T2R2 東京工業大学リサーチリポジトリ
Tokyo Tech Research Repository

論文 / 著書情報
Article / Book Information

題目(和文)	時空間相関の統計情報を利用したMIMOチャネル推定に関する研究
Title(English)	MIMO Channel Estimation Exploiting Spatiotemporal Correlation Statistics
著者(和文)	成瀬洋介
Author(English)	Yousuke Naruse
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第9464号, 授与年月日:2014年3月26日, 学位の種別:課程博士, 審査員:高田　潤一,高橋　邦夫,山下　幸彦,荒木　純道,府川　和彦
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第9464号, Conferred date:2014/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

Powered by T2R2 (Tokyo Institute Research Repository)

Doctoral Dissertation

MIMO Channel Estimation Exploiting Spatiotemporal Correlation Statistics

February 2014

Supervisor: Professor Jun-ichi Takada

Department of International Development Engineering
Graduate School of Engineering
Tokyo Institute of Technology

Yousuke Naruse

Abstract

Recently, MIMO transmission is attracting a lot of attention since it achieves much higher transmission capability by utilizing the spatial freedom of the waves in radio propagation. In MIMO systems, accuracy of the estimate of the channel state information (CSI) dramatically affects the performance of transmission because the estimation error leads to a decrease of SINR due to the degradation of the separation capability of each signal on the spatial freedom. Generally in pilot symbol aided channel estimation with a constant noise power, the more energy we put into the transmission of training symbols, the lower the estimation error can be. Especially for rapidly fading channels where the channel estimation must be frequently repeated in certain intervals to track the temporal changes of the channel, required energy and time for channel estimation consume a lot of resources of the wireless communication system. We address the issue of accuracy enhancement of MIMO channel estimation consuming fewer resources spared for channel observation. Regarding system model, we adopt the MIMO antenna selection system in order to generalize the problem formulation. In the system, we can consider a partial measurement problem where we only measure channels for only certain selected antenna elements, and the rest of the channels are predicted with the aid of prior spatial correlation information. We investigated the enhancement of channel estimation accuracy of MIMO channel with the aid of spatiotemporal channel statistics. This investigation consists of three steps.

First, under the situation that the channel's correlation statistics are known in advance, we considered how to obtain the best estimate by exploiting the statistics. The spatiotemporal correlations are incorporated to the estimation scheme by assuming the Gauss-Markov channel model. Under the MMSE criteria, the Kalman filter performs an iterative optimal estimation in such case. To take advantage of the enhanced estimation capability, we focused on the problem of channel estimation from a partial channel measurement in the MIMO antenna selection system. We discussed the optimal training sequence design, and also the optimal antenna subset selection for channel measurement based on the statistics. In a highly correlated channel, the estimation works even when the measurements from some antenna elements are omitted at each fading block.

Secondly, we consider how to estimate necessary statistics which is generally unknown in actual environments. A random-walk based Gauss-Markov model which has less parameters to be estimated is presented for the Kalman filtering. In order to obtain the channel's innovation statistical parameter, considering that the time evolution of the channel is a latent statistical variable, the EM algorithm is applied for accurate estimation. Numerical simulations revealed that the method is able to enhance estimation capability by exploiting spatiotemporal correlations, and works well under the conditions where small forgetting factor employed.

Thirdly, we discussed the problem from another point of view; in estimation problems, a property of the observation operator is generally a dominating factor for estimation accuracy. We discussed a method to control the observation operator by means of resource allocation of training sequences according to the channel's temporal correlation statistics. So far, we had

investigated that the channel estimation of MIMO antenna selection systems is improved by using the Kalman filter under the assumptions that a temporal correlation is perfectly characterized as Gauss-Markov model, and a spatial correlation is available. In such scheme, a concrete description of how the channel's time evolution can be expressed is necessary for channel estimation. However, such methods relying on spatial correlation statistics are fragile for a sudden change of spatial correlation which will sometimes happen in some propagation scenarios. We addressed the issue of MIMO channel estimation with the aid of only a rough estimate of temporal correlation statistics where MIMO antenna selection system is employed. Under the temporally correlated channel, the proposed method controls allocation of the length of training symbols for each antenna element so that the elements which are likely to be selected are estimated more accurately than the other elements. In order to utilize this estimation scheme effectively, we also proposed an antenna selection method which takes account of the difference of channel estimation accuracy among antenna elements. When the ML channel estimation was adopted, the proposed selection method is with almost the same computational cost as the conventional selection methods. The proposed method has robustness in the sense that it does not rely on certain mathematical models of temporal correlation, as well as spatial correlation which is sensitive to the movement of the receiver. The numerical simulation using 3GPP-SCM revealed that the proposed method works effectively under shorter training symbols. Though the improvement of channel capacity is only a few percent, the implementation of the scheme requires only a small modification to conventional antenna selection systems.

Acknowledgment

I would like to express my gratitude to Professor Jun-ichi Takada for his continuous supervision and encouragement.

I am grateful to Professor Kiyomichi Araki, Associate Professor Kazuhiko Fukawa, Associate Professor Yukihiko Yamashita, and Professor Kunio Takahashi for their insightful comments.

I would also like to thank all the members of the Takada laboratory.

Finally, I dedicate this thesis to my parents and friends who have supported me with long term encouragement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 A Channel Estimation Problem in MIMO Transmission Systems

The MIMO (Multiple-Input Multiple-Output) transmission scheme [1,2] is a recently emerging wireless transmission technique which enables an enhancement of channel capacity by putting information onto the spatial degree of freedom in wireless propagation channels. (Fig. 1.1)

Since the MIMO transmission utilizes orthogonality among eigen-vectors of a channel matrix in order to realize parallel transmission of independent signals, its transmission capability suffers significant degradation if the accuracy of channel estimation is insufficient; mismatch of the estimated and the actual eigen-vectors yield leaks between subchannels on each spatial degree of freedom, and thus results in degradation of SINR (Signal-to-Interference plus Noise power Ratio).

We address the issue of accuracy enhancement of MIMO channel estimation consuming less wireless communication resources spared for channel observation.

1.2 Exploitation of Channel Statistics in Channel Estimation Problem

A typical way to enhance estimation capability of the wireless propagation channel without increasing observation resources is utilization of constraints arising from mathematical modeling of propagation channel.

Intrinsically, if the radio propagation environment is static, the channel's transfer function is determined uniquely depending only on surrounding radio scattering environments, thus it does not fluctuate stochastically. Once we were able to obtain the precise concrete description of scattering environments, we do not have to estimate the transfer function. However, calculating the transfer function accurately from a perfect deterministic channel modeling is far from feasible due to the following reasons: The scatterers in a realistic natural environment possesses too much freedoms to be estimated, and the modeling becomes too complex since the transfer function is determined by mutual interactions among such scatterers. Furthermore, it is impossible for wireless communication systems that have poor physical resources i.e. antenna aperture, frequency bandwidth, and computational power, to estimate parameters for such precise environment modeling in a deterministic way.

Although the radio propagation channel is considered to be deterministic, it appears that the transfer function fluctuates randomly following a certain probability distribution if we assume

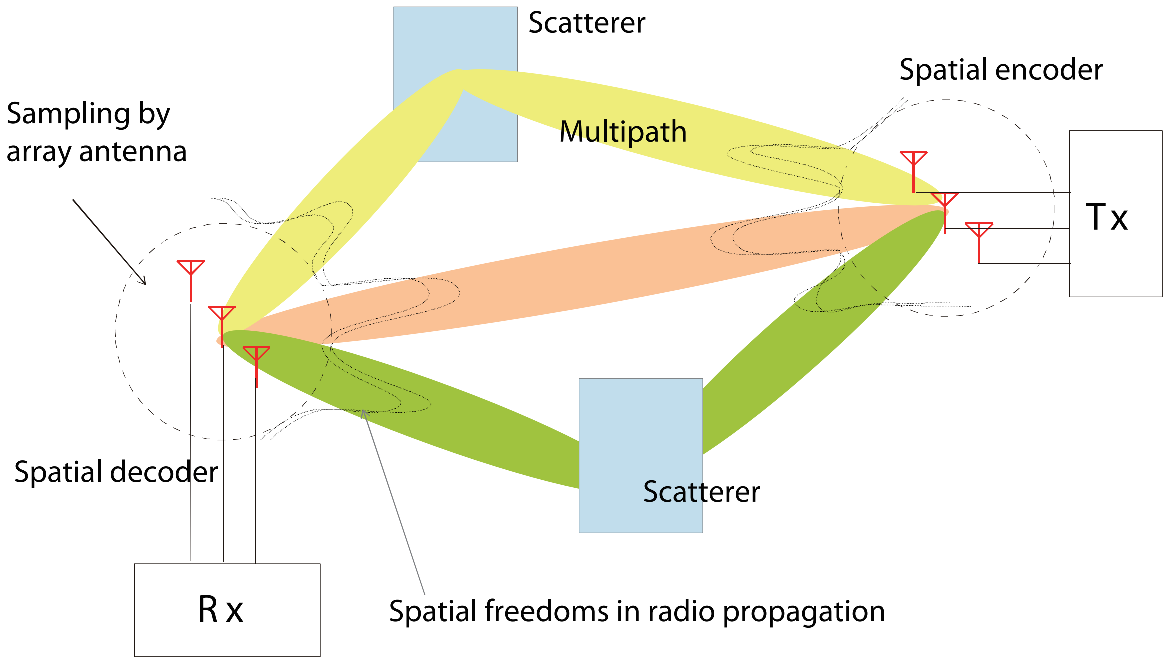

Figure 1.1: The MIMO transmission scheme.

some mobile communication scenarios. For the purpose of channel estimation enhancement in wireless communication system, simplified stochastic model is often used.

In our investigation, we focus on a situation where the channel fluctuation is not perfectly random, but has certain correlation on either the spatial dimension, temporal dimension or both. Assuming such scenario is considered to be valid; in fact, by using physical modeling and analysis of propagation phenomena, it has been explained that channel's spatial correlation is dominated by the spatial distribution of principal scatterers in propagation environment. Similarly, temporal correlation is dominated by moving velocity of MS.

1.3 MIMO Channel Model

1.3.1 Channel Bandwidth and Assumed Encoding Method

Throughout this paper, we will limit our discussion to the narrowband channel. The reason for this assumption is explained as follows.

If we consider realistic wireless communication scenarios, the wideband channel model should be employed. In such cases, it is necessary to deal with correlation of frequencies caused by frequency selective fading.

Inherently, frequency correlation and temporal correlation are physically the same phenomenon in the sense that they both deal with correlations on the time-axis. However, the term 'frequency correlation' in our context stands for correlation of frequency within temporally rectangular windowed signal sequence extracted at a certain symbol duration. In contrast, temporal correlation means correlation among a series of symbols (windowed signals) from larger view of time elapsed.

The same notion is seen in the field of time-frequency analysis, and there exists a so-called uncertainty principle between frequency and time. Each of them corresponds to each axis of a spectrogram.

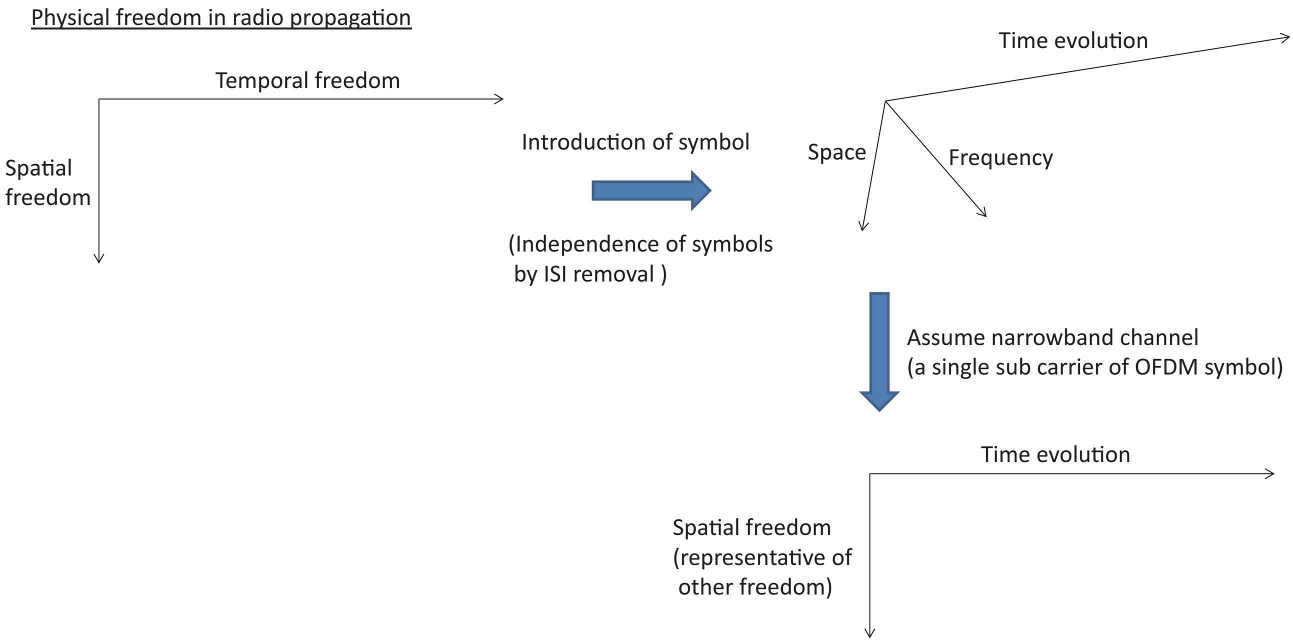

Figure 1.2: The freedom of MIMO-OFDM channel consists of three dimensions thanks to guard interval with cyclic prefix which prevents ISI due to multipath delays. For simplicity, we will consider a single sub carrier of OFDM channel.

Let us consider the case of the orthogonal frequency division multiplexing (OFDM) systems. If the length of guard interval associated with the symbol period is long enough to be able to include the impulse responses caused by multipath delays, we are able to avoid inter symbol interference (ISI). In such cases, information on each symbol become statistically independent of each other.(i.e. mutual information between symbol is zero.) Further, if symbol duration including guard interval is short enough to be regarded as time-invariant system (i.e. impulse response caused by multipath propagation remains constant during symbol duration in the sense of cyclic invariance for discrete time-delay systems), we can express the channel in the frequency domain, where sinusoidal functions are the basis functions, thanks to the guard interval with cyclic prefix of the OFDM. The OFDM puts information bits on such sinusoidal basis function according to the background described so far. The OFDM modulation method is widely used because of its easier implementation for managing frequency selective fading channels.

Thanks to the guard interval with cyclic prefix of OFDM, information on frequency axis of each symbol can be treated as independent among each other. In such a case, frequency characteristics of channel for each symbol evolve over time with certain correlation according to propagation scenarios. Due to this interpretation, we call temporal axis as temporal evolution axis from now on. The frequency axis means frequency within a symbol duration. Now, we must handle two kinds of correlation characteristics; the frequency correlation (channel correlation among sub carriers in OFDM) and the time evolution correlation among channels for each symbol.

If we consider MIMO-OFDM systems, spatial freedom and associated spatial correlation characteristics are additionally incorporated. In such systems, channel state on both spatial freedom and frequency freedom are assumed to evolve along the temporal evolution axis. The explanation is illustrated in Fig.1.2.

When addressing channel statistics of MIMO-OFDM systems, the frequency correlation should be considered in addition to the spatial correlation. In such cases, both correlation statistics can be treated in a similar way by assuming that they follow the multivariate complex normal distribution which is characterized by the second order moment. (If we were to discuss such channel statistics for MIMO-OFDM, it should be investigated whether the spatial and frequency correlation statistics are statistically independent or not. If they are independent, correlation matrix including both spatial and frequency correlation can be expressed as the Kronecker product of each correlation matrices, which results in the required computational cost to be rather small.)

On the other hand, if we do not employ MIMO systems but only OFDM systems, the same formulation in this paper is considered to be applicable by replacing the spatial correlation matrix with the frequency correlation matrix. (Of course, discussions and formulations about antenna selection system cannot be used in such cases.)

Based on what has been described so far, we employ the simplest formulation by assuming the narrowband channel in order to reduce the degrees of freedom (i.e. either spatial or frequency) of the channel for simplicity. When using a narrowband channel, we are allowed to consider only a flat fading channel even under the existence of frequency selective fading caused by multipath propagation, if the signal bandwidth is narrow enough to be regarded as flat fading. In the case of OFDM systems, this narrowband channel corresponds to single sub carrier within the symbol.

1.3.2 Statistical Spatiotemporal MIMO Channel Model

Regarding a concrete channel model for a narrowband channel, spatial correlation statistics are generally assumed to follow the multivariate complex normal distribution which is characterized by the second order moment. In the case of a wideband channel for OFDM systems, correlations among each frequency are also assumed to follow the multivariate complex normal distribution. As for temporal correlation (temporal evolution) of the channel, the first-order AR (autoregressive) model [39] is often assumed among the literatures [19–22, 54, 57]. A larger model order is generally desired when using the AR model. However, due to the limitation of the number of parameters to be estimated, and considering the tracking capability at fast fading channels which have a wideband Doppler frequency, the first-order modeling is often adopted.

In realistic situations, statistical parameter is unknown for communication systems, thus the statistical channel parameters must be estimated in advance prior to receiving the full benefit of estimation capability enhanced by the channel statistics.

1.4 Environmental Factors Dominating MIMO Channel Capacity

The MIMO transmission scheme enhances its capacity by exploiting spatial degrees of freedom of the radio propagation channel. This suggests that the MIMO channel capacity physically depends on how much the propagation environment is able to transfer spatial degrees of freedom inherently. One of the physical conditions dominating the channel capacity is that the multipath scattering is rich enough to achieve wider angular spread toward the array antenna elements. Also width of the array aperture against the angular spread is proven to be important for the MIMO scheme. According to the literatures [59, 60], transferable spatial degrees of freedom can be estimated as illustrated in Fig.1.3.

The reason that both the array aperture and the angular spread are important factors for

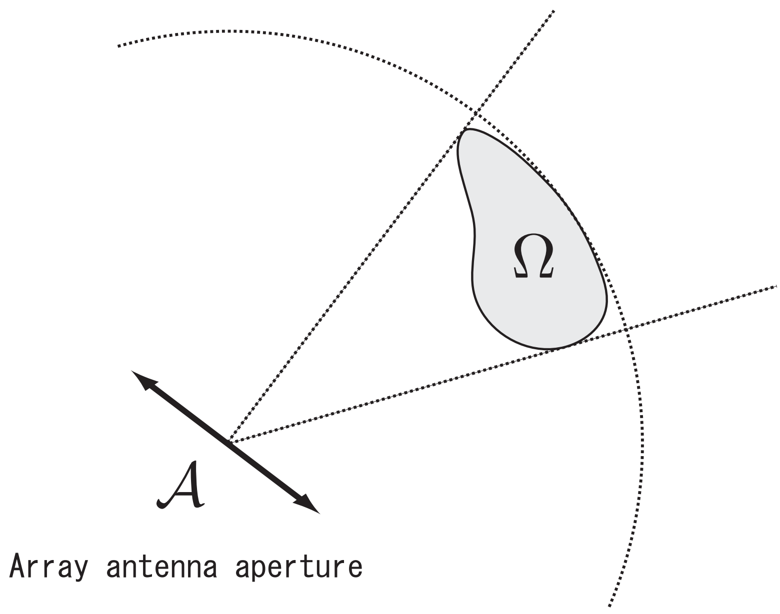

Figure 1.3: Upper limit of degree of spatial freedom is expressed as $N_F = \mathcal{A}|\Omega|$ according to [59, 60].

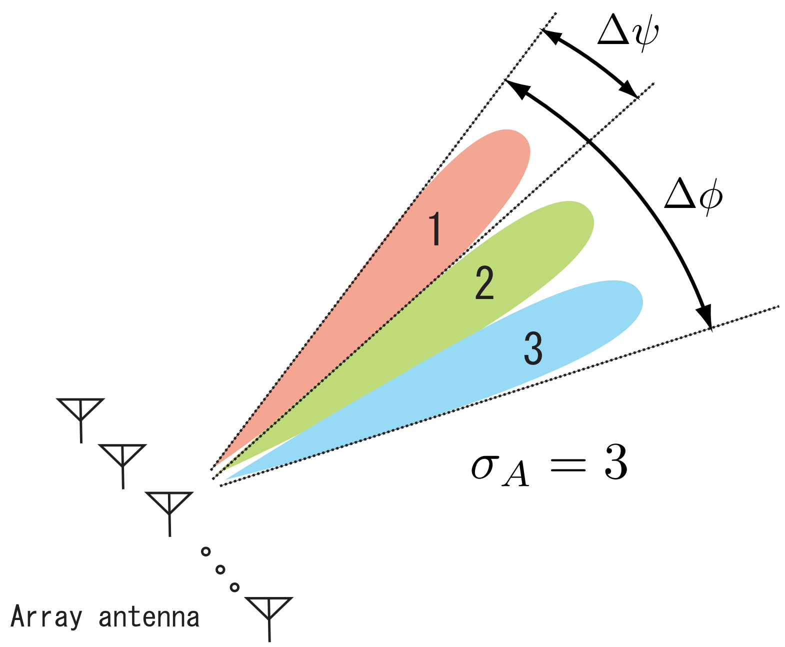

Figure 1.4: Normalized angular spread [61] is defined as $\sigma_A = \Delta\phi/\Delta\psi$ where $\Delta\phi$ and $\Delta\psi$ denote angular spread of incoming/outgoing wave, and array beamwidth of corresponding DOA/DOD, respectively.

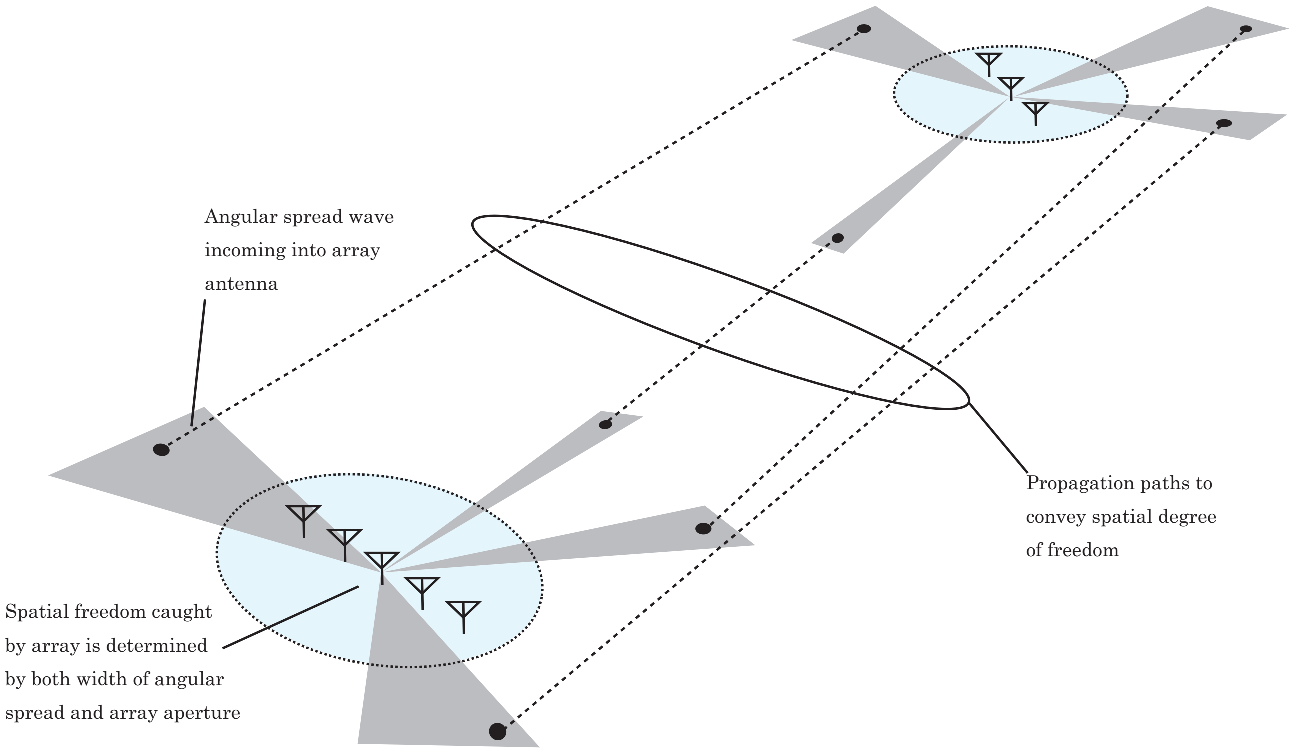

Figure 1.5: Important factors dominating MIMO channel capacity.

channel capacity is intuitively interpreted by considering how many beams corresponding to the array aperture of the direction are included into the angular spread of incoming/outgoing waves. This intuitive estimation gives the rough upper limit of spatial degrees of freedom transferable between radio wave and received signal. This notion was originally introduced in [61], and it is illustrated in Fig.1.4.

Now, it should be noted that this estimation offers only the upper limit. This is because if the arrival angular spread wave has correlation among different angles within them, the array antenna is unable to catch the freedom even with both large aperture and large number of elements. For example, considering a situation that the propagation path between the Tx and the Rx is not multi path, and the whole path connection is concentrated onto a spatially small area around some middle point, the channel capacity degrades even if the incoming/outgoing angular spread around array antenna is wide enough. This is known as the keyhole effect [62], and it means that the propagation environment has a certain upper limitation of channel capacity where the spatial freedom is not transferable enough.

The explanations given above are illustrated in Fig.1.5.

1.5 MIMO Antenna Selection System

In this study, we adopt the MIMO antenna selection systems as the system model. The system schematic is shown in Fig.1.6.

The reason that we adopt the antenna selection system is that the system is able to offer variety of observation models and problem formulations when discussing channel estimation problems, and allows us to discuss from several points of view. Thus the assumption leads to assist generalization of the problem formulation. For example, we can consider a partial mea-

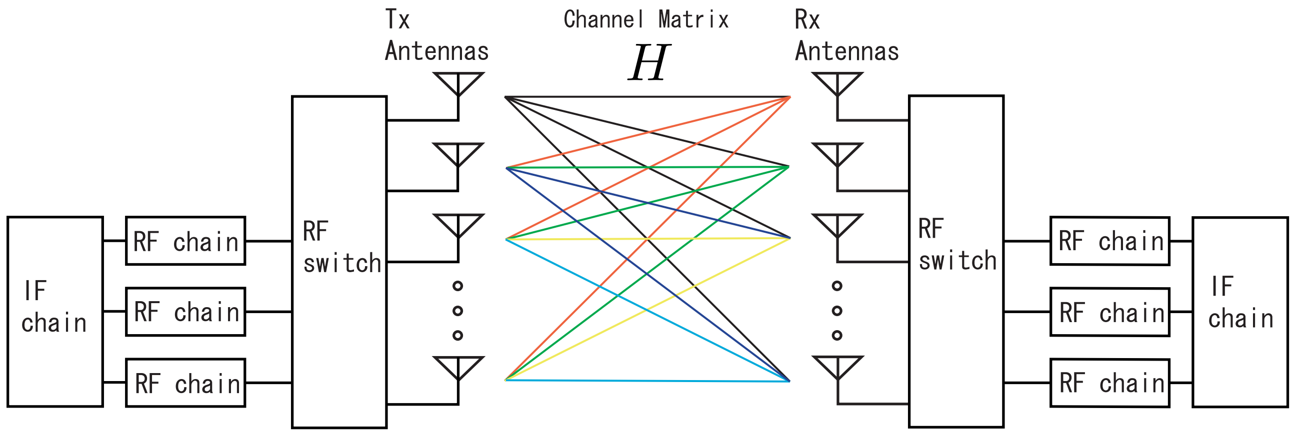

Figure 1.6: Schematic of MIMO antenna selection system. Either or both ends can have RF switch.

surement problem where we only measure channels for only certain selected antenna elements, and rest of the channels are predicted with the aid of prior spatial correlation information.

1.5.1 Principle for Capacity Improvement

As described above, we are unable to do anything with the channel capacity limitation caused by propagation matter about inherent transferable spatial degrees of freedom. However, to some extent, spatial degrees of freedom between angular spread wave and received signals via array antenna can be improved actively by using MIMO antenna selection systems [15].

Let us consider the situation to estimate parameters of a low dimensional parametric function by observation of discrete sampling under the presence of noise. Under badly configured sampling points, the observation operator becomes nearly singular, and the estimation problem becomes an ill-conditioned problem. Thus the estimation procedure requires some regularization technique in order to suppress amplification of noise. The cause of this phenomenon arises from rank-deficiency of the sampling operator due to inadequate sampling positions. If we view the low dimensional parametric function as a correlated angular spread wave, and also view the sampling position as the array antenna's each element position, then we can enhance the channel capacity (it corresponds to the estimation capability of parametric function) by actively adjusting the antenna positions corresponding to the concrete pattern of incoming/outgoing angular spread wave. (Fig.1.7) Based on such considerations, the MIMO antenna selection systems have been proposed; the system has more array antenna elements than the number of RF chains, and the system actively selects the best combination of antenna elements with the change of surrounding propagation environment.

Mathematically, MIMO channel capacity is determined by the distribution of singular values of a MIMO channel response matrix according to the Foschini-Telatar equation [1,2]. For channel capacity improvement, preferably uniform values are desired rather than nonuniform values. Roughly speaking, the reason for this property is that the channel capacity only increases in log order associated with increase of signal amplitude, but it increases in linear order associated with increase of another channels. This fact suggests that one should allocate power resources to create another channel rather than concentrating on certain existent channels.

In this manner, the channel capacity is uniquely determined by the channel response matrix. The channel response matrix is out of our control in normal conditions, but the MIMO antenna

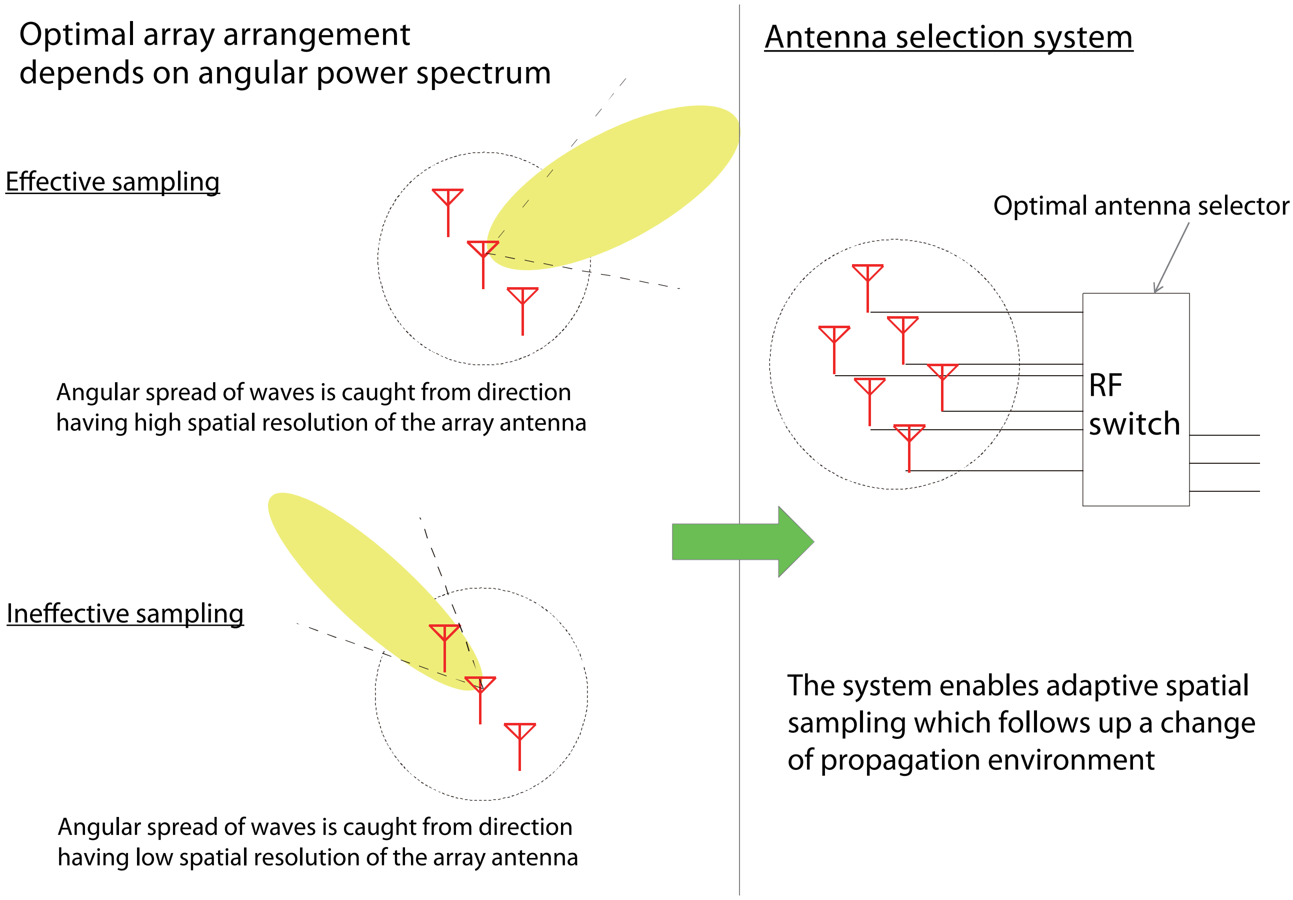

Figure 1.7: Mechanism of capacity improvement of antenna selection system in terms of optimal sampling problem.

selection system enables active change of the matrix in a sense that the system can choose any submatrix by selecting a combination of constituent row or column vectors from whole channel matrix. By employing the MIMO antenna selection system, it is known that the channel capacity is increased by 20 to 30 percent assuming independent and identically-distributed (i.i.d.) channel model. As previously explained, capacity improvement by the system has upper limitation because it is determined by physical characteristics such as array aperture and surrounding scattering environments.

1.5.2 Assumption Regarding Reradiation from Not-selected Elements

In antenna selection systems, if reradiation from antenna elements changes depending on selection, and the distance between antenna elements is small, the mutual coupling between selected and not-selected elements should be taken into account. If the mutual coupling occurs, selection of channel states does not become a simple extraction of row/column vectors of the whole channel matrix. In a strict sense, the channel state varies according to concrete selection combinations.

In this paper, we assume that such change of channel state caused by change of antenna selection does not occur. Concretely, this assumption means that the antenna termination of the not-selected elements is properly terminated by impedance matching in order to prevent reradiating from the antenna elements. Another condition to realize assumption is that the distance between antennas is large enough to be able to ignore mutual coupling.

1.6 Training Sequence Design Optimized for Correlation Statistics

In the field of the training sequence based MIMO channel estimation, under the situation that the spatial second-order correlation statistics are available, it has been discussed that the best estimate (in the least mean square error sense) is obtained by means of Wiener filtering. Furthermore, the best training sequence design has been also discussed when the Wiener filter was utilized. If we are able to design the training sequences freely, we can control distribution of singular values of the observation matrix to any desired values. Thus we can mitigate disturbance from noise by assigning large power onto the vector directions which are less reliable, or conceivably having a large estimation error according to the channel model. Mathematically, it is known that the optimal training sequence is designed by the Water-filling algorithm. Intuitively, the algorithm assigns a large observation power onto the channels having the larger estimation error. If the error is too large and is not cost-effective to allocate the power resources, the algorithm ceases to assign powers onto the channels.

1.7 Channel Estimation by the Kalman Filter

In addition to the above case, if the temporal correlation statistics are available under the assumption that the temporal evolution is modeled as AR (Autoregressive) model, then the best estimate in the least square error sense is obtained by the Kalman filtering if the statistical parameters are known in advance. A concrete description of how the channel's time evolution can be expressed is necessary for using the Kalman filter.

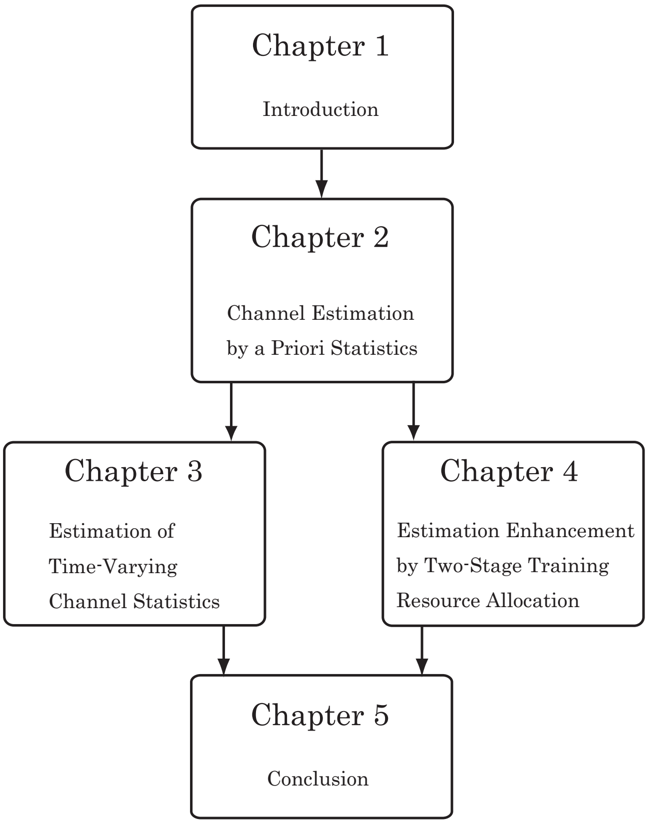

Figure 1.8: Thesis organization.

1.8 Outline of Thesis

We investigate the enhancement of channel estimation accuracy of MIMO channel with the aid of spatiotemporal channel statistics. This investigation consists of three parts.

In chapter 2, under an ideal situation that a channel's correlation statistics remain constant, an optimal channel estimation method is discussed after estimating the channel's statistics parameters by utilizing a sufficient number of observed channel realizations. This situation is almost identical to the one where channel's correlation statistics are known in advance. Mainly, this chapter discusses how to obtain the best estimate by exploiting the statistics.

Chapter 3 is devoted to investigate how to estimate channel statistics which is generally unknown but necessary for realizing the described method in the chapter 2 in actual environments. We focused on a situation that channel parameters vary depending on time elapsed. In chapter 4, under the assumption that a fine estimation of the required channel's statistical parameter is difficult, we tried another approach which does not depend on volatile spatial correlation parameter, but depends on only rough estimate of temporal correlation. We discussed the problem from another point of view; in estimation problems, a property of the observation operator is generally a dominating factor for estimation accuracy. We discussed a method to control the observation operator by means of resource allocation of training sequences according to the channel's temporal correlation statistics. Note that the proposed method in this chapter is devoted to MIMO antenna selection systems.

The whole thesis organization is shown in Fig.1.8.

1.8.1 Optimal Training Sequence Design for the Kalman Filter

We first formulate the training sequence based channel estimation by the Kalman filter assuming a typical spatial correlation model so-called the Kronecker model [43]. Physically, the Kronecker model means that the probability distribution with respect to DoD/DoA (Direction of Departure/Arrival) of the propagation path becomes statistically independent between the DoD and the DoA. This model is often used due to its mathematical simplicity. Mathematically, under this assumption, the correlation matrix of the channel matrix is represented as the Kronecker product of correlation matrices of the Tx side and the Rx side.

By assuming the Kronecker model based Gauss-Markov channel, we show that the filter calculation can be simplified under certain assumptions. We also discuss the optimal training sequence design for the Kalman filtering. The difference from the case of the Wiener filtering considering only spatial correlations is that in the case of the Kalman filtering, energy assignment of training sequences are decided based on both the temporal evolution from the previous time instant and the channel estimation error which has been accumulated so far during the past estimation process. A large energy is concentrated along the eigenvectors presumably having large estimation error accumulated so far. Resultantly, the optimal training sequence differs among different fading blocks. In order to implement the scheme, the both ends of the communication system must share the statistical information in advance, and they must calculate the channel's time evolution of statistics with time elapse to synchronize the state, respectively.

We verified the effectiveness of the estimation method by numerical simulation assuming typical correlation model. The estimation capability was confirmed to be enhanced under presence of high spatial and temporal correlation. We also confirmed that even under the condition that the spatial correlation does not strictly obey the Kronecker model assuming a LOS (Line of Sight) environment, the estimation capability becomes higher in spite of correlation model mismatch.

To take advantage of the enhanced estimation capability, we focus on the problem of channel estimation from a partial channel measurement in the MIMO antenna selection system. In a highly correlated channel, the estimation works even when the measurements from some antenna elements are omitted at each fading block.

1.8.2 Estimation of Channel Statistics by EM-based Algorithm

For channel estimation by the Kalman filter, statistical model parameters are necessary. We investigated a method to estimate the statistical information. In mobile communication scenarios, the movement of MS yields continuous change of direction and angular spread of incoming/outgoing waves. It means that the statistical parameters vary continuously according to time elapse.

In order to estimate both spatial and temporal correlation information, we first discussed simplification of the statistical channel model. In the first-order AR model, both the temporal and the spatial correlation coefficients are required. However, stably estimating both parameters is generally difficult, since the estimation error occurred at one side can be absorbed into the other side. For instance, a slight estimation error of temporal correlation coefficient can be absorbed in the spatial correlation matrix. Furthermore, the number of the parameters to be estimated is desired to be as small as possible in order to avoid overlearning.

Considering such points, we proposed a random walk based Gauss-Markov model which fixes the temporal correlation coefficient to 1.0. In other words, our model deals with the 1st order differentiation of channel sequence. In the model, the channel is characterized only by the innovation term assuming to have multivariate complex normal distribution.

Regarding the parameter estimation criterion, we adopt an exponentially weighted importance used similarly in derivation of the RLS (Recursive Least Squares) algorithm. The RLS algorithm is often used for tracking continuously changing statistics in which method filter coefficients are estimated by applying larger weights to temporally near channel realizations according to exponential function.

In this problem, we must consider how to obtain a realization of channel's time evolution which is necessary for ML parameter estimation. A naive and the simplest way is to use the estimated time evolution in the Kalman filter even though it is not an accurate value of realization.

In order to investigate the effects regarding unobservable realization, we evaluated the channel estimation capability when the statistics of the latent channel realizations are calculated by the correct estimation method. Generally in a statistical parameter estimation problem having latent statistical variables, the EM (Expectation - Maximization) algorithm is utilized where the expectation of the likelihood function is taken with respect to the latent variables according to its posterior distribution. We have constructed the EM-based estimation method, and evaluated its behavior by numerical simulation using the 3GPP-SCME channel model.

From the simulation, we confirmed that the proposed simplified channel model is able to enhance estimation capability by exploiting both the spatial and the temporal correlations. By employing the EM-based algorithm, it revealed that the estimation capability is enhanced where forgetting factor is small. For the forgetting factor close to 1.0, the estimation capability was not affected whether the EM-algorithm is employed or not.

A simulation and computational reduction in the case of antenna selection system was also discussed.

1.8.3 A Robust Estimation by Two-Stage Training Resource Allocation Exploiting Only Temporal Correlation

Generally in linear estimation problems, estimation capability is dominantly determined by the observation matrix of the system model. For example, if all the independent variables to be estimated are able to be observed with enough SNR, there is no concern about the estimation problem.

Now we try to consider a design problem of observation matrix (i.e. training resource allocation) from another point of view.

An approach to utilize prior statistical information for active improvement of the observation matrix is already taken in the optimal training sequence design from the spatial correlation statistics.

In the discussions so far, we have shown that the estimation capability can be significantly enhanced by optimal training sequence design which is dedicated for a given reliable channel statistics. However, this method has a problem in the sense that the method does not have robustness against accuracy of the spatial correlation statistics. For instance, if a channel is considered to have so high correlation that some eigenvalues were treated as zero, the situation arises where the powers of training sequence is not completely assigned onto the eigenvectors corresponding to such small eigenvalues. In such situations, if the assumed correlation statistics turns out to be wrong, increase of the estimation error becomes an enormous influence. In addition, sometimes the spatial correlation statistics are considered to be unstable since it highly depends on DOA/DOD which is affected by various sensitive factors such as geometry

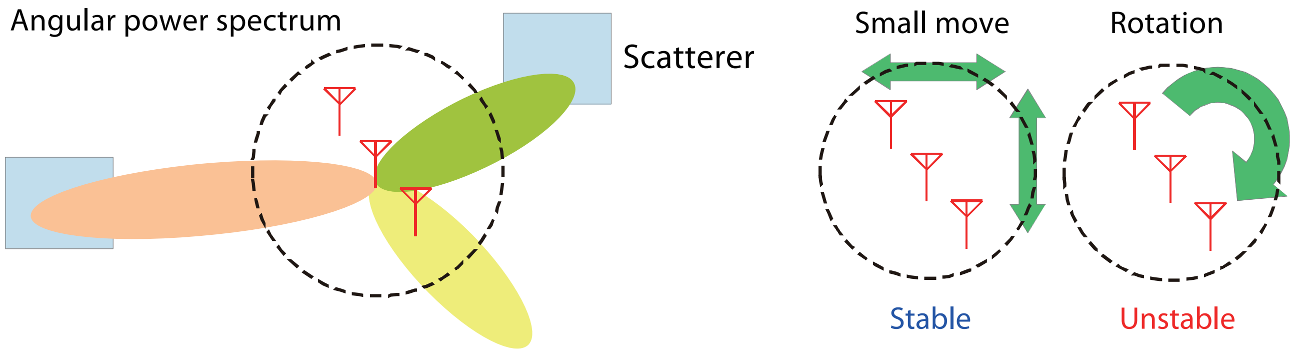

Figure 1.9: Spatial correlation statistics vary according to DOA/DOD. Hence they are unstable under certain movement of MS like rotation.

of surrounding scattering environment, and certain movement of MS like rotation. (Illustrated in Fig.1.9)

In this study, we try to construct a training resource allocation which can enhance estimation capability without using unreliable spatial correlation information but only rough temporal correlation information. In later discussions, we focus on the issue of enhancement of channel estimation efficiency in terms of control of observation matrix. In normal MIMO systems, the training sequence design is the only way to change the observation matrix. For more flexible design of the observation matrix, we employ MIMO antenna selection system for later discussion.

