2013-04-09

Improved Channel Probing for Secret Key Generation with Multiple Antenna Systems

Britton T. Quist

Brigham Young University - Provo

Follow this and additional works at: https://scholarsarchive.byu.edu/etd

Part of the Electrical and Computer Engineering Commons

BYU ScholarsArchive Citation

Quist, Britton T., "Improved Channel Probing for Secret Key Generation with Multiple Antenna Systems" (2013). All Theses and Dissertations. 3554.
https://scholarsarchive.byu.edu/etd/3554

This Dissertation is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Improved Channel Probing for Secret Key Generation
with Multiple Antenna Systems

Britton Quist

A dissertation submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Michael A. Jensen, Chair
  Brian D. Jeffs
  David G. Long
  Michael D. Rice
  Karl F. Warnick

Department of Electrical and Computer Engineering
Brigham Young University
March 2013

Copyright © 2013 Britton Quist
All Rights Reserved
ABSTRACT

Improved Channel Probing for Secret Key Generation with Multiple Antenna Systems

Britton Quist
Department of Electrical and Computer Engineering
Doctor of Philosophy

Establishing secret keys from the commonly-observed randomness of reciprocal wireless propagation channels has recently received considerable attention. In this work we propose improved strategies for channel estimation between MIMO or beamforming systems for secret key generation. The amount of mutual information that can be extracted from the channel matrix estimates is determined by the quality of channel matrix estimates. By allocating increased energy to channel estimation for higher gain beamforming combinations at the expense of low-gain combinations, key establishment performance can be increased. Formalizing the notion of preferential energy allocation to the most efficient excitations is the central theme of this dissertation. For probing with beamforming systems, we formulate a theoretically optimal probing strategy that upper bounds the number of key bits that can be generated from reciprocal channel observations. Specifically, we demonstrate that the eigenvectors of the channel spatial covariance matrix should be used as beamformer weights during channel estimation and we optimize the energy allocated to channel estimation for each beamformer weight under a total energy constraint. The optimal probing strategy is not directly implementable in practice, and therefore we propose two different modifications to the optimal algorithm based on a Kronecker approximation to the spatial covariance matrix. Though these approximations are suboptimal, they each perform well relative to the upper bound. To explore how effective an array is at extracting all of the information available in the propagation environment connecting two nodes, we apply the optimal beamformer probing strategy to a vector current basis function expansion on the array volume. We prove that the resulting key rate is a key rate spatial bound that upper bounds the key rate achievable by any set of antenna arrays probing the channel with the same total energy constraint. For MIMO systems we assume the channel is separable with a Kronecker model, and then for that model we propose an improved probing strategy that iteratively optimizes the energy allocation for each node using concave maximization. The performance of this iterative approach is better than that achieved using the traditional probing strategy in many realistic probing scenarios.

Keywords: cryptography, covariance matrices, security, array signal processing, MIMO, beamforming
ACKNOWLEDGMENTS

I would like to acknowledge and thank the NDSEG fellowship program for their generous support during my graduate studies. I would like to thank Brigham Young University and Church of Jesus Christ of Latter Day Saints for providing a tremendous opportunity to pursue a PhD in an environment where faith and academic pursuits each enhance the other. I would like to thank Ryan Smith of IMSAR for the flexibility of letting me work part time and on my terms through my PhD. I would like to thank and acknowledge my parents for their belief in me and their tremendous support of my educational pursuits throughout my entire life. I would like to thank my adviser in this research, Dr. Michael Jensen for his invitation to pursue a PhD at BYU and then for his tireless effort to always provide time, ideas, and expertise during this pursuit. He is a great friend and a great adviser. Finally I would like to thank my wife Kirsten for her love and her unconditional support of my professional pursuits. While she believes her support is ordinary, it truly is extraordinary.
Table of Contents

