Doctoral theses at NTNU, 2020:10

Fengyou Sun

Dependence Control in Wireless Communication

Doctoral Thesis

Fengyou Sun

ISBN 978-82-326-4388-2 (printed version)
ISBN 978-82-326-4389-9 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2020:10

NTNU
Norwegian University of
Science and Technology
Faculty of Information Technology
and Electrical Engineering
Department of Information Security
and Communication Technology

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

NTNU

Fengyou Sun

Dependence Control in Wireless Communication

Thesis for the degree of Philosophiae Doctor

Trondheim, February 2020

Norwegian University of Science and Technology
Faculty of Information Technology
and Electrical Engineering
Department of Information Security and Communication
Technology

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology
and Electrical Engineering
Department of Information Security and Communication
Technology

ISBN 978-82-326-4388-2 (printed version)
ISBN 978-82-326-4389-9 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2020:10

Printed by Skipnes Kommunikasjon as

Abstract

In the physical world, the system components or the system states are probabilistically related to each other spatially and temporally. This kind of relation is termed stochastic dependence or dependence, which is described in mathematics by the probability measures of the dependent elements. We regard the dependence as a physical reality as well as a mathematical property and propose to control the dependence for the system performance improvement. We build a theory of dependence control and apply the theory to the wireless communication system. Specifically, we prove that the wireless channel capacity is intrinsically light-tailed due to the passive nature of the wireless channel and the power limitation, the dependence of the stochastic process is transformable due to the existence of both uncontrollable and controllable random parameters, and dependence in the arrival process and service process of a queueing system are measurable and have a dual potency to influence the queueing system performance. Particularly, we summarize the dependence measures of the queueing system, the dual potency of the arrival and service processes, and the dependence transformability of the stochastic process as the three principles of the dependence control theory, i.e., the measurability, duality, and transformability.

Preface

This thesis is submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy at Norwegian University of Science and Technology. This thesis is my original work. The main contribution of this thesis lies in building a theory of dependence control.

The stochastic dependence or dependence for short describes the probabilistic interrelationships between the system states in physical systems. The dependence is an inherent property of the physical systems and stochastic processes, particularly, the dependence has a significant impact on the system performance. For example, the dependence in the arrival process and the service process, even with light-tailed marginal distributions, can induce a heavy-tailed distribution of the performance measures of the queueing system. In parallel with modeling the stochastic dependence in the physical system, this thesis proposes to control the dependence through the dependence manipulation techniques in order to obtain a better system performance. In principle, the control of dependence is feasible due to the fact that the system performance measures are usually determined by not only the uncontrollable random parameters but also the controllable random parameters, and a manipulation of the dependence in the stochastic processes of the controllable random parameters has a consequent influence on the system performance. The dependence control theory applies to the general stochastic systems, e.g., the queueing systems, specifically, we consider the case of wireless communication systems, of which the stochastic properties are studied in particular. The contents of this thesis are explicated as follows.

In Chapter 1, we provide an overview of this work. Specifically, we introduce the thoughts of dependence control, outline the structure of the dependence control theory, and provide the reason for choosing the mathematical techniques for proving the results in the dependence control theory. In addition, the related methodologies, like dependence modeling and dependence coupling, and the related application areas, like simulation and finance, are reviewed.

In Chapter 2, we provide a tail perspective on the wireless channel gain and the wireless channel capacity. We show that the wireless channel gain has finite moments in the stochastic channel models due to the passive nature of the wireless propagation environment. Furthermore, we show that the light-tail behavior is an intrinsic property of the wireless channel capacity considering the power constraints in the wireless communication system. The results provide a foundation for the mathematical analysis with respect to the moments of the wireless channel capacity, i.e., the moments of all orders exist.

In Chapter 3, we investigate how to control the dependence in a queueing system. We consider three fundamental questions raised by dependence control: what to measure the dependence; where to seek the dependence; and how to transform the dependence. By answering these questions, we formulate the underlying rules of dependence influence on the system performance as the three principles of dependence control, i.e., measurability, duality, and transformability, which verify the feasibility of dependence control and make up the building blocks of the dependence control theory. As a demonstration, we provide simulation and numerical results of the application of the dependence control theory to the wireless communication system.

In Chapter 4, we conclude this thesis and discuss the future research topics. Particularly, it is interesting to extend the dependence control concept to different application scenarios, different operational systems, and different probability structures.

Acknowledgement

I would like to thank my supervisor Prof. Yuming Jiang for his enlightening guidance in the early days of my research career, for his openness and support for letting me work on the topics that I find interesting, and for his generous help that has made many things possible for me. In addition, I would like to thank Prof. Jiang for proofreading and commenting the drafts of this thesis.

I would like to thank my department for being a wonderful workplace for doing research and I would like to thank everyone in the department for making such a workplace a reality. Specifically, I would like to thank Pål Sturla Sæther for playing darts with me in the past four years.

I would like to express my thanks for living in the great era and beautiful world. Particularly, I would like to thank China Scholarship Council for supporting my study in Norway.

Last but not least, I would like to thank my parents and brother for their enduring love. This thesis is dedicated to them.

Fengyou Sun

Trondheim, Norway
October, 2019

Contents

Abstract **i**

Preface **iii**

Notation **ix**

1 Introduction **1**

1.1 Dependence in Perspective 2

1.1.1 Dependence Here and There 3

1.1.2 Dependence in Mathematics 4

1.1.3 Dependence as Physical Resource 6

1.2 Essence of Dependence Control 8

1.2.1 The Light-Tail Property 8

1.2.2 The Dependence Market 11

1.2.3 The Three Principles 13

1.3 A Note on Methodologies 16

1.3.1 The Methodology 16

1.3.2 Related Methodologies 19

2 A Tale of Tails **23**

2.1 Single-Input-Single-Output Channel 24

2.1.1 Channel Capacity 24

2.1.2 Light-Tail Behavior 26
2.1.3 Dependence Refinement 31
2.2 Multiple-Input-Multiple-Output Channel 39
2.2.1 Deterministic Power Fluctuation 39
2.2.2 Random Power Fluctuation 50
2.3 Random Variable Arithmetic 55

3 The Facts of Dependence **69**
3.1 Foundation 70
3.1.1 Queueing Behavior 70
3.1.2 More Assumptions 71
3.1.3 Change of Measure 73
3.2 Dependence Mechanics 75
3.2.1 The General Rules 75
3.2.2 Markov Specialization 84
3.3 Dependence Manipulation 90
3.3.1 The Theoretical Reality 90
3.3.2 A Copula Approach 95
3.3.3 Manipulation at Large 108

4 Conclusion **121**

Appendices **125**

Notation

$\mathbb{R}$	the real number
$\mathbb{R}_{>0}$	the positive real number
$\mathbb{R}_{\geqslant 0}$	the non-negative real number
$\mathbb{N}$	the natural number
$\mathbb{C}$	the complex number
$\boldsymbol{H}^*$	the conjugate transpose of matrix $\boldsymbol{H}$
$\boldsymbol{I}$	the identity matrix
$\mathbb{P}$	the probability measure
$\mathbb{E}$	the mean
$\widetilde{\mathbb{P}}$	the new probability measure after change of measure
$\widetilde{\mathbb{E}}$	the mean for the new probability measure
$\mathbb{E}[X; A]$	$\mathbb{E}[X; A] = \mathbb{E}[X 1_A]$, where 1_A is the indicator of A
$\widehat{F}$	moment generating function
$N\big(\mu, \sigma^2\big)$	Gaussian random variable
$\mathcal{CN}\big(0, \sigma^2\big)$	circularly symmetric complex Gaussian random variable, with independent real and imaginary parts $N\big(0, \sigma^2/2\big)$
$\overset{d}{=}$	equal in distribution
$\leqslant_{\mathcal{F}}$	the integral stochastic order with generator $\mathcal{F}$
$\implies$	$A \implies B$ means A implies B
$\impliedby$	$A \impliedby B$ means A is implied by B
$\iff$	$A \iff B$ means A is equivalent to B

Chapter 1

Introduction

In real world, the stochastic dependence corresponds to the probabilistic interrelationship of the system states through time and space, and different forms of dependence result in different system performances. In mathematics, the stochastic dependence is a property of the dependent elements, specified by the probability measure, and independence is a special case with a product measure of probability. The dependence scenario, which is probably uncertain or is intractable to get an explicit mathematical expression, raises additional analytical issues that differ from the independence scenario. Considering the diverse characteristics and distinguishing effects of the stochastic dependence, it is intriguing to study how to control the dependence in a system in order to obtain an improved performance.

In this chapter, we introduce the stochastic dependence concept and sketch the dependence control theory. Specifically, we treat the dependence as a physical reality, we show how the idea of dependence control arises from the mathematical description of the dependence phenomena, we explain why the dependence is a tradable resource, and we elaborate on the three principles of dependence control. In addition, the mathematical methods to build the dependence control theory are discussed.

1.1 Dependence in Perspective

The dependence exists universally in physics, finance, and engineering. For example, the particles moving in a fluid is described by the Brownian motion [74][126], the stock and bond are correlated in finance [13][113], and the Ethernet traffic is self-similar exhibiting long-range dependence [84]. Particularly, we elaborate on the dependence in the wireless communication system.

Wireless communication has been around for over a hundred years, starting with Marconi's successful demonstration of wireless telegraphy in 1896 and transmission of the first wireless signals across the Atlantic in 1901 [101]. Since 1G in around 1980s [101], the cellular system carries on upgrading every decade, and the 5G in 2020s is supposed to advance mobile from largely a set of technologies connecting people-to-people and people-to-information to a unified connectivity fabric connecting people to everything, which endows 5G with the potential for thrusting mobile technology into the exclusive realm of general purpose technology [19], like electricity and automobile.

It has become a trend that a new generation of wireless systems is deployed every new decade and the theme of each generation is to increase the capacity and spectral efficiency of wireless channels. The trend is driven by the explosion of wireless traffic that is a rough reflection of people's demand on wireless communication, and the paradox of supply and demand [59] is kept relieving generation by generation through exploiting the physical resources, i.e., power, diversity, and degree of freedom [138]. Considering the trillions of devices to be connected to the wireless network, the high capacity demand, and the stringent latency requirement in the coming 5G and beyond [4], it is imperative to rethink the wireless channel resources. In the affirmative, we propose that the stochastic dependence is a new resource to exploit.

1.1.1 Dependence Here and There

We remark that the wireless communication system encompasses a series of physical parameters, deterministic or random, and the dependence in the wireless system is rooted in the dependence of these random parameters, both the spatial dependence and the temporal dependence. We show the dependence phenomena in the wireless channel fading and channel coding.

The wireless signals are electromagnetic radiations and the signal propagation environment is a passive medium with dissipation that is the loss of field energy due to absorption, and dispersion that is the variation of the refractive index in the medium [76][108][67]. The dissipation causes the energy loss of the signals on the path from the transmitter to the receiver [108]. This effect is termed the large-scale fading [118]. The dispersion causes the reflection, diffraction, and scattering of the transmitted signals [108], which result in the multipath interference and the Doppler shift, due to the mobility of the scatters or terminals, of the received signals. This effect is termed the small-scale fading [118]. As a characterization of the propagation channel, the channel gain is defined by the ratio of of the receiver-to-transmitter power, of which the reciprocal is defined as the channel loss. As a result of the energy conservation law, the channel gain is less than one or the channel loss is greater than one.

In a wireless propagation channel, the gain of received signal is correlated with the angle of arrival of the signal, because of the interference of the multiple signals that are dispersed through the air from the transmitter to the receiver. This effect is termed the spatial correlation, which is characterized by the channel gain matrix with dependent elements. The spatial correlation depends on both the scatter characteristics and the antenna parameters [128], particularly, the spatial correlation increases as the antenna distance decreases. The influences of the spatial dependence at the transmitter side are connected with the channel knowledge [72], i.e., the ergodic capacity decreases with the correlation between the transmitter antennas in the case of full or no channel side information, which also holds for the antenna correlation at the receiver side [27], and the ergodic capacity increase with the transmitter antenna correlation in case the covariance matrix is known. In addition, the similar impact of channel knowledge and spatial correlation on the symbol error rate is shown in [16].

On the other hand, the fading elements also bears temporal dependence, due to the temporal correlation of the signal strength or the propagation environment. The temporal dependence influences the variance of the partial sum of wireless channel capacity through time, i.e., a stronger dependence implies a greater variance, which further influences the channel performance, e.g., the latency and buffer size. In the wireless literature, the typical characterization of the temporal dependence is the autocorrelation function [109][95]. However, the autocorrelation concerns just the first-order and second order moments of the stochastic process [93], and the uncorrelation can not imply the stochastic independence. The temporal dependence is further studied in this work.

In addition, the multipath effect causes the intersymbol interference [103], due to the nonlinear frequency response of the wireless channel, e.g., the time delay spread or the limited bandwidth. The intersymbol interference further causes memory in the channel, i.e., an output symbol depends on multiple input symbols. In information theory, the coding techniques introduces the stochastic dependence between the input letters and the dependence is generally necessary to achieve reliable transmission [50].

1.1.2 Dependence in Mathematics

We focus on the measure theoretic probability theory [78] and we model the events and random variables through the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. We classify the stochastic dependence into three types, i.e., independence, positive dependence, and negative dependence. Other types of classification are shown in [71], e.g., weak dependence and strong dependence.

The random vector $\boldsymbol{X} = (X_1, \ldots, X_N)$ has an independence structure, if the probability measure satisfies

$$\mathbb{P}(\boldsymbol{X} \leqslant \boldsymbol{x}) = \prod_{i=1}^{N} \mathbb{P}\left(X_i \leqslant x_i\right). \tag{1.1}$$

A proper way to define the positive dependence and negative dependence is to compare the probability measures through stochastic orders [122]. Specifically, for random vectors with the same marginals, $\boldsymbol{X}_\perp$, $\boldsymbol{X}_+$, and $\boldsymbol{X}_-$, with the random vector $\boldsymbol{X}_\perp$ has an independence structure, if

$$\mathbb{P}_{\boldsymbol{X}_\perp} \leqslant_{\mathcal{F}} \mathbb{P}_{\boldsymbol{X}_+}, \tag{1.2}$$

we say $\boldsymbol{X}_+$ has a positive dependence structure with respect to the stochastic order $\leqslant_{\mathcal{F}}$, while if

$$\mathbb{P}_{\boldsymbol{X}_-} \leqslant_{\mathcal{F}} \mathbb{P}_{\boldsymbol{X}_\perp}, \tag{1.3}$$

we say $\boldsymbol{X}_-$ has a negative dependence structure with respect to the stochastic order $\leqslant_{\mathcal{F}}$. In view of the strength of dependence [36], we have the weak (or strong) positive (or negative) dependence. Explicit definitions of positive or negative dependence concepts are elaborated in [98][26][63][81][123]. Intuitively, the independence implies that the occurrence of one random variable doe not influence the occurrences of other random variables, the positive dependence implies that large or small values of random variables tend to occur together, and the negative dependence implies that large values of one variable tend to occur together with small values of others [36].

The negative dependence and the positive dependence are reversely symmetric for the two dimension random vector, but in general, the negative dependence is not a mirror reflection of the positive dependence [14]. Particularly, while the comonotonicity is agreed upon as the extreme positive dependence for random vectors of arbitrary dimensions, there are diverse notions of extreme negative dependence for random vectors of more than two dimensions [114].

On the other hand, the independence is of significant importance to the probability theory, e.g., the focus of the probabilists of the first half of the twentieth century was mostly on the study of the sums of independent random variables, the corresponding limit distributions, aside from the foundations of probability [92]. In classical (commutative) probability, there is only one definition of independence, in non-commutative probability, there are many concepts of independence and there can be more with a relaxed regulation on the axioms for being independence [99][49]. A unification of the different independence is proposed in the filtered probability [85].

1.1.3 Dependence as Physical Resource

We consider the physical world as the physical realm, e.g., the wireless communication channel, and the mathematical model as the mathematical realm, e.g., the probabilistic definition of dependence. We regard the stochastic dependence as a physical reality as well as a mathematical regulation. In parallel with taking advantage of the dependence information in the mathematical analysis, we propose to control the dependence in the physical system to improve the system performance. The dependence control in the physical realm is based on the corresponding analytical results in the mathematical realm. The correspondences between the mathematical realm and the physical realm are elaborated as follows.

• The differentiation of the physical and mathematical realms indicates that we can utilize the dependence as a physical resource as well as a mathematical property. The existence of both the uncontrollable and controllable random parameters in a physical system indicates that it is feasible to control the dependence in the system by manipulating the dependence in the controllable random parameters, because the system dependence is influenced by both the uncontrollable and controllable random parameters and is transformable from one form to another by inducing a different form of dependence through the controllable random parameters, e.g., transforming from positive dependence to negative dependence by inducing the negative dependence.

• The mathematics provides a way to describe the dependence phenomena and suggests approaches to utilize the dependence resource. Specifically, the stochastic order provides an approach to compare the dependence influences, on the other hand, the mathematical property of the stochastic orders with respect to different dependence forms explains the advantage of one form of dependence over another for system performance improvement in practice. For example, the mathematical property of the increasing convex order of the partial sum of negatively dependent random variables indicates that taking the advantage of negative dependence attains a better performance and preserves other resources in the physical world. In addition, the strength of dependence manipulation in the physical system corresponds to the reflexivity, transitivity, and antisymmetry of the stochastic orders in mathematics.

• The mathematics is a description of the physical world in a sense to show that the physical world behaves like the mathematical description, on the other hand, it is interesting to treat the mathematical description as a reality as well and engineer the physical world to behave in the way of the mathematical description. For example, the topology is a mathematical theory about space, while it is becoming exciting to build the physical systems that possess the topological properties, like using the quasiparticles in the topological materials to encode the quantum bits [20]. Thus, there are two types of reality, the natural reality and the artificial reality (or the fundamental reality and the emergent reality), which coexist on earth, like the lake and dam, and in return bolster the degree of freedom of engineering.

1.2 Essence of Dependence Control

We briefly introduce the theory of dependence control that is concerned with transforming the dependence structures of the stochastic processes in the system through dependence manipulation in order to improve the system performance, e.g., the backlog and delay of a queueing system. We provide a set of results with respect to the theory and application of dependence control. These are the light-tail behavior of the wireless channel capacity, which provides the basis for applying dependence control to the wireless communication system, the tradability of dependence, which is about the utility of dependence resource in the stochastic process, and the three principles of dependence control, which are about the dependence mechanics and dependence manipulation in the queueing system.

1.2.1 The Light-Tail Property

Consider the multiple-input-multiple-output channel model that is expressed as [138]

$$\mathbf{y}(t) = \boldsymbol{H}(t)\mathbf{x}(t) + \mathbf{w}(t),\ t \in \mathbb{N}, \tag{1.4}$$

where $\mathbf{x}(t) \in \mathbb{C}^{N_T}$, $\mathbf{y}(t) \in \mathbb{C}^{N_R}$, $N_T \in \mathbb{N}$, $N_R \in \mathbb{N}$, $\mathbf{w}(t) \sim \mathcal{CN}(0, N_0\boldsymbol{I}_{N_R})$, and $\boldsymbol{H}(t) \in \mathbb{C}^{N_R \times N_T}$ is the channel gain matrix. For simplification, we omit the time index. The instantaneous channel capacity $c \in \mathbb{R}$ is defined by the mutual information, which is a function $f : \mathbb{R} \times \mathbb{C}^{N_R \times N_R} \to \mathbb{R}$ of the product of the transmission power p and the channel matrix $\boldsymbol{H}\boldsymbol{H}^*$, i.e.,

$$f : p\boldsymbol{H}\boldsymbol{H}^* \mapsto c, \tag{1.5}$$

where we treat the instantaneous power as a random variable. Specifically, if the tail distribution function satisfies [7] $\overline{F}_X(x) = O\big(e^{-\theta x}\big)$, $\exists\theta > 0$, where $f(x) = O(g(x)) \iff \limsup_{x\to\infty} \frac{f(x)}{g(x)} < \infty$, equivalently, $\mathbb{E}\big[e^{\theta X}\big] < \infty, \exists\theta > 0$, then the distribution is light-tailed; otherwise, it is heavy-tailed. The heavy-tailed distribution indicates that extreme values occur with a relatively high probability [45]. Particularly, if the tail is super-heavy, it has no finite moments [57], e.g., the distributions with slowly varying tails. The class of slowly varying functions includes constants, logarithms, iterated logarithms, powers of logarithms [33].

We obtain that the sufficient condition for the light-tail wireless channel capacity is the existence of the mean value of the power law of the product of the random power and the maximum eigenvalue of the channel matrix, i.e.,

$$\overline{F}_c(x) = O\Big(e^{-\theta x}\Big),\ \exists\theta > 0 \impliedby \mathbb{E}\Big[(p\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0, \tag{1.6}$$

where $\lambda_{\max}$ is the maximum eigenvalue of $\boldsymbol{H}\boldsymbol{H}^*$ and the right hand side is equivalent to $\mathbb{E}\Big[(p\,\mathrm{Tr}\,[\boldsymbol{H}\boldsymbol{H}^*])^{\theta}\Big] < \infty,\ \exists\theta > 0$, where Tr denotes the trace of a square matrix, in terms of the tail behavior, they are equivalently expressed as $\overline{F}_{p\lambda_{\max}}(x) = O\big(x^{-\theta}\big),\ \exists\theta > 0$, and $\overline{F}_{p\,\mathrm{Tr}\,[\boldsymbol{H}\boldsymbol{H}^*]}(x) = O\big(x^{-\theta}\big),\ \exists\theta > 0$. Specifically, $p = 1$ corresponds to the deterministic power scenario. In addition, for the broadband channel scenario, the channel matrix is the diagonal matrix of each sub-channel matrices, i.e., $\boldsymbol{\mathcal{H}} = \mathrm{diag}\{\boldsymbol{H}_1, \ldots, \boldsymbol{H}_N\}$.

We observe that, for the typical stochastic channel models and the power supply systems in practice, the distribution of the capacity, which is a logarithm function of the product of the fading effects and random power, is light-tailed, because the logarithm function transforms a less than super-heavy-tailed distribution to a light-tailed distribution. The detailed explanations are as follows.

• The restriction that the passive channel gain is less than one excludes the existence of fading models with super-heavy tails. It is interesting to note that the typical large-scale fading distribution is heavy-tailed, e.g., the Lognormal distribution, while the typical small-scale fading distribution is light-tailed, e.g., the Rayleigh, Rice, and Nakagami distributions. Specifically, if a random variable is lognormal, then its reciprocal is also lognormal. The tail property indicates that the large-scale fading effects, like path loss and shadowing, are more likely to cause large values of both channel loss and gain, which may be due to the large shadow dynamics in the propagation environment; while the small-scale fading effects, like the multipath interference and Doopler shift, are less likely to cause large values of channel gain or the random values are more likely to be concentrated around the mean. Since both light-tailed and heavy-tailed distributions with finite mean are used to model the channel gain, the parametric distributions that can model both heavy-tailed and light-tailed distributions are of interest, e.g., the Weibull distribution [124][111]. These theoretical insights on the stochastic models match the empirical results [67]. In addition, since the random variables in the stochastic models, whether the light-tailed distribution or the heavy-tailed distribution, are unbounded, the stochastic models of the wireless channels are strictly not passive systems [76][89], because of the violation of the energy conservation law.

• Though the wireless system can be energy unlimited [90], the transmission power is unlikely to have an infinite mean, thus, the tail of the power distribution is lighter than the super-heavy distribution. When there are active relays in the wireless channels, the whole channel gain is the product of each individual channel gain. However, the tail of the product distribution can be asymptotically bounded above and below by the tail of a dominating random variable of the product for both independence and dependence scenarios [145][22][69][144]. In addition, the gain saturation also exclude the possibility of unlimited gain in active medium [102]. Thus, the whole channel gain is more likely to have a tail behavior lighter than the super-heavy tail. On the other hand, when the power in the capacity formula is set to be deterministic, e.g., the mean value of power, normalization is usually considered for the channel matrix. Specifically, if the channel description is based on the average transmitter power P_T [110], then, the channel matrix $\boldsymbol{H}$ is non-normalized; and if the description uses the average receiver power P_R, then the channel matrix $\overline{\boldsymbol{H}}$ is normalized [47][135]. Mathematically, the relationship is expressed as [47] $P_T^{1/2} \cdot \boldsymbol{H} = P_R^{1/2} \cdot \overline{\boldsymbol{H}}$. For example, the normalized channel gain of the Rayleigh fading channel is [47][138] $\overline{H}_{ij} \sim \mathcal{CN}(0,1)$ and $\mathbb{E}\left[\overline{H}_{ij}\overline{H}_{ij}^*\right] = 1$. The normalization indicates that the mean values of the matrix identities exist, which excludes the existence of the fading models with super-heavy tails.

1.2.2 The Dependence Market

We regard the stochastic process as a functional of random parameter processes, which are either uncontrollable or controllable, i.e., the stochastic process as a function of a set of random parameters, each of which is itself a stochastic process. We specify that the cardinality of the parameter set $\left(X_t^1, X_t^2, \ldots, X_t^n\right)$ is time-invariant and the function $f_t : \mathbb{R}^n \to \mathbb{R}$ is time-variant, i.e.,

$$X_t = f_t\left(X_t^1, X_t^2, \ldots, X_t^n\right). \tag{1.7}$$

In other words, we treat the stochastic process as a functional of a multivariate stochastic process and the functional maps the multivariate stochastic process to a univariate stochastic process. This functional specification is extensible to the general stochastic process on the Polish space. For example, in the wireless channel capacity, the uncontrollable parameters represent the property of the environment that can not be interfered, e.g., fading, and the controllable parameters represent the configurable property of the wireless system, e.g., power. In addition, this functional perspective is useful for studying the dependence impact of an individual arrival process on the aggregation of a set of multiplexed arrival processes.

We study how to transform the dependence in the functional process $\{X_t\}$, by manipulating the dependence in parameter processes $\{X_t^i\}$, $1 \leqslant i \leqslant n$. There are two ways to implement this dependence transform, i.e., one by transforming the dependence structure from the positive dependence to the negative dependence, and the other by transforming the marginal distributions. We highlight the following results, which provide insights for manipulating the dependence.

• The dependence is a resource that can be traded off, i.e., when the dependence is utilized, another form of resource can be saved, e.g., more amounts of negative dependence can exchange for less amounts of mean values. The chain relation, $\boldsymbol{X} \leqslant_{sm} \widetilde{\boldsymbol{X}} \implies \sum_{j=1}^{t} X_j \leqslant_{cx} \sum_{j=1}^{t} \widetilde{X}_j \implies \mathbb{E}\sum_{j=1}^{t} X_j = \mathbb{E}\sum_{j=1}^{t} \widetilde{X}_j$, means the supermodular order of the dependence structures implies the convex order of the variability of the partial sum with equal mean. To take into account both the mean and the variability, we use the increasing convex order for further elaboration. Specifically, the mean and the variability are exchangeable for each other, i.e., if the variability is relatively small, then a relatively greater mean can be tolerated while satisfying the increasing convex order, vice versa. The mathematical expressions are as follows, if $X \leqslant_{icx} Y$ and $\mathbb{E}X \leqslant \mathbb{E}Z' \leqslant \mathbb{E}Y$, then it is possible that $Z' \leqslant_{icx} Y$, because we have $X \leqslant_{icx} Y \iff X \leqslant_{st} Z \leqslant_{cx} Y$ [127]; and if $X \leqslant_{icx} Y$, then $X \leqslant_{cx} Z' \leqslant_{st} Y$ such that $Z' \leqslant_{icx} Y$, because we have $X \leqslant_{icx} Y \iff X \leqslant_{cx} Z \leqslant_{st} Y$ [127]. Complementary results hold in the sense of the increasing concave order [127].

• The manipulation of the marginal distributions has a dependence bias, while the manipulation of the dependence structure fixing the marginals has no such dependence bias. Specifically, the dependence bias means that, if a parameter process bears negative dependence, then the manipulation of each individual marginals with respect to the (increasing) convex order can not lead effectively to the (increasing) convex order of the partial sums, i.e., the (increasing) convex order of the marginals implies the (increasing) convex order of the partial sum holds for positive dependence and not for negative dependence [98]. The dependence bias of the marginals provides an opportunity for dependence control. Specifically, the dependence bias means that the increasing convex order of the partial sum is insensitive to the marginal manipulation of the parameter process with negative dependence, i.e., the increasing convex order still holds for a partial sum with smaller mean values of the marginals. For example, a better queueing system performance, in terms of backlog and delay, can be achieved in the scenario of negative dependence in the processes, even with a smaller mean value of the service process or a greater mean value of the arrival process.

1.2.3 The Three Principles

We consider a queueing system in the discrete-time setting, with the arrival process $a(t)$ and the service process $c(t)$, the instantaneous backlog in the system $B(t)$ is expressed as [21]

$$B(t+1) = [B(t) + X(t)]^+ , \tag{1.8}$$

where $[\cdot]^+ = \max(\cdot, 0)$ and $X(t) = a(t) - c(t)$ denotes the difference of the instantaneous arrival amount and the service amount. For a lossless system, the cumulative output $A^*(t) = A(t) - B(t)$ is the difference between the cumulative input $A(t) = \sum_{s=0}^{t} a(s)$ and backlog $B(t)$, and the delay is defined via the input-output relationship [29],

$$D(t) = \inf \{d \geqslant 0 : A(t-d) \leqslant A^*(t)\} , \tag{1.9}$$

which is the virtual delay that a hypothetical arrival has experienced on departure.

We formulate three principles of dependence control, namely measurability, duality, and transformability. Synthetically, the measurability talks about the performance measures for a queueing system, the duality talks about the impact consistency of the dependence of the arrival and service processes on the system performance, and the transformability talks about the dependence property of a stochastic process, e.g., the arrival process or the service process. The three principles are expounded as follows.

1. **Measurability.** *The asymptotic decay rate of the tail of delay or backlog is able to identify and quantify the dependence influence in the stochastic processes of the queueing system.*

Letting Z denote the backlog or delay, we prove that the decay rates of their tail distributions converge exactly to two respective constant values, i.e.,

$$\lim_{z\to\infty}\frac{1}{z}\log\mathbb{P}(Z>z)=-\gamma^*, \tag{1.10}$$

where $\gamma^*>0$, for light-tailed arrival and service processes with weak forms of dependence. We define the logarithmic asymptotic decay rates of the tails of backlog and delay as the measure identities and show that the measure identities, conditional on their existence, have a monotonic relationship with the dependence of the stochastic processes in the queueing system. With a manipulation, the asymptotic expression is equivalently written as $\lim_{z\to\infty}\frac{\log\mathbb{P}(Z>z)}{\log e^{-z\gamma^*}}=1$, which shows that the measure identities are the logarithmic asymptotics and capture only the dominant term in an asymptotic expression [7].

2. **Duality.** *The arrival process and the service process have a dual potency of transforming the dependence in the queue increment process, which further influences the system performance.*

Letting $\boldsymbol{X}=(X_t:t\in\mathbb{N})$ (also $\widetilde{\boldsymbol{X}}=(\widetilde{X}_t:t\in\mathbb{N})$) be the arrival process or the service process, and fixing one of the two processes and changing the other, the duality result is expressed as

$$\boldsymbol{X} \leqslant_{sm} \widetilde{\boldsymbol{X}} \implies \mathfrak{S}(t) \leqslant_{cx} \widetilde{\mathfrak{S}}(t) \overset{\forall t}{\implies} \gamma^* \geqslant \widetilde{\gamma}^*, \tag{1.11}$$

where $\leqslant_{sm}$ and $\leqslant_{cx}$ denote respectively the multivariate supermodular order and univariate convex order [127], $A \overset{C}{\implies} B$ denotes A implies B conditional on C, and $\mathfrak{S}(t) = A(t) - S(t)$, $\widetilde{\mathfrak{S}}(t) = A(t) - \widetilde{S}(t)$ or $\widetilde{\mathfrak{S}}(t) = \widetilde{A}(t) - S(t)$. The dual potency of arrival and service dependence indicates that if the dependence manipulation in the arrival process is not available, we can transfer to the dependence manipulation in the service process, vice versa. The supermodular order entails that the marginals on both sides of the inequality are identical, thus the influences are solely due to the dependence structure. Considering the influences of both dependence structure and marginals, a sufficient condition for the ordering of the measure identities for the arrival process is the increasing convex ordering $\sum_{j=1}^{t} X_j \leqslant_{icx} \sum_{j=1}^{t} \widetilde{X}_j$ and a sufficient condition for the service process is the increasing concave ordering $\sum_{j=1}^{t} X_j \geqslant_{icv} \sum_{j=1}^{t} \widetilde{X}_j$. This is coherent with the intuition that a smaller and less variable arrival process or a greater and less variable service process leads to a better system performance in terms of the backlog and delay.

3. **Transformability.** *The manipulation of the free dimensions of a stochastic process is able to transform the dependence of the process.*

For a stochastic process as a functional of uncontrollable or controllable random parameters, i.e., $X_t = f_t\left(X_t^1, X_t^2, \ldots, X_t^n\right)$, we specify that the dimension of the parameter set $\left(X_t^1, X_t^2, \ldots, X_t^n\right)$ is time-invariant and the function $f_t : \mathbb{R}^n \to \mathbb{R}$ is time-variant and is increasing or decreasing at X_t^i for all the time. We prove that the dependence in such a stochastic process is transformable from strong dependence to weak dependence in the sense of supermodular order, e.g., from positive dependence to independence or negative dependence, by manipulating the dependence in the controllable parameters, i.e.,

$$(X_1^i, X_2^i, \ldots, X_t^i) \leqslant_{sm} (\widetilde{X}_1^i, \widetilde{X}_2^i, \ldots, \widetilde{X}_t^i),\ \exists 1 \leqslant i \leqslant n$$
$$\Longrightarrow (X_1, X_2, \ldots, X_t) \leqslant_{sm} (\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_t), \quad (1.12)$$

where $\widetilde{X}_j = f_j(X_j^1, \ldots, X_j^{i-1}, \widetilde{X}_j^i, X_j^{i+1}, \ldots, X_j^n)$, $\forall 1 \leqslant j \leqslant t$. Considering the influences of both dependence structure and marginals, we prove the transformability with respect to the (increasing) directional convex order. Specifically, we show that the random parameters in the wireless channel capacity, the sub-channels of a compound wireless channel, and the random multiplexing mechanism of an aggregated arrival process, provide a chance to perform dependence manipulation in practice.

1.3 A Note on Methodologies

The construction of the dependence control theory is based on a series of mathematical techniques. We discuss the choice of the techniques and the related work of similar ideas.

1.3.1 The Methodology

Analytically, to build the dependence control theory, we adopt a few mathematical tools, which are necessary to describe different aspects of the dependence mechanics and dependence manipulation. Specifically, large deviation is used to find the measure identities, change of measure is used to explain the dual potency of arrival and service dependence, and stochastic order is used to prove the dependence transformability. In addition, the random matrix theory is used to study the tail property of the wireless channel capacity. The structure of the mathematical analysis and the reason for choosing the mathematical techniques are explicated as follows.

- We prove the measurability and duality principles integratively, due to the close relationship between the performance measures and the ordering of the measure identities. The measure identities are the logarithmic asymptotic decay rates of the performance measure distributions and are supposed to give an exact reflection of the strength of the stochastic dependence in the arrival process and the service process. Specifically, the change of measure is used to derive an upper bound and a lower bound of the identities through calculating the probability in a new probability space. The bounds are then proved to be asymptotically equal. The approach to derive the upper and lower bounds eases the difficulty of finding the exact value directly.

• We prove the dependence transformability by using the stochastic orders to compare the dependence between the original stochastic process and the stochastic process after dependence manipulation, instead of directly studying the impact of dependence manipulation on the measure identities. This approach allows to separate the transformability principle from the other two principles, with analytical flexibility and independent significance. For instance, it indicates that the transformability applies to the general stochastic processes beyond the arrival and service processes, and different stochastic orders indicate different dependence manipulation techniques.

• We build the connections between the tail of wireless channel capacity and the fading and power distributions, by utilizing the tail distributions to characterize the wireless channel property. Specifically, the random matrix is used to represent the multiple-input-multiple-output channel, of which the random scalar is a special case to represent the single-input-single-out channel, and the light-tail property of wireless channel capacity is shown to be determined by the maximum eigenvalue and the trace of the random matrix. This reaffirms the universality of the eigenvalue of the random matrix in the case of wireless channels. To further support the results, the asymptotic tail behaviors of the sum and product of the random variables are investigated, particularly, the regular variation, slow variation, and exponential variation of the tail distributions are studied.

As an application of the dependence control theory and as a justification to the assumptions in this work, we apply the theory to Markov additive process, which is capable of characterizing a large class of arrival processes, is versatile in capturing the dependence in the service processes, and is able to reflect the non-stationarity in the mobile wireless channels, and focus on the dependence manipulation of the wireless channel capacity. Particularly, martingale is used to derive the non-asymptotic tail bounds of the performance measures, and copula is used to represent the Markov property and the no-Granger causality, and is revived as a dependence manipulation technique. The application results are described as follows.

- For the performance measures of the queueing system, we derive the non-asymptotic and time-dependent tail probabilities of delay and backlog for Markov additive arrival process and Markov additive service process. The decay rates of the non-asymptotic results sufficiently imply the logarithmic asymptotic decay rates, on the other hand, the non-asymptotic results provide an overview of the tail behavior of performance measures in both the finite time regime and the infinite time regime. The analysis extends the single Markov additive process model in [7] to the double Markov additive process model.

- For the random parameters in the wireless channel, we treat the wireless fading as the uncontrollable random parameter, which is the inherent property of environment that can not be interfered, and we treat the power as the controllable random parameter, which remains constant during a coherence period and randomly fluctuates through different coherence periods. Specifically, the purpose of the random power is to induce negative dependence to the corresponding wireless channel capacity, and the new power allocation scheme based on the dependence control principle avoids the requirement of the channel side information, which is necessary in the traditional power allocation scheme [138].

• For the dependence manipulation of the stochastic process, we develop a copula manipulation technique for Markov process and use simulation to validate this technique. We model the random parameters as a multivariate Markov process. We use no-Granger causality [25] to model the relationship between the uncontrollable and controllable parameters, and provide sufficient and necessary condition of the no-Granger causality for Markov process. The no-Granger causality guarantees that if the random parameters form a Markov process then the uncontrollable parameters and controllable parameters each form a Markov process. This addresses the challenge in dependence control that the specific characteristics of the stochastic process under control must be known.

1.3.2 Related Methodologies

The research works on dependence have been focusing on the idea of dependence modeling [36], based on which the dependence influences are measured and compared [98][36][123]. For example, copula is used to model the dependence among multiple risks in actuarial theory [36] and the dependence among multiple arrival processes in stochastic network calculus [38]. The system model provides a domain of the problem and the dependence model provides a way to characterize the form of dependence. The dependence control advocates that the dependence in a certain system can not only be taken advantage of in a passive way due to the uncontrollable parameters but also be manipulated in an active way based on the controllable parameters and transformed from one form to another in order to improve the system performance.

A related concept is the dependence decoupling [34]. The decoupling reduces the mathematical problems on dependent variables to the problems on independent variables [34], e.g., through inequalities, thus the mathematical techniques for the independent random variable can be used for further analysis. The decoupling focuses on reducing the dependence situation to the independence situation in the mathematical realm, while the dependence control focuses on transforming the dependence structure of the stochastic processes in the physical realm. The dependence control treats the dependence as a physical resource that can be exploited for better performance, the dependence forms are classified beyond the independence, and the utility of different dependence forms are discussed. On the other hand, there are some overlapping between these two methodologies, especially in the sense of mathematical techniques that are used in the mathematical analysis, e.g., the inequalities. Thus, it is reciprocal to take advantage of the mathematical techniques in each methodology.

The dependence is taken advantage of in stochastic simulation [8]. For example, the antithetic variates, control variates, and common random numbers are used for variance reduction. Specifically, the antithetic variates, which are negatively dependent, are used to drive different runs of a simulation experiment [82]; the common random numbers, of which the functions are positively dependent [37], are used to drive the simulation in order to guarantee the similar experimental conditions when comparing different experiment configurations [82]; and both negative and positive dependence are used in the control variates [8]. It is worth noting that the simulation environment is completely controllable, hence the random processes in the simulation can be arbitrarily specified and interfered. On the other hand, the real system is more complex with both controllable and uncontrollable random parameters, and the dependence control in the real system manipulates the dependence structure of the controllable random parameter process in order to control the system performance that is a function of both the controllable and uncontrollable random parameters.

In finance and economics, the stock and bond have been either positively or negatively correlated, e.g., the correlation has turned from positive in the 1970s-1990s to negative in the 2000s-2010s [13][113]. In risk management [91], the correlation is useful for portfolio construction by creating diversified portfolios that can withstand market volatility and smooth out portfolio returns, e.g., the bond can be used to diversify against the stock. Specifically, the portfolio manager can use the negatively correlated assets to diversify the risk of a portfolio or hedge the portfolio to reduce the risk. Technically, the hedging requires a highly negative correlation and the diversification requires a correlation that is not highly positive.

Chapter 2

A Tale of Tails

An information theoretic measure of the wireless channel is the channel capacity, which defines the maximum transmission rate with arbitrarily small error probability. Since the wireless channel is time variant, the instantaneous capacity randomly fluctuates through time, in other words, the wireless channel capacity is a stochastic process, which brings about diverse features to the wireless channel and entails more measures to characterize the fundamental property of this stochastic process, e.g., the distributions.

In this chapter, we show that the tail distribution of wireless channel capacity is light-tailed. A simple explanation is that the capacity is a logarithm function of some random variables, so long as these random variables are not heavier than fat tails, the capacity is light-tailed. This property is fundamental as it holds for all typical wireless channel models, e.g., the Rayleigh, Rice, Nakagami, and Lognormal fading channels. Moreover, this property is extended from frequency-flat to frequency-selective fading channels, from instantaneous to cumulative time regimes, from single-hop to multiple-hop scenarios, and from single-input-single-output to multiple-input-multiple-output channels.

2.1 Single-Input-Single-Output Channel

2.1.1 Channel Capacity

We introduce the basic concepts of wireless channel capacity, including the ergodic capacity, instantaneous capacity, cumulative capacity, and transient capacity.

We consider the single-input-single-output channel with additive white Gaussian noise (AWGN). The complex baseband representation for a flat fading channel is [138]

$$y(t) = h(t)x(t) + n(t),\ t \in \mathbb{N}, \tag{2.1}$$

where $x(t)$ is the input, $y(t)$ is the output, $h(t)$ is the fading process, and $n(t) \sim \mathcal{CN}(0, N_0)$ is the noise process. Conditional on a realization of the fading process $h(t)$, the mutual information is expressed as [135]

$$I(X;Y|h(t)) = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \mathbb{P}(x,y|h(t)) \log_2 \frac{\mathbb{P}(x,y|h(t))}{\mathbb{P}(x|h(t))\mathbb{P}(y|h(t))}, \tag{2.2}$$

where X and Y are input and output random variables with alphabets $\mathcal{X}$ and $\mathcal{Y}$. The maximum mutual information over input distribution at t, denoted as $c(t)$, is defined as instantaneous capacity [30]:

$$c(t) = \max_{p(x)} I(X;Y|h(t)), \tag{2.3}$$

where the maximum is taken over all possible input distributions $p(x) = \mathbb{P}\{X = x\}$, $x \in \mathcal{X}$. Specifically, if the channel side information is only known at the receiver, the instantaneous capacity is expressed as [138]

$$c(t) = W \log_2 \left(1 + \gamma|h(t)|^2\right), \tag{2.4}$$

where $|h(t)|$ denotes the envelope of $h(t)$, $\gamma = \frac{P}{N_0 W}$ denotes the average received SNR per complex degree of freedom, P denotes the average transmission power per complex symbol, $\frac{N_0}{2}$ denotes the power spectral density of AWGN, and W denotes the channel bandwidth.

For a stationary process of instantaneous capacity, the average over the probability space is defined as ergodic capacity [135]:

$$\overline{c} = \mathbb{E}[c(t)]. \tag{2.5}$$

The definition implies that the ergodic capacity is a constant and is a concept for infinite code length in infinite time regime, i.e., it defines the maximum transmission rate of the channel with asymptotically small error probability for the code with sufficiently long length such that the received codewords is affected by all fading states [55].

To account for finite time regimes, the sum of instantaneous capacity over a time period $(s,t]$, denoted as $S(s,t)$, is defined as cumulative capacity:

$$S(s,t) = \sum_{i=s+1}^{t} c(i). \tag{2.6}$$

For $S(0,t)$, we use $S(t)$ as simplification. The time average of the cumulative capacity through $(0,t]$ is defined as transient capacity:

$$\overline{c}(t) = \frac{S(t)}{t}. \tag{2.7}$$

The transient capacity is random, which essentially defines the achievable capacity for a code with finite length such that the received codewords only experience partial fading states [138].

The probabilistic average of the transient capacity in a stationary process is expressed as

$$\mathbb{E}\left[\overline{c}(t)\right] = \overline{c}, \tag{2.8}$$

where $\overline{c}$ is the ergodic capacity of the channel. According to the law of large numbers, the transient capacity converges to the ergodic capacity when time goes to infinity, i.e.,

$$\mathbb{P}\Big\{\lim_{t\to\infty} \overline{c}(t) = \overline{c}\Big\} = 1, \tag{2.9}$$

for independent and identically distributed instantaneous capacity. However, the dependence in capacity may be unknown, and a more general result for the transient capacity on finite time horizon is expressed by the Chebyshev inequality [106],

$$\mathbb{P}\{|\overline{c}(t) - \overline{c}| \geqslant x\} \leqslant \frac{\mathrm{Var}[\overline{c}(t)]}{x^2}, \tag{2.10}$$

which is a basic result of concentration [17]. It indicates that, in view of temporal behavior, statistical properties of the cumulative process should be taken into account besides the instantaneous capacity.

2.1.2 Light-Tail Behavior

A distribution is said to be light-tailed, if the tail $\overline{F}(x) = 1 - F(x)$ is exponentially bounded, i.e.,

$$\overline{F}(x) = O(e^{-\theta x}),\ \exists\theta > 0, \tag{2.11}$$

where $f(x) = O(g(x)) \iff \limsup\limits_{x\to\infty} \frac{f(x)}{g(x)} < \infty$; equivalently, it means the moment generating function $\widehat{F}[\theta] = \int e^{\theta x} F(dx)$ is finite for some $\theta > 0$. Otherwise, the distribution is said to be heavy-tailed [7, 119]. Specifically, if $\overline{F}(x) \sim x^{-\theta}$, $\theta > 0$, $f(x) \sim g(x) \iff \lim\limits_{x\to\infty} \frac{f(x)}{g(x)} = 1$, it is defined to be fat-tailed; if $\overline{F}(x) = O\left(x^{-\theta}\right)$, $\exists\theta > 0$, it is defined to be fat-tail bounded.