In antenna selection system, receptions of training sequence are required several times while changing the connection of RF switches so that all the antenna elements are measured without fail. In our measurement scheme, we differ the training sequence length between the first and the second measurement, so that the elements measured in the first step are estimated more precisely than other elements. For the first step, if we could choose elements to be measured which are likely to be chosen in the next fading block, then the estimation accuracy of the resultant selected elements is expected to be improved.

Considering that the channel has temporal correlation, the antenna elements which are likely to be selected are chosen from those which were selected in the previous fading block. As the temporal correlation becomes stronger, we control the assignment of training sequences so that the measurement in the first step becomes more precise, expecting that the precisely measured elements are selected again.

The channel state information measured in such way has a characteristic that estimation accuracy varies by antenna elements. In order to increase performance in such situations, we must find the optimal antenna subset which maximizes channel capacity considering differences of estimation accuracy among antenna elements. To perform such selection with limited computational resources, we developed a method to maximize the lower bound of mutual information assuming that the channel has nonuniform estimation errors. This criterion means that the channel capacity under interference due to estimation error in the worst case scenario is to be maximized. The average channel capacity is enhanced as well by means of the criterion. Regarding computational cost, the proposed criterion can be optimized with almost the same computations as the case of conventional effective antenna selection algorithm.

The numerical simulation using 3GPP-SCM revealed that the proposed method works effectively under shorter training symbols employed.

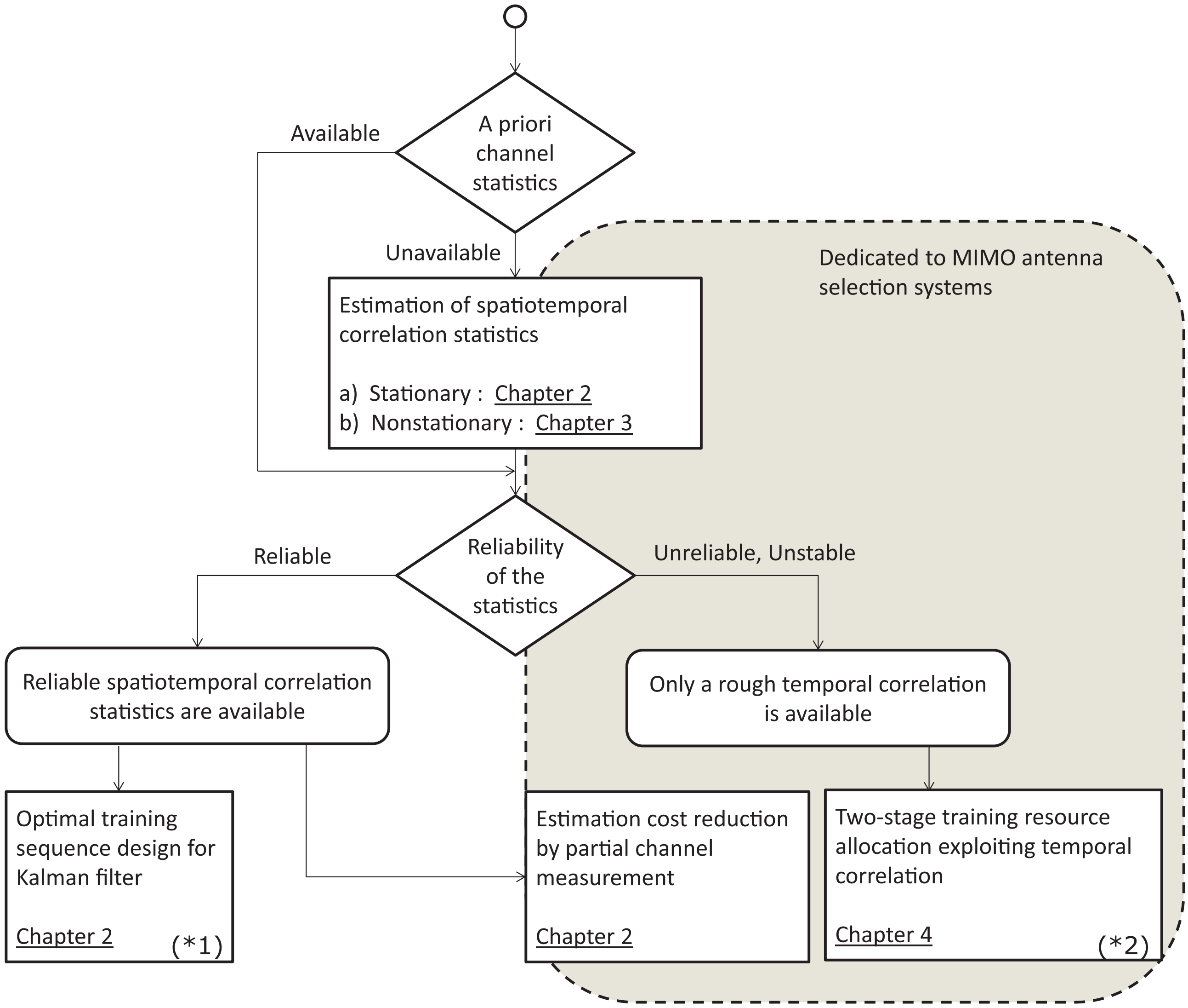

Figure 1.10: A flow chart indicating relationship between each discussion of the thesis. Note that (*1) and (*2) are not examined in conjunction with the channel statistics estimation in Chapter 3.

1.8.4 Relationship Between Individual Discussions

The flow chart of Fig.1.10 indicates relationships among technical investigations in this thesis. The chart provides conditional branching of each technique in terms of both applicable system requirements and available statistical information.

Firstly, depending on the situation whether a priori channel statistics are available or not, requirement of estimation of channel statistics is determined. If it is an ideal situation that the statistical parameter of channel remains constant, the estimation procedure is explained in Chapter 2. On the other hand, if the situation was rather realistic such that the channel is nonstationary having time-varying parameters, the estimation procedure is discussed in Chapter 3.

Next, based on the reliability of the channel statistics, types of channel estimation technique available is provided. If the statistics are fine enough, an optimal training sequence design and reduction of number of measurement in antenna selection system are available. On the other hand, if the statistics are coarse, an effective channel estimation method exploiting only rough temporal correlation information is available. However, this method is only applicable to MIMO antenna selection systems.

The filled area with gray color in the figure means that the included methods are dedicated to MIMO antenna selection systems. The proposed method in chapter 3 is available for both system configurations.

Chapter 2

MIMO Channel Estimation by the Kalman Filter

In this chapter, we address the issue of MIMO channel estimation with the aid of a priori temporal correlation statistics of the channel as well as the spatial correlation. The temporal correlations are incorporated to the estimation scheme by assuming the Gauss-Markov channel model. Under the MMSE criteria, the Kalman filter performs an iterative optimal estimation. To take advantage of the enhanced estimation capability, we focus on the problem of channel estimation from a partial channel measurement in the MIMO antenna selection system. We discuss the optimal training sequence design, and also the optimal antenna subset selection for channel measurement based on the statistics. In a highly correlated channel, the estimation works even when the measurements from some antenna elements are omitted at each fading block.

2.1 Introduction

Recently, MIMO transmission is attracting a lot of attention since it achieves much higher transmission capability by utilizing the spatial freedom of the waves in radio propagation [1,2]. In MIMO systems, precision of the estimate of the channel state information (CSI) dramatically affects the performance of transmission because the estimation error leads to a decrease of SINR due to the degradation of the separation capability of each signal on the spatial freedom. Generally in pilot symbol aided channel estimation with a constant noise power, the more energy we put into the transmission of training symbols, the lower the estimation error can be. Especially for rapidly fading channels where the channel estimation must be frequently repeated in certain intervals to track the temporal changes of the channel, required energy and time for channel estimation consume a lot of resources of the wireless communication system. In order to estimate channel accurately using a limited set of resources, optimal design of training symbol have been discussed by fully exploiting *a priori* channel statistics about the spatial correlation [26]. A MIMO channel with N_{Tx} transmit and N_{Rx} receive antenna possesses $N_{\mathrm{Tx}}N_{\mathrm{Rx}}$ independent freedoms to be estimated. In pilot symbol aided estimation, certain powers are allocated and transmitted for each freedom in order to perform their observation. Without the channel statistics, we have no choice but to treat them equally, which is called the orthogonal training sequence. On the other hand, if the channel is spatially correlated, each freedom of the channel has different variance, and then the one with larger variance is considered to be a more important unknown which contributes more to estimation capability. In order to estimate such important unknowns by priorty, optimal training symbols are designed such that the power in the training symbols is concentrated on the important unknown of the channel matrix. Physically, this can be interpreted as the energy of the training sequence is focused on the significant scatterer in the propagation environment [26].

In addition to the spatial correlation, by exploiting the nature of temporal correlation of the channel, a more effective estimation can be realized. A stochastic model for time-varying channel is required in order to incorporate the temporal correlation into the estimation scheme, and autoregressive (AR) modelling is widely accepted among these studies [3, 4, 18, 42]. In such context, MMSE estimate of the channel can be performed iteratively with the Kalman filter [17]. It should be noted that for the estimation of time-variant frequency-selective fading channels, Kalman-based methods are quite common in the literatures [18, 19].

In this paper, we extend the optimal design of training sequence to the Kalman-based estimation methods. The optimal training sequence of this case is determined respectively for each fading block. Here, the allocation of power for each freedom is determined similarly, but the variances refered are determined by accumulating the estimation errors occurred so far, based on the model of temporal channel transition.

Also we investigate the optimal partial measurements for MIMO antenna selection systems which can be discussed in the similar way as the optimal training sequence design. If the temporal and spatial correlations are high enough, we can partially omit the observation of the channel while satisfying the desired precision of the estimation. This approach might be useful especially in situations where measurement of the full channel state costs much.

In MIMO antenna selection systems with N_{RF} RF chains and M antenna elements, since the number of antenna elements is larger than the number of RF chains, it requires at least $\lceil M/N_{\mathrm{RF}} \rceil$ times measurements to cover all the antenna elements by selecting connected subsets of elements sequentially, where $\lceil x \rceil$ means the smallest integer not exceeded by x. With presence of highly correlated channel, measurements for all the antenna elements can be avoided by interpolating the not measured portion of the full channel matrix with the aid of a priori correlation statistics. This can reduce the repeated measurements in the antenna selection systems. In this paper, the method is formulated to replace the repeated measurements with a single measurement by utilizing the Kalman filter.

We have shown that the estimation can be also enhanced by optimally selecting the measuring elements, based on the channel statistics at each time. In general, filter calculations and optimal training sequence design, also the selection of measurement antenna subset requires much computational costs including the large matrix calculations. But, it was shown that with Kronecker model assumption about the spatial correlation, the filter calculations can be reduced to some small matrix calculations, and the training sequence is designed by the simple water-filling solutions. Also the near-optimal measurement antenna subset can be selected in a similar way as the fast antenna subset selection algorithm [5] with the aid of the matrix inversion lemma.

This paper is organized as follows. In section 2, the system model including the MIMO antenna selection is described. Also the stochastic channel model considering spatial and temporal correlation is explained. Section 3 is devoted to explain a channel estimation using the Kalman filter, and section 4 discusses the optimal training symbol design and measurement antenna selection to enhance the estimation capability. The effectiveness of the proposed method is verified by numerical simulation in section 5.

2.1.1 Mathematical Notations

Throughout this paper we will use bold-faced upper case letters to denote matrices, and bold-faced lower case letters for column vectors, light-faced letters for scalar quantities. The subscripts $\top$, $\mathcal{H}$, $*$ indicate transpose, Hermitian transpose (transpose and complex conjugate), and complex conjugate respectively. $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix. Also the inverse, Moore-Penrose pseudo inverse, trace, determinant, and Frobenius norm of the matrix $\boldsymbol{X}$ are denoted by $\boldsymbol{X}^{-1}$, $\boldsymbol{X}^{\dagger}$, $\operatorname{tr}\boldsymbol{X}$, $\det\boldsymbol{X}$, and $\|\boldsymbol{X}\|_F$, respectively. The m-th row and n-th column element of the matrix $\boldsymbol{X}$ is denoted by $[\boldsymbol{X}]_{m,n}$. E_x means the expectation with respect to x. The bracket $< \cdot, \cdot >$ is used for inner product on the column vector space. Also the vector Euclidean norm is expressed as $\|\cdot\|$. In this paper, since we often discuss correlations between each matrix element, it is convenient to treat matrix as one column vector that consists of all its elements. For any $m \times n$ matrix $\boldsymbol{A} = [\boldsymbol{a}_1\,\boldsymbol{a}_2\,\cdots\boldsymbol{a}_n]$, the vec operator generates a $mn \times 1$ vector defined as

$$\operatorname{vec}\boldsymbol{A} \triangleq \left[\boldsymbol{a}_1^\top\,\boldsymbol{a}_2^\top\,\cdots\boldsymbol{a}_n^\top\right]^\top \tag{2.1}$$

where $\triangleq$ means definition. The Kronecker product $\otimes$ is required with the use of the vec operator.

2.2 System Model

2.2.1 MIMO Antenna Selection System

For a narrowband MIMO channel with N_{Tx} transmit antennas and N_{Rx} receive antennas, received vector $\boldsymbol{y} \in \mathbb{C}^{N_{\mathrm{Rx}}}$ can be expressed as

$$\boldsymbol{y} = \sqrt{\frac{P_r}{N_{\mathrm{Tx}}}}\boldsymbol{H}_k\boldsymbol{x} + \boldsymbol{n} \tag{2.2}$$

where P_r is average receive signal power at each receive antenna. $\boldsymbol{n} \in \mathbb{C}^{N_{\mathrm{Rx}}}$ is the additive noise vector typically assumed to have a white complex normal distribution with average power σ_n^2. $\boldsymbol{x} \in \mathbb{C}^{N_{\mathrm{Tx}}}$ is the normalized transmit vector such that $\mathrm{E}_{\boldsymbol{x}}\boldsymbol{x}\boldsymbol{x}^{\mathcal{H}} = \boldsymbol{I}_{N_{\mathrm{Tx}}}$. $\boldsymbol{H}_k \in \mathbb{C}^{N_{\mathrm{Rx}} \times N_{\mathrm{Tx}}}$ is the normalized channel matrix for time instant k. Since $\boldsymbol{H}_k$ varies for each time instant k, we define the normalization as $\mathrm{E}_k\|\boldsymbol{H}_k\|_F^2 = N_{\mathrm{Tx}}N_{\mathrm{Rx}}$. In this formulation, average signal to noise ratio (SNR) per receive antenna can be expressed as follows:

$$\mathrm{SNR} = \frac{\mathrm{E}_{\boldsymbol{H}_k}\mathrm{E}_{\boldsymbol{x}}\left\|\sqrt{\frac{P_r}{N_{\mathrm{Tx}}}}\boldsymbol{H}_k\boldsymbol{x}\right\|^2}{\mathrm{E}_{\boldsymbol{n}}\|\boldsymbol{n}\|^2} = \frac{P_r}{\sigma_n^2} \tag{2.3}$$

We employ antenna selection system only for receiver side with N_{RF} RF chains satisfying $1 \le N_{\mathrm{RF}} < N_{\mathrm{Rx}}$. If we connect the i-th RF chains to the c_i-th $(1 \le c_i \le N_{\mathrm{Rx}})$ antenna element, only the $\{c_i\}_{i=1}^{N_{\mathrm{RF}}}$-th row vectors of $\boldsymbol{H}_k$ work for reception. If we define $\boldsymbol{\tau}_k$ as a vector specifying connection of RF switches for transmission at time instant k by putting together all $\{c_i\}_{i=1}^{N_{\mathrm{RF}}}$ into one vector as $\boldsymbol{\tau}_k \triangleq [c_1\,c_2\cdots c_{N_{\mathrm{RF}}}]^\top$, the matrix employed for transmission can be written as

$$\begin{bmatrix} [\boldsymbol{H}_k]_{c_1,:} \\ [\boldsymbol{H}_k]_{c_2,:} \\ \vdots \\ [\boldsymbol{H}_k]_{c_{N_{\mathrm{RF}}},:} \end{bmatrix} = \boldsymbol{A}_{\tau_k}\boldsymbol{H}_k, \quad \boldsymbol{A}_{\tau_k} \triangleq \sum_{i=1}^{N_{\mathrm{RF}}} \mathbf{f}_i\mathbf{e}^\top_{<\boldsymbol{\tau}_k,\mathbf{f}_i>} \tag{2.4}$$

where $[\boldsymbol{H}_k]_{c_n,:}$ means the c_n-th row vector of $\boldsymbol{H}_k$, and $\mathbf{e}_i, \mathbf{f}_j$ are the so-called standard basis of $\mathbb{C}^{N_{\mathrm{Rx}}}, \mathbb{C}^{N_{\mathrm{RF}}}$, respectively. We call $\boldsymbol{\tau}_k$ as the connection vector at k for later discussion. Corresponding $\boldsymbol{A}_{\tau_k}$ works as an extraction and permutation of row vectors interest on the time instant k. Similarly for antenna selection systems, channel model becomes $\boldsymbol{y} = \sqrt{P_r/N_{\mathrm{Tx}}}\boldsymbol{A}_{\tau_k}\boldsymbol{H}_k + \boldsymbol{n}$ $(\boldsymbol{y}, \boldsymbol{n} \in \mathbb{C}^{N_{\mathrm{RF}}})$.

Antenna selection is performed so that the Shannon capacity of the channel matrix $\sqrt{P_r/N_{\mathrm{Tx}}}\boldsymbol{A}_{\tau_k}\boldsymbol{H}_k$ becomes larger. Since the true channel matrix is not available directly, we determine the connection $\boldsymbol{\tau}_k$ by refering to the estimated channel matrix $\widehat{\boldsymbol{H}}_k$ instead. From the Foschini-Telatar equation [1, 2] for equal transmit power allocation, $\boldsymbol{\tau}_k$ is determined such that

$$\boldsymbol{\tau}_k = \arg\max_{\tau} \log_2 \det \left(\boldsymbol{I}_{N_{\mathrm{RF}}} + \frac{\mathrm{SNR}}{N_{\mathrm{Tx}}} \boldsymbol{A}_\tau \widehat{\boldsymbol{H}}_k \widehat{\boldsymbol{H}}_k^{\mathcal{H}} \boldsymbol{A}_\tau^{\mathcal{H}} \right)$$
$$\text{subject to:} \quad \operatorname{rank} \boldsymbol{A}_\tau = N_{\mathrm{RF}}\,. \tag{2.5}$$

The imposed condition $\operatorname{rank} \boldsymbol{A}_\tau = N_{\mathrm{RF}}$ suggests that the selected antenna elements should not be overlapped. Practical method to get a perfectly optimal solution of the above equation is not yet discovered. In order to obtain a quasi-optimal solution with lower computational complexity, many antenna selection algorithms have been proposed [5–8]. Many of them require full instant channel state $\widehat{\boldsymbol{H}}_k$ to get selection for each time instant k. In order to keep the optimal selection in the time-varying environment, channel measurements must be frequently repeated to track the temporal changes.

2.2.2 Spatially Correlated MIMO Channel Model

We consider quasi-static block fading channel where the channel remains constant during one transmit block. Let us denote the channel state in the k-th fading block by $\boldsymbol{H}_k$. Now we assume the channel is independent between each blocks. Hence we will suppress the time index k in $\boldsymbol{H}_k$ on such assumption.

We assume that all element of the $\boldsymbol{H}$ obeys multivariate complex normal distribution, therefore the second-order statistics provide enough information to characterize the model. We adopt the typical correlation model so-called the Kronecker model [43] which assumes that spatial Tx and Rx correlations are separable. This means that the channel is determined only by the Tx's and Rx's surrounding scattering environments, and they are independent each other. Hence the existence of dependence like direct-path waves is not considered. Generally, the Kronecker model is said to be suited to NLOS environments. A distance between antenna elements is assumed to be close enough to have correlation among them. Applicability of this model assumption to the LOS model is discussed later. Given the spatial correlation matrices for each side as $\boldsymbol{R}_{\mathrm{Tx}}$ and $\boldsymbol{R}_{\mathrm{Rx}}$, the channel response matrix can be generated as

$$\boldsymbol{H} = \boldsymbol{R}_{\mathrm{Rx}}^{1/2} \boldsymbol{G} \left(\boldsymbol{R}_{\mathrm{Tx}}^{1/2} \right)^{\top} \tag{2.6}$$

where $\boldsymbol{G}$ is a complex gaussian i.i.d. random matrix whose elements obey $\mathcal{CN}(0, 1)$. Also the covariance matrix for each element of the channel matrix can be expressed as follows.

$$\boldsymbol{R}_{\mathrm{MIMO}} \triangleq \mathrm{E}_{\boldsymbol{H}} \operatorname{vec} \boldsymbol{H} \left(\operatorname{vec} \boldsymbol{H} \right)^{\mathcal{H}} = \boldsymbol{R}_{\mathrm{Tx}} \otimes \boldsymbol{R}_{\mathrm{Rx}} \tag{2.7}$$

Due to the normalization condition imposed on $\boldsymbol{H}$, $\operatorname{tr} \boldsymbol{R}_{\mathrm{MIMO}} = (\operatorname{tr} \boldsymbol{R}_{\mathrm{Tx}})(\operatorname{tr} \boldsymbol{R}_{\mathrm{Rx}}) = N_{\mathrm{Tx}} N_{\mathrm{Rx}}$ shall be satisfied. In later discussion, we will impose $\operatorname{tr} \boldsymbol{R}_{\mathrm{Tx}} = N_{\mathrm{Tx}}$ and $\operatorname{tr} \boldsymbol{R}_{\mathrm{Rx}} = N_{\mathrm{Rx}}$. The Kronecker model was often assumed in theoretical analysis of the MIMO channel due to its mathematical simplicity. It is verified by measurements under certain environments in [40].

2.2.3 Stochastic Model for Time-Varying MIMO Channel

In this section, a time varying channel model over a frame interval is discussed based on [42]. Let ρ $(0 \leq \rho \leq 1)$ be a temporal correlation coefficient between adjacent blocks, channel responce matrix for time instant $k+1$ is generated from the previous channel matrix by the following single tap Gauss-Markov model:

$$\boldsymbol{H}_{k+1} = \rho \boldsymbol{H}_k + \sqrt{1-\rho^2}\boldsymbol{X} \tag{2.8}$$

$\sqrt{1-\rho^2}\boldsymbol{X}$ is called process noise. $\boldsymbol{X}$ is a random matrix whose elements obey a multivariate normal distribution, and its covariance matrix is defined as

$$\boldsymbol{Q}_k \triangleq \mathrm{E}_{\boldsymbol{X}} \operatorname{vec} \boldsymbol{X} \left(\operatorname{vec} \boldsymbol{X}\right)^{\mathcal{H}} . \tag{2.9}$$

$\boldsymbol{Q}_k$ shall be normalized such that $\operatorname{tr} \boldsymbol{Q}_k = N_{\mathrm{Tx}} N_{\mathrm{Rx}}$. With Jakes' model assumption [10], ρ can be modeled as $J_0(2\pi f_D T)$ where J_0, f_D and T are the zeroth order Bessel function of the first kind, maximum doppler frequency and time interval between the adjacent blocks.

In order to incorporate a spatial correlation into the sequence of $\{\boldsymbol{H}_k\}$, a spatial correlation is also imposed on the process noise term as well. In this paper, we assume that $\boldsymbol{Q}_k$ is constant during the period of our interest mainly due to mathematical simplicity. Since we address the estimation under the presence of higher temporal correlations, it is reasonable to assume the spatial correlations (i.e. DOAs) are nearly fixed among the fading blocks. Correlated random matrix $\boldsymbol{X}$ is generated with the Kronecker model assumption as (2.6). With this assumption, updating equation becomes,

$$\boldsymbol{H}_{k+1} = \rho \boldsymbol{H}_k + \sqrt{1-\rho^2}\boldsymbol{R}_{\mathrm{Rx}}^{1/2} \boldsymbol{G} \left(\boldsymbol{R}_{\mathrm{Tx}}^{1/2}\right)^{\top} \tag{2.10}$$

This model was originally assumed in [3], and employed for the practical performance evaluation of spatially correlated time-varying MIMO channel. From the nature of the model assumptions, if a sudden change of the spatial correlation statistics occures, the correlations must be remeasured.

At equilibrium, it becomes

$$\operatorname{vec} \boldsymbol{H}_k \sim \mathcal{CN}\Big(\boldsymbol{0},\, \boldsymbol{R}_{\mathrm{Tx}} \otimes \boldsymbol{R}_{\mathrm{Rx}}\Big) \tag{2.11}$$

regardless of initial channel state if $\rho \neq 1$.

2.3 Iterative Channel Estimation for Gauss-Markov Channel

2.3.1 Channel Observation Model

We formulate the observation model of the channel state for training symbol aided channel estimation with MIMO antenna selection systems employed.

Let $\boldsymbol{s}_1^{(k)}, \boldsymbol{s}_2^{(k)}, \cdots, \boldsymbol{s}_{N_t}^{(k)} \in \mathbb{C}^{N_{\mathrm{Tx}}}$ be normalized training sequences, each of them being launched from Tx side in number order at fading block k. They are normalized such that $\|\widetilde{\boldsymbol{S}}_k\|_F^2 = N_t N_{\mathrm{Tx}}$ where $\widetilde{\boldsymbol{S}}_k \triangleq [\boldsymbol{s}_1^{(k)}\ \boldsymbol{s}_2^{(k)}\ \cdots\ \boldsymbol{s}_{N_t}^{(k)}]$. Launched sequences are caught at Rx side with its RF switches

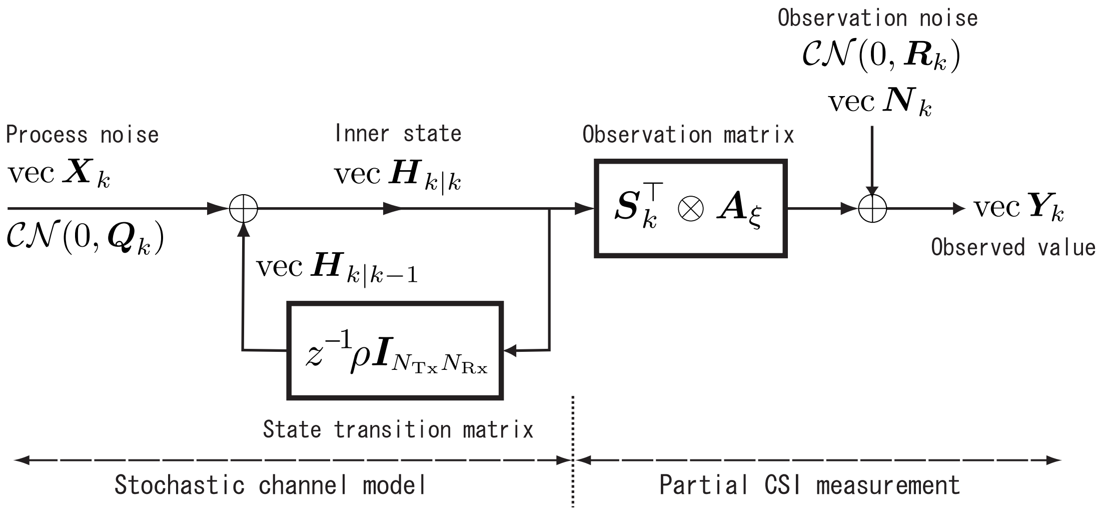

Figure 2.1: Schematic block diagram of the channel transition and observation model.

connected as the connection vector $\boldsymbol{\xi}_k$ prepared for channel measurement. Then, the received sequences are written as

$$[\boldsymbol{y}_1^{(k)}\,\boldsymbol{y}_2^{(k)}\,\cdots\,\boldsymbol{y}_{N_t}^{(k)}] \;=\sqrt{\tfrac{P_r}{N_{\mathrm{Tx}}}}\boldsymbol{A}_{\xi_k}\boldsymbol{H}_k[\boldsymbol{s}_1^{(k)}\,\boldsymbol{s}_2^{(k)}\,\cdots\,\boldsymbol{s}_{N_t}^{(k)}] + [\boldsymbol{n}_1\,\boldsymbol{n}_2\,\cdots\,\boldsymbol{n}_{N_t}] \tag{2.12}$$

where $\boldsymbol{n}_i \in \mathbb{C}^{N_{\mathrm{Rx}}}$ is the i-th additive noise vector and $\boldsymbol{y}_i^{(k)}$ is the received vector corresponding to the transmitted vector $\boldsymbol{s}_i^{(k)}$. If we do not employ antenna selection systems, $\boldsymbol{A}_{\xi_k}$ shall be altered by $\boldsymbol{I}_{N_{\mathrm{Rx}}}$. As for the design of training sequences, if we have no *a priori* knowledge of the channel, and also the noise is white gaussian, it has been proven that orthogonal training matrix defined as $\widetilde{\boldsymbol{S}}_k\widetilde{\boldsymbol{S}}_k^{\mathcal{H}} = N_t\boldsymbol{I}_{N_{\mathrm{Tx}}}$ provides the best estimation capability [13].

By letting $\boldsymbol{Y}_k$ and $\boldsymbol{N}_k$ be $N_{\mathrm{RF}} \times N_t$ matrices each of them consisting of received vectors and noise vectors, respectively, (2.12) is rewritten as

$$\boldsymbol{Y}_k = \boldsymbol{A}_{\xi_k}\boldsymbol{H}_k\boldsymbol{S}_k + \boldsymbol{N}_k \tag{2.13}$$

where $\boldsymbol{S}_k$ is defined as $\boldsymbol{S}_k \triangleq \sqrt{P_r/N_{\mathrm{Tx}}}\widetilde{\boldsymbol{S}}_k$ and hence $\|\boldsymbol{S}_k\|_F^2 = P_rN_t$ holds. Applying vec operator on both sides of the above equation yields the linear observation model as

$$\mathrm{vec}\,\boldsymbol{Y}_k = \left(\boldsymbol{S}_k^\top \otimes \boldsymbol{A}_{\xi_k}\right)\mathrm{vec}\,\boldsymbol{H}_k + \mathrm{vec}\,\boldsymbol{N}_k \tag{2.14}$$

where $\boldsymbol{S}_k^\top \otimes \boldsymbol{A}_{\xi_k}$ is an observation matrix, and channel state is observed through this linear mapping. Our concern is to solve this linear inverse problem under certain criteria.

2.3.2 Channel Estimation by Kalman Filter

By utilizing the temporal correlation characteristics of the channel in conjunction with the spatial correlations, we can further improve the estimation capability.

Under the MMSE condition, best linear estimate of channel state for each time instant k can be iteratively obtained by the Kalman filter [17] if the true channel state obeys Gauss-Markov model. We assume the channel model as (2.8).

This estimation scheme is available even when the whole channel matrix is not measured for each fading block. Now we consider to exploit this feature in order to omit the part of repeated measurements in MIMO antenna selection. In the later discussion, we assume only a single measurement of (2.14) is available at each fading block. In such context, $(N_{\mathrm{Rx}} - N_{\mathrm{RF}})N_{\mathrm{Tx}}$ variables of the channel matrix are not directly observed, but they are interpolated by the measurements of the different antenna elements, and also the one in the previous fading block.

Fig. 2.1 is a schematic block diagram that illustrates the whole system model. Let us denote observation matrix $\boldsymbol{S}_k^\top \otimes \boldsymbol{A}_{\xi_k}$ by $\boldsymbol{B}_k$ for simplification of notation, updating formula of the Kalman filter is expressed as follows:

$$\operatorname{vec}\widehat{\boldsymbol{H}}_{k|k-1} = \rho \operatorname{vec}\widehat{\boldsymbol{H}}_{k-1|k-1} \tag{2.15}$$

$$\boldsymbol{P}_{k|k-1} = \rho^2 \boldsymbol{P}_{k-1|k-1} + (1-\rho^2)\boldsymbol{Q}_k \tag{2.16}$$

$$\widetilde{\boldsymbol{z}}_k = \operatorname{vec}\boldsymbol{Y}_k - \boldsymbol{B}_k \operatorname{vec}\widehat{\boldsymbol{H}}_{k|k-1} \tag{2.17}$$

$$\boldsymbol{C}_k = \boldsymbol{B}_k \boldsymbol{P}_{k|k-1}\boldsymbol{B}_k^{\mathcal{H}} + \boldsymbol{R}_k \tag{2.18}$$

$$\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1}\boldsymbol{B}_k^{\mathcal{H}}\boldsymbol{C}_k^{\dagger} \tag{2.19}$$

$$\operatorname{vec}\widehat{\boldsymbol{H}}_{k|k} = \operatorname{vec}\widehat{\boldsymbol{H}}_{k|k-1} + \boldsymbol{K}_k\widetilde{\boldsymbol{z}}_k \tag{2.20}$$

$$\boldsymbol{P}_{k|k} = \Big[\boldsymbol{I}_{N_{\mathrm{Tx}}N_{\mathrm{Rx}}} - \boldsymbol{K}_k\boldsymbol{B}_k\Big]\boldsymbol{P}_{k|k-1} \tag{2.21}$$

$\widehat{\boldsymbol{H}}_{k|k}$ is the estimated channel matrix at time k, and $\boldsymbol{R}_k$ is the observation noise covariance matrix defined as $\boldsymbol{R}_k \triangleq \mathrm{E}_{\boldsymbol{N}_k} \operatorname{vec}\boldsymbol{N}_k(\operatorname{vec}\boldsymbol{N}_k)^{\mathcal{H}}$. $\boldsymbol{K}_k$ and $\widetilde{\boldsymbol{z}}_k$ are called Kalman gain and innovation term, respectively. $\boldsymbol{P}_{k|k}$ and $\boldsymbol{P}_{k|k-1}$ are error covariance matrices defined as

$$\boldsymbol{P}_{k|l} \triangleq \mathrm{E}_{\boldsymbol{n}}\mathrm{E}_{\boldsymbol{H}_k} \operatorname{vec}\Big(\boldsymbol{H}_k - \widehat{\boldsymbol{H}}_{k|l}\Big)\Big[\operatorname{vec}\Big(\boldsymbol{H}_k - \widehat{\boldsymbol{H}}_{k|l}\Big)\Big]^{\mathcal{H}} \tag{2.22}$$

where $l \in \{k, k-1\}$. In general formulation of the Kalman filter, $\boldsymbol{R}_k$ and $\boldsymbol{Q}_k$ can be dependent on k. But we assume they remain constant during the period of our interest, hence we will omit the index k from now on. If the noise is white gaussian, letting $\boldsymbol{R} = \sigma_n^2\boldsymbol{I}_{N_{\mathrm{RF}}}$ and substituting (2.18) (2.19) into (2.21), and with the aid of the matrix inversion lemma, (2.21) is rewritten as follows:

$$\boldsymbol{P}_{k|k} = \left(\boldsymbol{P}_{k|k-1}^{-1} + \frac{1}{\sigma_n^2}\boldsymbol{S}_k^*\boldsymbol{S}_k^\top \otimes \boldsymbol{A}_{\xi_k}^{\mathcal{H}}\boldsymbol{A}_{\xi_k}\right)^{-1} \tag{2.23}$$

In this paper, we assume model parameters ρ and $\boldsymbol{Q}$ are known in advance whereas they are not available in practical situations. Therefore we must get the estimate of them and utilize them instead of their true values. For the model parameter estimation, we require some sequences of full channel matrix measured in training period provided right before the use of the Kalman filter. A concrete estimation method is described in Appendix A.

2.4 Active Enhancement of Estimation

In this section, we incorporate the concept of active learning into our estimation scheme. Generally in linear inverse problems, estimation capability highly depends on the observation matrix in (2.14). In this case, since observation matrix is composed of connection of RF switches $\boldsymbol{\xi}_k$ and transmitted training sequences $\boldsymbol{S}_k$, we can enhance estimation quality by optimally designing $\boldsymbol{\xi}_k$ and $\boldsymbol{S}_k$ adaptivily to channel environments for each time instant k.

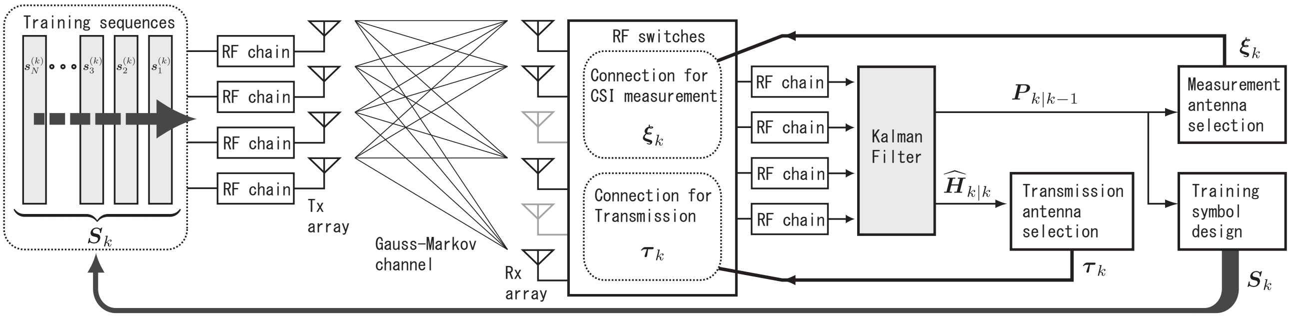

Figure 2.2: Schematic block diagram of the whole system model. Training sequences $\boldsymbol{S}_k$ and antenna subset $\boldsymbol{\xi}_k$ for CSI measurement are actively chosen in order to achieve better estimation. This figure is depicted where $\boldsymbol{S}_k$ and $\boldsymbol{\xi}_k$ are optimized simultaneously, but we do not discuss joint optimization for simplicity.

2.4.1 Problem Formulation

Let us define the MSE criteria as J_k. From (2.22), it is same as a trace of error covariance matrix kept in the Kalman filter. For observation matrix $\boldsymbol{S}^\top \otimes \boldsymbol{A}_\xi$, J_k is written as a functional with respect to ξ and $\boldsymbol{S}$:

$$\begin{aligned} J_k\,[\,\xi, \boldsymbol{S}\,] &\triangleq \mathrm{E}_{\boldsymbol{n}} \mathrm{E}_{\boldsymbol{H}_k} \left\| \boldsymbol{H}_k - \widehat{\boldsymbol{H}}_{k|k} \right\|_F^2 = \operatorname{tr} \boldsymbol{P}_{k|k} \\ &= \operatorname{tr} \left(\boldsymbol{P}_{k|k-1}^{-1} + \frac{1}{\sigma_n^2} \boldsymbol{S}^* \boldsymbol{S}^\top \otimes \boldsymbol{A}_\xi^{\mathcal{H}} \boldsymbol{A}_\xi \right)^{-1} \end{aligned} \tag{2.24}$$

In order to obtain the best estimate, $\boldsymbol{\xi}_k$ and $\boldsymbol{S}_k$ should be chosen such that

$$\begin{aligned} (\xi_k, \boldsymbol{S}_k) &= \arg\min_{\xi, \boldsymbol{S}} J_k\,[\,\xi, \boldsymbol{S}\,]\ , \\ \text{subject to:}& \quad \operatorname{rank} \boldsymbol{A}_\xi = N_{\mathrm{RF}}, \quad \|\boldsymbol{S}\|_F^2 = P_r N_t\,. \end{aligned} \tag{2.25}$$

Finding the solution to this conditional optimization problem requires much computational complexity. But relying on computational power is not suitable because we require this optimization for each fading block. In the following, only the specific cases of which we can handle are taken up for discussion. Schematic diagram of the system is shown in Fig. 2.2.

2.4.2 Optimal Training Sequence Design without Antenna Selection System

In this section, we consider the case in which antenna selection system is not employed. In this case, all $\boldsymbol{A}_{\xi_k}$ in equations shall be replaced by $\boldsymbol{I}_{N_{\mathrm{Rx}}}$.

Absence of Temporal Correlation ($\rho = 0$)

$\rho = 0$ means the estimation is not affected from the previous channel state, and in this case, the situation becomes identical to the MMSE estimation by Wiener filter. Optimal transmit signal design assuming this situation is already discussed in [26]. The optimal training sequence becomes the weighted eigenvectors of the spatial correlation matrix of Tx side whose power allocation is determined by water filling solution [11].

Presence of Temporal Correlation $(0 < \rho < 1)$

This case corresponds to a simple extension of [26] to Gauss-Markov channel model. Let us denote eigenvectors and eigen values of $\boldsymbol{R}_{\mathrm{Tx}}$ and $\boldsymbol{R}_{\mathrm{Rx}}$ as

$$\boldsymbol{R}_{\mathrm{Tx}}\boldsymbol{u}_i = \lambda_i^{(\mathrm{Tx})}\boldsymbol{u}_i, \qquad \boldsymbol{R}_{\mathrm{Rx}}\boldsymbol{v}_j = \lambda_j^{(\mathrm{Rx})}\boldsymbol{v}_j, \tag{2.26}$$

where $\lambda_i^{(\mathrm{Tx})}$ and $\lambda_j^{(\mathrm{Rx})}$ are sorted by descending order. If we assume a structure of $\boldsymbol{S}_k$ as

$$\boldsymbol{S}_k = \sum_{i=1}^{N_{\mathrm{Tx}}} \sqrt{\alpha_i^{(k)}}\boldsymbol{u}_i^*\boldsymbol{\varphi}_i^\top, \tag{2.27}$$

where $\{\boldsymbol{\varphi}_i\}_{i=1}^{N_{\mathrm{Tx}}}$ is any subset of arbitrary orthonormal basis of $\mathbb{C}^{N_t}$, and $\alpha_i^{(k)}$ is non-negative value specifying power allocation along the $\boldsymbol{u}_i$, then all updating equations of the error covariance matrix can be calculated independently for each component along each eigen vector, and $\boldsymbol{P}_{k|l}$ can be always expressed as

$$\boldsymbol{P}_{k|l} = \sum_{i=1}^{N_{\mathrm{Tx}}}\sum_{j=1}^{N_{\mathrm{Rx}}} \left[\boldsymbol{\Gamma}_{k|l}\right]_{i,j} \boldsymbol{u}_i\boldsymbol{u}_i^{\mathcal{H}} \otimes \boldsymbol{v}_j\boldsymbol{v}_j^{\mathcal{H}} \tag{2.28}$$

where $l \in \{k, k-1\}$, and $\boldsymbol{\Gamma}_{k|l} \in \mathbb{R}^{N_{\mathrm{Tx}}\times N_{\mathrm{Rx}}}$ contains all eigenvalues characterizing $\boldsymbol{P}_{k|l}$. $\left[\boldsymbol{\Gamma}_{k|l}\right]_{i,j}$ is the eigenvalue for the eigenvector $\boldsymbol{u}_i \otimes \boldsymbol{v}_j$. Now, updating equations are reduced to only arithmetic operations on the each element of $\boldsymbol{\Gamma}_{k|l}$, and hence much computations are saved. Inversely, if $\boldsymbol{P}_{k|k-1}$ is expressed as (2.28), the structure of the optimal training sequence becomes as a form of (2.27) (see Appendix C).