List of Figures viii

1 Introduction 1

1.1 Dissertation Organization .............................................. 4

2 Overview of Cryptography and Reciprocal Channel Probing 6

2.1 Symmetric Encryption ..................................................... 7

2.2 Key Distribution .......................................................... 8

2.2.1 Asymmetric Encryption .............................................. 9

2.2.2 Key Generation From Common Randomness ......................... 12

2.3 Mutual Information between Channel Estimates ...................... 12

2.4 Related Research .......................................................... 15

3 Optimal Channel Estimation in Beamformed Systems 18

3.1 Channel Estimation ....................................................... 19

3.1.1 Achievable Key Rate .............................................. 19

3.1.2 Optimal Basis .......................................................... 20

3.2 Energy Allocation .......................................................... 23

3.2.1 Concavity .............................................................. 23

3.2.2 Lagrange Multiplier Optimization .................................. 25

3.2.3 Implementation ....................................................... 30
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.4.2 Radiated Energy Constraint</td>
<td>74</td>
</tr>
<tr>
<td>5.4.3 OSEA Application</td>
<td>75</td>
</tr>
<tr>
<td>5.5 Finite Basis Function Expansion</td>
<td>76</td>
</tr>
<tr>
<td>5.6 Results</td>
<td>77</td>
</tr>
<tr>
<td>5.6.1 Constrained Energy vs. Current</td>
<td>78</td>
</tr>
<tr>
<td>5.6.2 Pulse Basis Functions</td>
<td>78</td>
</tr>
<tr>
<td>5.6.3 Simulated Convergence</td>
<td>80</td>
</tr>
<tr>
<td>5.7 Chapter Summary</td>
<td>83</td>
</tr>
<tr>
<td>6 Improved Channel Estimation Resource Allocation for MIMO Systems</td>
<td>85</td>
</tr>
<tr>
<td>6.1 Channel Estimation</td>
<td>85</td>
</tr>
<tr>
<td>6.1.1 System Model</td>
<td>85</td>
</tr>
<tr>
<td>6.1.2 Key Rate</td>
<td>87</td>
</tr>
<tr>
<td>6.1.3 Beamformer Matrices</td>
<td>88</td>
</tr>
<tr>
<td>6.1.4 Equal Energy Allocation</td>
<td>91</td>
</tr>
<tr>
<td>6.2 Iterative Optimization for Energy Allocation</td>
<td>91</td>
</tr>
<tr>
<td>6.3 Results</td>
<td>94</td>
</tr>
<tr>
<td>6.3.1 Array Size</td>
<td>95</td>
</tr>
<tr>
<td>6.3.2 Energy per Key Bit</td>
<td>95</td>
</tr>
<tr>
<td>6.3.3 Number of PAS Clusters</td>
<td>96</td>
</tr>
<tr>
<td>6.3.4 Line of Sight Propagation</td>
<td>98</td>
</tr>
<tr>
<td>6.4 Chapter Summary</td>
<td>101</td>
</tr>
<tr>
<td>7 Conclusion</td>
<td>103</td>
</tr>
<tr>
<td>7.1 Future Work</td>
<td>105</td>
</tr>
</tbody>
</table>
List of Figures

2.1 Symmetric encryption where the common session key is used for both encryption and decryption ............................................. 8

2.2 Symmetric encryption where the common session key is used for both encryption and decryption ............................................. 9

2.3 A diagram of the Diffie Hellman key exchange ............................................. 11

3.1 Normalized first and second partials of $I_k$ as a function of the normalized mode energy $\rho_i = p_i/\sigma_0^2$ for several values of $\Lambda_{ii}$............................................. 24

3.2 The functions $A_r(\rho)$, $A_s(\rho)$ and $C(\rho)$. The shaded portions represent integrals that show the impact of diverting energy from the higher-order mode $A_r(\rho)$ to the lower-order mode $A_s(\rho)$ in its convex interval. ............................................. 27

