950
User 4	13.5718	11.0017	17.1914	0	16.0565	30.1161
User 5	12.5164	22.9864	10.4037	16.0565	0	15.3680
User 6	11.9060	13.4486	10.0950	30.1161	15.3680	0

Table 2: Link parameters

Parameter	User 1	User 2	User 3	User 4	User 5	User 6
P_{av}/σ_n^2	5.336526 dB	8.244052 dB	3.542203 dB	7.182546 dB	2.219409 dB	0.112527 dB
ξ_1	-74.000026	-52.000048	-34.000066	-46.000054	-100.000000	-14.000086
ξ_2	35.080505	14.785183	23.487298	16.228476	80.410992	15.619362
η_t	0.897401	1.689059	4.629471	4.878653	3.401497	1.643604

Mild Interference Scenario We analyze the performance of the simple ad-hoc network with six interfering users. We assume that each link consists of transmitter with two antennas and receiver with two antennas. The link attenuations between users are shown in Table 1. The parameters of individual links are presented in Table 2.

The transmitter signal-to-noise ratio, that is, the ratio of the instantaneous transmission power of the ith user P_i to the noise variance σ_n^2, is shown in Figures 36 and 38. The achievable link spectral efficiencies η_r of specific links are plotted in Figures 37 and 39. The simulation results in Figures 36 and 37 suggest that the adaptive power control algorithm adapts relatively fast to changing environment and quickly converges to the equilibrium state. In Figure 40 we mark the risk-reward trade-off operating point of every user in interference-free and mild interference scenario with, respectively, asterisk and dot. The respective points for a given user are connected with a line. In general, the results in Figure 40 suggest that the achievable average link spectral efficiencies μ_r decrease and risk l_2^-, associated with transmitting below target rate η_t, increases under interference conditions.

Strong Interference Scenario We analyze the performance of the simple ad-hoc network with six interfering users. We assume that each link consists of transmitter with two antennas and receiver with two antennas. The link attenuations between users are presented in Table 3. The parameters of individual links are presented in Table 4.

The transmitter signal-to-noise ratio, that is, the ratio of the instantaneous transmission power of the ith user P_i to the noise variance σ_n^2, is shown in Figures 41 and 43. The achievable link spectral efficiencies η_r of specific links are plotted in Figures 42 and 44. The simulation results in Figures 41

Table 3: Link attenuation between respective users in dB

G_{ij}	User 1	User 2	User 3	User 4	User 5	User 6
User 1	0	1.3054	3.0054	1.1645	19.4908	5.4526
User 2	1.3054	0	1.0720	0.5829	1.6085	2.0308
User 3	3.0054	1.0720	0	1.2288	2.4931	7.6666
User 4	1.1645	0.5829	1.2288	0	1.9689	0.6061
User 5	19.4908	1.6085	2.4931	1.9689	0	11.6798
User 6	5.4526	2.0308	7.6666	0.6061	11.6798	0

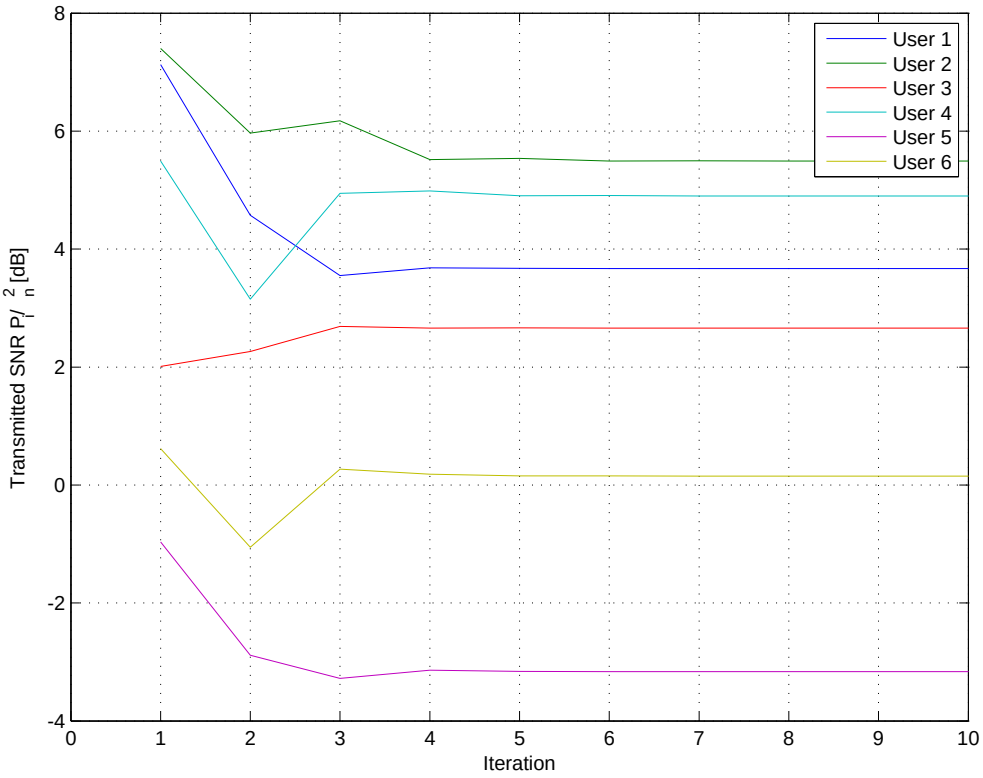

Figure 36: Stability analysis in mild interference scenario: Transmission power of respective users.

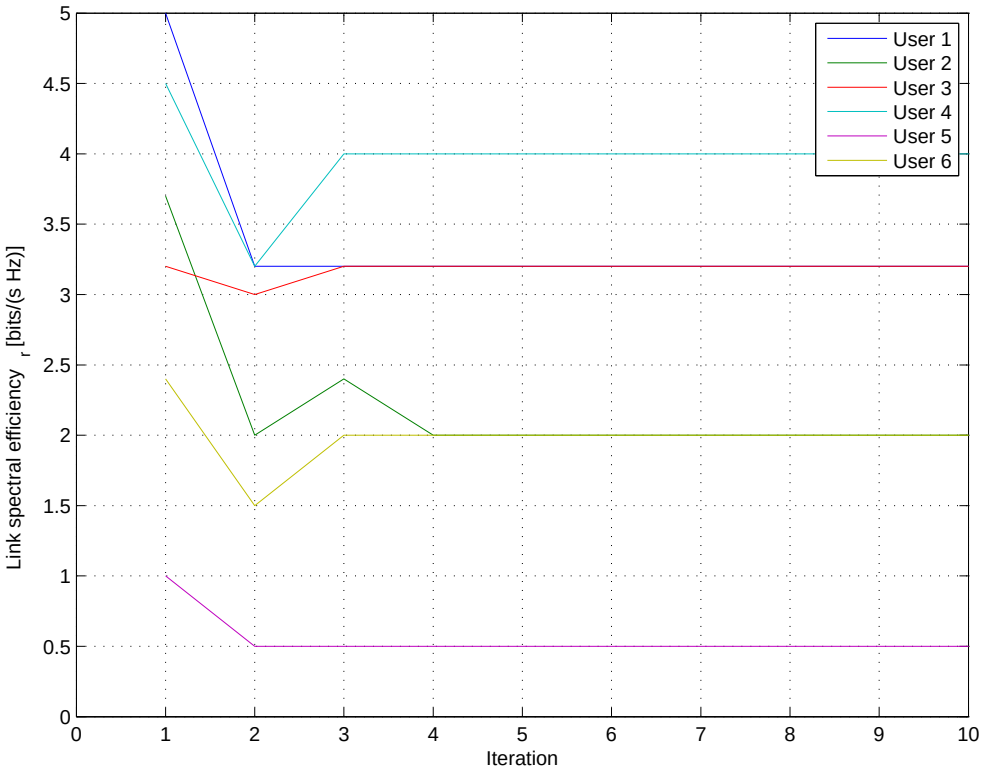

Figure 37: Stability analysis in mild interference scenario: Achievable link spectral efficiencies of respective users.