The following theorem gives the condition for the wireless channel capacity distribution to be light-tailed.

Theorem 1. *For flat fading, the instantaneous capacity is expressed as the logarithm transform of the instantaneous channel gain, i.e.,* $c(t) = W\log_2(1+\gamma h(t)^2)$, $\forall t$. *If the distribution of the fading process is fat-tail bounded, the distribution of the instantaneous capacity is light-tailed.*

Proof. For convenience, we omit the time index t and write $c = W\log_2(1+\gamma h^2)$. Correspondingly, the tail of the instantaneous capacity is a function of the tail of the channel gain, i.e.,

$$\overline{F}_c(x) = \overline{F}_h\left(\sqrt{\frac{2^{\frac{x}{W}}-1}{\gamma}}\right). \tag{2.12}$$

Let $r = \sqrt{\frac{2^{\frac{x}{W}}-1}{\gamma}}$, for some $\theta > 0$, $\overline{F}_c(x) = O(e^{-\theta x})$ entails

$$\overline{F}_h(r) = O\left(r^{-\theta}\right), \tag{2.13}$$

which completes the proof. □

The following corollary shows that the capacity distributions of the typical wireless fading channels are light-tailed.

Corollary 1. *If a wireless channel is Rayleigh, Rice, Nakagami-m, Weibull, or lognormal fading channel, its instantaneous capacity distribution is light-tailed.*

Proof. For Weibull fading channel, the tail of fading is expressed as

$$\overline{F}_h(r) = e^{-br^k}, \tag{2.14}$$

where $b > 0$ and $k > 0$ are constants. Applying Taylor's theorem to expend e^{br^k}, it is easily shown that, for some θ satisfying $k > \theta > 0$

$$\lim_{r\to\infty}\frac{e^{-br^k}}{r^{-\theta}} = \lim_{r\to\infty}\frac{r^\theta}{1+br^k+\dots} = 0. \tag{2.15}$$

This limit shows that though the Weibull distribution is heavy-tailed for $0 < k < 1$, it is lighter than the fat tail. Hence from Theorem 1, the instantaneous capacity under Weibull fading is light-tailed.

Rayleigh fading is a special case of Weibull fading with $k = 2$. The distribution of its instantaneous capacity is expressed as [62]

$$F(x) = 1 - e^{\frac{1-2^{\frac{x}{W}}}{\gamma}}. \tag{2.16}$$

It is trivial to show that the tail is exponentially bounded

$$\overline{F}(x) \leqslant e^{\frac{1}{\gamma}} e^{-\theta x}, \tag{2.17}$$

for $0 < \theta \leqslant \frac{1}{W\gamma} 2^{\frac{1}{\log 2}} \log 2$. Hence, the instantaneous capacity under Rayleigh fading is light-tailed.

For Rice fading channel, the tail of the instantaneous capacity is expressed as [117]

$$\overline{F}(x) = Q_1\left(\frac{s}{\sigma_0}, \frac{\sqrt{2^{x/W} - 1/\gamma_s s}}{{\sigma_0}^2}\right), \tag{2.18}$$

where W is the bandwidth, s the amplitude of the LOS (light of sight) component, ${\sigma_0}^2$ the variance of the underlying Gaussian process, and γ_s the average SNR. According to the exponential bound of the Marcum Q-function [129],

$$\alpha_F = \limsup_{x\to\infty} \frac{-\log \overline{F}(x)}{x} \geqslant \limsup_{x\to\infty} \frac{1}{2x}\left(\frac{\sqrt{2^{x/W} - 1/\gamma_s s}}{{\sigma_0}^2} - \frac{s}{\sigma_0}\right)^2 = \infty, \tag{2.19}$$

which means that the instantaneous capacity of a Rice fading channel is light-tailed [119][48].

For Nakagami-m fading channel [115], since the square of the Nakagami-m random variable follows a gamma distribution, which is light-tailed [7], the distribution of its instantaneous capacity is thus light-tailed.

For lognormal fading channel [116], since the lognormal distribution has all the moments, which means that it has a lighter tail than the fat-tailed distribution [57], the distribution of its instantaneous capacity is light-tailed. □

The rest of this subsection shows that the light-tailed property is extended from flat-fading to frequency-selective fading, from instantaneous to cumulative time regime, and from single-hop to multiple-hop scenarios.

Corollary 2. *For frequency-selective fading modeled by L parallel independent channels with the instantaneous capacity $c = \sum_{\ell=1}^{L} W_\ell \log_2(1 + \gamma h_\ell^2)$, if the distribution of the instantaneous capacity of each sub-channel $c_\ell = W_\ell \log_2(1+\gamma h_\ell^2)$ is light-tailed, so is the instantaneous capacity distribution of the frequency-selective fading channel.*

Proof. For this frequency-selective fading channel, its instantaneous capacity is by definition related to the instantaneous capacity of each sub-channel as

$$c = \sum_{\ell=1}^{L} c_\ell. \tag{2.20}$$

The tail of the distribution of the instantaneous capacity can then be expressed by [70]

$$\overline{F}_c(x) = 1 - F_{c_1} \circledast \ldots \circledast F_{c_L}(x) \tag{2.21}$$

$$\leqslant \overline{F}_{c_1} \oplus \ldots \oplus \overline{F}_{c_L}(x), \tag{2.22}$$

where $f \circledast g(x) = \int_{-\infty}^{\infty} f(x-y)dg(y)$ is the Stieltjes convolution and $f \oplus g(t) = \inf_{0 \leqslant s \leqslant t}\{f(s) + g(t-s)\}$ is the univariate min-plus convolution [12] or infimal convolution [123]. The first step results from sum of independent random variables, and the second step results from that the distribution of sum of independent random variables is upper bounded by the distribution of such a sum without dependence consideration [70]. As is illustrated in the proof of the next theorem, the latter is light-tailed. □

Corollary 3. *Consider a wireless channel, if the distribution of its instantaneous capacity at any time is light-tailed, the distribution of the cumulative capacity is light-tailed, and the distribution of the cumulative capacity of a concatenation of such wireless channels is light-tailed.*

Proof. Without considering any dependence constraint, the tail of the cumulative capacity, $S(t) = c(1) + \cdots + c(t)$, is bounded by [70]

$$\overline{F}_{S(t)}(x) \leqslant \overline{F}_{c(1)} \oplus \ldots \oplus \overline{F}_{c(t)}(x), \tag{2.23}$$

which is exactly the infimal convolution of the Fréchet upper bound [123]. If the instantaneous capacity is light tailed, i.e.,

$$\overline{F}_c(x) \leqslant ae^{-bx}, \tag{2.24}$$

applying a distribution bound for the sum of exponentially bounded random variables [70], the tail of the cumulative capacity is exponentially bounded, i.e.,

$$\overline{F}_{S(t)}(x) \leqslant \prod_{k=1}^{t} (a_k b_k w)^{\frac{1}{b_k w}} \cdot e^{\frac{-x}{w}}, \tag{2.25}$$

where $w = \sum_{k=1}^{t} \frac{1}{b_k}$.

For a concatenation of wireless channels, each with a cumulative capacity $S_i(s,t)$, the cumulative capacity process is essentially the service process of the channel, and the cumulative capacity of the concatenation channel is expressed as [70, 46]

$$S(s,t) = S_1 \otimes \ldots \otimes S_N(s,t), \tag{2.26}$$

where $f \otimes g(x) = \inf_{0 \leqslant y \leqslant x}\{f(y) + g(y,x)\}$ is the bivariate min-plus convolution [21]. Then, the tail is expressed as

$$
\begin{aligned}
\overline{F}_{S(t)}(x) &= \mathbb{P}\left\{S_1 \otimes \ldots \otimes S_N(t) \geqslant x\right\} && (2.27)\\
&= \mathbb{P}\left\{\inf_{\mathbf{u}\in\mathcal{U}(x)} \sum_{i=1}^{N} S_i(u_{i-1}, u_i) \geqslant x\right\} && (2.28)\\
&\leqslant \inf_{\mathbf{u}\in\mathcal{U}(x)} \mathbb{P}\left\{\sum_{i=1}^{N} S_i(u_{i-1}, u_i) \geqslant x\right\} && (2.29)\\
&\leqslant \inf_{\mathbf{u}\in\mathcal{U}(x)} \mathbb{E}\left[e^{\theta \sum_{i=1}^{N} S_i(u_{i-1}, u_i)}\right] \cdot e^{-\theta x}, && (2.30)
\end{aligned}
$$

where $\mathcal{U}(x) = \{\mathbf{u} = (u_0, u_1, \ldots, u_N) : u_0 = 0, u_N = t, 0 \leqslant u_1 \leqslant \ldots \leqslant u_{N-1} \leqslant t\}$, for some $\theta > 0$. □

2.1.3 Dependence Refinement

In general, the capacity is dependent over time, which results from the temporal dependence in the environment or in the controllable parameters of the system. Specifically for the cumulative capacity, the influence of stochastic dependence is characterized by the Fréchet bounds [123]

$$
\breve{F}_{S(t)}(x) \leqslant F_{S(t)}(x) \leqslant \widehat{F}_{S(t)}(x), \qquad (2.31)
$$

where

$$
\begin{aligned}
\breve{F}_{S(t)}(x) &= \left[\sup_{\mathbf{u}\in\mathcal{U}(x)} \sum_{i=1}^{t} F_{c(i)}(u_i) - (t-1)\right]^{+}, && (2.32)\\
\widehat{F}_{S(t)}(x) &= \left[\inf_{\mathbf{u}\in\mathcal{U}(x)} \sum_{i=1}^{t} F_{c(i)}(u_i)\right]_{1}, && (2.33)
\end{aligned}
$$

with $\mathcal{U}(x) = \left\{\mathbf{u} = (u_1, \ldots, u_t) : \sum_{i=1}^{t} u_i = x\right\}$, $[\cdot]_1 = \min(\cdot, 1)$, and $[\cdot]^+ = \max(\cdot, 0)$.

The Fréchet bounds hold generally, making use of specific dependence information among $c(1), c(2), \ldots,$ the bounds can be improved. To this end, three representative capacity processes are investigated in this subsection, which are comonotonic process, additive process, and Markov additive process.

2.1.3.1 Comonotonic Process

The upper Fréchet bound expresses the extremal positive dependence indicating the largest sum with respect to convex order, and the dependence structure is represented by the comonotonic copula [37, 41, 44], i.e.,

$$F(x_1, \ldots, x_t) = \min_{1 \leqslant i \leqslant t} F_{c(i)}(x_i); \tag{2.34}$$

equivalently [37], for a uniform random variable $U \sim U(0,1)$,

$$(c(1), \ldots, c(t)) \stackrel{d}{=} \left(F_{c(1)}^{-1}(U), \ldots, F_{c(t)}^{-1}(U) \right), \tag{2.35}$$

which implicates that comonotonic random variables are increasing functions of a common random variable [41].

If the increment of the cumulative capacity has comonotonicity, the cumulative capacity is defined as a comonotonic process in this work (which is different from a similar concept regarding the comonotonicity between different processes [73]). The distribution results of cumulative capacity and transient capacity are as follows.

Theorem 2. *For a strictly stationary capacity process, the distributions of the cumulative capacity and transient capacity with comonotonicity are expressed as*

$$F_{S(t)}(x) = F_c\left(\frac{x}{t}\right), \tag{2.36}$$

$$F_{\overline{c}(t)}(x) = F_c(x). \tag{2.37}$$

Proof. Since all the marginal distribution functions are identical $F_{c(i)} \sim F_c$, comonotonicity of $c(i)$ is equivalent to saying that $c(1) = c(2) = \ldots = c(t)$ holds almost surely [37]. In other words, the sample function of the capacity process is stationary and depends only on the initial value of the capacity in each realization. $\square$

2.1.3.2 Additive Process

The independence structure of an additive process is expressed by a product copula [43][44]

$$F(x_1, \ldots, x_t) = \prod_{i=1}^{t} F_{c(i)}(x_i), \tag{2.38}$$

and the distribution of the cumulative capacity is expressed via Stieltjes convolution as

$$F_{S(t)}(x) = F_{c(1)} \circledast \ldots \circledast F_{c(t)}(x). \tag{2.39}$$

The cumulative capacity with independent increment is modeled as an additive process [65]. The distribution bounds of cumulative capacity and transient capacity are as follows.

Theorem 3. *For an independent and identically distributed capacity process, the distribution of the cumulative capacity with independence is expressed as, for some $\theta > 0$,*

$$1 - e^{t\kappa(\theta) - \theta x} \leqslant F_{S(t)}(x) \leqslant e^{t\kappa(-\theta) + \theta x}, \tag{2.40}$$

where $\kappa(\theta) = \log \mathbb{E}\left[e^{\theta c(i)}\right]$ is the cumulant generating function of the instantaneous capacity, and the distribution of the transient capacity is expressed as

$$1 - e^{-y_l} \leqslant \mathbb{P}\left\{\bar{c}(t) \leqslant c^*\right\} \leqslant e^{-y_u}, \tag{2.41}$$

where $c^* = \frac{t\kappa(\theta^*) + y^*}{\theta^* t}$, with $y^* = y_u$ and $\theta^* < 0$ for the upper bound, and

$y^* = y_l$ and $\theta^* > 0$ for the lower bound.

Proof. Since all the marginal distribution functions are identical $F_{c(i)} \sim F_c$, a likelihood ratio process of the cumulative capacity is formulated and expressed as [6]

$$L(t) = e^{\theta S(t) - t\kappa(\theta)}, \tag{2.42}$$

where $L(t)$ is a mean-one martingale and $\kappa(\theta)$ is the cumulant generating function, i.e.,

$$\kappa(\theta) = \log \mathbb{E}\left[e^{\theta c(i)}\right] = \log \int e^{\theta x} F(dx), \tag{2.43}$$

where $\theta \in \Theta = \{\theta \in \mathbb{R} : \kappa(\theta) < \infty\}$.

According to Markov inequality, for any $\mu > 0$,

$$\mathbb{P}\{L(t) \geqslant \mu\} \leqslant \frac{1}{\mu}\mathbb{E}[L(t)] = \frac{1}{\mu}. \tag{2.44}$$

Letting $\mu = e^{-t\kappa(\theta)+\theta x}$, with a manipulation of $\mathbb{P}\{L(t) \geqslant \mu\}$, for $\theta \leqslant 0$, the distribution function of $S(t)$ is bounded by

$$\mathbb{P}\{S(t) \leqslant x\} \leqslant e^{t\kappa(\theta)-\theta x}, \tag{2.45}$$

while for $\theta > 0$, the tail distribution function of $S(t)$ is bounded by

$$\mathbb{P}\{S(t) \geqslant x\} \leqslant e^{t\kappa(\theta)-\theta x}, \tag{2.46}$$

which shows that the distribution has a light tail. Letting $-y^* = t\kappa(\theta) - \theta x \leqslant 0$, the distribution of the transient capacity $\overline{c}(t) = \frac{S(t)}{t}$ is bounded by

$$1 - e^{-y_l} \leqslant \mathbb{P}\left\{\overline{c}(t) \leqslant c^*\right\} \leqslant e^{-y_u}, \tag{2.47}$$

where $c^* = \frac{t\kappa(\theta)+y^*}{\theta t}$, $y^* = y_u$ for the upper bound with $\theta < 0$, and $y^* = y_l$ for the lower bound with $\theta > 0$. $\square$

Remark 1. *The upper and lower bound of $F_{S(t)}(x)$ do not hold simulta-*

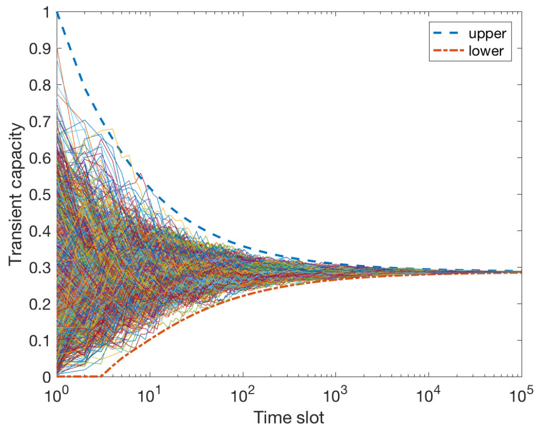
neously, the upper bound is useful for $x < \overline{S}(t)$*, the lower bound is useful for* $x > \overline{S}(t)$*, and both bounds are worthless for* $x = \overline{S}(t)$ *[51]. Considering* $\overline{F}_{S(t)}(x) = 1 - F_{S(t)}(t)$*, which means that the upper and lower bound can not decrease or increase simultaneously, this property holds in general. An indication of this property is that, for a fixed violation probability, the obtained bounds on* $S(t)$ *or* $\overline{c}(t)$ *based on the upper and lower distribution bounds are lower and upper bounds of* $S(t)$ *or* $\overline{c}(t)$ *with respect to their mean. It is illustrated in Fig. 2.1.*

2.1.3.3 Markov Additive Process

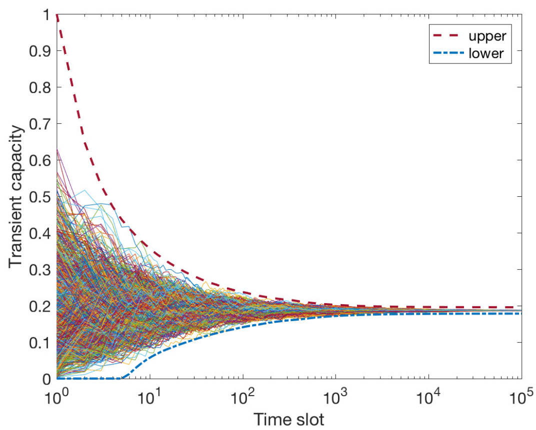
The Markov property is solely a dependence property that can be modeled exclusively in terms of copulas [32, 104]. As a consequence, starting with a Markov process, a multitude of other Markov processes can be constructed by just modifying the marginal distributions [32, 80, 104]. It is worth noting that the Markov property indicates both positive and negative dependence, which is determined by the underlying copula. For a Markov chain, the selection of the copula and the marginal distribution is coupled [32], the transition matrix can be expressed in terms of the copula and marginal distribution and vice versa. Particularly for an idempotent copula, the process is conditionally independently and identically distributed given the initial state [80].

Specifically, if the dependence in capacity follows a Markov process and the instantaneous capacity has a corresponding distribution with respect to a state transition, then the cumulative capacity is a Markov additive process, which is a bivariate process with strong Markov property and the increment process is conditionally independent given a realization of the underlying Markov process. A formal definition of Markov additive process is in Appendix B.2.

Theorem 4. *For a Markov additive process* $(S(t), J_t)$ *with state space* E*,*



(a) Additive process.



(b) Markov additive process.

Figure 2.1: Transient capacity of additive and Markov additive Rayleigh channel. According to the strong law of large numbers for the additive process and extended to the Markov additive process, the transient capacity converges to the mean as time goes to infinity, i.e., the convergence of sample paths. The large deviation results are upper bound and lower bound with respect to the mean. Results are normalized, with violation probability $\epsilon = 10^{-3}$, $W = 20\text{kHz}$, $SNR = e^{0.5}$ for the additive process, $\mathbf{SNR} = [e^{0.5}\ 0.9e^{0.5}; 0.8e^{0.5}\ 0.7e^{0.5}]$ and $\mathbf{P} = [0.3\ 0.7; 0.1\ 0.9]$ for the Markov additive process with initial distribution $\boldsymbol{\pi} = [0.5\ 0.5]$, and 1000 sample paths.

conditional on initial state J_0, the distribution of the cumulative capacity is expressed as, for some $\theta > 0$,

$$1 - \frac{h^{(\theta)}(J_0)e^{t\kappa(\theta)-\theta x}}{\min\limits_{j\in E}(h^{(\theta)}(J_j))} \leqslant F_{S(t)}(x) \leqslant \frac{h^{(-\theta)}(J_0)e^{t\kappa(-\theta)+\theta x}}{\min\limits_{j\in E}(h^{(-\theta)}(J_j))}, \tag{2.48}$$

where $\kappa(\theta)$ and $\mathbf{h}^{(\theta)}$ are respectively the logarithm of the Perron-Frobenius eigenvalue and the corresponding right eigenvector of the kernel for the Markov additive process $(S(t), J_t)$, i.e., $\widehat{\boldsymbol{F}}[\theta]$, and the distribution of the transient capacity is expressed as

$$1 - \frac{h^{(\theta)}(J_0)e^{-y_l}}{\min\limits_{j\in E}(h^{(\theta)}(J_j))} \leqslant \mathbb{P}\left\{\bar{c}(t) \leqslant c^*\right\} \leqslant \frac{h^{(-\theta)}(J_0)e^{-y_u}}{\min\limits_{j\in E}(h^{(-\theta)}(J_j))}, \tag{2.49}$$

where $c^* = \frac{t\kappa(\theta^*)+y^*}{\theta^* t}$, with $y^* = y_u$ and $\theta^* < 0$ for the upper bound, and $y^* = y_l$ and $\theta^* > 0$ for the lower bound.

Proof. Like the independent case, a likelihood ratio process is formulated with an exponential change of measure [6],

$$L(t) = \frac{h^{(\theta)}(J_t)}{h^{(\theta)}(J_0)}e^{\theta S(t)-t\kappa(\theta)}, \tag{2.50}$$

which is a mean-one martingale. $\kappa(\theta)$ and $\mathbf{h}^{(\theta)}$ are respectively the logarithm of the Perron-Frobenius eigenvalue and the corresponding right eigenvector of the kernel for the Markov additive process $(S(t), J_t)$, i.e., $\widehat{\mathbf{F}}[\theta]$. In order to provide exponential upper bound for the distribution of the cumulative capacity, define [51]

$$\underline{L}(t) = \frac{\min_{j\in E}(h^{(\theta)}(J_j))}{h^{(\theta)}(J_0)}e^{\theta S(t)-t\kappa(\theta)}, \tag{2.51}$$

where $\underline{L}(t) \leqslant L(t)$, i.e., $\mathbb{E}[\underline{L}(t)] \leqslant 1$. Apply Markov inequality to $\underline{L}(t)$ and get, for any $\mu > 0$,

$$\mathbb{P}\{\underline{L}(t) \geqslant \mu\} \leqslant \frac{1}{\mu}\mathbb{E}[\underline{L}(t)] \leqslant \frac{1}{\mu}. \tag{2.52}$$

Choosing $\mu = e^{-t\kappa(\theta)+\theta x} \cdot \frac{\min_{j\in E}(h^{(\theta)}(J_j))}{h^{(\theta)}(J_0)}$, based on the tail distribution function $\mathbb{P}\{\underline{L}(t) \geqslant \mu\}$, we get the distribution function of $\mathbb{P}\{S(t) \leqslant x\}$, for $\theta \leqslant 0$,

$$\mathbb{P}\{S(t) \leqslant x\} \leqslant \frac{h^{(\theta)}(J_0)}{\min_{j\in E}(h^{(\theta)}(J_j))} e^{t\kappa(\theta)-\theta x}, \tag{2.53}$$

while for $\theta > 0$,

$$\mathbb{P}\{S(t) \geqslant x\} \leqslant \frac{h^{(\theta)}(J_0)}{\min_{j\in E}(h^{(\theta)}(J_j))} e^{t\kappa(\theta)-\theta x}, \tag{2.54}$$

which indicates that the distribution has a light tail. Letting $-y^* = t\kappa(\theta) - \theta x \leqslant 0$, the distribution of the transient capacity $\overline{c}(t) = \frac{S(t)}{t}$ is bounded by

$$1 - \frac{h^{(\theta)}(J_0)e^{-y_l}}{\min\limits_{j\in E}(h^{(\theta)}(J_j))} \leqslant \mathbb{P}\left\{\overline{c}(t) \leqslant c^*\right\} \leqslant \frac{h^{(\theta)}(J_0)e^{-y_u}}{\min\limits_{j\in E}(h^{(\theta)}(J_j))}, \tag{2.55}$$

where $c^* = \frac{t\kappa(\theta)+y^*}{\theta t}$, $y^* = y_u$ for the upper bound with $\theta < 0$, and $y^* = y_l$ for the lower bound with $\theta > 0$. $\square$

Remark 2. *The Markov additive process can be seen as a non-stationary additive process defined on a Markov process. If the Markov process has only one state, then it reduces to a stationary additive process [28]. In addition, the strong law of large numbers applies to the Markov additive process [100], and the mean of transient capacity exists [7], i.e.,*

$$\lim_{t\to\infty} \frac{\mathbb{E}_{J_0\in E}[S(t)]}{t} = \kappa'(0). \tag{2.56}$$

It is demonstrated in Fig. 2.1.

2.2 Multiple-Input-Multiple-Output Channel

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, $m \in \mathbb{N}$, $n \in \mathbb{N}$, and $\boldsymbol{X} : \Omega \to \mathbb{C}^{m\times n}$ be measurable with respect to $\mathscr{F}$ and the Borel σ-algebra on $\mathbb{C}^{m\times n}$. Denote $\boldsymbol{\mathfrak{X}} = \{\boldsymbol{X} \in \mathbb{C}^{n\times n} : \boldsymbol{X} = \boldsymbol{X}^*\}$, where $*$ represents the conjugate transpose. Denote the cone [1] $\boldsymbol{\mathfrak{X}}_{\geqslant 0} = \{\boldsymbol{X} \in \boldsymbol{\mathfrak{X}} : \boldsymbol{X} \geqslant 0\}$, which introduces a partial order in $\boldsymbol{\mathfrak{X}}$, i.e., $\boldsymbol{X} \geqslant 0$ is equivalent to that all the eigenvalues of $\boldsymbol{X}$ are nonnegative [1]. Similarly, $\boldsymbol{\mathfrak{X}}_{>0} = \{\boldsymbol{X} \in \boldsymbol{\mathfrak{X}} : \boldsymbol{X} > 0\}$.

2.2.1 Deterministic Power Fluctuation

Consider the flat fading MIMO channel $\boldsymbol{H} \in \mathbb{C}^{N_R\times N_T}$, $N_T \in \mathbb{N}$, $N_R \in \mathbb{N}$, $\boldsymbol{H}\boldsymbol{H}^* \in \boldsymbol{\mathfrak{X}}_{\geqslant 0}$. The capacity, in bits per second, under total average transmit power constraint, is expressed as [110]

$$c = W \max_{\mathrm{Tr}[\boldsymbol{R_{ss}}]=N_T} \log_2 \det\left(\boldsymbol{I}_{N_R} + \frac{\rho}{N_T}\boldsymbol{H}\boldsymbol{R_{ss}}\boldsymbol{H}^*\right), \tag{2.57}$$

where W is the bandwidth, $\rho = \frac{P}{N_0 W}$, P is the total average transmit power, N_0 is the noise power spectral density, $\boldsymbol{R_{ss}} = \mathbb{E}[\boldsymbol{s}\boldsymbol{s}^*]$ is the covariance matrix for the transmitted signal $\boldsymbol{s} \in \mathbb{C}^{N_T\times 1}$.

The frequency-selective fading channel formulation requires a block diagonal extension of the flat fading channel model. The capacity, in bits per second, under total average transmit power constraint, is expressed as [110]

$$c = \frac{W}{N} \max_{\mathrm{Tr}[\boldsymbol{R_{\mathcal{S}\mathcal{S}}}]=N_T N} \log_2 \det\left(\boldsymbol{I}_{N_R N} + \frac{\rho}{N_T}\boldsymbol{\mathcal{H}}\boldsymbol{R_{\mathcal{S}\mathcal{S}}}\boldsymbol{\mathcal{H}}^*\right), \tag{2.58}$$

where W is the bandwidth, $\rho = \frac{P}{N_0 W}$, P is the total average transmit power, N_0 is the noise power spectral density, N is the number of sub-channels, $\boldsymbol{\mathcal{H}} \in \mathbb{C}^{N_T N\times N_R N}$ is the block diagonal matrix with $\boldsymbol{H}_i$ as the block diagonal elements, and $\boldsymbol{R_{\mathcal{S}\mathcal{S}}} = \mathbb{E}[\boldsymbol{\mathcal{S}}\boldsymbol{\mathcal{S}}^*]$ is the covariance matrix for the transmitted signal $\boldsymbol{\mathcal{S}} = [\boldsymbol{s}_1^T, \ldots, \boldsymbol{s}_N^T]^T \in \mathbb{C}^{N_T N \times 1}$, where $[\cdot]^T$ denotes the transpose of a matrix.

Remark 3. *The identity matrix* $\boldsymbol{I}$ *in the capacity formula implies that the capacity is non-negative, i.e.,* $c : \Omega \to \mathbb{R}_{\geqslant 0}$.

Remark 4. *The typical stochastic models of the channel gain are the Rayleigh, Rice, and Nakagami distributions [55]. The shadowing model is the Lognormal distribution [118][55], which is able to superimpose the path loss.*

Remark 5. *If* $\log Y \sim N\big(\mu, \sigma^2\big)$*, then* $a + b\log Y \sim N\big(a + b\mu, b^2\sigma^2\big)$*, where* $a, b \in \mathbb{R}$*, thus, if* Y *is lognormal, then* aY^b *is also lognormal in general. This result explains the product form of the combined effect of the multiple path interference, shadowing, and path loss [109].*

Remark 6. *Since the normal distribution with zero mean is symmetric, we have the equal lognormal distributions* $Y^{-1} \overset{d}{=} Y$*, because of* $-\log Y \overset{d}{=} \log Y \sim N\big(0, \sigma^2\big)$*. This result implies that the quotient* X/Y *of an arbitrary random variable* X *with a lognormal random variable* $\log Y \sim N\big(0, \sigma^2\big)$*, where* X *and* Y *are independent, equals in distribution the product* XY *of the two random variable, i.e.,* $X/Y \overset{d}{=} XY$*. This relation does not hold for the general normal distribution.*

Remark 7. *For a almost surely positive random variable,* X*, the right tail behavior of* $1/X$*,* $\mathbb{P}(1/X > x) = O\big(e^{-\theta x}\big)$*,* $\exists\theta > 0$*, and* $\mathbb{P}(1/X > x) = O\big(x^{-\theta}\big)$*,* $\exists\theta > 0$*, corresponds to left tail behavior of* X *[56][9],* $\mathbb{P}(X < 1/x) = O\big(e^{-\theta x}\big)$*,* $\exists\theta > 0$*, and* $\mathbb{P}(X < 1/x) = O\big(x^{-\theta}\big)$*,* $\exists\theta > 0$*, i.e.,* $\limsup\limits_{y\to 0} \frac{\mathbb{P}(X<y)}{e^{-\theta/y}} < \infty$*,* $\exists\theta > 0$*, and* $\limsup\limits_{y\to 0} \frac{\mathbb{P}(X<y)}{y^{\theta}} < \infty$*,* $\exists\theta > 0$*. Letting* $Y = 1/X$*, we obtain the complementary results. Considering the reciprocal relation between the channel loss* $\phi = P_T/P_R$ *and channel gain* $\psi = P_R/P_T$*, both the right tail and the left tail matter for the stochastic channel models.*

Remark 8. *The difference between the two dimensional and the three dimensional channel models lies in that there are different autocorrelation functions of the fading process and the same marginals of the fading envelopes [11][107]. In addition, it is shown that a horizontal separation of antennas has a superiority over the vertical separation, though the latter is able to increase the diversity gain [107][139]. The capacity of multi-user MIMO is more complex, i.e., the channel involves both multiple access channel and broadcasting channel, thus, the capacity expression has a diverse formulation [110].*

We present some equivalence results of the function of random variables.

Lemma 1. *Consider a flat MIMO channel $\boldsymbol{H} \in \mathbb{C}^{N_R \times N_T}$. The capacity is upper bounded by $c' = a \log_2(1 + b\lambda_{\max})$, where $a, b \in \mathbb{R}_{>0}$ and $\lambda_{\max}$ is the maximum eigenvalue of $\boldsymbol{H}\boldsymbol{H}^*$.*

1. For the tail property, we have the equivalent results

$$\begin{aligned} &\overline{F}_{c'}(x) = O(e^{-\theta x}),\ \exists \theta > 0 \\ &\qquad \Longleftrightarrow \overline{F}_{\lambda_{\max}}(x) = O(x^{-\theta}),\ \exists \theta > 0 \\ &\qquad\qquad \Longleftrightarrow \overline{F}_{\mathrm{Tr}\,[\boldsymbol{H}\boldsymbol{H}^*]}(x) = O(x^{-\theta}),\ \exists \theta > 0. \end{aligned} \tag{2.59}$$

2. For the power law function of the maximum eigenvalue, we have the equivalent expressions

$$\begin{aligned} &\mathbb{E}\Big[(1 + \Delta\lambda_{\max})^{\theta}\Big] < \infty,\ 0 < \Delta < \infty,\ \exists \theta > 0 \\ &\qquad \Longleftrightarrow \mathbb{E}\Big[(1 + \lambda_{\max})^{\theta}\Big] < \infty,\ \exists \theta > 0 \\ &\qquad\qquad \Longleftrightarrow \mathbb{E}\Big[(\lambda_{\max})^{\theta}\Big] < \infty,\ \exists \theta > 0. \end{aligned} \tag{2.60}$$

Specifically, $\theta = 1$ corresponds to $\mathbb{E}[\lambda_{\max}] < \infty$. In addition, we have

$\mathbb{E}\Big[(\lambda_{\max})^{\theta_1}\Big] = \infty \implies \mathbb{E}\Big[(\lambda_{\max})^{\theta_2}\Big] = \infty$, $\forall 0 < \theta_1 < \theta_2$.

3. In addition, we have another pair of equivalent expressions for the exponential function of the eigenvalue, i.e.,

$$\mathbb{E}\Big[e^{\theta \operatorname{Tr}[\boldsymbol{HH}^*]}\Big] < \infty,\ \exists\theta > 0 \iff \mathbb{E}\Big[e^{\theta\lambda_{\max}}\Big] < \infty,\ \exists\theta > 0. \tag{2.61}$$

Proof. 1. Considering the eigendecomposition of the channel matrix in the capacity formula [110], the capacity is upper bounded by $c' = a\log_2(1 + b\lambda_{\max})$, we have $\overline{F}_{c'}(x) = O\big(e^{-\theta x}\big) \iff \overline{F}_{\lambda_{\max}}(x) = O\big(x^{-\theta}\big) \iff \overline{F}_{\operatorname{Tr}[\boldsymbol{HH}^*]}(x) = O\big(x^{-\theta}\big)$. The first relationship follows the transform of random variables, i.e., $\overline{F}_{c'}(x) = O\big(e^{-\theta c}\big) \iff \overline{F}_{\lambda_{\max}}(x) = O\big((1+bx)^{-\theta}\big)$, and $(1+bx)^{-\theta} \sim (bx)^{-\theta}$ and $\overline{F}_{\lambda_{\max}}(x) = O\big(x^{-\theta}\big) \iff \overline{F}_{\lambda_{\max}}(x) = O\big((bx)^{-\theta}\big)$. The second relationship follows that $\lambda_{\max} \leqslant \operatorname{Tr}[\boldsymbol{HH}^*] \implies \overline{F}_{\lambda_{\max}}(x) \leqslant \overline{F}_{\operatorname{Tr}[\boldsymbol{HH}^*]}(x)$, thus, $\overline{F}_{\operatorname{Tr}[\boldsymbol{HH}^*]}(x) = O\big(x^{-\theta}\big) \implies \overline{F}_{\lambda_{\max}}(x) = O\big(x^{-\theta}\big)$; and $r(\boldsymbol{H})\lambda_{\max} \geqslant \operatorname{Tr}[\boldsymbol{HH}^*] \implies \overline{F}_{r(\boldsymbol{H})\lambda_{\max}}(x) \geqslant \overline{F}_{\operatorname{Tr}[\boldsymbol{HH}^*]}(x)$, thus, $\overline{F}_{\lambda_{\max}}(x) = O\big(x^{-\theta}\big) \implies \overline{F}_{\operatorname{Tr}[\boldsymbol{HH}^*]}(x) = O\big(x^{-\theta}\big)$.

2. First, we have the inequality, $\mathbb{E}\Big[(1+\Delta\lambda_{\max})^{\theta}\Big] \leqslant (1+\Delta)^{\theta}\mathbb{E}\Big[(1+\lambda_{\max})^{\theta}\Big]$, which implies that $\mathbb{E}\Big[(1+\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0 \implies \mathbb{E}\Big[(1+\Delta\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0$, because $(1+\Delta)^{\theta} < \infty$. Second, for $0 < \Delta < 1$, letting $\Delta \geqslant \frac{1}{m}$, $m \in \mathbb{N}$, we have $m^{\theta}\mathbb{E}\Big[(1+\Delta\lambda_{\max})^{\theta}\Big] \geqslant \mathbb{E}\Big[(1+\lambda_{\max})^{\theta}\Big]$, which implies that $\mathbb{E}\Big[(1+\Delta\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0 \implies \mathbb{E}\Big[(1+\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0$, because $m^{\theta} < \infty$; for $\Delta \geqslant 1$, it is trivial.

Since $(\lambda_{\max})^{\theta} \leqslant (1+\lambda_{\max})^{\theta}$, we have $\mathbb{E}\Big[(1+\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0 \implies \mathbb{E}\Big[(\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0$. According to Hölder's inequality, $\mathbb{E}[X^r] \leqslant (\mathbb{E}[X^s])^{r/s}$, $0 < r < s$, $X \in \mathbb{R}_{\geqslant 0}$, which implies that $\mathbb{E}[X^s] < \infty \implies \mathbb{E}[X^r] < \infty$. Specifically, we have $\mathbb{E}\Big[(\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 1 \implies \mathbb{E}\Big[(\lambda_{\max})^{\theta}\Big] < \infty,\ \forall 0 < \theta \leqslant 1$. Suppose $\mathbb{E}\Big[(\lambda_{\max})^{\theta}\Big] < \infty,\ \exists 0 < \theta \leqslant 1$, we have $\overline{F}_{\lambda_{\max}}(x) = o\big(x^{-\theta}\big)$ and $\int_0^{\infty} \overline{F}_{\lambda_{\max}}(x)x^{\theta-1}dx < \infty$, since

$x^{-\theta} \sim (1+x)^{-\theta}$ and $x^{\theta-1} \geqslant (1+x)^{\theta-1}$, we further have $\overline{F}_{\lambda_{\max}}(x) = o\big((1+x)^{-\theta}\big)$ and $\int_0^\infty \overline{F}_{\lambda_{\max}}(x)(1+x)^{\theta-1}dx < \infty$, which corresponds to $\mathbb{E}\Big[(1+\lambda_{\max})^\theta\Big] < \infty$. Thus, $\mathbb{E}\Big[(1+\lambda_{\max})^\theta\Big] < \infty,\ \exists 0 < \theta \leqslant 1 \Longleftarrow \mathbb{E}\Big[(\lambda_{\max})^\theta\Big] < \infty,\ \exists 0 < \theta \leqslant 1$. In all, we obtain $\mathbb{E}\Big[(1+\lambda_{\max})^\theta\Big] < \infty,\ \exists \theta > 0 \iff \mathbb{E}\Big[(\lambda_{\max})^\theta\Big] < \infty,\ \exists \theta > 0$.

If $\mathbb{E}[\lambda_{\max}] = \infty$, by Jensen's inequality, $\mathbb{E}\Big[(\lambda_{\max})^\theta\Big] \geqslant (\mathbb{E}[\lambda_{\max}])^\theta = \infty,\ \forall \theta \geqslant 1$, which implies that $\mathbb{E}\Big[(1+\lambda_{\max})^\theta\Big] = \infty,\ \forall \theta \geqslant 1$, because $\mathbb{E}\Big[(\lambda_{\max})^\theta\Big] \leqslant \mathbb{E}\Big[(1+\lambda_{\max})^\theta\Big]$.

3. Considering that the matrix $\boldsymbol{H}\boldsymbol{H}^* \in \mathfrak{X}_{\geqslant 0}$, we have $\lambda_{\max} \leqslant \mathrm{Tr}\,[\boldsymbol{H}\boldsymbol{H}^*]$ and $\mathrm{Tr}\,[\boldsymbol{H}\boldsymbol{H}^*] \leqslant r(\boldsymbol{H})\lambda_{\max}$. Then, the proof follows.

$\square$

Remark 9. *The parameters θ in the two equations, $\overline{F}_c(x) = O(e^{-\theta x})$ and $\overline{F}_{\lambda_{\max}}(x) = O(x^{-\theta})$, are not necessarily equal.*

Remark 10. *The above equivalent results indicate that, for $X \in \mathbb{R}_{\geqslant 0}$, $\mathbb{E}\big[X^\theta\big] < \infty,\ \exists \theta > 0 \iff \overline{F}_X(x) = O\big(x^{-\theta}\big),\ \exists \theta > 0$. However, it is interesting to notice that, for a common $\theta > 0$ and $X \in \mathbb{R}_{\geqslant 0}$, we only have $\mathbb{E}\big[X^\theta\big] < \infty \implies \overline{F}_X(x) = O\big(x^{-\theta}\big)$, and the reverse does not hold in general. Because it is shown in [125] that $\mathbb{E}\big[X^\theta\big] < \infty$, where $\theta > 0$ and X is a nonnegative random variable, if and only if $\overline{F}_X(x) = o\big(x^{-\theta}\big)$ and $\int_0^\infty \overline{F}_X(x)x^{\theta-1}dx < \infty$.*

Remark 11. *The mean identity $\mathbb{E}[X^\theta]$, for $X : \Omega \to \mathbb{R}_{\geqslant 0}$ and $\theta > 0$, is a special case of Mellin-Stieltjes transform [146] of the distribution function $F_X(x)$.*

Remark 12. *Suppose X is a regularly varying non-negative random variable with index $\alpha > 0$. Then [42], $\mathbb{E}\big[X^\beta\big] < \infty$, for $\beta < \alpha$; and $\mathbb{E}\big[X^\beta\big] = \infty$, for $\beta > \alpha$.*

2.2.1.1 Arbitrary Channel Side Information

We present the sufficient condition for the light-tailed capacity of the flat channel.

Theorem 5. *Consider a flat MIMO channel* $\boldsymbol{H} \in \mathbb{C}^{N_R \times N_T}$. *If the mean identity exists, i.e.,*

$$\mathbb{E}\Big[(\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0, \tag{2.62}$$

where $\lambda_{\max}$ *is the maximum eigenvalue of* $\boldsymbol{H}\boldsymbol{H}^*$, *the distribution of the capacity of the MIMO channel (with or without full channel side information) is light-tailed.*

Proof. First, consider the scenario with channel side information only at the receiver. We denote $\rho := \frac{P}{N_0 W}$. On the one hand, we have an upper bound of the capacity

$$c = W \sum\nolimits_{i=1}^{r(\boldsymbol{H})} \log_2\left(1 + \frac{\rho}{N_T}\lambda_i\right) \tag{2.63}$$

$$\leqslant W r(\boldsymbol{H}) \log_2\left(1 + \frac{\rho}{N_T}\lambda_{\max}\right), \tag{2.64}$$

where $r(\boldsymbol{H})$ is the rank of matrix $\boldsymbol{H}$, λ_i is the eigenvalue of the matrix $\boldsymbol{H}\boldsymbol{H}^*$, and the equality follows the eigenvalue expression of the capacity.

Second, consider the scenario with full channel side information. The capacity is upper bounded by

$$c = W \max_{\sum_{i=1}^{r(\boldsymbol{H})} \gamma_i = N_T} \sum\nolimits_{i=1}^{r(\boldsymbol{H})} \log_2\left(1 + \frac{\rho\gamma_i}{N_T}\lambda_i\right) \tag{2.65}$$

$$\leqslant W r(\boldsymbol{H}) \log_2\left(1 + \frac{\rho}{N_T}\gamma_{\max}\lambda_i\right) \tag{2.66}$$

$$\leqslant W r(\boldsymbol{H}) \log_2 (1 + \rho\lambda_{\max}). \tag{2.67}$$

It is easy to show that $\mathbb{E}\big[e^{\theta c}\big] < \infty,\ \exists\theta > 0$, entails $\mathbb{E}\Big[(1 + \rho\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0$. The rest of the proof follows Lemma 1.2. $\square$

We present the sufficient condition for the light-tailed capacity of the frequency-selective channel.

Theorem 6. *Consider a frequency-selective MIMO channel with the sub-channels* $\boldsymbol{H}_i \in \mathbb{C}^{N_R \times N_T}$, $i \in \{1, \ldots, N\}$, *and block diagonal matrix* $\boldsymbol{\mathcal{H}} = diag(\boldsymbol{H}_1, \ldots, \boldsymbol{H}_N)$. *For the scenarios where the channel side information is known or unknown at the transmitter, if the mean identity exists for each sub-channel, i.e.,*

$$\mathbb{E}\Big[\big(\lambda_{\max}^{i}\big)^{\theta}\Big] < \infty,\ \exists \theta > 0,\ \forall i \in \{1, \ldots, N\}, \tag{2.68}$$

where $\lambda_{\max}^{i}$ *is the maximum eigenvalue of* $\boldsymbol{H}_i \boldsymbol{H}_i^*$, *or the equivalent condition is satisfied, i.e.,*

$$\mathbb{E}\Big[\big(\lambda_{\max}\big)^{\theta}\Big] < \infty,\ \exists \theta > 0, \tag{2.69}$$

where $\lambda_{\max}$ *is the maximum eigenvalue of* $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^*$, *the distribution of the capacity is light-tailed.*

Proof. If the channel side information is unknown to the transmitter, then [110]

$$c = \frac{W}{N} \sum_{i=1}^{N} \log_2 \det \left(\boldsymbol{I}_{N_R} + \frac{\rho}{N_T} \boldsymbol{H}_i \boldsymbol{H}_i^* \right). \tag{2.70}$$

Since the light-tailed property is preserved under the sum operation, if the capacity distribution of each sub-channel is light-tailed, so is the total capacity distribution.