In this case, (2.16), (2.18), (2.19), and (2.21) can be rewritten with much smaller computations as follows:

$$\boldsymbol{\Gamma}_{k|k-1} = \rho^2\boldsymbol{\Gamma}_{k-1,k-1} + (1-\rho^2)[\lambda_1^{(\mathrm{Tx})}\,\lambda_2^{(\mathrm{Tx})}\,\cdots\,\lambda_{N_{\mathrm{Tx}}}^{(\mathrm{Tx})}]^\top[\lambda_1^{(\mathrm{Rx})}\,\lambda_2^{(\mathrm{Rx})}\,\cdots\,\lambda_{N_{\mathrm{Rx}}}^{(\mathrm{Rx})}] \tag{2.29}$$

$$\boldsymbol{C}_k = \sum_{i=1}^{N_{\mathrm{Tx}}}\sum_{j=1}^{N_{\mathrm{Rx}}} \left(\alpha_i^{(k)}\left[\boldsymbol{\Gamma}_{k|k-1}\right]_{i,j} + \sigma_n^2\right)\boldsymbol{\varphi}_i\boldsymbol{\varphi}_i^{\mathcal{H}} \otimes \boldsymbol{v}_j\boldsymbol{v}_j^{\mathcal{H}} \tag{2.30}$$

$$\begin{aligned}\boldsymbol{K}_k &= \boldsymbol{P}_{k|k-1}\left(\boldsymbol{S}_k^\top \otimes \boldsymbol{I}_{N_{\mathrm{Rx}}}\right)^{\mathcal{H}}\boldsymbol{C}_k^{-1} \\ &= \sum_{i=1}^{N_{\mathrm{Tx}}}\sum_{j=1}^{N_{\mathrm{Rx}}} \frac{\sqrt{\alpha_i^{(k)}}\left[\boldsymbol{\Gamma}_{k|k-1}\right]_{i,j}}{\alpha_i^{(k)}\left[\boldsymbol{\Gamma}_{k|k-1}\right]_{i,j} + \sigma_n^2}\boldsymbol{u}_i\boldsymbol{\varphi}_i^{\mathcal{H}} \otimes \boldsymbol{v}_j\boldsymbol{v}_j^{\mathcal{H}}\end{aligned} \tag{2.31}$$

$$\left[\boldsymbol{\Gamma}_{k|k}\right]_{i,j} = \frac{\sigma_n^2\left[\boldsymbol{\Gamma}_{k|k-1}\right]_{i,j}}{\alpha_i^{(k)}\left[\boldsymbol{\Gamma}_{k|k-1}\right]_{i,j} + \sigma_n^2} \tag{2.32}$$

The sum power constraint $\|\boldsymbol{S}_k\|_F^2 = P_rN_t$ is now reduced to $\sum_{i=1}^{N_{\mathrm{Tx}}}\alpha_i^{(k)} = P_rN_t$. Now, we want to minimize $J_k = \sum_{i=1}^{N_{\mathrm{Tx}}}\sum_{j=1}^{N_{\mathrm{Rx}}}\left[\boldsymbol{\Gamma}_{k|k}\right]_{i,j}$, but to simplify the problem, we will consider the minimization only for the first column of $\boldsymbol{\Gamma}_{k|k}$. From (2.29) to (2.32), we can confirm that if the initial value $\boldsymbol{\Gamma}_{1|1}$ satisfies $[\boldsymbol{\Gamma}_{1|1}]_{i,1} \geq [\boldsymbol{\Gamma}_{1|1}]_{i,2} \geq \cdots \geq [\boldsymbol{\Gamma}_{k|k}]_{i,N_{\mathrm{Rx}}}$, it always holds for all k that $[\boldsymbol{\Gamma}_{k|k}]_{i,1} \geq [\boldsymbol{\Gamma}_{k|k}]_{i,2} \geq \cdots \geq [\boldsymbol{\Gamma}_{k|k}]_{i,N_{\mathrm{Rx}}}$. Hence we can say that the first column of $\boldsymbol{\Gamma}_{k|k}$ has the most important effect on our criteria, and we can determine the upper bound of J_k which only depends on the first column of $\boldsymbol{\Gamma}_{k|k}$ as follows:

$$J_k[\boldsymbol{S}_k] = \sum_{i=1}^{N_{\mathrm{Tx}}}\sum_{j=1}^{N_{\mathrm{Rx}}}\left[\boldsymbol{\Gamma}_{k|k}\right]_{i,j} \leq N_{\mathrm{Rx}}\sum_{i=1}^{N_{\mathrm{Tx}}}\left[\boldsymbol{\Gamma}_{k|k}\right]_{i,1} \triangleq J_k'[\boldsymbol{S}_k] \tag{2.33}$$

Minimizing J'_k can be solved by using Lagrange multipliers and Kuhn-Tucker conditions. $\{\alpha_i^{(k)}\}_{i=1}^{N_{\mathrm{Tx}}}$ is determined as

$$\alpha_i^{(k)} = \left[W - \frac{\sigma_n^2}{\left[\boldsymbol{\Gamma}_{k|k-1}\right]_{i,1}} \right]^+ \tag{2.34}$$

where $[x]^+ \triangleq \max(0,x)$ and W is so-called the water line calculated by the water filling algorithm.

At equilibrium, $\boldsymbol{\Gamma}_{k|l}$ converges to the solution of the algebraic Riccati equation if the system is observable and controllable. In this paper, we will not discuss such steady-state Kalman filter, but it might be possible to calculate the optimal training sequence and corresponding filter gains in advance, prior to the actual iterative observations.

2.4.3 Optimal Measurement Antenna Selection

Now we consider the case of antenna selection system with its channel matrix being partially observed. A selection of partial measurement $\boldsymbol{\xi}_k$ is optimized for each iteration under the criteria of (2.25). In this time, $\boldsymbol{S}_k$ is not jointly optimized due to simplicity.

A naive implementation of (2.16) to (2.21) requires calculation of inverse of the $N_{\mathrm{Tx}}N_{\mathrm{Rx}} \times N_{\mathrm{Tx}}N_{\mathrm{Rx}}$ matrix for each step which results in much requirement of computations. Now we show that if we do not optimize on $\boldsymbol{S}_k$, and also it is given as a form of (2.27), we can implement the Kalman filter with smaller computations by avoiding the direct calculation of the inverse of the large error covariance matrix.

On this assumption, $\boldsymbol{P}_{k|l}$ is always expressed as

$$\boldsymbol{P}_{k|l} = \sum_{i=1}^{N_{\mathrm{Tx}}} \boldsymbol{u}_i \boldsymbol{u}_i^{\mathcal{H}} \otimes \boldsymbol{F}_{k|l}^{(i)} \tag{2.35}$$

where $\boldsymbol{F}_{k|l}^{(i)}$ is $N_{\mathrm{Rx}} \times N_{\mathrm{Rx}}$ matrix characterizing error covariance of Rx side which corresponds to Tx's component along $\boldsymbol{u}_i$. Then, thanks to the relationship of Appendix B, filter calculations are rewritten as:

$$\boldsymbol{F}_{k|k-1}^{(i)} = \rho^2 \boldsymbol{F}_{k-1|k-1}^{(i)} + (1-\rho^2)\lambda_i^{(\mathrm{Tx})} \boldsymbol{R}_{\mathrm{Rx}} \tag{2.36}$$

$$\boldsymbol{K}_k = \sum_{i=1}^{N_{\mathrm{Tx}}} \boldsymbol{u}_i \boldsymbol{\varphi}_i^{\mathcal{H}} \otimes \sqrt{\alpha_i^{(k)}} \boldsymbol{F}_{k|k-1}^{(i)} \boldsymbol{A}_{\xi_k}^{\mathcal{H}} \left(\alpha_i^{(k)} \boldsymbol{A}_{\xi_k} \boldsymbol{F}_{k|k-1}^{(i)} \boldsymbol{A}_{\xi_k}^{\mathcal{H}} + \sigma_n^2 \boldsymbol{I}_{N_{\mathrm{RF}}} \right)^{-1} \tag{2.37}$$

$$\boldsymbol{F}_{k|k}^{(i)} = \left[\left(\boldsymbol{F}_{k|k-1}^{(i)} \right)^{\dagger} + \frac{\alpha_i^{(k)}}{\sigma_n^2} \boldsymbol{A}_{\xi_k}^{\mathcal{H}} \boldsymbol{A}_{\xi_k} \right]^{-1} \tag{2.38}$$

By comparing (2.38) with (2.23), we can see the update of $\boldsymbol{F}_{k|k}^{(i)}$ is done by replacing the observation matrix by $\boldsymbol{A}_{\xi_k}$ with its noise power being scaled by $1/\alpha_i^{(k)}$. From (2.35), selection criteria is rewritten as follows:

$$J_k\left[\,\xi_k, \boldsymbol{S}_k\,\right] = \sum_{i=1}^{N_{\mathrm{Tx}}} \mathrm{tr}\, \boldsymbol{F}_{k|k}^{(i)} \tag{2.39}$$

We must compromise the best solution of $\boldsymbol{\xi}_k$, due to the required computations for optimization. In order to find a near optimal solution, we resort to Greedy algorithm which starts from an empty set of selected antennas, then add one antenna which best contributes to the selection criteria per step, continues until it reaches the desired number of selection. We show the implementation of the Greedy algorithm with much simpler calculations by utilizing a similar approach proposed in fast antenna subset selection [5].

Now we consider the $(t+1)$-th step of Greedy algorithm. Until the t-th $(0 \leq t < N_{\mathrm{RF}})$ step, t elements have been already selected, and we will denote $\boldsymbol{\omega}_t$ as a $t \times 1$ vector that contains already selected elements. Also we denote the extracted channel matrix specified by $\boldsymbol{\omega}_t$ as $\sqrt{P_r/N_{\mathrm{Tx}}}\boldsymbol{A}_{\omega_t}\boldsymbol{H}_k$. $\boldsymbol{A}_{\omega_t}$ is a $t \times N_{\mathrm{Rx}}$ matrix defined as $\boldsymbol{A}_{\omega_t} \triangleq \sum_{i=1}^{t} \mathbf{f}_i \mathbf{e}_{<\boldsymbol{\omega}_t,\mathbf{f}_i>}^{\top}$ where $\mathbf{e}_i$ and $\mathbf{f}_i$ are the i-th standard basis of $\mathbb{C}^{N_{\mathrm{Rx}}}$ and $\mathbb{C}^t$, respectivily. If we select x-th element at the $(t+1)$-th step, x is chosen so that

$$\begin{aligned} J_k\left[\,\boldsymbol{\omega}_{t+1}, \boldsymbol{S}_k\,\right] &= J_k\left[\,[\boldsymbol{\omega}_t^{\top}\; x]^{\top}, \boldsymbol{S}_k\,\right] \qquad (2.40) \\ &= \sum_{i=1}^{N_{\mathrm{Tx}}} \operatorname{tr} \boldsymbol{\Phi}_{\omega_{t+1}}^{(i)}, \quad \boldsymbol{\Phi}_{\omega_t}^{(i)} \triangleq \left[\left(\boldsymbol{F}_{k|k-1}^{(i)}\right)^{\dagger} + \frac{\alpha_i}{\sigma_n^2}\boldsymbol{A}_{\omega_t}^{\mathcal{H}}\boldsymbol{A}_{\omega_t}\right]^{-1} \end{aligned}$$

is minimized. Substituting the relationship of

$$\boldsymbol{A}_{\omega_{t+1}}^{\mathcal{H}}\boldsymbol{A}_{\omega_{t+1}} = \boldsymbol{A}_{\omega_t}^{\mathcal{H}}\boldsymbol{A}_{\omega_t} + \mathbf{e}_x\mathbf{e}_x^{\mathcal{H}} \qquad (2.41)$$

into (2.40), and thanks to the matrix inversion lemma, we have

$$\begin{aligned} \boldsymbol{\Phi}_{\omega_{t+1}}^{(i)} &= \left[\left(\boldsymbol{\Phi}_{\omega_t}^{(i)}\right)^{-1} + \frac{\alpha_i}{\sigma_n^2}\mathbf{e}_x\mathbf{e}_x^{\mathcal{H}}\right]^{-1} \qquad (2.42) \\ &= \boldsymbol{\Phi}_{\omega_t}^{(i)} - \left(\frac{\sigma_n^2}{\alpha_i} + \mathbf{e}_x^{\mathcal{H}}\boldsymbol{\Phi}_{\omega_t}^{(i)}\mathbf{e}_x\right)^{-1}\boldsymbol{\Phi}_{\omega_t}^{(i)}\mathbf{e}_x\mathbf{e}_x^{\mathcal{H}}\boldsymbol{\Phi}_{\omega_t}^{(i)} \end{aligned}$$

where $\boldsymbol{\Phi}_{\omega_0}^{(i)} = \boldsymbol{F}_{k|k-1}^{(i)}$. Finally, the element which best contributes to the criteria (2.40) is chosen as follows:

$$\begin{aligned} x_{\mathrm{opt}} &= \arg\min_x J_k\left[\,[\boldsymbol{\omega}_t^{\top}\; x]^{\top}, \boldsymbol{S}_k\,\right] \\ &= \arg\max_x \sum_{i=1}^{N_{\mathrm{Tx}}} \frac{\left\|\left[\boldsymbol{\Phi}_{\omega_t}^{(i)}\right]_{:\,,x}\right\|^2}{\sigma_n^2/\alpha_i + \left[\boldsymbol{\Phi}_{\omega_t}^{(i)}\right]_{x,x}} \qquad (2.43) \end{aligned}$$

After that, the selection result is stored into $\boldsymbol{\omega}_{t+1}$ as $\boldsymbol{\omega}_{t+1} \leftarrow [\boldsymbol{\omega}_t^{\top} x_{\mathrm{opt}}]^{\top}$. Also $\boldsymbol{\Phi}_{\omega_{t+1}}^{(i)}$ is generated by (2.42) and be used for the selection of the next step.

A brief explanation of the selection method is illustrated in Fig.2.4. Also, explanation of [5] is shown in FIg.2.3 for reference.

2.5 Simulation

We evaluated the effectiveness of the proposed scheme by numerical simulations with different parameters for spatial and temporal correlation. As a measure of the estimation capability, we employed the following three criteria.

- Evaluate the Frobenius norm of the estimation error, defined as (2.24).

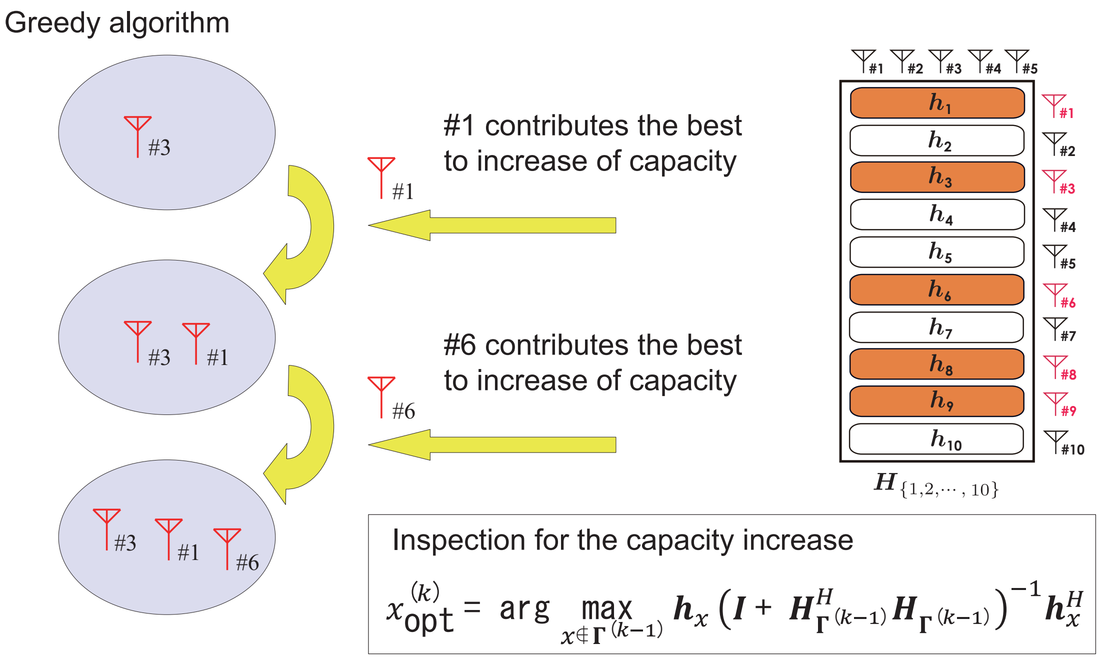

Figure 2.3: Brief explanation of the fast antenna subset selection method [5]. The method reduces computation thanks to the Woodbury formula for matrix determinant.

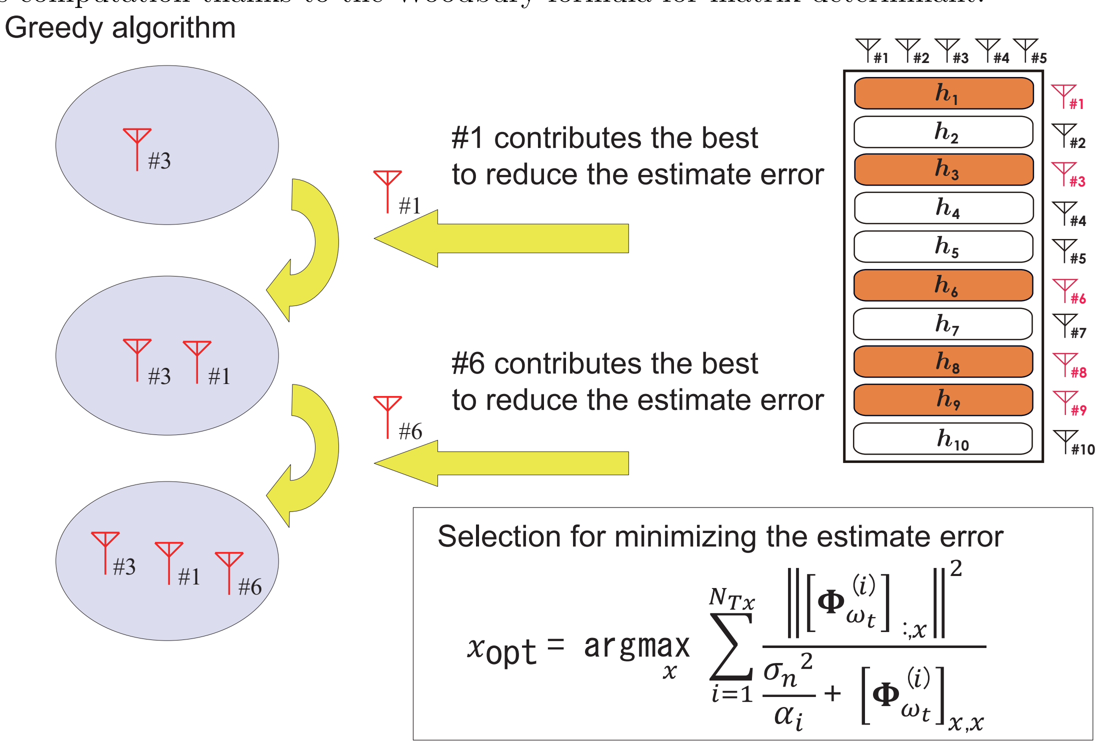

Figure 2.4: Brief explanation of the proposed quasi-optimal measurement antenna selection method. An approach similar to [5] is taken for reduction of computations.

- After the estimation by the Kalman filter, antenna subset used for transmission is chosen according to the information of the estimation result. Then the transmission starts by directly using the estimated channel state information. In this case, the contributing factor for the channel capacity degradation will be the validity of antenna selection and the decrease of SINR, both are caused by the estimation error by the Kalman filter.

- After the antenna selection based on the estimated channel, accurate channel estimation for which only the elements used for transmission is repeated with much longer training sequences and without using any channel statistics. In this case, only the validity of antenna selection can be evaluated.

The Shannon capacity of the MIMO channel is evaluated considering the estimation error of the channel under the MIMO eigenmode transmission system. Let singular value decomposition of the estimated channel matrix be $\boldsymbol{A}_{\tau_k}\widehat{\boldsymbol{H}} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathcal{H}}$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices and $\boldsymbol{S}$ is the diagonal matrix, respectivily. Then the diagonal element of $\boldsymbol{T} \triangleq \boldsymbol{U}^{\mathcal{H}}\boldsymbol{A}_{\tau_k}\boldsymbol{H}\boldsymbol{V}$ stands for the SNR for each sub channel, and $[\boldsymbol{T}]_{i,j}$ $(i \neq j)$ is the leak from the i-th sub channel to the j-th sub channel which is caused by the estimation error. The SINR of the i-th sub channel is expressed as

$$\mathrm{SINR}_i = \frac{P_r\,[\boldsymbol{T}]_{i,i}\,[\boldsymbol{T}]^*_{i,i}}{N_{\mathrm{Tx}}\sigma_n^2 + P_r\sum_{j\neq i}[\boldsymbol{T}]_{i,j}\,[\boldsymbol{T}]^*_{i,j}}\,. \tag{2.44}$$

From this, the Shannon capacity of the MIMO channel is derived as $C = \log_2 \Pi_{i=1}^{N_{\mathrm{RF}}}(1+\mathrm{SINR}_i)$.

The spatial correlation matrices of the Kronecker model was generated under the assumption that the waves incoming from the different angles are uncorrelated, and also have the same power. Let $\boldsymbol{a}(\theta)$ be the steering vector of the array antenna for a plane wave coming from the direction of θ. The spatial correlation matrix is generated as

$$\boldsymbol{R} = \frac{1}{\Delta\theta}\int_{\theta_0-\Delta\theta/2}^{\theta_0+\Delta\theta/2}\boldsymbol{a}(\theta)\boldsymbol{a}(\theta)^{\mathcal{H}}\mathrm{d}\theta \tag{2.45}$$

where θ_0 is the center of the arrival wave and $\Delta\theta$ is the angular spread. We assume a linear array whose distance between elements is half of the wavelength. θ_0 is chosen randomly for 8 times, and the resultant estimation error was averaged over them. We adopt $\Delta\theta$ as the indicator of the degree of spatial correlation.

For each simulation parameter, an iteration of the Kalman filter is repeated 60 times, and the result is averaged over them. The average SNR is set to 15dB. The number of the Tx elements N_{Tx} is fixed to 6. Also the length of the training sequence N_t is set to 32.

2.5.1 Simulation Procedure

The simulation procedure is shown below.

1. Initialize the channel model parameters including the correlation matrices generated as (2.45).

2. For the first N_s fading blocks, this period is devoted to learn the channel statistics required for the Kalman filter. The estimation method is described in Appendix A.

3. Update the true channel state by (2.10).

4. From $\boldsymbol{P}_{k|k-1}$, optimal training sequence $\boldsymbol{S}_k$ is calculated by (2.27) and (2.34). In the case of the antenna selection system, $\boldsymbol{\xi}_k$ is chosen by (2.43).

5. Based on $\boldsymbol{\xi}_k$ and $\boldsymbol{S}_k$, channel is observed by (2.14).

6. Obtain the estimate of the channel by the Kalman filter. Also $\boldsymbol{P}_{k|k}$ in the filter is updated.

7. Evaluate the estimation error $\boldsymbol{H}_k - \widehat{\boldsymbol{H}}_k$.

8. Based on $\widehat{\boldsymbol{H}}_k$, the connection $\boldsymbol{\tau}_k$ is chosen by the fast antenna subset selection algorithm [5].

9. Evaluate the Shannon capacity of $\boldsymbol{A}_{\tau_k}\widehat{\boldsymbol{H}}_k$ by using (2.44).

10. In order to evaluate only the antenna selection capability, evaluate the capacity of $\boldsymbol{A}_{\tau_k}\boldsymbol{H}_k$ as well.

11. Return to step 3 until the required number of samples are obtained.

In step 2 of the procedure, parameters of the Gauss-Markov model shall be estimated by using the N_s times measurements of the channel. Since the period that channel parameters remain stationary is limited, we must learn its statistics as fast as possible. Therefore N_s should not be too large as long as the required precision is obtained. Considering them, we choose N_s as 10 for the estimation of the spatial correlation matrices. However, as for the temporal correlation coefficient ρ, an accurate estimate cannot be obtained until the N_s is large enough. Therefore for the estimation of ρ, we selected N_s as 500. In this case, the estimation error becomes approximately less than 0.01. Later, we will show that the channel estimation capability does not depend very much on the precision of the estimate of ρ. Therefore, by sacrificing the precision of ρ, there might be another way that can reduce the required N_s without losing estimation capability significantly.

2.5.2 Effect of the Spatial Correlation

Fig. 2.5 depicts the mean squared error defined as (2.24) with different spatial correlations for various temporal correlations. The antenna selection system is not employed, hence the equations of (2.29) to (2.32) are used. The solid lines and the dotted lines indicate the case of optimal training sequence design by the water-filling solution, and for the case of orthogonal training sequence (i.e. $\alpha_i^{(k)} = P_r N_t / N_{\mathrm{Tx}}$), respectively. We also compared them with the case of maximal likelihood (ML) channel estimation [48] [49] without any channel statistics. An advantage of the optimal training sequence design was observed especially under the presence of stronger spatial correlations of the Tx side and weaker temporal correlations. The reason of this seems that these conditions tend to yield nonuniform error variances among each fading block which work advantageously for the optimal training sequence design.

2.5.3 Effect of Correlation Model Mismatch

Although the proposed method assumes the Kronecker model which is applicable to NLOS environments, if the true channel obeys the LOS model, the model mismatch might cause a degradation of estimation capability. In order to verify this, we generated the true channel by using the typical LOS channel model which considers specular and diffuse components. The channel was generated as

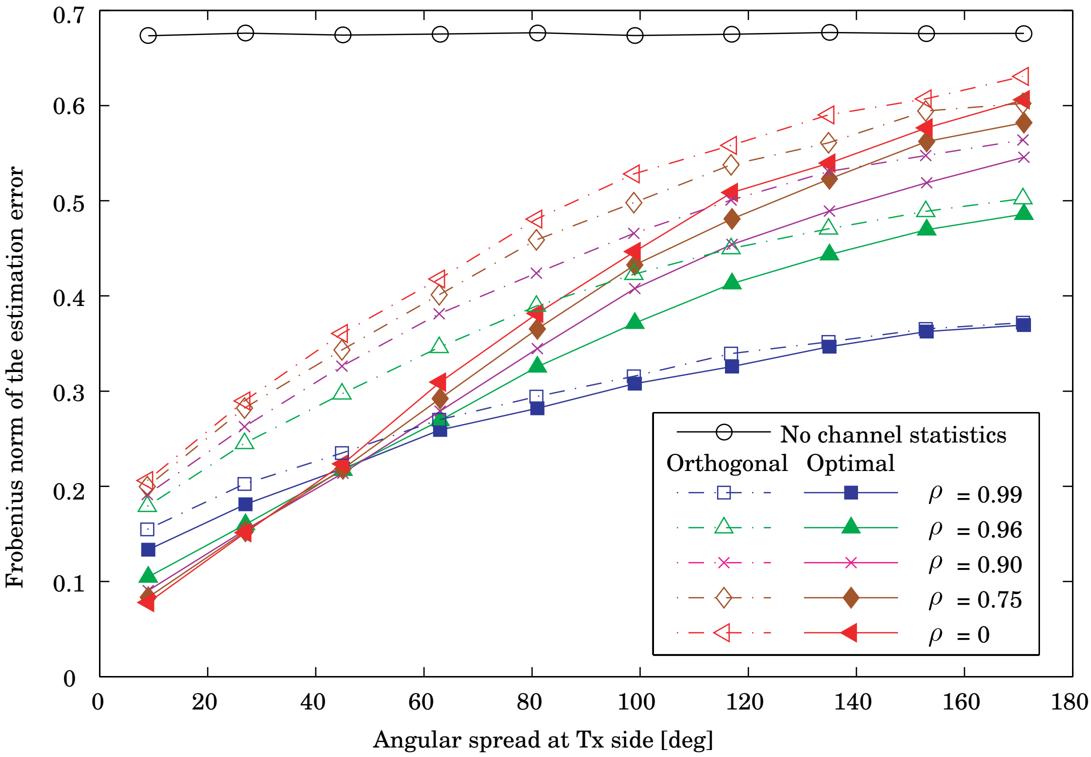

Figure 2.5: Mean squared error of the channel estimate for the different spatial and temporal correlations. ($N_{\mathrm{Rx}} = 6$)

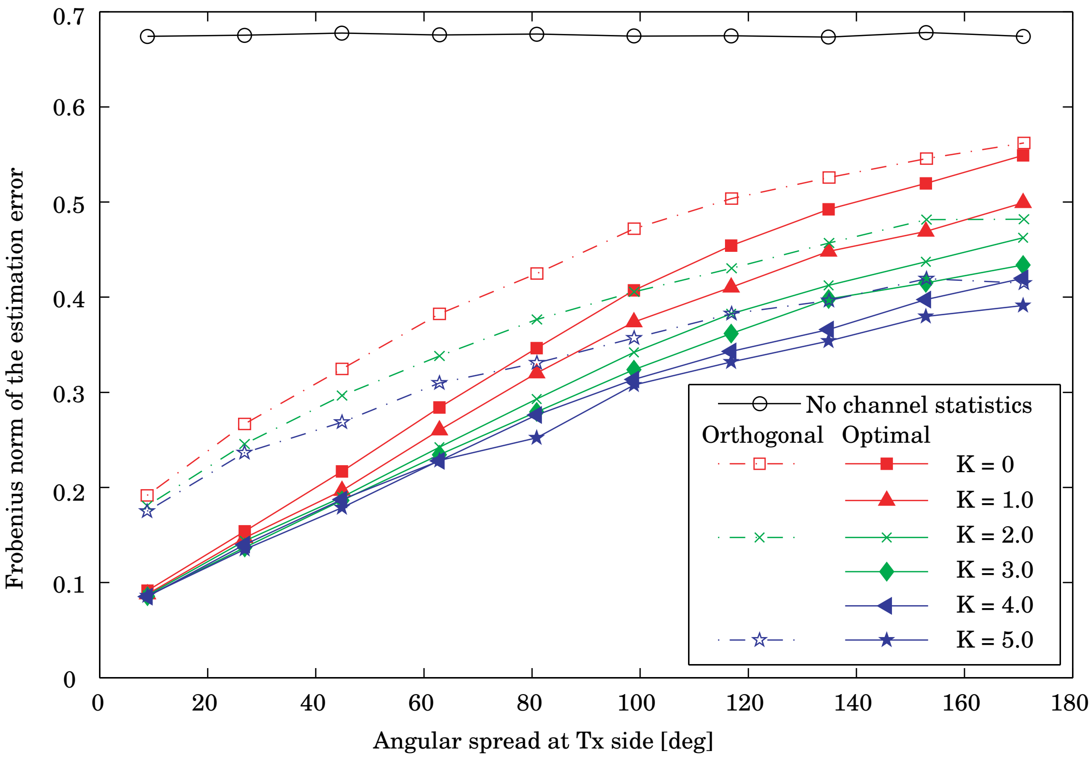

Figure 2.6: Mean squared error of the channel estimate for various Ricean K factors. ($N_{\mathrm{Rx}} = 6, \rho = 0.9, N_p = 4$)

$$\boldsymbol{H} = \sqrt{\frac{K}{K+1}}\sqrt{\frac{1}{N_p}}\sum_{i=1}^{N_p} g_i \cdot \boldsymbol{a}_{\mathrm{Rx}}(\theta_i^{(\mathrm{Rx})}) \left[\boldsymbol{a}_{\mathrm{Tx}}(\theta_i^{(\mathrm{Tx})})\right]^\top + \sqrt{\frac{1}{K+1}}\boldsymbol{R}_{\mathrm{DiffRx}}^{1/2}\boldsymbol{G}\left(\boldsymbol{R}_{\mathrm{DiffTx}}^{1/2}\right)^\top \quad (2.46)$$

where K is the Ricean K factor, N_p is the number of specular paths, and $\{g_i\}_{i=1}^{N_p} \sim \mathcal{CN}(0,1)$. $\boldsymbol{a}_{\mathrm{Tx}}$ and $\boldsymbol{a}_{\mathrm{Rx}}$ are steering vectors for each side. $\boldsymbol{R}_{\mathrm{DiffTx}}$ and $\boldsymbol{R}_{\mathrm{DiffRx}}$ are correlation matrices for diffuse component. The DOAs $\{\theta_i^{(\mathrm{Tx})}\}_{i=1}^{N_p}$ and $\{\theta_i^{(\mathrm{Rx})}\}_{i=1}^{N_p}$ are generated randomly within the range of the angular spread.

For various K factors, estimation error was evaluated in Fig. 2.6 using the optimal training sequence design. Since the diffuse component was modeled as the Kronecker model, K = 0 means the model matches exactly to our original assumption. Contrary to our expectations, the method works more effectively as the specular component beging stronger. It is considered that the presence of specular component yields strong spatial correlation which is superior to the proposed method rather than inferior due to the model mismatch.

2.5.4 Capacity Tracking Capability with the Time Elapse

Fig. 2.7 depicts the time varying channel capacity for each fading block.

We have compared the capacity of the following four cases; 1) Selection based on perfect CSI. 2) Selection based on the estimated CSI. 3) In addition to the previous case, also the transmissions are done by using the estimated CSI. 4) Random selection with perfect CSI.

The solid line indicates the capacity of the pefect channel matrix. The fluctuation of the capacity increases as the correletion coefficient ρ decreases. We can see that a nearly optimal selection is achieved by using the estimated channel as the selection criteria. But due to the estimation error caused by partial channel measurement, the capacity using the estimated channel degrades a lot. This result agrees with the description in [15] that the near optimal selection can be achieved even if the precision of the channel estimate is poor.

2.5.5 Effect of the Temporal Correlation

From now on, we employ the antenna selection system of $N_{\mathrm{Rx}} = 7$ and $N_{\mathrm{RF}} = 4$ or 6. The channel estimation is done by the equations of (2.36) to (2.38). The legend 'Kalman for transmission' in figures 2.8 and 2.9 show the degraded capacity with the change of temporal correlation of the actual channel. From the graphs, by utilizing the estimated channel by the Kalman filter, almost perfect selection can be achieved, but slight selection error is observed especially for the lower correlations. The legend 'Kalman for only selection' in the figures evaluate only the antenna selection capability of the estimated channel. If we directly employ the estimated channel for the transmission, the capacity is highly influenced by the temporal correlation. We also show the result obtained by the ML channel estimation without exploiting any channel statistics. In order to perform comparison under fair conditions, we imposed the condition where the received signal power of the training sequence is equal among all cases. In the case of $N_{\mathrm{RF}} = 6$, estimation by Kalman filter always exceeds the one without channel statistics. As for the case of $N_{\mathrm{RF}} = 4$, since almost half of the channel state is not observed directly, the estimation error becomes significant. We can read the large degradations with the correlations below 0.95 mainly caused by the leaks between sub channels.

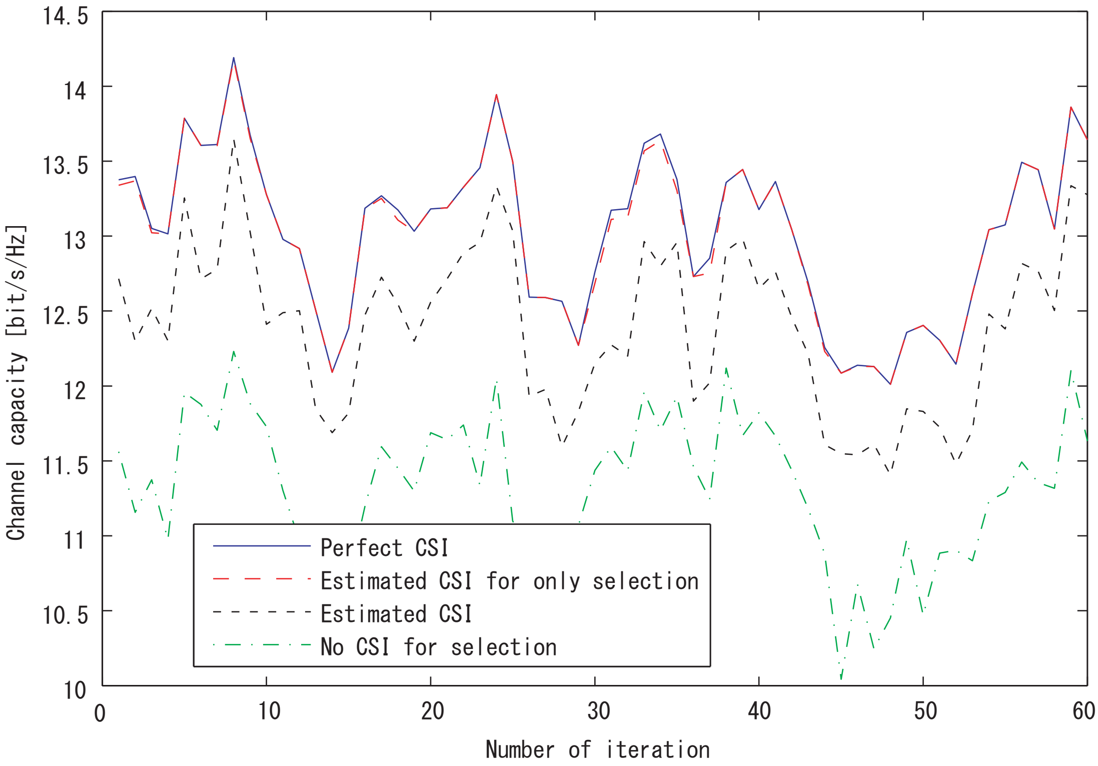

Figure 2.7: Time variation of the capacity. ($N_{\mathrm{Rx}} = 7, \rho = 0.97$,angular spread is 100[deg])

2.5.6 Effect of the Measurement Antenna Selection

We investigated the effect of the optimal antenna selection for the channel measurement. Fig. 2.10 shows the mean squared error of the channel matrix for different spatial correlations of the Rx side. For comparison, we also show the case of random selection which means $\boldsymbol{\xi}_k$ is determined randomly for each iteration.

The estimation quality of the optimal selection is always higher than other methods. Difference in estimation error becomes large when the spatial correlations are small, and also when the number of the unmeasured elements ($N_{\mathrm{Rx}} - N_{\mathrm{RF}}$) is large.

2.5.7 Robustness of the Estimate of ρ

Estimation error of the channel statistics causes the degradation of the channel estimation capability. In Fig. 2.11, the channel is generated with the temporal correlation coefficient being fixed to 0.98, although the estimator assumes its correlation coefficient as a value on the horizontal axis. The capacity with the estimated CSI seems not to be degraded even when the estimation error increases to 0.07. We can observe the robustness of the estimation of channel statistics, but from the discussions so far, it is shown that the estimation is sensitive to the actual temporal correlation itself.

2.6 Concluding Remarks

We proposed the optimal training sequence design for channel estimation by the Kalman filter. Also in order to reduce the measurement costs for antenna selection systems, partial channel

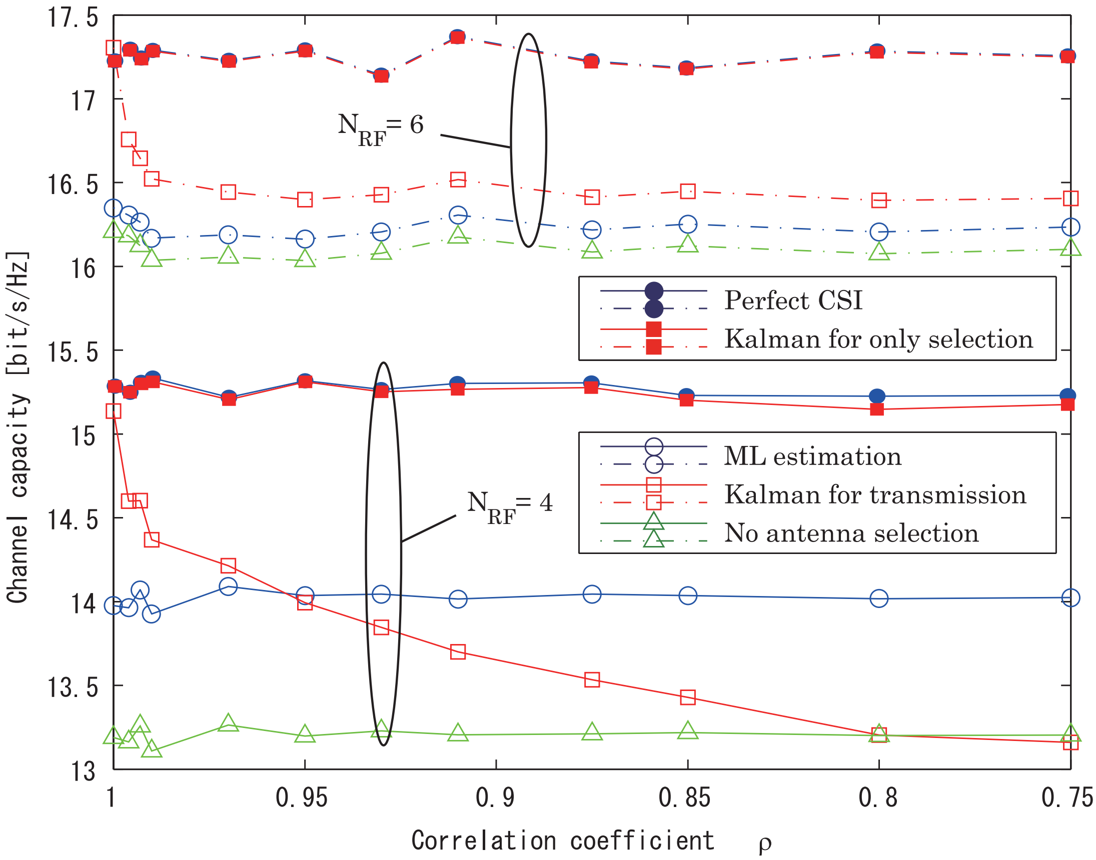

Figure 2.8: Capacity degradation caused by the leaks between sub channels in addition to the antenna selection capability. ($N_{\mathrm{Rx}} = 7$, SNR is 15dB, angular spread is 60[deg])

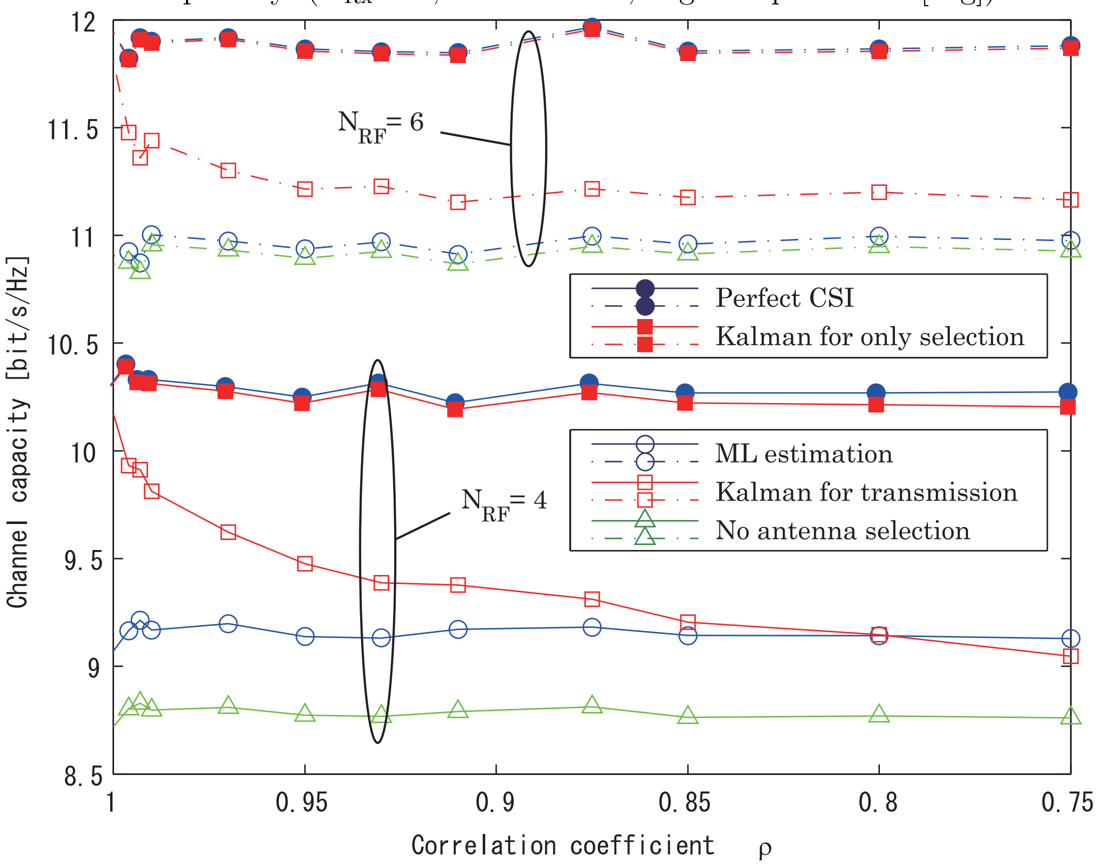

Figure 2.9: Capacity degradation caused by the leaks between sub channels in addition to the antenna selection capability. ($N_{\mathrm{Rx}} = 7$, SNR is 10dB, angular spread is 60[deg])

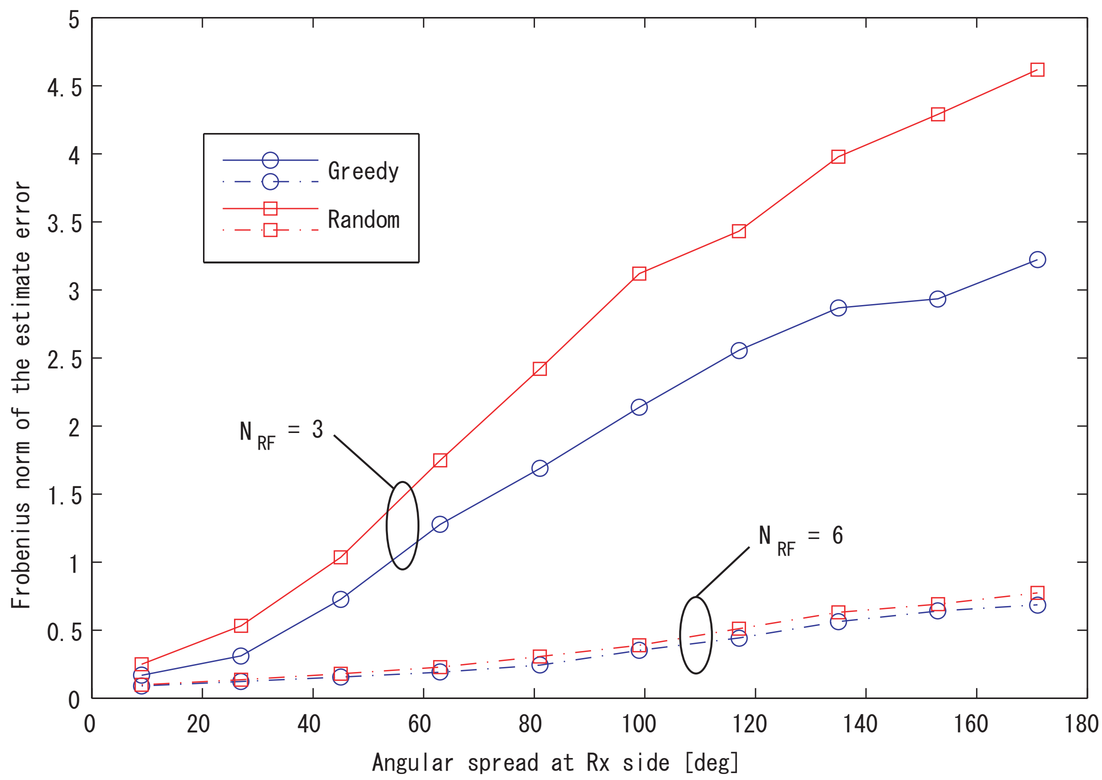

Figure 2.10: Comparison of the selection methods for channel measurement. ($N_{\mathrm{Rx}} = 7, N_{\mathrm{Tx}} = 3, \rho = 0.93$)

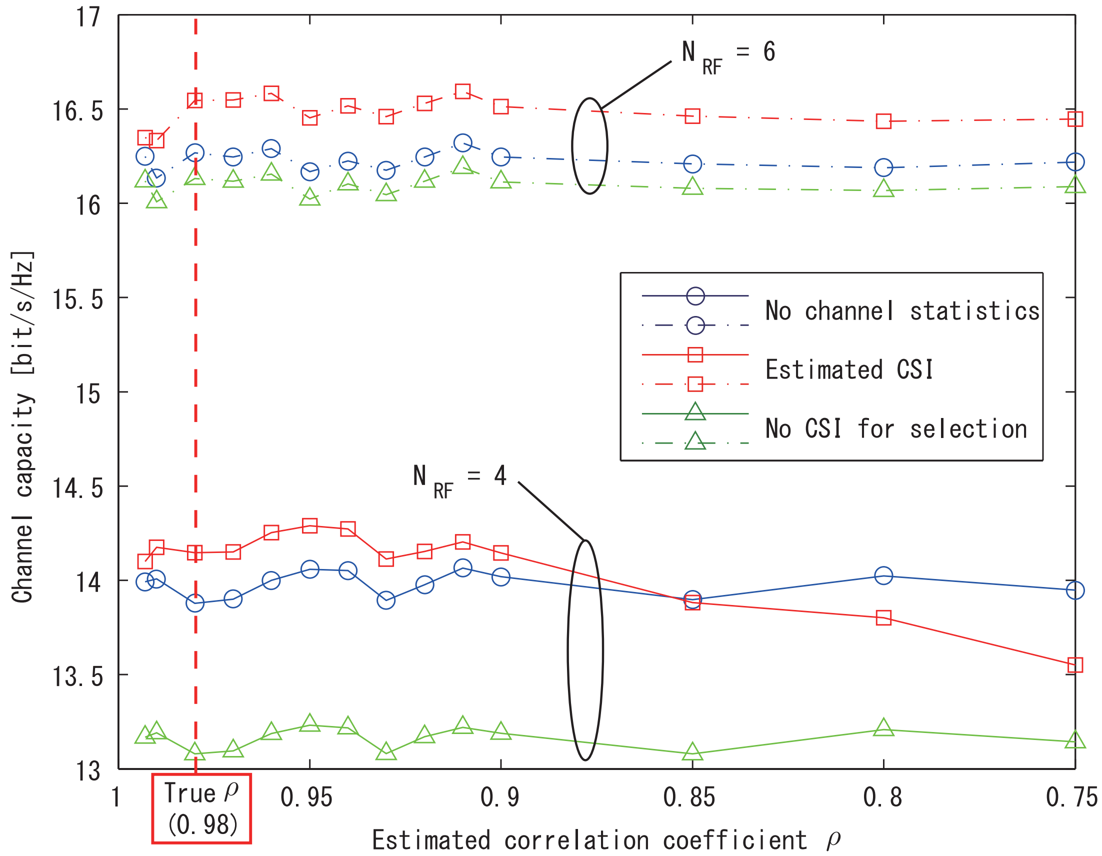

Figure 2.11: Effect of the estimation error of the temporal correlation coefficient. A true coefficient for channel generation is 0.98.

measurement and its optimal control based on the statistics were discussed.