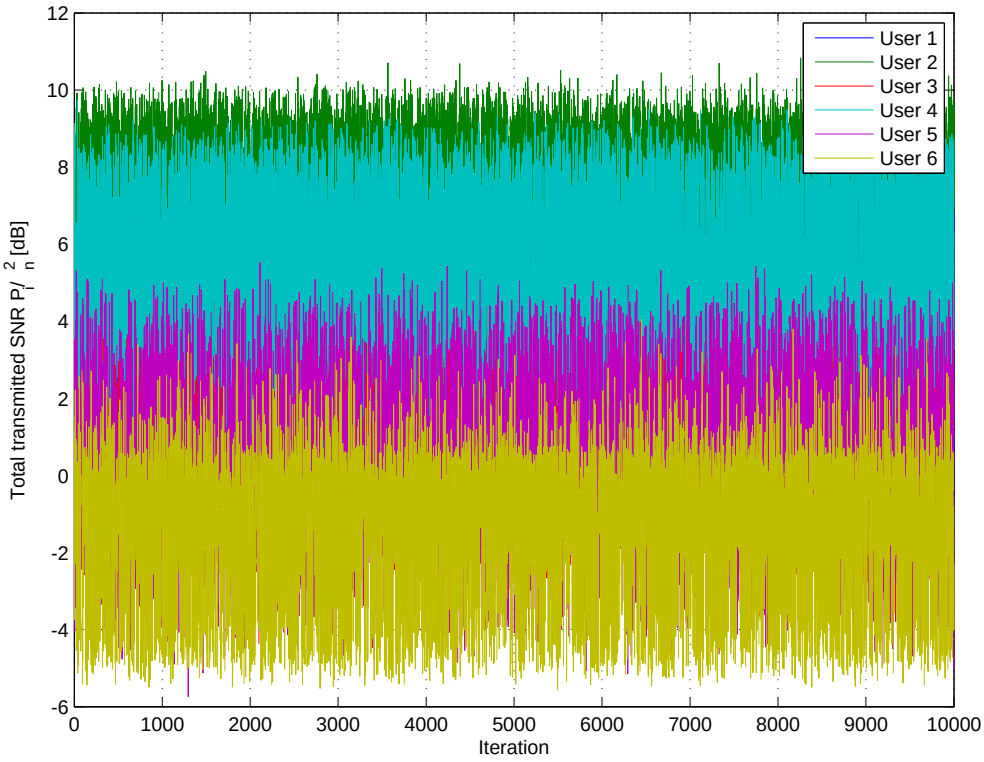

Figure 38: Long-term behavior analysis in mild interference scenario: Transmission power of respective users.

Table 4: Link parameters

Parameter	User 1	User 2	User 3	User 4	User 5	User 6
$P_{\mathrm{av}}/\sigma_n^2$	7.873924 dB	7.436599 dB	4.090288 dB	2.8880306 dB	0.910419 dB	9.123266 dB
ξ_1	-94.000006	-78.000022	-88.000012	-38.000062	-98.000002	-16.000084
ξ_2	28.591191	25.656438	51.467543	27.788194	100.000000	3.878274
η_t	0.510229	1.241310	0.330195	3.322590	4.947051	2.246864

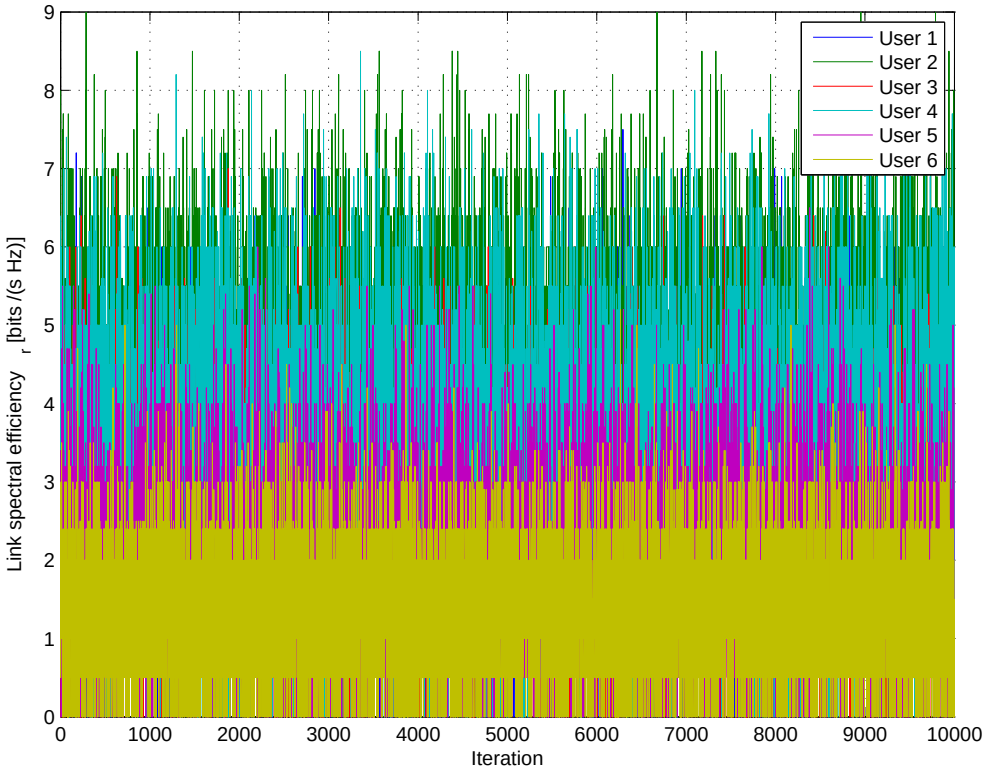

Figure 39: Long-term behavior analysis in mild interference scenario: Achievable link spectral efficiencies of respective users.