If the channel side information is known to the transmitter, then [110]

$$c = \frac{W}{N} \max_{\sum_{i=1}^{r(\boldsymbol{\mathcal{H}})} \gamma_i = N_T N} \sum_{i=1}^{r(\boldsymbol{\mathcal{H}})} \log_2 \left(1 + \frac{\rho \gamma_i}{N_T} \lambda_i(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^*)\right) \tag{2.71}$$

$$\leqslant \frac{W}{N} \sum_{i=1}^{r(\boldsymbol{\mathcal{H}})} \log_2 \left(1 + \frac{\rho \gamma_{\max}}{N_T} \lambda_i(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^*)\right) \tag{2.72}$$

$$\leqslant \frac{W}{N} r(\boldsymbol{\mathcal{H}}) \log_2 \left(1 + \rho N \lambda_{\max}(\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^*)\right). \tag{2.73}$$

Particularly, we have $\lambda_{\max} = \max\left(\lambda_{\max}^1, \ldots, \lambda_{\max}^N\right)$, which implies the equivalent expression of the condition. $\square$

Remark 13. *The equivalent expression of the sufficient condition means that it is equivalent to consider the block diagonal matrix of the frequency-selective channel as a whole or to consider the matrix of each sub-channel individually.*

Remark 14. *If the trace identity exists, i.e.,* $\mathbb{E}\left[\mathrm{Tr}\left[e^{\theta \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^*}\right]\right] < \infty$, $\exists \theta > 0$, *then the maximum eigenvalue distribution and the capacity distribution are light-tailed. Because the maximum eigenvalue distribution is exponentially bounded [137]*

$$\mathbb{P}(\lambda_{\max}(\boldsymbol{X}) \geqslant x) \leqslant e^{-\theta x} \cdot \mathbb{E}\left[\mathrm{Tr}\left[e^{\theta \boldsymbol{X}}\right]\right], \ \forall \theta > 0, \tag{2.74}$$

where $\boldsymbol{X} = \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^* \in \boldsymbol{\mathfrak{X}}$. *Since the matrix* $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^*$ *is block diagonal [61],* $e^{\theta \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^*} = \mathit{diag}\left(e^{\theta \boldsymbol{H}_1 \boldsymbol{H}_1^*}, \ldots, e^{\theta \boldsymbol{H}_N \boldsymbol{H}_N^*}\right)$, *and*

$$\mathbb{E}\left[\mathrm{Tr}\left[e^{\theta \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^*}\right]\right] = \sum_{i=1}^{N} \mathbb{E}\left[\mathrm{Tr}\left[e^{\theta \boldsymbol{H}_i \boldsymbol{H}_i^*}\right]\right]. \tag{2.75}$$

Thus, $\mathbb{E}\left[\mathrm{Tr}\left[e^{\theta \boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{H}}^*}\right]\right] < \infty$, $\exists \theta > 0$ *entails* $\mathbb{E}\left[\mathrm{Tr}\left[e^{\theta \boldsymbol{H}_i \boldsymbol{H}_i^*}\right]\right] < \infty$, $\exists \theta > 0$, $\forall i \in \{1, \ldots, N\}$, *which is non-negative.*

2.2.1.2 Without Channel Side Information at Transmitter

We present the sufficient and necessary condition for the flat channel capacity distribution to be light-tailed, when the channel side information is not known at the transmitter.

Theorem 7. *Consider that the channel side information is only known at the receiver. The capacity distribution of the flat channel is light tailed, if and only if the distribution of the determinant term is fat-tail bounded, i.e.,*

$$\mathbb{E}\Big[(\det(\boldsymbol{I}_{N_R} + \boldsymbol{\Lambda}))^{\theta}\Big] < \infty, \ \exists\theta > 0, \tag{2.76}$$

where $\boldsymbol{H}\boldsymbol{H}^* = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^*$, $\boldsymbol{Q}\boldsymbol{Q}^* = \boldsymbol{Q}^*\boldsymbol{Q} = \boldsymbol{I}_{N_R}$, *and* $\boldsymbol{\Lambda} = diag\{\lambda_1, \ldots, \lambda_{N_R}\}$, $\lambda_i \geqslant 0$. *Equivalently, the condition is expressed as*

$$\mathbb{E}\Big[(\lambda_{\max})^{\theta}\Big] < \infty, \ \exists\theta > 0, \tag{2.77}$$

where $\lambda_{\max} = \max_{1 \leqslant i \leqslant r(\boldsymbol{H})} \lambda_i$, $0 < \lambda_i \in \boldsymbol{\Lambda}$ *and* $r(\boldsymbol{H})$ *is the rank of* $\boldsymbol{H}$.

Proof. For the flat channel without channel side information at the transmitter, the capacity is expressed as [110]

$$c = W \log_2 \det\left(\boldsymbol{I}_{N_R} + \frac{\rho}{N_T}\boldsymbol{H}\boldsymbol{H}^*\right) \tag{2.78}$$

$$= W \log_2 \det\left(\boldsymbol{I}_{N_R} + \frac{\rho}{N_T}\boldsymbol{\Lambda}\right) \tag{2.79}$$

$$= W \sum\nolimits_{i=1}^{r(\boldsymbol{H})} \log_2\left(1 + \frac{\rho}{N_T}\lambda_i\right), \tag{2.80}$$

where $\boldsymbol{H}\boldsymbol{H}^* = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^*$ and $\boldsymbol{Q}\boldsymbol{Q}^* = \boldsymbol{I}$, the second equality follows that $\det(\boldsymbol{I}_m + \boldsymbol{A}\boldsymbol{B}) = \det(\boldsymbol{I}_n + \boldsymbol{B}\boldsymbol{A})$ for $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{C}^{n\times m}$, and the third equality is an equivalent expression.

The proof follows the light-tailed distribution definition, i.e.,

$$\mathbb{E}\left[e^{\theta W \log_2 \det\left(\boldsymbol{I}_{N_R}+\frac{\rho}{N_T}\boldsymbol{\Lambda}\right)}\right] < \infty,\ \exists\theta > 0.$$

Specifically, considering finite rank matrix, we have that, for $\Delta \geqslant 1$, $\Delta^{r(\boldsymbol{H})\theta}\mathbb{E}\left[\prod_{i=1}^{r(\boldsymbol{H})}(1+\lambda_i)^{\theta}\right] \geqslant \mathbb{E}\left[\prod_{i=1}^{r(\boldsymbol{H})}(1+\Delta\lambda_i)^{\theta}\right] \geqslant \mathbb{E}\left[\prod_{i=1}^{r(\boldsymbol{H})}(1+\lambda_i)^{\theta}\right]$; and for $0 < \Delta \leqslant 1$, $\Delta^{r(\boldsymbol{H})\theta}\mathbb{E}\left[\prod_{i=1}^{r(\boldsymbol{H})}(1+\lambda_i)^{\theta}\right] \leqslant \mathbb{E}\left[\prod_{i=1}^{r(\boldsymbol{H})}(1+\Delta\lambda_i)^{\theta}\right] \leqslant \mathbb{E}\left[\prod_{i=1}^{r(\boldsymbol{H})}(1+\lambda_i)^{\theta}\right]$. Thus, we have the following equivalent expressions

$$\begin{aligned}\mathbb{E}\left[\prod_{i=1}^{r(\boldsymbol{H})}(1+\Delta\lambda_i)^{\theta}\right] < \infty,\ 0<\Delta<\infty,\ \exists\theta>0 \\ \iff \mathbb{E}\left[\prod_{i=1}^{r(\boldsymbol{H})}(1+\lambda_i)^{\theta}\right] < \infty,\ \exists\theta > 0. \qquad (2.81)\end{aligned}$$

Considering the two inequalities, $\prod_{i=1}^{r(\boldsymbol{H})}(1+\lambda_i)^{\theta} \leqslant (1+\lambda_{\max})^{\theta r(\boldsymbol{H})}$ and $\prod_{i=1}^{r(\boldsymbol{H})}(1+\lambda_i)^{\theta} \geqslant (1+\lambda_{\max})^{\theta}$, where $\lambda_{\max} = \max_{1\leqslant i\leqslant r(\boldsymbol{H})}\lambda_i$, the sufficient and necessary condition is equivalently written as $\mathbb{E}\left[\prod_{i=1}^{r(\boldsymbol{H})}(1+\lambda_i)^{\theta}\right] < \infty,\ \exists\theta > 0 \iff \mathbb{E}\left[(1+\lambda_{\max})^{\theta}\right] < \infty,\ \exists\theta>0 \iff \mathbb{E}\left[(\lambda_{\max})^{\theta}\right] < \infty,\ \exists\theta > 0$. This completes the proof. □

Remark 15. *Considering the Fredholm determinant [54], for $|z|$ small enough, $\log\det(\boldsymbol{I}+z\boldsymbol{\Lambda}) = \operatorname{Tr}\log(\boldsymbol{I}+z\boldsymbol{\Lambda}) = \sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}z^k\operatorname{Tr}\left[\boldsymbol{\Lambda}^k\right]$, the condition is alternatively expressed as*

$$\mathbb{E}\left[e^{\theta\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\left(\frac{\rho}{N_T}\right)^k\operatorname{Tr}\left[\boldsymbol{\Lambda}^k\right]}\right] < \infty,\ \exists\theta>0, \qquad (2.82)$$

where $\boldsymbol{\Lambda} = \boldsymbol{H}\boldsymbol{H}^*$ or $\boldsymbol{H}\boldsymbol{H}^* = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^*$. Specifically, for $\boldsymbol{H}\boldsymbol{H}^* = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^*$, we have $\operatorname{Tr}\left[\boldsymbol{\Lambda}^k\right] = \sum_{i=1}^{r(\boldsymbol{\Lambda})}(\lambda_i(\boldsymbol{\Lambda}))^k$.

For arbitrary z*, according to the Plemelj-Smithies formulas [54], we have* $\det(\boldsymbol{I} + z\boldsymbol{\Lambda}) = 1 + \sum_{k=1}^{r(\boldsymbol{\Lambda})} \frac{d_k(\boldsymbol{\Lambda})}{k!} z^k$*, thus the condition is expressed as*

$$\mathbb{E}\left[\left(1 + \sum_{k=1}^{r(\boldsymbol{\Lambda})} \frac{d_k(\boldsymbol{\Lambda})}{k!}\left(\frac{\rho}{N_T}\right)^k\right)^\theta\right] < \infty,\ \theta > 0, \tag{2.83}$$

where

$$d_k(\boldsymbol{\Lambda}) = \begin{vmatrix} \operatorname{Tr}\boldsymbol{\Lambda} & k-1 & 0 & 0 & \dots & 0 & 0 \\ \operatorname{Tr}\boldsymbol{\Lambda}^2 & \operatorname{Tr}\boldsymbol{\Lambda} & k-2 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \operatorname{Tr}\boldsymbol{\Lambda}^{k-1} & \operatorname{Tr}\boldsymbol{\Lambda}^{k-2} & \operatorname{Tr}\boldsymbol{\Lambda}^{k-3} & \vdots & \dots & \operatorname{Tr}\boldsymbol{\Lambda} & 1 \\ \operatorname{Tr}\boldsymbol{\Lambda}^{k} & \operatorname{Tr}\boldsymbol{\Lambda}^{k-1} & \operatorname{Tr}\boldsymbol{\Lambda}^{k-2} & \vdots & \dots & \operatorname{Tr}\boldsymbol{\Lambda}^2 & \operatorname{Tr}\boldsymbol{\Lambda} \end{vmatrix}.$$

We present a sufficient condition for the light-tailed property of the frequency-selective channel capacity.

Theorem 8. *Consider that the channel side information is only known at the receiver. The capacity distribution of the frequency-selective channel is light tailed, if*

$$\mathbb{E}\Big[(\lambda_{\max})^\theta\Big] < \infty, \exists\theta > 0 \tag{2.84}$$

where $\lambda_{\max} = \max_{1 \leqslant j \leqslant N} \lambda_{\max}^j$*,* $\lambda_{\max}^j = \max_{1 \leqslant i \leqslant r(\boldsymbol{H}_j)} \lambda_i^j$*,* $\boldsymbol{H}_j$ *is the channel model of each sub-channel and* λ_i^j *is the corresponding eigenvalue of* $\boldsymbol{H}_j\boldsymbol{H}_j^*$*.*

Proof. The proof of the frequency-selective channel scenario follows that the light-tailed distribution of the capacity of each sub-channel implies the light-tailed distribution of the overall channel capacity. □

2.2.2 Random Power Fluctuation

We consider the channel scenario, where the channel side information is known at the receiver and is unknown at the transmitter, and the transmission power randomly fluctuates over the coherence periods and remains constant in each coherence period. Considering the analog to the deterministic power fluctuation, the obtained results are correspondingly the sufficient conditions for the channel scenario with arbitrary channel side information.

For the flat fading MIMO channel $\boldsymbol{H} \in \mathbb{C}^{N_R \times N_T}$, $N_T \in \mathbb{N}$, $N_R \in \mathbb{N}$, when the transmit power is allocated evenly across the transmit antennas during each coherence period, the capacity, in bits per second, is expressed as [110]

$$c_{p,\boldsymbol{H}} = W \log_2 \det \left(\boldsymbol{I}_{N_R} + \frac{1}{N_T N_0 W} p \boldsymbol{H}\boldsymbol{H}^* \right), \tag{2.85}$$

where W is the bandwidth, N_0 is the noise power spectral density, and p is the transmit power that is constant during each coherence period and randomly fluctuates over periods. Equivalently, the capacity is expressed as

$$\begin{aligned} c_{p,\boldsymbol{H}} &= W \textstyle\sum_{i=1}^{r(\boldsymbol{H})} \log_2 \left(1 + \frac{1}{N_T N_0 W} p \lambda_i \right) &(2.86)\\ &\leqslant W r(\boldsymbol{H}) \log_2 \left(1 + \frac{1}{N_T N_0 W} p \lambda_{\max} \right), &(2.87) \end{aligned}$$

where λ_i is the eigenvalue of the matrix $\boldsymbol{H}\boldsymbol{H}^*$ and $r(\boldsymbol{H})$ is the rank of $\boldsymbol{H}$.

We present some preliminary results considering the random power.

Lemma 2. *Consider a flat MIMO channel $\boldsymbol{H} \in \mathbb{C}^{N_R \times N_T}$. The capacity is upper bounded by $c' = a \log_2(1 + bp\lambda_{\max})$, where $a, b \in \mathbb{R}_{>0}$, $p : \Omega \to \mathbb{R}$ is the random power, and $\lambda_{\max}$ is the maximum eigenvalue of $\boldsymbol{H}\boldsymbol{H}^*$.*

1. For the tail property, we have the equivalent results

$$\begin{aligned}
&\overline{F}_{c'}(x) = O(e^{-\theta x}),\ \exists\theta > 0\\
&\qquad\qquad \iff \overline{F}_{p\lambda_{\max}}(x) = O(x^{-\theta}),\ \exists\theta > 0\\
&\qquad\qquad\qquad \iff \overline{F}_{p\,\mathrm{Tr}\,[\boldsymbol{HH^*}]}(x) = O(x^{-\theta}),\ \exists\theta > 0.
\end{aligned} \tag{2.88}$$

2. Alternatively, the tail property is expressed as $\mathbb{E}\Big[e^{\theta c'}\Big] < \infty$, $\exists\theta > 0$, *and we have the equivalent expressions*

$$\begin{aligned}
&\mathbb{E}\Big[(1 + bp\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0\\
&\qquad\qquad \iff \mathbb{E}\Big[(1 + p\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0\\
&\qquad\qquad\qquad \iff \mathbb{E}\Big[(p\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0.
\end{aligned} \tag{2.89}$$

If p and $\lambda_{\max}$ are independent, then $\mathbb{E}\big[(p\lambda_{\max})^{\theta}\big] = \mathbb{E}\big[p^{\theta}\big]\mathbb{E}\big[(\lambda_{\max})^{\theta}\big]$, *and the condition relaxes to* $\mathbb{E}[p] < \infty$ *and* $\mathbb{E}[\lambda_{\max}] < \infty$.

Proof. The proof is analog to the proof of the deterministic power scenario. □

We present the sufficient and necessary condition for the light-tailed property of the capacity.

Theorem 9. *The capacity distribution of the flat channel is light tailed, if and only if the distribution of the determinant term is fat-tail bounded, i.e.,*

$$\mathbb{E}\Big[(\det(\boldsymbol{I}_{N_R} + p\boldsymbol{\Lambda}))^{\theta}\Big] < \infty,\ \exists\theta > 0, \tag{2.90}$$

where $\boldsymbol{HH}^* = \boldsymbol{Q\Lambda Q}^*$, $\boldsymbol{QQ}^* = \boldsymbol{Q}^*\boldsymbol{Q} = \boldsymbol{I}_{N_R}$, *and* $\boldsymbol{\Lambda} = diag\{\lambda_1, \ldots, \lambda_{N_R}\}$, $\lambda_i \geqslant 0$. *Equivalently, the condition is expressed as*

$$\mathbb{E}\Big[(p\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0, \tag{2.91}$$

where $\lambda_{\max} = \max\limits_{1 \leqslant i \leqslant r(\boldsymbol{H})} \lambda_i$, $0 < \lambda_i \in \boldsymbol{\Lambda}$ *and* $r(\boldsymbol{H})$ *is the rank of* $\boldsymbol{H}$.

Proof. The proof is analog to the proof of the deterministic power scenario. $\square$

We present a set of sufficient conditions for the light-tailed capacity and their relationships.

Theorem 10. *Consider a flat MIMO channel* $\boldsymbol{H} : \Omega \to \mathbb{C}^{N_R \times N_T}$ *with random power fluctuation* $p : \Omega \to \mathbb{R}$. *We have the sufficient condition chain for the light-tailed property of the capacity.*

$$\begin{tikzcd}
 & & 0 \arrow[d, Leftrightarrow] & & \\
 & & 1 & & \\
 & 4 \arrow[ur, Rightarrow] & & 2 \arrow[ul, Rightarrow] & \\
5 \arrow[ur, Rightarrow] & & & & 3 \arrow[ul, Rightarrow]
\end{tikzcd}$$

$$0 := \mathbb{E}\Big[(1 + p\lambda_{\max})^{\theta}\Big] < \infty,\ \exists \theta > 0 \tag{2.92}$$

$$1 := \mathbb{E}\Big[(1 + p\,\mathrm{Tr}\,[\boldsymbol{H}\boldsymbol{H}^*])^{\theta}\Big] < \infty,\ \exists \theta > 0 \tag{2.93}$$

$$2 := \mathbb{E}\Big[(\mathrm{Tr}\,[\boldsymbol{I} + p\boldsymbol{H}\boldsymbol{H}^*])^{\theta}\Big] < \infty,\ \exists \theta > 0 \tag{2.94}$$

$$3 := \mathbb{E}\left[\left(\mathrm{Tr}\left[e^{p\boldsymbol{H}\boldsymbol{H}^*}\right]\right)^{\theta}\right] < \infty,\ \exists \theta > 0 \tag{2.95}$$

$$4 := \mathbb{E}\Big[e^{\theta p \lambda_{\max}}\Big] < \infty,\ \exists \theta > 0 \tag{2.96}$$

$$5 := \mathbb{E}\left[\mathrm{Tr}\left[e^{\theta p\boldsymbol{H}\boldsymbol{H}^*}\right]\right] < \infty,\ \exists \theta > 0 \tag{2.97}$$

$$6 := \mathbb{E}\Big[(p\lambda_{\max})^{\theta}\Big] < \infty,\ \exists \theta > 0 \tag{2.98}$$

$$7 := \mathbb{E}\Big[e^{\theta p\,\mathrm{Tr}\,[\boldsymbol{H}\boldsymbol{H}^*]}\Big] < \infty,\ \exists \theta > 0 \tag{2.99}$$

Note we have the equivalent conditions 0 $\iff$ 6 *and* 4 $\iff$ 7*. Particularly, letting $p = 1$, we obtain the corresponding sufficient conditions for the deterministic power fluctuation scenario of arbitrary channel side information.*

Proof. We present the proof of the deterministic power scenario, and the extension to the random power scenario is to replace the matrix $\boldsymbol{H}\boldsymbol{H}^*$ with the scalar multiplication $p\boldsymbol{H}\boldsymbol{H}^*$ of the random variable p and the random matrix $\boldsymbol{H}\boldsymbol{H}^*$.

We have a sufficient condition, i.e., $\mathbb{E}\Big[(1+\lambda_{\max})^{\theta}\Big] < \infty,\ \exists\theta > 0$, which is equivalent to $\mathbb{E}\Big[(1+\mathrm{Tr}\,[\boldsymbol{H}\boldsymbol{H}^*])^{\theta}\Big] < \infty,\ \exists\theta > 0$, which is further equivalent to $\mathbb{E}\left[\left(1+\sum_{i=1}^{r(\boldsymbol{H})}\lambda_i\right)^{\theta}\right] < \infty,\ \exists\theta > 0$. Furthermore, the condition relaxes to be $\mathbb{E}\Big[(\mathrm{Tr}\,[\boldsymbol{I}+\boldsymbol{H}\boldsymbol{H}^*])^{\theta}\Big] < \infty,\ \exists\theta > 0$.

Considering the transfer rule [137], the function inequality $(1+x)^{\vartheta} \leqslant e^{\vartheta x}$, $\vartheta > 0$, implies the partial order $e^{\vartheta \boldsymbol{H}\boldsymbol{H}^*} - (\boldsymbol{I}+\boldsymbol{H}\boldsymbol{H}^*)^{\vartheta} \geqslant 0$, thus $\mathrm{Tr}\left[e^{\vartheta \boldsymbol{H}\boldsymbol{H}^*} - (\boldsymbol{I}+\boldsymbol{H}\boldsymbol{H}^*)^{\vartheta}\right] \geqslant 0$, i.e., $\mathrm{Tr}\left[(\boldsymbol{I}+\boldsymbol{H}\boldsymbol{H}^*)^{\vartheta}\right] \leqslant \mathrm{Tr}\left[e^{\vartheta \boldsymbol{H}\boldsymbol{H}^*}\right]$, which implies a further relaxed condition $\mathbb{E}\left[\left(\mathrm{Tr}\left[e^{\boldsymbol{H}\boldsymbol{H}^*}\right]\right)^{\theta}\right] < \infty,\ \exists\theta > 0$.

In addition, according to the inequality, $(1+\rho\lambda_{\max})^{\vartheta} \leqslant e^{\vartheta\rho\lambda_{\max}} \leqslant \mathrm{Tr}\left[e^{\vartheta\rho \boldsymbol{H}\boldsymbol{H}^*}\right]$, where $\vartheta > 0$ and the last inequality follows that the spectral mapping theorem [137], thus we have the relaxed condition $\mathbb{E}\Big[\mathrm{Tr}\left[e^{\theta \boldsymbol{H}\boldsymbol{H}^*}\right]\Big] < \infty,\ \exists\theta > 0$. Similarly, $(1+\mathrm{Tr}\,[\boldsymbol{H}\boldsymbol{H}^*])^{\vartheta} \leqslant e^{\vartheta\,\mathrm{Tr}\,[\boldsymbol{H}\boldsymbol{H}^*]}$, $\vartheta > 0$, thus we have the relaxed condition $\mathbb{E}\Big[e^{\theta\,\mathrm{Tr}\,[\boldsymbol{H}\boldsymbol{H}^*]}\Big] < \infty,\ \exists\theta > 0$, which is equivalent to $\mathbb{E}\big[e^{\theta\lambda_{\max}}\big] < \infty,\ \exists\theta > 0$. □

Remark 16. *Particularly, we have* $\left(\mathrm{Tr}\left[e^{\boldsymbol{X}}\right]\right)^{\vartheta} \nleqslant \mathrm{Tr}\left[e^{\vartheta\boldsymbol{X}}\right],\ \exists\vartheta > 0$, $(\mathrm{Tr}\,[\boldsymbol{X}])^{\vartheta} \nleqslant \mathrm{Tr}\left[\boldsymbol{X}^{\vartheta}\right],\ \exists\vartheta > 0$, $e^{\vartheta\,\mathrm{Tr}[\boldsymbol{X}]} \nleqslant \mathrm{Tr}\left[e^{\vartheta\boldsymbol{X}}\right],\ \exists\vartheta > 0$, *where* $\boldsymbol{X} \in \mathfrak{X}_{\geqslant 0}$*. For example,* $\boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ *and* $\vartheta = 2$.

Remark 17. *The channel model $\boldsymbol{H}\boldsymbol{H}^*$ is the product formulation of the large-scale fading and small-scale fading effects.*

Theorem 11. *Consider the frequency-selective MIMO channel with random power fluctuation. If each sub-channel satisfies any one of the sufficient conditions in Theorem 10, then the distribution of the overall channel capacity is light-tailed.*

Proof. The proof follows that the light-tail property is preserved for the sum of random variables. □

Remark 18. *The frequency-selective channel provides an additional degree of freedom or a diversity for dependence control, i.e., the power can be randomly allocated to each parallel channel with a total power constraint [138, p. 182], which also provides the flexibility for each parallel channel to possess temporal dependence in power.*

Remark 19. *The signal frequency randomly varies with time, due to the fading of the in-phase and quadrature components of the signals [67]. On the other hand, the carrier frequency is free to configure such that it varies randomly through different periods, i.e., the random carrier frequency configuration provides an additional degree of freedom for dependence control.*

Remark 20. *The tail property of the capacity is determined by the product of the random power and the random eigenvalues of the channel matrix. Thus, it is necessary to investigate the tail property of the product of two random variables. It is reasonable to assume independence between these two random variables, because the channel side information is not necessarily known at the transmitter. On the other hand, it is interesting to take into account the dependence for refinement.*

2.3 Random Variable Arithmetic

For a simple wireless channel, the tail property of capacity is determined by the product of the random parameters of fading and the random power, for a compound wireless channel, the tail property of capacity is determined by the sum of the capacities of each sub-channel. Thus, it is interesting to study the tail property of the product and sum of random variables.

We study the impact of the tail property of one random variable on the overall product or sum distribution. We consider the nonnegative functions, $f(x)$ and $g(x)$, and define the asymptotic notations, $f(x) = O(g(x)) \iff \limsup\limits_{x\to\infty} \frac{f(x)}{g(x)} < \infty$, $f(x) = \Omega(g(x)) \iff \liminf\limits_{x\to\infty} \frac{f(x)}{g(x)} > 0$, $f(x) = \Theta(g(x)) \iff f(x) = O(g(x)) \bigcap f(x) = \Omega(g(x))$, $f(x) = o(g(x)) \iff \lim\limits_{x\to\infty} \frac{f(x)}{g(x)} = 0$, $f(x) = \omega(g(x)) \iff \lim\limits_{x\to\infty} \frac{f(x)}{g(x)} = \infty$, and $f(x) \sim g(x) \iff \lim\limits_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

We define a class of functions $\mathfrak{F}$, $\forall \varphi \in \mathfrak{F}$, $\varphi : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$, such that $\lim\limits_{x\to\infty} \varphi(x) = \infty$ and $\lim\limits_{x\to\infty} \frac{\varphi(x)}{x} = 0$, i.e.,

$$\mathfrak{F} = \left\{ \varphi : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}; \lim_{x\to\infty} \varphi(x) = \infty, \lim_{x\to\infty} \frac{\varphi(x)}{x} = 0 \right\}. \tag{2.100}$$

For example, $\varphi(x) = x^{\alpha}$, $0 < \alpha < 1$, or $\varphi(x) = \log(x)$. This class of functions are useful in decomposing the distribution function of the product or sum of random variables.

We study the asymptotic behavior of the composition of the function $\mathfrak{F}$ and some classes of tail distributions, e.g., the light-tail distribution, the regularly varying distribution $F \in \mathcal{R}_{\geqslant 0}$, and the long-tail distribution $F \in \mathcal{L}$ (containing the subexponential distribution as a subset).

Lemma 3. *Consider the independent random variables $X_i : \Omega \to \mathbb{R}_{\geqslant 0}$, $i \in \{1, 2\}$.*

1. If $F_1 \in \mathcal{L}$, i.e., $\lim\limits_{x\to\infty} \frac{\overline{F}_{X_1}(x-y)}{\overline{F}_{X_1}(x)} = 1$, $\forall y > 0$, then, $\overline{F}_{X_1}(x - \varphi(x)) \sim$

$\overline{F}_{X_1}(x)$ *and* $\overline{F}_{X_1}\Big(\log \frac{x}{\varphi(x)}\Big) \sim \overline{F}_{X_1}(\log x)$, $\forall \varphi \in \mathfrak{F}$. *We specify* $\varphi(x) = x^{\alpha}$, $0 < \alpha < 1$.

(a) *If* $\overline{F}_{X_1}(x) = \Omega\big(x^{-\theta_1}\big)$, $\theta_1 > 0$, *and* $\overline{F}_{X_2}(x) = O\big(e^{-\theta_2 x}\big)$, $\theta_2 > 0$, *then,* $\overline{F}_{X_2}(\varphi(x)) = O\big(\overline{F}_{X_1}(x)\big)$.

(b) *If* $\overline{F}_{X_1}(x) = \Omega\big(x^{-\theta_1}\big)$, $\theta_1 > 0$, *and* $\overline{F}_{X_2}(x) = O\big(x^{-\theta_2}\big)$, $\theta_2 > 0$, *and* $\alpha\theta_2 > \theta_1$, *then,* $\overline{F}_{X_2}(\varphi(x)) = O\big(\overline{F}_{X_1}(x)\big)$.

2. *If* $F_1 \in \mathcal{R}$, *i.e.,* $\overline{F}_{X_1}(x) = L_1(x)x^{-\theta_1}$, *where* $\theta_1 \geqslant 0$, *and* $\lim_{x\to\infty} \frac{L_1(tx)}{L_1(x)} = 1$, $\forall t > 0$, *then,* $\overline{F}_{X_1}(x - \varphi(x)) \sim \overline{F}_{X_1}(x)$, $\forall \theta_1 \geqslant 0$, *and* $\overline{F}_{X_1}\Big(\frac{x}{\varphi(x)}\Big) = \omega\big(\overline{F}_{X_1}(x)\big)$, $\forall \theta_1 > 0$ *and* $\overline{F}_{X_1}\Big(\frac{x}{\varphi(x)}\Big) \sim \overline{F}_{X_1}(x)$ *for* $\theta_1 = 0$, $\forall \varphi \in \mathfrak{F}$. *We specify* $\varphi(x) = x^{\alpha}$, $0 < \alpha < 1$.

(a) *If* $\overline{F}_{X_2}(x) = O\big(e^{-\theta_2 x}\big)$, $\theta_2 > 0$, *then,* $\overline{F}_{X_2}(\varphi(x)) = O\big(\overline{F}_{X_1}(x)\big)$. *If* $\overline{F}_{X_2}(x) = \Theta\big(e^{-\theta_2 x}\big)$, $\theta_2 > 0$, *then,* $\overline{F}_{X_2}(\varphi(x)) = o\big(\overline{F}_{X_1}(x)\big)$.

(b) *If* $\overline{F}_{X_2}(x) = O\big(x^{-\theta_2}\big)$, $\theta_2 > 0$, *and* $\alpha\theta_2 > \theta_1$, *then,* $\overline{F}_{X_2}(\varphi(x)) = O\big(\overline{F}_{X_1}(x)\big)$. *If* $\overline{F}_{X_2}(x) = \Theta\big(x^{-\theta_2}\big)$, $\theta_2 > 0$, *and* $\alpha\theta_2 > \theta_1$, *then,* $\overline{F}_{X_2}(\varphi(x)) = o\big(\overline{F}_{X_1}(x)\big)$.

3. *If* $\overline{F}_{X_1}(x) = \Theta\big(x^{-\theta_1}\big)$, $\theta_1 > 0$, *then,* $\overline{F}_{X_1}(x - \varphi(x)) = \Theta\big(\overline{F}_{X_1}(x)\big)$ *and* $\overline{F}_{X_1}\Big(\frac{x}{\varphi(x)}\Big) = \omega\big(\overline{F}_{X_1}(x)\big)$, $\forall \varphi \in \mathfrak{F}$. *We specify* $\varphi(x) = x^{\alpha}$, $0 < \alpha < 1$.

(a) *If* $\overline{F}_{X_2}(x) = O\big(e^{-\theta_2 x}\big)$, $\theta_2 > 0$, *then,* $\overline{F}_{X_2}(\varphi(x)) = O\big(\overline{F}_{X_1}(x)\big)$. *If* $\overline{F}_{X_2}(x) = \Theta\big(e^{-\theta_2 x}\big)$, $\theta_2 > 0$, *then,* $\overline{F}_{X_2}(\varphi(x)) = o\big(\overline{F}_{X_1}(x)\big)$.

(b) *If* $\overline{F}_{X_2}(x) = O\big(x^{-\theta_2}\big)$, $\theta_2 > 0$, *and* $\alpha\theta_2 \geqslant \theta_1$, *then,* $\overline{F}_{X_2}(\varphi(x)) = O\big(\overline{F}_{X_1}(x)\big)$. *If* $\overline{F}_{X_2}(x) = \Theta\big(x^{-\theta_2}\big)$, $\theta_2 > 0$, *and* $\alpha\theta_2 > \theta_1$, *then,* $\overline{F}_{X_2}(\varphi(x)) = o\big(\overline{F}_{X_1}(x)\big)$.

4. *If* $\overline{F}_{X_1}(x) = \Theta\big(e^{-\theta_1 x}\big)$, $\theta_1 > 0$, *then,* $\overline{F}_{X_1}\Big(\frac{x}{\varphi(x)}\Big) = \omega\big(\overline{F}_{X_1}(x)\big)$ *and* $\overline{F}_{X_1}(x - \varphi(x)) = \omega\big(\overline{F}_{X_1}(x)\big)$, $\forall \varphi \in \mathfrak{F}$.

(a) *If* $\overline{F}_{X_2}(x) = \Theta\big(e^{-\theta_2 x}\big)$, $\theta_2 > 0$, *then,* $\overline{F}_{X_2}(\varphi(x)) = \omega\big(\overline{F}_{X_1}(x)\big)$, $\forall \varphi \in \mathfrak{F}$.

(b) If $\overline{F}_{X_2}(x) = \Theta\big(x^{-\theta_2}\big)$, $\theta_2 > 0$, *then,* $\overline{F}_{X_2}(\varphi(x)) = \omega\big(\overline{F}_{X_1}(x)\big)$, $\forall \varphi \in \mathfrak{F}$.

Proof. The proofs follow the given conditions and the definition of the asymptotic symbols.

1. By the given condition, we have $\lim_{x\to\infty} \frac{\overline{F}_{X_1}(x-\varphi(x))}{\overline{F}_{X_1}(x)} = 1$, $\forall \varphi \in \mathfrak{F}$. Thus, $\overline{F}_{X_1}(x-\varphi(x)) \sim \overline{F}_{X_1}(x)$. Considering $F \circ \log \in \mathcal{R}_0$, $\forall F \in \mathcal{L}$, we have $\lim_{x\to\infty} \frac{\overline{F}_{X_1}\left(\log \frac{x}{\varphi(x)}\right)}{\overline{F}_{X_1}(\log x)} = \lim_{x\to\infty} \left(\frac{\log \frac{x}{\varphi(x)}}{\log x}\right)^{-\theta_1} = \lim_{x\to\infty} \left(1 - \frac{\log \varphi(x)}{\log x}\right)^{-\theta_1} = 1$, $\exists \theta_1 > 0$, $\forall \varphi \in \mathfrak{F}$. Thus, $\overline{F}_{X_1}\left(\log \frac{x}{\varphi(x)}\right) \sim \overline{F}_{X_1}(\log x)$.

Since $\overline{F}_{X_1}(x) = \Omega\big(x^{-\theta_1}\big)$, $\theta_1 > 0 \iff \exists x_1 > 0$, $\forall x > x_1$, $\exists C_1 > 0$, $\overline{F}_{X_1}(x) \geqslant C_1 x^{-\theta_1}$; $\overline{F}_{X_2}(x) = O\big(e^{-\theta_2 x}\big)$, $\theta_2 > 0 \iff \exists x_2 > 0$, $\forall x > x_2$, $\exists C_2 > 0$, $\overline{F}_{X_2}(x) \leqslant C_2 e^{-\theta_2 x}$. Letting $x_0 = \max(x_1, x_2)$ and $\overline{\lim_{x\to\infty}} := \inf_{x^* > x_0} \sup_{x \geqslant x^*}$, we obtain

$$\limsup_{x\to\infty} \frac{\overline{F}_{X_2}(\varphi(x))}{\overline{F}_{X_1}(x)} \leqslant \overline{\lim_{x\to\infty}} \frac{C_2 e^{-\theta_2 \varphi(x)}}{C_1 x^{-\theta_1}} = \overline{\lim_{x\to\infty}} \frac{\frac{C_2}{C_1} x^{\theta_1}}{\sum_{n=0}^{\infty} \frac{(\theta_2 x^{\alpha})^n}{n!}} = 0.$$

Thus, we obtain $\overline{F}_{X_2}(\varphi(x)) = O\big(\overline{F}_{X_1}(x)\big)$.

Similarly, if $\overline{F}_{X_1}(x) = \Omega\big(x^{-\theta_1}\big)$, $\exists \theta_1 > 0$, and $\overline{F}_{X_2}(x) = O\big(x^{-\theta_2}\big)$, $\exists \theta_2 > 0$, $\varphi(x) = x^{\alpha}$, $0 < \alpha < 1$, and $\alpha\theta_2 > \theta_1$, then

$$\limsup_{x\to\infty} \frac{\overline{F}_{X_2}(\varphi(x))}{\overline{F}_{X_1}(x)} \leqslant \inf_{x^* > x_0} \sup_{x \geqslant x^*} \frac{C_2 (\varphi(x))^{-\theta_2}}{C_1 x^{-\theta_1}} = \inf_{x^* > x_0} \sup_{x \geqslant x^*} \frac{C_2}{C_1} \frac{x^{\theta_1}}{x^{\alpha\theta_2}} = 0.$$

When $\alpha\theta_2 = \theta_1$, the limit is also finite. Thus, we obtain $\overline{F}_{X_2}(\varphi(x)) = O\big(\overline{F}_{X_1}(x)\big)$.

2. By the given condition, we have

$$\lim_{x\to\infty} \frac{\overline{F}_{X_1}(x-\varphi(x))}{\overline{F}_{X_1}(x)} = \lim_{x\to\infty} \frac{L_1(x-\varphi(x))}{L_1(x)} \left(1 - \frac{\varphi(x)}{x}\right)^{\theta_1} = 1,$$

$\forall \varphi \in \mathfrak{F}$, because we obtain $t = 1$ for $x - \varphi(x) = tx$ as $x \to \infty$. Thus, we obtain $\overline{F}_{X_1}(x - \varphi(x)) \sim \overline{F}_{X_1}(x)$. Similarly, we have $\lim\limits_{x\to\infty} \frac{\overline{F}_{X_1}\left(\frac{x}{\varphi(x)}\right)}{\overline{F}_{X_1}(x)} = \lim\limits_{x\to\infty} \frac{L_1\left(\frac{x}{\varphi(x)}\right)}{L_1(x)} (\varphi(x))^{\theta_1}$, $\forall \varphi \in \mathfrak{F}$. Thus, $\overline{F}_{X_1}\left(\frac{x}{\varphi(x)}\right) = \omega\big(\overline{F}_{X_1}(x)\big)$, $\forall \theta_1 > 0$ and $\overline{F}_{X_1}\left(\frac{x}{\varphi(x)}\right) \sim \overline{F}_{X_1}(x)$ for $\theta_1 = 0$.

Since $\overline{F}_{X_2}(x) = O\big(e^{-\theta_2 x}\big)$, $\theta_2 > 0 \iff \exists x_0 > 0$, $\forall x > x_0$, $\exists C_2 > 0$, $\overline{F}_{X_2}(x) \leqslant C_2 e^{-\theta_2 x}$. Then, letting $\overline{\lim\limits_{x\to\infty}} := \inf\limits_{x^* > x_0} \sup\limits_{x \geqslant x^*}$,

$$\limsup_{x\to\infty} \frac{\overline{F}_{X_2}(\varphi(x))}{\overline{F}_{X_1}(x)} \leqslant \overline{\lim_{x\to\infty}} \frac{C_2 e^{-\theta_2 \varphi(x)}}{L_1(x) x^{-\theta_1}} = \overline{\lim_{x\to\infty}} \frac{C_2}{L_1(x)} \frac{x^{\theta_1}}{\sum_{n=0}^{\infty} \frac{(\theta_2 x^{\alpha})^n}{n!}} = 0,$$

where $\lim_{x\to\infty} x^{\epsilon} L_1(x) = \infty$, $\forall \epsilon > 0$, follows the representation theorem of the slowly varying function. Thus, we obtain $\overline{F}_{X_2}(\varphi(x)) = O\big(\overline{F}_{X_1}(x)\big)$.

Similarly, for $\overline{F}_{X_2}(x) = O\big(x^{-\theta_2}\big)$, $\alpha\theta_2 > \theta_1$, we have

$$\limsup_{x\to\infty} \frac{\overline{F}_{X_2}(\varphi(x))}{\overline{F}_{X_1}(x)} \leqslant \inf_{x^* > x_0} \sup_{x \geqslant x^*} \frac{C_2 x^{-\theta_2 \alpha}}{L_1(x) x^{-\theta_1}} = 0.$$

The proof of the rest results follows the previous proofs, by considering the complementary $\liminf\limits_{x\to\infty}(\cdot)$ and by noticing the fact that, the limit $\lim\limits_{x\to\infty}(\cdot)$ exists if and only if $\liminf\limits_{x\to\infty}(\cdot) = \limsup\limits_{x\to\infty}(\cdot)$.

3. Since $\overline{F}_{X_1}(x) = \Theta\big(x^{-\theta_1}\big)$, $\theta_1 > 0$, we have, $\exists C_1^u > 0$, $\exists C_1^l > 0$, $\exists x_0 > 0$, $\forall x > x_0$, $C_1^l x^{-\theta_1} \leqslant \overline{F}_{X_1}(x) \leqslant C_1^u x^{-\theta_1}$, and $\exists C_1^{u'} > 0$, $\exists C_1^{l'} > 0$, $\exists x_0' > 0$, $\forall x > x_0'$, $C_1^{l'} x^{-\theta_1} \leqslant \overline{F}_{X_1}(x) \leqslant C_1^{u'} x^{-\theta_1}$. Let $x_0^* = \max(x_0, x_0')$. Then, we obtain $\limsup\limits_{x\to\infty} \frac{\overline{F}_{X_1}(x-\varphi(x))}{\overline{F}_{X_1}(x)} \leqslant \inf\limits_{x^* > x_0^*} \sup\limits_{x \geqslant x^*} \frac{C_1^{u'}}{C_1^l} \left(1 - \frac{\varphi(x)}{x}\right)^{\theta_1} = \frac{C_1^{u'}}{C_1^l}$, and $\liminf\limits_{x\to\infty} \frac{\overline{F}_{X_1}(x-\varphi(x))}{\overline{F}_{X_1}(x)} \geqslant \frac{C_1^{l'}}{C_1^u}$, where $\frac{C_1^{l'}}{C_1^u} \leqslant \frac{C_1^{u'}}{C_1^l}$, because $C_1^l \leqslant C_1^u$ and $C_1^{l'} \leqslant C_1^{u'}$. Thus, we obtain $\overline{F}_{X_1}(x - \varphi(x)) = \Theta\big(\overline{F}_{X_1}(x)\big)$.

Since $\overline{F}_{X_1}(x) = \Theta\big(x^{-\theta_1}\big)$, $\theta_1 > 0$, we have that $\limsup\limits_{x\to\infty} \frac{\overline{F}_{X_1}\left(\frac{x}{\varphi(x)}\right)}{\overline{F}_{X_1}(x)} \leqslant$

$\inf_{x^*>x_0} \sup_{x \geqslant x^*} C^*(\varphi(x))^{\theta_1} = \infty, \forall C^* > 0, \forall \varphi \in \mathfrak{F}$. Similarly, $\liminf_{x\to\infty} \frac{\overline{F}_{X_1}\left(\frac{x}{\varphi(x)}\right)}{\overline{F}_{X_1}(x)} \geqslant \infty$. Thus, we obtain $\overline{F}_{X_1}\left(\frac{x}{\varphi(x)}\right) = \omega\big(\overline{F}_{X_1}(x)\big)$.

The proofs of the rest results are analogical to the previous proofs, by considering $\lim_{x\to\infty}(\cdot)$, $\liminf_{x\to\infty}(\cdot)$, and $\limsup_{x\to\infty}(\cdot)$.

4. Since $\overline{F}_{X_1}(x) = \Theta\big(e^{-\theta_1 x}\big)$, $\theta_1 > 0$, we have

$$\limsup_{x\to\infty} \frac{\overline{F}_{X_1}(x/\varphi(x))}{\overline{F}_{X_1}(x)} \leqslant \inf_{x^*>x_0} \sup_{x\geqslant x^*} C^* e^{-\theta_1(x/\varphi(x)-x)} = \infty,$$

$\forall 0 < C^* < \infty, \forall \varphi \in \mathfrak{F}$; and $\liminf_{x\to\infty} \frac{\overline{F}_{X_1}\left(\frac{x}{\varphi(x)}\right)}{\overline{F}_{X_1}(x)} \geqslant \sup_{x^*>x_0} \inf_{x\geqslant x^*} C^\star e^{-\theta_1\left(\frac{x}{\varphi(x)}-x\right)} = \infty$, $\forall 0 < C^\star < \infty$ and $\forall \varphi \in \mathfrak{F}$. Thus, $\overline{F}_{X_1}\left(\frac{x}{\varphi(x)}\right) = \omega\big(\overline{F}_{X_1}(x)\big)$. The proof of the other result follows analogically.

Since $\overline{F}_{X_i}(x) = \Theta\big(e^{-\theta_i x}\big)$, $\theta_i > 0$, $i \in \{1, 2\}$, we have

$$\limsup_{x\to\infty} \frac{\overline{F}_{X_2}(\varphi(x))}{\overline{F}_{X_1}(x)} \leqslant \inf_{x^*>x_0} \sup_{x\geqslant x^*} C^* \frac{e^{-\theta_2\varphi(x)}}{e^{-\theta_1 x}} = \infty,$$

where the last step follows $\lim_{x\to\infty} \frac{x}{\varphi(x)} = \infty$. Similarly, $\liminf_{x\to\infty} \frac{\overline{F}_{X_2}(\varphi(x))}{\overline{F}_{X_1}(x)} \geqslant \infty$. Thus, $\overline{F}_{X_2}(\varphi(x)) = \omega\big(\overline{F}_{X_1}(x)\big)$.

Since $\overline{F}_{X_2}(x) = \Theta\big(x^{-\theta_2}\big)$, $\theta_2 > 0$, we have $\overline{F}_{X_2}(x) = \omega\big(\overline{F}_{X_1}(x)\big)$, $\forall \theta_2 > 0$. Letting $x = \varphi(y)$, then $\overline{F}_{X_2}(\varphi(y)) = \omega\big(\overline{F}_{X_1}(\varphi(y))\big)$, $\forall \varphi \in \mathfrak{F}$. Since $\overline{F}_{X_1}(\varphi(y)) = \overline{F}_{X_1}(y)$, thus, $\overline{F}_{X_2}(\varphi(y)) = \omega\big(\overline{F}_{X_1}(y)\big)$, $\forall \theta_2 > 0$, $\forall \varphi \in \mathfrak{F}$.