The numerical simulation assuming the Gauss-Markov channel revealed that the optimal training sequence design works effectively in the case of lower temporal correlation and higher spatial correlation of the Tx side, although the partial channel measurement is superior for higher temporal correlations and higher spatial correlations on the Rx side. If partial channel measurement was employed, the estimated channel is accurate enough as the coarse estimation which is utilized for only the selection criteria. Meanwhile, if we utilize the estimated channel to the transmission directly, the estimation error causes the degradation of the channel capacity to a large extent. In that case, it was shown that high spatial and temporal correlations are required to keep the accuracy of the method. Also the simulation revealed that the estimation capability does not suffer much degradation due to the estiamtion error of the temporal correlation statistics of the channel. This implies that the method has an error tolerance about temporal fluctuations that sometimes arise in actual propagation environments.

As future tasks, this method has a shortage that comes from the definition of the channel model about frequent temporal change of spatial correlation statistics. This might not be negligible especially in rapidly mobile wireless scenarios. Therefore, now we are investigating a more robust estimation scheme for antenna selection systems which have less dependence on the spatial correlations.

Chapter 3

EM-based Estimation of Spatiotemporal Correlation Statistics

In this chapter, we discussed a MIMO channel estimation method which exploits the channel's spatiotemporal correlation without the aid of a priori channel statistical information. A random-walk based Gauss-Markov model which has less parameters to be estimated is presented for the Kalman filter. In order to obtain the channel's innovation statistical parameter, considering that the time evolution of the channel is a latent statistical variable, the EM algorithm is applied for accurate estimation.Numerical simulations revealed that the method is able to enhance estimation capability by exploiting spatiotemporal correlations, and works well under the conditions where small forgetting factor employed.

3.1 Introduction

Exploitation of channel statistics for MIMO channel estimation is a challenging problem. Under certain stationary conditions, MIMO channel statistics provide reliable information based on physical meanings. For instance, spatial correlation statistics can be fully characterized by the distribution of surrounding principal scatterers in the propagation channel [26]. So far, many literatures [18–38, 44] reported that exploiting channel correlations yield significant enhancement of channel estimation capability. For the channel estimation, the MMSE criterion is often used, and the estimation methods are classified roughly into the two groups. One is the Wiener based methods [23–30] which only consider spatial or frequency correlations within a symbol. The other is the Kalman based methods [18–22] which additionally deal with inherent channel correlations between symbols. The Kalman based methods can take into account a degree of time evolution of the channel, but requires more computational cost.

Under the presence of a priori correlation statistics, or assumption that the channel obeys typical correlation model such as the Bessel function of the Rayleigh fading model, the optimal estimation method and detailed performance analysis have been discussed [22–24, 26–30]. In addition, the estimation capability can be further improved by designing the optimal training sequences considering the correlation statistics [22, 26–30]. In order to obtain benefit from the correlation statistics based on the methods above, a method to obtain accurate channel statistics is necessary, since a mismatch of the statistics cause degradation of estimation capability.

Regarding how to obtain the correlation statistics, many discussions have been made assuming various scenarios. The simplest way is to allocate a training period preceding data transmissions, and estimate by batch process [20–22]. A training sequence can be scattered pilots in OFDM systems [25]. In realistic situations, a method to stably-exploit the MIMO channel statistics is still not established. A major problem is difficulty in obtaining reliable statistics within non-stationary wireless channels where the mobile station does not remain a constant velocity, and moves irregularly through the scatterers. One of the methods to manage non-stationary is to use the RLS (Recursive Least Squares) criterion which can continuously alter the correlation statistics [31–34]. Even when using the RLS, there still exists a problem to balance tracking capability and estimation precision by controlling a forgetting factor. Other approaches have been proposed to establish a robust estimation method which works even under the existence of mismatch between assumed and actual channel statistics [35–37]. The authors have proposed a robust channel estimation method dedicated to the antenna selection systems which does not rely on unreliable spatial correlation statistics but only rough temporal correlation statistics [38].

In this paper, we address the issue of MIMO channel estimation including an estimation of channel statistics assuming non-stationary channels. Our method is devoted to obtain the channel's innovation statistics of the Kalman filter. The same problem is discussed in [20, 21] by utilizing batch estimation process using the training period, and incorporation of forgetting factor is remarked. We extend the estimation method considering the following point of view.

Generally, in order to obtain channel statistics accurately, many measurements of channel realizations are necessary. Originally however, if we could measure channel states accurately for each channel measurement, there is no demand for the aid of channel statistics. Inherently with limited set of resources, a true channel state is unable to be referenced directly, but it can be only indirectly predicted as a stochastic distribution due to the received training sequences. In addition, since the channel statistics do not remain constant, an interval update for tracking is required. In a strict sense, if we utilize the periodically updated channel statistics for the estimation of channel, the estimated channel state is not correct since the estimation is based on somewhat incorrect channel statistics which was also derived from incorrectly estimated channel state. In such way, utilizing channel statistics in a non-stationary environment has the potential to fall into a vicious cycle. Considering the circumstances, based on the formulation that we cannot measure the channel state directly, we developed a novel channel estimation method via the EM-algorithm. In the proposed method, the channel statistics are calculated based on a posterior distribution of the channel state after having observed the training sequences, which enables an estimation of the channel statistics where the channel state is not directly observable (or even partially not observable at extreme cases) as described above. We show that the channel estimation can be enhanced without a priori channel statistics by utilizing the proposed method.

As far as we know, this is the first attempt to estimate the channel's innovation statistics by regarding the estimated channel evolution as a latent statistical variable. Note that there has been many studies [50–53] for the purpose of channel state estimation by means of the EM algorithm. Though the same algorithm is utilized, our usage of the EM algorithm is different from theirs. These preceding studies deal with simultaneous estimation of both the channel state estimation and the data sequence estimation by regarding either of them as a latent variable. Intuitively, these methods iteratively calculate a posterior distribution of the transmitted data sequences, and then estimate the channel state so that its components along the directions having a large variance of the corresponding data sequences become small values since the estimated channel components corresponding to the unreliable data sequences are also unreliable. As just described, the preceding studies do not deal with the estimation of channel statistics though a formulation considering a latent channel state is similar. The differences of these EM-based channel estimation methods are summarized in Table 3.1 and Fig.3.1. These simultaneous estimations are often combined with DFE(Decision Feedback Equalizer) [54–57],

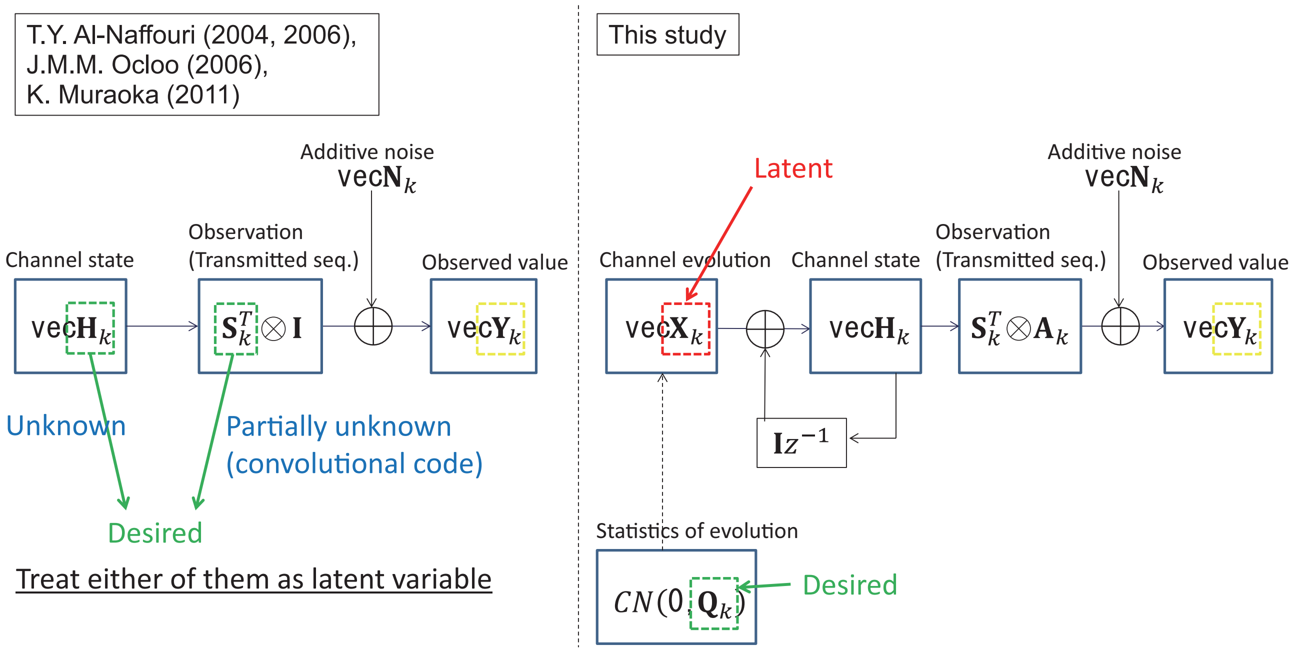

Figure 3.1: The difference of scheme with preceding studies about channel estimation using EM algorithm.

and a priori channel statistics such as the WSSUS assumption are often assumed. In this paper, we do not mention about the DFE, but exploitation of correlation information onto the DFE scheme is considered to be a future problem.

Literature	Time evolution model	Spatial correlation statistics	Frequency correlation statistics	Temporal correlation statistics	Training
This study	Gauss-Markov	Perform estimation	None (Narrowband)	Perform estimation	Known pilots
[50]	Gauss-Markov	None (Not MIMO)	a priori innovation statistics	a priori correlation coefficient	Initial known pilots and unknown data
[51]	Gauss-Markov	None (Not MIMO)	i.i.d. innovation statistics and a priori correlations	a priori correlation coefficient	Unknown data
[52]	Gauss-Markov	a priori innovation statistics	a priori innovation statistics	a priori correlation coefficient	Initial known pilots and unknown data
[53]	Gauss-Markov with inner differential states	None (Not MIMO)	a priori innovation statistics	a priori correlation coefficient	Initial known pilots and unknown data

Table 3.1: The studies on channel estimation via the EM algorithm.

This paper is organized as follows. In section 2, the system model including the stochastic channel model considering spatial and temporal correlation is explained. Section 3 shows the channel estimation method by the Kalman filter assuming that a priori channel statistics are available. Section 4 explains the proposed statistics estimation. Section 5 shows simulation results and considerations.

3.1.1 Mathematical Notations

Throughout this paper we will use bold-faced upper case letters to denote matrices, and bold-faced lower case letters for column vectors, light-faced letters for scalar quantities. The subscripts $\top$, $\mathcal{H}$, $*$ indicate transpose, Hermitian transpose (transpose and complex conjugate), and complex conjugate respectively. $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix. Also the inverse, Moore-Penrose pseudo inverse, trace, determinant, and Frobenius norm of the matrix $\boldsymbol{X}$ are denoted by $\boldsymbol{X}^{-1}$, $\boldsymbol{X}^{\dagger}$, $\operatorname{tr}\boldsymbol{X}$, $\det\boldsymbol{X}$, and $\|\boldsymbol{X}\|_F$, respectively. The m-th row and n-th column element of the matrix $\boldsymbol{X}$ is denoted by $[\boldsymbol{X}]_{m,n}$. $\mathbb{E}_x$ means the expectation with respect to x. The bracket $< \cdot,\cdot >$ is used for inner product on the column vector space. Also the vector Euclidean norm is expressed as $\|\cdot\|$. In this paper, since we often discuss correlations between each matrix element, it is convenient to treat matrix as one column vector that consists of all its elements. For any $m \times n$ matrix $\boldsymbol{A} = [\boldsymbol{a}_1\, \boldsymbol{a}_2\, \cdots \boldsymbol{a}_n]$, the vec operator generates a $mn \times 1$ vector defined as

$$\operatorname{vec}\boldsymbol{A} \triangleq \left[\boldsymbol{a}_1^\top\, \boldsymbol{a}_2^\top\, \cdots \boldsymbol{a}_n^\top\right]^\top \tag{3.1}$$

where $\triangleq$ means definition. The Kronecker product $\otimes$ is required with the use of the vec operator.

3.2 System Model

3.2.1 MIMO Channel Model

For a narrowband MIMO channel with N_{Tx} transmit antennas and N_{Rx} receive antennas, received vector $\boldsymbol{y} \in \mathbb{C}^{N_{\mathrm{Rx}}}$ can be expressed as

$$\boldsymbol{y} = \sqrt{\frac{P_r}{N_{\mathrm{Tx}}}}\boldsymbol{H}_k\boldsymbol{x} + \boldsymbol{n} \tag{3.2}$$

where P_r is average receive signal power at each receive antenna. $\boldsymbol{n} \in \mathbb{C}^{N_{\mathrm{Rx}}}$ is the additive noise vector typically assumed to have a white complex normal distribution with average power σ_n^2. $\boldsymbol{x} \in \mathbb{C}^{N_{\mathrm{Tx}}}$ is the normalized transmit vector such that $\mathbb{E}_{\boldsymbol{x}}\, \boldsymbol{x}\boldsymbol{x}^{\mathcal{H}} = \boldsymbol{I}_{N_{\mathrm{Tx}}}$. $\boldsymbol{H}_k \in \mathbb{C}^{N_{\mathrm{Rx}} \times N_{\mathrm{Tx}}}$ is the normalized channel matrix for time instant k. Since $\boldsymbol{H}_k$ varies for each time instant k, we define the normalization as $\mathbb{E}_k \|\boldsymbol{H}_k\|_F^2 = N_{\mathrm{Tx}}N_{\mathrm{Rx}}$. In this formulation, average signal to noise ratio (SNR) per receive antenna can be expressed as follows:

$$\mathrm{SNR} = \frac{\mathbb{E}_{\boldsymbol{H}_k}\,\mathbb{E}_{\boldsymbol{x}}\left\|\sqrt{\frac{P_r}{N_{\mathrm{Tx}}}}\boldsymbol{H}_k\boldsymbol{x}\right\|^2}{\mathbb{E}_{\boldsymbol{n}}\,\|\boldsymbol{n}\|^2} = \frac{P_r}{\sigma_n^2} \tag{3.3}$$

3.2.2 Stochastic Model for Time-Varying Correlated MIMO Channel

In order to exploit inherent correlation statistics of propagation channels, first we investigate a robust channel model which employs both spatial and temporal correlations simultaneously while not increasing the number of model parameters as far as possible. In the subsequent section, we consider a concrete estimation method based on the channel model introduced in this section.

We consider quasi-static block fading channel where the temporal channel evolution is small enough to be ignored during one transmit block. Let us denote the channel state in the k-th fading block by $\boldsymbol{H}_k$. A temporal transition assuming a wide-sense stationary random process is often modeled as the AR(Auto Regressive) model [39]. In MIMO case, the time evolution is expressed as

$$\begin{aligned}\operatorname{vec}\boldsymbol{H}_k &= \sum_{i=1}^{M}\boldsymbol{G}_i\operatorname{vec}\boldsymbol{H}_{k-i} + \boldsymbol{\psi}_k \\ \boldsymbol{\psi}_k &\sim \mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{\Psi}_k\right)\end{aligned} \tag{3.4}$$

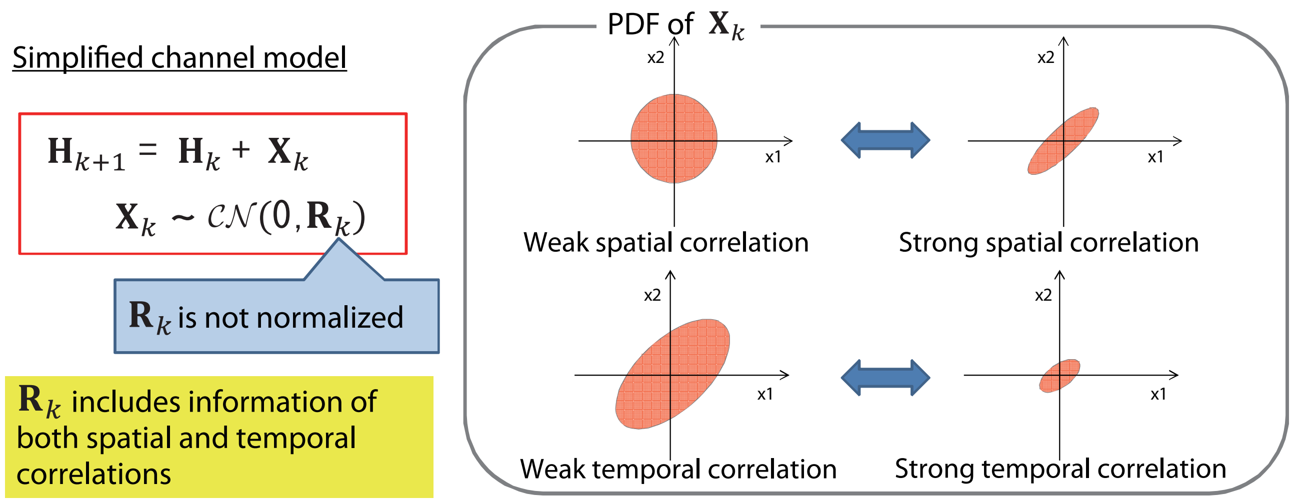

Figure 3.2: A brief explanation of why the proposed simplified random-walk based Gauss-Markov model is able to exploit both temporal and spatial correlation statistics.

where $\mathcal{CN}(\mathbf{0}, \mathbf{\Psi}_k)$ denotes the circularly-symmetric complex normal distribution with covariance of $\mathbf{\Psi}_k$. The tap number M is chosen depending on the shape of the power spectral density of time evolution. Considering the points that the number of model parameters should be small, and applicability for fast fading channels, we choose $M = 1$ and also $\boldsymbol{G}_1 = \rho \boldsymbol{I}$ where $\rho (0 \leq \rho \leq 1)$ denotes the correlation coefficient which depends on the maximum Doppler frequency and period of fading block. Such a first order AR model is widely used among the literatures [19–22, 54, 57] dealing with the similar problem. In the case of a non-stationary channel, since the correlation coefficient varies depending on time, we rewrite the channel model as:

$$\mathrm{vec}\, \boldsymbol{H}_k = \rho_k \,\mathrm{vec}\, \boldsymbol{H}_{k-i} + \boldsymbol{\psi}_k \tag{3.5}$$

The above model requires the parameters of both ρ_k and $\mathbf{\Psi}_k$. However, estimating both of them accurately using an insufficient number of channel realizations are generally difficult, since the estimation becomes an ill-conditioned problem, which can be illustrated by the fact that an estimation error of ρ_k can be absorbed into a choice of $\mathbf{\Phi}_k$. In order to avoid this matter while further reducing parameters, we adopt the following model.

$$\mathrm{vec}\, \boldsymbol{H}_k = \mathrm{vec}\, \boldsymbol{H}_{k-1} + \underbrace{(\rho_k - 1)\,\mathrm{vec}\, \boldsymbol{H}_{k-1} + \boldsymbol{\psi}_k}_{\boldsymbol{z}_k} \tag{3.6}$$
$$\boldsymbol{z}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{R}_k)$$

Now, the desired model parameters to be determined is reduced to only $\boldsymbol{R}_k$ which is a statistical parameter of first order differential value of channel. Note that we do not impose any normalization on $\boldsymbol{R}_k$, since absolute magnitude of its eigenvalues have meanings; tr $\boldsymbol{R}_k$ indicates strength of temporal correlation, and also unevenness of eigen values of $\boldsymbol{R}_k$ indicates strength of spatial correlation. Uniform eigenvalues ($\boldsymbol{R}_k = \alpha_k \boldsymbol{I}$) means that there exists no spatial correlations exploitable. (illustrated in Fig.3.2)

It might be pointed out that a simple random walk model of (3.6) would be incapable of expressing characteristics of true channel. In the random walk model, channel variance is simply accumulated as time elapsed, which suggests that expected SNR will monotonically increase,

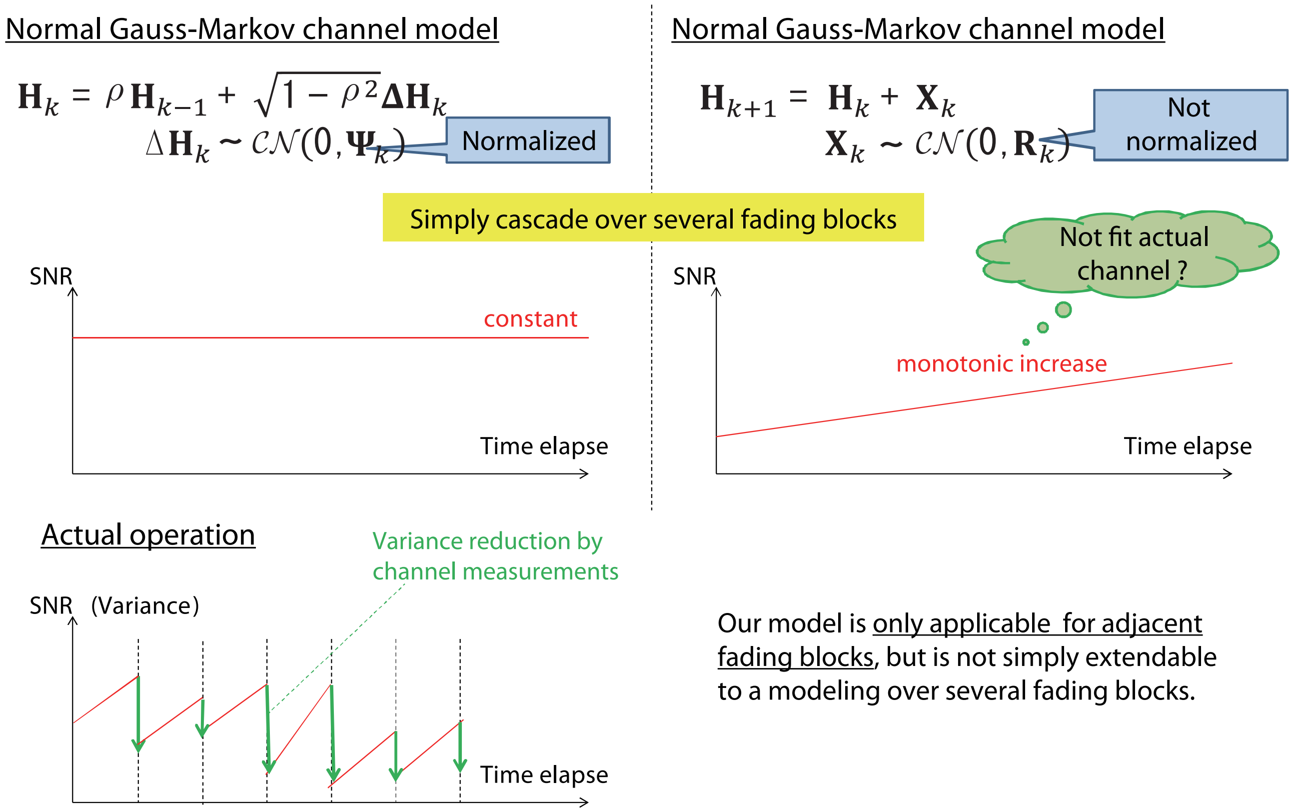

Figure 3.3: The stationarity of the proposed simplified random-walk based Gauss-Markov model when used in cascade connection.

thus the model does not belong to WSS. However, in actual operation, we assume that channel measurement is done for each time instant, and it decreases channel variances thus prevents SNR increase. In short, our model is only applicable for adjacent fading blocks, but is not simply extendable to a modeling over several fading blocks. (illustrated in Fig.3.3)

3.3 Channel Estimation for Gauss-Markov Channel

3.3.1 Channel Observation Model

Let $\boldsymbol{s}_1^{(k)}, \boldsymbol{s}_2^{(k)}, \cdots, \boldsymbol{s}_{N_t}^{(k)} \in \mathbb{C}^{N_{\mathrm{Tx}}}$ be normalized training sequences, each of them being launched from Tx side in number order at fading block k. They are normalized such that $\|\widetilde{\boldsymbol{S}}_k\|_F^2 = N_t N_{\mathrm{Tx}}$ where $\widetilde{\boldsymbol{S}}_k \triangleq [\boldsymbol{s}_1^{(k)}\ \boldsymbol{s}_2^{(k)}\ \cdots\ \boldsymbol{s}_{N_t}^{(k)}]$. Then, the received sequences are written as

$$[\boldsymbol{y}_1^{(k)}\ \boldsymbol{y}_2^{(k)}\ \cdots\ \boldsymbol{y}_{N_t}^{(k)}] = \sqrt{\frac{P_r}{N_{\mathrm{Tx}}}} \boldsymbol{H}_k [\boldsymbol{s}_1^{(k)}\ \boldsymbol{s}_2^{(k)}\ \cdots\ \boldsymbol{s}_{N_t}^{(k)}] + [\boldsymbol{n}_1\ \boldsymbol{n}_2\ \cdots\ \boldsymbol{n}_{N_t}] \tag{3.7}$$

where $\boldsymbol{n}_i \in \mathbb{C}^{N_{\mathrm{Rx}}}$ is the i-th additive noise vector and $\boldsymbol{y}_i^{(k)}$ is the received vector corresponding to the transmitted vector $\boldsymbol{s}_i^{(k)}$.

If the second order statistic of $\boldsymbol{H}_k$ is known in advance, optimal training sequences are determined uniquely according to the literatures [22, 26]. To simplify problems, we adopt the orthogonal training matrix defined as $\widetilde{\boldsymbol{S}}_k \widetilde{\boldsymbol{S}}_k^{\mathcal{H}} = N_t \boldsymbol{I}_{N_{\mathrm{Tx}}}$ from now on.

By letting $\boldsymbol{Y}_k$ and $\boldsymbol{N}_k$ be $N_{\mathrm{RF}} \times N_t$ matrices, each of them consisting of received vectors and noise vectors, respectively, (3.7) is rewritten as

$$\boldsymbol{Y}_k = \boldsymbol{H}_k \boldsymbol{S}_k + \boldsymbol{N}_k \tag{3.8}$$

where $\boldsymbol{S}_k$ is defined as $\boldsymbol{S}_k \triangleq \sqrt{P_r/N_{\mathrm{Tx}}}\widetilde{\boldsymbol{S}}_k$ and hence $\|\boldsymbol{S}_k\|_F^2 = P_r N_t$ holds. Applying vec operator on both sides of the above equation yields the linear observation model as

$$\operatorname{vec} \boldsymbol{Y}_k = \boldsymbol{B}_k \operatorname{vec} \boldsymbol{H}_k + \operatorname{vec} \boldsymbol{N}_k \tag{3.9}$$

where $\boldsymbol{B}_k \triangleq \boldsymbol{S}_k^{\top} \otimes \boldsymbol{I}_{N_{\mathrm{Tx}}N_{\mathrm{Rx}}}$ is an observation matrix, and an actual channel state is observed through this linear mapping.

3.3.2 Channel Estimation by the Kalman Filter

Under the MMSE condition, best linear estimate of channel state for each time instant k can be iteratively obtained by the Kalman filter [17] when the true channel state obeys Gauss-Markov model. Based on the channel's time evolution model defined as (3.6), updating formula of the Kalman filter is expressed as follows:

$$\boldsymbol{P}_{k|k-1} = \boldsymbol{P}_{k-1|k-1} + \boldsymbol{R}_k \tag{3.10}$$

$$\boldsymbol{\delta}_k = \operatorname{vec} \boldsymbol{Y}_k - \boldsymbol{B}_k \operatorname{vec} \widehat{\boldsymbol{H}}_{k-1} \tag{3.11}$$

$$\boldsymbol{K}_k = \boldsymbol{P}_{k|k-1} \boldsymbol{B}_k^{\mathcal{H}} \left(\boldsymbol{B}_k \boldsymbol{P}_{k|k-1} \boldsymbol{B}_k^{\mathcal{H}} + \boldsymbol{Q}_k\right)^{-1} \tag{3.12}$$

$$\hat{\boldsymbol{z}}_k = \boldsymbol{K}_k \boldsymbol{\delta}_k \tag{3.13}$$

$$\operatorname{vec} \widehat{\boldsymbol{H}}_k = \operatorname{vec} \widehat{\boldsymbol{H}}_{k-1} + \hat{\boldsymbol{z}}_k \tag{3.14}$$

$$\boldsymbol{P}_{k|k} = \left[\boldsymbol{I}_{N_{\mathrm{Tx}}N_{\mathrm{Rx}}} - \boldsymbol{K}_k \boldsymbol{B}_k\right] \boldsymbol{P}_{k|k-1} \tag{3.15}$$

$\widehat{\boldsymbol{H}}_k$ is the estimated channel matrix at time k, and $\boldsymbol{Q}_k$ is the observation noise covariance matrix defined as $\boldsymbol{Q}_k \triangleq \mathbb{E}_{\boldsymbol{N}_k} \operatorname{vec} \boldsymbol{N}_k (\operatorname{vec} \boldsymbol{N}_k)^{\mathcal{H}}$. $\hat{\boldsymbol{z}}_k$ is an estimate of latent time evolution of channel on which we will focus in later discussion. $\boldsymbol{K}_k$ is so-called the Kalman gain which restores channel evolution from a difference of channel observations. $\boldsymbol{P}_{k|k}$ and $\boldsymbol{P}_{k|k-1}$ are error covariance matrices defined as

$$\boldsymbol{P}_{k|l} \triangleq \mathop{\mathbb{E}}_{\boldsymbol{n}} \mathop{\mathbb{E}}_{\boldsymbol{H}_k} \operatorname{vec}\left(\boldsymbol{H}_k - \widehat{\boldsymbol{H}}_l\right) \left[\operatorname{vec}\left(\boldsymbol{H}_k - \widehat{\boldsymbol{H}}_l\right)\right]^{\mathcal{H}} \tag{3.16}$$

where $l \in \{k, k-1\}$.

3.4 Recursive Estimation of Channel Statistics

In this section, we discuss a method for estimating covariance matrix $\boldsymbol{R}_k$ from its earlier observations $\{\boldsymbol{Y}_k\}_{k=1}^{N}$. Preceded by the problem formulation, we show the two methods for solving the problem. One is a naive and simple method with lower computational cost, and the other is an EM-based estimation method.

3.4.1 ML Estimation Criteria

We formulate estimation problem for time-varying statistical model parameters in terms of the maximum-likelihood (ML) estimation. Though the resulting method is only an exponentially weighted averaging which is seen in derivation of the RLS filter, our purpose is to incorporate an essence of the RLS into our estimation scheme via framework of the ML estimation. Also we consider the case of constant model parameters as comprehensible example.

Case of $\boldsymbol{R}_k$ Remains Constant

In this case, we suppress the subscript k of $\boldsymbol{R}_k$. Assuming that $\{\boldsymbol{z}_i\}_{i=1}^{N}$ are independent and zero mean, we have:

$$L(\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_N \mid \boldsymbol{R}) = \log \prod_{i=1}^{N} \mathcal{CN}\left(\boldsymbol{z}_i \mid \boldsymbol{0}, \boldsymbol{R}\right) \tag{3.17}$$

Here the problem is how to obtain the realizations of channel evolution $\boldsymbol{z}_i$ which are not directly observable. One of the simplest way for this is to use estimated value $\hat{\boldsymbol{z}}_i$ of (3.14) instead. It should be noted that this convenient substitution is not accurate, and our proposed method we discuss later is an improvement concerning this problem.

Maximization of the above equation by using $\hat{\boldsymbol{z}}_i$ is achieved by its covariance matrix as follows:

$$\widehat{\boldsymbol{R}} = \frac{1}{N} \sum_{i=1}^{N} \hat{\boldsymbol{z}}_i \hat{\boldsymbol{z}}_i^{\mathcal{H}} \tag{3.18}$$

where N denotes a number of realizations available.

Case of $\boldsymbol{R}_k$ Varies Depending on k

Generally for estimation of time varying statistical parameters, we attach more importance to recently occurred realizations. Since $\{\boldsymbol{z}_i\}_{i=1}^N$ are assumed to be independent having zero mean, our concern is to adjust importance weights for each realization in (3.17). In this paper, we formulate the following exponentially weighted likelihood function so that its maximization result matches an exponentially weighted averaging used in derivation of the RLS algorithm in order to reduce computations for maximization. For estimation at time instant $k\,(N \geq k)$, the function is expressed as:

$$L(\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_N \,|\, \boldsymbol{R}_k) = \log \prod_{i=1}^{k} \mathcal{CN}\left(\boldsymbol{z}_i \,|\, \boldsymbol{0}, \boldsymbol{R}_k\right)^{\lambda^{k-i}} \tag{3.19}$$

where λ is so-called forgetting factor which is a given hyperparameter that indicates variation speed of channel statistics. If temporal variation of channel statistics is small, larger λ is suited. For faster variation, small λ is suited but too small λ leads to lose accuracy of parameters due to shortage of a number of realizations having large importance. In this paper, we do not discuss estimation algorithm of the optimal λ.

Similarly discussed in the above section, if we employ $\{\hat{\boldsymbol{z}}_i\}_{i=1}^N$ instead of actual realizations, the parameter maximizing above function can be obtained recursively as the following equation:

$$\widehat{\boldsymbol{R}}_k = \lambda \widehat{\boldsymbol{R}}_{k-1} + (1-\lambda)\hat{\boldsymbol{z}}_i \hat{\boldsymbol{z}}_i^{\mathcal{H}} \tag{3.20}$$

Here, we suppose k is large enough to hold approximation of $\lambda^{k-1} \approx 0$. The detailed derivation is shown in the Appendix D.

Since $\{\hat{\boldsymbol{z}}_i\}_{i=1}^N$ are calculated by the Kalman filter assuming that the previous channel statistics are still correct even at the current time instant, we call this method as a stationarity assumption method in the rest of the paper.

3.4.2 The EM-based Estimation

The EM algorithm is an iterative parameter estimation method for finding suitable parameters according to the criterion of maximum-likelihood, where part of the stochastic variables are unobservable, or not directly accessible.

The method consists of an Expectation step and a Maximization step. In the Expectation step, the Q function which is the conditional expectation of the log-likelihood function is created using the current parameter $\boldsymbol{\theta}^{(t)}$. In the Maximization step, the updated parameter $\boldsymbol{\theta}^{t+1}$ is calculated by maximization of the Q function. These steps are alternately iterated until the parameter $\theta^{(t)}$ converges. Let $\boldsymbol{x}$ and $\boldsymbol{z}$ be observed and latent stochastic variables, respectively. The above described steps are expressed as

$$\begin{aligned}
\mathrm{E-Step}\ :\ Q(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{(t)}) &= \mathop{\mathbb{E}}_{\boldsymbol{z} \,|\, \boldsymbol{x}, \boldsymbol{\theta}^{(t)}} \log L(\boldsymbol{\theta} \,|\, \boldsymbol{x}, \boldsymbol{z}) \\
\mathrm{M-Step}\ :\ \boldsymbol{\theta}^{(t+1)} &= \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta} \,|\, \boldsymbol{\theta}^{(t)})
\end{aligned} \tag{3.21}$$

where $\mathbb{E}_{\boldsymbol{a} \,|\, \boldsymbol{b}}$ means conditional expectation with respect to $\boldsymbol{a}$ under the condition of $\boldsymbol{b}$.

In our problem, only the received symbols $\boldsymbol{Y}_k$ are directly observable. Meanwhile, the $\boldsymbol{z}_k$ corresponds to the latent variable since a direct observation of realized channel evolution $\boldsymbol{z}_i$ is unavailable.

Posterior Distribution of Latent Channel Evolution

When using the EM algorithm, one of the difficulties is that computational cost of the E-Step is intractable in most cases. Fortunately, Q function of our problem can be calculated easily since all distributions are the Gaussian normal. In this section, we calculate the posterior probability of the latent variable $\boldsymbol{z}_k$ for preparation of the evaluation of conditional expectation to be described.

Let the stochastic variables $\boldsymbol{z}_k, \boldsymbol{H}_{k-1}, \boldsymbol{N}_k$ be independent among each other. Then the joint distribution function of them can be expressed as follows:

$$\begin{aligned} P(\boldsymbol{z}_k, \boldsymbol{N}_k, \boldsymbol{H}_{k-1}) =& \mathcal{CN}(\boldsymbol{z}_k \,|\, \boldsymbol{0}, \boldsymbol{R}_k)\mathcal{CN}(\operatorname{vec}\boldsymbol{N}_k \,|\, \boldsymbol{0}, \boldsymbol{Q}_k) \\ & \times \mathcal{CN}(\operatorname{vec}\boldsymbol{H}_{k-1} \,|\, \operatorname{vec}\widehat{\boldsymbol{H}}_{k-1}, \boldsymbol{P}_{k-1\,|\,k-1}) \end{aligned} \tag{3.22}$$

After having observed $\boldsymbol{Y}_k$ whose definition is expressed according to (3.6) (3.9) as

$$\operatorname{vec}\boldsymbol{Y}_k = \boldsymbol{B}_k(\operatorname{vec}\boldsymbol{H}_{k-1} + \boldsymbol{z}_k) + \operatorname{vec}\boldsymbol{N}_k\,, \tag{3.23}$$

the posterior distribution of $\boldsymbol{z}_k$ is obtained by marginalizing the above joint distribution function with respect to $\boldsymbol{H}_{k-1}$.

$$\begin{aligned} P(\boldsymbol{z}_k \,|\, \boldsymbol{Y}_k) &= \int_{\mathbb{C}^{N_{\mathrm{Rx}}\times N_{\mathrm{Tx}}}} P(\boldsymbol{z}_k, \boldsymbol{N}_k, \boldsymbol{H}_{k-1})\,\mathrm{d}\boldsymbol{H}_{k-1} \\ &= \int_{\mathbb{C}^{N_{\mathrm{Rx}}N_{\mathrm{Tx}}}} \mathcal{CN}(\boldsymbol{z}_k \,|\, \boldsymbol{0}, \boldsymbol{R}_k)\,\mathcal{CN}(\boldsymbol{h}_k \,|\, \boldsymbol{0}, \boldsymbol{P}_{k-1\,|\,k-1}) \\ &\qquad \times \mathcal{CN}(\boldsymbol{\delta}_k - \boldsymbol{B}_k(\boldsymbol{h}_{k-1} + \boldsymbol{z}_k) \,|\, \boldsymbol{0}, \boldsymbol{Q}_k)\,\mathrm{d}\boldsymbol{h}_{k-1} \end{aligned} \tag{3.24}$$

where

$$\begin{aligned} \boldsymbol{h}_{k-1} &\triangleq \operatorname{vec}\left(\boldsymbol{H}_{k-1} - \widehat{\boldsymbol{H}}_{k-1}\right) \\ \boldsymbol{\delta}_l &\triangleq \operatorname{vec}\boldsymbol{Y}_k - \boldsymbol{B}_k \operatorname{vec}\widehat{\boldsymbol{H}}_{k-1} = \boldsymbol{B}_k(\boldsymbol{h}_{k-1} + \boldsymbol{z}_k) + \boldsymbol{N}_k \end{aligned} \tag{3.25}$$

The above integration of the Gaussian (exponential quadratic) function is tractable, and the posterior distribution is obtained as

$$P(\boldsymbol{z}_k \,|\, \boldsymbol{Y}_k) = \mathcal{CN}\left(\boldsymbol{z}_k \,|\, \boldsymbol{\varphi}_k, \left(\boldsymbol{R}_k^{-1} + \boldsymbol{F}_k\boldsymbol{C}_k\right)^{-1}\right) \tag{3.26}$$

where

$$\boldsymbol{\varphi}_k \triangleq \left(\boldsymbol{R}_k^{-1} + \boldsymbol{F}_k\boldsymbol{C}_k\right)^{-1} \boldsymbol{F}_k\boldsymbol{B}_k^{\mathcal{H}}\boldsymbol{Q}_k^{-1}\boldsymbol{\delta}_k \tag{3.27}$$

$$\boldsymbol{C}_k \triangleq \boldsymbol{B}_k\boldsymbol{Q}_k^{-1}\boldsymbol{B}_k \tag{3.28}$$

$$\boldsymbol{F}_k \triangleq \boldsymbol{I} - \boldsymbol{C}_k\left(\boldsymbol{P}_{k-1\,|\,k-1}^{-1} + \boldsymbol{C}_k\right)^{-1} \tag{3.29}$$

A detailed derivation is shown in the Appendix E.

EM-based Recursive Estimation of Latent Channel Statistics

Based on the discussions so far, we design the concrete updating equation of the EM algorithm.