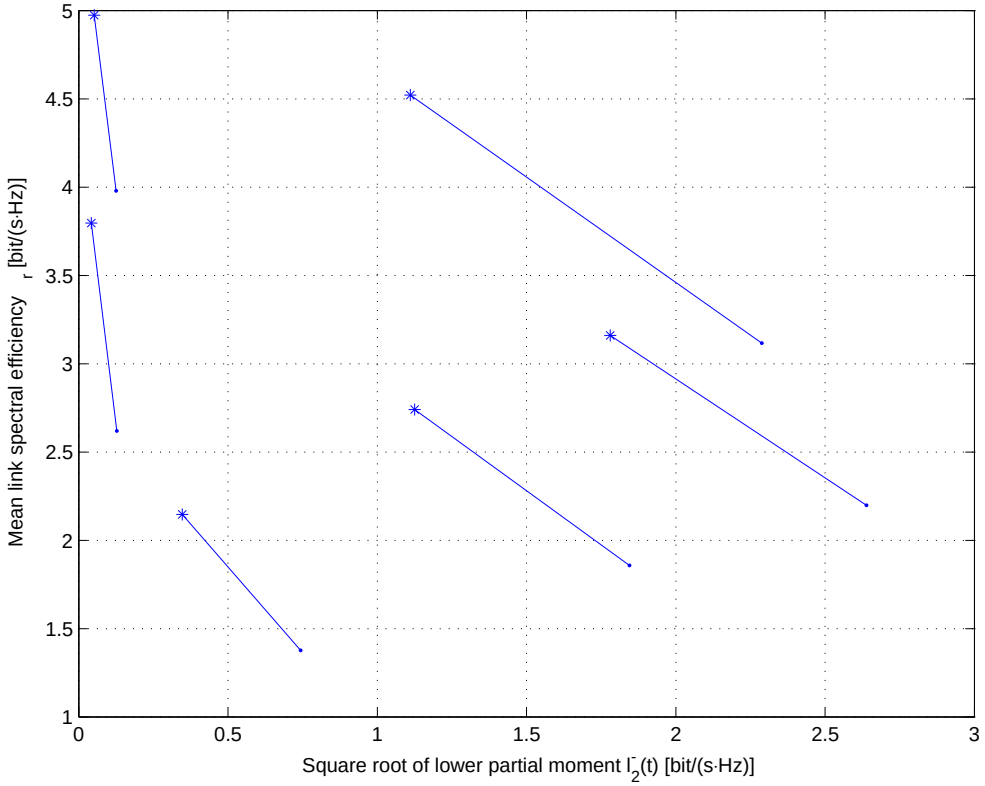

Figure 40: Risk-reward trade-off of respective users in mild interference scenario.

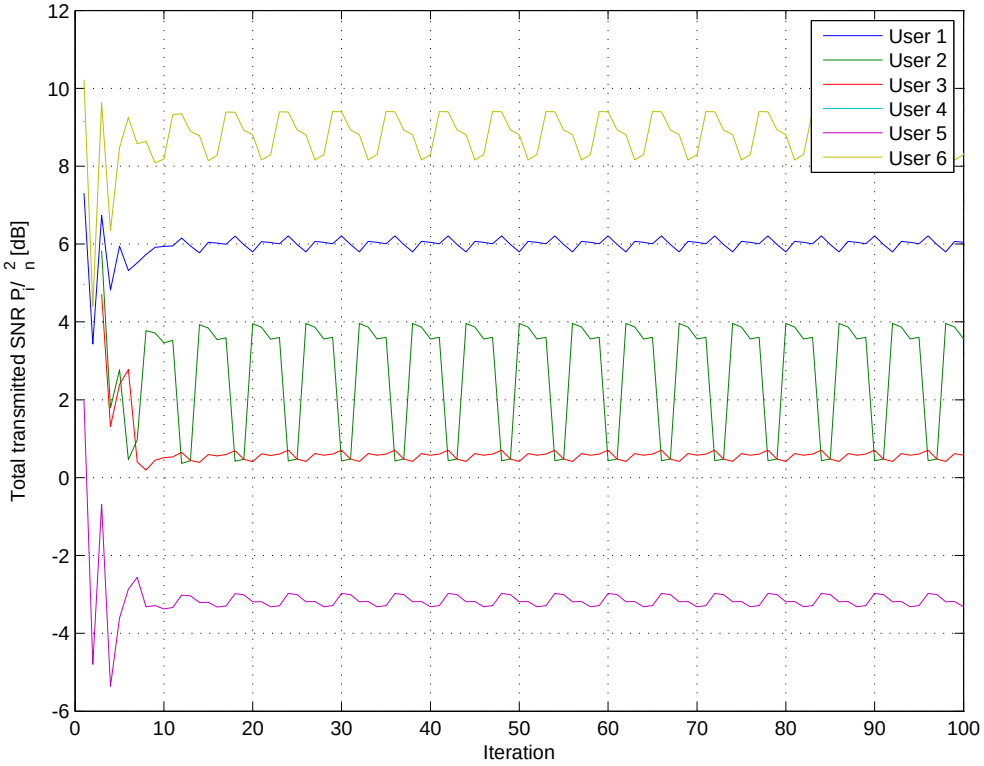

Figure 41: Stability analysis in strong interference scenario: Transmission power of respective users.

and 42 suggest that the adaptive power control algorithm adapts relatively fast to interference and exhibits quasi-stable behavior. In particular, as shown in Figures 41 and 42 the transmission power P_i oscillates between two boundary values whereas link spectral efficiency flips between two neighboring values or, equivalently, modulation and coding schemes. In Figure 45 we mark the risk-reward trade-off operating point of every user in interference-free and strong interference scenario with, respectively, asterisk and dot. The respective points for a given user are connected with a line. In general, the results suggest that the achievable average link spectral efficiencies μ_r decrease and risk l_2^-, associated with transmitting below target rate η_t, increases under interference conditions.

5 Conclusions

We propose a novel approach to analysis and optimization of the performance of adaptive transmission schemes in information-unstable channels. The new approach is based on observation that the instantaneous rate of reliable transmission is a random variable and the best adaptive transmission strategy should be determined by ordering probability distributions. We propose to use methods and tools of theory of rational decision-making to construct relevant representation and uniqueness theorems as well as valid performance indicators. In particular, we suggest using riskiness and expected utility as relevant performance indicators. We rigorously justify their use in wireless communication scenarios using similarity of information-theoretic problems to various decision-theoretic problems. The application of decision theory brings a new intuition to the understanding of the performance of the adaptive transmission in information-unstable channels.

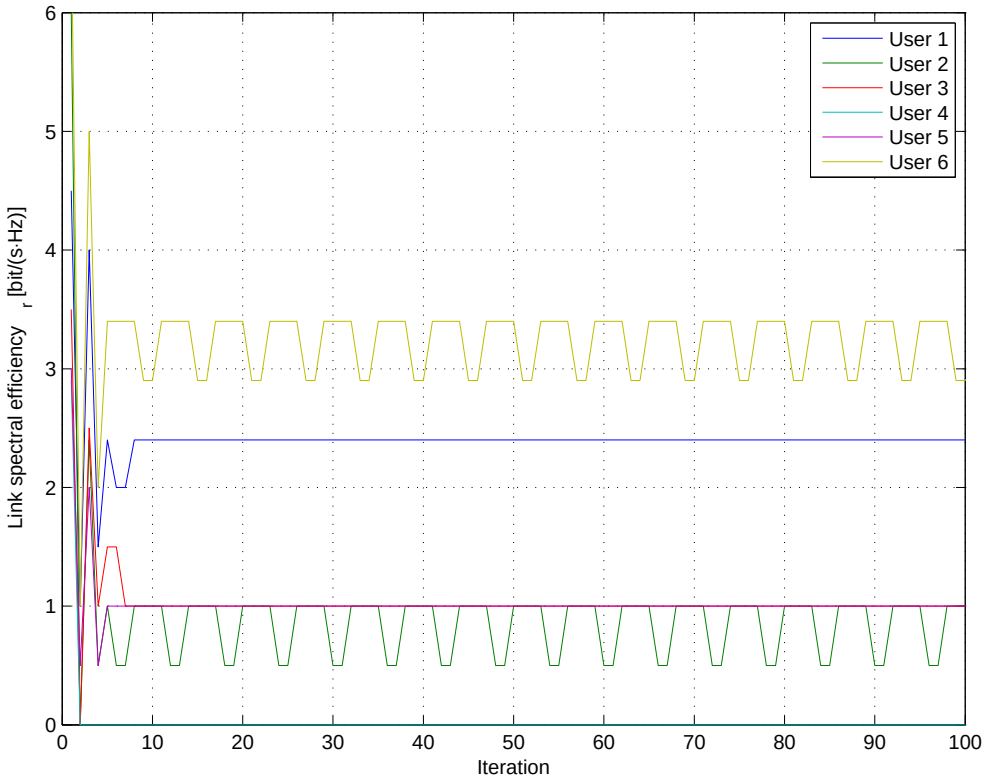

Figure 42: Stability analysis in strong interference scenario: Achievable link spectral efficiencies of respective users.