□

Remark 21. *It is interesting to notice that, from the light-tail to the heavy-tail distributions, the tail behaviors go from* $\overline{F}_{X_1}\left(\frac{x}{\varphi(x)}\right) = \omega\big(\overline{F}_{X_1}(x)\big)$ *and* $\overline{F}_{X_1}(x - \varphi(x)) = \omega\big(\overline{F}_{X_1}(x)\big)$ *to* $\overline{F}_{X_1}\left(\frac{x}{\varphi(x)}\right) \sim \overline{F}_{X_1}(x)$ *and* $\overline{F}_{X_1}(x - \varphi(x)) \sim$

$\overline{F}_{X_1}(x)$, $\forall\varphi \in \mathfrak{F}$. It is impossible to have $\overline{F}_{X_1}\Big(\frac{x}{\varphi(x)}\Big) = o\big(\overline{F}_{X_1}(x)\big)$ or $\overline{F}_{X_1}(x-\varphi(x)) = o\big(\overline{F}_{X_1}(x)\big)$, because the complementary cumulative distribution function is non-increasing.

Remark 22. *It is known that [42] the distribution function $F \in \mathcal{L}$ if and only if $\overline{F} \circ \log \in \mathcal{R}_0$, where $(f \circ g)(x) = f(g(x))$. The distribution function F is regularly varying, i.e., $F \in \mathcal{R}_{>0}$, if and only if, there exists a positive function, $a(t)$, such that [45]*

$$\lim_{t\to\infty} \frac{F(tx) - F(t)}{a(t)} = \frac{1-x^{\alpha}}{\alpha}, \; x > 0. \tag{2.101}$$

These distribution functions have polynomially decaying tail. Letting $\alpha \to 0$, we obtain [45]

$$\lim_{t\to\infty} \frac{F(tx) - F(t)}{a(t)} = \log(x), \; x > 0, \tag{2.102}$$

which characterizes a class of super-heavy distribution functions with slowly varying tails $\mathcal{R}_0$.

Remark 23. *It is interesting to define and study a new tail behavior, $\limsup_{x\to\infty} e^{\epsilon\varphi(x)}\overline{F}(x) = \infty$, $\forall\epsilon > 0$, $\exists\varphi \in \mathfrak{F}$. Note the function $\overline{F}(x)$ is heavy-tailed [48], if and only if $\limsup_{x\to\infty} e^{\epsilon x}\overline{F}(x) = \infty$, $\forall\epsilon > 0$.*

We present a necessary condition for the product of random variables to be light-tailed.

Theorem 12. *Consider the independent random variables $X_i : \Omega \to \mathbb{R}$, $i \in \{1, \ldots, N\}$. The necessary condition for the existence of the moment generating function of the product of the random variables, $X = \prod_{i=1}^{N} X_i$, is the existence of the means of all the random variables, i.e.,*

$$\mathbb{E}\Big[e^{\theta X}\Big] < \infty, \; \exists\theta > 0 \implies \mathbb{E}[X_i] < \infty, \; \forall i \in \{1, \ldots, N\}. \tag{2.103}$$

Proof. It is easy to show that $\mathbb{E}_X\big[e^{\theta X}\big] = \mathbb{E}_{X_1}\big[\mathbb{E}_{X_2}\big[\ldots\mathbb{E}_{X_N}\big[e^{\theta X}\big]\big]\big] \geqslant e^{\theta\prod_{i=1}^N \mathbb{E}[X_i]}$, where the equality follows the independence assumption and the inequality follows the Jensen's inequality. □

Remark 24. *The existence of the mean of each random variables is not a sufficient condition for the existence of the moment generating function of the product. For example, there can be no moment generating function for the product of $N \geqslant 3$ independent standard normal random variables [131], except the product of $N = 2$ normal random variables [31].*

We present a sufficient condition on a random variable, whose product with a fat-tail type random variable remains a fat-tail type random variable.

Theorem 13. *Consider the independent random variables $X_i : \Omega \to \mathbb{R}_{\geqslant 0}$, $i \in \{1, \ldots, N\}$. Suppose $\overline{F}_{X_1}(x) = \Theta\big(x^{-\theta}\big)$, and $\mathbb{E}\Big[X_j^\theta\Big] < \infty$, $\forall 2 \leqslant j \leqslant N$, $\exists\theta > 0$. Then, we have*

$$\overline{F}_{\prod_{i=1}^N X_i}(x) = \Theta\Big(x^{-\theta}\Big). \tag{2.104}$$

Proof. We prove the case of $N = 2$ and the proof of the case $N > 2$ follows by the iteration of the same procedure.

Considering the independence between X_1 and X_2, for $x > 0$, we have $\overline{F}_{X_1X_2}(x) = \mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)\Big]$, which is reformulated as

$$\mathbb{E}\left[\overline{F}_{X_1}\left(\frac{x}{X_2}\right)\right] = \mathbb{E}\left[\overline{F}_{X_1}\left(\frac{x}{X_2}\right)1_{0<X_2\leqslant\varphi_2(x)}\right] + \mathbb{E}\left[\overline{F}_{X_1}\left(\frac{x}{X_2}\right)1_{X_2>\varphi_2(x)}\right].$$

Since $\overline{F}_{X_1}(x) = \Theta\big(x^{-\theta}\big)$, we have, $\exists x_0 > 0$, $\exists C_1, C_1' > 0$, $\forall x > x_0$ and $\forall x/x' > x_0$, $\overline{F}_{X_1}(x) \geqslant C_1 x^{-\theta}$ and $\overline{F}_{X_1}(x/x') \leqslant C_1'(x/x')^{-\theta}$, and

$\limsup_{x\to\infty} \frac{\overline{F}_{X_1}(x/x')}{\overline{F}_{X_1}(x)} \leqslant \frac{C_1'}{C_1}(x')^\theta$. Then,

$$\limsup_{x\to\infty} \lim_{\varphi_2(x)\to\infty} \int_{(\varphi_2(x),\infty)} \frac{\mathbb{P}(X_1 > x/x')}{\mathbb{P}(X_1 > x)} \mathbb{P}_{X_2}(dx')$$
$$= \lim_{\varphi_2(x)\to\infty} \int_{(\varphi_2(x),\infty)} \limsup_{x\to\infty} \frac{\mathbb{P}(X_1 > x/x')}{\mathbb{P}(X_1 > x)} \mathbb{P}_{X_2}(dx') \leqslant 0.$$

Similarly, $\liminf_{x\to\infty} \lim_{\varphi_2(x)\to\infty} \int_{(\varphi_2(x),\infty)} \frac{\mathbb{P}(X_1>x/x')}{\mathbb{P}(X_1>x)} \mathbb{P}_{X_2}(dx') \geqslant 0$. Thus, we obtain $\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{X_2>\varphi_2(x)}\Big] = o\big(x^{-\theta}\big)$.

Similarly, we have

$$\limsup_{x\to\infty} \lim_{\varphi_2(x)\to\infty} \int_{(0,\varphi_2(x)]} \frac{\mathbb{P}(X_1 > x/x')}{\mathbb{P}(X_1 > x)} \mathbb{P}_{X_2}(dx')$$
$$= \lim_{\varphi_2(x)\to\infty} \int_{(0,\varphi_2(x)]} \limsup_{x\to\infty} \frac{\mathbb{P}(X_1 > x/x')}{\mathbb{P}(X_1 > x)} \mathbb{P}_{X_2}(dx') \leqslant C^*\mathbb{E}\Big[(X_2)^\theta\Big],$$

$\exists C^* > 0$. On the other hand, $\liminf_{x\to\infty} \lim_{\varphi_2(x)\to\infty} \int_{(0,\varphi_2(x)]} \frac{\mathbb{P}(X_1>x/x')}{\mathbb{P}(X_1>x)} \mathbb{P}_{X_2}(dx') \geqslant C^\star\mathbb{E}\Big[(X_2)^\theta\Big]$, $\exists C^\star > 0$. Thus, $\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{0<X_2\leqslant\varphi_2(x)}\Big] = \Theta\big(x^{-\theta}\big)$.

The proof completes by the fact that, if $f(x) = o(h(x))$ and $g(x) = \Theta(h(x))$ then $f(x) + g(x) = \Theta(h(x))$. □

Remark 25. *If there exists one and only one random variable,* $\mathbb{E}\Big[X_j^\theta\Big] = 0$, $j \in \{2, 3, \ldots, N\}$, $\forall\theta > 0$, *then* $\overline{F}_{\prod_{i=1}^N X_i}(x) = o\big(x^{-\theta}\big)$. *Letting this random variable be the last one for multiplication yields the proof.*

Remark 26. *Since* $f(x) \sim g(x) \implies f(x) = \Theta(g(x))$, *if* $\overline{F}_{X_1}(x) \sim C_1x^{-\theta}$, $\exists C_1 > 0$, *and* $\mathbb{E}\Big[X_j^\theta\Big] < \infty$, $\forall 2 \leqslant j \leqslant N$, $\exists\theta > 0$, *then,* $\overline{F}_{\prod_{i=1}^N X_i}(x) = \Theta\big(x^{-\theta}\big)$.

Remark 27. *If* $\overline{F}_{X_1}(x) \sim C_1x^{-\theta}$, $\exists C_1 > 0$, *and* $\mathbb{E}\Big[X_j^\theta\Big] < \infty$, $\forall 2 \leqslant j \leqslant N$,

$\exists\theta > 0$, then,

$$\overline{F}_{\prod_{i=1}^{N} X_i}(x) \sim \prod_{j=2}^{N} \mathbb{E}\Big[X_j^{\theta}\Big] \cdot C_1 x^{-\theta}. \tag{2.105}$$

The proof of the case $N = 2$ is available in [18] and the proof of the general case follows by iteration.

We present a sufficient condition on a random variable, whose product with a fat-tail upper bounded random variable remains a fat-tail upper bounded random variable.

Theorem 14. *Consider the independent random variables $X_i : \Omega \to \mathbb{R}_{\geqslant 0}$, $i \in \{1, \ldots, N\}$. Suppose $\overline{F}_{X_1}(x) = O\big(x^{-\theta}\big)$, $\exists\varphi_j \in \mathfrak{F}$, $\overline{F}_{X_j}(\varphi_j(x)) = O\big(x^{-\theta}\big)$, and $\mathbb{E}\Big[X_j^{\theta}\Big] < \infty$, $\forall 2 \leqslant j \leqslant N$, $\exists\theta > 0$. Then, we have*

$$\overline{F}_{\prod_{i=1}^{N} X_i}(x) = O\Big(x^{-\theta}\Big). \tag{2.106}$$

Proof. We prove the case of $N = 2$ and the proof of the case $N > 2$ follows by the iteration of the same procedure.

Considering the independence between X_1 and X_2, for $x > 0$, we have $\overline{F}_{X_1X_2}(x) = \mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)\Big]$, which is reformulated as

$$\mathbb{E}\left[\overline{F}_{X_1}\left(\frac{x}{X_2}\right)\right] = \mathbb{E}\left[\overline{F}_{X_1}\left(\frac{x}{X_2}\right)1_{0<X_2\leqslant\varphi_2(x)}\right] + \mathbb{E}\left[\overline{F}_{X_1}\left(\frac{x}{X_2}\right)1_{X_2>\varphi_2(x)}\right].$$

Since $0 \leqslant \overline{F}_{X_1}(x) \leqslant 1$, we have that $0 \leqslant \mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{X_2>\varphi_2(x)}\Big] \leqslant \mathbb{P}(X_2 > \varphi_2(x)) = \overline{F}_{X_2}(\varphi_2(x))$. Thus, $\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{X_2>\varphi_2(x)}\Big] = O\big(x^{-\theta}\big)$.

Since $\overline{F}_{X_1}(x) = O\big(x^{-\theta}\big)$ and $\lim_{x\to\infty} \frac{x}{X_2} \geqslant \lim_{x\to\infty} \frac{x}{\varphi_2(x)} = \infty$, we have, $\exists x_0 > 0$, $\forall x > x_0$, $\exists C_1 > 0$, $\exists\theta > 0$, $\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{0<X_2\leqslant\varphi_2(x)}\Big] \leqslant \mathbb{E}\Big[C_1\Big(\frac{x}{X_2}\Big)^{-\theta}\Big] = C_1\mathbb{E}\big[X_2^{\theta}\big]x^{-\theta}$. Thus, $\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{0<X_2\leqslant\varphi_2(x)}\Big] = O\big(x^{-\theta}\big)$.

Considering the $O(\cdot)$ polynomial [60], i.e., if $f_1(x) = O(g(x))$ and $f_2(x) = O(g(x))$ then $f_1(x) + f_2(x) = O(g(x))$, we have $\overline{F}_{X_1X_2}(x) =$

$\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)\Big] = O\big(x^{-\theta}\big).$ □

Remark 28. *Since* $\lim_{x\to\infty} \frac{\varphi_2(x)}{x} = 0$*, we have,* $\exists x_0 > 0$*,* $\forall x > x_0$*,* $x \geqslant \varphi_2(x)$ *and* $\overline{F}_{X_2}(x) \leqslant \overline{F}_{X_2}(\varphi_2(x))$*. Thus, if* $\overline{F}_{X_2}(\varphi_2(x)) = O\big(x^{-\theta}\big)$*, where* $\lim_{x\to\infty} \varphi_2(x) = \infty$*, then* $\overline{F}_{X_2}(x) = O\big(x^{-\theta}\big)$*.*

Remark 29. *A related result concerning the sharp approximation,* $\overline{F}(x) \sim Cx^{-\theta}$*, where* $\theta > 0$ *and* $C > 0$ *is a constant, is available in [125]. Another reference is [68].*

Remark 30. *For the tail behavior of the capacity, it is sufficient to study the* $f(x) = O(g(x))$ *approximation rather than the sharper* $f(x) \sim g(x)$ *approximation, because a light-tailed distribution of capacity can bear as heavy as fat-tailed distributions of the random power and the random fade in the logarithm function of the capacity.*

We present the results of the tail upper bounds of the product and sum of random variables.

Theorem 15. *Consider the independent random variables* $X_i : \Omega \to \mathbb{R}_{\geqslant 0}$*,* $i \in \{1, \ldots, N\}$*.*

1. Suppose, $\exists \varphi_j \in \mathfrak{F}$*,* $\overline{F}_{X_j}(\varphi_j(x)) = O\big(\overline{F}_{X_1}(x)\big)$*, and* $\overline{F}_{X_1}(x/\varphi_j(x)) = O\big(\overline{F}_{X_1}(x)\big)$*,* $\forall 2 \leqslant j \leqslant N$*. Then, we have*

$$\overline{F}_{\prod_{i=1}^N X_i}(x) = O\big(\overline{F}_{X_1}(x)\big). \tag{2.107}$$

2. Suppose, $\exists \varphi_j \in \mathfrak{F}$*,* $\overline{F}_{X_j}(\varphi_j(x)) = O\big(\overline{F}_{X_1}(x)\big)$*, and* $\overline{F}_{X_1}(x - \varphi_j(x)) = O\big(\overline{F}_{X_1}(x)\big)$*,* $\forall 2 \leqslant j \leqslant N$*. Then, we have*

$$\overline{F}_{\sum_{i=1}^N X_i}(x) = O\big(\overline{F}_{X_1}(x)\big). \tag{2.108}$$

Proof. We prove the case of $N = 2$ and the proof of the case $N > 2$ follows by the iteration of the same procedure and by the fact that [60], if $f(x) = O(g(x))$ and $g(x) = O(h(x))$ then $f(x) = O(h(x))$.

Considering the independence between X_1 and X_2, for $x > 0$, we have $\overline{F}_{X_1X_2}(x) = \mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)\Big]$, which is reformulated as

$$\mathbb{E}\left[\overline{F}_{X_1}\left(\frac{x}{X_2}\right)\right] = \mathbb{E}\left[\overline{F}_{X_1}\left(\frac{x}{X_2}\right)1_{0<X_2\leqslant\varphi_2(x)}\right] + \mathbb{E}\left[\overline{F}_{X_1}\left(\frac{x}{X_2}\right)1_{X_2>\varphi_2(x)}\right].$$

Since $0 \leqslant \overline{F}_{X_1}(x) \leqslant 1$, we have that $0 \leqslant \mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{X_2>\varphi_2(x)}\Big] \leqslant \mathbb{P}(X_2 > \varphi_2(x)) = \overline{F}_{X_2}(\varphi_2(x))$. Thus, $\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{X_2>\varphi_2(x)}\Big] = O\big(\overline{F}_{X_1}(x)\big)$.

Since $\overline{F}_{X_1}(x)$ is nonincreasing, we have $\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{0<X_2\leqslant\varphi_2(x)}\Big] \leqslant \overline{F}_{X_1}\Big(\frac{x}{\varphi_2(x)}\Big)$. Thus, $\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{0<X_2\leqslant\varphi_2(x)}\Big] = O\big(\overline{F}_{X_1}(x)\big)$.

Considering the $O(\cdot)$ polynomial [60], i.e., if $f_1(x) = O(g(x))$ and $f_2(x) = O(g(x))$ then $f_1(x) + f_2(x) = O(g(x))$, we have $\overline{F}_{X_1X_2}(x) = \mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)\Big] = O\big(x^{-\theta}\big)$.

The proof of the result for the sum of random variables follows analogically. $\square$

Remark 31. *Since* $\lim_{x\to\infty}\frac{\varphi_2(x)}{x} = 0$*, we have,* $\exists x_0 > 0$*,* $\forall x > x_0$*,* $x \geqslant \varphi_2(x)$ *and* $\overline{F}_{X_2}(x) \leqslant \overline{F}_{X_2}(\varphi_2(x))$*. Thus, if* $\overline{F}_{X_2}(\varphi_2(x)) = O\big(\overline{F}_{X_1}(x)\big)$*, where* $\lim_{x\to\infty}\varphi_2(x) = \infty$*, then* $\overline{F}_{X_2}(x) = O\big(\overline{F}_{X_1}(x)\big)$*.*

We present a further result, which indicates that the tail behavior of the random variables product can not be effectively transformed from heavy to light, by the product or sum with other random variables, when there is a slowly varying distribution $F \in \mathcal{R}_0$ in the product or a regularly varying distribution $F \in \mathcal{R}$ in the sum.

Theorem 16. *Consider the independent random variables* $X_i : \Omega \to \mathbb{R}_{\geqslant 0}$*,* $i \in \{1, \ldots, N\}$*.*

1. Suppose, $\exists\varphi_j \in \mathfrak{F}$*,* $\overline{F}_{X_j}(\varphi_j(x)) = o\big(\overline{F}_{X_1}(x)\big)$*,* $F_1 \in \mathcal{R}_0$*, i.e.,* $\overline{F}_{X_1}(tx) \sim \overline{F}_{X_1}(x)$*,* $\forall t > 0$*,* $\forall 2 \leqslant j \leqslant N$*. Then, we have*

$$\overline{F}_{\prod_{i=1}^N X_i}(x) \sim \overline{F}_{X_1}(x). \tag{2.109}$$

2. Suppose, $\exists \varphi_j \in \mathfrak{F}$, $\overline{F}_{X_j}(\varphi_j(x)) = o\big(\overline{F}_{X_1}(x)\big)$, $F_{X_1} \in \mathcal{L}$, *i.e.,* $\overline{F}_{X_1}(x-t) \sim \overline{F}_{X_1}(x)$, $\forall t > 0$, $\forall 2 \leqslant j \leqslant N$. *Then, we have*

$$\overline{F}_{\sum_{i=1}^{N} X_i}(x) \sim \overline{F}_{X_1}(x). \tag{2.110}$$

Proof. We prove the case of $N = 2$ and the proof of the case $N > 2$ follows by the iteration of the same procedure.

Considering the independence between X_1 and X_2, for $x > 0$, we have $\overline{F}_{X_1X_2}(x) = \mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)\Big]$, which is reformulated as

$$\mathbb{E}\left[\overline{F}_{X_1}\left(\frac{x}{X_2}\right)\right] = \mathbb{E}\left[\overline{F}_{X_1}\left(\frac{x}{X_2}\right)1_{0<X_2\leqslant\varphi_2(x)}\right] + \mathbb{E}\left[\overline{F}_{X_1}\left(\frac{x}{X_2}\right)1_{X_2>\varphi_2(x)}\right].$$

Since $0 \leqslant \overline{F}_{X_1}(x) \leqslant 1$, we have that $0 \leqslant \mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{X_2>\varphi_2(x)}\Big] \leqslant \mathbb{P}(X_2 > \varphi_2(x)) = \overline{F}_{X_2}(\varphi_2(x))$. Thus, $\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{X_2>\varphi_2(x)}\Big] = o\big(\overline{F}_{X_1}(x)\big)$.

In addition, we have $\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{0<X_2\leqslant\varphi_2(x)}\Big] \sim \mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)\Big]$, which follows the Lebesgue dominated convergence theorem. Since $\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)\Big] = \int \overline{F}_{X_1}\Big(\frac{x}{x_2}\Big)dF_{X_2}(x_2) \sim \int \overline{F}_{X_1}(x)dF_{X_2}(x_2) = \overline{F}_{X_1}(x)$, which follows that $\overline{F}_{X_1}$ is slowly varying, i.e., $\overline{F}_{X_1}(tx) \sim \overline{F}_{X_1}(x)$, $\forall t > 0$, thus, we obtain $\mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)1_{0<X_2\leqslant\varphi_2(x)}\Big] \sim \overline{F}_{X_1}(x)$.

Considering the asymptotics polynomial [60], i.e., if $f_1(x) = o(g(x))$ and $f_2(x) \sim g(x)$ then $f_1(x) + f_2(x) \sim g(x)$, we obtain that $\overline{F}_{X_1X_2}(x) = \mathbb{E}\Big[\overline{F}_{X_1}\Big(\frac{x}{X_2}\Big)\Big] \sim \overline{F}_{X_1}(x)$.

The proof of the result for the sum of random variables follows analogically. □

Remark 32. *According to Lemma 3, the situation for the product of random variables appears for slowly varying distributions* $F_1 \in \mathcal{R}_0$, *and* $F_j(x) = \Theta\big(x^{-\theta_j}\big)$ *or* $F_j(x) = \Theta\big(e^{-\theta_j x}\big)$, $\theta_j > 0$, $j \in \{2, \ldots, N\}$; *and the situation for the sum of random variables appears for regularly varying dis-*

tributions $F_1 \in \mathcal{R}_{\geqslant 0}$*, and* $F_j(x) = \Theta\big(x^{-\theta_j}\big)$ *or* $F_j(x) = \Theta\big(e^{-\theta_j x}\big)$, $\theta_j > 0$, $j \in \{2, \ldots, N\}$.

Remark 33. *It is interesting to study the possibility of transforming the tail heaviness of a random variable from heavy to light through some functions with other random variables, e.g., for some special cases.*

Remark 34. *The asymptotic behavior of the right tail of the sum of the random variables can be insensible to both positive and negative dependence [2][133][77][10][52], while the asymptotic behavior of the left tail can be connected with the dependence structures [56][134]. In addition, there are scenarios, where the right tail of the sum distribution is sensitive to the dependence structures [2].*

Remark 35. *The tail behavior of the product distribution is more complicated. For example, it is shown that the product distribution of two independent random variables with exponential distributions is subexponential [132][88]. Particularly, the dependence among the random variables are crucial for the tail behavior of the product distribution [69][58][22], e.g., the dependence can either decrease or increase the product distribution tail heaviness compared to the independence scenario [69][144]. In addition, the tail of the product distribution with dependence can be asymptotically bounded above and below by the tail of a dominating random variable [145][22] or can be asymptotically bounded above and below by the tail with assumption of independence [69][144].*

Remark 36. *It is interesting to investigate the extreme influence of the dependence among the random parameters in the wireless channel capacity on the tail behavior of the marginal distribution of the capacity, e.g., whether or not the dependence between two light-tailed or heavy-tailed random variables can cause a super-heavy tail of the product or sum distribution. For example, considering the comonotonic random variables with identical distributions [37],* $X_i \sim X$, $1 \leqslant i \leqslant N$*, we have* $\overline{F}_{\sum\limits_{1\leqslant i\leqslant N} X_i}(x) = \overline{F}_X(x/N)$

and $\overline{F}_{\prod\limits_{1\leqslant i\leqslant N} X_i}(x) = \overline{F}_X\big(x^{1/N}\big)$*, compared to the distribution* $\overline{F}_X(x)$*, the sum distribution* $\overline{F}_X(x/N)$ *is scale invariant for Pareto Type I distribution and asymptotically scale invariant for regular varying distributions, and the product distribution* $\overline{F}_X\big(x^{1/N}\big)$ *has a smaller tail index for Pareto Type I distribution.*

Remark 37. *The tail asymptotic is investigated in [125] for the product and sum of random variables in terms of the asymptotic equality* $f(x) \sim g(x)$*. In this work, we extend the analysis to specific heavy-tailed and light-tailed distribution classes, e.g., the long-tail distribution, the regular varying distribution, and the light-tailed distribution. Specifically, we find that the slowly varying distribution can dominate the tail behavior for the sum and product distribution. Moreover, we extend the analysis beyond the asymptotic equality to more asymptotic notations, e.g.,* $f(x) = O(g(x))$*,* $f(x) = \Theta(g(x))$*,* $f(x) = \omega(g(x))$*, and* $f(x) = o(g(x))$*. Since the capacity is a logarithm transform of the product of the power and the fading random variable, the less strict asymptotic bound provides more flexibility than the asymptotic equality, i.e., it has less restriction and can capture more distribution scenarios, most importantly, it is sufficiently enough to investigate the light-tail behavior that is defined by the asymptotic bound* $f(x) = O(g(x))$*. Another related work is [132], which provides conditions for the product of a light-tailed random variable and a heavy-tailed random variable to be heavy-tailed. In contrast to the result with asymptotic precision up to some distribution classes in [132], we show results of the exact tail domination with respect to a certain distribution function in this work.*

Chapter 3

The Facts of Dependence

There are two kinds of stochastic processes in the stochastic systems, the uncontrollable parameter processes and controllable parameter processes, which correspond to two kinds of dependence, the uncontrollable dependence and controllable dependence, which provide a degree of freedom to trade off for the system performance benefits. For example, the dependence that exists in the fading process of the wireless channels is uncontrollable, and the dependence that appears in the power process is controllable.

In this chapter, we consider the performance analysis of the queueing system, e.g., the wireless channel, we study the influence of dependence in the arrival process and service process on the performance measures, i.e., the backlog and delay, and we develop the dependence manipulation techniques to transform the dependence structure of the stochastic processes. The measurability of the dependence with respect to the performance measures, the duality of the dependence in the arrival process and service process, and the transformability of the dependence structure constitute the theory of dependence control.

3.1 Foundation

3.1.1 Queueing Behavior

We consider a queueing system with arrival process $a(t)$, service process $c(t)$, and the temporal increment in the system

$$X(t) = a(t) - c(t). \tag{3.1}$$

The queueing behavior is expressed through the backlog $B(t)$ in the system, which is a reflected process of the temporal increment $X(t)$ [6], i.e.,

$$B(t+1) = [B(t) + X(t)]^{+}. \tag{3.2}$$

Assuming $B(0) = 0$, the backlog function is then expressed as

$$B(t) = \sup_{0 \leqslant s \leqslant t} (A(s,t) - S(s,t)), \tag{3.3}$$

where $A(s,t) = \sum_{i=s+1}^{t} a(i)$ and $S(s,t) = \sum_{i=s+1}^{t} c(i)$ are respectively the cumulative arrival process and the cumulative service process. Specifically, we denote $A(0,t) \equiv A(t)$ and $S(0,t) \equiv S(t)$. For a lossless system, the output $A^*(t)$ is the difference between the arrival input and backlog, i.e.,

$$A^*(t) = A(t) - B(t), \tag{3.4}$$

and the delay is defined via the input-output relationship, i.e.,

$$D(t) = \inf \{d \geqslant 0 : A(t-d) \leqslant A^*(t)\}, \tag{3.5}$$

which is the virtual delay that a hypothetical arrival has experienced on departure.

Consider the infinite time horizon and the time reversal assumption that is introduced in Appendix A.3. The delay tail probability is expressed as

$$\mathbb{P}(D > d) = \mathbb{P}\left\{\sup_{t\geqslant d}\{A(d,t) - S(0,t)\} > 0\right\}, \tag{3.6}$$

which follows that $\mathbb{P}(D(t) > d) = \mathbb{P}\{A(t-d) > A^*(t)\} = \mathbb{P}\{A(t-d) > \inf_{0\leqslant s\leqslant t}\{A(0,s) + S(s,t)\}\}$ and the last step follows the time reversal with respect to time $\tau = \infty$. The backlog tail probability is expressed as

$$\mathbb{P}(B > b) = \mathbb{P}\left\{\sup_{t\geqslant 0}(A(0,t) - S(0,t)) > b\right\}, \tag{3.7}$$

which follows that $\mathbb{P}(B(t) > b) = \mathbb{P}\{\sup_{0\leqslant s\leqslant t}(A(s,t) - S(s,t)) > b\}$ and the last step follows the time reversal with respect to time $\tau = \infty$.

3.1.2 More Assumptions

In addition to the time reversal assumption in the tail probability expressions of delay and backlog, we specify the cumulant generating functions of the cumulative arrival process $A(t)$, the cumulative service process $S(t)$, and the increment process of the queue $A(t) - S(t)$.

The assumption for the queue increment process is as follows [53], without assumption on the dependence between the arrival process and service process.

Assumption 1. *Denote $\mathfrak{S}(t) = A(t) - S(t)$ and $X(t) = a(t) - c(t)$. Assume that there exist $\gamma, \epsilon > 0$ such that*

1. $\kappa_t(\theta) = \log \mathbb{E}e^{\theta\mathfrak{S}(t)}$ *is well-defined and finite for* $\gamma - \epsilon < \theta < \gamma + \epsilon$*;*
2. $\limsup_{t\to\infty} \mathbb{E}e^{\theta X(t)} < \infty$ *for* $-\epsilon < \theta < \epsilon$*;*
3. $\kappa(\theta) = \lim_{t\to\infty} \frac{1}{t}\kappa_t(\theta)$ *exists and is finite for* $\gamma - \epsilon < \theta < \gamma + \epsilon$*;*
4. $\kappa(\gamma) = 0$ *and* κ *is differentiable at* γ *with* $0 < \dot{\kappa}(\gamma) < \infty$*.*

Assuming independence between the arrival and service process, we get an alternative expression [53].

Assumption 2. *Assume independence between the sequences of $a(t)$ and $c(t)$, $t \geqslant 0$. Let $\gamma, \epsilon > 0$ be as in Assumption 1 such that*

1. $\kappa_t^A(\theta) = \log \mathbb{E}e^{\theta A(t)}$ *is well-defined and finite for* $\gamma - \epsilon < \theta < \gamma + \epsilon$*;*
2. $\limsup_{t\to\infty} \mathbb{E}e^{\theta a(t)} < \infty$ *for* $-\epsilon < \theta < \epsilon$*;*
3. $\kappa^A(\theta) = \lim_{t\to\infty} \frac{1}{t}\kappa_t^A(\theta)$ *exists, is differentiable at* γ*, and is finite for* $\gamma - \epsilon < \theta < \gamma + \epsilon$*;*
4. $\kappa_t^{-S}(\theta) = \log \mathbb{E}e^{-\theta S(t)}$ *is well-defined and finite for* $\gamma - \epsilon < \theta < \gamma + \epsilon$*;*
5. $\limsup_{t\to\infty} \mathbb{E}e^{-\theta c(t)} < \infty$ *for* $-\epsilon < \theta < \epsilon$*;*
6. $\kappa^{-S}(\theta) = \lim_{t\to\infty} \frac{1}{t}\kappa_t^{-S}(\theta)$ *exists, is differentiable at* γ*, and is finite for* $\gamma - \epsilon < \theta < \gamma + \epsilon$*;*
7. $\kappa(\theta) = \kappa^A(\theta) + \kappa^{-S}(\theta)$.

The Assumption 1 and Assumption 2 apply to the scenarios [7], where there are weak forms of dependence, e.g., Markov dependence, and the average of the cumulant generating function exists and converges, e.g., light-tailed process.

Remark 38. *The asymptotic independence [52] is a weak form dependence, which is neither too positively nor too negatively dependent, and the strong forms of dependence are association and regression [83][39]. Another weak form of dependence is shown in [77]. It is interesting to provide a review of the dependence concepts in the literature.*

3.1.3 Change of Measure

We introduce a change of measure for $a_1, \ldots, a_n, c_1, \ldots, c_n$,

$$\begin{aligned}&\widetilde{F}_n\left(da_1, \ldots, da_n, dc_1, \ldots, dc_n\right)\\&= e^{\gamma s_n - \kappa_n(\gamma)} F_n\left(da_1, \ldots, da_n, dc_1, \ldots, dc_n\right),\end{aligned} \tag{3.8}$$

where F_n is the distribution of $a_1, \ldots, a_n, c_1, \ldots, c_n$ and $s_n = a_1 - c_1 + \ldots + a_n - c_n$. Assuming independence between $a_1, \ldots, a_n$ and $c_1, \ldots, c_n$, the distributions of $a_1, \ldots, a_n$ and $c_1, \ldots, c_n$ in the new probability measure are given by

$$\widetilde{F}_n^A\left(da_1, \ldots, da_n\right) = e^{\gamma s_n^A - \kappa_n^A(\gamma)} F_n\left(da_1, \ldots, da_n, \mathbf{1}\right), \tag{3.9}$$
$$\widetilde{F}_n^S\left(dc_1, \ldots, dc_n\right) = e^{\gamma s_n^{-S} - \kappa_n^{-S}(\gamma)} F_n\left(\mathbf{1}, dc_1, \ldots, dc_n\right), \tag{3.10}$$

where $s_n^A = a_1 + \ldots + a_n$, $s_n^{-S} = -(c_1 + \ldots + c_n)$, and

$$\kappa_n(\gamma) = \kappa_n^A(\gamma) + \kappa_n^{-S}(\gamma). \tag{3.11}$$

In the new probability measure $\widetilde{\mathbb{P}}_n$, we show that, for fixed $d, k \in \mathbb{N}$, $\frac{\mathfrak{S}(d,n-k)}{n} = \frac{A(d,n-k)-S(d,n-k)}{n}$ converges in probability to $\dot{\kappa}(\gamma) \equiv \frac{\partial}{\partial\gamma}\kappa(\gamma)$.

Theorem 17. *Let $n, d, k \in \mathbb{N}$ and $d, k < \infty$, and $\widetilde{\mu} \equiv \dot{\kappa}(\gamma)$. Then,*

$$\lim_{n\to\infty} \widetilde{\mathbb{P}}_n\left(\left|\frac{\mathfrak{S}(d, n-k)}{n} - \widetilde{\mu}\right| > \eta\right) = 0, \ \forall\eta > 0. \tag{3.12}$$

An equivalent expression of the above theorem is the following theorem. Particularly, the probability z^n is set to facilitate the proof of Theorem 19.

Theorem 18. *Let $n, d, k \in \mathbb{N}$ and $d, k < \infty$, and $\widetilde{\mu} \equiv \dot{\kappa}(\gamma)$. For each $\eta > 0$, there exist $z \equiv z(\eta) \in (0, 1)$ and n_0 such that*

$$\widetilde{\mathbb{P}}_n\left(\left|\frac{\mathfrak{S}(d, n-k)}{n} - \widetilde{\mu}\right| > \eta\right) \leqslant z^n, \ \textit{for } n \geqslant n_0. \tag{3.13}$$

Proof. Let $0 < \theta < \epsilon$, where ϵ is as in Assumption 1. Note $\mathbb{E}\left[e^{\theta(\mathfrak{S}(n)-\mathfrak{S}(n-k))}\right] < \infty$ for all $|\theta| < \delta$ for some $\delta > 0$ by Assumption 1 (2).

According to Chernoff bound,

$$\widetilde{\mathbb{P}}_n\left(\frac{\mathfrak{S}(n-k)-\mathfrak{S}(d)}{n} - \widetilde{\mu} > \eta\right) \tag{3.14}$$

$$\leqslant e^{-\theta n(\widetilde{\mu}+\eta)}\widetilde{\mathbb{E}}_n\left[e^{\theta(\mathfrak{S}(n-k)-\mathfrak{S}(d))}\right] \tag{3.15}$$

$$= e^{-\theta n(\widetilde{\mu}+\eta)}\mathbb{E}_n\left[e^{\theta(\mathfrak{S}(n-k)-\mathfrak{S}(d))}\cdot e^{\gamma\mathfrak{S}(n)-\kappa_n(\gamma)}\right] \tag{3.16}$$

$$= e^{-\theta n(\widetilde{\mu}+\eta)-\kappa_n(\gamma)}\mathbb{E}_n\left[e^{(\theta+\gamma)\mathfrak{S}(n)-\theta\mathfrak{S}(d)-\theta\mathfrak{S}(n-k,n)}\right] \tag{3.17}$$

$$\leqslant e^{-\theta n(\widetilde{\mu}+\eta)-\kappa_n(\gamma)}\left[\left[\mathbb{E}_n e^{\hat{p}p(\theta+\gamma)\mathfrak{S}(n)}\right]^{1/\hat{p}}\left[\mathbb{E}_n e^{-\hat{q}p\theta\mathfrak{S}(d)}\right]^{1/\hat{q}}\right]^{1/p}$$

$$\cdot\left[\mathbb{E}_n e^{-q\theta\mathfrak{S}(n-k,n)}\right]^{1/q} \tag{3.18}$$

$$= e^{-\theta n(\widetilde{\mu}+\eta)-\kappa_n(\gamma)+\kappa_n(\hat{p}p(\theta+\gamma))/(\hat{p}p)}\left[\mathbb{E}_n e^{-\hat{q}p\theta\mathfrak{S}(d)}\right]^{1/(\hat{q}p)}$$

$$\cdot\left[\mathbb{E}_n e^{-q\theta\mathfrak{S}(n-k,n)}\right]^{1/q}, \tag{3.19}$$

where we used Hölder's inequality twice, for positive p and q with $p^{-1} + q^{-1} = 1$, and $\hat{p}$ and $\hat{q}$ with $\hat{p}^{-1} + \hat{q}^{-1} = 1$, and we choose p and $\hat{p}$ close enough to 1 and θ close enough to 0 that $|\hat{p}p(\theta+\gamma)-\gamma| < \epsilon$ and $|-\hat{q}p\theta-\gamma| < \epsilon$. Particularly, for $k = 0$ or $d = 0$, the proof needs to use Hölder's inequality only once; for $k = 0$ and $d = 0$, the proof needs no Hölder's inequality.

By Assumption 1 (1), $\mathbb{E}_n[e^{-\hat{q}p\theta\mathfrak{S}(d)}] < \infty$, and by Assumption 1 (2), $\mathbb{E}_n\left[e^{-q\theta(\mathfrak{S}(n)-\mathfrak{S}(n-k))}\right]^{1/q} < \infty$ for large n, we get

$$\limsup_{n\to\infty}\frac{1}{n}\log\widetilde{\mathbb{P}}_n\left(\frac{\mathfrak{S}(n-k)-\mathfrak{S}(d)}{n} - \widetilde{\mu} > \eta\right)$$

$$\leqslant \kappa(\hat{p}p(\theta+\gamma))/(\hat{p}p) - \kappa(\gamma) - \theta\left(\widetilde{\mu}+\eta\right), \tag{3.20}$$

by Taylor expansion, it is easy to see that the right hand side can be chosen strictly negative by setting p and $\hat{p}$ close enough to 1 and θ close enough to 0. This establishes $\widetilde{\mathbb{P}}_n(\mathfrak{S}(d, n-k)/n - \widetilde{\mu} > \eta) \leqslant z^n$, correspondingly, $\widetilde{\mathbb{P}}_n(\mathfrak{S}(d, n-k)/n - \widetilde{\mu} < -\eta) \leqslant z^n$ follows by symmetry. $\square$

To facilitate the proof of Theorem 19, we have the following Corollary.

Corollary 4. *Let $n, d, k \in \mathbb{N}$ and $d, k < \infty$, and $\widetilde{\mu} \equiv \dot{\kappa}(\gamma)$. Let $x \in \mathbb{R}$ and $x < \infty$. For each $\eta > 0$, there exist $z \equiv z(\eta) \in (0,1)$ and n_0 such that*

$$\widetilde{\mathbb{P}}_n\left(\left|\frac{\mathfrak{S}(n-k) - A(d) + x}{n} - \widetilde{\mu}\right| > \eta\right) \leqslant z^n, \text{ for } n \geqslant n_0. \tag{3.21}$$

Proof. Theorem 18 shows that the convergence is insensible to the head and tail of the sequence. Replacing $\mathfrak{S}(d)$ with $A(d)$ and a finite constant x completes the proof. $\square$

3.2 Dependence Mechanics

3.2.1 The General Rules

As a measure to identify and quantify the dependence influence, we are interested in the asymptotic tail behavior of the delay and backlog in the queue and we show that this measure has a monotonic relationship with the dependence in the arrival process and the service process. Particularly, this monotonic property is useful for understanding the trend of the dependence influence, when the explicit results are not tractable in complex dependence scenarios.

We derive the exact results of the asymptotic decay rate of delay and backlog, and the logarithmic approximation implies that the asymptotic behavior of the delay and backlog tail probability is exponential for weak forms of dependence and light-tailed processes, which is characterized by Assumption 1 and Assumption 2.

Theorem 19. *Under the conditions in Assumption 2, the asymptotic decay rates of delay and backlog are respectively*

$$\lim_{d\to\infty}\frac{1}{d}\log\mathbb{P}(D>d) = -\kappa^A(\gamma), \tag{3.22}$$
$$\lim_{b\to\infty}\frac{1}{b}\log\mathbb{P}(B>b) = -\gamma, \tag{3.23}$$

where $\gamma=\theta>0$ *is the root to the stability equation* $\kappa^A(\theta)+\kappa^{-S}(\theta)=0$.

Proof. We only provide proof for the delay result, since the backlog result is a trivial reduction of the delay proof. The proof is inspired by [53, 7], by defining a new change of measure, and by noting the following result, for large enough n,

$$\widetilde{\mathbb{P}}_n\left(\left|\frac{\mathfrak{S}(n-k)-A(d)+d}{n}-\widetilde{\mu}\right|>\eta\right)\leqslant z^n. \tag{3.24}$$

We first show that $\liminf\limits_{d\to\infty}\frac{1}{d}\log\mathbb{P}(D>d)\geqslant-\kappa^A(\gamma)$. Given $\eta>0$ and let $m\equiv m(\eta)=\lfloor d(1+\eta)/\widetilde{\mu}\rfloor+1$. Then

$$\begin{aligned}
&\mathbb{P}(D>d)\geqslant\mathbb{P}(\mathfrak{S}(m)>A(d)) && (3.25)\\
&=\widetilde{\mathbb{E}}_m\left[e^{-\gamma\mathfrak{S}(m)+\kappa_m(\gamma)};\mathfrak{S}(m)-A(d)+d>d\right] && (3.26)\\
&\geqslant\widetilde{\mathbb{E}}_m\left[e^{-\gamma\mathfrak{S}(m)+\kappa_m(\gamma)};\frac{\mathfrak{S}(m)-A(d)+d}{m}-\widetilde{\mu}>-\frac{\widetilde{\mu}\eta}{1+\eta}\right] && (3.27)\\
&\geqslant\widetilde{\mathbb{E}}_m\left[e^{-\gamma\mathfrak{S}(m)+\kappa_m(\gamma)};\left|\frac{\mathfrak{S}(m)-A(d)+d}{m}-\widetilde{\mu}\right|<\frac{\widetilde{\mu}\eta}{1+\eta}\right] && (3.28)\\
&\geqslant\widetilde{\mathbb{E}}_m\left[e^{-\gamma\left(\widetilde{\mu}\frac{1+2\eta}{1+\eta}m+A(d)-d\right)+\kappa_m(\gamma)}\right]\\
&\quad\cdot\widetilde{\mathbb{P}}_m\left(\left|\frac{\mathfrak{S}(m)-A(d)+d}{m}-\widetilde{\mu}\right|<\frac{\widetilde{\mu}\eta}{1+\eta}\right) && (3.29)\\
&=e^{-\kappa_d^A\gamma-\gamma\widetilde{\mu}\frac{1+2\eta}{1+\eta}m+\gamma d+\kappa_m(\gamma)}\\
&\quad\cdot\widetilde{\mathbb{P}}_m\left(\left|\frac{\mathfrak{S}(m)-A(d)+d}{m}-\widetilde{\mu}\right|<\frac{\widetilde{\mu}\eta}{1+\eta}\right), && (3.30)
\end{aligned}$$

where $\widetilde{\mathbb{P}}_m(\cdot)$ goes to 1 according to Corollary 4. Since $\kappa_m(\gamma)/d \to 0$ and $m/d \to (1+\eta)/\widetilde{\mu}$, we get

$$\liminf_{d\to\infty} \frac{1}{d} \log \mathbb{P}(D > d) \geqslant -\kappa^A(\gamma) - 2\eta. \tag{3.31}$$

Letting $\eta \downarrow 0$ yields $\liminf\limits_{d\to\infty} \frac{1}{d} \log \mathbb{P}(D > d) \geqslant -\kappa^A(\gamma)$.

We then show that $\limsup\limits_{d\to\infty} \frac{1}{d} \log \mathbb{P}(D > d) \leqslant -\kappa^A(\gamma)$. Let $\tau(d) = \inf\{n : \mathfrak{S}(n) > A(d)\}$, then $\tau(d) > d$ and $\mathbb{P}(D > d) = \mathbb{P}(\tau(d) < \infty)$, i.e.,

$$\mathbb{P}(D > d) = \sum_{n=d+1}^{\infty} \mathbb{P}(\tau(d) = n) = I_1 + I_2 + I_3 + I_4, \tag{3.32}$$

where $I_1 = \sum_{n=d+1}^{n(\delta)} \mathbb{P}(\tau(d) = n)$, $I_2 = \sum_{n=n(\delta)+1}^{\lfloor d(1-\delta)/\widetilde{\mu}\rfloor} \mathbb{P}(\tau(d) = n)$, $I_3 = \sum_{n=\lfloor d(1-\delta)/\widetilde{\mu}\rfloor+1}^{\lfloor d(1+\delta)/\widetilde{\mu}\rfloor} \mathbb{P}(\tau(d) = n)$, $I_4 = \sum_{n=\lfloor d(1+\delta)/\widetilde{\mu}\rfloor+1}^{\infty} \mathbb{P}(\tau(d) = n)$, $n(\delta)$ is chosen such that $\kappa_n(\gamma)/n < \min\{\delta, (-\log z)/2\}$ and

$$\widetilde{\mathbb{P}}_n\left(\left|\frac{\mathfrak{S}(n-k) - A(d) + d}{n} - \widetilde{\mu}\right| > \frac{\delta\widetilde{\mu}}{1+\delta}\right) \leqslant z^n, \text{ for } k \leqslant 1, \tag{3.33}$$

for some $z < 1$ and all $n > n(\delta)$. This is possible by Assumption 1 (3) and Assumption 1 (4) and Corollary 4.