Case of $\boldsymbol{R}_k$ Remains Constant

By taking conditional expectation of (3.17) with respect to $\boldsymbol{z}_i$, the Q function is calculated as

$$\begin{aligned} Q(\boldsymbol{R} \mid \boldsymbol{R}^{(t)}) &= \mathop{\mathbb{E}}_{\{\boldsymbol{z}_i \mid \boldsymbol{Y}_i, \boldsymbol{R}^{(t)}\}_{i=1}^N} \log \prod_{i=1}^N \mathcal{CN}\left(\boldsymbol{z}_i \mid \boldsymbol{0}, \boldsymbol{R}\right) \\ &= N \log\det \boldsymbol{R}^{-1} - \left\langle \boldsymbol{R}^{-1}, \sum_{i=1}^N \mathop{\mathbb{E}}_{\boldsymbol{z}_i \mid \boldsymbol{Y}_i, \boldsymbol{R}^{(t)}} \boldsymbol{z}_i \boldsymbol{z}_i^{\mathcal{H}} \right\rangle_{\mathrm{HS}} + \mathrm{const} \end{aligned} \tag{3.30}$$

where the notation $\mathbb{E}_{\{\boldsymbol{x}_i\}_{i=1}^N}$ means $\mathbb{E}_{\boldsymbol{x}_1} \mathbb{E}_{\boldsymbol{x}_2} \cdots \mathbb{E}_{\boldsymbol{x}_N}$. Maximization with respect to $\boldsymbol{R}$ is achieved in a similar way to the Appendix D. From (3.26) to (3.29), the resultant update equation is

$$\begin{aligned} \boldsymbol{R}^{(t+1)} &= \arg\max_{\boldsymbol{R}} Q(\boldsymbol{R} \mid \boldsymbol{R}^{(t)}) \\ &= \frac{1}{N} \sum_{i=1}^N \mathop{\mathbb{E}}_{\boldsymbol{z}_i \mid \boldsymbol{Y}_i, \boldsymbol{R}^{(t)}} \boldsymbol{z}_i \boldsymbol{z}_i^{\mathcal{H}} \\ &= \frac{1}{N} \sum_{i=1}^N \mathop{\mathbb{E}}_{\Delta \boldsymbol{z}_i} \left(\boldsymbol{\varphi}_i + \Delta \boldsymbol{z}_i\right) \left(\boldsymbol{\varphi}_i + \Delta \boldsymbol{z}_i\right)^{\mathcal{H}}, \\ &\quad \Delta \boldsymbol{z}_i \sim \mathcal{CN}\left(\boldsymbol{0}, \left[(\boldsymbol{R}^{(t)})^{-1} + \boldsymbol{F}_i \boldsymbol{C}_i\right]^{-1}\right) \\ &= \frac{1}{N} \left(\sum_{i=1}^N \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^{\mathcal{H}} + \left[\left(\boldsymbol{R}^{(t)}\right)^{-1} + \boldsymbol{F}_i \boldsymbol{C}_i\right]^{-1} \right) \end{aligned} \tag{3.31}$$

where $\boldsymbol{\varphi}_i$ is defined as (3.27) except that the $\boldsymbol{R}_k$ is replaced with $\boldsymbol{R}^{(t)}$. Unfortunately, calculation of (3.31) takes a lot of computation since $\boldsymbol{\varphi}_i$ and $\boldsymbol{F}_i$ require $\boldsymbol{P}_{k-1 \mid k-1}$ and $\boldsymbol{\delta}_k$ which are obtained by Kalman iteration under the updated parameter $\boldsymbol{R}^{(t)}$. As shown in Fig.3.4, it means that the past N Kalman estimation procedures must be recalculated for each iteration of the EM algorithm. In the following time-varying statistics case, we do not have to face this problem.

Case of $\boldsymbol{R}_k$ Varies Depending on k

In the case of time varying parameter, derivation is similar to the above section except that the log-likelihood function is replaced with (3.19). For time instant k satisfying $N \geq k$, the Q function is

$$Q\left(\boldsymbol{R}_k \mid \boldsymbol{R}_1^{(t)}, \boldsymbol{R}_2^{(t)}, \cdots, \boldsymbol{R}_N^{(t)}\right) = \mathop{\mathbb{E}}_{\{\boldsymbol{z}_i \mid \boldsymbol{Y}_i, \boldsymbol{R}_i^{(t)}\}_{i=1}^k} \log \prod_{i=1}^k \mathcal{CN}\left(\boldsymbol{z}_i \mid \boldsymbol{0}, \boldsymbol{R}_k\right)^{\lambda^{k-i}} \tag{3.32}$$

A straightforwardly associated procedure from the above is that after fixing all $\{\boldsymbol{R}_k^{(t)}\}_{k=1}^N$, maximization of $\{\boldsymbol{R}_k^{(t+1)}\}_{k=1}^N$ at the next iteration can be finally started. As shown below, since optimization with respect to the $\boldsymbol{R}_k^{(t+1)}$ only depends on $\boldsymbol{R}_{k-1}^{(t+1)}$ and $\boldsymbol{R}_k^{(t)}$, it is possible to

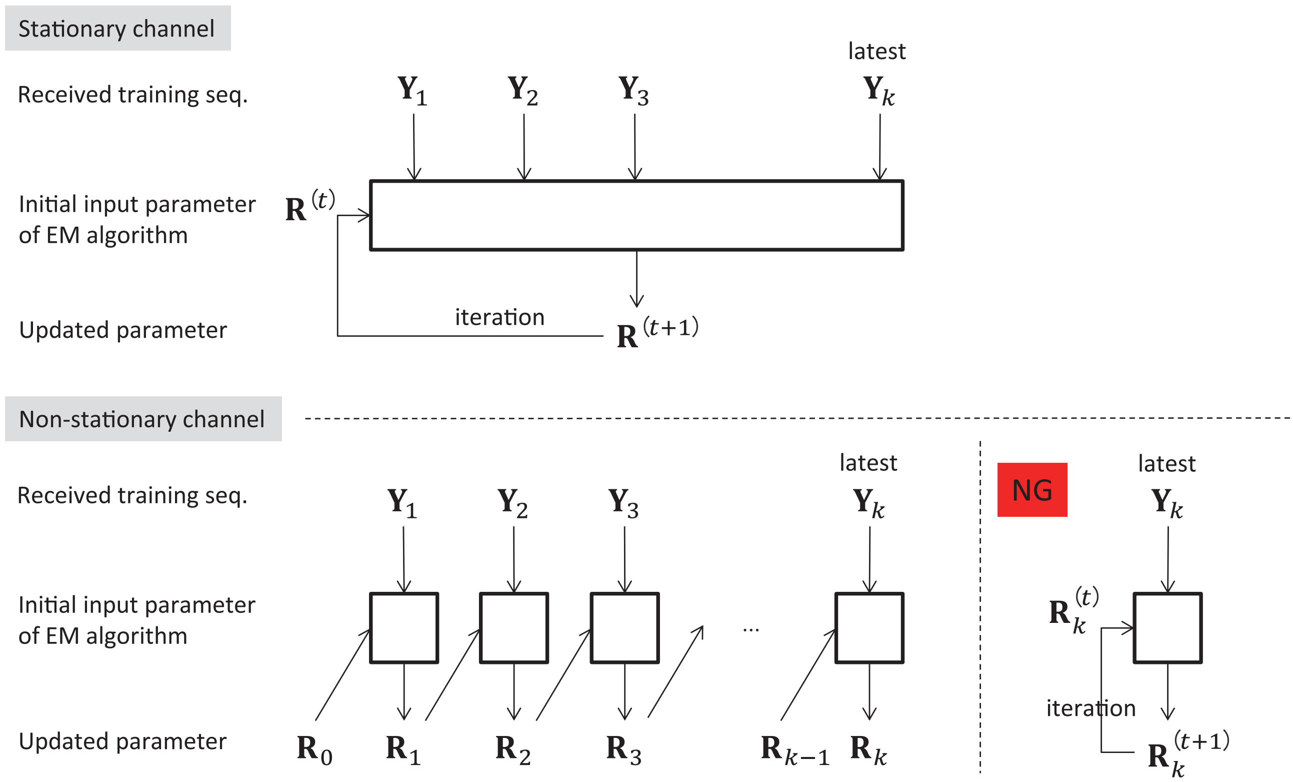

Figure 3.4: Schematics explaining parameter updating in EM algorithm. In stationary case, EM iteration can start by using all the observations. In non-stationary case, EM iteration can only utilize the latest observations, so doing many iterations result in too much dependence on the latest observation.

proceed iteration of t without updating the future ones subsequent to k.

$$\begin{aligned}\boldsymbol{R}_k^{(t+1)} &= \arg\max_{\boldsymbol{R}_k} Q\left(\boldsymbol{R}_k \mid \boldsymbol{R}_1^{(t)}, \boldsymbol{R}_2^{(t)}, \cdots, \boldsymbol{R}_N^{(t)}\right) \\ &= \frac{1}{\tau}\sum_{i=1}^{k} \lambda^{k-i}\left(\boldsymbol{\varphi}_i\boldsymbol{\varphi}_i^{\mathcal{H}} + \left[\left(\boldsymbol{R}_i^{(t)}\right)^{-1} + \boldsymbol{F}_i\boldsymbol{C}_i\right]^{-1}\right) \\ &= \lambda\boldsymbol{R}_{k-1}^{(t+1)} + (1-\lambda)\left(\boldsymbol{\varphi}_k\boldsymbol{\varphi}_k^{\mathcal{H}} + \left[\left(\boldsymbol{R}_k^{(t)}\right)^{-1} + \boldsymbol{F}_k\boldsymbol{C}_k\right]^{-1}\right)\end{aligned} \tag{3.33}$$

where τ is the same definition as used in (D.2). Equality of the last equation of (3.33) holds when $\lambda^{k-1} \approx 0$. For initial iteration at $k = 1$, $\boldsymbol{R}_{k-1}^{(t+1)}$ should be replaced with an initial guess $\widehat{\boldsymbol{R}}_0$.

Originally, the scheme of the EM algorithm states that the parameter update of (3.33) should be iterated until convergence. However, because of the following two reasons, we employ only a single iteration.

The first reason is to prevent increase of computational cost. Due to temporal correlation, the desired parameter $\boldsymbol{R}_k^{(t+1)}$ would be near to the one of the previous time instant $\boldsymbol{R}_{k-1}^{(t+1)}$. In order to start the iteration with as good parameter as possible, the initial value of the iteration is set to $\boldsymbol{R}_{k-1}^{(t+1)}$. Thus the single EM-update increases likelihood of the latest parameter based on criterion of the latest likelihood function.

The second reason comes from the formulation of the problem. Since each $\boldsymbol{R}_i^{(t+1)}$ only depends on the past inputs $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_i$ until time instant i, as illustrated in Fig.3.4, the EM iteration for determination of $\boldsymbol{R}_k^{(t+1)}$ is only affected by the latest input $\boldsymbol{Y}_k$. Therefore, doing many iterations results in too much dependence on the latest single realization.

From now on, we focus on the scenario described in this section. The derived parameter updating steps are executed followed by the Kalman steps, thereby the following updating equations are inserted before (3.10).

$$\boldsymbol{\delta}_k = \operatorname{vec}\boldsymbol{Y}_k - \boldsymbol{B}_k \operatorname{vec}\widehat{\boldsymbol{H}}_{k-1} \tag{3.34}$$

$$\boldsymbol{C}_k \triangleq \boldsymbol{B}_k\boldsymbol{Q}_k^{-1}\boldsymbol{B}_k \tag{3.35}$$

$$\boldsymbol{F}_k \triangleq \boldsymbol{I} - \boldsymbol{C}_k\left(\boldsymbol{P}_{k-1\,|\,k-1}^{-1} + \boldsymbol{C}_k\right)^{-1} \tag{3.36}$$

$$\boldsymbol{\varphi}_k \triangleq \left(\boldsymbol{R}_k^{-1} + \boldsymbol{F}_k\boldsymbol{C}_k\right)^{-1}\boldsymbol{F}_k\boldsymbol{B}_k^{\mathcal{H}}\boldsymbol{Q}_k^{-1}\boldsymbol{\delta}_k \tag{3.37}$$

$$\boldsymbol{R}_k = \lambda\boldsymbol{R}_{k-1} + (1-\lambda)\left(\boldsymbol{\varphi}_k\boldsymbol{\varphi}_k^{\mathcal{H}} + \left[\boldsymbol{R}_{k-1}^{-1} + \boldsymbol{F}_k\boldsymbol{C}_k\right]^{-1}\right) \tag{3.38}$$

3.4.3 Reduction of Computations by Parametric Covariance Modeling

The derived updating equations above includes many matrix-inverses, which results in consumption of considerable amount of computational resources. Those matrix inversions are required for accurate calculation of non-stationary spatial covariance matrix. However, if we abandon some estimation accuracy by confining the form of covariance matrix to a certain parametric model, computations of matrix-inverse can be avoided, and significant computations are saved.

If there holds a situation where eigenvectors of both spatial covariance matrix and observation matrix are identical among all the time instants, matrix calculations can be substituted by scalar arithmetic operations for eigenvalues corresponding to the fixed eigenvectors. The assumption that eigenvectors of correlation matrix remain constant is also used in [44].

Since the observation matrix has the form of $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B} = \boldsymbol{S}^{*}\boldsymbol{S}^{\top} \otimes \boldsymbol{I}$, its eigenvectors should belong to $\mathbb{C}^{N_{\mathrm{Tx}}} \otimes \mathbb{C}^{N_{\mathrm{Rx}}}$. Let $\{\boldsymbol{\phi}_j \otimes \boldsymbol{\psi}_i\}_{i=1,j=1}^{N_{\mathrm{Tx}},N_{\mathrm{Rx}}}$ be common fixed eigenvectors of $\boldsymbol{B}^{\mathcal{H}}\boldsymbol{B}, \boldsymbol{R}_k$, and $\boldsymbol{P}_{k\,|\,l}$. Now, $\boldsymbol{\psi}_i$ can be an arbitrary orthonormal basis, and $\{\boldsymbol{\phi}_j\}_{j=1}^{N_{\mathrm{Tx}}}$ should be eigenvectors of $\boldsymbol{S}^{*}\boldsymbol{S}^{\top}$. If $\boldsymbol{S}$ is the orthogonal training sequence, also a choice of $\boldsymbol{\phi}_j$ can be an arbitrary orthonormal basis, and in such a case, $\boldsymbol{\phi}_j$ and $\boldsymbol{\psi}_i$ are decided by other aspects such as a priori correlation statistics. Considering this, we assume that the spatial correlation matrix can be expressed as

$$\boldsymbol{R}_k = \sum_{j=1}^{N_{\mathrm{Tx}}}\sum_{i=1}^{N_{\mathrm{Rx}}} r_{i,j}^{(k)} \boldsymbol{\phi}_j \boldsymbol{\phi}_j^{\mathcal{H}} \otimes \boldsymbol{\psi}_i \boldsymbol{\psi}_i^{\mathcal{H}} \tag{3.39}$$

where $\boldsymbol{\phi}_j$ and $\boldsymbol{\psi}_i$ are eigenvectors of a priori correlation matrices of Tx and Rx, respectively. Here, we can say that $\boldsymbol{R}_k$ is parameterized with the parameters of $\{r_{i,j}^{(k)}\}_{i=1,j=1}^{N_{\mathrm{Rx}},N_{\mathrm{Tx}}}$, thus we call this simplification as parametric covariance modeling for later discussion. The degrees of freedom of $\boldsymbol{R}_k$ is reduced to $N_{\mathrm{Tx}}N_{\mathrm{Rx}}$ from $(N_{\mathrm{Tx}}^2 N_{\mathrm{Rx}}^2 - N_{\mathrm{Tx}}N_{\mathrm{Rx}})/2 + N_{\mathrm{Tx}}N_{\mathrm{Rx}}$. If there holds a special case that $r_{i,j}^{(k)}$ are determined independently with respect to combinations of i and j (i.e. if $r_{i,j}^{(k)} = s_i^{(k)} t_j^{(k)}$ holds), (3.39) is identical to the Kronecker model assumption [40, 41] which means correlations between Tx and Rx are independent.

For more generalized formulation, we assume that the observation matrix can be expressed as follows:

$$\boldsymbol{B}_k^{\mathcal{H}}\boldsymbol{B}_k = \sum_{j=1}^{N_{\mathrm{Tx}}}\sum_{i=1}^{N_{\mathrm{Rx}}} \beta_{i,j}^{(k)} \boldsymbol{\phi}_j \boldsymbol{\phi}_j^{\mathcal{H}} \otimes \boldsymbol{\psi}_i \boldsymbol{\psi}_i^{\mathcal{H}} \tag{3.40}$$

If $\boldsymbol{B} = \boldsymbol{S}^{\top} \otimes \boldsymbol{I}$ holds, $\beta_{i,j}^{(k)}$ becomes the j-th eigenvalue of $\boldsymbol{S}^{*}\boldsymbol{S}^{\top}$. Under these assumptions, the error covariance matrices are always expressed as:

$$\boldsymbol{P}_{k\,|\,l} = \sum_{j=1}^{N_{\mathrm{Tx}}}\sum_{i=1}^{N_{\mathrm{Rx}}} \nu_{i,j}^{(k\,|\,l)} \boldsymbol{\phi}_j \boldsymbol{\phi}_j^{\mathcal{H}} \otimes \boldsymbol{\psi}_i \boldsymbol{\psi}_i^{\mathcal{H}} \tag{3.41}$$

Let $\alpha_{i,j}^{(k)}$, $f_{i,j}^{(k)}$, and $\eta_{i,j}^{(k)}$ be eigenvalues corresponding to eigenvector $\boldsymbol{\phi}_j \otimes \boldsymbol{\psi}_i$ of matrices $\boldsymbol{C}_k$, $\boldsymbol{F}_k$, $(\boldsymbol{R}_{k-1}^{-1} + \boldsymbol{F}_k \boldsymbol{C}_k)^{-1}$, respectively. Assuming that observation noise is Gaussian white ($\boldsymbol{Q}_k = \sigma_n^2 \boldsymbol{I}$), and $\boldsymbol{R}_k$ is nonsingular, the update equations of (3.10) (3.38) and (3.15) are rewritten with small computations as follows:

$$\nu_{i,j}{}^{(k\,|\,k-1)} = \nu_{i,j}^{(k-1\,|\,k-1)} + r_{i,j}^{(k)} \tag{3.42}$$

$$\alpha_{i,j}^{k} = \beta_{i,j}^{k}/\sigma_n^2 \tag{3.43}$$

$$f_{i,j}^{(k)} = \frac{1}{1 + \nu_{i,j}^{(k-1\,|\,k-1)} \alpha_{i,j}^{(k)}} \tag{3.44}$$

$$\eta_{i,j}^{(k)} = \frac{r_{i,j}^{(k-1)}(1 + \nu_{i,j}^{(k-1\,|\,k-1)} \alpha_{i,j}^{(k)})}{1 + \nu_{i,j}^{(k-1\,|\,k-1)} \alpha_{i,j}^{(k)} + r_{i,j}^{(k-1)} \alpha_{i,j}^{(k)}} \tag{3.45}$$

$$r_{i,j}^{(k)} = \lambda r_{i,j}^{(k-1)} + (1-\lambda)\left[\left(\gamma_{i,j}^{(k)}\right)^2 \left\| < \boldsymbol{\phi}_j \otimes \boldsymbol{\psi}_i, \boldsymbol{B}_k^{\mathcal{H}} \delta_k > \right\|^2 + \eta_{i,j}^{(k)}\right] \tag{3.46}$$

$$\nu_{i,j}^{(k\,|\,k)} = \left(\frac{1}{\nu_{i,j}^{(k\,|\,k-1)}} + \alpha_{i,j}^{(k)}\right)^{-1} \tag{3.47}$$

where

$$\gamma_{i,j}^{(k)} \triangleq \frac{r_{i,j}^{(k-1)}}{(1+\nu_{i,j}^{(k-1\,|\,k-1)}\alpha_{i,j}^{(k)} + r_{i,j}^{(k-1)}\alpha_{i,j}^{(k)})\sigma_n^2} \tag{3.48}$$

A detailed calculation of the maximization step required in the derivation of (3.46) is described in the Appendix F.

3.4.4 Parametric Covariance Modeling in the Case of Antenna Selection Systems

In the previous section, eigenvectors of the parametric covariance matrix are assumed by a priori channel assumptions. In contrast, if eigenvectors of the observation matrix are determined due to some system configurations, eigenvectors of the parametric covariance matrix are imposed to be ones of the observation matrix regardless of a priori channel assumptions, since they should be common in order to apply this computational reduction. In this case, degradation of estimation capability is expected because of mismatch between the true and the assumed eigenvectors.

One of the examples of this situation arise when we employ antenna selection systems on either Tx or Rx. In antenna selection systems, since the number of RF chains are smaller than the number of antenna elements, only several rows or columns of channel matrix are measured at single reception of training sequences. According to [22] [38], observation equation of antenna selection system on Rx side is expressed as:

$$\boldsymbol{Y}_k = \boldsymbol{A}_{\tau_k}\boldsymbol{H}_k\boldsymbol{S}_k + \boldsymbol{N}_k \tag{3.49}$$

where $\boldsymbol{A}_{\tau_k}$ is antenna selection matrix which performs extraction and permutation of row vectors of channel matrix. By using the antenna connection vector $\boldsymbol{\tau}_k \triangleq [c_1\, c_2 \cdots c_{N_{\mathrm{RF}}}]^\top$, $\boldsymbol{A}_{\tau_k}$ is expressed as

$$\begin{bmatrix} [\boldsymbol{H}_k]_{c_1,:} \\ [\boldsymbol{H}_k]_{c_2,:} \\ \vdots \\ [\boldsymbol{H}_k]_{c_{N_{\mathrm{RF}}},:} \end{bmatrix} = \boldsymbol{A}_{\tau_k}\boldsymbol{H}_k, \quad \boldsymbol{A}_{\tau_k} \triangleq \sum_{i=1}^{N_{\mathrm{RF}}} \mathbf{f}_i \mathbf{e}_{<\boldsymbol{\tau}_k,\mathbf{f}_i>}^\top$$

where $[\boldsymbol{H}_k]_{c_n,:}$ means the c_n-th row vector of $\boldsymbol{H}_k$, and $\mathbf{e}_i, \mathbf{f}_j$ are the so-called standard basis of $\mathbb{C}^{N_{\mathrm{Rx}}}, \mathbb{C}^{N_{\mathrm{RF}}}$, respectively. In this situation, considering the relationship of $\boldsymbol{B}_k^{\mathcal{H}}\boldsymbol{B}_k = \boldsymbol{S}_k^*\boldsymbol{S}_k^\top \otimes \boldsymbol{A}_{\tau_k}^{\mathcal{H}}\boldsymbol{A}_{\tau_k}$, and $\boldsymbol{A}_{\tau_k}^{\mathcal{H}}\boldsymbol{A}_{\tau_k}$ is a diagonal matrix, $\boldsymbol{\psi}_i$ should be the i-th standard basis of $\mathbb{C}^{N_{\mathrm{Rx}}}$.

(3.49) expresses a single measurement which measures only a part of the whole channel matrix.Another canonical case is that, in order to cover all the elements, measurements are repeated several times assuming that the channel remains constant during our time period of interest. In such cases, if we employ orthogonal training sequences, the observation matrix can be expressed as

$$\boldsymbol{B}_k^{\mathcal{H}}\boldsymbol{B}_k = \frac{1}{N_{\mathrm{Tx}}}\boldsymbol{I}_{N_{\mathrm{Tx}}} \otimes \mathrm{diag}\,[\,T_1,\, T_2,\, \cdots,\, T_{N_{\mathrm{Rx}}}\,] \tag{3.50}$$

where T_i denotes the total number of received training sequences of the i-th element during the measurements.

As discussed above, if we employ antenna selection systems on Rx, the common eigenvectors are forced to be $\phi_j \otimes \mathrm{e}_i$ where e_i is the i-th standard basis, and this constraint results in degradation of estimation capability. Similarly, if antenna selection is applied on Tx, the common eigenvectors are forced to be $\mathrm{e}_j \otimes \psi_i$.

3.5 Simulation

We evaluated the effectiveness of the proposed scheme by numerical simulations with different parameters for spatial and temporal correlation. As a measure of the estimation capability, we employed the squared error of channel matrix defined as $\|\boldsymbol{H}_k - \widehat{\boldsymbol{H}}_k\|_F^2$ where $\widehat{\boldsymbol{H}}_k$ is estimate of the true channel state $\boldsymbol{H}_k$. Note that the channel state is normalized such that $\mathrm{E}\|\boldsymbol{H}_k\|_F^2 = N_{\mathrm{Tx}}N_{\mathrm{Rx}}$.

3.5.1 Verification by the 3GPP SCME

In order to generate spatially and temporally correlated channel, we resort to the 3GPP Spatial Channel Model Extended (SCME) [46]. The SCME has features such as time-variant shadow fading and time-variant angles and delays, which coincide with our channel assumption that the channel is non-stationary. A scenario is selected to "Urban micro," and all other parameters are default settings except a number of antenna elements, sampling interval, and velocity of the mobile station. The path delays calculated by the SCME are transformed into a narrowband transfer function at 2GHz. This operation corresponds to deal with a single carrier of the OFDM systems under frequency selective fading channel.

For each simulation parameter, an iteration of the Kalman filter is repeated 5000 times, and the result is averaged over them. The average SNR is set to 15dB. The number of Tx and Rx elements (N_{Tx} and N_{Rx}) are fixed to 7. Also the length of the training sequence N_t is set to 8. Fig.3.5 depicts an ensemble average and a maximum value of squared error for different velocity of the mobile station. In the figure, two estimation methods are compared. One is the EM-based estimation method described in 3.4.2, and the other is the stationarity assumption method which is calculated as (3.20).

For both methods, as the velocity becomes faster, smaller estimation error is achieved, which suggests that the methods can exploit the temporal correlation characteristics of the propagation channel. The EM-based method marks better estimation performance than the stationarity assumption method, especially on the point of the maximum error value, which means that the proposed method can better suppress errors when the worst mistakes happen. This phenomenon would be a difference of convergence speed of parameters, and we analyze the reason for this by using s simple channel model in the following sections.

3.5.2 Analysis by Simple Channel Model

In order to obtain more comprehensible behavior of the algorithms, we also investigated them by using the single tapped Gauss-Markov model. To generate a non-stationary channel assuming change of channel statistics due to shadowing or irregular movements of mobile station, we update the statistical parameters every certain period (every N_p time instants). In each period, spatial correlation matrices are interpolated by means of bilinear method so that the statistics change continuously. The updating equation is expressed as

$$\boldsymbol{H}_k = \rho_t \boldsymbol{H}_{k-1} + \sqrt{1-\rho_t^2}\,(\boldsymbol{R}_{\mathrm{Rx}})^{1/2}\boldsymbol{X}_k\left[(\boldsymbol{R}_{\mathrm{Tx}})^{1/2}\right]^\top \tag{3.51}$$

$$\boldsymbol{R}_{\mathrm{Rx}} \triangleq \frac{N_p - s}{N_p}\boldsymbol{R}_{\mathrm{Rx}}^{(t)} + \frac{s}{N_p}\boldsymbol{R}_{\mathrm{Rx}}^{(t+1)}, \qquad s \triangleq k - N_p \times t \tag{3.52}$$

$$\boldsymbol{R}_{\mathrm{Tx}} \triangleq \frac{N_p - s}{N_p}\boldsymbol{R}_{\mathrm{Tx}}^{(t)} + \frac{s}{N_p}\boldsymbol{R}_{\mathrm{Tx}}^{(t+1)} \tag{3.53}$$

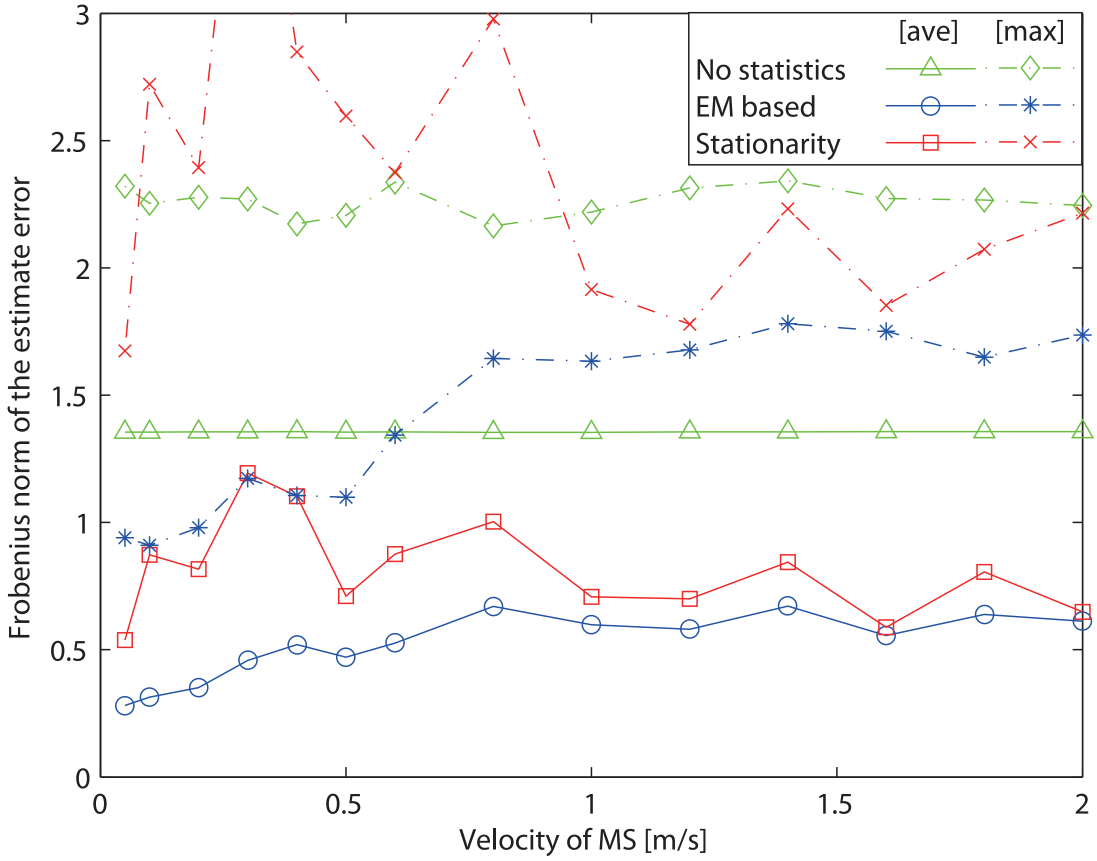

Figure 3.5: Channel estimate error of the 3GPP-SCME channel model.(SNR is 15[dB], $N_t = 8$, $\lambda = 0.98$)

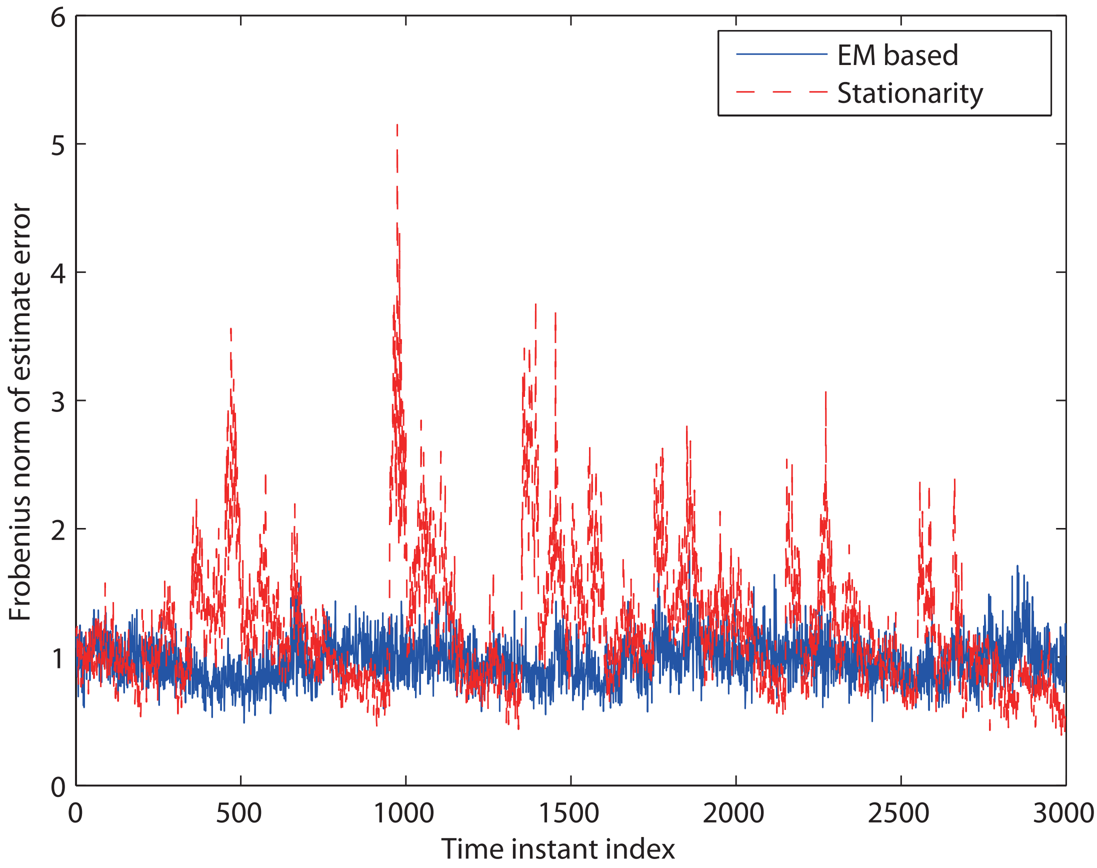

Figure 3.6: Squared error of channel estimate with elapse of time.(SNR is 15[dB], $N_t = 8$, $N_p = 50$, $\lambda = 0.98$)

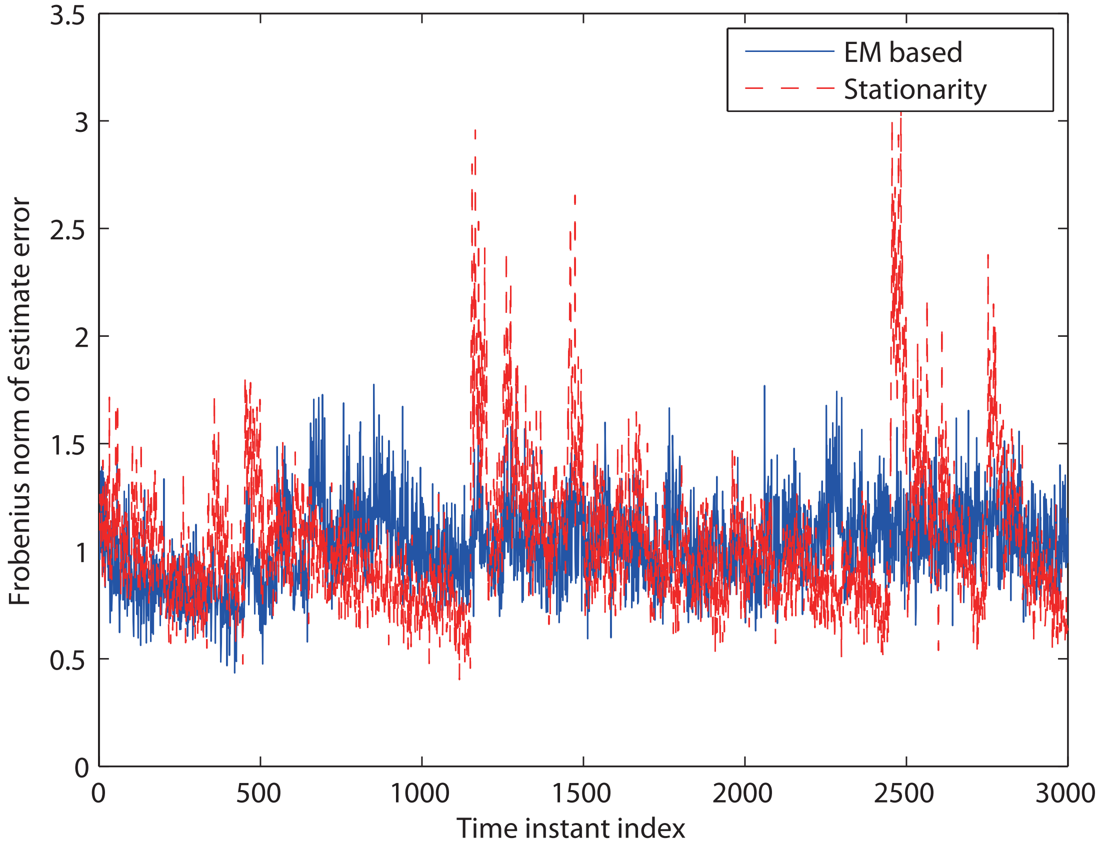

Figure 3.7: Squared error of channel estimate with elapse of time.(SNR is 15[dB], $N_t = 8$, $N_p = 50$, $\lambda = 0.99$)

where ρ_t $(0 \leq \rho \leq 1)$ is the temporal correlation coefficient between adjacent fading blocks, and the subscript t indicates the parameter index defined as $t \triangleq \lfloor k/N_p \rfloor$. $\lfloor x \rfloor$ denotes integer part of x. $\boldsymbol{X}_k$ is a random matrix which obeys $\mathrm{vec}\,\boldsymbol{X}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_{N_{\mathrm{Tx}}N_{\mathrm{Rx}}})$. ρ_t is generated randomly by $\mathcal{N}(\mu_\rho, \sigma_\rho)$, and then clipped into [0,1]. (3.51) is the same channel model as described in [42] if ρ does not depend on time.

Fig.3.6, Fig.3.7 shows squared channel estimation error of both methods with elapse of time. From the graph, we can see that the stationarity assumption method has weak estimation capability on the switching points of channel statistics (occurs per N_p samples), but the EM-based method has smaller estimation error in such points. This indicates that the EM-based method has superiority for variation of channel statistics, and is suited for non-stationary channels.

In general, if the number of channel realizations for statistics estimation is not enough, estimation will become unstable. In contrast, the EM-based method is able to estimate stable parameters with relatively fewer channel realizations. The reason is considered that an expectation of estimation error is taken in the expectation step, thus virtually acts as using many channel realizations. In the notion of the bias-variance tradeoff [58], this advantage is interpreted as decrease of variance which is brought by expectation step using a posterior probability. However, incorporating a bias like this would lead to increase estimation error if there is a mismatch in estimation model of the posterior probability.

In order to investigate these characteristics in detail, we evaluated the average and maximum value of the estimation errors for various degrees of stationarities indicated by N_p. Fig.3.9 shows the mean squared error of channel estimate for different forgetting factors ranging from $\lambda = 1.0$ to 0.9. Also Fig.3.8 shows its magnification of the range from $\lambda = 1.0$ to 0.975. Fig.3.10,3.11 shows the maximum value of estimation error for the same conditions. Note that $\lambda = 1.0$ means that the channel statistics are not updated and they are fixed to its initial values.

In these configurations, considering that $N_{\mathrm{Tx}} = N_{\mathrm{Rx}} = 7$, the rank of the spatial correlation matrix is up to 49. Therefore, the forgetting factor requires at least about 0.975 ($0.975^{49} \approx 0.29$) in order to prevent rank-deficiency.

In the case of non-stationary channels($N_p = 50, 100$), the EM-based method provides better estimation capability especially for smaller forgetting factors. The EM-based method suffers from less influence of the smaller forgetting factors, thanks to the effect of variance suppression as previously mentioned.

In the case of stationary channels($N_p = \infty$), if the forgetting factor is large enough, the stationarity assumption method performs better estimation than the EM-based method. Otherwise, where the forgetting factor is small, both methods have almost same estimation capability. If the channel is stationary, even the stationarity assumption method can achieve precise estimation since enough channel realizations are available by employing large forgetting factors. In this way, it is difficult for the EM-based method to exhibit its superiority of obtaining better performance when using fewer channel realizations.

It should be noted that the EM-based method has inferior performance to the stationarity assumption method when larger forgetting factors employed, but the EM-based method achieves same performance for smaller forgetting factors (about 0.91). In this manner, both methods have a different choice of the optimal parameters, respectively. The EM-based methods tends to perform better when the dependency on the past channel realizations is small.

Exploitation of Spatiotemporal Correlation

We confirmed that both the proposed and the stationarity assumption methods are surely able to exploit spatiotemporal correlations. For this observation, in order to specify degree of

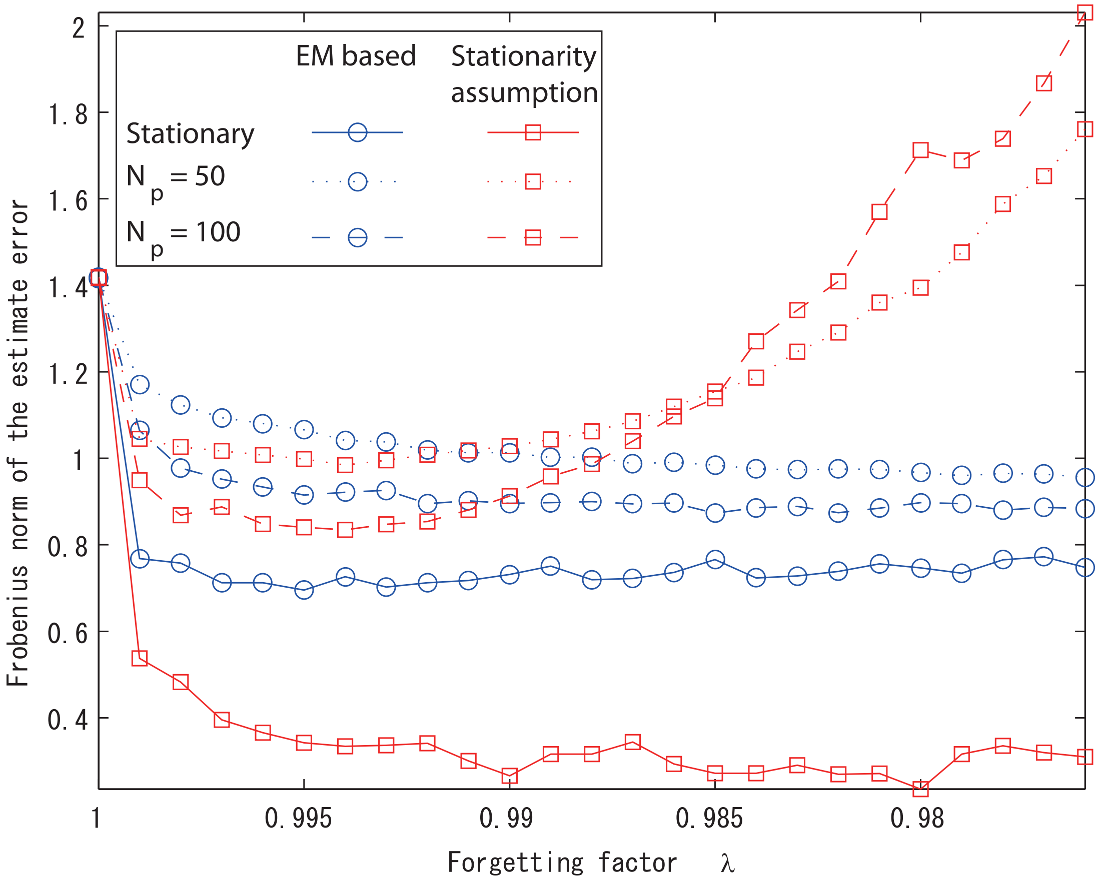

Figure 3.8: Mean squared error of channel estimate for different forgetting factors.

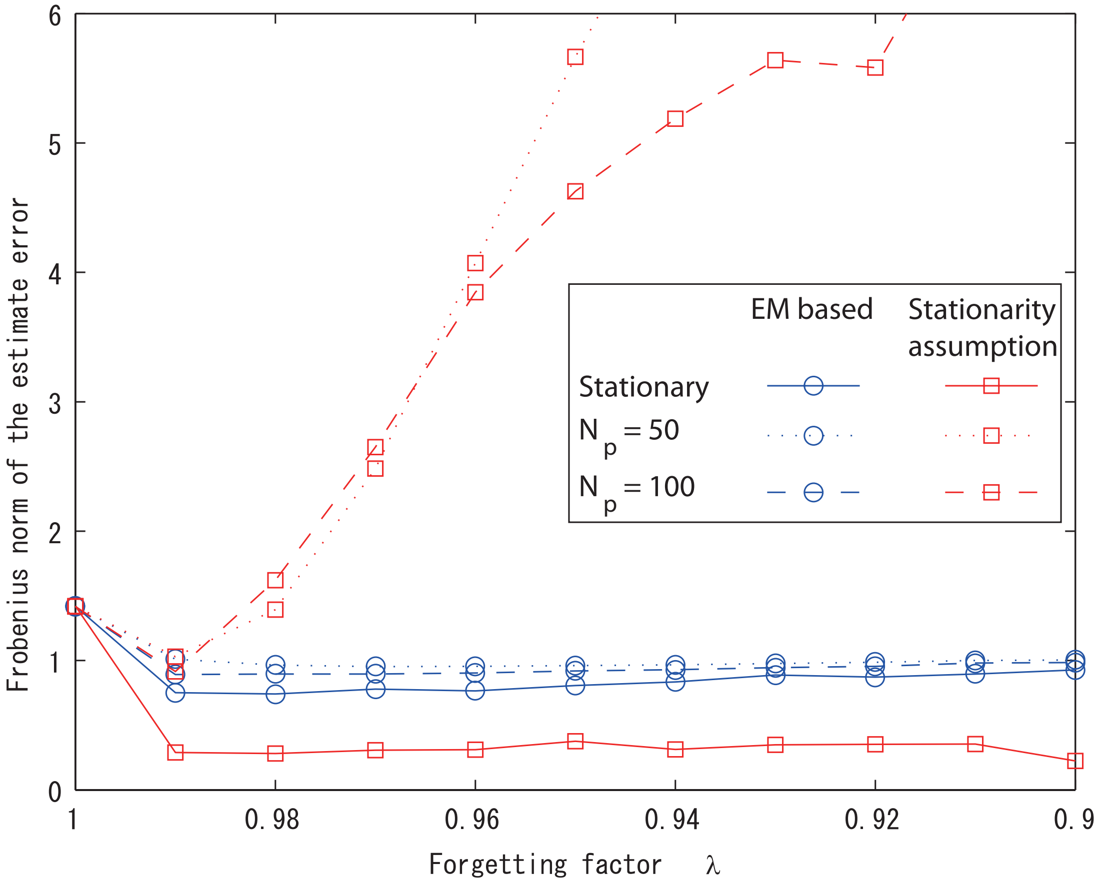

Figure 3.9: Mean squared error of channel estimate for different forgetting factors.