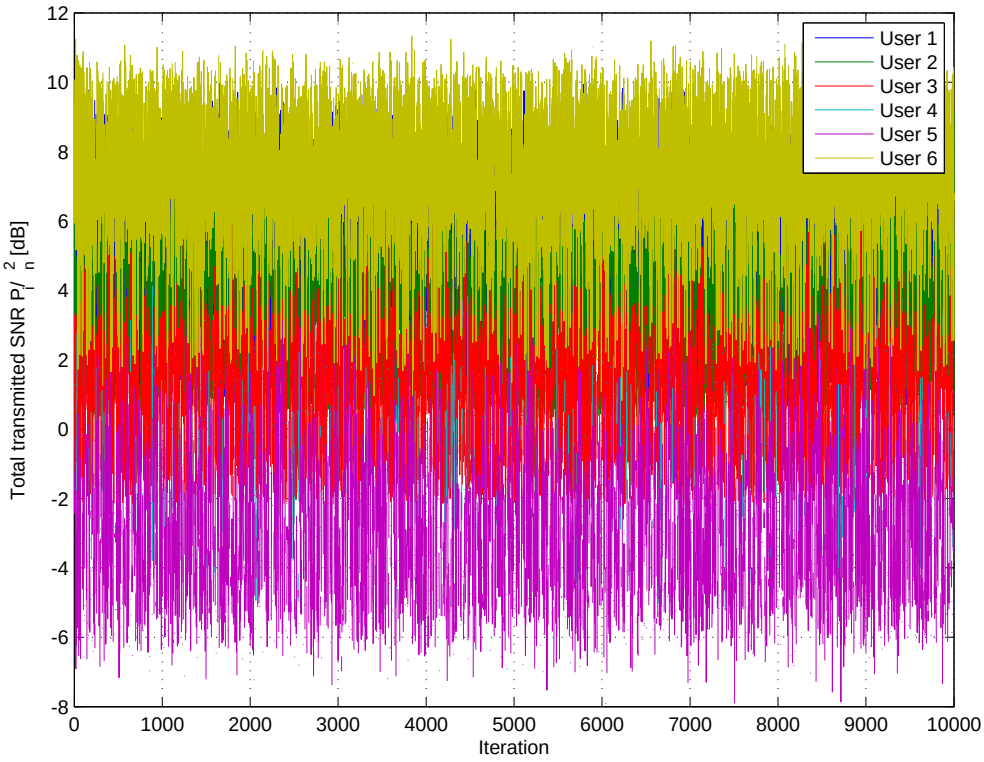

Figure 43: Long-term behavior analysis in strong interference scenario: Transmission power of respective users.

A Proofs of propositions

A.1 Proof of Proposition 7

If $\boldsymbol{\lambda} \in Q$ and $\eta_r(\boldsymbol{\lambda}, \boldsymbol{\rho}) \geq \eta_t > 0$, then $F(\boldsymbol{\lambda}, \boldsymbol{\rho}) = 0$. Consequently, the set of Euler-Lagrange equations (165) becomes

$$\rho_k \left(\frac{\xi_1 \lambda_k}{\sigma_n^2 + \rho_k^2 \lambda_k} + \xi_2 \ln 2 \right) = 0, \quad k = 1, 2, \ldots, m. \tag{185}$$

A trivial solution, $\boldsymbol{\rho} = \mathbf{0}$ for all $\boldsymbol{\lambda} \in T$, does not satisfy the second Weierstrass-Erdmann condition (168) which implies that $\boldsymbol{\rho} \neq \mathbf{0}$ on the boundary of T. It is easily seen that the choice

$$q_k = \rho_k^2 = \sigma_n^2 \left(\theta_T^{-1} - \lambda_k^{-1} \right)_+, \quad k = 1, 2, \ldots, m \tag{186}$$

where

$$\theta_T = -\sigma_n^2 \xi_1^{-1} \xi_2 \ln 2 \tag{187}$$

satisfies (185) for all $\boldsymbol{\lambda}$ such that $F(\boldsymbol{\lambda}, \boldsymbol{\rho}) = 0$.

Let us assume that $\theta_T < 0$, which implies that $q_k \geq 0$ if and only if $\lambda_k < \theta_T$. Since $\lambda_k > 0$ by definition, we obtain $\rho_k = 0$ and $\eta_r(\boldsymbol{\lambda}, \boldsymbol{\rho}) = 0$ for all $\boldsymbol{\lambda}$, which contradicts the assumption $\eta_r(\boldsymbol{\lambda}, \boldsymbol{\rho}) \geq \eta_t > 0$. Therefore, we conclude that $\theta_T > 0$.

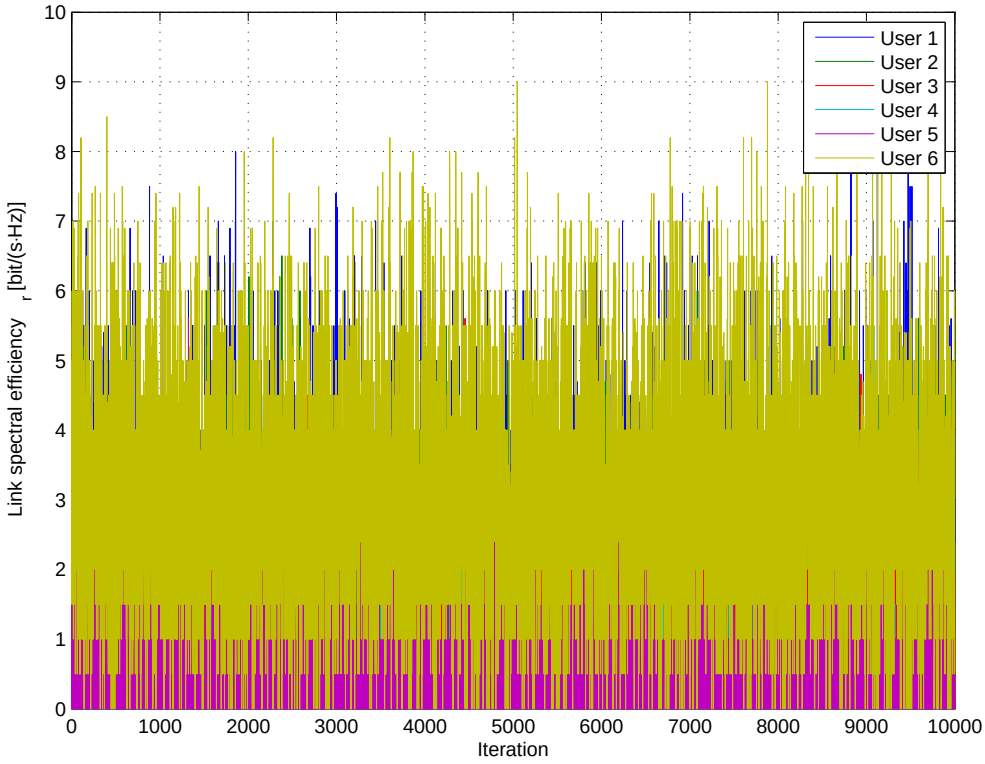

Figure 44: Long-term behavior analysis in strong interference scenario: Achievable link spectral efficiencies of respective users.