3.3 Top plot: Comparison of energy allocated using OSEA, waterfilling, and equal allocation. Bottom plot: Difference between mutual information achieved with OSEA, waterfilling, or allocating all energy to a single mode and that achieved using equal energy allocation ($I_{k,E}$). ............................................. 32

3.4 The mutual information achieved using OSEA compared to that achieved using equal energy allocation as a function of the number of antennas in Alice’s array. ............................................. 34

3.5 The ratio of the mutual information achieved using OSEA to that achieved with equal energy allocation as a function of the normalized total available energy $P_T/\sigma_0^2$. ............................................. 35

3.6 Block diagram of the beamforming system used for channel estimation and establishment of secret encryption keys. ............................................. 35

3.7 Comparison between OSEA with the Kronecker approximation and the full covariance for the situation considered in Fig. 3.3. ............................................. 37
3.8 The mutual information achieved using OSEA compared to that achieved using equal energy allocation as a function of the number of antennas in Alice’s array when the simulation includes mutual coupling. 40

4.1 Key rate achieved using OSEA with the full covariance, OSEA using a Kronecker approximation to the covariance, and equal energy allocation as a function of the number of antennas in Alice’s array for several values of SNR with Rayleigh fading. 44

4.2 Key rate achieved using OSEA with the full covariance, OSEA using a Kronecker approximation to the covariance, and equal energy allocation as a function of the number of antennas in Alice’s array for several values of SNR with Rician fading. 45

4.3 Key rate achieved using OSEA with the full covariance, Kronecker approximation, and modified Kronecker approximation as a function of the number of antennas in Alice’s array for several values of SNR with Rician fading. 47

4.4 Key rate achieved using OSEA with the full covariance, Kronecker approximation, and modified Kronecker approximation as a function of the number of antennas in Alice’s array for several values of SNR with Rayleigh fading. 48

4.5 Key rate achieved using OSEA with the full covariance, Kronecker approximation, and modified Kronecker approximation as well as with equal energy allocation as a function of the number of multipath clusters for several values of SNR with Rician fading. 50

4.6 Key rate achieved using OSEA with the full covariance, Kronecker approximation, and modified Kronecker approximation as well as with equal energy allocation as a function of the number of multipath clusters for several values of SNR with Rayleigh fading. 51

4.7 Key rate achieved using OSEA with the full covariance, Kronecker approximation, and modified Kronecker approximation as well as with equal energy allocation as a function of the Rician K-factor for several values of SNR. 53

4.8 The number of active modes used in OSEA as a function of the Rician K-factor for several values of SNR. 54

4.9 Normalized difference between the number of bits enabled by OSEA and by the Kronecker approximation (top) or by the modified Kronecker approximation (bottom) as a function of the Rician K-factor. 55

4.10 Normalized difference between the number of bits enabled by OSEA and by the Kronecker approximation or the modified Kronecker approximation with Rayleigh propagation as a function of SNR. 56
5.1 The key rate spatial bound constructed when the energy or the current is constrained as a function of the aperture size.

5.2 Ratio of the key rate achieved using OSEA for an $M \times M$ array of square current pulses and the key rate spatial bound as a function of aperture size for an SNR of 5 dB.

5.3 The OSEA key rate of an $M \times M$ dipole array modeled by square current pulses relative to key rate spatial bound as a function of SNR.

5.4 The number of active modes obtained from the OSEA solution for an $M \times M$ array of pulse functions as well as for the key rate spatial bound as a function of SNR.

5.5 Normalized difference between the key rate $I_{k,Q_{\max}}$ achieved using $Q_1 = Q_2 = Q_{\max}$ and the key rate $I_{k,Q}$ obtained using $Q_1 = Q_2 = Q$.

5.6 Normalized difference between the key rate $I_{k,Q_{\max}}$ achieved using $Q_1 = Q_2 = Q_{\max}$ and the key rate $I_{k,Q}$ obtained when sweeping $Q_1$ or $Q_2$ (or both) at an SNR of 5 dB.