Note

$$\begin{aligned}
\mathbb{P}(\tau(d) = n) &\leqslant \mathbb{P}(\mathfrak{S}(n) > A(d)) && (3.34)\\
&= \widetilde{\mathbb{E}}_n\left[e^{-\gamma\mathfrak{S}(n)+\kappa_n(\gamma)}; \mathfrak{S}(n) > A(d)\right] && (3.35)\\
&\leqslant e^{-\kappa_d^A(\gamma)} \cdot e^{\kappa_n(\gamma)} \cdot \widetilde{\mathbb{P}}_n(\mathfrak{S}(n) > A(d)), && (3.36)
\end{aligned}$$

so that

$$I_1 \leqslant e^{-\kappa_d^A(\gamma)} \sum_{n=d+1}^{n(\delta)} e^{\kappa_n(\gamma)}, \tag{3.37}$$

$$\begin{aligned}
I_2 &\leqslant e^{-\kappa_d^A(\gamma)} \sum_{n=n(\delta)+1}^{\lfloor d(1-\delta)/\widetilde{\mu}\rfloor} e^{\kappa_n(\gamma)} \widetilde{\mathbb{P}}_n(\mathfrak{S}(n) > A(d)) && (3.38)\\
&\leqslant e^{-\kappa_d^A(\gamma)} \sum_{n=n(\delta)+1}^{\lfloor d(1-\delta)/\widetilde{\mu}\rfloor} e^{-n\log z/2} \cdot \widetilde{\mathbb{P}}_n\left(\left|\frac{\mathfrak{S}(n) - A(d) + d}{n} - \widetilde{\mu}\right| > \frac{\delta\widetilde{\mu}}{1+\delta}\right)\\
&\leqslant e^{-\kappa_d^A(\gamma)} \sum_{n=n(\delta)+1}^{\lfloor d(1-\delta)/\widetilde{\mu}\rfloor} \frac{1}{z^{n/2}} z^n && (3.39)\\
&\leqslant e^{-\kappa_d^A(\gamma)} \sum_{n=0}^{\infty} z^{n/2} && (3.40)\\
&= e^{-\kappa_d^A(\gamma)} \frac{1}{1 - z^{1/2}}, && (3.41)\\
I_3 &\leqslant e^{-\kappa_d^A(\gamma)} \sum_{n=\lfloor d(1-\delta)/\widetilde{\mu}\rfloor+1}^{\lfloor d(1+\delta)/\widetilde{\mu}\rfloor} e^{\kappa_n(\gamma)} && (3.42)\\
&\leqslant e^{-\kappa_d^A(\gamma)} \sum_{n=\lfloor d(1-\delta)/\widetilde{\mu}\rfloor+1}^{\lfloor d(1+\delta)/\widetilde{\mu}\rfloor} e^{n\delta} && (3.43)\\
&\leqslant e^{-\kappa_d^A(\gamma)} \left(\frac{2\delta d}{\widetilde{\mu}} + 1\right) e^{\delta d(1+\delta)/\widetilde{\mu}}. && (3.44)
\end{aligned}$$

Finally, let $\mathfrak{S}_{n-1}^n(d) \equiv \{\mathfrak{S}(n-1) \leqslant A(d), \mathfrak{S}(n) > A(d)\}$,

$$\begin{aligned}
I_4 &\leqslant \sum_{n=\lfloor d(1+\delta)/\widetilde{\mu}\rfloor+1}^{\infty} \mathbb{P}\left(\mathfrak{S}_{n-1}^n(d)\right) && (3.45)\\
&= \sum_{n=\lfloor d(1+\delta)/\widetilde{\mu}\rfloor+1}^{\infty} \widetilde{\mathbb{E}}_n\left[e^{-\gamma\mathfrak{S}(n)+\kappa_n(\gamma)}; \mathfrak{S}_{n-1}^n(d)\right] && (3.46)\\
&\leqslant e^{-\kappa_d^A(\gamma)} \sum_{n=\lfloor d(1+\delta)/\widetilde{\mu}\rfloor+1}^{\infty} e^{\kappa_n(\gamma)} \cdot \widetilde{\mathbb{P}}_n\left(\left|\frac{\mathfrak{S}(n-1) - A(d) + d}{n} - \widetilde{\mu}\right| > \frac{\delta\widetilde{\mu}}{1+\delta}\right)\\
&\leqslant e^{-\kappa_d^A(\gamma)} \sum_{n=\lfloor d(1+\delta)/\widetilde{\mu}\rfloor+1}^{\infty} \frac{1}{z^{n/2}} z^n && (3.47)\\
&\leqslant e^{-\kappa_d^A(\gamma)} \frac{1}{1 - z^{1/2}}. && (3.48)
\end{aligned}$$

By Assumption 1 (1) and Assumption 2 (1), we get

$$\limsup_{d\to\infty} \frac{1}{d}\log \mathbb{P}(D > d) \leqslant -\kappa^A(\gamma) + \frac{\delta(1+\delta)}{\widetilde{\mu}}. \tag{3.49}$$

Letting $\delta \downarrow 0$ yields $\limsup\limits_{d\to\infty} \frac{1}{d}\log \mathbb{P}(D > d) \leqslant -\kappa^A(\gamma)$. □

Remark 39. *It is interesting to notice that the event* $\mathfrak{S}(t) > A(d)$ *has zero probability for* $t \leqslant d$*, i.e.,* $\mathbb{P}(\tau(d) \leqslant d) = 0$*, where* $\tau(d) = \inf\{t : \mathfrak{S}(t) > A(d)\}$*. It is interesting to notice that the probability* $\mathbb{P}(D > d) = \mathbb{P}(d < \tau(d) < \infty) = \mathbb{P}(\tau(d) < \infty)$*, where the last equality follows that* $\mathbb{P}(\tau(d) < \infty) = \mathbb{P}(0 \leqslant \tau(d) \leqslant d) + \mathbb{P}(d < \tau(d) < \infty)$ *and* $\mathbb{P}(0 \leqslant \tau(d) \leqslant d) = 0$*.*

Remark 40. *The change of measure implies that the original probability measure and the new probability measure are equivalent, i.e., sharing the same set of null events. Specifically, the event* $\mathfrak{S}(t) > A(d)$ *on timeline* $0 \leqslant t \leqslant d$ *is a null event, which holds in both the original probability measure and the new probability measure.*

Remark 41. *In addition to the asymptotic tail behavior of the right tail* $\mathbb{P}(X > x)$*, which is about the large value violations, it is interesting to investigate the asymptotic tail behavior of the left tail* $\mathbb{P}(X \leqslant x)$ *[56][9], which is important to understand the phenomena of small values.*

We consider two queue increment processes with respective arrival process and service process, $\mathfrak{S}(t) = A(t) - S(t)$ and $\widetilde{\mathfrak{S}}(t) = \widetilde{A}(t) - \widetilde{S}(t)$. Let γ, $\kappa^{-A}(\gamma)$, and $\kappa^{-S}(\gamma)$ be the asymptotic decay rates with respect to $\mathfrak{S}(t)$, and let $\widetilde{\gamma}$, $\widetilde{\kappa}^{-A}(\widetilde{\gamma})$, and $\widetilde{\kappa}^{-S}(\widetilde{\gamma})$ be the asymptotic decay rates with respect to $\widetilde{\mathfrak{S}}(t)$. The "positive" queue increment process $\mathfrak{S}(t) = A(t) - S(t)$ indicates the change of measure on the positive parameter axis, $\widetilde{F}_n(da_1, \ldots, da_n, dc_1, \ldots, dc_n) = e^{\gamma\mathfrak{S}_n - \kappa_n^{\mathfrak{S}}(\gamma)} F_n(da_1, \ldots, da_n, dc_1, \ldots, dc_n)$, $\gamma > 0$, with the stability equation $\kappa^A(\gamma) + \kappa^{-S}(\gamma) = 0$, $\gamma > 0$. The "negative" queue increment process $-\mathfrak{S}(t) = S(t) - A(t)$ indicates the change of measure on the negative parameter axis, $\widetilde{F}_n(da_1, \ldots, da_n, dc_1, \ldots, dc_n) = e^{-\gamma\mathfrak{S}_n - \kappa_n^{-\mathfrak{S}}(\gamma)} F_n(da_1, \ldots, da_n, dc_1, \ldots, dc_n)$, $\gamma < 0$, with the stability equation $\kappa^S(\gamma) + \kappa^{-A}(\gamma) = 0$, $\gamma < 0$. We use change of measure to explain the duality principle.

First, we show that the ordering of the "positive" queue increment process or "negative" queue increment process implies the ordering of the absolute decay rate of the dependence measure identities. This result means that when considering the absolute value of the measure identities, the two ordering rules of the measure identities on the positive and negative axes are exactly the same, because of the change of measure respectively on the positive and negative parameter axes.

Theorem 20. *Let $\mathfrak{S}(t) = A(t) - S(t)$ and $\widetilde{\mathfrak{S}}(t) = \widetilde{A}(t) - \widetilde{S}(t)$.*

- *If $\mathfrak{S}(t) \leqslant_{cx} \widetilde{\mathfrak{S}}(t)$, $\forall t \in \mathbb{N}$, then $0 < \widetilde{\gamma} \leqslant \gamma$; furthermore, if $\mathfrak{S}(t)$ and $\widetilde{\mathfrak{S}}(t)$ have identical service process $S(t) = \widetilde{S}(t)$, then $\kappa^{-S}(\gamma) \leqslant \widetilde{\kappa}^{-S}(\widetilde{\gamma}) < 0$.*
- *If $-\mathfrak{S}(t) \leqslant_{cx} -\widetilde{\mathfrak{S}}(t)$, $\forall t \in \mathbb{N}$, then $0 > \widetilde{\gamma} \geqslant \gamma$; furthermore, if $\mathfrak{S}(t)$ and $\widetilde{\mathfrak{S}}(t)$ have identical arrival process $A(t) = \widetilde{A}(t)$, then $\kappa^{-A}(\gamma) \geqslant \widetilde{\kappa}^{-A}(\widetilde{\gamma}) > 0$.*

Proof. We follow and extend the approach in [96]. Since the exponential function is convex, $\mathfrak{S}(t) \leqslant_{cx} \widetilde{\mathfrak{S}}(t)$ implies $\mathbb{E}e^{\theta\mathfrak{S}(t)} \leqslant \mathbb{E}e^{\theta\widetilde{\mathfrak{S}}(t)}$, thus $\kappa_t(\theta) \leqslant \widetilde{\kappa}_t(\theta)$ and subsequently $\kappa(\theta) \leqslant \widetilde{\kappa}(\theta)$, $\forall\theta \in \mathbb{R}$. Since κ is convex with $\kappa(0) = \kappa(\gamma) = 0$, which implies $0 < \kappa(\theta) \leqslant \widetilde{\kappa}(\theta)$, $\forall\theta > \gamma$, therefore, we must have $0 < \widetilde{\gamma} \leqslant \gamma$. Since κ^{-S} is decreasing, we get $\kappa^{-S}(\gamma) \leqslant \widetilde{\kappa}^{-S}(\widetilde{\gamma}) < 0$. The other results follow analogically. □

Remark 42. *The change of measure on the positive parameter axis $\theta > 0$ with respect to $\mathfrak{S}(t)$ and $\widetilde{\mathfrak{S}}(t)$, and the change of measure on the negative parameter axis $\theta < 0$ with respect to $-\mathfrak{S}(t)$ and $-\widetilde{\mathfrak{S}}(t)$, aims to preserve the absolute invariance, $\{\theta\mathfrak{S}(t),\ \theta > 0\} \equiv \{-\theta\mathfrak{S}(t),\ \theta < 0\}$ and $\left\{\theta\widetilde{\mathfrak{S}}(t),\ \theta > 0\right\} \equiv \left\{-\theta\widetilde{\mathfrak{S}}(t),\ \theta < 0\right\}$.*

Remark 43. *The closure property of convex order implies that* $-\mathfrak{S}(t) \leqslant_{cx} -\widetilde{\mathfrak{S}}(t) \iff \mathfrak{S}(t) \leqslant_{cx} \widetilde{\mathfrak{S}}(t)$. *This is an alternative perspective on the dependence duality to the change of measure perspective. In addition, it indicates that the change of measure on the positive parameter axis and on the negative parameter axis are equivalent.*

Remark 44. *The requirement of* $\mathbb{E}e^{\theta\mathfrak{S}(t)} \leqslant \mathbb{E}e^{\theta\widetilde{\mathfrak{S}}(t)}$, $\theta > 0$, *considering the change of measure on the positive parameter axis in the proof, implies that the sufficient condition of convex order* $\mathfrak{S}(t) \leqslant_{cx} \widetilde{\mathfrak{S}}(t)$ *can be replaced by the increasing convex order* $\mathfrak{S}(t) \leqslant_{icx} \widetilde{\mathfrak{S}}(t)$.

Remark 45. *For the change of measure on the negative parameter axis, the requirement of* $\mathbb{E}e^{-\theta\mathfrak{S}(t)} \leqslant \mathbb{E}e^{-\theta\widetilde{\mathfrak{S}}(t)}$, $\theta < 0$ *implies that the sufficient condition of convex order* $-\mathfrak{S}(t) \leqslant_{cx} -\widetilde{\mathfrak{S}}(t)$ *can be replaced by the increasing concave order* $-\mathfrak{S}(t) \geqslant_{icv} -\widetilde{\mathfrak{S}}(t)$.

Remark 46. *The equivalence* $-\mathfrak{S}(t) \geqslant_{icv} -\widetilde{\mathfrak{S}}(t) \iff \mathfrak{S}(t) \leqslant_{icx} \widetilde{\mathfrak{S}}(t)$ *implies the equivalence of the change of measure on the positive parameter axis and on the negative parameter axis.*

Second, we show that the ordering of the arrival process and the service process implies respectively the ordering of the "positive" queue increment process and "negative" queue increment process.

Theorem 21. *Let* $\mathfrak{S}(t) = A(t) - S(t)$ *and* $\widetilde{\mathfrak{S}}(t) = \widetilde{A}(t) - \widetilde{S}(t)$.

- *If* $A(t) \leqslant_{cx} \widetilde{A}(t)$, *and* $\mathfrak{S}(t)$ *and* $\widetilde{\mathfrak{S}}(t)$ *have identical service process* $S(t) = \widetilde{S}(t)$, *then* $\mathfrak{S}(t) \leqslant_{cx} \widetilde{\mathfrak{S}}(t)$.
- *If* $S(t) \leqslant_{cx} \widetilde{S}(t)$, *and* $\mathfrak{S}(t)$ *and* $\widetilde{\mathfrak{S}}(t)$ *have identical arrival process* $A(t) = \widetilde{A}(t)$, *then* $-\mathfrak{S}(t) \leqslant_{cx} -\widetilde{\mathfrak{S}}(t)$.

Proof. Consider a convex function f. Let $X = S(t)$, $Y = \widetilde{S}(t)$, and $Z = A(t)$, $\forall t$. Let $g(z) = \mathbb{E}[f(X - z)]$ and $h(z) = \mathbb{E}[f(Y - z)]$. As the function

$x \mapsto f(x-z)$ is convex for all $z \in \mathbb{R}$, $X \leqslant_{cx} Y$ implies $g(z) \leqslant h(z)$ for all $z \in \mathbb{R}$. Thus, $\mathbb{E}[f(X-Z)] = \mathbb{E}[g(Z)] \leqslant \mathbb{E}[h(Z)] = \mathbb{E}[f(Y-Z)]$, i.e., $S(t) - A(t) \leqslant_{cx} \widetilde{S}(t) - A(t)$. The other result follows analogically. $\square$

Remark 47. *Considering* $\mathfrak{S}(t) = A(t) - S(t)$ *and* $\widetilde{\mathfrak{S}}(t) = \widetilde{A}(t) - \widetilde{S}(t)$. *If* $A(t) \leqslant_{icx} \widetilde{A}(t)$, *and* $\mathfrak{S}(t)$ *and* $\widetilde{\mathfrak{S}}(t)$ *have identical service process* $S(t) = \widetilde{S}(t)$, *then* $\mathfrak{S}(t) \leqslant_{icx} \widetilde{\mathfrak{S}}(t)$. *If* $S(t) \geqslant_{icv} \widetilde{S}(t)$, *and* $\mathfrak{S}(t)$ *and* $\widetilde{\mathfrak{S}}(t)$ *have identical arrival process* $A(t) = \widetilde{A}(t)$, *then* $-\mathfrak{S}(t) \geqslant_{icv} -\widetilde{\mathfrak{S}}(t)$.

Remark 48. *The increasing convex order* $X \leqslant_{icx} Y$ *means* X *is smaller and less variable than* Y *in some stochastic sense, and* $X \leqslant_{icx} Y \iff \mathbb{E}\big[[X-a]^+\big] \leqslant \mathbb{E}\big[[Y-a]^+\big]$, $\forall a \in \mathbb{R}$ *indicates that* $\mathbb{E}X \leqslant \mathbb{E}Y$ *is a necessary and not a sufficient condition for* $X \leqslant_{icx} Y$.

Remark 49. *If* $X \leqslant_{icx} Y$ *and* $\mathbb{E}X \leqslant \mathbb{E}Z' \leqslant \mathbb{E}Y$, *then, it is possible that* $Z' \leqslant_{icx} Y$, *because we have* $X \leqslant_{icx} Y \iff X \leqslant_{st} Z \leqslant_{cx} Y$ *[127]. If* $X \leqslant_{icx} Y$, *then,* $X \leqslant_{cx} Z' \leqslant_{st} Y$ *such that* $Z' \leqslant_{icx} Y$, *because we have* $X \leqslant_{icx} Y \iff X \leqslant_{cx} Z \leqslant_{st} Y$ *[127]. Complementary results hold in the sense of the increasing concave order [127]. The mathematical relations indicate that the mean and the variability can trade off each other, i.e., if the variability is relatively small, then a relatively greater mean can be tolerated while satisfying the increasing convex order, and if the mean is relatively small, then a relatively greater variability can be tolerated while satisfying the increasing convex order.*

Remark 50. *The mean and variability trade-off with respect to the increasing convex or concave order tallies with the intuition, i.e., a smaller and less variable arrival process or a greater and less variable service process leads to a better performance. In addition, the consistency of the less variability of the arrival process and the service process tallies the dependence duality of the arrival process and service process. On the other hand, it indicates that the convex order describes the dependence duality more precisely than the increasing convex or concave order.*

Remark 51. *The increasing convex (concave) order is a sufficient but not necessary condition, in other words, the order of the mean values is not a necessary condition. A stricter sufficient condition is the moment generating function order* $\leqslant_{mgf}$ *(for the arrival process) and the Laplace transform order* $\leqslant_{Lt}$ *(for the service process) [35][36][127]. In addition, the moments order implies the moment generating function order while the reverse is not true, i.e.,* $X \leqslant_{mom} Y \implies X \leqslant_{mgf} Y$.

Remark 52. *Since the (increasing) convex order of a partial sum of a random vector is respectively insensitive and sensitive to the marginal variations in negative dependence scenarios and in the positive dependence scenarios [98], the negative dependence can be utilized to maintain a stable system that is robust to the bad marginal conditions, e.g., with small mean value, and the positive dependence can be utilized in case of good marginal conditions, e.g., with big mean value. Specifically, conditional on the identical marginals, the negative dependence has an advantage strictly over the positive dependence.*

Last, we show that the ordering of the arrival process and the service process consistently implies the ordering of the absolute decay rate. This result means that the dependence impact of the arrival process and service process on the measure identities are dual to each other, i.e., the same type of dependence therein influence the system performance in the same way.

Corollary 5. *Let* $\mathfrak{S}(t) = A(t) - S(t)$ *and* $\widetilde{\mathfrak{S}}(t) = \widetilde{A}(t) - \widetilde{S}(t)$.

- *If* $A(t) \leqslant_{cx} \widetilde{A}(t)$, $\forall t \in \mathbb{N}$, *and* $\mathfrak{S}(t)$ *and* $\widetilde{\mathfrak{S}}(t)$ *have identical service process* $S(t) = \widetilde{S}(t)$, *then* $0 < \widetilde{\gamma} \leqslant \gamma$ *and* $\kappa^{-S}(\gamma) \leqslant \widetilde{\kappa}^{-S}(\widetilde{\gamma}) < 0$.

- *If* $S(t) \leqslant_{cx} \widetilde{S}(t)$, $\forall t \in \mathbb{N}$, *and* $\mathfrak{S}(t)$ *and* $\widetilde{\mathfrak{S}}(t)$ *have identical arrival process* $A(t) = \widetilde{A}(t)$, *then* $0 > \widetilde{\gamma} \geqslant \gamma$ *and* $\kappa^{-A}(\gamma) \geqslant \widetilde{\kappa}^{-A}(\widetilde{\gamma}) > 0$.

Proof. The proof follows Theorem 20 and Theorem 21. □

In the extreme case, when both arrival process and service process are deterministic, the asymptotic decay rate is infinity. Generally, for any convex function ϕ, Jensen's inequality is expressed as $\phi(\mathbb{E}[\sum_{i=1}^{n} X_i]) \leqslant \mathbb{E}[\phi(\sum_{i=1}^{n} X_i)]$, by noting $\phi(\mathbb{E}[\sum_{i=1}^{n} X_i]) = \mathbb{E}[\phi(\mathbb{E}[\sum_{i=1}^{n} X_i])]$, we obtain that the mean of random variables has the minimum sum in convex order, i.e., $\sum_{i=1}^{n} \mathbb{E}[X_i] \leqslant_{cx} \sum_{i=1}^{n} X_i$, which indicates that deterministic arrival process or service process can cause the maximum asymptotic decay rate.

Corollary 6. *Under the conditions in Assumption 2. For (stationary or non-stationary) arrival and service processes, if the mean of the time average exists, i.e.,* $\mathbb{E}[X] = \lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{i=1}^{n} \mathbb{E}[X_i]$, *then the situation with the constant arrival process with the same mean or the constant service process with the same mean has the largest asymptotic decay rate of delay and of backlog.*

Proof. According to Jensen's inequality, $\phi(\mathbb{E}[\sum_{i=1}^{n} X_i]) \leqslant \mathbb{E}[\phi(\sum_{i=1}^{n} X_i)]$, $\forall\phi$, where ϕ is a convex function. For an arbitrary sequence of random variables, $X_1, X_2, \ldots, X_n$, we have $\sum_{i=1}^{n} \mathbb{E}\left[X_i\right] \leqslant_{cx} \sum_{i=1}^{n} X_i$, i.e., the mean of these random variables has the minimum sum in convex order. Specifically, $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}e^{\theta \sum_{i=1}^{n} X_i} \geqslant \theta \cdot \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\sum_{i=1}^{n} X_i]$, where the right hand side is equivalent to a constant process if the limit exists. This completes the proof. $\square$

3.2.2 Markov Specialization

We adopt the definition of Markov additive process that is introduced in Appendix B.2. We derive the non-asymptotic tail probabilities of delay and backlog on infinite time horizon and on finite time horizon, for Markov additive arrival process and Markov additive service process. The decay rate in the non-asymptotic probability results from sharp approximation and sufficiently implies the asymptotic decay rate resulting from logarithmic approximation. For both infinite time and finite time results, we only give results of the conditional tail probability, which is sufficient to calculate the tail probability by averaging over the initial state space.

We consider the time reversed arrival and service processes in the following results. Specifically, for the Markov additive process, the relations between the Perron-Frobenius eigenvalues and eigenvectors of the original process and the time reversed process are elaborated in Appendix B.2.

The following theorem presents the delay and backlog tail probabilities on infinite time horizon. The results have time-invariant decay rates, because they use a stability condition in the asymptotic time regime.

Theorem 22. *Consider a Markov additive arrival process $\left(J_t^A, A(t)\right)$ with state space E and initial state distribution $\boldsymbol{\varpi}_0^A$, and a Markov additive service process $\left(J_t^S, S(t)\right)$ with state space E' and initial distribution $\boldsymbol{\varpi}_0^S$. Specifically, given the initial state distribution, the state distribution at time t is $\boldsymbol{\varpi}_t = \boldsymbol{\varpi}_0 \boldsymbol{P}^t$, where $\boldsymbol{P}$ is the transition matrix. Denote $\kappa(\cdot)$ as the logarithmic eigenvalue and $\boldsymbol{h}(\cdot)$ as the right eigenvector of the kernel matrix. Assume independence between the arrival process and the service process.*

The delay tail probability, conditional on the initial state $\boldsymbol{J}_{d,0} = \boldsymbol{i}$, i.e., $\left\{J_d^A, J_0^{-S}\right\} = \left\{i^A, i^{-S}\right\}$, is expressed as

$$H_-^D \cdot h_{J_0}^{-S}(\theta) \cdot e^{-d\kappa(\theta)} \leqslant \mathbb{P}_{\boldsymbol{i}}(D > d) \leqslant H_+^D \cdot h_{J_0}^{-S}(\theta) \cdot e^{-d\kappa(\theta)}, \tag{3.50}$$

where $H_-^D = e^{-\kappa^A(\theta)} \cdot \left(\frac{\min_{j \in E} h_j^A(\theta)}{\max_{j \in E} h_j^A(\theta)} \right)^2 \cdot \frac{1}{\max_{j \in E'} h_j^{-S}(\theta)}$ *and* $H_+^D = \frac{\max_{j \in E} h_j^A(\theta)}{\min_{j \in E} h_j^A(\theta)} \cdot \frac{1}{\min_{j \in E'} h_j^{-S}(\theta)}$.

The backlog tail probability, conditional on the initial state $\boldsymbol{J}_{0,0} = \boldsymbol{i}$, i.e., $\left\{J_0^A, J_0^{-S}\right\} = \left\{i^A, i^{-S}\right\}$, is expressed as

$$H_-^B \cdot h_{J_0}^A(\theta) h_{J_0}^{-S}(\theta) \cdot e^{-\theta b} \leqslant \mathbb{P}_{\boldsymbol{i}}(B > b) \leqslant H_+^B \cdot h_{J_0}^A(\theta) h_{J_0}^{-S}(\theta) \cdot e^{-\theta b}, \tag{3.51}$$

where $H_-^B = e^{-\kappa^A(\theta)} \cdot \dfrac{\min\limits_{j\in E} h_j^A(\theta)}{\left(\max_{j\in E} h_j^A(\theta)\right)^2} \cdot \dfrac{1}{\max\limits_{j\in E'} h_j^{-S}(\theta)}$ *and* $H_+^B = \dfrac{1}{\min_{j\in E} h_j^A(\theta)} \cdot \dfrac{1}{\min_{j\in E'} h_j^{-S}(\theta)}$.

For delay and backlog, $\theta = \left\{\theta > 0 : \kappa^A(\theta) + \kappa^{-S}(\theta) = 0\right\}$ *and* $\kappa(\theta) := \kappa^A(\theta) = -\kappa^{-S}(\theta)$.

Proof. The idea of the proof is to find a likelihood ratio martingale of the process $A(d,t) - S(0,t)$ for delay and $A(t) - S(t)$ for backlog, change the measure, by the likelihood ratio identity, we obtain a likelihood ratio representation of the probability in the new measure. We only present the proof of delay. The proof of backlog follows analogically.

Recall the definition of the Markov additive process $\mathbb{E}[f(S(t+s) - S(t))g(J_{t+s})|\mathscr{F}_t] = \mathbb{E}_{J_t,0}[f(S(s))g(J_s)]$, which indicates that the time shift of the process is only dependent on the state at the shift epoch, specifically, for $\theta > 0$, the likelihood ratio martingale [7] of the arrival process $A(d,t)$ is expressed as $L_{t-d}^A \circ \theta_d = \frac{h_{J_t}^A(\theta)}{h_{J_d}^A(\theta)} e^{\theta A(d,t) - (t-d)\kappa^A(\theta)}$, where θ_d is the shift operator; and the likelihood ratio martingale [7] of the service process $-S(t)$ is $L_t^{-S} = \frac{h_{J_t}^{-S}(\theta)}{h_{J_0}^{-S}(\theta)} e^{-\theta S(0,t) - t\kappa^{-S}(\theta)}$. Assume the arrival process and the service process are independent, then the product of the martingales

$$L_{d,t}^{A-S} = \left(L_{t-d}^A \circ \theta_d\right) \cdot L_t^{-S} \tag{3.52}$$

is also a martingale [23], and $\mathbb{E}\left[L_{d,t}^{A-S}\right] = \mathbb{E}\left[L_{t-d}^A \circ \theta_d\right] \cdot \mathbb{E}\left[L_t^{-S}\right] = 1$.

Define the stopping time $\tau(d) = \inf\{t \geqslant d : A(d,t) - S(0,t) > 0\}$. Let $H(\theta) = \frac{h_{J_d}^A(\theta)}{h_{J_{\tau(d)}}^A(\theta)} \frac{h_{J_0}^{-S}(\theta)}{h_{J_{\tau(d)}}^{-S}(\theta)}$. The delay tail probability, conditional on the initial state $\boldsymbol{J}_{d,0} = \boldsymbol{i}$, i.e., $\left\{J_d^A, J_0^{-S}\right\} = \left\{i^A, i^{-S}\right\}$, is expressed as

$$\mathbb{P}_{\boldsymbol{i}}(D > d) = \mathbb{P}_{\boldsymbol{i}}(\tau(d) < \infty) \tag{3.53}$$

$$= \widetilde{\mathbb{E}}_{\boldsymbol{i}}\left[H(\theta) e^{-\theta\xi_{\tau(d)} + (\tau(d)-d)\kappa^A(\theta) + \tau(d)\kappa^{-S}(\theta)};\ \tau(d) < \infty\right], \tag{3.54}$$

where θ is the root to the stability equation $\kappa^A(\theta) + \kappa^{-S}(\theta) = 0$, and $\xi_{\tau(d)} > 0$ is the overshoot at the hitting time, which is bounded by

$$0 < \xi_{\tau(d)} < A(\tau(d) - 1, \tau(d)). \tag{3.55}$$

The delay upper bound is expressed as

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{i}}(D > d) = \mathbb{P}_{\boldsymbol{i}}(\tau(d) < \infty) &(3.56)\\
&\leqslant \widetilde{\mathbb{E}}_{\boldsymbol{i}}\left[H(\theta)e^{(\tau(d)-d)\kappa^A(\theta)+\tau(d)\kappa^{-S}(\theta)};\ \tau(d) < \infty\right] &(3.57)\\
&\leqslant H_+ \cdot h^{-S}_{J_0}(\theta)\widetilde{\mathbb{E}}_{\boldsymbol{i}}\left[e^{(\tau(d)-d)\kappa^A(\theta)+\tau(d)\kappa^{-S}(\theta)};\ \tau(d) < \infty\right] &(3.58)\\
&= H_+ \cdot h^{-S}_{J_0}(\theta) \cdot e^{-d\kappa^A(\theta)}, &(3.59)
\end{aligned}$$

where $H_+ = \frac{\max_{j\in E} h^A_j(\theta)}{\min_{j\in E} h^A_j(\theta)} \cdot \frac{1}{\min_{j\in E'} h^{-S}_j(\theta)}$.

The delay lower bound is expressed as

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{i}}(D > d) = \mathbb{P}_{\boldsymbol{i}}(\tau(d) < \infty) &(3.60)\\
&\geqslant \widetilde{\mathbb{E}}_{\boldsymbol{i}}\left[H(\theta)e^{-\theta A(\tau(d)-1,\tau(d))-d\kappa^A(\theta)};\ \tau(d) < \infty\right] &(3.61)\\
&\geqslant \widetilde{\mathbb{E}}_{\boldsymbol{i}}\left[e^{-\theta A(\tau(d)-1,\tau(d))};\ \tau(d) < \infty\right] \cdot \hat{H}_- \cdot h^{-S}_{J_0}(\theta) \cdot e^{-d\kappa^A(\theta)}, &(3.62)
\end{aligned}$$

where $\hat{H}_- = \frac{\min_{j\in E} h^A_j(\theta)}{\max_{j\in E} h^A_j(\theta)} \cdot \frac{1}{\max_{j\in E'} h^{-S}_j(\theta)}$ and

$$\begin{aligned}
&\widetilde{\mathbb{E}}_{\boldsymbol{i}}\left[e^{-\theta A(\tau(d)-1,\tau(d))};\ \tau(d) < \infty\right] &(3.63)\\
&= \mathbb{E}_{i^A}\left[e^{-\theta A(\tau(d)-1,\tau(d))} \cdot \left(L^A_{\tau(d)-d} \circ \theta_d\right);\ \tau(d) < \infty\right] &(3.64)\\
&= \mathbb{E}_{i^A}\left[\frac{h^A_{J_{\tau(d)}}}{h^A_{J_{\tau(d)-1}}} \cdot \left(L^A_{(\tau(d)-1)-d} \circ \theta_d\right) \cdot e^{-\kappa^A(\theta)};\ \tau(d) < \infty\right] &(3.65)\\
&\geqslant \frac{\min_{j\in E} h^A_j(\theta)}{\max_{j\in E} h^A_j(\theta)} \cdot e^{-\kappa^A(\theta)}, &(3.66)
\end{aligned}$$

where the first equality is due to the assumption of independence between the arrival process and service process, and the last inequality follows that $\left(L^A_{(\tau(d)-1)-d} \circ \theta_d\right)$ is a mean-one martingale. $\square$

Remark 53. *It is interesting to find tighter double-sided bounds with the Wiener-Hopf factorization [5]. However, this pair of bounds are sufficient for comparing the asymptotic decay rates of the distributions.*

The following theorem presents the time-dependent delay and backlog tail probability, the probability is a function of the violated delay and backlog, and the decay rates are time-variant, which results from the time-dependent optimization condition. If the asymptotic stability condition is used, the tail distribution bounds have an identical decay rate in finite time regime and infinite time regime.

Theorem 23. *Consider the same specification as in Theorem 22. Denote $\dot{\kappa}(x) \equiv \partial\kappa(x)/\partial x$.*

For delay, let γ be the root to $\kappa^A(\theta) + \kappa^{-S}(\theta) = 0$, $y_\gamma = \frac{\dot{\kappa}^A(\gamma)}{\dot{\kappa}^A(\gamma)+\dot{\kappa}^{-S}(\gamma)}$; given any fixed $y > 1$, θ is the root to $y\dot{\kappa}^{-S}(\theta) = -(y-1)\dot{\kappa}^A(\theta)$, and $\theta_y = -y\kappa^{-S}(\theta) - (y-1)\kappa^A(\theta)$, then

$$\mathbb{P}_{\boldsymbol{i}}(D(t) > d; t \leqslant yd) \leqslant H^D(\theta)e^{-d\theta_y}, \ y < y_\gamma, \tag{3.67}$$

$$\mathbb{P}_{\boldsymbol{i}}(D > d) - P_{\boldsymbol{i}}(D(t) > d; t \leqslant yd) \leqslant H^D(\theta)e^{-d\theta_y}, \ y > y_\gamma, \tag{3.68}$$

where $H^D(\theta) = H^D_+ h^{-S}_{J_0}(\theta)$ and $H^D_+ = \frac{\max_{j\in E} h^A_j(\theta)}{\min_{j\in E} h^A_j(\theta)} \cdot \frac{1}{\min_{j\in E'} h^{-S}_j(\theta)}$.

For backlog, let γ be the root to $\kappa^A(\theta) + \kappa^{-S}(\theta) = 0$, $y_\gamma = \frac{1}{\dot{\kappa}^A(\gamma)+\dot{\kappa}^{-S}(\gamma)}$; given any fixed $y > 0$, θ is the root to $y(\dot{\kappa}^A(\theta) + \dot{\kappa}^{-S}(\theta)) = 1$, and $\theta_y = \theta - y\left(\kappa^A(\theta) + \kappa^{-S}(\theta)\right)$, then

$$\mathbb{P}_{\boldsymbol{i}}(B(t) > b; t \leqslant yb) \leqslant H^B(\theta)e^{-b\theta_y}, \ y < y_\gamma, \tag{3.69}$$

$$\mathbb{P}_{\boldsymbol{i}}(B > b) - \mathbb{P}_{\boldsymbol{i}}(B(t) > b; t \leqslant yb) \leqslant H^B(\theta)e^{-b\theta_y}, \ y > y_\gamma, \tag{3.70}$$

where $H^B(\theta) = H^B_+ h^A_{J_0}(\theta) h^{-S}_{J_0}(\theta)$ and $H^B_+ = \frac{1}{\min_{j\in E} h^A_j(\theta)} \cdot \frac{1}{\min_{j\in E'} h^{-S}_j(\theta)}$.

Proof. The proof follows two phases. In the first phase, we provide the condition that the inequalities hold. In the second phase, we provide the setting of y that satisfies the condition. We only present the proof of delay. The proof of backlog follows analogically.

First, we prove that the inequalities hold under a condition on $\kappa^A(\theta)+\kappa^{-S}(\theta)$. Let $H(\theta)=\frac{h^A_{J_d}(\theta)}{h^A_{J_{\tau(d)}}(\theta)}\frac{h^{-S}_{J_0}(\theta)}{h^{-S}_{J_{\tau(d)}}(\theta)}$. For any $\theta>0$, $\kappa^A(\theta)+\kappa^{-S}(\theta)>0$.

$$
\begin{aligned}
&P_{\boldsymbol{i}}(D(t)>d; t\leqslant yd) && (3.71)\\
&=\widetilde{\mathbb{E}}_{\boldsymbol{i}}\left[H(\theta)e^{-\theta\xi_{\tau(d)}+(\tau(d)-d)\kappa^A(\theta)+\tau(d)\kappa^{-S}(\theta)};\ \tau(d)\leqslant yd\right] && (3.72)\\
&\leqslant\widetilde{\mathbb{E}}_{\boldsymbol{i}}\left[H(\theta)e^{(\tau(d)-d)\kappa^A(\theta)+\tau(d)\kappa^{-S}(\theta)};\ \tau(d)\leqslant yd\right] && (3.73)\\
&\leqslant H_+h^{-S}_{J_0}(\theta)\widetilde{\mathbb{E}}_{\boldsymbol{i}}\left[e^{(\tau(d)-d)\kappa^A(\theta)+\tau(d)\kappa^{-S}(\theta)};\ \tau(d)\leqslant yd\right] && (3.74)\\
&\leqslant H_+h^{-S}_{J_0}(\theta)e^{-d(-y\kappa^{-S}(\theta)-(y-1)\kappa^A(\theta))}. && (3.75)
\end{aligned}
$$

For any $\theta>0$, $\kappa^A(\theta)+\kappa^{-S}(\theta)<0$.

$$
\begin{aligned}
&P_{\boldsymbol{i}}(D>d)-P_{\boldsymbol{i}}(D(t)>d; t\leqslant yd) && (3.76)\\
&=\widetilde{\mathbb{E}}_{\boldsymbol{i}}\left[H(\theta)e^{-\theta\xi_{\tau(d)}+(\tau(d)-d)\kappa^A(\theta)+\tau(d)\kappa^{-S}(\theta)};\ yd<\tau(d)<\infty\right] && (3.77)\\
&\leqslant H_+h^{-S}_{J_0}(\theta)\widetilde{\mathbb{E}}_{\boldsymbol{i}}\left[e^{(\tau(d)-d)\kappa^A(\theta)+\tau(d)\kappa^{-S}(\theta)};\ yd<\tau(d)<\infty\right] && (3.78)\\
&\leqslant H_+h^{-S}_{J_0}(\theta)e^{-d(-y\kappa^{-S}(\theta)-(y-1)\kappa^A(\theta))}. && (3.79)
\end{aligned}
$$

Second, we link y to the $\kappa^A(\theta)+\kappa^{-S}(\theta)$ condition. Denote $\theta_y=-y\kappa^{-S}(\theta)-(y-1)\kappa^A(\theta)$, which is a concave function of θ for any fixed $y>1$. Thus, the optimal θ^* to maximize θ_y is the root to the derivative equation $\dot{\theta}_y=0$, i.e., $\theta^*=\left\{\theta: y\dot{\kappa}^{-S}(\theta)=-(y-1)\dot{\kappa}^A(\theta)\right\}$. Consider the equation $\frac{1}{y}=1+\frac{\dot{\kappa}^{-S}(\theta)}{\dot{\kappa}^A(\theta)}$, since $\frac{\partial}{\partial\theta}\left(\frac{\dot{\kappa}^{-S}(\theta)}{\dot{\kappa}^A(\theta)}\right)=\frac{\ddot{\kappa}^{-S}(\theta)\cdot\dot{\kappa}^A(\theta)-\ddot{\kappa}^A(\theta)\cdot\dot{\kappa}^{-S}(\theta)}{[\dot{\kappa}^A(\theta)]^2}\geqslant 0$, which indicates that the decrease of y maps to the increase of θ, it follows, if $y<\frac{\dot{\kappa}^A(\gamma)}{\dot{\kappa}^A(\gamma)+\dot{\kappa}^{-S}(\gamma)}$, then $\theta>\gamma$ and $\kappa^A(\theta)+\kappa^{-S}(\theta)>0$, vice versa. □

3.3 Dependence Manipulation

3.3.1 The Theoretical Reality

We treat a stochastic process as a function of a set of random parameters, each of which is itself a stochastic process. In other words, we treat the stochastic process as a functional of a multivariate stochastic process and the functional maps the multivariate stochastic process to a univariate stochastic process. For wireless channel capacity, the random parameters are either uncontrollable or controllable, the uncontrollable parameters represent the property of the environment that can not be interfered, e.g., fading, and the controllable parameters represent the configurable property of the wireless system, e.g., power. For the arrival, this functional perspective is useful for studying the dependence impact of an individual arrival process on the aggregation of a set of multiplexed arrival processes.

We specify that the dimension of the parameter set $\left(X_t^1, X_t^2, \ldots, X_t^n\right)$ is time-invariant, the function $f_t : \mathbb{R}^n \to \mathbb{R}$ is time-variant, and the function f_t is increasing or decreasing at X_t^i for all the time, i.e.,

$$X_t = f_t\left(X_t^1, X_t^2, \ldots, X_t^n\right). \tag{3.80}$$

Leveraging this functional perspective, on the one hand, we prove the dependence transformability, on the other hand, we provide additional results of supermodular order to the literature [98][127].

The following theorem shows that the manipulation of the dependence in the controllable parameter processes transforms the dependence structure of the stochastic process.

Theorem 24. *Assume the random parameters are spatially independent and temporally dependent. The supermodular ordering of a parameter series implies the ordering of the stochastic process, i.e., for any* $1 \leqslant i \leqslant n$, *if*

$$\left(X_1^i, X_2^i, \ldots, X_t^i\right) \leqslant_{sm} \left(\widetilde{X}_1^i, \widetilde{X}_2^i, \ldots, \widetilde{X}_t^i\right), \tag{3.81}$$

then

$$\left(X_1, X_2, \ldots, X_t\right) \leqslant_{sm} \left(\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_t\right), \tag{3.82}$$

where $\widetilde{X}_j = f_j(X_j^1, \ldots, X_j^{i-1}, \widetilde{X}_j^i, X_j^{i+1}, \ldots, X_j^n)$, $\forall 1 \leqslant j \leqslant t$.

Proof. Without loss of generalization, we consider the supermodular order of the random parameters with index 1 in the proof.

For all increasing or all decreasing functions $g_i : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, t$, we have $(X_1^1, X_2^1, \ldots, X_t^1) \leqslant_{sm} (\widetilde{X}_1^1, \widetilde{X}_2^1, \ldots, \widetilde{X}_t^1) \implies \left(g_1\left(X_1^1\right), \ldots, g_t\left(X_t^1\right)\right)$ $\leqslant_{sm} \left(g_1\left(\widetilde{X}_1^1\right), \ldots, g_t\left(\widetilde{X}_t^1\right)\right)$, because a composition of a supermodular function with coordinatewise functions, that are all increasing or are all decreasing, is a supermodular function [127].

Let $\boldsymbol{Z}_t = \left(X_t^2, X_t^3, \ldots, X_t^n\right)$ and assume $\boldsymbol{Z}_t$ is independent of X_t^1, $\forall t$. Then, $\left(g_1\left(X_1^1\right), \ldots, g_t\left(X_t^1\right)\right) \leqslant_{sm} \left(g_1\left(\widetilde{X}_1^1\right), \ldots, g_t\left(\widetilde{X}_t^1\right)\right)$ implies

$$\begin{aligned}&\left(f_1\left(X_1^1, \boldsymbol{Z}_1\right), \ldots, f_t\left(X_t^1, \boldsymbol{Z}_t\right) \middle| \left(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_t\right) = \boldsymbol{z}\right)\\ &\qquad\leqslant_{sm} \left(f_1\left(\widetilde{X}_1^1, \boldsymbol{Z}_1\right), \ldots, f_t\left(\widetilde{X}_t^1, \boldsymbol{Z}_t\right) \middle| \left(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_t\right) = \boldsymbol{z}\right), \forall \boldsymbol{z},\end{aligned}$$

when $f_i(x_i^1, \boldsymbol{z}_i)$, $\forall i$, are all increasing or are all decreasing in x_i^1 for every $\boldsymbol{z}_i$. Finally, $(f_1(X_1^1, \boldsymbol{Z}_1), \ldots, f_t(X_t^1, \boldsymbol{Z}_t)) \leqslant_{sm} (f_1(\widetilde{X}_1^1, \boldsymbol{Z}_1), \ldots, f_t(\widetilde{X}_t^1, \boldsymbol{Z}_t))$, because the sumermodular order is closed under mixtures [127]. □

Remark 54. *If the dimension of the parameter set is time-variant, the function must take into account the dimension being manipulated, while it is not necessary to take into account other dimensions. Specifically, for the dependence in the manipulated dimension to take effect, at least two time epochs should be taken into account by the functions and it is not necessary for the functions to take into account every time epoch, because the composition of a supermodular function with coordinatewise functions, which are* *all increasing or are all decreasing, is a supermodular function, and strict monotonnity is not necessary. An example is random multiplexing, where the number of integrated arrivals is random at each time epoch.*

Remark 55. *Since the supermodular order is closed with respect to the permutation of the time index [98], the dependence transformability result is robust to the time reversibility assumption.*

The following theorem shows that a greater number of controllable random parameter processes brings a stronger transform to the dependence structure of the stochastic process.