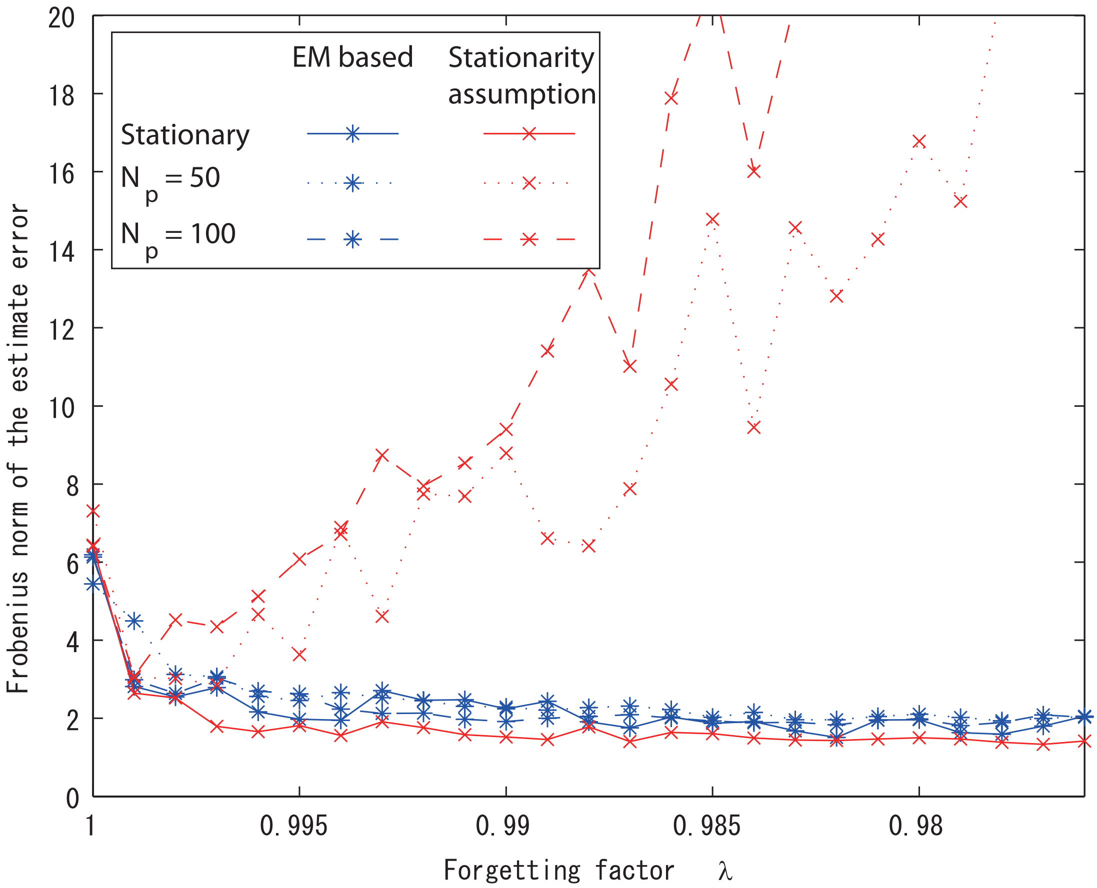

Figure 3.10: Maximum squared error of channel estimate for different forgetting factors.

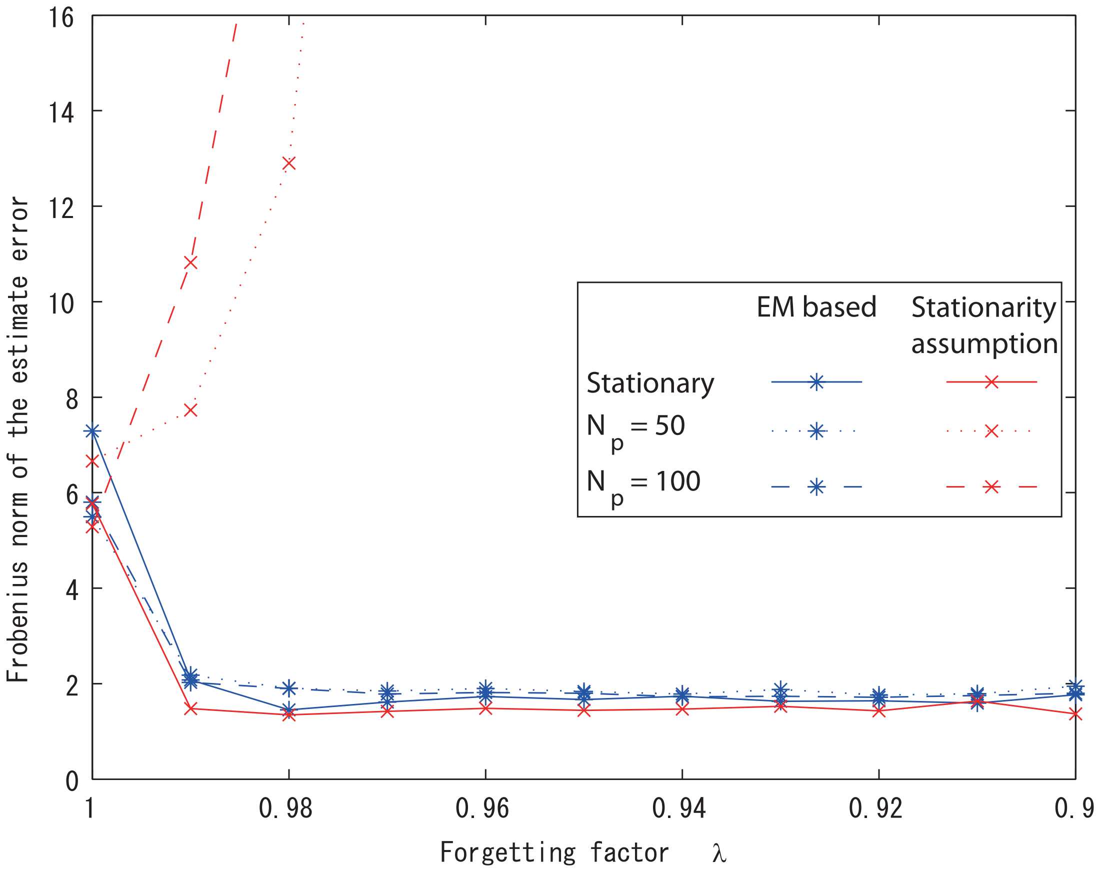

Figure 3.11: Maximum squared error of channel estimate for different forgetting factors.

correlation, whole snapshots are generated by the same parameters thus the channel is set to a stationary process (i.e. $N_p = \infty$). Fig.3.12 and Fig.3.13 show estimation capability for various degrees of spatial and temporal correlations, respectively. In the figures, the "No channel statistics" means that the channel is estimated without any statistical information by means of the ML estimation [48] [49] whose procedure is only a multiplication of pseudoinverse of the training sequence. Also the "A priori statistics" means that the estimation is done by the Kalman filter with true statistical parameters which are inaccessible in practical situations. We can see that both algorithms can reduce estimation error under presence of higher spatial correlations. On the other hand, temporal correlations can also be exploited, but its effect is rather insignificant compared to spatial correlations.

Reduction of Computations by Parametric Covariance Matrix

Fig.3.14 shows channel estimation capability when the parametric covariance matrix described in the section 3.4.3 was employed. The true channel states are generated by the SCME. The fixed eigenvectors are determined from the covariance matrix defined as

$$\boldsymbol{R}_{\mathrm{Rx}} = \int_0^{2\pi} \boldsymbol{a}_{\mathrm{Rx}}(\theta)\boldsymbol{a}_{\mathrm{Rx}}^{\mathcal{H}}(\theta)\,\mathrm{d}\theta \tag{3.54}$$

where $\boldsymbol{a}_{\mathrm{Rx}}(\theta)$ denotes the steering vector at the Rx side for an incident plane wave from the direction of θ. It means that the arrival waves are assumed to be uncorrelated and uniformly distributed over all directions. Note that we will uniformly use such derived eigenvectors regardless of the actual correlation statistics. (In this case, the actual correlation model within the SCME is different from the assumed correlation model.) The intention of using such derived eigenvectors is that it would contain correlation of the array antenna configuration itself to some extent, which is independent of the distribution of arrival waves.

From Fig.3.14, the result of the parametric covariance matrix is not largely degraded, but limited to the case of the slower MS velocity in order to keep superiority to the stationarity assumption method.

Parametric Covariance Matrix for Antenna Selection Systems

Fig. 3.15 shows channel estimation capability when the parametric covariance matrix for antenna selection systems described in the section 3.4.4 was employed. As expressed in (3.49), we employ a partial observation which measures only selected antenna elements, and rest of the unobserved channels are inferred via channel statistics. In the simulation, we choose $N_{\mathrm{Rx}} = 7$ and $N_{\mathrm{RF}} = 3$.

In this scenario, since not all the channel elements are directly observed, channel estimation error increases rapidly as the temporal correlation becomes weaker due to increase of the MS velocity. In the result, this kind of partial observation method is only useful where the MS velocity is slow enough. This result is same with the result of [22] which discusses on similar system configurations except that the ideal channel statistics are available.

The simulation also revealed that the EM based parameter estimation scheme works effectively for partial measurement problem in antenna selection system. In the partial channel measurement, since the scheme has the characteristic that the channel estimation error varies by antenna elements, the acquisition of its statistics becomes a difficult problem. This is because, we have to acquire correlation between elements while the elements are not observed simultaneously in uniform precision. The reason for the effectiveness of the EM based method is considered as follows. In the EM based method, the information of observation matrix is

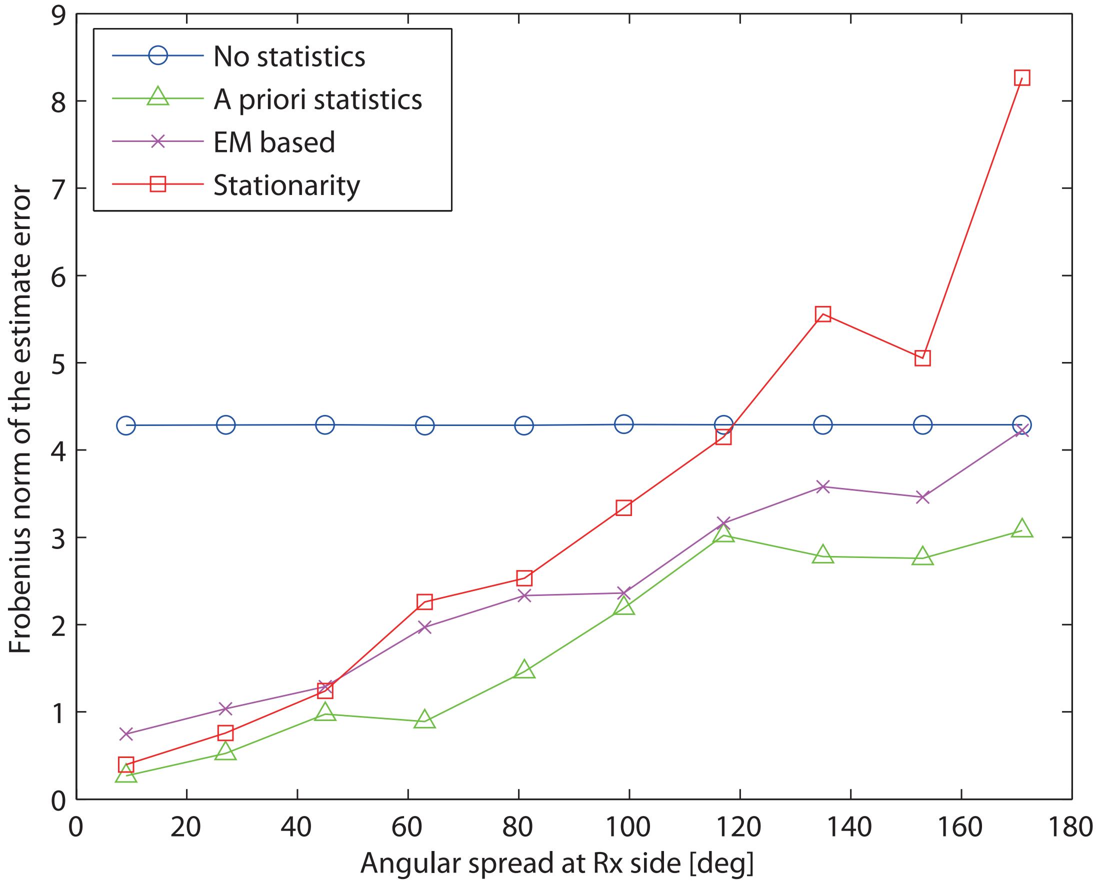

Figure 3.12: Mean squared error of channel estimate for different spatial correlations.

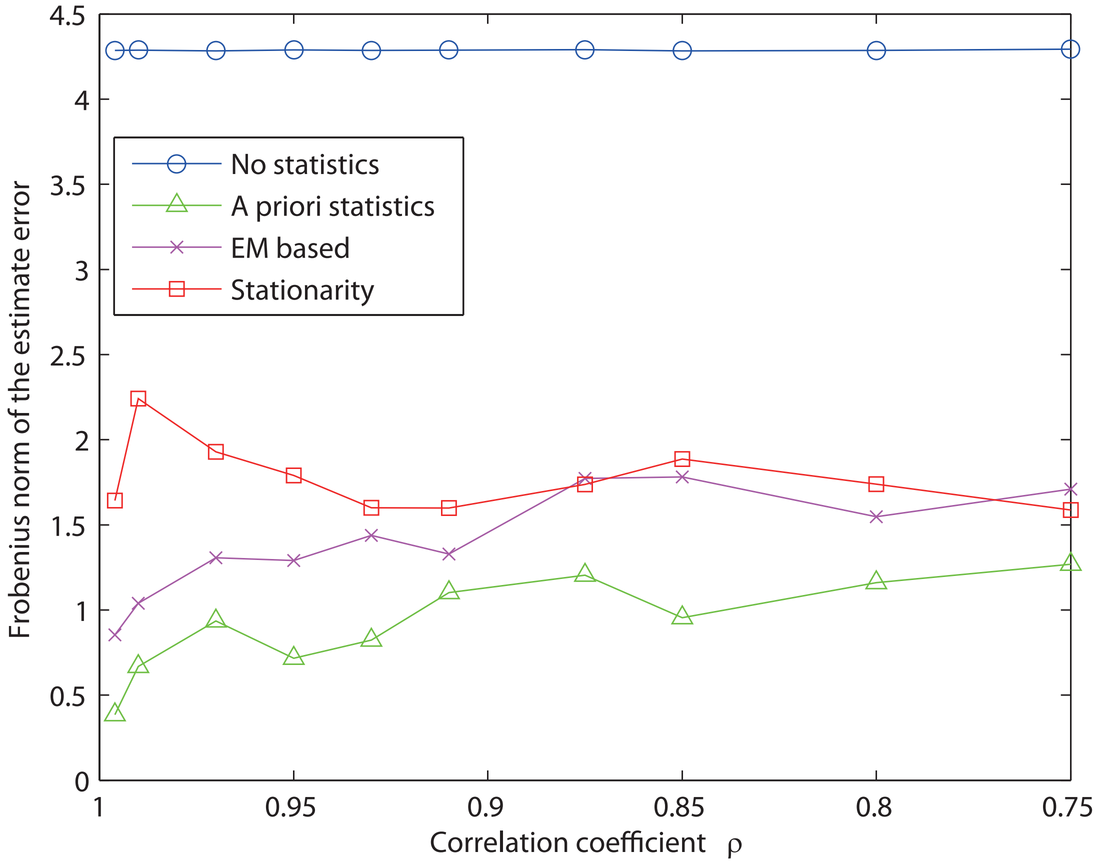

Figure 3.13: Mean squared error of channel estimate for different temporal correlations.

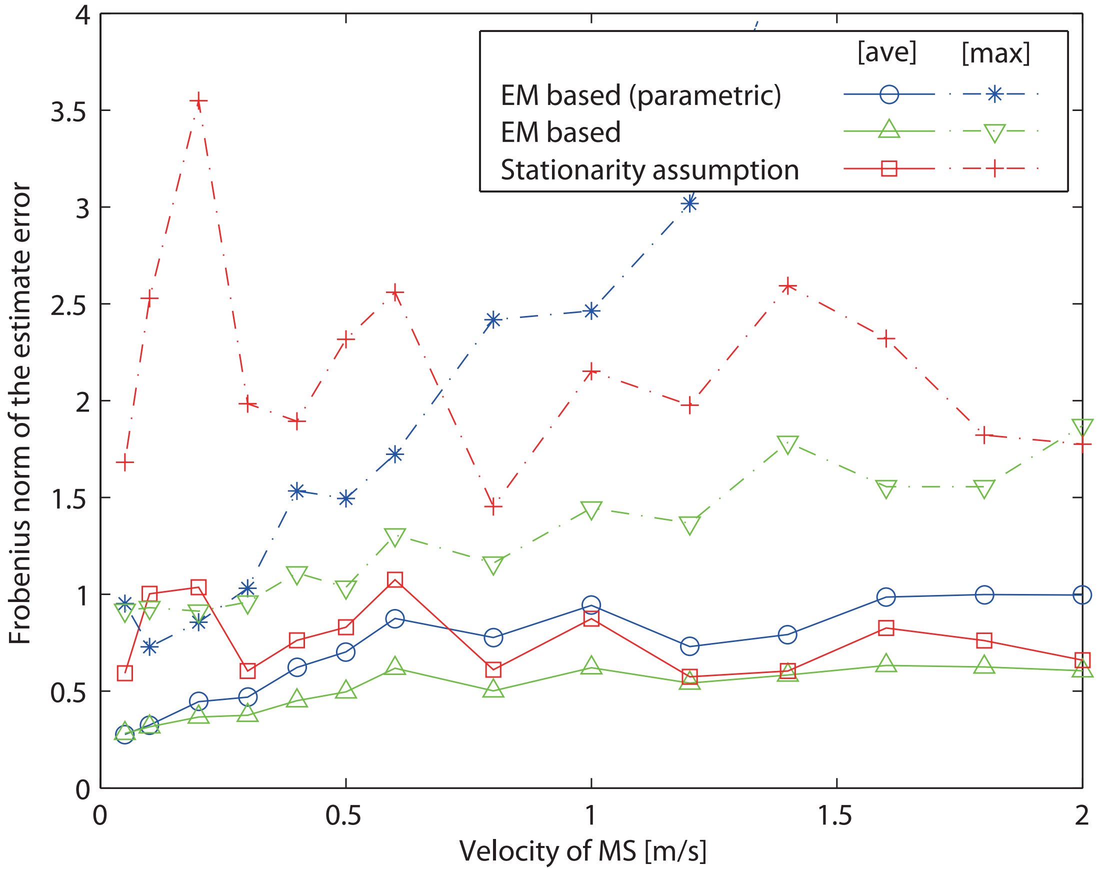

Figure 3.14: Mean squared error of channel estimate with parametric covariance matrix employed.

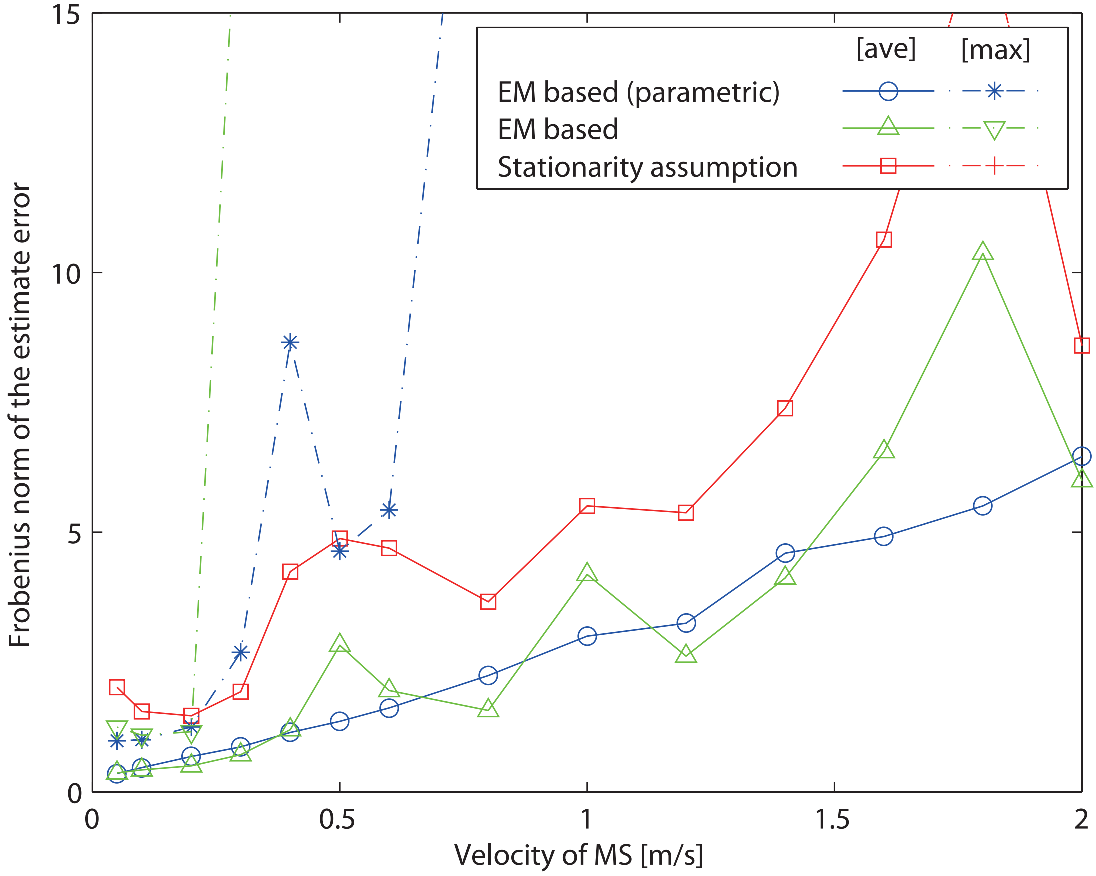

Figure 3.15: Mean squared error of channel estimate with parametric covariance matrix employed together with antenna selection systems.

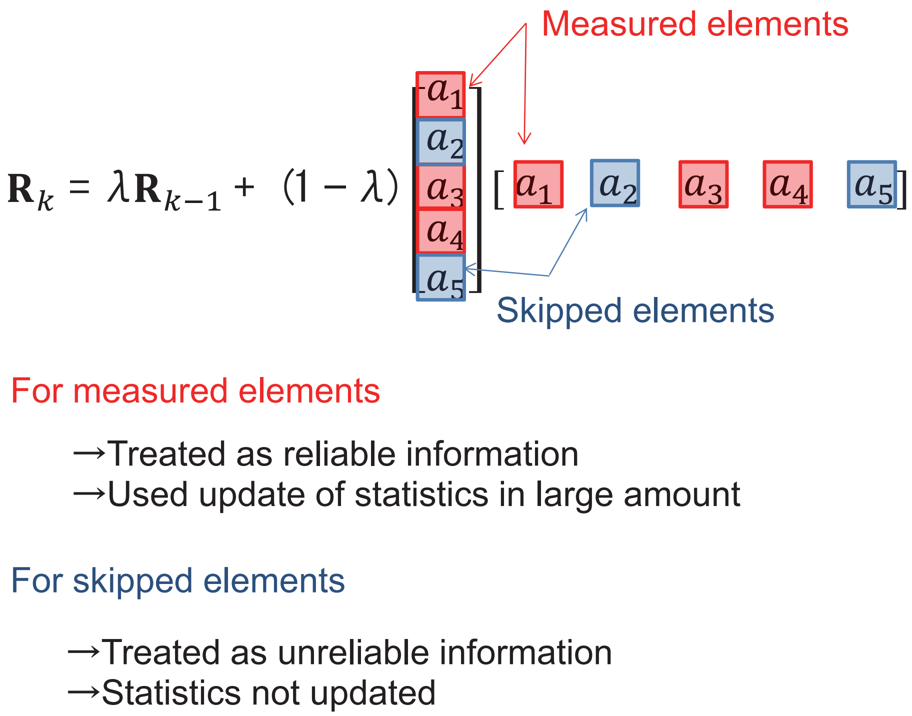

Figure 3.16: A brief explanation of why the EM-based method has benefit when time-varying partial measurement matrix was employed.

utilized for estimation of the channel statistics. In the EM based parameter estimation, the channels for measured antenna elements are treated as highly reliable information, thus it is utilized for a large update of existing channel statistic. In contrast, the channels for unmeasured antenna elements are treated as unreliable, so the information is only utilized for a small update of existing statistics. Such mechanism as so far described is considered to perform effectively under the scenario involved with antenna selection system in the sense that it has varying observation matrices. (illustrated in Fig.3.16)

By utilizing the parametric covariance matrix, the maximum estimation error is reduced and becomes stable regardless of the existence of correlation model mismatch. Since the parametric covariance matrix has the smaller amount of model parameters, in the notion of bias-variance tradeoff [58], it has smaller variance which is advantageous for such a severe statistics estimation problem.

3.6 Concluding Remarks

We proposed a random-walk based Gauss-Markov model which does not require a correlation coefficient between fading blocks, and confirmed that the model is able to exploit the channel's spatiotemporal correlation properties. In order to obtain the channel's innovation statistical parameter, we also proposed a novel online estimation method by means of the EM algorithm noticing that an observed channel state is a latent statistical variable, and discussed its characteristics.

The numerical simulation revealed that the method works well under the conditions where small forgetting factor employed. A small forgetting factor has a merit in rapid tracking capability, but causes degradation of statistical parameter because the number of channel realizations

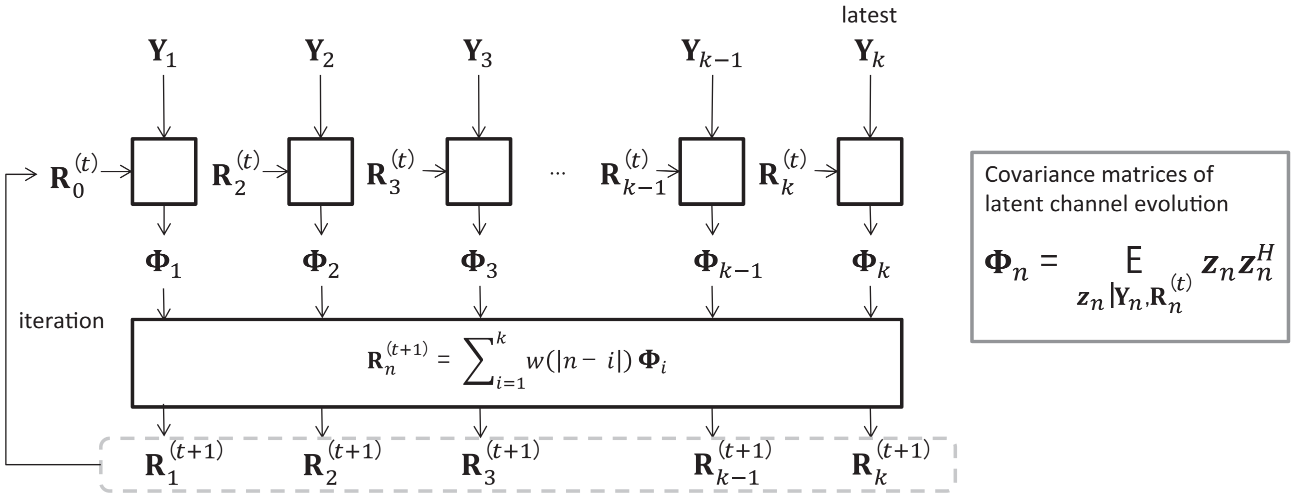

Figure 3.17: Schematics explaining improvement of parameter updating in EM algorithm.In the figure, $\mathbf{\Phi}_i \triangleq \mathrm{E}_{\boldsymbol{z}_i|\boldsymbol{Y}_i|\boldsymbol{R}_i^{(t)}} \boldsymbol{z}_i \boldsymbol{z}_i^{\mathcal{H}} = \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^{\mathcal{H}} + [(\boldsymbol{R}_i^{(t)})^{-1} + \boldsymbol{F}_i \boldsymbol{C}_i]^{-1}$. Dependence on future observations enables update of past parameters for each iteration.

that contribute to the statistics estimation becomes small. The proposed method is considered to mitigate such effects by expectation average of the latent channel realizations. This characteristics lead to a rapid convergence of channel statistics even if a sudden change of statistics happens caused by shadowing, etc. On the other hand, if the channel is stationary enough to allow large forgetting factor, the method has no advantage over the stationarity assumption method which does not have the E-step.

The proposed method has shortcoming in computational costs. We investigated that the computations can be reduced by imposing constraints onto spatial correlation modeling. Due to the model mismatch, estimation capability degrades to some extent, but the approach is considered to be still useful depending on system model and situations. For practical implementation, further reduction will be preferred.

As future tasks, the proposed method is considered to be extendable to the problem of both data sequence and channel state estimation without using fixed training sequence as discussed in the literatures [50–53].

As explained in the derivation of (3.33), in the case of estimation of time-varying channel, only a single EM iteration was employed since the innovation statistics only depends on the past inputs. In order to improve this, it is considered that the dependence of the statistics estimation should be extended to future inputs as well. This means that the likelihood function of (3.19) is modified as follows:

$$L(\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_N \mid \boldsymbol{R}_k) = \log \prod_{i=1}^{N} \mathcal{CN}\left(\boldsymbol{z}_i \mid \boldsymbol{0}, \boldsymbol{R}_k\right)^{w(|i-k|)} \tag{3.55}$$

where $w(i), i \geq 0$ denotes decaying weight function for each realization. $w(i) = \lambda^i$ was utilized in (3.19). As illustrated in Fig.3.17, if we modify weights so that the statistical parameter at time instant depends on observed inputs at future, the latest input $\boldsymbol{Y}_k$ would be able to update not only $\boldsymbol{R}_k^{(t+1)}$ but also $\boldsymbol{R}_{k-1}^{(t+1)}, \boldsymbol{R}_{k-2}^{(t+1)}, \cdots$ during EM iterations. By such improvement, we can expect enhancement of the estimation capability by increasing the number of the EM iterations. However, similar to the case of stationary in Fig.3.4, required computational cost would become very large for each iteration. An idea to reduce computational cost should be necessary for practical application.

Chapter 4

Two-Stage Training Resource Allocation for Antenna Selection Systems

This chapter addresses the issue of MIMO channel estimation with the aid of a rough estimate of temporal correlation statistics where MIMO antenna selection system is employed. Under the temporally correlated channel, proposed method controls allocation of length of training symbols for each antenna element so that the elements which are likely to be selected are estimated more accurately than the other elements. In order to utilize this estimation scheme effectively, this paper also proposes the antenna selection method which takes account of the difference of channel estimation accuracy among antenna elements. When the ML channel estimation was adopted, the proposed selection method has almost the same computational cost as the conventional selection methods. The proposed method does not rely on spatial correlation statistics which is not always stable in real propagation scenarios. The numerical simulation using 3GPP-SCM revealed that the proposed method works effectively under shorter training symbols employed.

4.1 Introduction

MIMO transmission attracts much attention since it achieves much higher transmission capability by carrying an information via spatial degrees of freedom of the radio propagation channel [1, 2]. Especially for the multi stream MIMO transmission, its performance is dramatically affected by whether the surrounding propagation environment is capable of conveying spatial degrees of freedom effectively or not. In order to improve such inherent channel capabilities, MIMO antenna selection system has been proposed which enables to control the channel conditions by actively changing the antenna subset to use so that the rank of the channel matrix can be improved [14]. In addition, the system can save hardware cost by reducing expensive RF chains while not losing its performance to a large extent. In the system, the antenna subset for transmission is optimally chosen to enhance the MIMO channel capacity. A number of antenna selection algorithms have been proposed in order to perform a near-optimal selection with a smaller computational complexity [5, 5–9]. Many of them require a full channel state information, so the channel measurement must precede the antenna selection.

Generally in pilot symbol aided channel estimation with a constant noise power, the more energy is put into the transmission of training symbols, the lower the estimation error can be. The authors' interest is how to estimate the channel accurately within a limited set of resources.

The authors had investigated that the channel estimation of MIMO antenna selection systems is improved by using the Kalman filter under the assumptions that the temporal correlation is perfectly characterized as the Gauss-Markov model, and the spatial correlation is available [22]. However, as many conditions are imposed on the channel, the estimation method loses its robustness. Furthermore, for realistic use, estimations of many statistical parameters are also required. However, such methods relying on spatial correlation statistics are fragile for a sudden change of spatial correlation which will sometimes happen in some propagation scenarios.

This paper proposes a novel channel estimation scheme for MIMO antenna selection systems which exploits *only* the temporal correlation statistics. In antenna selection systems, the channel state information for all the antenna elements is required as a selection criteria. However, if the channel state information is estimated within the same precision for all the antenna elements, the measurements for elements which turned out to be not selected become in vain since they are only discarded. Therefore, we propose an estimation scheme which measures the channel accurately only for the elements which are likely to be selected in the next step, and the rest of the elements are coarsely estimated by shorter training symbols. In order to utilize this scheme effectively, we also consider an antenna selection method which takes account of the difference of estimation precision of each antenna element. If the ML channel estimation is adopted, the proposed selection method requires almost the same computational cost as the conventional selection method. The proposed method has robustness in the sense that it does not rely on certain mathematical models of temporal correlation, as well as spatial correlation which is sensitive to the movement of the receiver.

Literature(s)	Method	Channel statistics	Required statistical parameters	Advantage(s)	Shortcoming(s)
[23–30]	Wiener based	Spatial	Exact spatial correlation matrix	High performance	Requires exact spatial correlation statistics
[18–22]	Kalman based	Spatial & Temporal	Exact spatial correlation matrix and exact parameters of channel transition model	High performance	Requires exact model of temporal channel transition
Proposed	Training resource allocation	Temporal	Rough temporal correlation coefficient	Not require exact model of temporal channel transition	Only available for antenna selection systems, Moderate performance

Table 4.1: MIMO channel estimation methods exploiting a priori channel statistics.

It should be noted that for the estimation of time-variant frequency selective fading channels, exploiting temporal correlation is quite common among literatures. [18, 19]. However, the proposed method is substantially different from them in the points that the proposed method deals with only the resource allocation of channel estimation dedicated for antenna selection systems, and has nothing in common with the Kalman-based methods. In fact, the currently estimated channel state itself is independent with the previously estimated one. This feature is advantageous in a way that it can prevent degradation of capacity even when sudden change of channel state happens induced by shadowing, rotation of mobile station, etc. The differences of the proposed method compared to the other correlation based estimation methods are summarized in Table 4.1.

4.1.1 Mathematical Notations

Throughout this paper we will use bold-faced upper case letters to denote matrices, and bold-faced lower case letters for column vectors, light-faced letters for scalar quantities. The subscripts $\top$, $\mathcal{H}$, $*$ indicate transpose, Hermitian transpose (transpose and complex conjugate), and complex conjugate respectively. $\boldsymbol{I}_N$ denotes the $N \times N$ identity matrix. Also the inverse, trace, determinant, and Frobenius norm of the matrix $\boldsymbol{X}$ are denoted by $\boldsymbol{X}^{-1}$, $\operatorname{tr}\boldsymbol{X}$, $\det \boldsymbol{X}$, and $\|\boldsymbol{X}\|_{\mathrm{F}}$, respectively. A diagonal matrix having elements of $x_{1,1}, x_{2,2}, \cdots, x_{n,n}$ is denoted by $\operatorname{diag}[x_{1,1}, x_{2,2}, \cdots, x_{n,n}]$. The m-th row and n-th column element of the matrix $\boldsymbol{X}$ is denoted by $[\boldsymbol{X}]_{m,n}$. E_x means the expectation with respect to x. The bracket $< \cdot, \cdot >$ is used for inner product on the column vector space. Also the vector Euclidean norm is expressed as $\|\cdot\|$. Since we often discuss correlations between each matrix element, it is convenient to treat matrix as one column vector that consists of all its elements. For any $m \times n$ matrix $\boldsymbol{A} = [\boldsymbol{a}_1\, \boldsymbol{a}_2\, \cdots \boldsymbol{a}_n]$, the vec operator generates a $mn \times 1$ vector defined as

$$\operatorname{vec} \boldsymbol{A} \triangleq \left[\boldsymbol{a}_1^\top\, \boldsymbol{a}_2^\top\, \cdots \boldsymbol{a}_n^\top\right]^\top \tag{4.1}$$

where $\triangleq$ means definition. The Kronecker product $\otimes$ is required with the use of the vec operator.

4.2 System Model

4.2.1 MIMO Antenna Selection System

For simplicity, we consider antenna selection system only for the receiver side with N_{Tx} transmit antennas and N_{Rx} receive antennas, and N_{RF} RF chains satisfying $1 \leq N_{\mathrm{RF}} < N_{\mathrm{Rx}}$. For a narrowband frequency nonselective MIMO channel, if we connect the i-th RF chains to the c_i-th ($1 \leq c_i \leq N_{\mathrm{Rx}}$) antenna element, the received vector $\boldsymbol{y} \in \mathbb{C}^{N_{\mathrm{RF}}}$ can be expressed as

$$\boldsymbol{y} = \boldsymbol{A}_{\tau_k} \boldsymbol{H}_k \boldsymbol{x} + \boldsymbol{n} \tag{4.2}$$

where $\boldsymbol{n} \in \mathbb{C}^{N_{\mathrm{RF}}}$ is additive noise vector typically assumed to have a white complex Gaussian distribution with average power σ_n^2, and $\boldsymbol{x} \in \mathbb{C}^{N_{\mathrm{Tx}}}$ is the normalized transmit vector such that $\mathrm{E}_{\boldsymbol{x}} \boldsymbol{x}\boldsymbol{x}^{\mathcal{H}} = \boldsymbol{I}_{N_{\mathrm{Tx}}}$. $\boldsymbol{H}_k \in \mathbb{C}^{N_{\mathrm{Rx}} \times N_{\mathrm{Tx}}}$ is the complex channel gain matrix at time instant k. The channel is normalized such that $\mathrm{E}\|\boldsymbol{H}_k\|_{\mathrm{F}}^2 = P_r N_{\mathrm{Rx}}$ where P_r is average receive signal power for each receive antenna. Average signal to noise ratio (SNR) per receive antenna can be expressed as P_r/σ_n^2. $\boldsymbol{A}_{\tau_k}$ performs extraction and permutation of row vectors of $\boldsymbol{H}_k$. By using the antenna connection vector $\boldsymbol{\tau}_k \triangleq [c_1\, c_2 \cdots c_{N_{\mathrm{RF}}}]^\top$, $\boldsymbol{A}_{\tau_k}$ is expressed as

$$\begin{bmatrix} [\boldsymbol{H}_k]_{c_1,:} \\ [\boldsymbol{H}_k]_{c_2,:} \\ \vdots \\ [\boldsymbol{H}_k]_{c_{N_{\mathrm{RF}}},:} \end{bmatrix} = \boldsymbol{A}_{\tau_k} \boldsymbol{H}_k, \quad \boldsymbol{A}_{\tau_k} \triangleq \sum_{i=1}^{N_{\mathrm{RF}}} \mathbf{f}_i \mathbf{e}^\top_{<\boldsymbol{\tau}_k, \mathbf{f}_i>}$$

where $[\boldsymbol{H}_k]_{c_n,:}$ means the c_n-th row vector of $\boldsymbol{H}_k$, and $\mathbf{e}_i, \mathbf{f}_j$ are the so-called standard basis of $\mathbb{C}^{N_{\mathrm{Rx}}}, \mathbb{C}^{N_{\mathrm{RF}}}$, respectively. Antenna subset is selected so that the Shannon capacity of extracted channel matrix $\boldsymbol{A}_{\tau_k}\boldsymbol{H}_k$ becomes the largest of all the combinations. Since the true channel matrix is not available directly, we determine the connection $\boldsymbol{\tau}_k$ by referring to the estimated

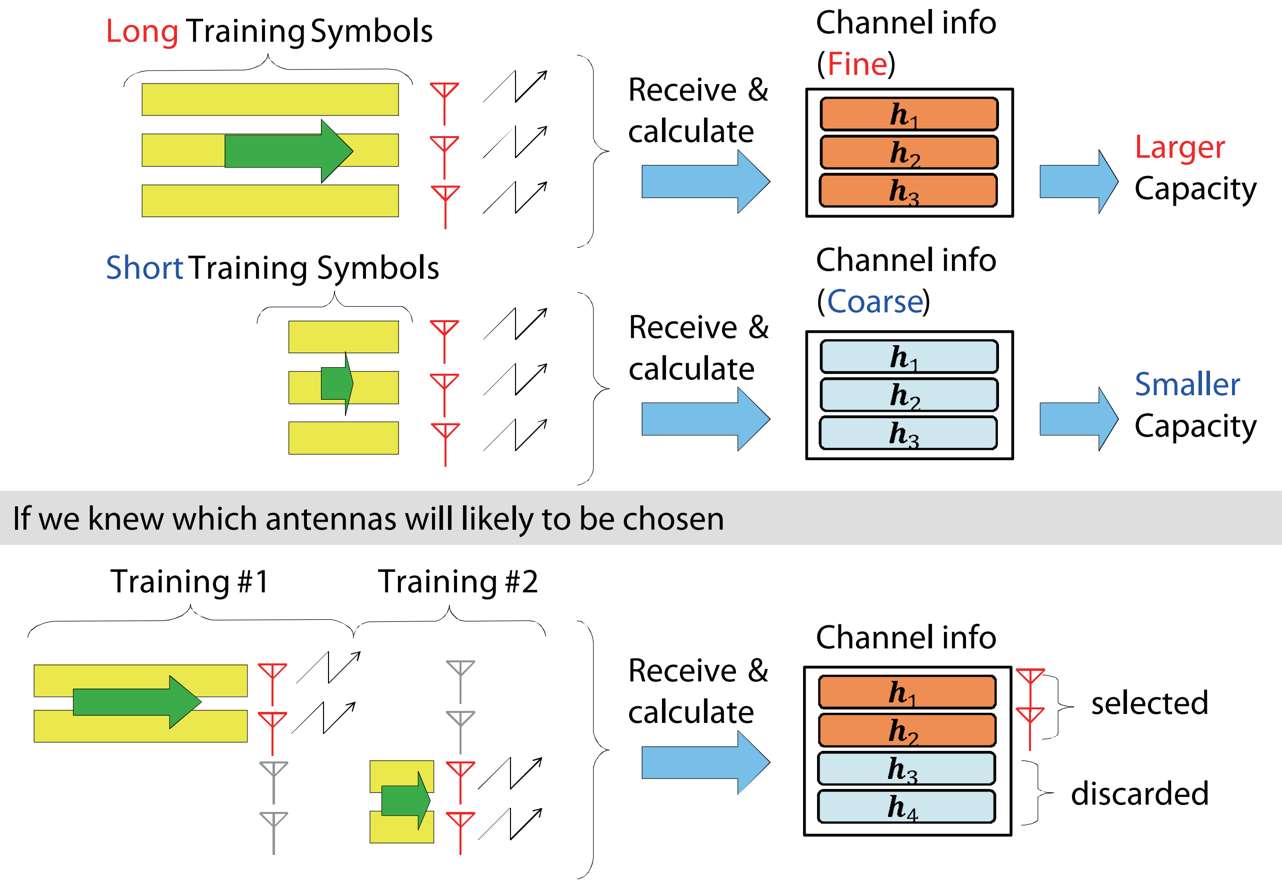

Figure 4.1: Basic idea of the proposed channel estimation scheme.

channel matrix $\widehat{\boldsymbol{H}}_k$ instead. From the Foschini-Telatar equation for equal transmit power allocation, $\boldsymbol{\tau}_k$ is determined such that:

$$\boldsymbol{\tau}_k = \arg\max_{\tau} \log_2 \det\left(\boldsymbol{I}_{N_{\mathrm{RF}}} + \boldsymbol{A}_\tau \widehat{\boldsymbol{H}}_k \widehat{\boldsymbol{H}}_k^{\mathcal{H}} \boldsymbol{A}_\tau^{\mathcal{H}}\right) \tag{4.3}$$

In order to obtain a near-optimal solution with lower computational complexity, many antenna selection algorithms have been proposed. Many of them require full instant channel state $\widehat{\boldsymbol{H}}_k$ to get selection for each time instant k. In order to keep the optimal selection in the time-varying environment, channel measurements must be frequently repeated to track the temporal changes.

4.3 Channel Estimation with a Selection Bias

In this section, we design a channel estimation scheme based on the idea of how to save the energy to measure the channels which are not to be selected. (illustrated in Fig.4.1) Concretely, the proposed method assigns longer training symbols to the antenna elements which are likely to be selected in the next step, and estimate them more accurately compared to the other antenna elements.