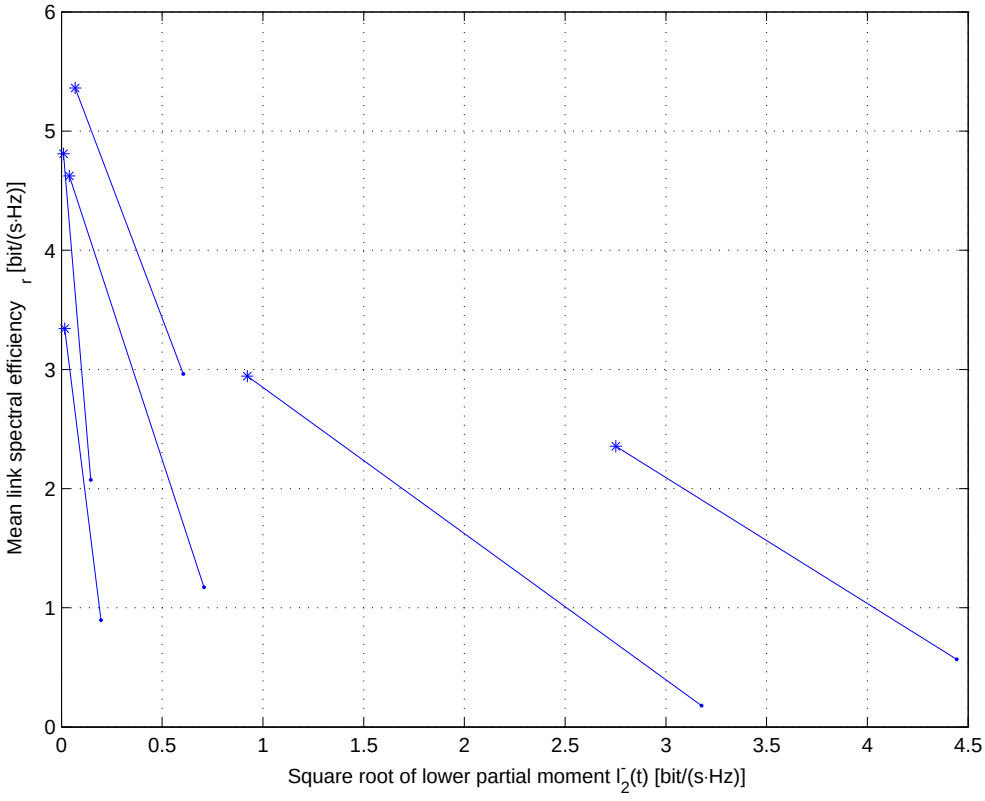

Figure 45: Risk-reward trade-off of respective users in strong interference scenario.

A.2 Proof of Proposition 8

If $\boldsymbol{\lambda} \in Q$ and $0 \leq \eta_r(\boldsymbol{\lambda}, \boldsymbol{\rho}) < \eta_t$, then $F(\boldsymbol{\lambda}, \boldsymbol{\rho}) \neq 0$. Consequently, the set of Euler-Lagrange equations (165) becomes

$$\frac{2\rho_k\lambda_k}{\sigma_n^2 + \rho_k^2\lambda_k}\left[\frac{2\eta_t - \xi_1}{2} - \sum_{k=1}^{m}\log_2\left(1 + \frac{\rho_k^2\lambda_k}{\sigma_n^2}\right)\right] - \frac{\rho_k \ln 2}{\xi_2^{-1}} = 0 \tag{188}$$

where $k = 1, 2, \ldots, m$. A trivial solution, $\boldsymbol{\rho} = \mathbf{0}$ for all $\boldsymbol{\lambda} \in R$, does not satisfy the second Weierstrass-Erdmann condition (168) which implies that $\boldsymbol{\rho} \neq \mathbf{0}$ on the boundary of T. Consequently, $\boldsymbol{\rho} = \mathbf{0}$, which implies that $\eta_r(\boldsymbol{\lambda}, \boldsymbol{\rho}) = 0$, can be a solution to (188) only in some subset $R_0 \subset R$. Let us then assume that $0 < \eta_r(\boldsymbol{\lambda}, \boldsymbol{\rho}) < \eta_t$ and that (188) has a nontrivial and unique solution

$$\boldsymbol{\rho} = (\rho_1, \ldots, \rho_w, 0, \ldots, 0), \quad w = 1, 2, \ldots, m \tag{189}$$

for some $\boldsymbol{\lambda} \in R \backslash R_0$. Since only w components of $\boldsymbol{\rho}$ are positive, it is sufficient to consider only w equations

$$\frac{2\lambda_k}{\sigma_n^2 + \rho_k^2\lambda_k}\left[\frac{2\eta_t - \xi_1}{2} - \sum_{k=1}^{w}\log_2\left(1 + \frac{\rho_k^2\lambda_k}{\sigma_n^2}\right)\right] - \frac{\ln 2}{\xi_2^{-1}} = 0. \tag{190}$$

We observe that the existence of a unique solution (189) implies

$$\sum_{k=1}^{m}\log_2\left(1 + \frac{\rho_k^2\lambda_k}{\sigma_n^2}\right) = \sum_{k=1}^{w}\log_2\left(1 + \frac{\rho_k^2\lambda_k}{\sigma_n^2}\right) = \eta_{\boldsymbol{\lambda}} \tag{191}$$

where $0 < \eta_{\boldsymbol{\lambda}} < \eta_t$ is some constant. Consequently,

$$\frac{\lambda_k}{\sigma_n^2 + \rho_k^2\lambda_k} = \frac{\theta_R}{\sigma_n^2}, \quad k = 1, 2, \ldots, w \tag{192}$$

where $\theta_R > 0$ is some positive constant because the left-hand side of (192) is always positive. By rearranging the terms in (192) and using (191), we obtain

$$q_j = \rho_k^2 = \sigma_n^2\left(\theta_R^{-1} - \lambda_k^{-1}\right)_+, \quad k = 1, \ldots, m \tag{193}$$

where w is a unique integer in $\{1, 2, \ldots, m\}$ such that $\lambda_w > \theta_R \geq \lambda_{w+1}$. The existence and uniqueness of the solution (193) can be proved by repeating the steps of the proof of Lemma 1 in [6, Appendix C], which are related to the existence and uniqueness of the solution (193) to the equation of the form (191).