5.7 The number of basis functions satisfying the constraint $\Theta_{a,nn} > 1/Q_1$ when $Q_2 = Q_{\max}$.

5.8 The number of virtual elements satisfying the constraints $\Theta_{a,nn} > 1/Q_1$ and $\Lambda_{a,vv} > 1/Q_2$ thresholds.

6.1 System diagram showing Alice transmitting the vector $a_n x_{a,n}$ from her $N_a$ antennas and Bob applying the beamforming matrix $B$ to the received signal to produce $N_b$ channel coefficient estimates $\hat{w}_{b,mm}$ for $1 \leq m \leq N_b$.

6.2 Average normalized difference between the key rate at algorithm convergence and the key rate at the current iteration number as a function of the iteration number for 500 randomly-realized systems.

6.3 Average key rate as a function of $N_a$ when $N_b = 5$ for different values of the SNR $P_T/\sigma_0^2$.

6.4 Ratio of the key rate achieved with equal energy allocation ($I_{k,E}$) to that achieved using the iterative optimization for different values of $N_a$ when $N_b = 5$.

6.5 Energy per key bit realized as a function of the total energy allocated to probing (for fixed noise) for different values of $N_a$ when $N_b = 5$. 

x
6.6 Key rate as a function of the number of PAS clusters for two different SNR values when $N_a = N_b = 5$. .................................................. 99

6.7 Relative key rate as a function of the number of PAS clusters for two different SNR values when $N_a = N_b = 5$. .................................................. 100

6.8 The performance of the iterative optimization technique relative to equal allocation for Ricean and Rayleigh propagation when $N_a = N_b = 5$. ........... 101
Chapter 1

Introduction

The pervasive nature of wireless communication has motivated increased efforts to develop methods for preserving the confidentiality of sensitive data transmissions. Because wireless communication cannot prevent unwanted access to the transmitted data stream, secure protocols must encrypt data prior to transmission. While a variety of methods for encryption exist, it is most common to use a symmetric approach in which both the data encryption at the transmitter and decryption at the receiver use a commonly-known secret key [1] that is often established using the Diffie-Hellman exchange [2, 3]. While this exchange publicly discloses all information required to uniquely identify the key, it is considered to be computationally secure since the discrete logarithm required to identify the key from the observation is prohibitively costly for currently available hardware and algorithms. Naturally, the security of this key establishment technique will be compromised if efficient methods are discovered for efficiently computing such logarithms. Furthermore, the numbers used to generate the keys are frequently derived from deterministic pseudo-random sequences, introducing an additional security weakness [4].

One novel approach for key establishment between two nodes is key generation based on commonly observed randomness [5, 6]. One source of such randomness is the reciprocal nature of multipath electromagnetic propagation. In a wireless channel, the impulse response or channel response between two antennas is the same, independent of which antenna is transmitting. In a multipath environment, there will be several independent signal paths and each will have a unique complex gain. The channel response between two antenna positions is the coherent sum of the gain from these signal paths. Since the phase of each path will vary with antenna position, the exact channel measurement is highly dependent on the exact antenna placement. In fact, channel measurements usually decorrelate within
only a few wavelengths, preventing key estimation by a nearby eavesdropper. Additionally, the channel response is derived from the inherent randomness in nature which makes the corresponding keys equally non-deterministic with any existing model.

When two nodes estimate a channel, the key rate, computed as the mutual information between the channel estimates at the two nodes, represents the maximum number of key bits that can be generated from a single channel estimate. Measuring performance relative to this upper bound, initial work in this area focused on single-input single-output (SISO) communication systems [7], with the main emphasis placed on practical algorithms for generating key bits from realized channel estimates [8, 9, 10, 11]. More recent work on SISO systems explores techniques for implementing the channel estimation to increase the underlying key rate [12]. Additional recent research extends analysis of the key rate and development of practical key establishment techniques from available channel estimates to multiple-input multiple-output (MIMO) systems [13, 14, 15, 16]. Related work in optimal channel probing has also demonstrated techniques for maximizing the accuracy of the channel estimate for standard multiple-input multiple-output (MIMO) communication [17], but this differs from the idea of maximizing the mutual information between channel estimates observed at two distinct locations.