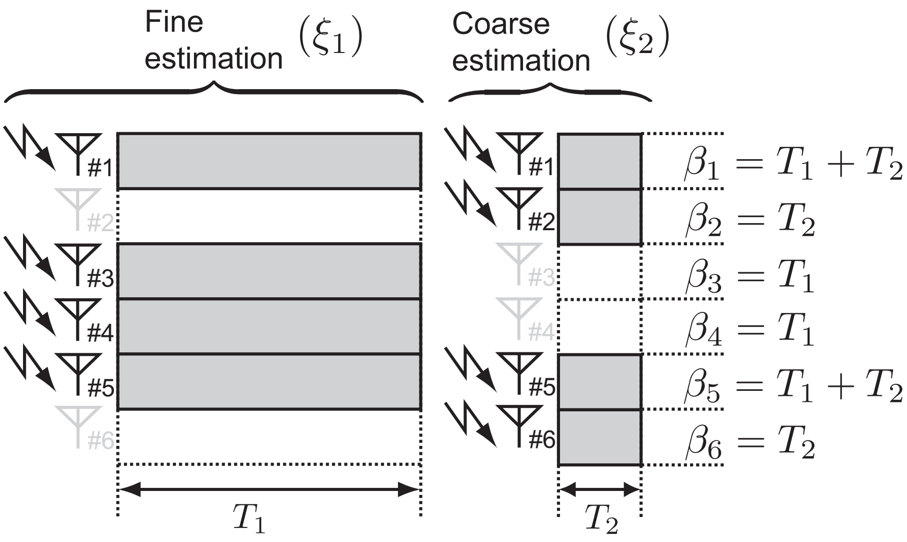

Figure 4.2: Precision control considering temporal correlation. In this example, we assume $\boldsymbol{\xi}_1 = (1, 3, 4, 5)^\top$ and $\boldsymbol{\xi}_2 = (1, 2, 5, 6)^\top$.

4.3.1 Fine and Coarse Estimation

We introduce a two stage channel estimation which consists of a fine estimation phase and a coarse estimation phase. In antenna selection systems, since the number of antenna elements is larger than the number of RF chains, receptions of training symbols must be repeated several times while changing the connection of RF switches. The repeated trainings are required to be at least $N_m = \lceil N_{\mathrm{Rx}}/N_{\mathrm{RF}} \rceil$ times where $\lceil x \rceil$ means the smallest integer greater than or equal to x. In the fine estimation phase, the measurement is done by longer training symbols, and the channel is estimated finely. The rest of measurement is utilized for the coarse estimation phase by using shorter training symbols.

In the fine estimation phase, a measured antenna subset should be chosen in such a way that they have a high probability to be selected in the next frame. In order to realize this, we introduced the assumption that "In temporally correlated channels, the antenna subset which was selected as the best combination in the $(k-1)$-th fading block is likely to have high transmission capability also in the k-th fading block". The assumption is considered reasonable because in temporally correlated channels, the channel state information does not change significantly between adjacent fading blocks.

This plot is illustrated in Fig.4.2. In the figure, we assume $N_{\mathrm{RF}} < N_{\mathrm{Rx}} \leq 2N_{\mathrm{RF}}$, so it requires at least two times measurements in order to measure all the antenna elements. Let us assume that the subset $\boldsymbol{\tau}_{k-1} = (1, 3, 4, 5)^\top$ of receiver elements was chosen for transmission at the $(k-1)$-th fading block. Then, in the k-th fading block, same subset ($\boldsymbol{\xi}_1 = \boldsymbol{\tau}_{k-1}$) is finely measured by the training symbols with length of T_1. The rest of the elements, $\boldsymbol{\xi}_2 = (1, 2, 5, 6)^\top$ are coarsely measured with length of T_2 ($T_1 > T_2$). Since N_{Rx} is not divisible by N_{RF}, $2N_{\mathrm{RF}} - N_{\mathrm{Rx}}$ elements (in the figure, the 1st and 5th elements) receive training sequence in both the fine and the coarse phases. We do not define which elements should receive symbols twice in such cases. (In simulation, we select them randomly.)

In later discussions, we denote the total length of training symbols as $N_t \triangleq T_1 + T_2$. Also we introduce a parameter α which indicates ratio between T_1 and T_2 as follows:

$$T_1 = \alpha N_t, \quad T_2 = (1-\alpha) N_t \quad (0.5 \leq \alpha < 1) \tag{4.4}$$

If $\alpha = 0.5$, the proposed method is equivalent to the conventional method.

4.4 Antenna Selection Considering Estimate Error

The channel estimation scheme explained so far has a characteristic that the estimation error varies by antenna element. In order to efficiently exploit this feature, we cannot utilize directly the antenna selection criteria of (4.3) because the criteria assumes uniform estimation error at each element.

In later discussions, we consider the best antenna selection method when the channel estimation error is generally expressed as the multivariate Gaussian distribution. Let us denote the true channel of time instant k as $\boldsymbol{H}_k$, and the corresponding estimate as $\widehat{\boldsymbol{H}}_k$. The behavior of the estimation error is fully characterized by the error covariance matrix defined as

$$\boldsymbol{P}_k \triangleq \mathrm{E}_{\boldsymbol{n}} \mathrm{E}_{\boldsymbol{H}_k} \operatorname{vec}\left(\boldsymbol{H}_k - \widehat{\boldsymbol{H}}_k\right) \left[\operatorname{vec}\left(\boldsymbol{H}_k - \widehat{\boldsymbol{H}}_k\right)\right]^{\mathcal{H}} \tag{4.5}$$

where E means expectation with respect to its subscript. Although obtaining $\boldsymbol{P}_k$ seems difficult, we will show later that it can be calculated easily like (4.20) by only using length of training sequences if the ML channel estimation was adopted.

4.4.1 Antenna Selection Method Maximizing Lower-Bound of Mutual Information

Although how to incorporate the estimation error into the antenna selection criterion can be considered in various ways, desired properties are its validity and feasibility of optimization. In this paper, we propose to maximize the lower bound of mutual information. This means that antenna subset is chosen such that the capacity degradation in the worst case scenario is minimized.

4.4.2 Lower-Bound of Mutual Information

According to [3], we derive the lower-bound [12] of degraded channel capacity caused by the estimate error of channel. This method is originally introduced to evaluate the degraded capacity under the continuous fading channel characterized by temporal correlation coefficient.

If the receiver employes a synchronized detection, i.e. $\widehat{\boldsymbol{H}}_k$ is used at the Rx side as if it were a true channel matrix, (4.2) can be separated into the two terms [16] as

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{A}_{\tau_k} \widehat{\boldsymbol{H}}_k \boldsymbol{x} + \boldsymbol{A}_{\tau_k} \left(\boldsymbol{H}_k - \widehat{\boldsymbol{H}}_k\right) \boldsymbol{x} + \boldsymbol{n} \\ &= \boldsymbol{A}_{\tau_k} \widehat{\boldsymbol{H}}_k \boldsymbol{x} + \hat{\boldsymbol{n}} \end{aligned} \tag{4.6}$$

where the first term contains information received at the Rx, and the second term $\hat{\boldsymbol{n}}$ is called the effective noise term which is regarded as an additive noise caused by the estimate error, because a synchronized detection virtually cannot exploit the information in this term. The covariance matrix of $\hat{\boldsymbol{n}}$ is expressed as

$$\begin{aligned} \boldsymbol{\Phi} &\triangleq \mathrm{E}_{\boldsymbol{x}} \mathrm{E}_{\boldsymbol{n}} \mathrm{E}_{\widehat{\boldsymbol{H}}_k} \hat{\boldsymbol{n}} \hat{\boldsymbol{n}}^{\mathcal{H}} \\ &= \mathrm{E}_{\widehat{\boldsymbol{H}}_k} \boldsymbol{A}_{\tau_k} \widehat{\boldsymbol{H}}_k \widehat{\boldsymbol{H}}_k^{\mathcal{H}} \boldsymbol{A}_{\tau_k}^{\mathcal{H}} + \sigma_n^2 \boldsymbol{I}_{N_{\mathrm{RF}}} \\ &= \boldsymbol{A}_{\tau_k} \boldsymbol{R}_k \boldsymbol{A}_{\tau_k}^{\mathcal{H}} + \sigma_n^2 \boldsymbol{I}_{N_{\mathrm{RF}}} \end{aligned} \tag{4.7}$$

where $\boldsymbol{R}_k$ is determined by $\boldsymbol{P}_k$ of (4.5) as described later. The mutual information of this channel has the lower-bound expressed as

$$\mathcal{I}(x, y|\widehat{\boldsymbol{H}}_k) \geq \log_2 \det \left(\boldsymbol{I}_{N_{\mathrm{RF}}} + \boldsymbol{A}_{\tau_k} \widehat{\boldsymbol{H}}_k \widehat{\boldsymbol{H}}_k^{\mathcal{H}} \boldsymbol{A}_{\tau_k}^{\mathcal{H}} \boldsymbol{\Phi}^{-1} \right) \triangleq C_{\mathrm{LB}} \,. \tag{4.8}$$

Instead of directly maximizing the instantaneous channel capacity from the estimated channel as in (4.3), our proposal is to maximize C_{LB} of (4.8).

As for the optimization method, the greedy algorithm is available as well as the case of (4.3). In the case of criteria (4.3), the greedy algorithm can be implemented with significantly small computations by utilizing the Sherman-Morrison formula [5]. However, in the case of C_{LB}, generally we cannot simplify the computation like that due to the additional term of $\boldsymbol{\Phi}^{-1}$.

Let us discuss the special case when $\boldsymbol{R}_k$ is expressed as a diagonal form. Generally, such situation arises when $\boldsymbol{P}_k$ has the form of

$$\boldsymbol{P}_k = \sum_{i=1}^{N} \boldsymbol{B}_i \otimes \boldsymbol{D}_i \tag{4.9}$$

where N is any positive integer, $\{\boldsymbol{B}_i\}_{i=1}^{N}$ is any $N_{\mathrm{Tx}} \times N_{\mathrm{Tx}}$ positive semi-definite matrices, and $\{\boldsymbol{D}_i\}_{i=1}^{N}$ is any $N_{\mathrm{Rx}} \times N_{\mathrm{Rx}}$ diagonal matrices, respectively. In this case, substituting the relationship of

$$\boldsymbol{\Phi}^{-1/2} = \boldsymbol{A}_{\tau_k} \left(\boldsymbol{R}_k + \sigma_n^2 \boldsymbol{I}_{N_{\mathrm{Rx}}} \right)^{-1/2} \boldsymbol{A}_{\tau_k}^{\mathcal{H}} \tag{4.10}$$

into (4.8) yields,

$$\begin{aligned} C_{\mathrm{LB}} = \log_2 \det \Big[& \boldsymbol{I}_{N_{\mathrm{RF}}} + \boldsymbol{A}_{\tau_k} \left(\boldsymbol{R}_k + \sigma_n^2 \boldsymbol{I}_{N_{\mathrm{Rx}}} \right)^{-1/2} \\ & \boldsymbol{A}_{\tau_k}^{\mathcal{H}} \boldsymbol{A}_{\tau_k} \widehat{\boldsymbol{H}}_k \widehat{\boldsymbol{H}}_k^{\mathcal{H}} \boldsymbol{A}_{\tau_k}^{\mathcal{H}} \boldsymbol{A}_{\tau_k} \left(\boldsymbol{R}_k + \sigma_n^2 \boldsymbol{I}_{N_{\mathrm{Rx}}} \right)^{-1/2} \boldsymbol{A}_{\tau_k}^{\mathcal{H}} \Big] \,. \end{aligned} \tag{4.11}$$

Since $\boldsymbol{A}_{\tau_k} \left(\boldsymbol{R}_k + \sigma_n^2 \boldsymbol{I}_{N_{\mathrm{Rx}}} \right)^{-1/2} \boldsymbol{A}_{\tau_k}^{\mathcal{H}} \boldsymbol{A}_{\tau_k}$ is equivalent to $\boldsymbol{A}_{\tau_k} \left(\boldsymbol{R}_k + \sigma_n^2 \boldsymbol{I}_{N_{\mathrm{Rx}}} \right)^{-1/2}$, replacing this and denoting $\boldsymbol{R}_k = \mathrm{diag}[r_1 \, r_2 \, \cdots \, r_{N_{\mathrm{Rx}}}]$ yields,

$$C_{\mathrm{LB}} = \log_2 \det \left(\boldsymbol{I}_{N_{\mathrm{RF}}} + \boldsymbol{A}_{\tau_k} \widetilde{\boldsymbol{H}}_k \widetilde{\boldsymbol{H}}_k^{\mathcal{H}} \boldsymbol{A}_{\tau_k}^{\mathcal{H}} \right) \tag{4.12}$$

$\widetilde{\boldsymbol{H}}_k$ is defined as

$$\widetilde{\boldsymbol{H}}_k \triangleq \begin{bmatrix} [\widehat{\boldsymbol{H}}_k]_{1,:} \, / \sqrt{r_1 + \sigma_n^2} \\ [\widehat{\boldsymbol{H}}_k]_{2,:} \, / \sqrt{r_2 + \sigma_n^2} \\ \vdots \\ [\widehat{\boldsymbol{H}}_k]_{N_{\mathrm{Rx}},:} \, / \sqrt{r_{N_{\mathrm{RF}}} + \sigma_n^2} \end{bmatrix} . \tag{4.13}$$

where r_i is the i-th diagonal element of $\boldsymbol{R}_k$. (4.12) is exactly the same form as (4.3) except that $\widehat{\boldsymbol{H}}_k$ is replaced by $\widetilde{\boldsymbol{H}}_k$. This implies that the maximization of C_{LB} can be achieved by exactly the same method as a number of conventional antenna selection methods proposed so far. For example, applying the fast antenna subset selection [5] can save much computational costs. Additional calclations required for (4.13) is only a division of row vectors by corresponding effective noise.

Intuitively, (4.13) can be interpreted that the antenna elements which have larger estimation error tend not to be selected as compared to the other elements.

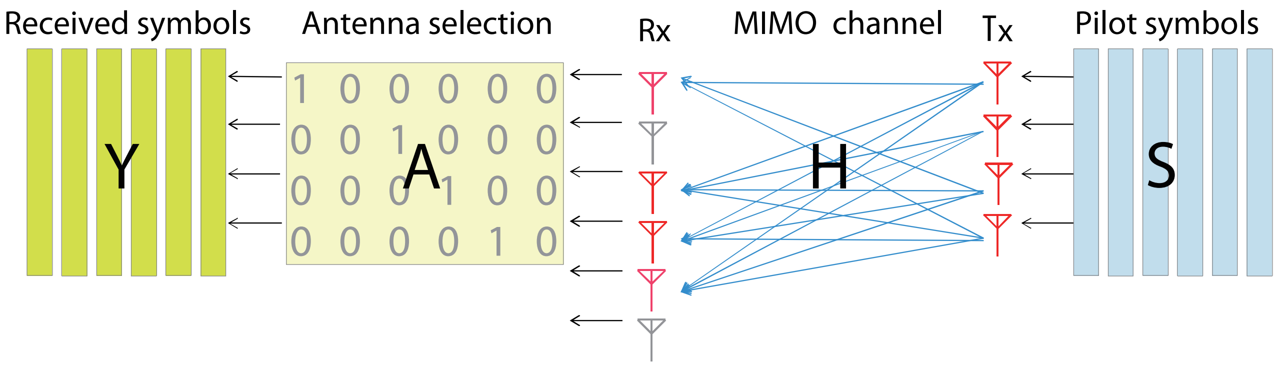

Figure 4.3: Pilot symbol aided channel estimation.

4.5 Fast Antenna Subset Selection in the Case of ML Channel Estimation

In MIMO antenna selection systems, because we have only a small set of RF chains available at the same time compared to the all antenna elements, if we want to obtain the full channel state, measurements should be repeated while changing the connection of the RF switches, required at least $N_m = \lceil N_{\mathrm{Rx}}/N_{\mathrm{RF}} \rceil$ times where $\lceil x \rceil$ means the smallest integer not exceeded by x. Now we consider the case where length of training sequence is different for each repeated measurement. Furthermore, if N_{RF} is not divisible by N_{RF}, some elements receive training sequence two times. Such overlapped measurements also yield the differences of the estimate error for each element. In this section we show that $\boldsymbol{R}_k$ becomes diagonal if the ML channel estimation is employed under a block fading channel.

4.5.1 Channel Observation Model under a Block Fading Channel

We assume for each fading block, training is repeated N_m times while changing the antenna connection as $\boldsymbol{\xi}_p\,(1 \le p \le N_m)$. Let $\boldsymbol{s}_1^{(p)}, \boldsymbol{s}_2^{(p)}, \cdots, \boldsymbol{s}_{T_p}^{(p)} \in \mathbb{C}^{N_{\mathrm{Tx}}}$ be normalized training symbols, each of them being launched from Tx side in number order at fading block k. And let T_p denote the length of training symbols for the p-th measurement. They are normalized such that $\|\boldsymbol{s}_i^{(p)}\|^2 = 1$ i.e. $\|\boldsymbol{S}_p\|_F^2 = T_p$ where $\boldsymbol{S}_p \triangleq [\boldsymbol{s}_1^{(p)}\ \boldsymbol{s}_2^{(p)}\ \cdots\ \boldsymbol{s}_{T_p}^{(p)}]$. Launched symbols are caught at the Rx side with its RF switches connected as $\boldsymbol{\xi}_p$. Then, the received sequences are written as

$$\boldsymbol{Y}_p = \boldsymbol{A}_{\xi_p} \boldsymbol{H}_k \boldsymbol{S}_p + \boldsymbol{N}_p\,. \tag{4.14}$$

where $\boldsymbol{Y}_p \triangleq [\boldsymbol{y}_1^{(p)}\ \boldsymbol{y}_2^{(p)}\ \cdots\ \boldsymbol{y}_{T_p}^{(p)}]$ and $\boldsymbol{N}_p$ are $N_{\mathrm{RF}} \times T_p$ matrices and each of them consisting of received vectors and noise vectors, respectively. (illustrated as Fig.4.3) $\boldsymbol{y}_i^{(p)}$ is the received vector corresponding to the transmitted vector $\boldsymbol{s}_i^{(p)}$. Applying vec operator on both sides of the above equation yields,

$$\operatorname{vec} \boldsymbol{Y}_p = \left(\boldsymbol{S}_p^\top \otimes \boldsymbol{A}_{\xi_p}\right) \operatorname{vec} \boldsymbol{H}_k + \operatorname{vec} \boldsymbol{N}_p\,. \tag{4.15}$$

Now, our aim is to obtain the best estimate of $\boldsymbol{H}_k$ from the set of received signals $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots, \boldsymbol{Y}_{N_m}$. The matrix equations of (4.15) can be united as

$$\boldsymbol{\varphi} = \boldsymbol{\Psi} \operatorname{vec} \boldsymbol{H}_k + \boldsymbol{u} \tag{4.16}$$

where $\boldsymbol{\varphi}$, $\boldsymbol{\Psi}$, and $\boldsymbol{u}$ are defined as

$$\boldsymbol{\varphi} \triangleq \sum_{i=1}^{N_m} \mathbf{e}_i \otimes \operatorname{vec} \boldsymbol{Y}_i, \quad \boldsymbol{\Psi} \triangleq \sum_{i=1}^{N_m} \mathbf{e}_i \otimes \boldsymbol{S}_i^\top \otimes \boldsymbol{A}_{\xi_i}$$
$$\boldsymbol{u} \triangleq \sum_{i=1}^{N_m} \mathbf{e}_i \otimes \operatorname{vec} \boldsymbol{N}_i\,, \tag{4.17}$$

respectively. $\mathbf{e}_i$ is the i-th standard basis of column vector space $\mathbb{R}^{N_m}$, and $\boldsymbol{\Psi}$ is an observation matrix of this observation model.

4.5.2 ML Channel Estimation

If the spatial correlation characteristics of the channel is not available, and also the noise is white gaussian, it has been proven that the best linear estimate of a channel is obtained by pseudoinverse of $\left(\boldsymbol{S}_p^\top \otimes \boldsymbol{A}_{\xi_p}\right)$in conjunction with orthogonal training sequence ($\boldsymbol{S}_p \boldsymbol{S}_p^{\mathcal{H}} = T_p / N_{\mathrm{Tx}} \boldsymbol{I}_{N_{\mathrm{Tx}}}$) [13]. Such estimation is called the maximal likelihood (ML) estimation [48] [49]. In this case, the origin of channel estimation error is only additive noise term $\boldsymbol{N}_p$ of (4.15). The estimate is obtained as

$$\operatorname{vec} \widehat{\boldsymbol{H}} = \boldsymbol{X}_{\mathrm{ML}}\, \boldsymbol{\varphi} \tag{4.18}$$

where $\boldsymbol{X}_{\mathrm{ML}}$ is a $N_{\mathrm{Tx}} N_{\mathrm{Rx}} \times N_{\mathrm{RF}}(T_1 + \cdots + T_p)$ matrix. If the orthogonal training sequence ($\boldsymbol{S}_p \boldsymbol{S}_p^{\mathcal{H}} = T_p / N_{\mathrm{Tx}} \boldsymbol{I}_{N_{\mathrm{Tx}}}$) was employed, such $\boldsymbol{X}_{\mathrm{ML}}$ is expressed as

$$\begin{aligned} \boldsymbol{X}_{\mathrm{ML}} = \boldsymbol{\Psi}^\dagger = \left(\boldsymbol{\Psi}^{\mathcal{H}} \boldsymbol{\Psi}\right)^\dagger \boldsymbol{\Psi}^{\mathcal{H}} &= \left(\sum_{i=1}^{N_m} \boldsymbol{S}_i^* \boldsymbol{S}_i^\top \otimes \boldsymbol{A}_{\xi_i}^{\mathcal{H}} \boldsymbol{A}_{\xi_i} \right)^\dagger \boldsymbol{\Psi}^{\mathcal{H}} \\ &= N_{\mathrm{Tx}} \sum_{i=1}^{N_m} \mathbf{e}_i^\top \otimes \boldsymbol{S}_i^* \otimes \operatorname{diag} \left[\frac{1}{\beta_1} \, \frac{1}{\beta_2} \, \frac{1}{\beta_3} \, \cdots \, \frac{1}{\beta_{N_{\mathrm{Rx}}}} \right] \boldsymbol{A}_{\xi_i}^{\mathcal{H}} \end{aligned} \tag{4.19}$$

Therefore $\boldsymbol{P}_k$ of (4.5) is

$$\begin{aligned} \boldsymbol{P}_k &= \sigma_n^2 \boldsymbol{X}_{\mathrm{ML}} \left(\boldsymbol{X}_{\mathrm{ML}}\right)^{\mathcal{H}} = \sigma_n^2 \left(\boldsymbol{\Psi}^{\mathcal{H}} \boldsymbol{\Psi}\right)^\dagger \boldsymbol{\Psi}^{\mathcal{H}} \boldsymbol{\Psi} \left(\boldsymbol{\Psi}^{\mathcal{H}} \boldsymbol{\Psi}\right)^\dagger \\ &= \sigma_n^2 \left(\boldsymbol{\Psi}^{\mathcal{H}} \boldsymbol{\Psi}\right)^\dagger = \sigma_n^2 N_{\mathrm{Tx}} \boldsymbol{I}_{N_{\mathrm{Tx}}} \otimes \operatorname{diag} \left[\frac{1}{\beta_1} \, \frac{1}{\beta_2} \, \frac{1}{\beta_3} \, \cdots \, \frac{1}{\beta_{N_{\mathrm{Rx}}}} \right] \end{aligned} \tag{4.20}$$

where β_i means the total length of received training symbols at the i-th Rx element in the k-th fading block. If β_i equals to zero, the term $1/\beta_i$ should be replaced by zero. For any positive semi-definite matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, if $\boldsymbol{X}$ obeys $\operatorname{vec} \boldsymbol{X} \sim \mathcal{CN}\left(\mathbf{0},\, \boldsymbol{A} \otimes \boldsymbol{B}\right)$, it holds that $\mathrm{E}_{\boldsymbol{X}} \boldsymbol{X} \boldsymbol{X}^{\mathcal{H}} = (\operatorname{tr} \boldsymbol{A}) \boldsymbol{B}$. By utilizing this relationship, we can obtain the $\boldsymbol{R}_k$ in (4.7) as follows:

$$\boldsymbol{R}_k = \sigma_n^2 N_{\mathrm{Tx}}^2 \operatorname{diag} \left[\frac{1}{\beta_1} \, \frac{1}{\beta_2} \, \frac{1}{\beta_3} \, \cdots \, \frac{1}{\beta_{N_{\mathrm{Rx}}}} \right] \tag{4.21}$$

Since $\boldsymbol{R}_k$ is diagonal, the optimization with respect to the criteria C_{LB} is tractable with much smaller computations. In this case, r_i in (4.13) is expressed as follows:

$$r_i = \sigma_n^2 N_{\mathrm{Tx}}^2 / \beta_i \tag{4.22}$$

4.6 Simulation

4.6.1 Verification by 3GPP SCM

We verified the effectiveness of the proposed method by numerical simulation. In order to generate a temporally correlated channel, we resort to the 3GPP Spatial Channel Model [45]. The "Urban micro" scenario was selected, and all other parameters are default settings except a number of antenna elements, sampling interval, and velocity of the mobile station. The path delays calculated by the SCM are transformed into a narrowband transfer function at 2GHz. This operation corresponds to deal with a single carrier of the OFDM systems under frequency selective fading channel.

Fig.4.4 depicts an ensemble average of channel capacity and the lower-bound of capacity for different velocities of the mobile station. The channel capacity was calculated by changing initial channel state 400 times, and correlated channels are generated 80 times for each initial channel state. The results are obtained by averaging over them. In the conventional method, we let $\alpha = 0.5$, and antenna subset is selected not considering estimate error as (4.3). In the proposed method, in order to exploit the temporal correlation, we let $\alpha = 0.8$, and utilized the selection criteria of (4.12). For fair comparison, the number of channel measurements per unit time is chosen as the same value. From the graph, we can see that as the velocity of the mobile station becomes slower, the capacity is enhanced.

4.6.2 Analysis by Simple Channel Model

In order to obtain more comprehensible results, we also investigated by using a simple single tapped Gauss-Markov model. The updating equation is

$$\boldsymbol{H}_k = \rho \boldsymbol{H}_{k-1} + \sqrt{1-\rho^2}\boldsymbol{X} \tag{4.23}$$

where ρ $(0 \leq \rho \leq 1)$ is the temporal correlation coefficient between adjacent fading blocks, and $\boldsymbol{X}$ is a random matrix which obeys $\operatorname{vec}\boldsymbol{X} \sim \mathcal{CN}(\mathbf{0}, P_r/N_{\mathrm{Tx}}\boldsymbol{I}_{N_{\mathrm{Tx}}N_{\mathrm{Rx}}})$.

4.6.3 Channel Estimation

Estimated channel is calculated as

$$\operatorname{vec}\widehat{\boldsymbol{H}}_k = \operatorname{vec}\boldsymbol{H}_k + \boldsymbol{z}, \quad \boldsymbol{z} \sim (\mathbf{0}, \boldsymbol{P}_k) \tag{4.24}$$

where $\boldsymbol{z}$ is additive estimate error. $\boldsymbol{P}_k$ is directly calculated by using (4.20). Therefore, N_t is treated only as the index of total power of training sequences. In realistic situation, in order to observe all degrees of freedom in the channel, the length of training sequence must be chosen not to be less than N_{Tx}. For example, if $\alpha = 0.8$ and $N_{\mathrm{Tx}} = 4$, in order to satisfy $T_2 \geq 4$, $N_t \geq 20$ is necessary as long as the transmitted power is fixed to P_r.

Fig.4.5 depicts an ensemble average of channel capacity and the lower-bound of capacity for different temporal correlation(ρ) where $N_t = 20$. Fig.4.7 depicts the capacity for different N_t where $\rho = 0.95$. For both simulations, other configurations are the same as the case of Fig.4.5.

As the temporal correlation becomes higher, the proposed method achieves better performance than the conventional method. We can observe 3 to 5 percent improvement for both ensemble mean and lower-bound of capacity. In the method 2, although the criteria which we directly maximize is the lower-bound of capacity, by selecting antennas to raise the worst case of capacity, the ensemble mean of capacity increased as well.

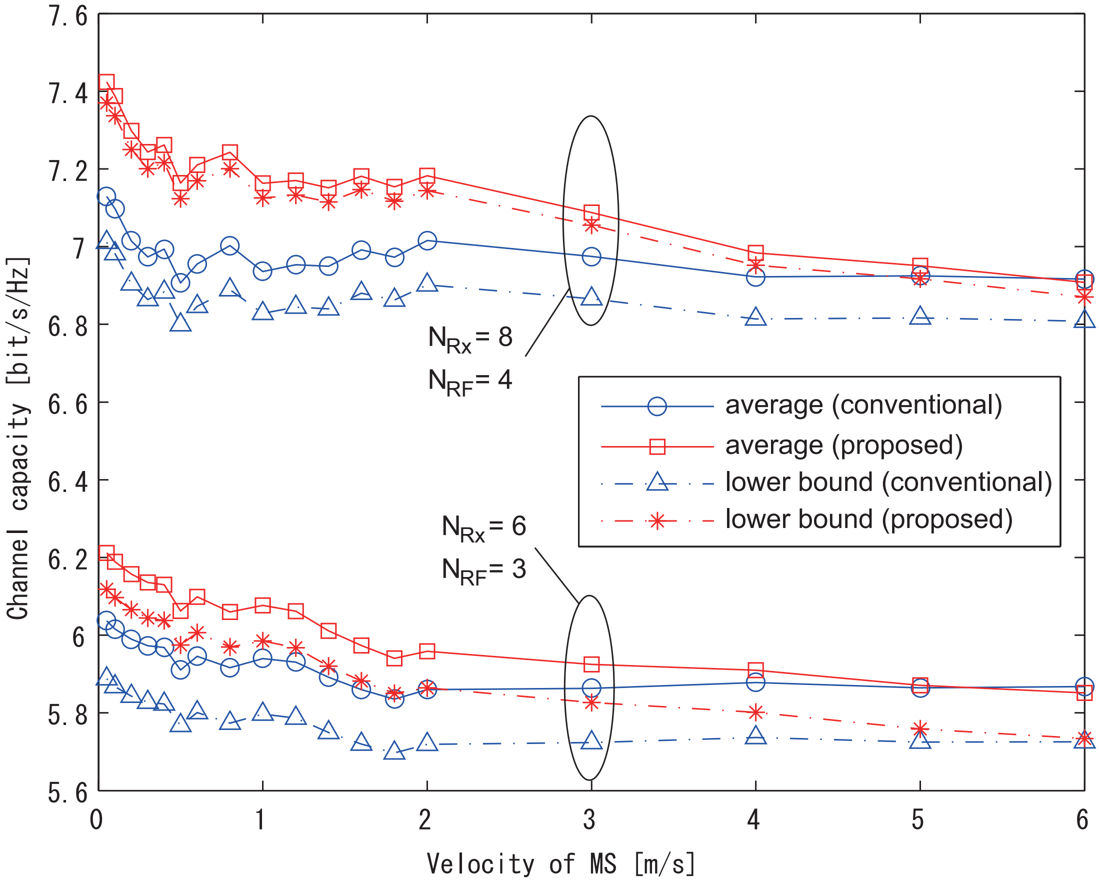

Figure 4.4: Channel capacity for different velocities of MS. ($\alpha = 0.8$, SNR=8dB)

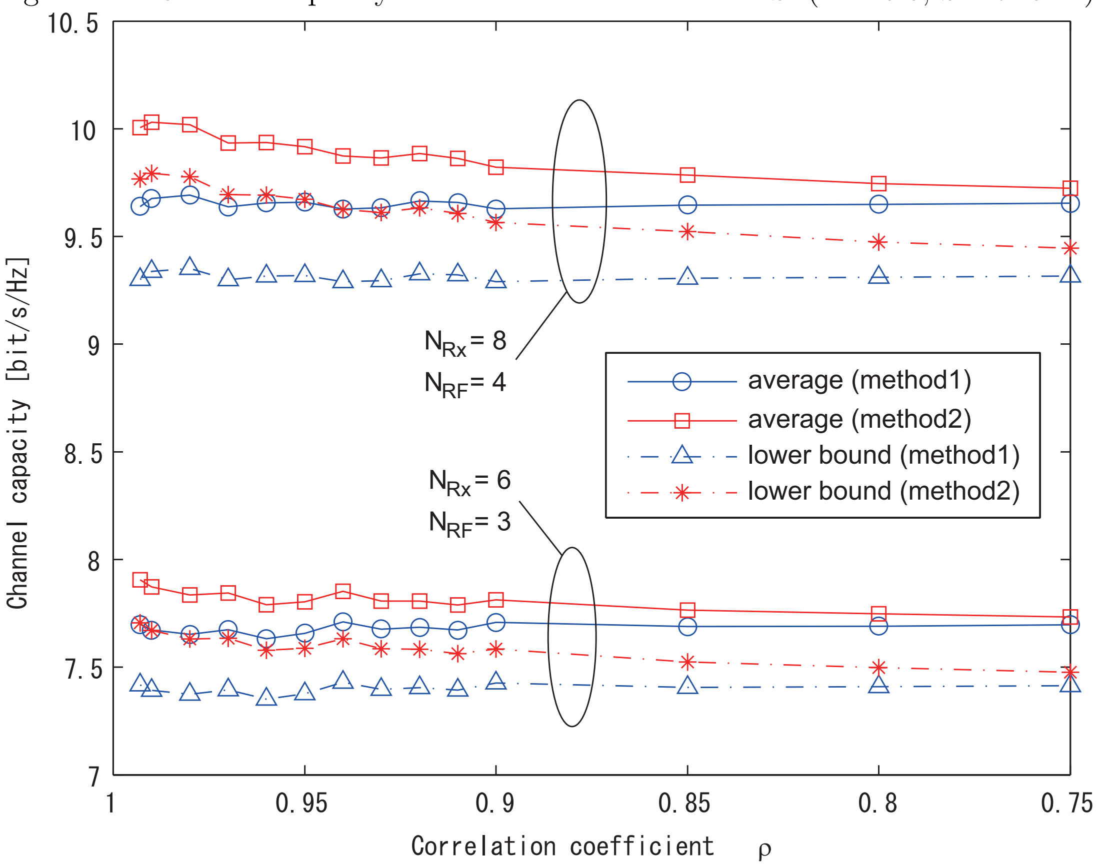

Figure 4.5: Channel capacity for different ρ. ($\alpha = 0.8$, SNR=8dB)

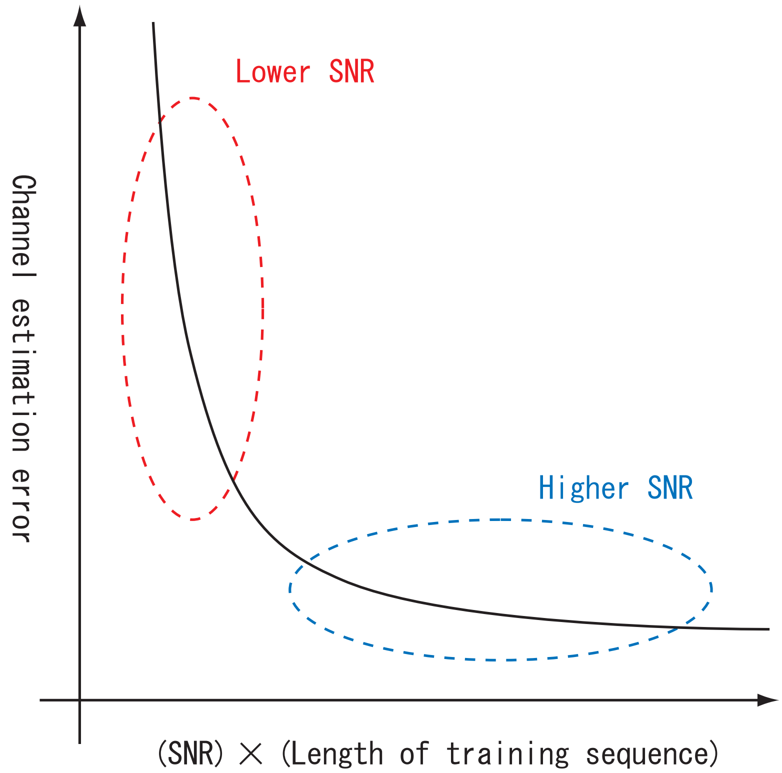

Figure 4.6: Tradeoff between total power for measurement and estimation error. A reciprocal relationship holds in case of ML estimation.

From Fig.4.7, as the length of training symbols becomes larger, the effectiveness of proposed method becomes weaker. As for the average SNR, more improvement of capacity was observed in the lower SNR conditions (less than 8dB). The superiority of this method is weakened under relatively high SNR conditions (more than 15dB). The reason for this phenomenon is explained as follows. Under situations having either high SNR or longer training symbols, channel estimation can be achieved easily with high estimation accuracy. (Fig. 4.6) In such easy cases, it is difficult for the proposed method to exhibit its superiority since both estimation stages (fine and coarse) have small differences in estimation precision.

The effectiveness of this method depends on channel conditions such as ρ and SNR. In such poor conditions, the method marks almost the same capacity as the naive antenna selection. But even in the worst case, the capacity of proposed method does not become worse than the naive method unless the choice of α becomes extremely close to either 0 or 1.

4.6.4 Optimal Choice of α

Fig.4.8 depicts the lower-bound of capacity for different temporal correlations and α. We can find that as the temporal correlation of channel becomes higher, larger α is suitable. Presumably, the reason of this phenomenon is that the antennas measured in the fine estimation phase tend to be selected. Even at extremely high temporal correlations, the optimal α appears to be at most 0.75. The reason for this is that using too large α results in inhibition of change of antenna combination. This is because, in such situations, the channels estimated in the coarse estimation stage become too coarse to be selected. Therefore, too large α spoils merits of antenna selection systems.

On the other hand, $\alpha = 0.5$ is suitable for the lower temporal correlations. This makes sense because we cannot predict in advance which antennas are selected in that case, hence the channel should be estimated equally for all the antenna elements. In the presence of temporal correlations below 0.5, $\alpha = 0.5$ appears to be the optimal.

We can observe that the proposed method works the best at especially high temporal correlations such as $\rho > 0.95$. From the figure, small deviation of α from its optimal value does not yield much degradations in capacity. It is advantageous for us, because choosing α has robustness. This implies that a precise estimation of temporal correlation is not necessary in order to determine α.

4.7 Concluding Remarks

We proposed a novel channel estimation scheme for MIMO antenna selection systems as well as a new antenna selection criteria dedicated for the method. The simulation revealed that the proposed method works more effectively when shorter length of training sequence is used under the lower average SNR conditions. Though an improvement of channel capacity is only a few percent, the implementation of the scheme requires only a small modification to conventional antenna selection systems.

As future tasks, an easy method which measures the degree of temporal correlation by small computation should be investigated in order to implement this scheme.

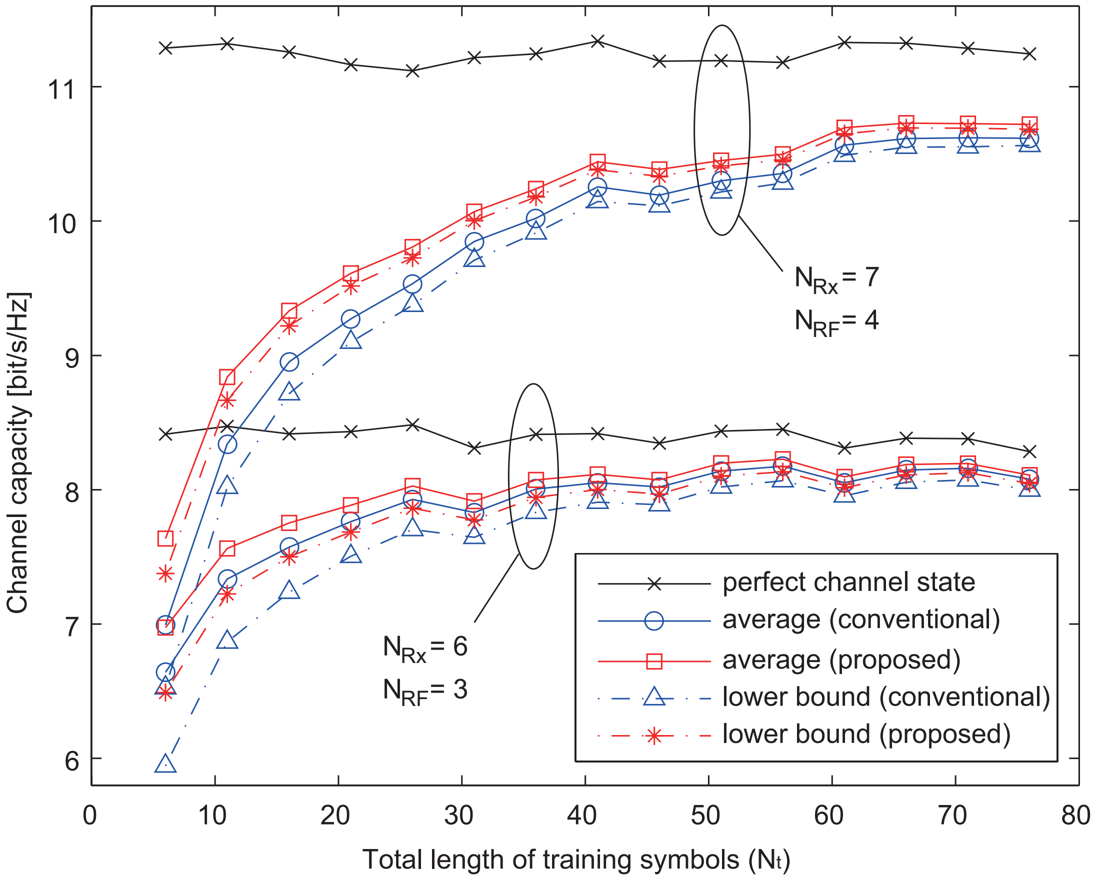

Figure 4.7: Channel capacity for different N_t. ($\alpha = 0.8$, SNR=8dB)

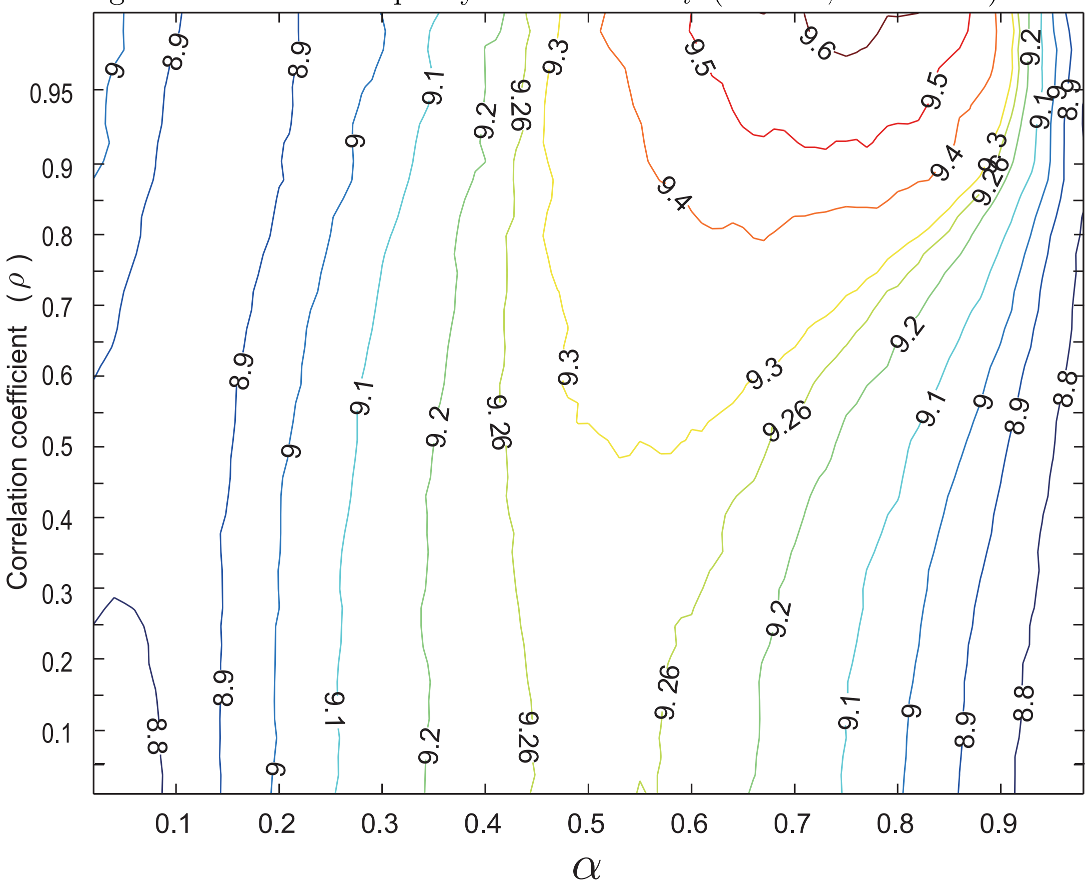

Figure 4.8: Lower-bound of channel capacity for different ρ and α. (SNR=8dB, $N_t = 20$) Scale of ρ from 0.9 to 1 is magnified by two times compared to other areas.

Chapter 5

Conclusion

In this chapter, we summarize important points and findings obtained by the discussions so far.

5.1 Optimal Training Sequence Design for the Kalman Filter

We showed that an enhancement of channel capacity exploiting spatiotemporal correlation characteristics can be realized under channel environments whose temporal evolution can be known and expressed as the Gauss-Markov model and also has spatiotemporal correlation with known parameters.