The value of θ_R can be found by solving (190). In particular, by substituting (193) into (190), we obtain

$$\sum_{k=1}^{w}\log_2\left(\frac{\lambda_k}{\theta_R}\right) - \frac{2\eta_t - \xi_1}{2} = -\frac{\sigma_n^2 \ln 2}{2\theta_R\xi_2^{-1}} \tag{194}$$

which after some additional algebra is equivalent to

$$\underbrace{\left[\frac{\sigma_n^2(\ln 2)^2}{2\theta_R\xi_2^{-1}}\right]^w}_{a}\exp\underbrace{\left[\frac{\sigma_n^2(\ln 2)^2}{2\theta_R\xi_2^{-1}}\right]}_{a} = \underbrace{\left[\frac{\sigma_n^2(\ln 2)^2}{2\xi_2^{-1}}\right]^w\frac{2^{(2\eta_t - \xi_1)/2}}{\prod_{k=1}^{w}\lambda_k}}_{b}. \tag{195}$$

Equation (195), which has the form $a^w e^a = b$, can be transformed into an equation of the form $\tilde{a}e^{\tilde{a}} = \tilde{b}$, whose solution is given by the Lambert W function, i.e., $\tilde{a} = W(\tilde{b})$ [7]. In particular, we use the following change of variables $\tilde{a} = a/w$ and $\tilde{b} = \sqrt[w]{b}/w$ to obtain $\tilde{a}e^{\tilde{a}} = \tilde{b}$. Consequently,

$$\theta_R(w, \boldsymbol{\lambda}) = \frac{\sigma_n^2(\ln 2)^2}{2w\xi_2^{-1}}\left\{W\left[\frac{\sigma_n^2(\ln 2)^2}{2w\xi_2^{-1}}\sqrt[w]{\frac{2^{(2\eta_t - \xi_1)/2}}{\prod_{k=1}^{w}\lambda_k}}\right]\right\}^{-1}. \tag{196}$$

Furthermore, since $\tilde{b} \geq -1/e$ [7], we conclude that

$$\xi_2 \geq -\frac{2we^{-1}}{\sigma_n^2 (\ln 2)^2} \sqrt[w]{\frac{\prod_{k=1}^{w} \lambda_k}{2^{(2\eta_t - \xi_1)/2}}}. \tag{197}$$

On the other hand, substituting (193) into (191) yields

$$0 < \prod\nolimits_{k=1}^{w} \lambda_k < 2^{\eta_t} (\theta_R)^w . \tag{198}$$

Since the inequalities (197) and (198) must be satisfied for any $\boldsymbol{\lambda} \in R$, we conclude that $\xi_2 > 0$ and $\tilde{b} > 0$. The condition $\tilde{b} > 0$ implies that $W(\tilde{b}) > 0$ for all $\boldsymbol{\lambda} \in R$, i.e., we use only the principal branch of the Lambert W function [7].

A.3 Proof of Proposition 9

The target link spectral efficiency η_t can be achieved by transmitting in exactly $w = 1, 2, \ldots, m$ subchannels, i.e., on the boundary of one of the regions T_w

$$S_w = \left\{ \boldsymbol{\lambda} \in T_w : \sum\nolimits_{k=1}^{w} \log_2\!\left(1 + q_k \lambda_k / \sigma_n^2\right) = \eta_t \right\}. \tag{199}$$

By substituting (169) into (199), we obtain

$$S_w = \left\{ \boldsymbol{\lambda} \in T_w : \prod\nolimits_{k=1}^{w} \lambda_k = 2^{\eta_t} (\theta_T)^w \right\}. \tag{200}$$

Since the boundary S is union of S_w, we need to show that component functions ρ_k satisfy Weierstrass-Erdmann conditions (167) and (168) for all possible values of w.

First, we observe that the variational integrand (166) does not depend on the first order derivatives ρ_k'. Consequently, the condition (167) is always satisfied. It is easily seen that (168) is satisfied if $\theta_R(w, \boldsymbol{\lambda}) = \theta_T$ for all $\boldsymbol{\lambda} \in S_w$. If $\boldsymbol{\lambda} \in S_w$, then by substituting (200) and (170) into (172), we obtain

$$\begin{aligned} \theta_R(w, \boldsymbol{\lambda}) &= \frac{\sigma_n^2 (\ln 2)^2}{2w\xi_2^{-1}} \left\{ W\!\left[-\frac{\xi_1 \ln 2}{2w} \exp\!\left(-\frac{\xi_1 \ln 2}{2w} \right) \right] \right\}^{-1} \\ &= -\sigma_n^2 \xi_1^{-1} \xi_2 \ln 2 = \theta_T \end{aligned} \tag{201}$$

because $W(ze^z) = z$ [7].

The proof of the second part of the proposition is based on the observation that the integrand $L = L(\boldsymbol{\lambda}, \boldsymbol{q})$ given by (166) is pointwise convex in $\mathbb{R}_+^m$. Namely, we show that the Hessian matrix

$$\mathcal{H} = \begin{bmatrix} \dfrac{\partial^2 L}{\partial q_1^2} & \dfrac{\partial^2 L}{\partial q_1 \partial q_2} & \dfrac{\partial^2 L}{\partial q_1 \partial q_3} & \cdots & \dfrac{\partial^2 L}{\partial q_1 \partial q_m} \\ \dfrac{\partial^2 L}{\partial q_2 \partial q_1} & \dfrac{\partial^2 L}{\partial q_2^2} & \dfrac{\partial^2 L}{\partial q_2 \partial q_3} & \cdots & \dfrac{\partial^2 L}{\partial q_2 \partial q_m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial^2 L}{\partial q_m \partial q_1} & \dfrac{\partial^2 L}{\partial q_m \partial q_2} & \dfrac{\partial^2 L}{\partial q_m \partial q_3} & \cdots & \dfrac{\partial^2 L}{\partial q_m^2} \end{bmatrix} \tag{202}$$

is positive semidefinite, that is, the determinant of $\mathcal{H}$ is nonnegative.

If $\boldsymbol{\lambda} \in R$, then the partial derivatives are

$$\frac{\partial^2 L}{\partial q_k^2} = \frac{2\lambda_k^2}{\left(\sigma_n^2 + q_k \lambda_k\right)^2 (\ln 2)^2} \left[1 + \eta_t \ln 2 - \sum_{k=1}^{m} \ln\!\left(1 + \frac{q_k \lambda_k}{\sigma_n^2} \right) - \frac{\xi_1}{2} \ln 2 \right] p(\boldsymbol{\lambda}) \tag{203}$$

and

$$\frac{\partial^2 L}{\partial q_i \partial q_j} = \frac{2\lambda_i \lambda_j}{\left(\sigma_n^2 + q_i\lambda_i\right)\left(\sigma_n^2 + q_j\lambda_j\right)\left(\ln 2\right)^2} p(\boldsymbol{\lambda}) \tag{204}$$

for all $i, j, k = 1, 2, \ldots, m$ and all $\boldsymbol{q} \in \mathbb{R}_+^m$ and $\boldsymbol{\lambda} \in R$. To simplify the calculation of $\det \mathcal{H}$, we use the following substitutions

$$c_k = \frac{\sqrt{2}\lambda_k}{\left(\sigma_n^2 + q_k\lambda_k\right)\ln 2}\sqrt{p(\boldsymbol{\lambda})} \tag{205}$$

and

$$d = 1 + \eta_t \ln 2 - \sum_{k=1}^{m} \ln\left(1 + \frac{q_k\lambda_k}{\sigma_n^2}\right) - \frac{\xi_1}{2}\ln 2. \tag{206}$$

Consequently, the Hessian matrix becomes

$$\mathcal{H} = \begin{bmatrix} c_1^2 d & c_1c_2 & c_1c_3 & \ldots & c_1c_m \\ c_2c_1 & c_2^2 d & c_2c_3 & \ldots & c_2c_m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_mc_1 & c_mc_2 & c_mc_3 & \ldots & c_m^2 d \end{bmatrix}. \tag{207}$$