In maximizing the key rate between channel matrix estimates, we consider two legitimate nodes, Alice and Bob, each using multiple antennas to estimate the channel. When each node estimates the channel (probes) with a beamforming system, the antenna array scales the signal transmitted or received with each array element, resulting in SISO communication. This differs from MIMO systems which are capable of transmitting or receiving an independent signal from each array element with an arbitrary complex gain. When probing with different array element combinations, some configurations will be more effective than others at transferring energy between the two nodes. Intuitively, performance can be improved if more effective array configurations are preferentially allocated energy at the expense of the less productive configurations. The central objective of this dissertation is to formalize the preferential energy allocation for beamforming and MIMO systems, with the objective being maximization of the key rate.
When probing with a beamforming system, the vector of complex gains of all the array elements at a single point in time constitutes the beamformer weight. The probing scheme is defined by the set of beamformer weights used to estimate all the desired channel coefficients. A primary contribution of this research is the optimal sounding energy allocation (OSEA), which is an optimal probing scheme that upper bounds the key rate achievable with any set of beamformer weights. Unfortunately the OSEA scheme does not provide a method for mapping the optimal probing vectors to actual beamformer weights. To address this, two different suboptimal approximations to OSEA are developed that each achieve performance near the upper bound. Extensive simulations are provided to characterize the performance of OSEA and its practical adaptations relative to the traditional probing strategy.

The OSEA key rate provides an upper bound on the key rate that can be realized between two beamformer fed arrays in a given propagation environment. This leaves open the question of whether performance could be improved with a better choice of antenna arrays. To answer this question, we develop a spatial bound on key rate that upper bounds the performance achievable between two antenna arrays within constrained volumes. This bound is formulated by applying OSEA to an antenna array where each element is actually a vector current represented by Fourier basis functions. We prove that as the number of basis functions becomes large, the solution upper bounds the key rate achievable in the propagation environment by any other set of square integrable currents in the same volume satisfying a common energy constraint.

MIMO systems are capable of estimating the channel response from all receive channels simultaneously, which results in more energy efficient key bit generation. One byproduct of this parallel estimation is an asymmetric channel estimate signal-to-noise ratio (SNR) at the two nodes. When trying to optimize the energy allocation at each node, this SNR asymmetry creates an interdependency between the energy allocation at each node that prevents direct application of the technique developed for optimal beamformer probing. To circumvent this interdependence, this work proposes an iterative strategy where the energy allocation at one node is fixed while that at the other node is optimized using traditional concave maximization. An iterative procedure based on this principle allows specification of the asymmetric probing strategy that in many cases increases the key rate.
1.1 Dissertation Organization

To provide context for the ideas presented in this dissertation, Chapter 2 presents a high level view of current encryption and key exchange techniques as well as a more detailed overview of key generation from reciprocal wireless probing. This discussion includes a derivation of the key rate between two channel matrix estimates. This expression serves as a launch point from which the proposed probing strategies are developed.

The derivation of the optimal beamformer probing strategy is provided in Chapter 3. The first step in this development is a proof of the optimality of probing the channel with beamformer weightings derived from eigenvectors of the full spatial covariance matrix. This is followed by a proof of the optimality of the energy allocation strategy used in OSEA. Since this algorithm is cannot be directly implemented in practical systems, we next provide a physically realizable approximation – referred to as the Kronecker approximation – that achieves performance near the upper bound. This is followed by analysis of how the proposed probing strategy can be adapted to include the effects of mutual coupling.

In Chapter 4, the key rate achieved by OSEA and the Kronecker approximation are evaluated through simulation of a number of propagation environments. The results demonstrate that when a line of sight propagation path is present, the performance of the Kronecker approximation degrades. To address this shortcoming, we propose a modification to the Kronecker approximation that performs well with or without a line of sight signal path present.