Also, we confirmed that the estimation assuming the Kronecker spatial correlation model is capable of enhancing estimation accuracy even when the actual environment does not obey the Kronecker model, i.e. the channel has LOS components so it has strong correlation dependency between the Tx and the Rx.

We discussed whole channel estimation from partial channel measurement in MIMO antenna selection system. The numerical simulation revealed that if we directly utilized the estimated channel for MIMO transmission, a high temporal correlation coefficient ($\rho > 0.98$) is essential in order to receive the benefit of MIMO antenna selection system compared to the case without antenna selection system.

This indicates that channel measurements for all the antenna elements are preferable rather than the partial measurements under absence of sufficient spatiotemporal correlation such that the channel roughly remains static.

5.2 Estimation of Channel Statistics by EM-based Algorithm

We investigated the estimation of time-varying channel statistics by means of the EM algorithm. The investigation consists of simplification of the channel model and parameter estimation by the EM algorithm. We confirmed that enhancement of the estimation capability can be achieved by proposed methods. The proposed parameter estimation method enhances estimation capability when smaller forgetting factors are employed.

We also applied this estimation scheme to the problem of whole channel estimation from partial channel measurement at MIMO antenna selection systems. The result of numerical simulation revealed that the estimation works well when spatiotemporal correlation of the actual channel is large enough. The estimation capability is almost same as the case of using a prior statistics.

Even if a prior statistics are available, achievement of accurate estimation by the partial measurement requires a high temporal correlation. The proposed parameter estimation scheme cannot surpass the estimation capability of the case of using a prior statistics.

The simulation also revealed that the EM based parameter estimation scheme works effectively for partial measurement problem in antenna selection system. The reason of this is considered that the information of observation matrix is utilized for estimation of the channel statistics. In the EM based parameter estimation, the channels for measured antenna elements are treated as highly reliable information, thus it is utilized for a large update of existing channel statistic. In contrast, the channels for unmeasured antenna elements are treated as unreliable, thus the information is only utilized for a small update of existing statistic. Such mechanism as so far described is considered to perform effectively under the scenario involved with antenna selection system in the sense that it has varying observation matrices.

The reason for estimation enhancement under small forgetting factors can be explained as an effect of expectation of unobserved realizations performed in the EM algorithm.

Presumably, treating the forgetting factor as a latent statistical variable may be one of the breakthroughs for further improvement of the estimation capability. A similar idea is already available as a variable forgetting factor [63]. Adaptive optimization of the forgetting factor in the scheme of the EM algorithm based parameter estimation may lead to an acceptable result.

5.3 A Robust Training Resource Allocation Exploiting Only Temporal Correlation

Focusing on the problem of power resource allocation of training sequences, we discussed a robust channel estimation method which is insusceptible to a change of the channel's temporal evolution model and spatial correlation statistic. The proposed method enhances channel capacity by controlling the channel estimation accuracy for each antenna element according to the degree of the channel's temporal correlation. The method is able to enhance channel capacity by a few percent, but its required hardware and computations for antenna selection is almost the same as the case of conventional antenna selection algorithms.

From a channel estimation technique point of view, the proposed method conducts only a simple ML estimation without exploiting any correlation statistics. The temporal correlation statistic is only utilized for determination of observation matrix via control of training sequences. This leads to an idea that it is possible to incorporate the aid of correlation information into the channel estimation method itself in the proposed scheme.

The situation enabling easy computation of our proposed antenna selection criterion which deals with uneven estimation error is realized because the simple ML estimation method is employed for channel estimation. If we utilized a spatial correlation aided channel estimation method, it might be difficult to conduct the proposed antenna selection method with lower computational cost. Furthermore, it is concerned that by employing correlation based estimation enhancement, the channel estimation error among antenna elements becomes small which might result in inefficiency of the proposed method.

The idea of the proposed method is based on the optimal control of power resource allocation of observation system based on correlation information. On this occasion, we realize such concept by exploiting an observation model structure of antenna selection system. However, such approach is not always possible in general observation model of wireless communication systems.

5.4 Important Factors for Correlation Based Enhancement of Channel Estimation

Let us consider important factors for correlation based enhancement of channel estimation by reflecting the essential points of the discussions so far. Effective and realistic methods from the viewpoint of implementation are as follows:

- The Kalman based estimation methods can exploit a spatiotemporal correlation.
- The proposed random walk based Gauss-Markov model is useful for channel modeling. The model can reduce necessary temporal correlation parameters without losing estimation capability to a large extent.
- If computational resources are acceptable, the optimal training sequence design based on correlation statistics is effective.
- The proposed power resource allocation method considering temporal correlation offers robust enhancement of channel capacity but it is dedicated to antenna selection systems.
- For antenna selection systems, partial measurement scheme is usable only where spatiotemporal correlation is high enough. Otherwise, at least all the elements are recommended to be measured even if some elements are coarsely estimated by small observation power.

5.5 Future Works

The following points are perspectives for continuing this study.

5.5.1 Improvement of Estimation Method of Statistical Model Parameters

Regarding estimation of statistical parameters, factors that dominantly influence the estimation capability has not been analyzed sufficiently. For instance, the behavior of tradeoff between tracking capability and estimation accuracy where the forgetting factor becomes large is not fully analyzed yet. Detailed analysis is required in order to find the bottleneck of estimation enhancement. Also, an idea for improvement of EM iteration by using future observations expressed in (3.55) should be discussed.

We should also consider the possibility that the limit of estimation capability of the linear estimation methods based on Gaussian modeling and least square criterion has already been reached. Also, an incorporation of DFE into our estimation scheme is considered to be worth exploring.

5.5.2 Channel Prediction for Channel Aging Problem in Massive MIMO Systems

Recently, in the field of future cellular network, a novel multiuser MIMO system called massive MIMO [66–68] has been got a lot of attention. The massive MIMO system employs hundreds of antennas at the base station serving tens of uses, and it has merits of simplified multiuser processing and reduced transmit power.

Meanwhile, its performance is sensitive to channel aging effects which is progressive over time. In order to overcome the channel aging effects, channel prediction exploiting spatiotemporal correlation statistics has been studied [69]. So far, such study assumes a priori channel statistics, and system investigation under realistic environments is not enough. Contribution to solve such problem based on our study is considered.

5.5.3 Exploitation of Sparsity of the Channel

There is another approach for estimation enhancement exploiting statistical behavior of the channel. The approach is to change the statistical model of the channel.

Recently, modeling of natural signals by means of a sparse statistical distribution is attracting much attention.

It has been experimentally and theoretically proven that the behavior of neuron activity in the brain, and coefficients of wavelet transform of natural images and sounds obey the sparse distribution. Also, it has been studied that a signal having sparsity can be estimated by Bayesian estimation assuming a sparse distribution. Generally, imposing strong sparsity results in an increase of computational cost, but introducing appropriate mitigation such as Laplace distribution enables the optimization problem to be easily computed by the L1-norm optimization by means of the linear programming. Such discussions are made in the field of compressed sensing, sparse coding, sparse regularization, and independent component analysis, etc.

Under the situation that correlation information such as the second order moment is known in advance, once the signal is known to obey the sparse distribution, more accurate estimation can be achieved since the sparse distribution has less entropy than the Gauss distribution. It suggests that if we manage to find a suitable model of the signal to be estimated, more fine results might be obtained.

In the field of wireless communications, there have been many studies incorporating this concept into channel estimation problems. One of them takes note of the characteristics that the propagation channel's impulse response function obeys the sparse distribution, and extends the RLS algorithm with the aid of sparse channel estimation, which is known as the sparse RLS [64, 65].

For sparse signals, it is possible to apply a novel sampling theorem called the compressed sensing [70–73]. The theorem states that the sparse signals can be accurately reconstructed from an even much smaller number of sampling observations (by means of particular observations such as random projections) compared to the conventional Nyquist sampling theorem. For example, many applications have been studied such as CT, MRI image reconstructions [77], and a single-pixel digital camera [75–77].

Originally, the idea of the compressed sensing comes from in the field of image processing. Until now, active introductions into the field of wireless communication have been made [78, 79] since sparse characteristics are often found in the wireless channel scenarios. One of the examples is the decoding problem of UWB signals. Since UWB signals have wide bandwidth, a full observation of them requires many samplings. However, by employing the compressed measurement, it is possible to reduce the number of samplings for signal reconstruction. Thus we can reduce hardware cost for samplings. [80, 81].

Many studies have emerged that incorporate the sparse estimation into the issue of MIMO channel estimation in order to improve estimation capability with smaller observation resources [82–84]. We expect that such novel attempts can be a breakthrough for solving the problem.

Appendix A

Estimation of Channel Statistics in Batch Process

We require N_s sequences of full channel matrix in order to estimate the parameters in the Kalman filter. Estimates $\hat{\rho}$ and $\widehat{\boldsymbol{Q}}$ are given as:

$$\hat{\rho} = \frac{1}{N_{\mathrm{Tx}} N_{\mathrm{Rx}} (N_s - 1)} \sum_{i=1}^{N_{\mathrm{Tx}}} \sum_{j=1}^{N_{\mathrm{Rx}}} \sum_{k=1}^{N_s - 1} \frac{[\boldsymbol{H}_{k+1}]_{i,j} [\boldsymbol{H}_k^*]_{i,j}}{[\boldsymbol{H}_k]_{i,j} [\boldsymbol{H}_k^*]_{i,j}} \tag{A.1}$$

$$\begin{aligned}
2(1-\rho)\widehat{\boldsymbol{Q}} &= 2(1-\rho)\mathrm{E}_k \ \mathrm{vec}\, \boldsymbol{X}_k (\mathrm{vec}\, \boldsymbol{X}_k)^{\mathcal{H}} \\
&= (1-\rho)^2 \mathrm{E}_k \ \mathrm{vec}\, \boldsymbol{H}_k (\mathrm{vec}\, \boldsymbol{H}_k)^{\mathcal{H}} + (1-\rho^2)\mathrm{E}_k \ \mathrm{vec}\, \boldsymbol{X}_k (\mathrm{vec}\, \boldsymbol{X}_k)^{\mathcal{H}} \\
&\qquad \left(\because \ \mathrm{E}_k \ \mathrm{vec}\, \boldsymbol{H}_k (\mathrm{vec}\, \boldsymbol{H}_k)^{\mathcal{H}} = \mathrm{E}_k \ \mathrm{vec}\, \boldsymbol{X}_k (\mathrm{vec}\, \boldsymbol{X}_k)^{\mathcal{H}}\right) \\
&= \mathrm{E}_k \ \mathrm{vec}\,(\boldsymbol{H}_k - \boldsymbol{H}_{k+1}) \left[\mathrm{vec}\,(\boldsymbol{H}_k - \boldsymbol{H}_{k+1})\right]^{\mathcal{H}} \\
&\qquad \left(\because \quad \mathrm{E}_k \ \mathrm{vec}\, \boldsymbol{H}_k \,(\mathrm{vec}\, \boldsymbol{X}_k)^{\mathcal{H}} = \boldsymbol{0}\right)
\end{aligned} \tag{A.2}$$

In addition, if channel obeys (2.10), spatial correlation matrices of the Kronecker model $\widehat{\boldsymbol{R}}_{\mathrm{Tx}}$ and $\widehat{\boldsymbol{R}}_{\mathrm{Rx}}$ can be estimated as follows:

$$\widetilde{\boldsymbol{R}}_{\mathrm{Tx}} \triangleq \sum_{k=1}^{N_s-1} (\boldsymbol{H}_k - \boldsymbol{H}_{k+1})^{\top} (\boldsymbol{H}_k - \boldsymbol{H}_{k+1})^{*} \tag{A.3}$$

$$\widetilde{\boldsymbol{R}}_{\mathrm{Rx}} \triangleq \sum_{k=1}^{N_s-1} (\boldsymbol{H}_k - \boldsymbol{H}_{k+1}) (\boldsymbol{H}_k - \boldsymbol{H}_{k+1})^{\mathcal{H}} \tag{A.4}$$

$$\widehat{\boldsymbol{R}}_{\mathrm{Tx}} = N_{\mathrm{Tx}} \widetilde{\boldsymbol{R}}_{\mathrm{Tx}} / (\mathrm{tr}\, \widetilde{\boldsymbol{R}}_{\mathrm{Tx}}) \tag{A.5}$$

$$\widehat{\boldsymbol{R}}_{\mathrm{Rx}} = N_{\mathrm{Rx}} \widetilde{\boldsymbol{R}}_{\mathrm{Rx}} / (\mathrm{tr}\, \widetilde{\boldsymbol{R}}_{\mathrm{Rx}}) \tag{A.6}$$

Appendix B

Proof of Equation for Filter Calculations

Let $\{\boldsymbol{\varphi}_i\}_{i=1}^M$ be any subset of arbitrary orthonormal basis of $\mathbb{C}^N$ $(N \geq M)$, and $\{\boldsymbol{X}_i\}_{i=1}^M$ be a set of any matrices of the same size, following relationship holds:

$$\left(\sum_{i=1}^{M} \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^{\mathcal{H}} \otimes \boldsymbol{X}_i\right)^{\dagger} = \sum_{i=1}^{M} \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^{\mathcal{H}} \otimes \boldsymbol{X}_i^{\dagger} \tag{B.1}$$

(Proof) Let $\{\boldsymbol{A}_i\}_{i=1}^M$ and $\{\boldsymbol{B}_i\}_{i=1}^M$ be sets of any matrices of the conforming size, then we have:

$$\left(\sum_{i=1}^{M} \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^{\mathcal{H}} \otimes \boldsymbol{A}_i\right) \left(\sum_{i=1}^{M} \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^{\mathcal{H}} \otimes \boldsymbol{B}_i\right) = \sum_{i=1}^{M} \boldsymbol{\varphi}_i \boldsymbol{\varphi}_i^{\mathcal{H}} \otimes \boldsymbol{A}_i \boldsymbol{B}_i \tag{B.2}$$

From the above relationship, we can confirm that the right side of (B.1) satisfies the four conditions of the Moore-Penrose generalized inverse, i.e. $\boldsymbol{X}\boldsymbol{X}^{\dagger}\boldsymbol{X} = \boldsymbol{X}$, $\boldsymbol{X}^{\dagger}\boldsymbol{X}\boldsymbol{X}^{\dagger} = \boldsymbol{X}^{\dagger}$, $\boldsymbol{X}\boldsymbol{X}^{\dagger} = (\boldsymbol{X}\boldsymbol{X}^{\dagger})^{\mathcal{H}}$, and $\boldsymbol{X}^{\dagger}\boldsymbol{X} = (\boldsymbol{X}^{\dagger}\boldsymbol{X})^{\mathcal{H}}$.

Appendix C

Optimal Training Sequence Structure for Kalman Filter

Let singular value decomposition of $\boldsymbol{S}_k^\top$ be $\boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Psi}^{\mathcal{H}}$ where $\boldsymbol{\Phi}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Sigma}$ are $N_t \times N_t$, $N_{\mathrm{Tx}} \times N_{\mathrm{Tx}}$ unitary matrices and $N_t \times N_{\mathrm{Tx}}$ diagonal matrix, respectively. Substituting it into (2.24) yields,

$$J_k[\boldsymbol{S}_k] = \mathrm{tr}\left(\boldsymbol{P}_{k|k-1}^{-1} + \frac{1}{\sigma_n^2}\boldsymbol{\Psi}\boldsymbol{\Sigma}^{\mathcal{H}}\boldsymbol{\Sigma}\boldsymbol{\Psi}^{\mathcal{H}} \otimes \boldsymbol{I}_{N_{\mathrm{Rx}}}\right)^{-1}. \tag{C.1}$$

This implies that $\boldsymbol{\Phi}$ can be chosen arbitraily. For any $N_{\mathrm{Rx}} \times N_{\mathrm{Rx}}$ unitary matrix $\boldsymbol{V}$, J_k can be transformed into:

$$J_k = \mathrm{tr}\left[(\boldsymbol{\Psi} \otimes \boldsymbol{V})^{\mathcal{H}}\boldsymbol{P}_{k|k-1}^{-1}(\boldsymbol{\Psi} \otimes \boldsymbol{V}) + \frac{1}{\sigma_n^2}\boldsymbol{\Sigma}^{\mathcal{H}}\boldsymbol{\Sigma} \otimes \boldsymbol{I}_{N_{\mathrm{Rx}}}\right]^{-1} \tag{C.2}$$

It is proven that J_k is minimized if inside of $[\,\cdot\,]$ in (C.2) becomes diagonal form ([26] : Theorem 1). If $\boldsymbol{P}_{k|k-1}$ have the form of (2.28), diagonalization of (C.2) is achieved if and only if $\boldsymbol{\Psi} = [\boldsymbol{u}_1\,\boldsymbol{u}_2\,\cdots\,\boldsymbol{u}_{N_{\mathrm{Tx}}}]$ and $\boldsymbol{V} = [\boldsymbol{v}_1\,\boldsymbol{v}_2\,\cdots\,\boldsymbol{v}_{N_{\mathrm{Rx}}}]$. Denoting the i-th column vector of $\boldsymbol{\Phi}$ as $\boldsymbol{\varphi}_i$, and $\boldsymbol{\Sigma}^{\mathcal{H}}\boldsymbol{\Sigma} = \mathrm{diag}\,[\alpha_1\,\alpha_2\,\cdots\,\alpha_{N_{\mathrm{Tx}}}]$ yield (2.27).

Appendix D

ML Estimation of (3.19)

Let $\{\hat{z}_i\}_{i=1}^N$ be available estimated realization of latent channel evolution where its indexes indicate time instant. Based on our definition of likelihood function, in order to find covariance parameter at time instant k, we want to maximize

$$
\begin{aligned}
L(\hat{\boldsymbol{z}}_1, \hat{\boldsymbol{z}}_2, \cdots, \hat{\boldsymbol{z}}_N \,|\, \boldsymbol{R}_k) &= \log \prod_{i=1}^{k} \mathcal{CN}\left(\hat{\boldsymbol{z}}_i \,|\, \boldsymbol{0}, \boldsymbol{R}_k\right)^{\lambda^{k-i}} \\
&= \left(\sum_{i=1}^{k} \lambda^{k-i}\right) \left[-\log \pi^{N_{\mathrm{Tx}} N_{\mathrm{Rx}}} + \log\det \boldsymbol{R}_k^{-1}\right] - \sum_{i=1}^{k} \left\langle \boldsymbol{R}_k^{-1}, \lambda^{k-i} \hat{\boldsymbol{z}}_i \hat{\boldsymbol{z}}_i^{\mathcal{H}} \right\rangle_{\mathrm{HS}}
\end{aligned} \tag{D.1}
$$

where $< \cdot, \cdot >_{\mathrm{HS}}$ is the Hilbert-Schmidt inner product defined as $< \boldsymbol{A}, \boldsymbol{B} >_{\mathrm{HS}} \triangleq \operatorname{tr} \boldsymbol{A}\boldsymbol{B}^{\mathcal{H}}$. To find the extremal value, differentiating with respect to $\boldsymbol{R}_k^{-1}$ yields,

$$
\begin{aligned}
\tau \boldsymbol{R}_k &= \sum_{i=1}^{k-1} \lambda^{k-i} \hat{\boldsymbol{z}}_i \hat{\boldsymbol{z}}_i^{\mathcal{H}} \qquad \left(\tau \triangleq \sum_{i=1}^{k} \lambda^{k-i} = \sum_{i=0}^{k-1} \lambda^i\right) \\
&= \lambda \sum_{i=1}^{k-1} \lambda^{k-i-1} \hat{\boldsymbol{z}}_i \hat{\boldsymbol{z}}_i^{\mathcal{H}} + \hat{\boldsymbol{z}}_k \hat{\boldsymbol{z}}_k^{\mathcal{H}} \\
&= \lambda \left(\sum_{i=0}^{k-2} \lambda^i\right) \boldsymbol{R}_{k-1} + \hat{\boldsymbol{z}}_k \hat{\boldsymbol{z}}_k^{\mathcal{H}} \\
&= \lambda(\tau - \lambda^{k-1}) \boldsymbol{R}_{k-1} + \hat{\boldsymbol{z}}_k \hat{\boldsymbol{z}}_k^{\mathcal{H}}
\end{aligned} \tag{D.2}
$$

If k is large enough to approximate $\lambda^{k-1} \approx 0$, substituting it and multiplying $1/\tau = (1-\lambda)/(1-\lambda^k) \approx 1 - \lambda$ on both side yields (3.20).

Appendix E

Posterior Probability of Latent Channel Evolution

Note that the notation of variables in this section is different from other sections for simplicity. Let us consider the following linear observation model where the statistical variables $\boldsymbol{x}$ and $\boldsymbol{y}$ are observed through matrix $\boldsymbol{A}$.

$$\boldsymbol{z} = \boldsymbol{A}(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{n} \tag{E.1}$$

We assume that $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{n}$ are independent, and their prior distributions are available as follows:

$$\begin{aligned} \boldsymbol{x} &\sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{R}) \\ \boldsymbol{y} &\sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{S}) \\ \boldsymbol{n} &\sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{Q}) \end{aligned} \tag{E.2}$$

$$P(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{n}) = \mathcal{CN}(\boldsymbol{x} \,|\, \boldsymbol{0}, \boldsymbol{R})\, \mathcal{CN}(\boldsymbol{y} \,|\, \boldsymbol{0}, \boldsymbol{S})\, \mathcal{CN}(\boldsymbol{n} \,|\, \boldsymbol{0}, \boldsymbol{Q}) \tag{E.3}$$

After having observed $\boldsymbol{z}$, a posterior distribution of $\boldsymbol{x}$ is obtained by marginalization of the simultaneous distribution with respect to $\boldsymbol{y}$ as follows:

$$P(\boldsymbol{x} \,|\, \boldsymbol{z}) = \int P\left(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} - \boldsymbol{A}(\boldsymbol{x} + \boldsymbol{y})\right) \mathrm{d}\boldsymbol{y} \tag{E.4}$$

Now, the above integrand can be decomposed as

$$\begin{aligned} P\left(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} - \boldsymbol{A}(\boldsymbol{x} + \boldsymbol{y})\right) &= C \times P(\boldsymbol{y} \,|\, \boldsymbol{x}, \boldsymbol{z}) P(\boldsymbol{x} \,|\, \boldsymbol{z}) \\ &= C \times \mathcal{CN}(\boldsymbol{y} \,|\, \boldsymbol{\mu}_y, \boldsymbol{\Sigma}_y)\, \mathcal{CN}(\boldsymbol{x} \,|\, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \end{aligned} \tag{E.5}$$

where C denotes constant, and

$$\boldsymbol{\mu}_y = \boldsymbol{\Sigma}_y \boldsymbol{A}^{\mathcal{H}} \boldsymbol{Q}^{-1} (\boldsymbol{z} - \boldsymbol{A}\boldsymbol{x}) \tag{E.6}$$

$$\boldsymbol{\Sigma}_y = \left(\boldsymbol{S}^{-1} + \boldsymbol{A}^{\mathcal{H}} \boldsymbol{Q}^{-1} \boldsymbol{A}\right)^{-1} \tag{E.7}$$

$$\boldsymbol{\mu}_x = \boldsymbol{\Sigma}_x \left[\boldsymbol{I} - \boldsymbol{A}^{\mathcal{H}} \boldsymbol{Q}^{-1} \boldsymbol{A} \boldsymbol{\Sigma}_y\right] \boldsymbol{A}^{\mathcal{H}} \boldsymbol{Q}^{-1} \boldsymbol{z} \tag{E.8}$$

$$\boldsymbol{\Sigma}_x = \left(\boldsymbol{R}^{-1} + \left[\boldsymbol{I} - \boldsymbol{A}^{\mathcal{H}} \boldsymbol{Q}^{-1} \boldsymbol{A} \boldsymbol{\Sigma}_y\right] \boldsymbol{A}^{\mathcal{H}} \boldsymbol{Q}^{-1} \boldsymbol{A}\right)^{-1} \tag{E.9}$$

From (E.4) and (E.5), we obtain the following result.

$$P(\boldsymbol{x} \,|\, \boldsymbol{z}) = \mathcal{CN}(\boldsymbol{x} \,|\, \boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x) \tag{E.10}$$

Appendix F

ML Estimation of the Parametric Covariance Matrix

Let $r_{i,j}^{(k,t)}$ be the eigenvalue corresponding to the eigenvector $\boldsymbol{\phi}_j \otimes \boldsymbol{\psi}_i$ of $\boldsymbol{R}_k^{(t)}$.

$$\begin{aligned} r_{m,n}^{(k,t+1)} &= \arg\max_{r_{m,n}} Q\left(\sum_{j=1}^{N_{\mathrm{Tx}}} \sum_{i=1}^{N_{\mathrm{Rx}}} r_{i,j} \boldsymbol{\phi}_j \boldsymbol{\phi}_j^{\mathcal{H}} \otimes \boldsymbol{\psi}_i \boldsymbol{\psi}_i^{\mathcal{H}} \mid \boldsymbol{R}_k^{(t)} \right) \\ &= \arg\max_{r_{m,n}} \tau \left(-N_{\mathrm{Tx}} N_{\mathrm{Rx}} \log \pi + \sum_{i=1,j=1}^{N_{\mathrm{Tx}},N_{\mathrm{Rx}}} \log \frac{1}{r_{i,j}} \right) \\ &\qquad\qquad - \sum_{i,j,p} \frac{\lambda^{k-p}}{r_{i,j}} \left[\left(\gamma_{i,j}^{(p)} \right)^2 \left\| < \boldsymbol{\phi}_j \otimes \boldsymbol{\psi}_i, \boldsymbol{B}_p^{\mathcal{H}} \delta_p > \right\|^2 + \eta_{i,j}^{(p)} \right] \end{aligned} \tag{F.1}$$

where τ is the same definition as one used in (D.2). To find the extremal value, differentiating with respect to $1/r_{m,n}$ yields,

$$\tau r_{m,n} = \sum_{p=1}^{k-1} \lambda^{k-p} \left[\left(\gamma_{m,n}^{(p)} \right)^2 \left\| < \boldsymbol{\phi}_n \otimes \boldsymbol{\psi}_m, \boldsymbol{B}_p^{\mathcal{H}} \delta_p > \right\|^2 + \eta_{m,n}^{(p)} \right] \tag{F.2}$$

Similarly as in the Appendix D, if k is large enough to approximate $\lambda^{k-1} \approx 0$, this yields (3.46).

Appendix G

List of Publications

Journal Papers

1. Yousuke Naruse and Jun-ichi Takada, "Iterative Channel Estimation in MIMO Antenna Selection Systems for Correlated Gauss-Markov Channel," *IEICE Trans.Commun.*, vol.E92B, no.3, pp.922-932, Mar.2009.

2. Yousuke Naruse and Jun-ichi Takada, "EM-based Recursive Estimation of Spatio-temporal Correlation Statistics for Non-stationary MIMO Channel," to be submitted to *IEICE Trans.Commun.*

International Conference

1. Yousuke Naruse and Jun-ichi Takada, "Efficient Channel Estimation Method for MIMO Antenna Selection Systems Exploiting Temporal Correlation of Channel," *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, pp.1-5, 6-9 May 2012.

Domestic Conference

1. Yousuke Naruse and Jun-ichi Takada, "Channel estimation method for MIMO antenna selection system exploiting temporal correlation of the channel,"(In Japanese) *Technical Report of IEICE*, pp.269-274, 27 Feb.2008.

Bibliography

[1] G.J. Foschini and M.J Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311-355, 1998.

[2] I.E. Telatar, "Capacity of multi-antenna Gaussian channels," *AT&T Bell Lab. Internal Tech. Memo.*, June. 1995. (*European Trans. Telecom.*, vol. 10, pp. 585-595, Nov. 1999.)

[3] S. H. Ting, K. Sakaguchi and K. Araki, "Performance Analysis of MIMO Eigenmode Transmission System under Realistic Channel and System Conditions," *IEICE Trans. Commun.*, vol.E87-B, no.8, pp.2222-2232, Oct. 2004.

[4] K. J. Kim and R. A. Iltis, "Joint detection and channel estimation algorithms for QS-CDMA signals over time-varying channels," *IEEE Trans. on Commun.*, vol. 50, no. 5, pp. 845-855, May. 2002.

[5] M. Gharavi-Alkhansari and A.B. Gershman, "Fast antenna subset selection in MIMO systems," *IEEE Trans. Signal Processing*, vol. 52, no. 2, pp. 339-347, Feb. 2004.

[6] S. Sandhu, R. Nabar, D. A. Gore, and A. J. Paulraj, "Near-optimal selection of transmit antennas for a MIMO channel based on shannon capacity," *34th Asilomar Conferenceon Signals, Systems and Computers*, vol. 1, pp. 567-571, Nov. 2000.

[7] A. Gorokhov, D. A. Gore, and A. J. Paulraj, "Receive Antenna Selection for MIMO Spatial Multiplexing: Theory and Algorithms," *IEEE Trans. Signal Processing*, vol. 51, no. 11, pp. 2796-2807, Nov. 2003.

[8] H. Zhang and H. Dai, "Fast transmit antenna selection algorithms for MIMO systems with fading correlation," *Vehicular Technology Conference*, 2004. VTC2004-Fall.

[9] Inaki Berenguer, Xiaodong Wang, Vikram Krishnamurthy, "Adaptive MIMO antenna selection via discrete stochastic optimization," *IEEE Trans. Signal Processing*, vol. 53, no. 11, Nov. 2005.

[10] W.C. Jakes, *Microwave Mobile Communications*, New York: Wiley, 1974.

[11] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[12] Medard, M, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Information Theory*, vol.46, no.3, pp933-946, May 2000.

[13] M. Sugiyama, and H. Ogawa, "Pseudo orthogonal bases give the optimal generalization capability in neural network learning," in Proceedings of SPIE, *Wavelet Applications in Signal and Image Processing VII*, vol.3813, pp.526-537, Jul. 1999.

[14] A. F. Molisch, M. Z. Win, and J. H. Winters, "Capacity of MIMO systems with antenna selection," in *Proc. IEEE Int. Conf. Commun.*, pp. 570-574, Jun. 2001.

[15] A.F. Molisch, "MIMO systems with antenna selection - an overview," in *Proc. IEEE Radio and Wireless Conf. (RAWCON '03)*, pp. 167-170, 10-13 Aug. 2003.

[16] Jens Baltersee, Gunnar Fock, and Heinrich Meyr, "An Information Theoretic Foundation of Synchronized Detection," *IEEE Transactions on Communications*, vol. 49, no. 12, pp. 2115-2123, Dec. 2001.

[17] S. Haykin, "*Adaptive Filter Theory, 4th Edition*," Prentice hall information and system sciences series, Sep. 2001.

[18] C. Komninakis, C. Fragouli, A. H. Sayed and R. D. Wesel, "Multi-Input Multi-Output Fading Channel Tracking and Equalization Using Kalman Estimation," *IEEE Trans. Signal Processing*, vol. 50, no. 5, May. 2002.

[19] D. Schafhuber, G. Matz, F. Hlawatsch, "Kalman Tracking of Time-Varying Channels in Wireless MIMO-OFDM Systems," invited paper in *Proc. 37th Asilomar Conf. Signals, Systems, Computers, Pacific Grove (CA)*, pp. 1261-1265, Nov. 2003.

[20] M. Enescu, M. Herdin, T. Roman and V. Koivunen, "Parameter estimation of measured channels in mobile MIMO OFDM system," in *Proc. of the Fourth IEEE Int. Symp. on*, pp.123-126, 18-21 Dec. 2004.

[21] M. Enescu, T. Roman and V. Koivunen, "Channel estimation and tracking in spatially correlated MIMO OFDM systems," *Workshop on Stat. Signal Processing, 2003 IEEE*, pp.347-350, 28 Sept. - 1 Oct. 2003.

[22] Y. Naruse and J. Takada "Iterative Channel Estimation in MIMO Antenna Selection Systems for Correlated Gauss-Markov Channel," *IEICE Trans. Commun.*, vol.E92B, no.3, pp.922-932, Mar.2009.

[23] M.K. Ozdemir, H. Arslan and E. Arvas "MIMO-OFDM Channel Estimation for Correlated Fading Channels," *IEEE Int. Conf. Commun. 2005 (ICC 2005)*, vol.4, pp.2626-2630, 16-20 May 2005.

[24] H. Miao and M.J.Juntti, "Space-time MMSE channel estimation for MIMO-OFDM system with spatial correlation," *Vehicular Technology Conference (VTC Spring), 2004 IEEE 59th*, vol. 3, pp.1806-1810, 17-19 May 2004.

[25] J. Choi and Y. Lee, "Complexity-Reduced Channel Estimation in Spatially Correlated MIMO-OFDM Systems," *IEICE Trans. Commun.*, vol.E90B, no.9, pp.2609-2612, Sep.2007.

[26] J.H. Kotecha and J.H.Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Trans. Signal Processing*, vol. 52, pp. 546-557, Feb. 2004.

[27] E. Bjornson and B. Ottersten, "A Framework for Training-Based Estimation in Arbitrarily Correlated Rician MIMO Channels with Rician Disturbance," *IEEE Trans. Signal Processing*, vol. 58, no. 3, pp.1807-1820, Mar. 2010.

[28] H. Nooralizadeh and S.S. Moghaddam, "Performance Improvement in Estimation of Spatially Correlated Rician Fading MIMO Channels Using a New LMMSE Estimator," *Int. J. Communications, Network and System Sciences*, vol. 3, pp.962-971, Dec. 2010.

[29] M. Kiessling, J. Speidel and Y. Chen, "MIMO Channel Estimation in Correlated Fading Environments," *Vehicular Technology Conference (VTC Fall), 2003 IEEE 58th*, vol.2, pp.1187-1191, 6-9 Oct. 2003.

[30] O. Weikert and U. Zolzer, "Efficient MIMO Channel Estimation With Optimal Training Sequences," in *Proc. of 1st Workshop on Commercial MIMO-Components and -Systems (CMCS 2007)*, Duisburg, Germany, 13-14 Sept. 2007.

[31] M. A. Saeed, N. K. Noordin, B. M. Ali, S. Khatun and M. Ismail, "RLS Channel Estimation and Tracking for MIMO-extended IEEE 802.11a WLANs," *Int. J. Computer Science and Network Security (IJCSNS)*, vol.8, no.2, Feb. 2008.

[32] D. Schafhuber, G. Matz and F. Hlawatsch, "Adaptive Wiener Filters For Time-Varying Channel Estimation In Wireless OFDM Systems," in *Proc. IEEE ICASSP-03*, vol. IV, pp. 688-691, Hong Kong, Apr. 2003.

[33] Y.C. Chen and Y.T. Su, "MIMO channel estimation in spatially correlated environments," *Personal, Indoor and Mobile Radio Commun. (PIMRC), 2004 IEEE 15th*, vol.1, pp.498-502, 5-8 Sept. 2004.

[34] X. Hou, S. Li, C. Yin and G. Yue, "Two-dimensional recursive least square adaptive channel estimation for OFDM systems," in *Proc. Int. Conf. Wireless Commun., Networking and Mobile Computing*, vol.1, pp.232-236, 23-26 Sept. 2005.

[35] M.D. Nisar, W. Utschick and T. Hindelang, "Maximally Robust 2-D Channel Estimation for OFDM Systems," *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 3163-3172, June 2010.

[36] Z. Luo and D. Huang, "Optimal and Robust MMSE Channel Estimation for MIMO-OFDM Systems," *Personal, Indoor and Mobile Radio Commun. (PIMRC), 2008 IEEE 19th*, pp.1-5, 15-18 Sept. 2008.

[37] Y. Li, L.J. Cimini and N.R. Sollenberger, "Robust channel estimation for OFDM systems with rapid dispersive fading channels," *IEEE Trans. Commun.*, vol. 46, no. 7, pp.902-915, Jul. 1998.

[38] Y. Naruse and J. Takada "Efficient Channel Estimation Method for MIMO Antenna Selection Systems Exploiting Temporal Correlation of Channel," *Vehicular Technology Conference (VTC Spring), 2012 IEEE 75th*, pp.1-5, 6-9 May 2012.

[39] F. Hlawatsch and G. Matz "Fundamentals of Time-Varying Communication Channels," in *Wireless Communications over Rapidly Time-Varying Channels*, Chapter 1, pp.1-63, Amsterdam, The Netherlands: Academic Press, 2011.

[40] D. Chizhik, J. Ling, P. W. Wolniansky, R. A. Valenzuela, N. Costa and K. Huber "Multiple-input multiple-output measurements and modeling in Manhattan," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 321-331, Apr. 2003.

[41] J.P. Kermoal, L. Schumacher, K.I. Pedersen, P.E. Mogensen and F. Frederiksen, "A Stochastic MIMO Radio Channel Model With Experimental Validation," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1211-1226, Aug. 2002.

[42] S. H. Ting, K. Sakaguchi and K. Araki, "A Markov-Kronecker model for analysis of closed-loop MIMO systems," *IEEE Commun. Letters*, vol. 10, no. 8, pp. 617-619, Aug. 2006.

[43] K. I. Pedersen, J. B. Andersen, J. P. Kermoal and P. Mogensen, "A stochastic multiple-input multiple-output radio channel model for evaluation of space-time coding algorithms," *Vehicular Technology Conference, 2000. IEEE-VTS Fall VTC 2000. 52nd*, vol.2, pp.893-897, 2000.

[44] Y.C. Chen and Y.T. Su, "MIMO Channel Estimation in Correlated Fading Environments," *IEEE Trans. Wireless Commun.*, vol.9, no.3, pp.1108-1119, Mar. 2010.

[45] J. Salo, G. Del Galdo, J. Salmi, P. Kyosti, M. Milojevic, D. Laselva, and C. Schneider, "MATLAB implementation of the 3GPP Spatial Channel Model (3GPP TR 25.996) [Online]," Available: http://www.tkk.fi/Units/Radio/scm/ (2005, Jan.)

[46] D. S. Baum, J. Salo, M. Milojevic, P. Kyosti and J. Hansen, "MATLAB implementation of the interim channel model for beyond-3G systems (SCME)," *[Online].* May 2005.

[47] M. Biguesh and A.B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Processing*, vol.54, no.3, pp.884-893, Mar. 2006.

[48] T. L. Marzetta, "BLAST training: Estimation channel characteristics for high-capacity space-time wireless," in *Proc. 37th Annu. Allerton Conf. Commun., Comput., Control*, Sep. 1999.

[49] Q. Sun, D. C. Cox, A. Lozano and H. C. Huang, "Training-based channel estimation for continuous flat fading BLAST," in *Proc. Int. Conf. Commun., New York*, Apr. 2002.

[50] T. Y. Al-Naffouri, O. Awoniyi, O. Oteri and A. Paulraj, "Receiver Design for MIMO-OFDM Transmission over Time Variant Channels," in *Proc. IEEE GLOBECOM*, vol. 4, pp. 2487-2492, Nov. 2004.

[51] J.M.M. Ocloo and F. Alberge, "OFDM channel estimation by a linear EM-MAP algorithm," *IEEE Trans. Signal Processing*, vol. 4, pp. 109-112, May 2006.

[52] Tareq Y. Al-Naffouri, "An EM-Based Forward-Backward Kalman Filter for the Estimation of Time-Variant Channels in OFDM," *IEEE Trans. Signal Processing*, vol. 1, no. 11, Nov. 2006.

[53] K. Muraoka, K. Fukawa, H. Suzuki and S. Suyama, "Joint signal detection and channel estimation using differential models via EM algorithm for OFDM mobile communications," *IEICE Trans. Commun.*, vol. E94-B, no. 2, pp. 533-545, Feb. 2011.

[54] C. Komninakis, C. Fragouli, A. H. Sayed and R.D. Wesel, "Multi-Input Multi-Output Fading Channel Tracking and Equalization Using Kalman Estimation," *IEEE Trans. Signal Processing*, vol.50, no.5, pp.1065-1076, May 2002.

[55] Y.H. Chan and X. Yu, "A reduced-rank MMSE-DFE receiver for space-time coded DS-CDMA systems," *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th,* vol.5, pp.3659-3663, 26-29 Sept. 2004.

[56] N. Al-Dhahir and A.H. Sayed, "The finite-length multi-input multi-output MMSE-DFE," *IEEE Trans. Signal Processing*, vol.48, no.10, pp.2921-2936, Oct. 2000.

[57] F.H. Chiu, S.H. Wu and C.C.J. Kuo, "Joint MIMO Channel Tracking and Symbol Detection with EM Algorithm and Soft Decoding," in *Proc. IEEE GLOBECOM*, vol.4, pp.2404-2408, 2 Dec. 2005.

[58] S. Geman, E. Bienenstock and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Computation*, vol. 4, no. 1, pp.1-58, 1992.

[59] A. S. Y. Poon, R. W. Brodersen, and D. N. C. Tse, "Degrees of freedom in multiple-antenna channels: a signal space approach," *IEEE Trans. Information Theory*, vol. 51, no. 2, pp.523-536, Feb. 2005.

[60] A. S. Y. Poon, D. N. C. Tse, and R. W. Brodersen, "Impact of scattering on the capacity, diversity, and propagation range of multiple-antenna channels," *IEEE Trans. Information Theory*, vol. 52, no. 3, pp.1087-1100, Mar. 2006.

[61] T. Inoue, and Y. Karasawa, "Theoretical Analysis on the Performance of Optimal Combining for Multipath Waves Distributed in Spatial and Time Domains," *Technical report of IEICE* (in Japanese), CS98-25/RCS98-25, May 1998.

[62] P. Almers, F. Tufvesson, and A.F. Molisch, "Keyhole Effect in MIMO Wireless Channels: Measurements and Theory," *Wireless Commun., IEEE Trans. on*, vol.5, no.12, pp.3596-3604, Dec. 2006.

[63] C. Paleologu, J. Benesty, and S. Ciochina, "A Robust Variable Forgetting Factor Recursive Least-Squares Algorithm for System Identification," *IEEE Signal Processing Letters*, vol.15, pp.597-600, 2008.

[64] B. Babadi, N. Kalouptsidis, and V. Tarokh, "SPARLS: The Sparse RLS Algorithm," *IEEE Trans. Signal Processing*, vol.58, no.8, pp.4013-4025, Aug. 2010.

[65] B. Dumitrescu, A. Onose, P. Helin, and I. Tabus, "Greedy Sparse RLS," *IEEE Trans. Signal Processing*, vol.60, no.5, pp.2194,2207, May 2012.

[66] R. W. Heath, Jr., "What is the Role of MIMO in Future Cellular Networks: Massive? Coordinated? mmWave?," *IEEE CTS COMSOC/SP and I&M Joint DLT Meeting*, Austin, TX, October 21, 2013.

[67] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp.3590-3600, Nov. 2010.

[68] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Mag.*, vol. 30, no. 1, pp. 40-60, Jan. 2013.

[69] K. T. Truong and R. W. Heath, Jr., "Effects of Channel Aging in Massive MIMO Systems," *IEEE/KICS Journal of Communications and Networks, Special Issue on Massive MIMO*, vol. 15, no. 4, pp. 338-351, Aug. 2013.

[70] D.L. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol.52, no.4, pp.1289-1306, Apr. 2006.

[71] E.J. Candes, and M.B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, vol.25, no.2, pp.21-30, Mar. 2008.

[72] E.J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, vol.52, no.2, pp.489-509, Feb. 2006.

[73] J.A. Tropp, and A.C. Gilbert, "Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit," *IEEE Trans. Information Theory*, vol.53, no.12, pp.4655-4666, Dec. 2007.

[74] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk, "A New Compressive Imaging Camera Architecture using Optical-Domain Compression," *Proc. IS&T/SPIE Computational Imaging IV*, Jan. 2006.

[75] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83-91, 2008.

[76] Gang Huang, Hong Jiang, Kim Matthews, and Paul A. Wilford, "Lensless Imaging by Compressive Sensing," *DBLP, http://dblp.uni-trier.de [Online]*, Available: http://arxiv.org/abs/1305.7181 , 2013.

[77] M. Lustig, D.L. Donoho, J.M. Santos, and J.M. Pauly, "Compressed Sensing MRI," *IEEE Signal Processing Magazine*, vol.25, no.2, pp.72-82, Mar. 2008.

[78] Chun-Yang Chen, and P.P. Vaidyanathan, "Compressed sensing in MIMO radar," *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*, pp.41-44, 26-29 Oct. 2008.

[79] T. Zhi, and G.B. Giannakis, "Compressed Sensing for Wideband Cognitive Radios," *IEEE Trans. Acoust. Speech Signal Process. (ICASSP 2007)*, vol.4, pp.IV1357-IV1360, 15-20 Apr. 2007.

[80] J.L. Paredes, G.R. Arce, and Z.M. Wang, "Ultra-wideband compressed sensing: channel estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol.1, no.3, pp.383-395, Oct. 2007.

[81] N.T. Dung, N.V. Sinh, N.T. Hoa, and N.T. Hieu, "Application of compressive sensing in time hopping multi-user UWB system," *2011 International Conference on IEEE Advanced Technologies for Communications (ATC). Danang, Vietnam*, pp.248-251, 2-4 Aug. 2011.

[82] W.U. Bajwa, J. Haupt, A.M. Sayeed, and R. Nowak, "Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels," *Proc. of the IEEE*, vol.98, no.6, pp.1058-1076, Jun. 2010.

[83] G. Taubock, and F. Hlawatsch, "A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots," *IEEE Trans. Acoust. Speech Signal Process. (ICASSP 2007)*, pp.2885-2888, Mar. 31 2008-Apr. 4 2008.

[84] C.R. Berger, Zhaohui Wang, Jianzhong Huang, and Shengli Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Communications Magazine*, vol.48, no.11, pp.164-174, Nov. 2010.