Theorem 25. *Assume the random parameters are spatially independent and temporally dependent. Consider there are i, $1 \leqslant i \leqslant n$, controllable random parameters. If*

$$\left(X_1^j, X_2^j, \ldots, X_t^j\right) \leqslant_{sm} \left(\widetilde{X}_1^j, \widetilde{X}_2^j, \ldots, \widetilde{X}_t^j\right), \ \forall 1 \leqslant j \leqslant i, \tag{3.83}$$

then

$$\widetilde{\boldsymbol{X}}_t^k \leqslant_{sm} \widetilde{\boldsymbol{X}}_t^j, \ \forall 0 \leqslant k \leqslant j \leqslant i, \tag{3.84}$$

with $\widetilde{X}_{t_m}^l = f_m(\widetilde{X}_m^1, \ldots, \widetilde{X}_m^l, X_m^{l+1}, \ldots, X_m^n)$, $1 \leqslant m \leqslant t$, and $\widetilde{\boldsymbol{X}}_t^l = (\widetilde{X}_{t_1}^l, \ldots, \widetilde{X}_{t_t}^l)$, $l \in \{k, j\}$.

Proof. According to Theorem 24,

$$\begin{aligned}
&\left(f_1\left(X_1^1, X_1^2, \ldots, X_1^n\right), \ldots, f_t\left(X_t^1, X_t^2, \ldots, X_t^n\right)\right) \\
&\leqslant_{sm} \left(f_1\left(\widetilde{X}_1^1, X_1^2, \ldots, X_1^n\right), \ldots, f_t\left(\widetilde{X}_t^1, X_t^2, \ldots, X_t^n\right)\right) \\
&\leqslant_{sm} \left(f_1\left(\widetilde{X}_1^1, \widetilde{X}_1^2, X_1^3, \ldots, X_1^n\right), \ldots, f_t\left(\widetilde{X}_t^1, \widetilde{X}_t^2, X_t^3, \ldots, X_t^n\right)\right),
\end{aligned}$$

and the result follows iteratively because of the transitivity property of supermodular order [98]. □

In practice, a wireless channel can be a composition of sub-channels and the wireless channel capacity is a function of each sub-channel's capacity, e.g., multiple-input-multiple-output channel and frequency-selective channel. The following theorem shows that the manipulation of dependence in a sub-channel capacity transforms the dependence structure of the overall channel capacity, the more number of manipulated sub-channels the stronger dependence transform strength on the overall capacity.

Theorem 26. *Consider a wireless channel composing of M independent sub-channels, the instantaneous capacity of the overall channel, denoted by $\mathfrak{C}_t \equiv \mathfrak{C}(t)$, is a function of the instantaneous capacity of each sub-channels, denoted by $c_t^i \equiv c^i(t)$, $1 \leqslant i \leqslant M$, i.e., $\mathfrak{C}_t = f_t\left(c_t^1, \ldots, c_t^M\right)$. For example, the overall capacity is the sum of the capacity of each sub-channel, i.e., $\left(c_t^1, \ldots, c_t^M\right) \mapsto f_t\left(c_t^1, \ldots, c_t^M\right) = \sum_{m=1}^{M} c_t^m$. Assume the function is always increasing or decreasing at the instantaneous capacity of each sub-channels.*

If $\left(c_1^i, \ldots, c_t^i\right) \leqslant_{sm} \left(\tilde{c}_1^i, \ldots, \tilde{c}_t^i\right)$, $\forall 1 \leqslant i \leqslant M$, then

$$(\mathfrak{C}_1, \ldots, \mathfrak{C}_t) \leqslant_{sm} (\widetilde{\mathfrak{C}}_1, \ldots, \widetilde{\mathfrak{C}}_t), \tag{3.85}$$

where $\widetilde{\mathfrak{C}}_j = f_j\left(c_j^1, \ldots, \tilde{c}_j^i, c_j^{i+1}, \ldots, c_j^M\right)$, $1 \leqslant j \leqslant t$.

If $\left(c_1^i, \ldots, c_t^i\right) \leqslant_{sm} \left(\tilde{c}_1^i, \ldots, \tilde{c}_t^i\right)$, $\forall 1 \leqslant i \leqslant M$, then

$$\widetilde{\boldsymbol{\mathfrak{C}}}_t^k \leqslant_{sm} \widetilde{\boldsymbol{\mathfrak{C}}}_t^j, \ \forall 0 \leqslant k \leqslant j \leqslant M, \tag{3.86}$$

with $\widetilde{\mathfrak{C}}_{t_m}^l = f_m\left(\tilde{c}_m^1, \ldots, \tilde{c}_m^l, c_m^{l+1}, \ldots, c_m^M\right)$ and $\widetilde{\boldsymbol{\mathfrak{C}}}_t^l = \left(\widetilde{\mathfrak{C}}_{t_1}^l, \ldots, \widetilde{\mathfrak{C}}_{t_t}^l\right)$, where $1 \leqslant m \leqslant t$ and $l \in \{k, j\}$.

Proof. The proof follows Theorem 24 and Theorem 25. □

Multiplexing of arrival processes involves deterministic multiplexing and random multiplexing. Specifically, the aggregated arrival process is expressed as $\mathfrak{A}(t) = \sum_{i=1}^{N(t)} a_i(t)$, where $a_i(t)$ is the individual arrival process, and $N(t)$ is a deterministic process for the deterministic multiplexing and a stochastic process for the random multiplexing. Theorem 26 has a direct application to the deterministic multiplexing, i.e., the dependence manipulation in an individual arrival process transforms the dependence structure of the aggregate process, the more number of manipulated processes the stronger strength of dependence transform.

For random multiplexing, in addition to that the dependence manipulation in an individual process transforms the dependence structure of the randomly multiplexed process, the following theorem shows that the dependence in the random multiplexing mechanism also has an impact on the dependence structure of the aggregated process.

Theorem 27. *Let $\boldsymbol{X}_j = (X_{j,1}, \ldots, X_{j,m})$ and $\boldsymbol{Y}_j = (Y_{j,1}, \ldots, Y_{j,m})$, $j = 1, 2, \ldots$, be two independent sequences of non-negative random vectors, and let $\boldsymbol{M} = (M_1, M_2, \ldots, M_m)$ and $\boldsymbol{N} = (N_1, N_2, \ldots, N_m)$ be two vectors of non-negative integer-valued random variables. Assume that both $\boldsymbol{M}$ and $\boldsymbol{N}$ are independent of the $\boldsymbol{X}_j$'s and $\boldsymbol{Y}_j$'s.*

If $\boldsymbol{M} \leqslant_{sm} \boldsymbol{N}$, then

$$\left(\sum_{j=1}^{M_1} X_{j,1}, \ldots, \sum_{j=1}^{M_m} X_{j,m}\right) \leqslant_{sm} \left(\sum_{j=1}^{N_1} X_{j,1}, \ldots, \sum_{j=1}^{N_m} X_{j,m}\right). \tag{3.87}$$

If $\boldsymbol{X}_j \leqslant_{sm} \boldsymbol{Y}_j$, $\forall j$, then

$$\left(\sum_{j=1}^{N_1} X_{j,1}, \ldots, \sum_{j=1}^{N_m} X_{j,m}\right) \leqslant_{sm} \left(\sum_{j=1}^{N_1} Y_{j,1}, \ldots, \sum_{j=1}^{N_m} Y_{j,m}\right). \tag{3.88}$$

If $\boldsymbol{M} \leqslant_{sm} \boldsymbol{N}$ and $\boldsymbol{X}_j \leqslant_{sm} \boldsymbol{Y}_j$, $\forall j$, then

$$\left(\sum_{j=1}^{M_1} X_{j,1}, \ldots, \sum_{j=1}^{M_m} X_{j,m}\right) \leqslant_{sm} \left(\sum_{j=1}^{N_1} Y_{j,1}, \ldots, \sum_{j=1}^{N_m} Y_{j,m}\right). \tag{3.89}$$

Proof. The first result is proved in [127]. The second result follows Theorem 24 and Theorem 25, by conditioning on $(N_1, \ldots, N_m) = (n_1, \ldots, n_m)$ and integrating for the expectation. Specifically, since

$$\begin{aligned}
&\mathbb{E}\left[\phi\left(\sum_{j=1}^{N_1} X_{j,1}, \ldots, \sum_{j=1}^{N_m} X_{j,m}\right) \Big| (N_1, \ldots, N_m) = (n_1, \ldots, n_m)\right] \\
&\leqslant \mathbb{E}\left[\phi\left(\sum_{j=1}^{N_1} Y_{j,1}, \ldots, \sum_{j=1}^{N_m} Y_{j,m}\right) \Big| (N_1, \ldots, N_m) = (n_1, \ldots, n_m)\right],
\end{aligned}$$

for any supermodular function ϕ, thus

$$\mathbb{E}\left[\phi\left(\sum_{j=1}^{N_1} X_{j,1}, \ldots, \sum_{j=1}^{N_m} X_{j,m}\right)\right] \leqslant \mathbb{E}\left[\phi\left(\sum_{j=1}^{N_1} Y_{j,1}, \ldots, \sum_{j=1}^{N_m} Y_{j,m}\right)\right].$$

By considering the transitivity property of supermodular order, the third result follows the first and second results. □

Remark 56. *The arrival process under manipulation must be multiplexed at least twice at different time epochs, in order to take into account the dependence effect of this arrival process.*

Remark 57. *It is interesting to extend the results to the increasing supermodular order $\leqslant_{ism}$ and the symmetric supermodular order $\leqslant_{symsm}$.*

3.3.2 A Copula Approach

We develop a dependence manipulation technique based on copula, which is introduced in Appendix B.3. We consider the stochastic process that is modeled by a multivariate Markov process of uncontrollable parameters and controllable parameters. The challenge with dependence control lies in that, if we model the whole system of the uncontrollable and controllable parameters as a Markov process, we need to know respectively the exact characterizations of the uncontrollable and controllable parameters processes, because we need an exact process for dependence manipulation. To address this challenge, we postulate no-Granger causality among the random parameter processes.

No-Granger causality refers to a multivariate dynamic system in which each variable is determined by its own lagged values and no further information is provided by the lagged values of other variables [25]. On the one hand, no-Granger causality and Markov property of each process with respect to its natural filtration together imply the whole system as a Markov process, on the other hand, a multidimensional Markov process together with no-Granger causality implies that each of its components is a Markov process with respect to its own natural filtration [24, 25].

Mathematically, for a n-dimensional process $\boldsymbol{X}$, we say $\boldsymbol{X}^1, \ldots, \boldsymbol{X}^{i-1}$, $\boldsymbol{X}^{i+1}, \ldots, \boldsymbol{X}^n$ do not Granger cause $\boldsymbol{X}^i$, if, for any t_k and x,

$$\mathbb{P}\left(X^i_{t_{k+1}} \leqslant x | \mathscr{F}^{\boldsymbol{X}^1,\ldots,\boldsymbol{X}^n}_{t_k}\right) = \mathbb{P}\left(X^i_{t_{k+1}} \leqslant x | \mathscr{F}^{\boldsymbol{X}^i}_{t_k}\right), \tag{3.90}$$

where $\mathscr{F}^{\boldsymbol{X}^1,\ldots,\boldsymbol{X}^n}_{t_k}$ is the natural filtration of process $\boldsymbol{X}$ and $\mathscr{F}^{\boldsymbol{X}^i}_{t_k}$ is the natural filtration of process $\boldsymbol{X}^i$ [24, 25]. In addition, we introduce a copula operator. Denote $A(\mathbf{x}, d\mathbf{y}) = A_{,C}(\mathbf{x}, \mathbf{y})C(d\mathbf{y})$ and $B(d\mathbf{x}, \mathbf{y}) = B_{C,}(\mathbf{x}, \mathbf{y})C(d\mathbf{x})$. The operator $\overset{C(\cdot)}{\star}$ is defined as [104]

$$(A \overset{C(\mathbf{z})}{\star} B)(\mathbf{x}, \mathbf{y}) = \int_0^{\mathbf{z}} A_{,C}(\mathbf{x}, \mathbf{r}) \cdot B_{C,}(\mathbf{r}, \mathbf{y})C(d\mathbf{r}), \tag{3.91}$$

where A is a $(k+n)$-dimensional copula, B is a $(n+l)$-dimensional copula, C is a n-dimensional copula. Specifically, $A \star B$ is written as [32] $A \star B\,(\mathbf{x}, z, \mathbf{y}) = \int_0^z A_{,r}(\mathbf{x}, r)B_{r,}(r, \mathbf{y})dr$, for 1-dimensional copula C. We refer to Appendix B.3 and [32, 104] for elaboration on the copula representation of Markov process.

The copula for Markov process with no-Granger causality is expressed as follows, which is extension of the 2-dimensional result in [25].

Theorem 28. *For a n-dimensional Markov process $\boldsymbol{X}$ consisting of two sets $\overline{\boldsymbol{X}}$ and $\underline{\boldsymbol{X}}$, $\boldsymbol{X} = \overline{\boldsymbol{X}} \bigcup \underline{\boldsymbol{X}}$, mapping respectively to the uncontrollable and controllable parameters, $\overline{\boldsymbol{X}}$ does not Granger cause $\underline{\boldsymbol{X}}$, if and only if*

$$\begin{aligned}
&C_{j,j+1}\left(\boldsymbol{u}_{\underline{\boldsymbol{X}}_j}, \boldsymbol{u}_{\overline{\boldsymbol{X}}_j}, \boldsymbol{u}_{\underline{\boldsymbol{X}}_{j+1}}, \mathbf{1}_{\boldsymbol{u}_{\overline{\boldsymbol{X}}_{j+1}}}\right) \\
&= C_{\overline{\boldsymbol{X}}_j \underline{\boldsymbol{X}}_j} \overset{C_{\underline{\boldsymbol{X}}_j}\left(\boldsymbol{u}_{\underline{\boldsymbol{X}}_j}\right)}{\star} C_{\underline{\boldsymbol{X}}_j \underline{\boldsymbol{X}}_{j+1}}\left(\boldsymbol{u}_{\overline{\boldsymbol{X}}_j}, \boldsymbol{u}_{\underline{\boldsymbol{X}}_{j+1}}\right),
\end{aligned} \tag{3.92}$$

$\underline{\boldsymbol{X}}$ does not Granger cause $\overline{\boldsymbol{X}}$, if and only if

$$\begin{aligned}
&C_{j,j+1}\left(\boldsymbol{u}_{\underline{\boldsymbol{X}}_j}, \boldsymbol{u}_{\overline{\boldsymbol{X}}_j}, \mathbf{1}_{\boldsymbol{u}_{\underline{\boldsymbol{X}}_{j+1}}}, \boldsymbol{u}_{\overline{\boldsymbol{X}}_{j+1}}\right) \\
&= C_{\underline{\boldsymbol{X}}_j \overline{\boldsymbol{X}}_j} \overset{C_{\overline{\boldsymbol{X}}_j}\left(\boldsymbol{u}_{\overline{\boldsymbol{X}}_j}\right)}{\star} C_{\overline{\boldsymbol{X}}_j \overline{\boldsymbol{X}}_{j+1}}\left(\boldsymbol{u}_{\underline{\boldsymbol{X}}_j}, \boldsymbol{u}_{\overline{\boldsymbol{X}}_{j+1}}\right).
\end{aligned} \tag{3.93}$$

Proof. Since

$$\begin{aligned}
\mathbb{P}\left(\underline{\boldsymbol{X}}_{j+1} \leqslant \boldsymbol{x} | \boldsymbol{X}_j\right) &= C_{j,j+1_{C_{\underline{\boldsymbol{X}}_j \overline{\boldsymbol{X}}_j},}}\left(F_{\underline{\boldsymbol{X}}_j}(\underline{\boldsymbol{X}}_j), F_{\overline{\boldsymbol{X}}_j}(\overline{\boldsymbol{X}}_j), F_{\underline{\boldsymbol{X}}_{j+1}}(\boldsymbol{x}), \mathbf{1}_{F_{\overline{\boldsymbol{X}}_{j+1}}}\right), \\
\mathbb{P}\left(\underline{\boldsymbol{X}}_{j+1} \leqslant \boldsymbol{x} | \underline{\boldsymbol{X}}_j\right) &= C_{j,j+1_{C_{\underline{\boldsymbol{X}}_j},}}\left(F_{\underline{\boldsymbol{X}}_j}(\underline{\boldsymbol{X}}_j), \mathbf{1}_{F_{\overline{\boldsymbol{X}}_j}}, F_{\underline{\boldsymbol{X}}_{j+1}}(\boldsymbol{x}), \mathbf{1}_{F_{\overline{\boldsymbol{X}}_{j+1}}}\right),
\end{aligned}$$

the no-Granger causality holds, if and only if

$$\begin{aligned}
&C_{j,j+1_{C_{\underline{\boldsymbol{X}}_j \overline{\boldsymbol{X}}_j},}}\left(\boldsymbol{u}_{\underline{\boldsymbol{X}}_j}, \boldsymbol{u}_{\overline{\boldsymbol{X}}_j}, \boldsymbol{u}_{\underline{\boldsymbol{X}}_{j+1}}, \mathbf{1}_{\boldsymbol{u}_{\overline{\boldsymbol{X}}_{j+1}}}\right) \\
&\qquad = C_{j,j+1_{C_{\underline{\boldsymbol{X}}_j},}}\left(\boldsymbol{u}_{\underline{\boldsymbol{X}}_j}, \mathbf{1}_{\boldsymbol{u}_{\overline{\boldsymbol{X}}_j}}, \boldsymbol{u}_{\underline{\boldsymbol{X}}_{j+1}}, \mathbf{1}_{\boldsymbol{u}_{\overline{\boldsymbol{X}}_{j+1}}}\right).
\end{aligned}$$

By integrating, we obtain

$$\begin{aligned}
&C_{j,j+1}\left(\boldsymbol{u}_{\underline{\boldsymbol{X}}_j}, \boldsymbol{u}_{\overline{\boldsymbol{X}}_j}, \boldsymbol{u}_{\underline{\boldsymbol{X}}_{j+1}}, \mathbf{1}_{\boldsymbol{u}_{\overline{\boldsymbol{X}}_{j+1}}}\right) \\
&= \int_{\mathbf{0}}^{\boldsymbol{u}_{\underline{\boldsymbol{X}}_j}} C_{\overline{\boldsymbol{X}}_j \underline{\boldsymbol{X}}_j, C_{\underline{\boldsymbol{X}}_j}}\left(\boldsymbol{u}_{\overline{\boldsymbol{X}}_j}, \boldsymbol{u}_{\underline{\boldsymbol{X}}}\right) \cdot C_{\underline{\boldsymbol{X}}_j \underline{\boldsymbol{X}}_{j+1_{C_{\underline{\boldsymbol{X}}_j},}}}\left(\boldsymbol{u}_{\underline{\boldsymbol{X}}}, \boldsymbol{u}_{\underline{\boldsymbol{X}}_{j+1}}\right) C_{\underline{\boldsymbol{X}}_j}\left(d\boldsymbol{u}_{\underline{\boldsymbol{X}}}\right) \\
&= C_{\overline{\boldsymbol{X}}_j \underline{\boldsymbol{X}}_j} \overset{C_{\underline{\boldsymbol{X}}_j}(\boldsymbol{u}_{\underline{\boldsymbol{X}}_j})}{\star} C_{\underline{\boldsymbol{X}}_j \underline{\boldsymbol{X}}_{j+1}}\left(\boldsymbol{u}_{\overline{\boldsymbol{X}}_j}, \boldsymbol{u}_{\underline{\boldsymbol{X}}_{j+1}}\right).
\end{aligned}$$

The other result follows analogically. □

Remark 58. *The result is extensible to high order Markov process [104][64]. For a Markov process of order $k \geqslant 1$, letting $j(k) \equiv \{j-k+1, \ldots, j\}$, $\overline{\boldsymbol{X}}$ does not Granger cause $\underline{\boldsymbol{X}}$, if and only if*

$$\begin{aligned}
& C_{j(k),j+1}\left(\boldsymbol{u}_{\underline{\boldsymbol{X}}_{j(k)}}, \boldsymbol{u}_{\overline{\boldsymbol{X}}_{j(k)}}, \boldsymbol{u}_{\underline{\boldsymbol{X}}_{j+1}}, \mathbf{1}_{\boldsymbol{u}_{\overline{\boldsymbol{X}}_{j+1}}}\right) \\
& = C_{\overline{\boldsymbol{X}}_{j(k)}\underline{\boldsymbol{X}}_{j(k)}} \overset{C_{\underline{\boldsymbol{X}}_{j(k)}}\left(\boldsymbol{u}_{\underline{\boldsymbol{X}}_{j(k)}}\right)}{\star} C_{\underline{\boldsymbol{X}}_{j(k)}\underline{\boldsymbol{X}}_{j+1}}\left(\boldsymbol{u}_{\overline{\boldsymbol{X}}_{j(k)}}, \boldsymbol{u}_{\underline{\boldsymbol{X}}_{j+1}}\right),
\end{aligned} \tag{3.94}$$

and $\underline{\boldsymbol{X}}$ does not Granger cause $\overline{\boldsymbol{X}}$, if and only if

$$\begin{aligned}
& C_{j(k),j+1}\left(\boldsymbol{u}_{\underline{\boldsymbol{X}}_{j(k)}}, \boldsymbol{u}_{\overline{\boldsymbol{X}}_{j(k)}}, \mathbf{1}_{\boldsymbol{u}_{\underline{\boldsymbol{X}}_{j+1}}}, \boldsymbol{u}_{\overline{\boldsymbol{X}}_{j+1}}\right) \\
& = C_{\underline{\boldsymbol{X}}_{j(k)}\overline{\boldsymbol{X}}_{j(k)}} \overset{C_{\overline{\boldsymbol{X}}_{j(k)}}\left(\boldsymbol{u}_{\overline{\boldsymbol{X}}_{j(k)}}\right)}{\star} C_{\overline{\boldsymbol{X}}_{j(k)}\overline{\boldsymbol{X}}_{j+1}}\left(\boldsymbol{u}_{\underline{\boldsymbol{X}}_{j(k)}}, \boldsymbol{u}_{\overline{\boldsymbol{X}}_{j+1}}\right).
\end{aligned} \tag{3.95}$$

A stronger restriction is that all the 1-dimensional Markov processes do not Granger cause each other, and the results are as follows.

Theorem 29. *For a n-dimensional Markov process $\boldsymbol{X}$ with temporal copula $C_{j,j+1}$ and spatial copula C_j,*

$$\mathbb{P}\left(X^i_{t_{k+1}} \leqslant x | \boldsymbol{X}^1_{t_k}, \ldots, \boldsymbol{X}^n_{t_k}\right) = \mathbb{P}\left(X^i_{t_{k+1}} \leqslant x | \boldsymbol{X}^i_{t_k}\right), \tag{3.96}$$

if and only if

$$\begin{aligned}
& C_{j,j+1}\left(x^1_j, \ldots, x^n_j, 1, \ldots, x^i_{j+1}, \ldots, 1\right) \\
& = C'^{i}_j \star C^i_{j,j+1}\left(x^1_j, \ldots, x^{i-1}_j, x^{i+1}_j, \ldots, x^n_j, x^i_j, x^i_{j+1}\right),
\end{aligned} \tag{3.97}$$

where C'^{i}_j is the reordered spatial copula, and $C^i_{j,j+1}$ is the temporal copula of the 1-dimensional Markov process $\boldsymbol{X}^i$.

Proof. The proof analogically follows Theorem 28's. □

Example 1. *For a* 2*-dimensional Markov process* $\boldsymbol{X}$, $\boldsymbol{X}^2$ *does not Granger cause* $\boldsymbol{X}^1$, *if and only if [25]*

$$C_{j,j+1}(u_1, v_1, u_2, 1) = C_{X_j^2, X_j^1} \star C_{X_j^1, X_{j+1}^1}(v_1, u_1, u_2), \tag{3.98}$$

and $\boldsymbol{X}^1$ *does not Granger cause* $\boldsymbol{X}^2$, *if and only if [25]*

$$C_{j,j+1}(u_1, v_1, 1, v_2) = C_{X_j^1, X_j^2} \star C_{X_j^2, X_{j+1}^2}(u_1, v_1, u_2). \tag{3.99}$$

In the special case, if the spatial dependence is expressed by the product copula, then

$$C_{j,j+1}(u_1, v_1, u_2, 1) = v_1 C_{X_j^1 X_{j+1}^1}(u_1, u_2), \tag{3.100}$$

$$C_{j,j+1}(u_1, v_1, 1, v_2) = u_1 C_{X_j^2 X_{j+1}^2}(v_1, v_2). \tag{3.101}$$

To perform dependence manipulation, we calculate the transition matrix of the controllable parameter Markov process based on its copula, which is an inverse procedure of constructing a copula from the transition matrix of a Markov process [32]. Since the Markov property is a pure property of copula, it is flexible to use different copula functions to character and configure the negative or positive dependence in the uncontrollable and controllable parameters. The calculation approach is summarized in the following theorem.

Theorem 30. *For a* 1*-dimensional Markov process with finite state space* E *and initial distribution* $\boldsymbol{\varpi}$, *given the copula between successive levels* $C_{j,j+1}$,

$$\sum_{s_j \leqslant x} \boldsymbol{\varpi}_j(s_j) \boldsymbol{P}_j(s_j, s_{j+1} \leqslant y) = C_{j,j+1}\left(F_j(x), F_{j+1}(y)\right), \ \forall x, y \in E, \tag{3.102}$$

the state distribution at j *is* $\boldsymbol{\varpi}_j = \boldsymbol{\varpi} \prod_{0 \leqslant k \leqslant j} \boldsymbol{P}_k$, *and* $F_j(s_j) = \sum_{s_k \leqslant s_j} \boldsymbol{\varpi}_j(s_k)$

and $F_{j+1} = \boldsymbol{\varpi}_j \boldsymbol{P}_j$*. Together with the unity property of transition matrix* $\sum_{j \in E} p_{ij} = 1$*,* $\forall i \in E$*, the transition probabilities* $\boldsymbol{P}_j$ *are obtained.*

Proof. For random variables X and Y with the copula C, $\mathbb{E}\left(I_{Y<y}|X\right)(\omega) = C_{1,}\left(F_X(X(\omega)), F_Y(y)\right)$, a.s. [32], by integrating, $\sum\limits_{\xi \leqslant x} \mathbb{P}\left(X_t \leqslant y | X_s = \xi\right) = C\left(F_s(x), F_t(y)\right)$, where $x \in E$ and $y \in E$. The result directly follows. $\square$

Example 2. *For a 2-state homogeneous Markov process, the equations are expressed as*

$$\begin{cases} \pi_0 p_{00} = C\left(F(0), F(0)\right), & \text{(3.103a)} \\ \pi_0 p_{00} + \pi_1 p_{10} = C\left(F(1), F(0)\right), & \text{(3.103b)} \\ \pi_0\left(p_{00} + p_{01}\right) + \pi_1\left(p_{10} + p_{11}\right) = C\left(F(1), F(1)\right), & \text{(3.103c)} \\ \pi_0\left(p_{00} + p_{01}\right) = C\left(F(0), F(1)\right). & \text{(3.103d)} \end{cases}$$

Given a stationary distribution $[\pi_0\ \pi_1]$*,* $F(0) = \pi_0$ *and* $F(1) = \pi_0 + \pi_1$*, we obtain the values of* p_{00} *and* p_{10} *from the equations, and we further obtain* $p_{01} = 1 - p_{00}$ *and* $p_{11} = 1 - p_{10}$ *from the unity property.*

Remark 59. *Regarding the computation of transition matrix from copula functions, as an alternative to the parametric copula function, where the negative and positive dependence is indicated by the parameters, the smaller or greater copula, with respect to independence, based on the supermodular order is an additional approach. Specifically, the supermodular ordering of copulas is elaborated in [141].*

3.3.2.1 Applying to Wireless Channel

We apply the dependence control theory to the wireless channel, where the fading represents the inherent characteristic of the propagation environment and is an uncontrollable parameter, and the power represents the free dimension of the wireless system and is a controllable parameter for dependence manipulation. Particularly, we model the fading $h(t)$ as an independent process, the power $P(t)$ as a Markov process, and the accumulation of the wireless channel capacity as a Markov additive process $(S(t), P(t))$, where $S(t) = \sum_{i=0}^{t} c(t)$ and $c(t) = W \log_2 \left(1 + \frac{P(t)}{N_0 W} |h(t)|^2\right)$, i.e., the evolution of the capacity is modulated by the random power. As regards the dependence manipulation, we calculate the transition matrix of the Markov process using the Fréchet copula and Gaussian copula, of which the dependence parameter is an indicator of the dependence strength ranging from negative to positive.

We provide a simulation example of the dependence manipulation in wireless channel. We specify the fading as the uncontrollable parameter, which is an independent process, and the power as the controllable parameter, which is a Markov process. We manipulate the power into bearing respectively negative dependence and positive dependence for two experiments. We study the time series and the lag-1 series of an arbitrary sample path of the functional processes, i.e., the instantaneous capacity $c(t)$, the cumulative capacity $S(t)$, and the transient capacity $\overline{c}(t)$. We calculate the correlation coefficient and the probability value [140]. It is reasonable to consider the correlation coefficient, which requires the existence of the variance, because the capacity is light-tailed. The correlation coefficient of the series X and Y is defined as

$$corr = \frac{\sum_i (X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_i (X_i - \overline{X})^2}\sqrt{\sum_i (Y_i - \overline{Y})^2}}, \tag{3.104}$$

where $\overline{X}$ and $\overline{Y}$ are respectively the sample means of X and Y. The probability value is a measure of the evidence against the null hypothesis that the time series is uncorrelated. The null hypothesis test uses the Student's t-distribution. If the probability value $pval < 0.05$, it rejects the null hypothesis, otherwise, it shows the evidence against the alternative hypothesis that the time series is correlated.

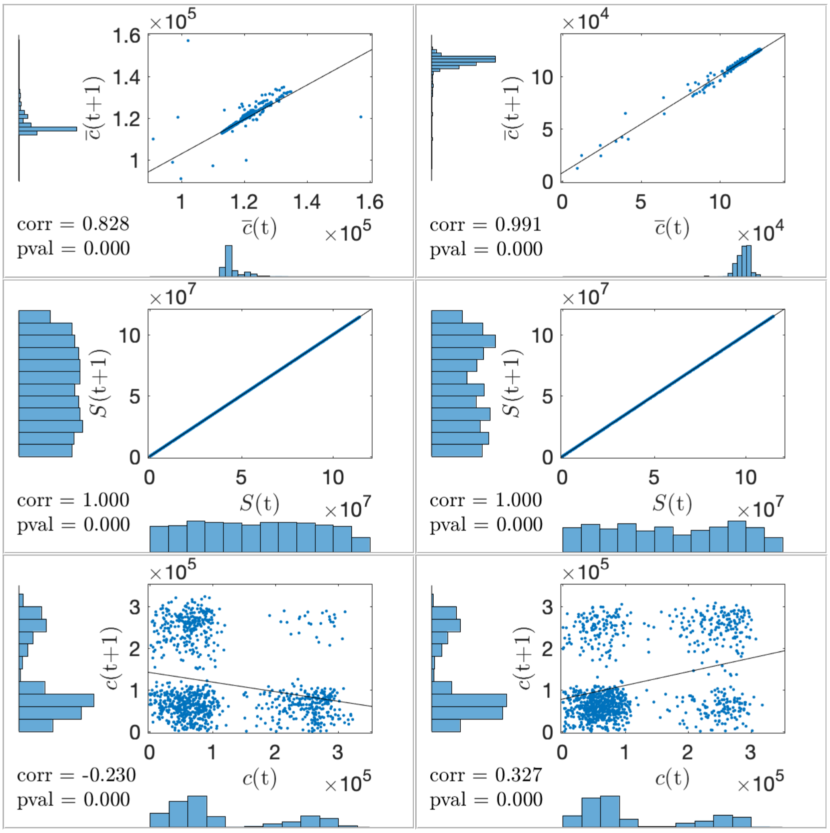


Figure 3.1: Wireless channel capacity of Rayleigh channel. The uncontrollable parameter is the fading process with one state and the controllable parameter is the power that is a Markov process with two states. The Markov process is time homogeneous without Granger causality. The dependence structure is given by Gaussian copula with correlation matrix $\mathbf{\Sigma} = [1\ 0\ \ -0.5\ 0; 0\ 1\ 0\ 0; -0.5\ 0\ 1\ 0; 0\ 0\ 0\ 1]$ as negative dependence (left column), $\mathbf{\Sigma} = [1\ 0\ 0.5\ 0; 0\ 1\ 0\ 0; 0.5\ 0\ 1\ 0; 0\ 0\ 0\ 1]$ as positive dependence (right column), initial distribution $\boldsymbol{\varpi} = [0.3\ 0.7]$, stationary distribution $\boldsymbol{\pi} = [0.3\ 0.7]$. $W = 20$kHz and $P/N = [10^4\ 10^4; 10\ 10]$. 1000 time slots. The correlation coefficient and probability value between the time series and lag-1 series are provided.

The simulation results are shown in Fig. 3.1. For the instantaneous capacity $c(t)$, the result shows that the times series exhibits respectively negative dependence that is shown in the left column with negative correlation coefficient, and positive dependence that is shown in the right column with positive correlation coefficient. The probability values are both zero, which rejects the null hypothesis that no correlation exists. It is evident that the left column has more density in the left-upper and right-lower clusters than the right column, while the right column has more density in the left-lower and right-upper clusters than the left column. This phenomena is coherent with the intuition that the large values of a random variable tends to occur together small values of other random variables for negative dependence, while large values or small values tend to occur together for positive dependence. It is interesting to notice that the marginal distribution of the instantaneous capacity is bimodal, which results from the mixture of two distribution functions. Since the instantaneous capacity is non-negative, the cumulative capacity $S(t)$ exhibits extremely positive dependence, no matter the negative or positive dependence in the instantaneous capacity, and the influence of the dependence in the instantaneous capacity is manifested in the transient capacity $\overline{c}(t)$. The simulation results validate the dependence manipulation analysis.

We regard the wireless channel as a queueing system for the information transmission of arrivals and the wireless channel capacity as the service process. In order to study the impact of the dependence in the wireless channels capacity, we consider a constant arrival process. According to Corollary 6, the wireless channel attains the best performance for constant arrival process in terms of the asymptotic decay identities. Thus, the constant arrival is fit for investigating the ultimate quality of service that the wireless channel can provide. Moreover, it indicates that the ultimate wireless channel performance is solely determined by the statistical properties of the wireless channel regardless of the arrivals, in terms of some measure identities. Specifically, the performance analysis results with respect to the constant arrival process are as follows.

Corollary 7. *Consider a constant arrival process $A(t) = \lambda t$ and Markov additive service process $(J_t, S(t))$ with state space E' and initial distribution $\boldsymbol{\varpi}$. The delay conditional on the initial state $J_0 = i$ is bounded by*

$$\frac{e^{-\theta A(1)} \cdot h_{J_0}(-\theta)}{\max_{j \in E'} h_j(-\theta)} \cdot e^{-\theta\lambda d} \leqslant \mathbb{P}_i(D > d) \leqslant \frac{h_{J_0}(-\theta)}{\min_{j \in E'} h_j(-\theta)} \cdot e^{-\theta\lambda d}, \tag{3.105}$$

where $A(1)$ is the amount of arrival in one unit time, $-\theta$ is the negative root of $\kappa(\theta) = 0$ of the Markov additive process $(J_t, S(t) - \lambda t)$ and $\boldsymbol{h}(-\theta)$ is the corresponding right eigenvector.

Proof. The proof of this special case follows that of Theorem 22. □

Remark 60. *For a constant arrival process $A(t) = \lambda t$ and independent and identically distributed service process $c(t) \stackrel{d}{=} C$, the delay is bounded by [119, p. 257]*

$$C_- e^{-\theta\lambda d} \leqslant P(D > d) \leqslant C_+ e^{-\theta\lambda d}, \tag{3.106}$$

where

$$C_- = \inf_{x \in [0, x_0)} \frac{\overline{B}(x)}{\int_x^\infty e^{\theta(y-x)} B(dy)}, \tag{3.107}$$

$$C_+ = \sup_{x \in [0, x_0)} \frac{\overline{B}(x)}{\int_x^\infty e^{\theta(y-x)} B(dy)}, \tag{3.108}$$

and B is the distribution of $\lambda - C$ and $x_0 = \sup\{x : B(x) < 1\}$.

Remark 61. *For a queueing system with constant arrival process with arrival rate λ, the relationship between delay and backlog is expressed as $\mathbb{P}(B > b) = \mathbb{P}(D > b/\lambda)$, and the asymptotic tail decay rate of delay γ_D equals that of backlog γ_B multiplied by the arrival rate λ, i.e., $\gamma_D = \lambda \cdot \gamma_B$.*

We provide the numerical results of the tail bounds of the performance measure delay, which are illustrated in Fig. 3.2. We fix the noise power $N = WN_0$, and randomize the signal power P in SNR $= P/N$. The results indicate that the dependence resource is exchangeable for other resources, e.g., the power resource. It is shown, by manipulating the dependence and marginals through the power process, the wireless channel attains a better performance with a smaller power mean, equivalently a smaller capacity, compared to the independent power case. Moreover, it shows the increasing convex order is a sufficient but not necessary condition, i.e., the order of the mean values is not a necessary condition for the order of the asymptotic decay rates. Considering the performance gain of the non-stationary Markov additive process over the stationary additive process, it indicates that the forms of distribution functions are exchangeable for the mean values.

The asymptotic decay rates with respect to the dependence parameters of the Fréchet copula and Gaussian copula are shown in Fig. 3.3. The results show that the Markov additive process bearing negative dependence, which results in a smoother fluctuation of the measure identities than the positive dependence scenario, is less sensitive to the marginal variations, i.e., while a stronger marginal can have a positive effect in the positive dependence scenario, a weaker marginal does not necessarily have a negative effect in the negative dependence scenario. However, the impact of the marginals counts, because the asymptotic decay rate is a joint effect of the marginal distributions and the dependence structure of the arrival and service processes, and the supermodular ordering of the arrival process or service process is a sufficient but not necessary condition. Particularly, it is shown, for the positive dependence, the upper envelope and the lower envelope are both decreasing functions, especially, the lower envelope continues the decreasing trend of the negative dependence scenario. Since the Markov additive process is a non-stationary process, with respect to

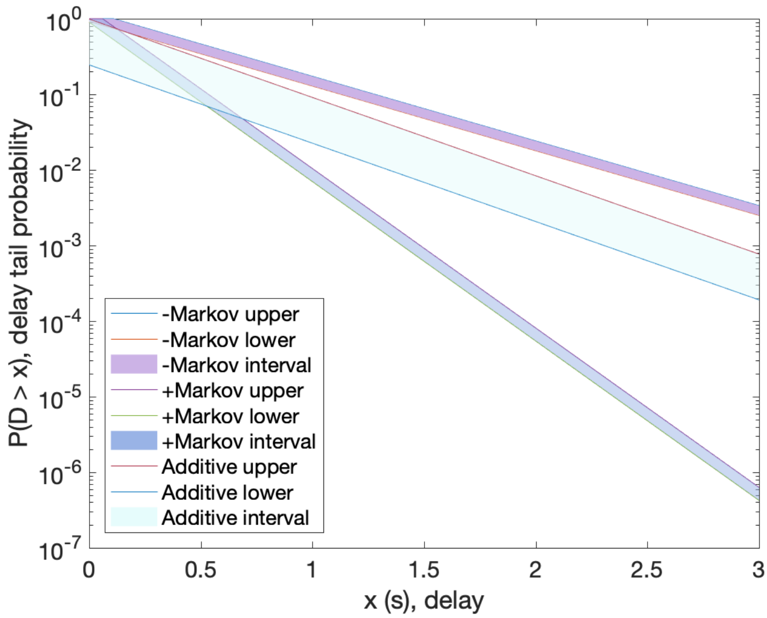
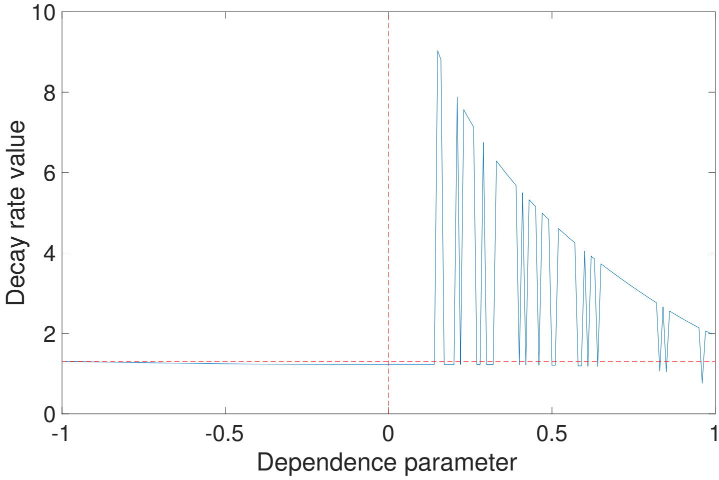
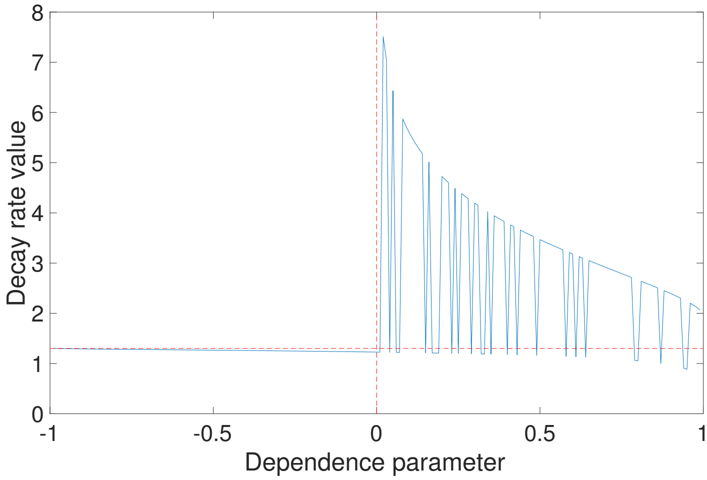


Figure 3.2: Delay tail distribution of Rayleigh channel. "-" and "+" depict respectively negative and positive dependence in capacity, the lines depict the double-sided bounds with the intervals depicted as the shaded areas. λ = 10kbits, W = 20kHz, $SNR = e^{0.5}$ for the independence case of additive capacity process, $\mathbf{SNR} = [e^{0.5}\ e^{0.5}; 0.7e^{0.5}\ 0.7e^{0.5}]$, $\mathbf{P} = [0.4125\ 0.5875; 0.2518\ 0.7482]$ calculated from Fréchet copula with $\alpha = 0.5$ for $\lambda - c(t)$ indicating positive dependence in capacity and $\mathbf{P} = [0.2875\ 0.7125; 0.3054\ 0.6946]$ calculated from Fréchet copula with $\alpha = -0.5$ for $\lambda - c(t)$ indicating negative dependence in capacity, for the dependence case of Markov additive capacity process with initial distribution $\boldsymbol{\varpi} = [0.3\ 0.7]$ and stationary distribution $\boldsymbol{\pi} = [0.3\ 0.7]$. Note when the initial and stationary distributions are $\boldsymbol{\varpi} = \boldsymbol{\pi} = [0.5\ 0.5]$, we get two interesting matrices, i.e., $\mathbf{P} = [0.5625\ 0.4375; 0.4375\ 0.5625]$ for positive dependence and $\mathbf{P} = [0.4375\ 0.5625; 0.5625\ 0.4375]$ for negative dependence in capacity, with a similar tail phenomena.



(a) Fréchet copula.



(b) Gaussian copula.

Figure 3.3: Decay trend of measure identities of Rayleigh channel with Markov additive capacity. $\lambda = 1$bit, $W = 1$Hz, $\mathbf{SNR} = [3\ 3; 1.5\ 1.5]$, with initial distribution $\boldsymbol{\varpi} = [0.5\ 0.5]$ and stationary distribution $\boldsymbol{\pi} = [0.5\ 0.5]$.

different dependence parameters, it can have different marginals such that they have a stronger impact than the dependence structure, which explains the discontinuity of the trend function, on the other hand, the continuity of the trend function results from that if the difference between the marginals with respect to different dependence parameters is small, then the dependence structure has the dominant effect rather than the marginals. In addition, the numerical results indicate that if the marginals are identical then the extremely negative dependence results in the optimal asymptotic decay rate, otherwise, the asymptotic decay rate is optimized at the limit from the weakly positive dependence to the independence.

3.3.3 Manipulation at Large

We provide the dependence manipulation techniques for both the spatial dependence and temporal dependence of a stochastic process. The manipulation of the spatial dependence means the dependence manipulation of the random parameters of the stochastic process at some time epochs, while the manipulation of the temporal dependence means the manipulation of some random parameters on the time line.

3.3.3.1 Identical Marginals

We present a sufficient condition that the composition of a supermodular function with some multivariate functions is a supermodular function.

Lemma 4. *Let* $f_t : \mathbb{R}^n \to \mathbb{R}$, $\forall t \geqslant 1$. *If* $g : \mathbb{R}^t \to \mathbb{R}$ *is supermodular,* $g' := g(f_1, \dots, f_t) : \mathbb{R}^{t\times n} \to \mathbb{R}$, *and* $g((f_1(\boldsymbol{x}_1), \dots, f_t(\boldsymbol{x}_t)) \wedge (f_1(\boldsymbol{y}_1), \dots, f_t(\boldsymbol{y}_t))) + g((f_1(\boldsymbol{x}_1), \dots, f_t(\boldsymbol{x}_t)) \vee (f_1(\boldsymbol{y}_1), \dots, f_t(\boldsymbol{y}_t))) \leqslant g'((\boldsymbol{x}_1, \dots, \boldsymbol{x}_t) \wedge (\boldsymbol{y}_1, \dots, \boldsymbol{y}_t)) + g'((\boldsymbol{x}_1, \dots, \boldsymbol{x}_t) \vee (\boldsymbol{y}_1, \dots, \boldsymbol{y}_t))$, *then* g' *is supermodular.*

Proof. Considering

$$
\begin{aligned}
g((f_1(\boldsymbol{x}_1), \ldots, f_t(\boldsymbol{x}_t)) \wedge (f_1(\boldsymbol{y}_1), \ldots, f_t(\boldsymbol{y}_t))) \\
+ g((f_1(\boldsymbol{x}_1), \ldots, f_t(\boldsymbol{x}_t)) \vee (f_1(\boldsymbol{y}_1), \ldots, f_t(\boldsymbol{y}_t))) \\
\geqslant g'(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t) + g'(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t),
\end{aligned}
$$

the proof follows directly. $\square$

Remark 62. *The implicit function theorem [130][120] gives a sufficient condition on the functions for the existence of their inverse in general.*

We investigate the scenario, where the multidimensional process is temporally independent and spatially dependent.