Since the terms c_k, $k = 1, 2, \ldots, m$ are common terms both in the kth row and kth column, we obtain

$$\det \mathcal{H} = \left(c_1c_2\cdots c_m\right)^2 \det \begin{bmatrix} d & 1 & 1 & \ldots & 1 \\ 1 & d & 1 & \ldots & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & \ldots & d \end{bmatrix} = \left(c_1c_2\cdots c_m\right)^2 \det\left[(d-1)\,\boldsymbol{I}_m + \boldsymbol{e}\boldsymbol{e}^{\mathrm{T}}\right] \tag{208}$$

where $\boldsymbol{I}_m$ is an $m \times m$ identity matrix and $\boldsymbol{e}$ is the all-one vector, that is, $\boldsymbol{e} = [1\ 1 \ldots 1]^{\mathrm{T}}$. By using matrix determinant lemma [14, p. 420], we obtain

$$\det \mathcal{H} = \left(c_1c_2\cdots c_m\right)^2 (d-1)^{(m-1)}\,(d-1+m) > 0 \tag{209}$$

because $\xi_1 < 0$ and thus

$$d - 1 = \underbrace{\eta_t \ln 2 - \sum_{k=1}^{m} \ln\left(1 + \frac{q_k\lambda_k}{\sigma_n^2}\right)}_{\geq 0} - \frac{\xi_1}{2}\ln 2 > 0. \tag{210}$$

If $\boldsymbol{\lambda} \in T$, then the partial derivatives are

$$\frac{\partial^2 L}{\partial q_k^2} = -\frac{\xi_1 \lambda_k^2}{\left(\sigma_n^2 + q_k\lambda_k\right)^2 \ln 2} p(\boldsymbol{\lambda}) \tag{211}$$

and

$$\frac{\partial^2 L}{\partial q_i \partial q_j} = 0 \tag{212}$$

for all $i, j, k = 1, 2, \ldots, m$ and all $\boldsymbol{q} \in \mathbb{R}_+^m$ and $\boldsymbol{\lambda} \in T$. The Hessian matrix $\mathcal{H}$ is a diagonal matrix, whose determinant is

$$\det \mathcal{H} = \prod_{k=1}^{m} \left[-\frac{\xi_1 \lambda_k^2}{\left(\sigma_n^2 + q_k\lambda_k\right)^2 \ln 2} p(\boldsymbol{\lambda})\right] > 0 \tag{213}$$

because $\xi_1 < 0$.

From (209) and (213), we can conclude that the Hessian matrix $\mathcal{H}$ is actually a positive definite matrix. Consequently, any $\boldsymbol{q}$ that satisfies Euler-Lagrange equations (165) is an absolute minimizer [13, p. 238].

A.4 Proof of Proposition 6

The proof of Proposition 6 essentially follows the same steps as the proof of Proposition 8 with target link spectral efficiency η_t replaced by the target mean link spectral efficiency r_{av}.

A.5 Proof of Proposition 5

If $\boldsymbol{\lambda} \in R$, then

$$L(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \sum_{k=1}^{m} \int_R \left[\xi_1 + \xi_2 \rho_k^2 - \log_2 \left(1 + \frac{\rho_k^2 \lambda_k}{\sigma_n^2} \right) \right] p(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}. \tag{214}$$

Consequently, the set of Euler-Lagrange equations (165) becomes

$$\rho_k \left(\frac{\lambda_k}{\sigma_n^2 + \rho_k^2 \lambda_k} - \xi_2 \ln 2 \right) = 0, \quad k = 1, 2, \ldots, m. \tag{215}$$

It is easily seen that the choice

$$q_k = \rho_k^2 = \sigma_n^2 \left(\theta_R^{-1} - \lambda_k^{-1} \right)_+, \quad k = 1, 2, \ldots, m \tag{216}$$

where

$$\theta_R = \sigma_n^2 \xi_2 \ln 2 \tag{217}$$

satisfies (215) for all $\boldsymbol{\lambda} \in R$.

On the other hand, if $\boldsymbol{\lambda} \in T$, then

$$L(\boldsymbol{\lambda}, \boldsymbol{\rho}) = \sum_{k=1}^{m} \int_T \left[\xi_3 \rho_k^2 - \log_2 \left(1 + \frac{\rho_k^2 \lambda_k}{\sigma_n^2} \right) \right] p(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}. \tag{218}$$

Consequently, the set of Euler-Lagrange equations (165) becomes

$$\rho_k \left(\frac{\lambda_k}{\sigma_n^2 + \rho_k^2 \lambda_k} - \xi_3 \ln 2 \right) = 0, \quad k = 1, 2, \ldots, m. \tag{219}$$

It is easily seen that the choice

$$q_k = \rho_k^2 = \sigma_n^2 \left(\theta_T^{-1} - \lambda_k^{-1} \right)_+, \quad k = 1, 2, \ldots, m \tag{220}$$

where

$$\theta_T = \sigma_n^2 \xi_3 \ln 2 \tag{221}$$

satisfies (219) for all $\boldsymbol{\lambda} \in R$. A trivial solution, $\boldsymbol{\rho} = \mathbf{0}$ for all $\boldsymbol{\lambda} \in T$, does not satisfy the second Weierstrass-Erdmann condition (130) which implies that $\boldsymbol{\rho} \neq \mathbf{0}$ on the boundary of T.

Let us assume that $\theta_T < 0$, which implies that $q_k \geq 0$ if and only if $\lambda_k < \theta_T$. Since $\lambda_k > 0$ by definition, we obtain $\rho_k = 0$ and $\eta_r(\boldsymbol{\lambda}, \boldsymbol{\rho}) = 0$ for all $\boldsymbol{\lambda}$, which contradicts the assumption $\eta_t > 0$. Therefore, we conclude that $\theta_T > 0$.

References

[1] K. J. Arrow, “The use of unbounded utility functions in expected utility maximization: Response,” *Quarterly Journal of Economics*, vol. 88, pp. 136–138, Feb. 1974.

[2] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, “Coherent measures of risk,” *Mathematical Finance*, vol. 9, pp. 203-228, July 1999.

[3] V. S. Bawa, “Optimal rules for ordering uncertain prospects,” Journal of Financial Economics, vol. 2, pp. 95-121, March 1975.

[4] D. Bernoulli, “Specimen theoriae novae de mensura sortis,” *Commentarii Academiae Scientiarum Imperialis Petropolitanae*, vol. 5, pp. 175–192, 1738. English translation in *Econometrica*, vol. 22, pp. 23–36, Jan. 1954.

[5] E. Biglieri, J. Proakis, and S. Shamai (Shitz), “Fading channels: Information-theoretic and communication aspects,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

[6] G. Caire, G. Taricco, and E. Biglieri, “Optimum power control over fading channels,” *IEEE Transactions on Information Theory*, vol. 45, pp. 1468–1489, Jul. 1999.

[7] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, “On the Lambert W function,” *Advances in Computational Mathematics*, vol. 5, pp. 329–359, Dec. 1996.

[8] P. C. Fishburn, “Utility theory,” *Management Science*, vol. 14, pp. 335-378, January 1968.

[9] P. C. Fishburn, “Unbounded expected utility,” *Annals of Statistics*, vol. 3, pp. 884–896, Jul. 1975.

[10] P. C. Fishburn, “Mean-risk analysis with risk associated with below-target returns,” *American Economic Review*, vol. 67, pp. 116–126, Mar. 1977.