Theorem 31. *Assume the random parameters are spatially dependent and temporally independent. Assume $f_t : \mathbb{R}^n \to \mathbb{R}$,*

$$f_t(\boldsymbol{x}_t) \diamond f_t(\boldsymbol{y}_t) = f_t(\boldsymbol{x}_t \hat{\diamond} \boldsymbol{y}_t), \ \forall \boldsymbol{x}_t, \boldsymbol{y}_t \in \mathbb{R}^n, \ \forall t \geqslant 0, \tag{3.109}$$

where $\diamond, \hat{\diamond} \in \{\wedge, \vee\}$ preserves one of the following three relations for all $t \geqslant 0$: $\diamond = \hat{\diamond}$, $\{\diamond = \wedge | \hat{\diamond} = \vee\}$, and $\{\diamond = \vee | \hat{\diamond} = \wedge\}$.

If, for any $1 \leqslant j \leqslant t$,

$$\left(X_j^1, X_j^2, \ldots, X_j^n\right) \leqslant_{sm} \left(\widetilde{X}_j^1, \widetilde{X}_j^2, \ldots, \widetilde{X}_j^n\right), \tag{3.110}$$

and $\left(X_j^1, \ldots, X_j^n\right)$ and $\left(\widetilde{X}_j^1, \ldots, \widetilde{X}_j^n\right)$, and $\left(X_k^1, \ldots, X_k^n\right)$ are independent for all $j \neq k$, then

$$(X_1, X_2, \ldots, X_t) \leqslant_{sm} \left(X_1, X_2, \ldots, \widetilde{X}_j, \ldots, X_t\right), \tag{3.111}$$

where $X_i = f_i\left(X_i^1, \ldots, X_i^n\right)$, $\forall 1 \leqslant i \leqslant t$, and $\widetilde{X}_j = f_j\left(\widetilde{X}_j^1, \ldots, \widetilde{X}_j^n\right)$, $1 \leqslant j \leqslant t$.

If

$$\left(X_j^1, X_j^2, \ldots, X_j^n\right) \leqslant_{sm} \left(\widetilde{X}_j^1, \widetilde{X}_j^2, \ldots, \widetilde{X}_j^n\right), \ \forall 1 \leqslant j \leqslant t, \tag{3.112}$$

and $\left(X_j^1, \ldots, X_j^n\right)$ *and* $\left(X_k^1, \ldots, X_k^n\right)$ *are independent for all* $j \neq k$*, so are* $\left(\widetilde{X}_j^1, \ldots, \widetilde{X}_j^n\right)$ *and* $\left(\widetilde{X}_k^1, \ldots, \widetilde{X}_k^n\right)$*, then,* $\forall 1 \leqslant k \leqslant j \leqslant t$*,*

$$\left(\widetilde{X}_1, \ldots, \widetilde{X}_k, X_{k+1}, \ldots, X_t\right) \leqslant_{sm} \left(\widetilde{X}_1, \ldots, \widetilde{X}_j, X_{j+1}, \ldots, X_t\right), \tag{3.113}$$

where $X_j = f_j\left(X_j^1, \ldots, X_j^n\right)$ *and* $\widetilde{X}_j = f_j\left(\widetilde{X}_j^1, \ldots, \widetilde{X}_j^n\right)$, $\forall 1 \leqslant j \leqslant t$*.*

Proof. Considering the temporal independence assumption and the conjunction property of supermodular order [127], we have

$$\begin{aligned}&\left(X_1^1, \ldots, X_1^n, \ldots, X_t^1, \ldots, X_t^n\right) \\ &\qquad \leqslant_{sm} \left(X_1^1, \ldots, X_1^n, \ldots, \widetilde{X}_j^1, \ldots, \widetilde{X}_j^n, \ldots, X_t^1, \ldots, X_t^n\right).\end{aligned}$$

Letting $g : \mathbb{R}^t \to \mathbb{R}$ be supermodular and denote $g' := g(f_1, \ldots, f_t) : \mathbb{R}^{t\times n} \to \mathbb{R}$, we have g' is supermodular, which follows Lemma 4. Thus, it directly implies

$$\begin{aligned}&\left(f_1\left(X_1^1, \ldots, X_1^n\right), \ldots, f_t\left(X_t^1, \ldots, X_t^n\right)\right) \\ &\quad \leqslant_{sm} \left(f_1\left(X_1^1, \ldots, X_1^n\right), \ldots, f_j\left(\widetilde{X}_j^1, \ldots, \widetilde{X}_j^n\right), \ldots, f_t\left(X_t^1, \ldots, X_t^n\right)\right).\end{aligned}$$

The proof of the other result follows the reflexivity and transitivity property of supermodular order [98]. □

Remark 63. *The results indicate that the spatial dependence of the random parameters also influences the dependence of the stochastic process and more manipulations of the spatial dependence has more strength to transform the dependence of the stochastic process.*

Remark 64. *It is interesting to investigate the relationship between the requirement of the functional and the spatial dependence of the random parameters. An example is the comonotonicity dependence structure with identical marginal distribution, i.e., the random parameters are equal almost surely, thus the requirement of the function reduces to the scenario of the requirement of the univariate functional scenario, i.e., decreasing or increasing for each variate on the function domain.*

We present a result without specification on the spatial and temporal dependence. Note the relaxation of the specification on dependence is replaced by the additional conditions on the functionals.

Theorem 32. *Assume $f_t : \mathbb{R}^n \to \mathbb{R}$ and $f_t\big(\boldsymbol{X}_t^i | \boldsymbol{Z}_t^i = \boldsymbol{z}_t^i\big)$ are all increasing or all decreasing at each component of $\boldsymbol{X}^i = \big(X_1^i, \ldots, X_t^i\big)$, for any $\boldsymbol{z}_t^i = \big(x_t^1, \ldots, x_t^{i-1}, x_t^{i+1}, \ldots, x_t^n\big)$ in support of $\boldsymbol{Z}_t^i = \big(X_t^1, \ldots, X_t^{i-1}, X_t^{i+1}, \ldots, X_t^n\big)$, $\forall 1 \leqslant i \leqslant n$, $\forall t \geqslant 1$. Denote $\boldsymbol{z}^i = \big\{\boldsymbol{z}_1^i, \ldots, \boldsymbol{z}_t^i\big\}$, $\boldsymbol{Z}^i = \big\{\boldsymbol{Z}_1^i, \ldots, \boldsymbol{Z}_t^i\big\}$, and $\big(\boldsymbol{X}^i | \boldsymbol{Z}^i\big) \leqslant_{sm} \Big(\widetilde{\boldsymbol{X}}^i | \boldsymbol{Z}^i\Big) \iff \big(\boldsymbol{X}^i | \boldsymbol{Z}^i = \boldsymbol{z}^i\big) \leqslant_{sm} \Big(\widetilde{\boldsymbol{X}}^i | \boldsymbol{Z}^i = \boldsymbol{z}^i\Big)$, $\forall \boldsymbol{z}^i \in \boldsymbol{Z}^i$.*

If, for any $1 \leqslant i \leqslant n$,

$$\big(X_1^i, X_2^i, \ldots, X_t^i | \boldsymbol{Z}^i\big) \leqslant_{sm} \Big(\widetilde{X}_1^i, \widetilde{X}_2^i, \ldots, \widetilde{X}_t^i | \boldsymbol{Z}^i\Big), \tag{3.114}$$

then

$$(X_1, X_2, \ldots, X_t) \leqslant_{sm} \Big(\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_t\Big), \tag{3.115}$$

where $\widetilde{X}_j = f_j(X_j^1, \ldots, X_j^{i-1}, \widetilde{X}_j^i, X_j^{i+1}, \ldots, X_j^n)$, $\forall 1 \leqslant j \leqslant t$.

If, $\forall 1 \leqslant j \leqslant i$,

$$\Big(X_1^j, X_2^j, \ldots, X_t^j | \boldsymbol{Z}^j\Big) \leqslant_{sm} \Big(\widetilde{X}_1^j, \widetilde{X}_2^j, \ldots, \widetilde{X}_t^j | \boldsymbol{Z}^j\Big), \tag{3.116}$$

then

$$\widetilde{\boldsymbol{X}}_t^k \leqslant_{sm} \widetilde{\boldsymbol{X}}_t^j, \ \forall 0 \leqslant k \leqslant j \leqslant i, \tag{3.117}$$

with $\widetilde{X}^l_{t_m} = f_m(\widetilde{X}^1_m, \dots, \widetilde{X}^l_m, X^{l+1}_m, \dots, X^n_m)$, $1 \leqslant m \leqslant t$, *and* $\widetilde{\boldsymbol{X}}^l_t = (\widetilde{X}^l_{t_1}, \dots, \widetilde{X}^l_{t_t})$, $l \in \{k, j\}$.

Proof. The proof follows analogically the proof of the independence scenario, by using the conditional probability. □

Remark 65. *The orders* $(\boldsymbol{X}|\boldsymbol{Z} = \boldsymbol{z}) \leqslant_{sm} (\boldsymbol{Y}|\boldsymbol{Z} = \boldsymbol{z})$ *and* $\mathbb{E}(\boldsymbol{X}|\boldsymbol{Z}) \leqslant_{sm} \mathbb{E}(\boldsymbol{Y}|\boldsymbol{Z})$ *correspond to the conditional supermodular order in the sense of the uniform conditional ordering [142][121]. The conditional formulation influences the stochastic ordering of the probability measures, moreover, it influences the property of the functions of the random variables, e.g., the monotonicity.*

Remark 66. *There is an implicit condition that the spatial dependence must not influence the temporal dependence ordering, or the temporal dependence ordering is conditional on the spatial dependence. Specifically, if spatial independence is assumed, the conditional probability disappears. On the other hand, it is interesting to investigate what type of spatial dependence sufficiently imply the conditional ordering.*

Remark 67. *It is interesting to investigate the conditional probability and conditional stochastic order expression of the spatial dependence manipulation scenario, e.g., Theorem 31.*

Remark 68. *As an example of conditional probability, for the function of two random variables* $f(X, Y)$*, the independence assumption implies* $\mathbb{E}[f(X, Y)] = \mathbb{E}_X \mathbb{E}_Y [f(X, Y)]$*, while the absence of independence implies that* $\mathbb{E}[f(X, Y)] = \mathbb{E}_X \mathbb{E}_{Y|X=x}[f(X, Y)|X = x]$*.*

Remark 69. *As an example of conditional monotonicity of deterministic functions, letting* $f(x, y) = x^y$*,* $x > 0$*, then* $f(x, y|y > 0)$ *is increasing at* x*,* $f(x, y|y < 0)$ *is decreasing at* x*, and* $f(x, y|y = 0) = 1$ *is constant.*

Remark 70. *It is interesting to extend the results in Theorem 32 and the corresponding results without conditional probability to the increasing supermodular order* $\leqslant_{ism}$ *and the symmetric supermodular order* $\leqslant_{symsm}$.

3.3.3.2 Different Marginals

We present a result about the manipulation of stochastic process based on the marginals.

Theorem 33. *Assume the random parameters are spatially independent and temporally dependent.*

If, for any $1 \leqslant i \leqslant n$, $\left(X_1^i, X_2^i, \ldots, X_t^i\right) \leqslant_{dcx} \left(\widetilde{X}_1^i, \widetilde{X}_2^i, \ldots, \widetilde{X}_t^i\right)$, *and* $f_j\left(X_j^1 \ldots, X_j^n\right)$, $\forall 1 \leqslant j \leqslant t$, *are increasing and affine functions, then*

$$(X_1, X_2, \ldots, X_t) \leqslant_{dcx} \left(\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_t\right), \tag{3.118}$$

where $\widetilde{X}_j = f_j(X_j^1, \ldots, X_j^{i-1}, \widetilde{X}_j^i, X_j^{i+1}, \ldots, X_j^n)$, $\forall 1 \leqslant j \leqslant t$, *and*

$$\sum_{j=1}^{t} \alpha_j X_j \leqslant_{cx} \sum_{j=1}^{t} \alpha_j \widetilde{X}_j. \tag{3.119}$$

where $\alpha_j \in \mathbb{R}_{\geqslant 0}$, $\forall 1 \leqslant j \leqslant t$.

If, for all $1 \leqslant i \leqslant n$, $\left(X_1^i, X_2^i, \ldots, X_t^i\right) \leqslant_{dcx} \left(\widetilde{X}_1^i, \widetilde{X}_2^i, \ldots, \widetilde{X}_t^i\right)$, *and* $f_j\left(X_j^1 \ldots, X_j^n\right)$, $\forall 1 \leqslant j \leqslant t$, *are increasing and affine functions, then*

$$\widetilde{\boldsymbol{X}}_t^k \leqslant_{dcx} \widetilde{\boldsymbol{X}}_t^{k'}, \ \forall 0 \leqslant k \leqslant k' \leqslant n, \tag{3.120}$$

with $\widetilde{X}_{t_m}^l = f_m(\widetilde{X}_m^1, \ldots, \widetilde{X}_m^l, X_m^{l+1}, \ldots, X_m^n)$, $1 \leqslant m \leqslant t$, *and* $\widetilde{\boldsymbol{X}}_t^l = (\widetilde{X}_{t_1}^l, \ldots, \widetilde{X}_{t_t}^l)$, $l \in \{k, k'\}$, *and*

$$\sum_{j=1}^{t} \alpha_j X_j^k \leqslant_{cx} \sum_{j=1}^{t} \alpha_j \widetilde{X}_j^{k'}. \tag{3.121}$$

where $\alpha_j \in \mathbb{R}_{\geqslant 0}$, $\forall 1 \leqslant j \leqslant t$.

Proof. We first prove that the composition of a directionally convex function $g : \mathbb{R}^t \to \mathbb{R}$ with the coordinatewise increasing and affine functions $f_j : \mathbb{R} \to \mathbb{R}$, $\forall 1 \leqslant j \leqslant t$, is a directionally convex function $g(f_1, \ldots, f_t)$. Considering $\boldsymbol{x}_i \in \mathbb{R}^t$, $i = 1, 2, 3, 4$, with $\boldsymbol{x}_1 \leqslant \boldsymbol{x}_2 \leqslant \boldsymbol{x}_4$, $\boldsymbol{x}_1 \leqslant \boldsymbol{x}_3 \leqslant \boldsymbol{x}_4$, and $\boldsymbol{x}_1 + \boldsymbol{x}_4 = \boldsymbol{x}_2 + \boldsymbol{x}_3$, since f_j, $\forall 1 \leqslant j \leqslant t$, are increasing and affine, we have $f^*(\boldsymbol{x}_1) \leqslant f^*(\boldsymbol{x}_2) \leqslant f^*(\boldsymbol{x}_4)$, $f^*(\boldsymbol{x}_1) \leqslant f^*(\boldsymbol{x}_3) \leqslant f^*(\boldsymbol{x}_4)$, and $f^*(\boldsymbol{x}_1) + f^*(\boldsymbol{x}_4) = f^*(\boldsymbol{x}_2) + f^*(\boldsymbol{x}_3)$, where $f^*(\boldsymbol{x}_i) = \big(f_1(x_i^1), \ldots, f_t(x_i^t)\big)$, $\boldsymbol{x}_i = \big(x_i^1, \ldots, x_i^t\big)$, $i = 1, 2, 3, 4$. Thus, $g(f_1, \ldots, f_t)(\boldsymbol{x}_1) + g(f_1, \ldots, f_t)(\boldsymbol{x}_4) \leqslant g(f_1, \ldots, f_t)(\boldsymbol{x}_2) + g(f_1, \ldots, f_t)(\boldsymbol{x}_3)$, which means that $g(f_1, \ldots, f_t)$ is directionally convex [97].

Without loss of generality, we consider the first variate. Since the functional $\{f_j\}_{1 \leqslant j \leqslant t}$ is increasing and affine, we obtain

$$\begin{aligned}&\big(f_1\big(X_1^1, x_1^2, \ldots, x_1^n\big), \ldots, f_t\big(X_t^1, x_t^2, \ldots, x_t^n\big)\big)\\ &\qquad\qquad \leqslant_{dcx} \left(f_1\left(\widetilde{X}_1^1, x_1^2, \ldots, x_1^n\right), \ldots, f_t\left(\widetilde{X}_t^1, x_t^2, \ldots, x_t^n\right)\right),\end{aligned}$$

$\forall \left(x_j^2, \ldots, x_j^n\right) \in \left(X_j^2, \ldots, X_j^n\right)$, $\forall 1 \leqslant j \leqslant t$. By taking the expectation, we obtain $(X_1, X_2, \ldots, X_t) \leqslant_{dcx} \left(\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_t\right)$, which further implies the convex order of the weighted sum [97].

The rest of the results follows the transitivity of the directionally convex order. $\square$

Remark 71. *The results show that the manipulation of the marginal distributions of one dimension is able to transform the distribution ordering properties of the overall stochastic process, the more dimensions the more manipulation strength.*

Remark 72. *To implement the stochastic ordering* $(X_1^i, \ldots, X_t^i) \leqslant_{dcx} (\widetilde{X}_1^i, \ldots, \widetilde{X}_t^i)$, *it is sufficient to consider that [97][98]* $\big(X_1^i, X_2^i, \ldots, X_t^i\big)$

and $\left(\widetilde{X}_1^i, \widetilde{X}_2^i, \ldots, \widetilde{X}_t^i\right)$ *have the common conditionally increasing copula* $C_t^i\big(u_1^i, \ldots, u_t^i\big)$, *and* $X_j^i \leqslant_{cx} \widetilde{X}_j^i$, $\forall 1 \leqslant j \leqslant t$. *Similar result [15] holds for the increasing directionally convex order* $\leqslant_{idcx}$.

Remark 73. *The marginal distribution manipulation is feasible only for positive dependence, while the dependence structure manipulation has no dependence bias. In other words, the marginal distribution manipulation with respect to negative dependence should be avoided in practice.*

The following theorems provide results of more general functionals with respect to the increasing directionally convex order.

Theorem 34. *Assume the random parameters are spatially independent and temporally dependent. If* $f_j\left(X_j^1 \ldots, X_j^n\right)$, $\forall 1 \leqslant j \leqslant t$, *are increasing and directionally convex, and* $\left(X_1^i, X_2^i, \ldots, X_t^i\right) \leqslant_{idcx} \left(\widetilde{X}_1^i, \widetilde{X}_2^i, \ldots, \widetilde{X}_t^i\right)$, $\forall 1 \leqslant i \leqslant n$, *the results in Theorem 33 extend to the increasing directionally convex order* $\leqslant_{idcx}$ *of the random vectors and the corresponding increasing convex order* $\leqslant_{icx}$ *of the weighted partial sums.*

Proof. The function $g(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \ldots, f_t(\boldsymbol{x}))$, $g : \mathbb{R}^{n\times t} \to \mathbb{R}^t$, is increasing and directionally convex if the functions $f_j : \mathbb{R}^{n\times t} \to \mathbb{R}$, $\forall 1 \leqslant j \leqslant t$, are increasing and directionally convex [127]. The composition of an increasing and directionally convex function $g : \mathbb{R}^t \to \mathbb{R}$ and an increasing and directionally convex function $f : \mathbb{R}^{n\times t} \to \mathbb{R}^t$ is an increasing and directionally convex function $g \circ f$ [127]. The proof follows that the increasing directionally convex order is closed under conjunction [127]. $\square$

Theorem 35. *Assume the random parameters are spatially dependent and temporally independent. If* $f_j\left(X_j^1 \ldots, X_j^n\right)$, $\forall 1 \leqslant j \leqslant t$, *are increasing and directionally convex, and* $\left(X_j^1, X_j^2, \ldots, X_j^n\right) \leqslant_{idcx} \left(\widetilde{X}_j^1, \widetilde{X}_j^2, \ldots, \widetilde{X}_j^n\right)$, $\forall 1 \leqslant j \leqslant t$, *then the same results hold as in Theorem 31, but in the sense of the increasing and directionally convex order* $\leqslant_{idcx}$.

Proof. Note the composition of an increasing and convex function $g : \mathbb{R} \to \mathbb{R}$ and an increasing and directionally convex function $f : \mathbb{R}^n \to \mathbb{R}$ is an increasing and directionally convex function $g \circ f$ [127]. Then, the proof follows that the increasing and convex order of each elements implies the increasing and directionally convex order of the random vector with independent elements. $\square$

Remark 74. *It is interesting to extend the temporal and spatial manipulation results to the conditional (increasing) directionally convex order.*

Assuming independence among the random vectors, we present the directionally convex order result for random sums, which are not necessarily independent.

Theorem 36. *Let $\boldsymbol{X}_j = (X_{j,1}, \ldots, X_{j,m})$ and $\mathbf{Y}_j = (Y_{j,1}, \ldots, Y_{j,m})$, $j = 1, 2, \ldots$, be two sequences of non-negative random vectors with independence among components, and let $\boldsymbol{M} = (M_1, M_2, \ldots, M_m)$ and $\boldsymbol{N} = (N_1, N_2, \ldots, N_m)$ be two vectors of non-negative integer-valued random variables. Assume that both $\boldsymbol{M}$ and $\boldsymbol{N}$ are independent of the $\boldsymbol{X}_j$'s and $\mathbf{Y}_j$'s. Assume that $X_{j,i} \leqslant_{cx} X_{j+1,i}$, $\forall 1 \leqslant i \leqslant m$, $\forall j \geqslant 1$.*

If $\boldsymbol{M} \leqslant_{dcx} \boldsymbol{N}$, then

$$\left(\sum_{j=1}^{M_1} X_{j,1}, \ldots, \sum_{j=1}^{M_m} X_{j,m} \right) \leqslant_{dcx} \left(\sum_{j=1}^{N_1} X_{j,1}, \ldots, \sum_{j=1}^{N_m} X_{j,m} \right). \tag{3.122}$$

If $\boldsymbol{X}_j \leqslant_{dcx} \mathbf{Y}_j$, $\forall j$, then

$$\left(\sum_{j=1}^{N_1} X_{j,1}, \ldots, \sum_{j=1}^{N_m} X_{j,m} \right) \leqslant_{dcx} \left(\sum_{j=1}^{N_1} Y_{j,1}, \ldots, \sum_{j=1}^{N_m} Y_{j,m} \right). \tag{3.123}$$

If $\boldsymbol{M} \leqslant_{dcx} \boldsymbol{N}$ *and* $\boldsymbol{X}_j \leqslant_{dcx} \boldsymbol{Y}_j$, $\forall j$, *then*

$$\left(\sum_{j=1}^{M_1} X_{j,1}, \ldots, \sum_{j=1}^{M_m} X_{j,m}\right) \leqslant_{dcx} \left(\sum_{j=1}^{N_1} Y_{j,1}, \ldots, \sum_{j=1}^{N_m} Y_{j,m}\right). \tag{3.124}$$

Proof. The first result is available in [112][127]. For the second result, $\boldsymbol{X}_j \leqslant_{dcx} \boldsymbol{Y}_j \implies X_{j,i} \leqslant_{cx} Y_{j,i}$, $\forall 1 \leqslant i \leqslant m$, the independence assumption implies that the convex order is closed under convolutions [127], i.e., $\sum_{j=1}^{n_i} X_{j,i} \leqslant_{cx} \sum_{j=1}^{n_i} Y_{j,i}$, $\forall 1 \leqslant i \leqslant m$, furthermore, it implies

$$\begin{aligned} &\mathbb{E}\left[\phi\left(\sum_{j=1}^{N_1} X_{j,1}, \ldots, \sum_{j=1}^{N_m} X_{j,m}\right) |\boldsymbol{N} = (n_1, \ldots, n_m)\right] \\ &\qquad \leqslant \mathbb{E}\left[\phi\left(\sum_{j=1}^{N_1} Y_{j,1}, \ldots, \sum_{j=1}^{N_m} Y_{j,m}\right) |\boldsymbol{N} = (n_1, \ldots, n_m)\right], \end{aligned}$$

where ϕ is directionally convex. By integrating for expectation, we obtain the final result. The third result follows the transitivity of the directionally convex order. □

Corollary 8. *With proper revisions, the results in Theorem 36 extend to the increasing directionally convex order* $\leqslant_{idcx}$ *and (increasing) componentwise convex order* $(\leqslant_{iccx}) \leqslant_{ccx}$. *Specifically, the corresponding revisions are* $X_{j,i} \leqslant_{icx} (\leqslant_{cx}, \leqslant_{icx}) X_{j+1,i}$, $\boldsymbol{M} \leqslant_{idcx} (\leqslant_{ccx}, \leqslant_{iccx}) \boldsymbol{N}$, *and* $\boldsymbol{X}_j \leqslant_{idcx} (\leqslant_{ccx}, \leqslant_{iccx}) \boldsymbol{Y}_j$.

Proof. The proof follows the properties of each stochastic orders and preliminary results in [112][127]. □

Remark 75. *It is interesting to notice the fact that: Let* $\boldsymbol{X} = (X_1, \ldots, X_m)$ *be a set of independent random variables and let* $\boldsymbol{Y} = (Y_1, \ldots, Y_m)$ *be another set of independent random variables, then,* $\boldsymbol{X} \leqslant_{dcx} (\leqslant_{idcx}) \boldsymbol{Y} \iff X_i \leqslant_{cx} (\leqslant_{icx}) Y_i$, $\forall 1 \leqslant i \leqslant m \iff \boldsymbol{X} \leqslant_{ccx} (\leqslant_{iccx}) \boldsymbol{Y}$.

We study the ordering property of the partial sums under the ordering condition of the sequences.

Theorem 37. *If* $(X_{t_1}, \ldots, X_{t_k}) \leqslant_{idcx} \left(\widetilde{X}_{t_1}, \ldots, \widetilde{X}_{t_k}\right)$, $\forall t_1, \ldots, t_k \in \mathbb{N}$, $\forall k \in \mathbb{N}$, *then we have*

$$\left(\sum_{j_1 \in \mathcal{T}_1} X_{j_1}, \ldots, \sum_{j_k \in \mathcal{T}_k} X_{j_k}\right) \leqslant_{idcx} \left(\sum_{j_1 \in \mathcal{T}_1} \widetilde{X}_{j_1}, \ldots, \sum_{j_k \in \mathcal{T}_k} \widetilde{X}_{j_k}\right), \tag{3.125}$$

for any disjoint subsets $\mathcal{T}_1, \ldots, \mathcal{T}_k \in \mathbb{N}$.

Proof. The proof follows that, if $f : \mathbb{R}^m \to \mathbb{R}^k$ is increasing and directionally convex and $g : \mathbb{R}^n \to \mathbb{R}^m$ is increasing and directionally convex, then the composition $f \circ g$ is increasing and directionally convex [127]. □

Remark 76. *An alternative approach is to treat the functional stochastic process as a random field on* $\mathbb{N}^n \times \mathbb{R}$, *then the comparison result directly follows the comparison result of random field in [94][127].*

Remark 77. *The result indicates that the* $\leqslant_{idcx}$ *ordering of the instantaneous values implies the* $\leqslant_{idcx}$ *ordering of the accumulated values.*

Remark 78. *It is interesting to study the corresponding property of the supermodular order or the counter examples.*

Since the usual stochastic order has a direct indication on the mean values, i.e., $\boldsymbol{X} \leqslant_{st} \boldsymbol{Y} \implies \mathbb{E}\boldsymbol{X} \leqslant \mathbb{E}\boldsymbol{Y} \implies \sum \mathbb{E}X_i \leqslant \sum \mathbb{E}Y_i$, $X_i \in \boldsymbol{X}$, $Y_i \in \boldsymbol{Y}$, it is interesting to consider the dependence manipulation with respect to the usual stochastic order when the mean value is the objective measure.

Theorem 38. *Assume the random parameters are spatially independent and temporally dependent. If* $f_j\left(X_j^1 \ldots, X_j^n\right)$, $\forall 1 \leqslant j \leqslant t$, *are increasing, and* $\left(X_1^i, X_2^i, \ldots, X_t^i\right) \leqslant_{st} \left(\widetilde{X}_1^i, \widetilde{X}_2^i, \ldots, \widetilde{X}_t^i\right)$, $\forall 1 \leqslant i \leqslant n$, *then*

$$(X_1, X_2, \ldots, X_t) \leqslant_{st} \left(\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_t\right), \tag{3.126}$$

where $\widetilde{X}_j = f_j(X_j^1, \ldots, X_j^{i-1}, \widetilde{X}_j^i, X_j^{i+1}, \ldots, X_j^n)$, $\forall 1 \leqslant j \leqslant t$, *for any* $1 \leqslant i \leqslant n$; *and*

$$\widetilde{\boldsymbol{X}}_t^k \leqslant_{st} \widetilde{\boldsymbol{X}}_t^{k'}, \ \forall 0 \leqslant k \leqslant k' \leqslant n, \tag{3.127}$$

where $\widetilde{X}_{t_m}^l = f_m(\widetilde{X}_m^1, \ldots, \widetilde{X}_m^l, X_m^{l+1}, \ldots, X_m^n)$, $1 \leqslant m \leqslant t$, *and* $\widetilde{\boldsymbol{X}}_t^l = (\widetilde{X}_{t_1}^l, \ldots, \widetilde{X}_{t_t}^l)$, $l \in \{k, k'\}$.

Proof. The spatial independence implies the conjunction property of the usual stochastic order $\left(\boldsymbol{X}^1, \ldots, \boldsymbol{X}^i, \ldots, \boldsymbol{X}^n\right) \leqslant_{st} \left(\boldsymbol{X}^1, \ldots, \widetilde{\boldsymbol{X}}^i, \ldots, \boldsymbol{X}^n\right)$, where $\boldsymbol{X}^i = \left(X_1^i, X_2^i, \ldots, X_t^i\right)$, then the first result directly follows the closure property of the usual stochastic order [127]. The second result follows the transitivity of the usual stochastic order. □

Remark 79. *Particularly, if* $\left(X_1^i, X_2^i, \ldots, X_t^i\right)$ *and* $\left(\widetilde{X}_1^i, \widetilde{X}_2^i, \ldots, \widetilde{X}_t^i\right)$ *have the common copula* $C_t^i\left(u_1^i, \ldots, u_t^i\right)$, *the order condition* $\left(X_1^i, X_2^i, \ldots, X_t^i\right) \leqslant_{st} \left(\widetilde{X}_1^i, \widetilde{X}_2^i, \ldots, \widetilde{X}_t^i\right)$ *can be replaced by* $X_j^i \leqslant_{st} \widetilde{X}_j^i$, $\forall 1 \leqslant j \leqslant t$. *This result is available in [127, p. 272].*

Corollary 9. *Assume the random parameters are spatially dependent and temporally independent. If* $f_j\left(X_j^1 \ldots, X_j^n\right)$, $\forall 1 \leqslant j \leqslant t$, *are increasing, and* $\left(X_j^1, X_j^2, \ldots, X_j^n\right) \leqslant_{st} \left(\widetilde{X}_j^1, \widetilde{X}_j^2, \ldots, \widetilde{X}_j^n\right)$, $\forall 1 \leqslant j \leqslant t$, *then the same results hold as in Theorem 31, but in the sense of the usual stochastic order* $\leqslant_{st}$.

Proof. The results follow the closure property and transitivity of the usual stochastic order [127]. □

Remark 80. *It is interesting to extend the results to the scenario without spatial and temporal dependence specification and express the results in terms of the conditional probability and conditional stochastic order as in Theorem 32.*

We have the following results of the random sums with respect to the usual stochastic order.

Remark 81. *Let* $\boldsymbol{X}_j = (X_{j,1}, \ldots, X_{j,m})$ *and* $\mathbf{Y}_j = (Y_{j,1}, \ldots, Y_{j,m})$, $j = 1, 2, \ldots$, *be two sequences of non-negative random vectors, and let* $\boldsymbol{M} = (M_1, M_2, \ldots, M_m)$ *and* $\boldsymbol{N} = (N_1, N_2, \ldots, N_m)$ *be two vectors of non-negative integer-valued random variables. Assume that both* $\boldsymbol{M}$ *and* $\boldsymbol{N}$ *are independent of the* $\boldsymbol{X}_j$*'s and* $\mathbf{Y}_j$*'s.*

If $\{\boldsymbol{X}_j, j \in \mathbb{N}\} \leqslant_{st} \{\mathbf{Y}_j, j \in \mathbb{N}\}$ *and* $\boldsymbol{M} \leqslant_{st} \boldsymbol{N}$*, then*

$$\left(\sum_{j=1}^{M_1} X_{j,1}, \ldots, \sum_{j=1}^{M_m} X_{j,m}\right) \leqslant_{st} \left(\sum_{j=1}^{N_1} Y_{j,1}, \ldots, \sum_{j=1}^{N_m} Y_{j,m}\right). \tag{3.128}$$

This result is available in [127]. Specifically, the proof follows the transitivity of the following results,

$$\left(\sum_{j=1}^{M_1} X_{j,1}, \ldots, \sum_{j=1}^{M_m} X_{j,m}\right) \leqslant_{st} \left(\sum_{j=1}^{N_1} X_{j,1}, \ldots, \sum_{j=1}^{N_m} X_{j,m}\right), \tag{3.129}$$

which is provided in [112], and

$$\left(\sum_{j=1}^{N_1} X_{j,1}, \ldots, \sum_{j=1}^{N_m} X_{j,m}\right) \leqslant_{st} \left(\sum_{j=1}^{N_1} Y_{j,1}, \ldots, \sum_{j=1}^{N_m} Y_{j,m}\right), \tag{3.130}$$

which follows the closure property of the usual stochastic order, by conditioning on $(N_1, \ldots, N_m) = (n_1, \ldots, n_m)$ *and integrating for expectation.*

Chapter 4

Conclusion

This work advocates the research on dependence control, which concerns transforming the dependence structure of a stochastic process in the system through dependence manipulation to improve the system performance. Specifically, we develop a dependence control theory and formulate three principles of dependence control, namely measurability, duality, and transformability. In the development of the theory, we adopt various mathematical techniques, like large deviation, change of measure, stochastic order, martingale, and copula. In addition, several assumptions are made, which allow to characterize weak forms of dependence and light-tailed process, and an example is the Markov additive process. We remark that, among the three principles, these assumptions are necessary only for the first. For the second principle, it relies on the assumptions when the dual potency of arrival and service dependence implies the ordering of the measure identities, and for the third principle, the results do not require these assumptions provided that the stochastic orders make sense.

We apply the dependence control theory to the wireless channels. We find that the light-tail behavior is an intrinsic property of the wireless channel capacity, which is due to the passive nature of the wireless propagation environment and the power limitation in the practical systems. It is interesting to note that the marginal manipulation has a dependence bias, i.e., it is effective with respect to positive dependence rather than negative dependence. Considering the trade-off between the dependence and the marginal distributions, it is reasonable that a process with positive dependence has a better effect on system performance than a process with negative dependence and vice versa. This property indicates that the dependence is a tradable resource in the physical world and it sheds light on the development of new wireless technologies to trade off the dependence resource and other resources, e.g., exchanging dependence for transmission power. While the focus of the application is on dependence control in wireless channel capacity, many of the obtained results hold for general stochastic processes and queueing systems.

An Outlook on Future Research

We believe that dependence control is a new direction for research. On the one hand, we hope this work paves the way for the development of the dependence control theory, on the other hand, we hope the theoretical results can stimulate the development of new mathematical techniques and new engineering technologies. Specifically, the potential research directions are outlined as follows.

1. It is interesting to study the tail property of the wireless channel capacity in the wireless network scenarios, e.g., the multi-user scenario of multiple access channel and the broadcast channel. Moreover, it is interesting to take into account the tail behavior of the noise process.

2. It is interesting to study the extreme strength of dependence transform, e.g., whether there are some strong forms of dependence that are unable to be transformed from strong to weak, or one form of dependence that

can not be transformed from one form to another form. It is interesting to study the mixing property of the Markov additive process when the dependence of the Markov chain goes from positive dependence or negative dependence to independence.

3. It is interesting to define a probability metric to quantify the dependence advantage, e.g., the distance between average power of different dependence schemes or the distance between the power distributions in the wireless channel capacity, or the distance between the asymptotic decay rates of the performance measures of a queueing sytem. While the stochastic orders provide comparison results in a qualitative way [98], it is interesting to investigate the quantitative properties regarding the comparison results, e.g., the probability metrics [36]. It is interesting to investigate the dependence advantage in different wireless channel situations or scenarios, e.g., of different signal-to-noise ratios.

4. It is interesting to consider the dependence control principles in the queueing network, where the dynamics are expressed as a high-order Lindley equation [75], which is an extension of the simple Lindley equation of the single-server queue. In addition, it is interesting to consider the scenario of a multi-server queue, where the dynamics are expressed as one type of recursive equations [79][3]. It is interesting to notice the potential duality among the multi-class arrivals [143], parallel queues [86], and multi-component shock models [87].

5. It is interesting to study new ideas of dependence control, e.g., new perspectives on dependence and new methods to transform the dependence. Specifically, it is interesting to study the additional perspectives on the measure identity, diverse manipulation techniques to transform the dependence structure, and more application scenarios. For example, it is interesting to study the left tail of the performance measures besides the right tail and it is interesting to extend the three principles of dependence

control to other forms of dependence and heavy-tailed processes.

6. It is interesting to apply the dependence control theory to simulation, e.g., for generating statistically independent random variables and for variance reduction. Specifically, the simulation provides a controllable environment, which eases the difficulty of manipulating the dependence of the stochastic process.

7. It is interesting to apply the dependence control to different physical systems, e.g., the real-time system that has strict delay requirement, the data center services, and the future Internet architecture. It is interesting to apply the dependence control concept to different research fields, e.g., economics, finance, and insurance. For instance, an analogical result is to improve the ruin probability by controlling the dependence in the risk process. In addition, the application of dependence control to reliability theory is also of interest.

8. It is interesting to develop new technologies that implement the dependence control theory as proof-of-principle of the mathematical analysis in this work. For example, by analogy with power allocation in wireless communication, it is interesting to define new dependence measures and study the mechanism of dependence allocation. It is interesting to study the possibility to encode information into dependence and decode the information from dependence. It is interesting to study the dependence management and dependence engineering.

9. It is interesting to extend the mathematics of the dependence control theory to other probability foundations, e.g., from the concrete probability space to the abstract probability space, from commutative probability to non-commutative probability, and from quantitative probability to qualitative probability. The extension requires the definitions of new concepts and relations, which are the opportunities to create new mathematics.

Appendices

Appendix A

Probability Preliminaries

A.1 Basics

A probability space is a triplet $(\Omega, \mathscr{F}, \mathbb{P})$, where $(\Omega, \mathscr{F})$ is a measurable space and $\mathbb{P}$ is a probability measure. A filtration $\{\mathscr{F}_t\}_{t\in\mathbb{T}}$, $\mathbb{T} = \mathbb{N}$ or $\mathbb{T} = \mathbb{R}_{\geqslant 0}$, is an increasing and right-continuous family of sub-σ-field of $\mathscr{F}$, i.e., $\mathscr{F}_s \subset \mathscr{F}_t$ for $s \leqslant t$ and $\mathscr{F}_t = \bigcap_{s>t} \mathscr{F}_s$.

A stochastic basis $(\Omega, \mathscr{F}, \{\mathscr{F}_t\}_{t\in\mathbb{T}}, \mathbb{P})$, which is also called a filtered probability space, is a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ equipped with a filtration $\{\mathscr{F}_t\}_{t\in\mathbb{T}}$.

A random time $\tau \leqslant \infty$ is a stopping time with respect to the filtration $\{\mathscr{F}_t\}_{t\in\mathbb{T}}$ if $\{\tau \leqslant t\} \in \mathscr{F}_t$ for all $t \in \mathbb{T}$.

More about the stochastic basis and stopping time are available in [66][6].

A.2 Transforms

For a random variable X with distribution function $F(x)$, the moment generating function is defined as $\widehat{F}[\theta] = \mathbb{E}\big[e^{\theta X}\big] = \int e^{\theta x} F(dx)$ and the cumulant generating function is defined as $\kappa(\theta) = \log \widehat{F}[\theta]$.

The kth derivative of the moment generating function or cumulant generating function at 0 is the kth moment or cumulant of the distribution. Specifically, the first cumulant is the mean, the second cumulant is the variance, and the third cumulant is the central third moment [6].

A.3 Time Reversal

Consider a stochastic process X_t. The reversed process X_t^* with respect to time τ is defined as

$$X_t^* = X_{\tau-t}. \tag{A.1}$$

If the reversed process X_t^* and the original process X_t are statistically indistinguishable, i.e.,

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{\tau-t_1}, X_{\tau-t_2}, \ldots, X_{\tau-t_n}) \text{ for all } t_1, t_2, \ldots, t_n \text{ and } n \tag{A.2}$$

the process X_t is said to be time reversible with respect to time τ.

A.4 Stochastic Order

The stochastic orders are special cases of the partial orders [98].

Definition 1. *A binary relation $\leqslant_X$ on an arbitrary set X is called a partial order if it satisfies the following three properties.*

1. *Reflexivity: $x \leqslant_X x$ for any $x \in X$;*
2. *Transitivity: if $x, y, z \in X$ are such that $x \leqslant_X y$ and $y \leqslant_X z$, then $x \leqslant_X z$;*
3. *Antisymmetry: if $x, y \in X$ are such that $x \leqslant_X y$ and $y \leqslant_X x$, then $x = y$.*

A partially ordered set is a set X together with a partial order $\leqslant_X$ on X.

We introduce the univariate usual stochastic order, the convex order, and the multivariate usual stochastic order.

Definition 2. *The random variable X is said to be smaller than the random variable Y with respect to the usual stochastic order $\leqslant_{st}$, if*

$$\mathbb{P}(X > x) \leqslant \mathbb{P}(Y > x), \ \forall x \in (-\infty, \infty), \tag{A.3}$$

or equivalently,

$$\mathbb{E}[\phi(X)] \leqslant \mathbb{E}[\phi(Y)] \text{ for all increasing functions } \phi : \mathbb{R} \to \mathbb{R}, \tag{A.4}$$

provided the expectations exist.

Definition 3. *The random variable X is said to be smaller than the random variable Y with respect to the convex order $\leqslant_{cx}$, if*

$$\mathbb{E}[\phi(X)] \leqslant \mathbb{E}[\phi(Y)] \text{ for all convex functions } \phi : \mathbb{R} \to \mathbb{R}, \tag{A.5}$$

provided the expectations exist.

Remark 82. *Similarly, the increasing convex order $\leqslant_{icx}$, the concave order $\leqslant_{cv}$, and the increasing concave order $\leqslant_{icv}$ are defined in terms of the increasing convex functions, the concave functions, and the increasing concave functions.*

Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$. We denote $\boldsymbol{x} \leqslant \boldsymbol{y}$ if $x_i \leqslant y_i$ for $i = 1, 2, \ldots, n$. We say that the function $\phi : \mathbb{R}^n \to \mathbb{R}$ is increasing (decreasing), if $\phi(\boldsymbol{x}) \leqslant (\geqslant) \phi(y)$ whenever $\boldsymbol{x} \leqslant \boldsymbol{y}$.

Definition 4. *The random vector $\boldsymbol{X}$ is said to be smaller than the random vector $\boldsymbol{Y}$ with respect to the usual stochastic order $\leqslant_{st}$, if*

$$\mathbb{E}[\phi(\boldsymbol{X})] \leqslant \mathbb{E}[\phi(\boldsymbol{Y})] \text{ for all increasing functions } \phi : \mathbb{R}^n \to \mathbb{R}, \tag{A.6}$$

provided the expectations exist.

Appendix B

Elements of Dependence

B.1 Dependence Order

The dependence orders are reviewed. There are many concepts of dependence corresponding to different dependence orders [98][123].

For a function, $f : \mathbb{R}^n \to \mathbb{R}$, $\varepsilon > 0$, the difference operator is defined as [98][123]

$$\Delta_i^{\varepsilon} f(\boldsymbol{x}) = f(\boldsymbol{x} + \varepsilon \boldsymbol{e}_i) - f(\boldsymbol{x}),\ i \in \{1, \ldots, n\}, \tag{B.1}$$

where $\boldsymbol{e}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ denotes the i-th unit vector.

Let $f : \mathbb{R}^n \to \mathbb{R}$. f is Δ-monotone [98][123], if for every subset $\{i_1, \ldots, i_k\} \in \{1, \ldots, n\}$ and $\varepsilon_1, \ldots, \varepsilon_k > 0$,

$$\Delta_{i_1}^{\varepsilon_1} \ldots \Delta_{i_k}^{\varepsilon_k} f(\boldsymbol{x}) \geqslant 0,\ \forall \boldsymbol{x}. \tag{B.2}$$

Corresponding to the Δ-monotone functions $\mathfrak{F}_{\Delta}$, the $\underline{\Delta}$-monotone functions $\mathfrak{F}_{\underline{\Delta}}$ is defined as [123] $\mathfrak{F}_{\underline{\Delta}} = \{h : \mathbb{R}^n \to \mathbb{R};\ \exists f \in \mathfrak{F}_{\Delta},\ h(\boldsymbol{x}) = f(-\boldsymbol{x})\}$. The functions $h \in \mathfrak{F}_{\underline{\Delta}}$ are decreasing and satisfy $(-1)^k \Delta_{i_1}^{\varepsilon_1} \ldots \Delta_{i_k}^{\varepsilon_k} h(\boldsymbol{x}) \geqslant 0,\ \forall \boldsymbol{x}$, for every subset $\{i_1, \ldots, i_k\} \in \{1, \ldots, n\}$ and $\varepsilon_1, \ldots, \varepsilon_k > 0$. In brief, $f(\boldsymbol{x}) \in \mathfrak{F}_{\underline{\Delta}}$ if and only if $-f(-\boldsymbol{x}) \in \mathfrak{F}_{\Delta}$.

The function $f : \mathbb{R}^n \to \mathbb{R}$ is supermodular [98][123], if for all $1 \leqslant i < j \leqslant n$, $\varepsilon, \delta > 0$,

$$\Delta_i^{\varepsilon} \Delta_j^{\delta} f(\boldsymbol{x}) \geqslant 0, \ \forall \boldsymbol{x}. \tag{B.3}$$

Equivalently, the function is supermodular, if

$$f(\boldsymbol{x} \wedge \boldsymbol{y}) + f(\boldsymbol{x} \vee \boldsymbol{y}) \geqslant f(\boldsymbol{x}) + f(\boldsymbol{y}), \ \forall \boldsymbol{x}, \boldsymbol{y}, \tag{B.4}$$

where $\boldsymbol{x} \wedge \boldsymbol{y} = (\min\{x_1, y_1\}, \ldots, \min\{x_n, y_n\})$ and $\boldsymbol{x} \vee \boldsymbol{y} = (\max\{x_1, y_1\}, \ldots, \max\{x_n, y_n\})$.