[11] S. French, *Decision Theory: An Introduction to the Mathematics of Rationality*, Chichester: Ellis Horwood, 1988.

[12] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. Englewood Cliffs, NJ: Prentice-Hall, 1963.

[13] M. Giaquinta and S. Hildebrandt, *Calculus of Variations I*. Berlin: Springer-Verlag, 1996.

[14] D. A. Harville, *Matrix Algebra From a Statistician’s Perspective*. Berlin: Springer-Verlag, 2008.

[15] P. D. Kaplan and J. A. Knowles, “Kappa: A generalized downside risk-adjusted performance measure,” *Journal of Performance Measurement*, vol. 8, no. 3, pp. 42–54, 2004.

[16] S. Kataoka, “A stochastic programming model,” *Econometrica*, vol. 31, pp. 181-196, January 1963.

[17] G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientist and Engineers: Definitions, Theorems, and Formulas for Reference and Review*. Mineola, NY: Dover, 2000.

[18] A. Kotelba, A. Mämmelä, and D. P. Taylor, “Normalization of linear vector channels,” in *Proceedings of the IEEE Global Telecommunications Conference*, Washington, DC, 26–30 Nov. 2007, pp. 4537–4542.

[19] A. Kotelba and A. Mämmelä, “Application of financial risk-reward theory to adaptive transmission,” in *Proceedings of the 67th Vehicular Technology Conference*, Marina Bay, Singapore, 11-14 May 2008, pp. 1756-1760.

[20] A. Kotelba and A. Mämmelä, "Efficient power control over fading channels," in *Proceedings of the 51st IEEE Global Telecommunications Conference*, New Orleans, USA, 30 Nov.-4 Dec. 2008.

[21] D. H. Krantz, R. D. Luce, P. Supes, and A. Tversky, *Foundations of Measurement: Additive and Polynomial Representation*, San Diego, CA: Academic Press, 1971.

[22] D. M. Kreps, *Notes on the Theory of Choice*, Boulder, CO: Westview Press, 1988.

[23] K. R. Kumar and G. Caire, "Information theoretic foundations of adaptive coded modulation,", *Proceedings of the IEEE*, vol. 95, no. 12, pp. 2274–2298, Dec. 2007.

[24] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Transactions on Information Theory*, vol. 44, pp. 2148-2177, October 1998.

[25] H. Levy, *Stochastic Dominance: Investment Decision Making under Uncertainty*, New York, NY: Springer Science, 2nd edition, 2006.

[26] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. 2nd ed. Malden: Blackwell, 1991.

[27] H. M. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, Malden, MA: Blackwell, 2nd edition, 1991.

[28] K. Menger, "Das Unsicherheitsmoment in der Wertlehre: Betrachtungen in Anschluss an das sogenannte Petersburger Spiel," *Zeitschrift für Nationalökonomie*, vol. 5, pp. 459–485, Aug. 1934.

[29] D. N. Nawrocki, "A brief history of downside risk measures," *Journal of Investing*, pp. 9–26, Fall 1999.

[30] J. von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton, MA: Princeton University Press, 1944.

[31] A. Pollatsek and A. Tversky, "A theory of risk," *Journal of Mathematical Psychology*, vol. 7, pp. 540-553, October 1970.

[32] A. D. Roy, "Safety-first and the holding of assets," *Econometrica*, vol. 20, pp. 431-449, July 1952.

[33] T. M. Ryan, "The use of unbounded utility functions in expected utility maximization: Comment," *Quarterly Journal of Economics*, vol. 88, pp. 133–135, Feb. 1974.

[34] P. A. Samuelson, "St. Petersburg paradoxes: Defanged, dissected, and historically described," *Journal of Economic Literature*, vol. 15, pp. 24–55, Mar. 1977.

[35] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, Jul.,Oct. 1948.

[36] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecomm.*, vol. 10, pp. 585–595, Nov./Dec. 1999.

[37] L. G. Telser, "Safety first and hedging," *Review of Economic Studies*, vol. 23, pp. 1–16, Jan. 1955.

[38] S. Verdu and T. S. Han, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, pp. 1147-1157, July 1994.

[39] WiMAX Forum, “Mobile WiMAX - Part I: A Technical Overview and Performance Evaluation,” February 2006.

[40] G. L. Stüber, *Principles of Mobile Communication*, 2nd ed., Norwell, MA: Kluwer, 2001.

[41] A. Mämmelä, A. Kotelba, M. Höyhtyä, and D. P. Taylor, “Relationship of average transmitted and received energies in adaptive transmission,” *IEEE Trans. Veh. Technol.*, vol. 59, pp. 1257–1268, Mar. 2010.

[42] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, NY: New York, 2005.

[43] L. A. Goodman, “On the exact variance of products,” *Journal of the American Statistical Association*, vol. 55, pp. 708–713, Dec. 1960.

[44] W. Feller, *An Introduction to Probability Theory and Its Applications*, vol. 2, New York: Wiley, 1971.

[45] P. Bias, S. Hedman, and D. Rose, “Boundary distributions with respect to Chebyshev’s inequality,” *Journal of Mathematics and Statistics*, vol. 6, pp. 47–51, Jan. 2010.

[46] H. Bottomley. (2002, Apr.) One tailed version of Chebyshev’s inequality. [Online] Available: http://www.btinternet.com/∼se16/hgb/cheb.htm

Symbols and abbreviations

CDI	channel distribution information
CSI	channel state information
FFT	fast Fourier transform
IEEE	Institute of Electrical and Electronics Engineers
IFFT	inverse fast Fourier transform
OFDM	orthogonal frequency-division multiplexing
QAM	quadrature amplitude modulation
Qos	quality of servise
SINR	Signal-to-Noise-plus-Interference Ratio
TDD	time division duplex
WiMAX	Worldwide Interoperability for Microwave Access
WSSUS	wide-sense stationary uncorrelated scattering

$l_n^-(\cdot)$	nth order lower partial moment
$\varepsilon(\cdot)$	outage probability
$\kappa(\cdot)$	reward-to-semivariablity ratio
η_r	link spectral efficiency
η_t	target link spectral efficiency
ξ	Lagrange multiplier
$\succ$	strict ordering relation
$\succcurlyeq$	weak ordering relation
$\sim$	equivalence relation
$\circ$	concatenation
$\neg$	logical contradiction
$\vee$	logical disjunction
$\Rightarrow$	logical implication
$\Leftrightarrow$	logical equivalence
$\subseteq$	inclusion relation
$\subset$	proper inclusion relation
$\mathbb{Q}$	set of rational numbers
$\mathbb{R}$	set of real numbers
$\mathbb{R}_+$	set of nonnegative real numbers
$\mathbb{R}^n$	n-dimensional Euclidean space
$\mathbb{R}_+^n$	nonnegative orthant of n-dimensional Euclidean space
$E[\cdot]$	expectation operator
$R(\cdot)$	risk measure
$u(\cdot)$	elementary utility function
$U(\cdot)$	von Neumann-Morgenstern utility function
$V(\cdot)$	risk index
$\mathrm{Var}[\cdot]$	variance operator
$W(\cdot)$	Lambert W function
$\mathcal{A}^c$	complement of set $\mathcal{A}$
$\mathcal{F}$	algebra of sets
$\mathcal{P}$	set of probability measures
$\mathcal{P}_d$	set of one-point probability measures
$\mathcal{P}_s$	set of simple probability measures