If the function $\log f(\boldsymbol{x})$ is supermodular, i.e., $\log f(\boldsymbol{x}) \in \mathfrak{F}_{sm}$, then the function $f(\boldsymbol{x})$ is multivariate totally positive of order 2 [123], i.e., $f(\boldsymbol{x}) \in \mathfrak{F}_{MTP_2}$.

If $\Delta_i^{\varepsilon} \Delta_j^{\delta} f(\boldsymbol{x}) \geqslant 0$ holds for all $i \leqslant j$, then f is directionally convex [98][123]. A function is directionally convex if it is supermosular and componentwise convex [98].

Definition 5. *The stochastic orders generated by the supermodular function, the increasing supermodular function, the symmetric supermodular function, the directionally convex function, and the increasing directionally convex function are respectively defined as the supermodular order $\leqslant_{sm}$, the increasing supermodular order $\leqslant_{ism}$, the symmetric supermodular order $\leqslant_{symsm}$, the increasing directional convex order $\leqslant_{dcx}$, and the increasing directional convex order $\leqslant_{idcx}$.*

Remark 83. *Since the supermodular order is invariant to both componentwise increasing transforms and componentwise decreasing transforms [98], we have $\boldsymbol{X} \leqslant_{sm} \boldsymbol{Y} \iff -\boldsymbol{X} \leqslant_{sm} -\boldsymbol{Y}$. Thus, we have the chain rule, $\boldsymbol{X} \leqslant_{sm} \boldsymbol{Y} \iff -\boldsymbol{X} \leqslant_{sm} -\boldsymbol{Y} \implies \sum X \leqslant_{cx} \sum Y \iff -\sum X \leqslant_{cx} -\sum Y$.*

The classical dependence ordering is based on the orthant ordering [123]. For two random vectors $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^n$, the upper orthant order is defined as $\boldsymbol{X} \leqslant_{uo} \boldsymbol{Y}$, if $\overline{F}_{\boldsymbol{X}}(\boldsymbol{x}) \leqslant \overline{F}_{\boldsymbol{Y}}(\boldsymbol{x})$, $\forall \boldsymbol{x} \in \mathbb{R}^n$; the lower orthant order is defined as $\boldsymbol{X} \leqslant_{lo} \boldsymbol{Y}$, if $F_{\boldsymbol{X}}(\boldsymbol{x}) \leqslant F_{\boldsymbol{Y}}(\boldsymbol{x})$, $\forall \boldsymbol{x} \in \mathbb{R}^n$; and the concordance order is defined as $\boldsymbol{X} \leqslant_c \boldsymbol{Y}$, if $\boldsymbol{X} \leqslant_{uo} \boldsymbol{Y}$ and $\boldsymbol{X} \leqslant_{lo} \boldsymbol{Y}$. Particularly, we have [98], if $\boldsymbol{X} \leqslant_{st} \boldsymbol{Y}$, then $\boldsymbol{X} \geqslant_{lo} \boldsymbol{Y}$ and $\boldsymbol{X} \leqslant_{uo} \boldsymbol{Y}$. Note $\leqslant_{lo}$ and $\leqslant_{uo}$ are denoted as the lower concordance order and upper concordance order in [98]. We denote the stochastic order $\leqslant_{\Delta}$ and $\leqslant_{\underline{\Delta}}$, which are respectively generated by the Δ-monotone functions and $\underline{\Delta}$-monotone functions. Then, the upper orthant order is equivalently expressed as [123]

$$\boldsymbol{X} \leqslant_{uo} \boldsymbol{Y} \iff \boldsymbol{X} \leqslant_{\Delta} \boldsymbol{Y}, \tag{B.5}$$

and the lower orthant order is equivalently expressed as [123]

$$\boldsymbol{X} \leqslant_{lo} \boldsymbol{Y} \iff \boldsymbol{X} \leqslant_{\underline{\Delta}} \boldsymbol{Y}. \tag{B.6}$$

In addition, the positive orthant order and negative orthant order are defined by comparing the probability measure with the probability measure of independence [98].

Since the multivariate distribution function and survival function are supermodular functions, for $\boldsymbol{X}, \boldsymbol{Y} \in \mathbb{R}^n$, we have the follows relation [98][123]

$$\boldsymbol{X} \leqslant_{sm} \boldsymbol{Y} \implies \boldsymbol{X} \leqslant_c \boldsymbol{Y}, \ \forall n \geqslant 2, \tag{B.7}$$

particularly, $\boldsymbol{X} \leqslant_{sm} \boldsymbol{Y} \iff \boldsymbol{X} \leqslant_c \boldsymbol{Y}$ and $\boldsymbol{X} \leqslant_{uo} \boldsymbol{Y} \iff \boldsymbol{X} \leqslant_{lo} \boldsymbol{Y} \iff \boldsymbol{X} \leqslant_c \boldsymbol{Y}$, for $n = 2$. It indicates that the comparison based on the supermodular order is stronger than the comparision based on the orthant order [123].

Remark 84. *The relationships between the supermodular order, (weak) association, and (weakly) conditional increasing are elaborated in [98][26][123]. Another reference is [81]. In case of positive dependence, the dependence concepts ranging from strong to weak are as follows [123]: multivariate* *totally positive of order* 2*, conditional increasing, conditional increasing in sequence, associated, weak associated, weak associated in sequence, and positive supermodular dependent. In case of negative dependence, the strong to weak ranging is as follows [98][26][63]: negatively associated, negatively supermodular dependent, negative concordance, and negatively lower and upper orthant dependence.*

Remark 85. *The directionally convex order, compared to the convex order, is more suitable for the comparison of random vectors with a common copula [97], moreover, the convex order of the marginals can not imply the convex order of the partial sum, if the components are negatively dependent [98]. Thus, a common conditionally increasing copula, which indicates positive dependence, is assumed for the comparison of the partial sum.*

Remark 86. *For a Polish space with a closed partial order, that the space is totally ordered is equivalent to that every probability measure on this space is associated [98, p. 124], i.e., exhibiting positive dependence.*

Remark 87. *It is interesting to find the extreme negative dependence by symmetric of the extreme positive dependence in terms of some stochastic orders, i.e., letting $\boldsymbol{X}^*$ be with the extreme positive dependence and $T(\boldsymbol{X}^*)$ be with the extreme negative dependence for some transform T, such that $T(\boldsymbol{X}^*) \leqslant_{\mathcal{F}} \boldsymbol{X} \leqslant_{\mathcal{F}} \boldsymbol{X}^*$ for any $\boldsymbol{X}$, provided the order $\leqslant_{\mathcal{F}}$ exists. Thus, to find the symmetric negative dependence, it is fundamental to find the transform that is symmetric with respect to the dependence structure and equivalent with respect to the stochastic order, which boils down to the definition of the symmetry and equivalence.*

Remark 88. *It is interesting to give an overview of the different dependence concepts and link the dependence concepts to the stochastic orders. It is interesting to investigate the the extension of the dependence control with respect to different dependence orders, which manifests the characteristics of different dependence definitions.*

B.2 Markov Process

Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in\mathbb{N}}, \mathbb{P})$ be a filtered probability space and $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$ be an adapted stochastic process. $\boldsymbol{X}$ is a Markov process if and only if

$$\mathbb{P}\left(\boldsymbol{X}_t \leqslant x | \boldsymbol{X}_{t-1}, \boldsymbol{X}_{t-2}, \ldots, \boldsymbol{X}_0\right) = \mathbb{P}\left(\boldsymbol{X}_t \leqslant x | \boldsymbol{X}_{t-1}\right). \tag{B.8}$$

A Markov additive process is defined as a bivariate Markov process $\{X_t\} = \{(J_t, S(t))\}$ where $\{J_t\}$ is a Markov process with state space E and the increments of $\{S(t)\}$ are governed by $\{J_t\}$ in the sense that [7]

$$\mathbb{E}\left[f(S(t+s) - S(t))g(J_{t+s}) | \mathscr{F}_t\right] = \mathbb{E}_{J_t,0}\left[f(S(s))g(J_s)\right]. \tag{B.9}$$

We consider the finite state space and discrete time scenario [6][7].

In discrete time, a Markov additive process is specified by the measure-valued matrix (kernel) $\mathbf{F}(dx)$ whose ijth element is the defective probability distribution

$$F_{ij}(dx) = \mathbb{P}_{i,0}(J_1 = j, Y_1 \in dx), \tag{B.10}$$

where $Y_t = S(t) - S(t-1)$. An alternative description is in terms of the transition matrix $\mathbf{P} = (p_{ij})_{i,j\in E}$, $p_{ij} = \mathbb{P}_i(J_1 = j)$, and the probability measures

$$H_{ij}(dx) = \mathbb{P}(Y_1 \in dx | J_0 = i, J_1 = j) = \frac{F_{ij}(dx)}{p_{ij}}. \tag{B.11}$$

With respect to a transition probability p_{ij}, the increment of $\{S_t\}$ has a distribution B_{ij}.

Consider the matrix $\widehat{\mathbf{F}}_t[\theta] = (\mathbb{E}_i[e^{\theta S(t)}; J_t = j])_{i,j\in E}$. In discrete time,

$$\widehat{\mathbf{F}}_t[\theta] = \widehat{\mathbf{F}}[\theta]^t, \tag{B.12}$$

where $\widehat{\mathbf{F}}[\theta] = \widehat{\mathbf{F}}_1[\theta]$ is a $E \times E$ matrix with ijth element $\widehat{F}^{(ij)}[\theta] =$

$p_{ij} \int e^{\theta x} H^{(ij)}(dx)$, and $\theta \in \Theta = \{\theta \in R : \int e^{\theta x} H^{(ij)}(dx) < \infty\}$ [6]. By Perron-Frobenius theorem, $\widehat{\mathbf{F}}[\theta]$ has a positive real eigenvalue with maximal absolute value $e^{\kappa(\theta)}$. The corresponding right and left eigenvectors are respectively $\mathbf{h}(\theta) = (h_i(\theta))_{i \in E}$ and $\mathbf{v}(\theta) = (v_i(\theta))_{i \in E}$, with the normalization in particular, $\mathbf{v}(\theta)\mathbf{h}(\theta) = 1$ and $\boldsymbol{\pi}\mathbf{h}(\theta) = 1$, where $\boldsymbol{\pi} = \mathbf{v}(0)$ is the stationary distribution and $\mathbf{h}(0) = \mathbf{1}$.

If the stationary distribution $\boldsymbol{\pi} = (\pi(j))_{j \in E}$ exists, the time reversed process, $\{(J_n^*, X_n^*)\}$, is represented by [5]

$$F^*(i, j; A) = \mathbb{P}_{\boldsymbol{\pi}}(J_0 = j, X_1 \in A | J_1 = i) = \frac{\pi(j)}{\pi(i)} F(j, i; A), \tag{B.13}$$

i.e., $\pi(i)F^*(i, j; A) = \pi(j)F(j, i; A)$. The time-reversed transition probability, when looking at $\{J_n^*\}$ alone, is $p_{ij}^* = \frac{\pi(j)}{\pi(i)} p_{ji}$, i.e., $\boldsymbol{P}^* = \boldsymbol{\Delta}_{\boldsymbol{\pi}}^{-1} \boldsymbol{P}^T \boldsymbol{\Delta}_{\boldsymbol{\pi}}$, where $\boldsymbol{\Delta}_{\boldsymbol{\pi}}$ is the diagonal matrix with $\boldsymbol{\pi}$ on the diagonal, see [6, p. 314], and that the conditional distribution of X_1^* given $J_0^* = i$ and $J_1^* = j$ is the same as the conditional distribution of X_1 given $J_0 = j$ and $J_1 = i$, i.e., $H^*(i, j; A) = H(j, i; A)$. Considering the matrix $\widehat{\mathbf{F}}^*[\theta] = \boldsymbol{\Delta}_{\boldsymbol{\pi}}^{-1} \widehat{\mathbf{F}}[\theta]^T \boldsymbol{\Delta}_{\boldsymbol{\pi}}$, the roots of $\det\left(\widehat{\mathbf{F}}[\theta] - \boldsymbol{I}\right) = 1$ and $\det\left(\widehat{\mathbf{F}}^*[\theta] - \boldsymbol{I}\right) = 1$ are the same [6, p. 331], i.e., the same eigenvalues with the Perron-Frobenius eigenvalue in particular

$$\lambda^*(\theta) = \lambda(\theta). \tag{B.14}$$

Letting $\widehat{\mathbf{F}}[\theta]\boldsymbol{h}^*(\theta) = \lambda(\theta)\boldsymbol{h}^*(\theta)$, then $\boldsymbol{h}^*(\theta)^T \boldsymbol{\Delta}_{\boldsymbol{\pi}} \widehat{\mathbf{F}}[\theta] = \lambda(\theta)\boldsymbol{h}^*(\theta)^T \boldsymbol{\Delta}_{\boldsymbol{\pi}}$. Thus, if the matrix $\widehat{\mathbf{F}}[\theta]$ has the Perron-Frobenius eigenvalue $\lambda(\theta)$ with the corresponding left eigenvector $\boldsymbol{\nu}(\theta)$, the matrix $\widehat{\mathbf{F}}^*[\theta]$ has the same Perron-Frobenius eigenvalue $\lambda(\theta)$ with the corresponding right eigenvector

$$\boldsymbol{h}^*(\theta) = \boldsymbol{\Delta}_{\boldsymbol{\pi}}^{-1} \boldsymbol{\nu}(\theta)^T. \tag{B.15}$$

Similarly, if the matrix $\widehat{\mathbf{F}}[\theta]$ has the Perron-Frobenius eigenvalue $\lambda(\theta)$ with the corresponding right eigenvector $\boldsymbol{h}(\theta)$, the matrix $\widehat{\mathbf{F}}^*[\theta]$ has the same

Perron-Frobenius eigenvalue $\lambda(\theta)$ with the corresponding left eigenvector

$$\boldsymbol{\nu}^*(\theta) = \boldsymbol{\nu}(\theta)^T \boldsymbol{\Delta}_{\boldsymbol{\pi}}. \tag{B.16}$$

Remark 89. *The fact that the time reversed process and the original process have the same Perron-Frobenius eigenvalue for a Markov additive process, which further determines the asymptotic decay rate of the performance measures in the queueing system, implies that the influence of dependence is robust to the time reversibility assumption.*

Considering a Markov additive process $(J_t, S(t))$ with the kernel $\mathbf{F}(A)$ and a linear transformation $T : \mathbb{R} \to \mathbb{R}$, letting $S^\star(t) = T(S(t))$, then $(J_, S^\star(t))$ is a Markov additive process with the kernel $\mathbf{F}^\star(B)$ such that [105]

$$F^\star_{ij}(B) = F_{ij}\big(T^{-1}(B)\big). \tag{B.17}$$

The more general formation is shown in [105]. The result is useful for studying the dual process of the queueing processes, e.g., the dual process of the service process or the arrival process. The problem lies in the specification of the process, e.g., letting the dual process be $(J_t, -S(t))$, the transition matrix remains the same as the original process. Particularly, for the kernel matrix of the moment generating functions, considering the dual process on the positive parameter axis is equivalent to considering the original process on the negative parameter axis.

Remark 90. *It is interesting to study the marginal distribution of the Markov additive process. Since the Markov additive process is not stationary in general, it can not be compared with an independently and identically distributed process based on the supermodular order. On the other hand, it is interesting to study whether or not the Markov additive processes with respectively negative and positive dependence can be compared based on the supermodular order.*

B.3 Copula

Consider a joint distribution $F(X_1, \ldots, X_n)$ with marginal distribution $F_i(X_i)$, $i = 1, \ldots, n$. Denote $u_i = F_i(X_i)$, which has a uniform distribution, then

$$F(X_1, \ldots, X_n) = F\left(F_1^{-1}(u_1), \ldots, F_n^{-1}(u_n)\right) \tag{B.18}$$

$$\equiv C(u_1, \ldots, u_n), \tag{B.19}$$

where C is a copula, with standard uniform marginals [40]. Specifically, if the marginals are continuous, the copula is unique. It is interesting to study the relationship between copula and other dependence concepts, e.g., autocorrelation, regression and association.

It is well known that the copula is invariant to the strictly increasing transforms. However, for the strictly decreasing transform, we have the following result [43]. Let $(X_1, \ldots, X_n)$ be a vector of continuous random variables with copula $C_{X_1,\ldots,X_n}$. Let $f_1, \ldots, f_n$ be strictly monotone on the range $\mathrm{Ran}X_1, \ldots, \mathrm{Ran}X_n$, respectively, and let $(f_1(X_1), \ldots, f_n(X_n))$ have copula $C_{f_1(X_1),\ldots,f_n(X_n)}$. Furthermore let f_k be strictly decreasing for some k. Without loss of generality let $k = 1$. Then [43],

$$\begin{aligned} C_{f_1(X_1),\ldots,f_n(X_n)}(u_1, \ldots, u_n) = {} & C_{f_2(X_2),\ldots,f_n(X_n)}(u_2, \ldots, u_n) \\ & - C_{X_1,f_2(X_2),\ldots,f_n(X_n)}(1 - u_1, u_2, \ldots, u_n). \end{aligned} \tag{B.20}$$

Recursively, the copula $C_{f_1(X_1),\ldots,f_n(X_n)}$ can be expressed in terms of the copula $C_{X_1,\ldots,X_n}$ and its lower-dimensional marginals [43].

Example 3. *For the 2-dimensional case, let f_1 and f_2 be strictly decreasing, then [43]*

$$C_{f_1(X_1),f_2(X_2)}(u_1, u_2) = u_1 + u_2 - 1 + C_{X_1,X_2}(1 - u_1, 1 - u_2). \tag{B.21}$$

The Markov property is solely a property of copula [32, 104]. The n-dimensional process $\mathbf{X}$ is a Markov process, if and only if, for all $t_1 < t_2 < \ldots < t_p$, the copula $C_{t_1,\ldots,t_p}$ of $(\mathbf{X}_{t_1},\ldots,\mathbf{X}_{t_p})$ satisfies [104]

$$C_{t_1,\ldots,t_p} = C_{t_1,t_2} \overset{C_{t_2}(\cdot)}{\star} C_{t_2,t_3} \overset{C_{t_3}(\cdot)}{\star} \ldots \overset{C_{t_{p-1}}(\cdot)}{\star} C_{t_{p-1},t_p}, \tag{B.22}$$

where $C_{t_i,t_{i+1}}$ is the (n + n)-dimensional copula between $\boldsymbol{X}_{t_i}$ and $\boldsymbol{X}_{t_{i+1}}$, and $C_{t_{i+1}}$ is the copula of $\boldsymbol{X}_{t_{i+1}}$. Provided that the integral exists for all $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$, the operator $\overset{C(\cdot)}{\star}$ is defined by

$$(A \overset{C(\mathbf{z})}{\star} B)(\mathbf{x},\mathbf{y}) = \int_0^{\mathbf{z}} A_{,C}(\mathbf{x},\mathbf{r}) \cdot B_{C,}(\mathbf{r},\mathbf{y})C(d\mathbf{r}), \tag{B.23}$$

where A is a $(k+n)$-dimensional copula, B is a $(n+l)$-dimensional copula, C is a n-dimensional copula, $A_{,C}(\mathbf{x},\mathbf{y})$ and $B_{C,}(\mathbf{x},\mathbf{y})$ are defined by the derivative of the copula $A(\mathbf{x},\cdot)$ and $B(\cdot,\mathbf{y})$ with respect to the copula C, i.e, $A(\mathbf{x},d\mathbf{y}) = A_{,C}(\mathbf{x},\mathbf{y})C(d\mathbf{y})$ and $B(d\mathbf{x},\mathbf{y}) = B_{C,}(\mathbf{x},\mathbf{y})C(d\mathbf{x})$. $A_{,C}$ and $B_{C,}$ are well-defined. Specifically, for 1-dimensional Markov process, the copula is expressed by [32]

$$C_{t_1,\ldots,t_p} = C_{t_1,t_2} \star C_{t_2,t_3} \star \ldots \star C_{t_{p-1},t_p}, \tag{B.24}$$

where $C_{t_1,\ldots,t_p}$ is the copula of $\left(X_{t_1},\ldots,X_{t_p}\right)$, C_{t_{k-1},t_k} is the copula of $\left(X_{t_{k-1}},X_{t_k}\right)$, and $A \star B$ is written as

$$\begin{aligned} &A \star B\,(x_1,\ldots,x_{m+n-1}) \\ &= \int_0^{x_m} A_{,\xi}(x_1,\ldots,x_{m-1},\xi)B_{\xi,}(\xi,x_{m+1},\ldots,x_{m+n-1})d\xi, \end{aligned} \tag{B.25}$$

for m-dimensional copula A and n-dimensional copula B.

Remark 91. *Considering the supermodular order of the Markov process, $\boldsymbol{X} \leqslant_{sm} \widetilde{\boldsymbol{X}}$, the question reduces to the supermodular order of the cop-*

ula of the Markov process, i.e., $C \leqslant_{sm} \widetilde{C}$. *It is interesting to study the stochastic ordering of the Markov copula. Specifically, it is interesting to study the relationship between* $C_{t_i,t_{i+1}} \leqslant_{sm} \widetilde{C}_{t_i,t_{i+1}}$, $\forall 0 < t_i < t_p < t$, *and* $C_{t_1,\ldots,t_p} \leqslant_{sm} \widetilde{C}_{t_1,\ldots,t_p}$.

The copula representation of the Markov property requires the Markov family copula [32][104][64]. Examples of Markov family copula are Gaussian copula and Fréchet copula [104]. The 2-dimensional Fréchet copula is available in [32] and the n-dimensional extension is available in [104].

Example 4. *The* n*-dimensional Gaussian copula is written as [104]*

$$C_{\mathbf{\Sigma}}(\boldsymbol{u}) = \Phi_{\mathbf{\Sigma}}\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)\right), \tag{B.26}$$

where $\Phi_{\mathbf{\Sigma}}$ *denotes the joint distribution of the* n*-dimensional standard normal distribution with linear correlation matrix* $\mathbf{\Sigma}$, *and* Φ^{-1} *denotes the inverse of the distribution function of the* 1*-dimensional standard normal distribution. The Gaussian copula allows for equal degrees of positive and negative dependence. [136].*

Remark 92. *Considering the two Gaussian random vectors,* $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$ *and* $\widetilde{\boldsymbol{X}} \sim \mathcal{N}\left(\widetilde{\boldsymbol{\mu}}, \widetilde{\mathbf{\Sigma}}\right)$, $\boldsymbol{X} \leqslant_{sm} \widetilde{\boldsymbol{X}}$ *if and only if* $\boldsymbol{X}$ *and* $\widetilde{\boldsymbol{X}}$ *have the same marginals and* $\sigma_{ij} \leqslant \widetilde{\sigma}_{ij}$ *for all* i, j *[98, p. 144]. Specifically, the Gaussian copulas are supermodular ordered if and only if the linear correlation matrices are correspondingly ordered.*

Example 5. *The extremely positive dependence, independence, and extremely negative dependence are expressed by copulas. For* 2*-dimensional copula, the extremely positive copula, product copula (independence), and extremely negative copula are defined as* $M(x,y) = \min(x,y)$, $P(x,y) = xy$, *and* $W(x,y) = \max(x+y-1, 0)$.

A convex combination of M, P, *and* W *is a Fréchet copula [32], i.e.,*

$$C_{st} = \alpha(s,t)W + (1 - \alpha(s,t) - \beta(s,t))P + \beta(s,t)M, \tag{B.27}$$

if and only if [32, 104], for $s < u < t$,

$$\alpha(s,t) = \beta(s,u)\alpha(u,t) + \alpha(s,u)\beta(u,t), \tag{B.28}$$

$$\beta(s,t) = \alpha(s,u)\alpha(u,t) + \beta(s,u)\beta(u,t), \tag{B.29}$$

where $\alpha(s,t) \geqslant 0$, $\beta(s,t) \geqslant 0$, *and* $\alpha(s,t) + \beta(s,t) \leqslant 1$. *For homogeneous case,* $\alpha(s,t) = \alpha(t-s)$ *and* $\beta(s,t) = \beta(t-s)$, *a solution is as follows*

$$\alpha(h) = e^{-2h}(1-e^{-h})/2, \tag{B.30}$$

$$\beta(h) = e^{-2h}(1+e^{-h})/2. \tag{B.31}$$

Let $\alpha = e^{-h}$, *it's a one-parameter copula [32]*

$$C_\alpha = \frac{\alpha^2(1-\alpha)}{2}W + (1-\alpha^2)P + \frac{\alpha^2(1+\alpha)}{2}M, \tag{B.32}$$

where $-1 \leqslant \alpha \leqslant 1$, *if* $|\alpha|$ *is small, independence is indicated, if* α *is near* 1, *strongly positive dependence is indicated, and if* α *is near* -1, *strongly negative dependence is indicated. The* n*-dimensional extension is available in [104].*

Bibliography

[1] R. Ahlswede and A. Winter. Strong converse for identification via quantum channels. *IEEE Transactions on Information Theory*, 48(3):569–579, 2002.

[2] H. Albrecher, S. Asmussen, and D. Kortschak. Tail asymptotics for the sum of two heavy-tailed dependent risks. *Extremes*, 9(2):107–130, 2006.

[3] D. J. Aldous, A. Bandyopadhyay, et al. A survey of max-type recursive distributional equations. *The Annals of Applied Probability*, 15(2):1047–1110, 2005.

[4] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang. What will 5g be? *IEEE Journal on selected areas in communications*, 32(6):1065–1082, 2014.

[5] S. Asmussen. Aspects of matrix wiener-hopf factorisation in applied probability. *Math. Sci*, 14(2):101–116, 1989.

[6] S. Asmussen. *Applied Probability and Queues*, volume 51. Springer Science & Business Media, 2003.

[7] S. Asmussen and H. Albrecher. *Ruin probabilities*, volume 14. World scientific, 2010.

[8] S. Asmussen and P. W. Glynn. *Stochastic simulation: algorithms and analysis*, volume 57. Springer Science & Business Media, 2007.

[9] S. Asmussen, J. L. Jensen, and L. Rojas-Nandayapa. Exponential family techniques for the lognormal left tail. *Scandinavian Journal of Statistics*, 43(3):774–787, 2016.

[10] S. Asmussen and L. Rojas-Nandayapa. Asymptotics of sums of lognormal random variables with gaussian copula. *Statistics & Probability Letters*, 78(16):2709–2714, 2008.

[11] T. Aulin. A modified model for the fading signal at a mobile radio channel. *IEEE Transactions on Vehicular Technology*, 28(3):182–203, 1979.

[12] F. Baccelli, G. Cohen, G. J. Olsder, and J.-P. Quadrat. *Synchronization and Linearity: an algebra for discrete event systems.* John Wiley & Sons, 1992.

[13] L. Baele and F. Van Holle. Stock-bond correlations, macroeconomic regimes and monetary policy. *Available at SSRN 3075816*, 2017.

[14] N. Balakrishna and C. D. Lai. Concepts of stochastic dependence. In *Continuous Bivariate Distributions*, pages 105–140. Springer, 2009.

[15] N. Balakrishnan, F. Belzunce, M. A. Sordo, and A. Suárez-Llorens. Increasing directionally convex orderings of random vectors having the same copula, and their use in comparing ordered data. *Journal of Multivariate Analysis*, 105(1):45–54, 2012.

[16] E. Bjornson, E. Jorswieck, and B. Ottersten. Impact of spatial correlation and precoding design in ostbc mimo systems. *IEEE Transactions on Wireless Communications*, 9(11):3578–3589, 2010.

[17] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[18] D. Buraczewski, E. Damek, T. Mikosch, et al. *Stochastic models with power-law tails*. Springer, 2016.

[19] K. Campbell, J. Diffley, B. Flanagan, B. Morelli, B. O'Neil, and F. Sideco. The 5g economy: How 5g technology will contribute to the global economy. *IHS Economics and IHS Technology*, 2017.

[20] D. Castelvecchi. The strange topology that is reshaping physics. *Nature News*, 547(7663):272, 2017.

[21] C.-S. Chang. *Performance guarantees in communication networks*. Springer, 2000.

[22] Y. Chen, D. Chen, and W. Gao. Extensions of breiman's theorem of product of dependent random variables with applications to ruin theory. *Communications in Mathematics and Statistics*, Jun 2018.

[23] A. Cherny. Some particular problems of martingale theory. *From Stochastic Calculus to Mathematical Finance*, pages 109–124, 2006.

[24] U. Cherubini, F. Gobbi, S. Mulinacci, and S. Romagnoli. A copula-based model for spatial and temporal dependence of equity markets. *Copula Theory and Its Applications*, pages 257–265, 2010.

[25] U. Cherubini, S. Mulinacci, and S. Romagnoli. A copula-based model of speculative price dynamics in discrete time. *Journal of Multivariate Analysis*, 102(6):1047–1063, 2011.

[26] T. C. Christofides, E. Vaggelatou, et al. A connection between supermodular ordering and positive/negative association. *Journal of Multivariate analysis*, 88(1):138–151, 2004.

[27] C.-N. Chuah, D. N. C. Tse, J. M. Kahn, and R. A. Valenzuela. Capacity scaling in mimo wireless systems under correlated fading. *IEEE Transactions on Information theory*, 48(3):637–650, 2002.

[28] E. Çinlar. Theory of continuous storage with markov additive inputs and a general release rule. *Journal of Mathematical Analysis and Applications*, 43(1):207–231, 1973.

[29] F. Ciucu. *Scaling properties in the stochastic network calculus*. PhD thesis, University of Virginia, 2007.

[30] N. Costa and S. Haykin. *Multiple-input multiple-output channel models: Theory and practice*, volume 65. John Wiley & Sons, 2010.

[31] C. C. Craig. On the frequency function of xy. *The Annals of Mathematical Statistics*, 7(1):1–15, 1936.

[32] W. F. Darsow, B. Nguyen, E. T. Olsen, et al. Copulas and markov processes. *Illinois Journal of Mathematics*, 36(4):600–642, 1992.

[33] R. A. Davis and T. Mikosch. Probabilistic properties of stochastic volatility models. In *Handbook of financial time series*, pages 255–267. Springer, 2009.

[34] V. H. DelaPeña, V. de la Peña, and E. Giné. *Decoupling: From Dependence to Independence.* Springer Science & Business Media, 1999.

[35] M. Denuit. Laplace transform ordering of actuarial quantities. *Insurance: Mathematics and Economics*, 29(1):83–102, 2001.

[36] M. Denuit, J. Dhaene, M. Goovaerts, and R. Kaas. *Actuarial theory for dependent risks: measures, orders and models.* John Wiley & Sons, 2006.

[37] J. Dhaene, M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics*, 31(1):3–33, 2002.

[38] F. Dong, K. Wu, and V. Srinivasan. Copula analysis for statistical network calculus. In *Computer Communications (INFOCOM), 2015 IEEE Conference on*, pages 1535–1543. IEEE, 2015.

[39] D. Dubhashi and D. Ranjan. Balls and bins: A study in negative dependence. *Random Structures & Algorithms*, 13(2):99–124, 1998.

[40] P. Embrechts. Copulas: A personal view. *Journal of Risk and Insurance*, 76(3):639–650, 2009.

[41] P. Embrechts, A. Höing, and A. Juri. Using copulae to bound the value-at-risk for functions of dependent risks. *Finance and Stochastics*, 7(2):145–167, 2003.

[42] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media, 1997.

[43] P. Embrechts, F. Lindskog, and A. McNeil. Modelling dependence with copulas. *Rapport technique, Département de mathématiques, Institut Fédéral de Technologie de Zurich, Zurich*, 2001.

[44] P. Embrechts and G. Puccetti. Bounds for functions of dependent risks. *Finance and Stochastics*, 10(3):341–352, 2006.

[45] M. Falk, J. Hüsler, and R.-D. Reiss. *Laws of small numbers: extremes and rare events*. Springer Science & Business Media, 2010.

[46] M. Fidler. An end-to-end probabilistic network calculus with moment generating functions. In *Quality of Service, 2006. IWQoS 2006. 14th IEEE International Workshop on*, pages 261–270. IEEE, 2006.

[47] G. J. Foschini and M. J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless personal communications*, 6(3):311–335, 1998.

[48] S. Foss, D. Korshunov, and S. Zachary. *An Introduction to Heavy-Tailed and Subexponential Distributions.* Springer Science & Business Media, 2013.

[49] U. Franz. What is stochastic independence? In *Non-Commutativity, Infinite-Dimensionality and Probability at the Crossroads*, pages 254–274. World Scientific, 2002.

[50] R. G. Gallager. *Information Theory and Reliable Communication.* John Wiley & Sons, Inc., 1968.

[51] R. G. Gallager. *Stochastic processes: theory for applications.* Cambridge University Press, 2013.

[52] J. Geluk and Q. Tang. Asymptotic tail probabilities of sums of dependent subexponential random variables. *Journal of Theoretical Probability*, 22(4):871, 2009.

[53] P. W. Glynn and W. Whitt. Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. *Journal of Applied Probability*, 31(A):131–156, 1994.

[54] I. Gohberg, S. Goldberg, and N. Krupnik. *Traces and determinants of linear operators*, volume 116. Birkhäuser, 2012.

[55] A. Goldsmith. *Wireless communications.* Cambridge university press, 2005.

[56] A. Gulisashvili, P. Tankov, et al. Tail behavior of sums and differences of log-normal random variables. *Bernoulli*, 22(1):444–493, 2016.

[57] L. J. Halliwell. Classifying the tails of loss distributions. In *Casualty Actuarial Society E-Forum, Spring 2013 Volume 2*, page 1.

[58] E. Hashorva and Z. Weng. Tail asymptotic of weibull-type risks. *Statistics*, 48(5):1155–1165, 2014.

[59] J. Hecht et al. The bandwidth bottleneck. *Nature*, 536(7615):139–142, 2016.

[60] J. L. Hein. *Discrete Structures, Logic, and Computability.* Jones & Bartlett Publishers, 2015.

[61] N. J. Higham. *Functions of matrices: theory and computation*, volume 104. Siam, 2008.

[62] B. Hogstad and M. Patzold. Exact closed-form expressions for the distribution, level-crossing rate, and average duration of fades of the capacity of mimo channels. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 455–460, April 2007.

[63] T. Hu and J. Yang. Further developments on sufficient conditions for negative dependence of random variables. *Statistics & probability letters*, 66(3):369–381, 2004.

[64] R. Ibragimov. Copula-based characterizations for higher order markov processes. *Econometric Theory*, 25(03):819–846, 2009.

[65] K. Itō. *Essentials of stochastic processes*, volume 231. American Mathematical Soc., 2006.

[66] J. Jacod and A. Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer Science & Business Media, 2003.

[67] W. C. Jakes and D. C. Cox. *Microwave mobile communications.* Wiley-IEEE Press, 1994.

[68] H. A. Jessen and T. Mikosch. Regularly varying functions. *Publications de l'Institut Mathematique*, 80(94):171–192, 2006.

[69] J. Jiang and Q. Tang. The product of two dependent random variables with regularly varying or rapidly varying tails. *Statistics & Probability Letters*, 81(8):957–961, 2011.

[70] Y. Jiang and Y. Liu. *Stochastic network calculus*. Springer, 2008.

[71] H. Joe. *Multivariate models and multivariate dependence concepts*. Chapman and Hall/CRC, 1997.

[72] E. A. Jorswieck and H. Boche. Optimal transmission strategies and impact of correlation in multiantenna systems with different types of channel state information. *IEEE Transactions on Signal Processing*, 52(12):3440–3453, 2004.

[73] E. Jouini and C. Napp. Comonotonic processes. *Insurance: Mathematics and Economics*, 32(2):255–265, 2003.

[74] I. Karatzas and S. E. Shreve. Brownian motion. In *Brownian Motion and Stochastic Calculus*, pages 47–127. Springer, 1998.

[75] F. Karpelevich, M. Y. Kelbert, and Y. M. Suhov. Higher-order lindley equations. *Stochastic Processes and their Applications*, 53(1):65–96, 1994.

[76] J. R. Klauder. Signal transmission in passive media. *IEE Proceedings-Radar, Sonar and Navigation*, 152(1):23–28, 2005.

[77] B. Ko and Q. Tang. Sums of dependent nonnegative random variables with subexponential tails. *Journal of Applied Probability*, 45(1):85–94, 2008.

[78] A. N. Kolmogorov. *Foundations of the theory of probability*. Chelsea, 1956.

[79] N. K. Krivulin. A recursive equations based representation for the g/g/m queue. *Applied Mathematics Letters*, 7(3):73–77, 1994.

[80] A. N. Lagerås. Copulas for markovian dependence. *Bernoulli*, pages 331–342, 2010.

[81] C. D. Lai and M. Xie. *Stochastic ageing and dependence for reliability*. Springer Science & Business Media, 2006.

[82] A. M. Law. *Simulation modeling and analysis*. McGraw-Hill, 2015.

[83] E. L. Lehmann. Some concepts of dependence. *The Annals of Mathematical Statistics*, pages 1137–1153, 1966.

[84] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson. On the self-similar nature of ethernet traffic. In *ACM SIGCOMM computer communication review*, volume 23, pages 183–193. ACM, 1993.

[85] R. Lenczewski. Filtered probability. *Journal of Mathematical Sciences*, 106(1):2704–2712, 2001.

[86] H. Li and S. H. Xu. On the dependence structure and bounds of correlated parallel queues and their applications to synchronized stochastic systems. *Journal of Applied Probability*, 37(4):1020–1043, 2000.

[87] H. Li and S. H. Xu. Stochastic bounds and dependence properties of survival times in a multicomponent shock model. *Journal of Multivariate Analysis*, 76(1):63–89, 2001.

[88] Y. Liu and Q. Tang. The subexponential product convolution of two weibull-type distributions. *Journal of the Australian Mathematical Society*, 89(2):277–288, 2010.

[89] A. Mammela, M. Hoyhtya, and D. P. Taylor. Normalization of a fading channel. In *IEEE Globecom 2006*, pages 1–6, Nov 2006.

[90] A. Mammela, A. Kotelba, M. Hoyhtya, and D. P. Taylor. Relationship of average transmitted and received energies in adaptive transmission. *IEEE Transactions on Vehicular Technology*, 59(3):1257–1268, 2010.

[91] A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative risk management: Concepts, techniques and tools.* Princeton university press, 2015.

[92] P.-A. Meyer. Stochastic processes from 1950 to the present. *Electronic Journal for History of Probability and Statistics*, 5(1):1–42, 2009.

[93] T. C. Mills and R. N. Markellos. *The econometric modelling of financial time series.* Cambridge University Press, 2008.

[94] N. Miyoshi. A note on bounds and monotonicity of spatial stationary cox shot noises. *Probability in the Engineering and Informational Sciences*, 18(4):561–571, 2004.

[95] A. F. Molisch. *Wireless communications*, volume 34. John Wiley & Sons, 2012.

[96] A. Müller and G. Pflug. Asymptotic ruin probabilities for risk processes with dependent increments. *Insurance: Mathematics and Economics*, 28(3):381–392, 2001.

[97] A. Müller and M. Scarsini. Stochastic comparison of random vectors with a common copula. *Mathematics of operations research*, 26(4):723–740, 2001.

[98] A. Müller and D. Stoyan. *Comparison methods for stochastic models and risks.* John Wiley & Sons Inc, 2002.

[99] N. Muraki. The five independences as natural products. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 6(03):337–371, 2003.

[100] P. Ney and E. Nummelin. Markov additive processes i. eigenvalue properties and limit theorems. *The Annals of Probability*, pages 561–592, 1987.

[101] E. C. Niehenke. Wireless communications: Present and future. *IEEE Microwave Magazine*, 15(2):26–35, 2014.

[102] B. Nistad and J. Skaar. Causality and electromagnetic properties of active media. *Physical Review E*, 78(3):036603, 2008.

[103] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab. *Signals and Systems*. Prentice Hall, 1997.

[104] L. Overbeck and W. M. Schmidt. Multivariate markov families of copulas. *Dependence Modeling*, 3(1), 2015.

[105] A. Pacheco and N. U. Prabhu. Markov-additive processes of arrivals. Technical report, Cornell University Operations Research and Industrial Engineering, 1994.

[106] A. Papoulis and S. U. Pillai. *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education, 2002.

[107] J. Parsons and A. Turkmani. Characterisation of mobile radio signals: model description. *IEE Proceedings I-Communications, Speech and Vision*, 138(6):549–556, 1991.

[108] J. D. Parsons and P. J. D. Parsons. *The mobile radio propagation channel*. Pentech press London, 1992.

[109] M. Pätzold. *Mobile radio channels*. John Wiley & Sons, 2011.

[110] A. Paulraj, R. Nabar, and D. Gore. *Introduction to space-time wireless communications.* Cambridge university press, 2003.

[111] D. Č. Pavlović, N. M. Sekulović, G. V. Milovanović, A. S. Panajotović, M. Č. Stefanović, and Z. J. Popović. Statistics for ratios of rayleigh, rician, nakagami-, and weibull distributed random variables. *Mathematical Problems in Engineering*, 2013, 2013.

[112] F. Pellerey. Stochastic comparisons for multivariate shock models. *Journal of Multivariate Analysis*, 71(1):42–55, 1999.

[113] M. Pericoli. Macroeconomics determinants of the correlation between stocks and bonds. Technical report, Bank of Italy, Economic Research and International Relations Area, 2018.

[114] G. Puccetti, R. Wang, et al. Extremal dependence concepts. *Statistical Science*, 30(4):485–517, 2015.

[115] G. Rafiq, V. Kontorovich, and M. Pa tzold. The influence of severity of fading on the statistical properties of the capacity of nakagami-m channels with mrc and egc. In *Wireless Conference (EW), 2010 European*, pages 406–410. IEEE, 2010.

[116] G. Rafiq and M. Patzold. The influence of the severity of fading and shadowing on the statistical properties of the capacity of nakagami-lognormal channels. In *Global Telecommunications Conference, 2008. IEEE GLOBECOM 2008. IEEE*, pages 1–6. IEEE, 2008.

[117] G. Rafiq and M. Patzold. Statistical properties of the capacity of rice channels with mrc and egc. In *Wireless Communications & Signal Processing, 2009. WCSP 2009. International Conference on*, pages 1–5. IEEE, 2009.

[118] T. Rappaport. *Wireless Communications: Principles and Practice.* Prentice Hall PTR, 2001.

[119] T. Rolski, H. Schmidli, V. Schmidt, and J. L. Teugels. *Stochastic Processes for Insurance and Finance.* John Wiley & Sons, 1998.

[120] W. Rudin et al. *Principles of mathematical analysis*, volume 3. McGraw-hill New York, 1976.

[121] L. Rüschendorf. On conditional stochastic ordering of distributions. *Advances in Applied Probability*, 23(1):46–63, 1991.

[122] L. Rüschendorf. Stochastic ordering of risks, influence of dependence and as constructions. *Advances on Models, Characterizations and Applications*, pages 19–56, 2005.

[123] L. Rüschendorf. *Mathematical Risk Analysis: Dependence, Risk Bounds, Optimal Allocations and Portfolios.* Springer Science & Business Media, 2013.

[124] N. C. Sagias and G. K. Karagiannidis. Gaussian class multivariate weibull distributions: theory and applications in fading channels. *IEEE Transactions on Information Theory*, 51(10):3608–3619, 2005.

[125] A. Sarantsev. Tail asymptotic of sum and product of random variables with applications in the theory of extremes of conditionally gaussian processes. *arXiv preprint arXiv:1107.3869*, 2011.

[126] V. Schmitz. *Copulas and stochastic processes.* PhD thesis, Bibliothek der RWTH Aachen, 2003.

[127] M. Shaked and J. G. Shanthikumar. *Stochastic orders.* Springer Science & Business Media, 2007.

[128] D.-S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn. Fading correlation and its effect on the capacity of multielement antenna systems. *IEEE Transactions on communications*, 48(3):502–513, 2000.

[129] M. K. Simon and M.-S. Alouini. Exponential-type bounds on the generalized marcum q-function with application to error probability analysis over fading channels. *IEEE Transactions on Communications*, 48(3):359–366, 2000.

[130] M. Spivak. *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus.* Westview Press, 1965.

[131] Ž. Stojanac, D. Suess, and M. Kliesch. On products of gaussian random variables. *arXiv preprint arXiv:1711.10516*, 2017.

[132] Q. Tang. From light tails to heavy tails through multiplier. *Extremes*, 11(4):379, 2008.

[133] Q. Tang. Insensitivity to negative dependence of asymptotic tail probabilities of sums and maxima of sums. *Stochastic Analysis and Applications*, 26(3):435–450, 2008.

[134] P. Tankov. Tails of weakly dependent random vectors. *Journal of Multivariate Analysis*, 145:73–86, 2016.

[135] E. Telatar. Capacity of multi-antenna gaussian channels. *European transactions on telecommunications*, 10(6):585–595, 1999.

[136] P. K. Trivedi, D. M. Zimmer, et al. Copula modeling: an introduction for practitioners. *Foundations and Trends® in Econometrics*, 1(1):1–111, 2007.

[137] J. A. Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.

[138] D. Tse and P. Viswanath. *Fundamentals of wireless communication.* Cambridge university press, 2005.

[139] A. Turkmani and J. Parsons. Characterisation of mobile radio signals: base station crosscorrelation. *IEE Proceedings I-Communications, Speech and Vision*, 138(6):557–565, 1991.

[140] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference.* Springer Science & Business Media, 2004.

[141] G. Wei and T. Hu. Supermodular dependence ordering on a class of multivariate copulas. *Statistics & probability letters*, 57(4):375–385, 2002.

[142] W. Whitt. Uniform conditional stochastic order. *Journal of Applied Probability*, 17(1):112–123, 1980.

[143] S. H. Xu. Structural analysis of a queueing system with multiclasses of correlated arrivals and blocking. *Operations Research*, 47(2):264–276, 1999.

[144] H. Yang and S. Sun. Subexponentiality of the product of dependent random variables. *Statistics & Probability Letters*, 83(9):2039–2044, 2013.

[145] Y. Yang, S. Hu, and T. Wu. The tail probability of the product of dependent random variables from max-domains of attraction. *Statistics & Probability Letters*, 81(12):1876–1882, 2011.

[146] V. M. Zolotarev. Mellin-stieltjes transforms in probability theory. *Theory of Probability & Its Applications*, 2(4):433–460, 1957.

ISBN 978-82-326-4388-2 (printed version)
ISBN 978-82-326-4389-9 (electronic version)
ISSN 1503-8181

NTNU
Norwegian University of
Science and Technology

Doctoral Thesis

Fengyou Sun

Doctoral theses at NTNU, 2020:10

NTNU
Norwegian University of
Science and Technology
Faculty of Information Technology
and Electrical Engineering
Department of Information Security
and Communication Technology

NTNU

Doctoral theses at NTNU, 2020:10

Fengyou Sun

Dependence Control in Wireless Communication

NTNU
Norwegian University of
Science and Technology