The presentation of the beamformer key rate spatial bound is given in Chapter 5. The treatment first demonstrates that the OSEA key rate applied to an expansion of the current using Fourier basis functions converges as the number of basis functions becomes large. This is followed by a proof that this limiting value upper bounds what is achievable by any set of square integrable current distributions satisfying the same energy constraint. We then provide a generalization to OSEA that yields the optimal probing strategy when the energy constraint accounts for the mutual resistance of the array elements. Results then illustrate a number of aspects of the key rate spatial bound including performance achieved with the alternative energy constraint as well as the algorithm convergence as a function of the number of basis functions included in simulation.
Chapter 6 provides the iteratively-optimized MIMO channel probing strategy. This include a mathematical justification of the approach and a detailed discussion of the convergence of the algorithm. Results then explore the performance achieved with the algorithm as a function of SNR and propagation environment description. This is followed by conclusions and a discussion of future work in Chapter 7.
Chapter 2

Overview of Cryptography and Reciprocal Channel Probing

When considering the security of an encipherment technique, it is important to understand the level of protection desired. One notion of security, called unconditional security is that a malicious node with unlimited time and unlimited computational resources would be unable to decipher the concealed message. In his seminal work on security, Shannon [18] showed that when a message $M$, often called plain text, is encrypted into a cypher text $C$, the message is unconditionally secure if and only if

$$H(C|M) = H(C)$$  \hspace{1cm} (2.1)$$

where $H(\cdot)$ is the entropy function. This statement requires the mapping between the plain text and the cypher text to be as random as the cypher text itself. This requirement is very restrictive, and to date only the one-time pad [19, 20] has been shown to meet this criteria. The approach requires that the entropy of the key be greater than the entropy of the message and that each key be used only once. Since this requires both nodes to possess a common secret key with more information than the data stream they want to exchange, the approach has not been widely adopted in modern wireless communication.

If the requirements for unconditional security are overly taxing for a given system configuration, it is still possible to ensure the privacy of communication within realistic constraints. One standard by which security can be measured is whether or not the encryption technique prevents a malicious attacker with realistic computational resources from accessing the data for as long as the secrecy of the encrypted message is important. This form of security is called computational security [2]. Since this metric is defined based on existing technology and algorithms, developments in either can erode the security of an algorithm
previously considered secure [21]. This is in contrast to unconditional security that does not weaken with advancing capabilities. In spite of this important nuance, the widespread need for practical encryption strategies motivates this as the most useful method for evaluating the security of encryption algorithms.

2.1 Symmetric Encryption

For centuries, computational security has been primarily achieved through symmetric encryption [22], which is diagrammed in Fig. 2.0. In this diagram, Alice and Bob are the legitimate nodes trying to maintain private communication in the presence of a malicious eavesdropper called Eve. In symmetric encryption, Alice and Bob both possess a common session key. This key is used to encrypt the plain text into the cypher text and then to decrypt the cypher text back into the original plain text. With Eve having no access to the session key, she will be unable to convert the cypher text into the original message. When the keys are securely obtained, the greatest weakness of symmetric encryption come from linear analysis [23] or differential analysis [24, 25] where Eve attempts to glean information about a key by observing the relationship between similar input plain texts and the corresponding outputs. While such an attack can reduce the time it takes for a malicious node to illegitimately decipher a message, such vulnerabilities can often be overcome by simply increasing the key length. This is because most attacks in practice only reduce the search space required to decrypt the message from a brute force attack [26]. With the remaining search space still quite large, widespread algorithms such as AES [27] are considered computationally secure as long as key lengths are sufficient.

Presuming the encryption method provides a satisfactory level of security after accounting for any known attacks, the primary vulnerability of symmetric encryption arises from the generation and distribution of keys. If the keys are formed from deterministic pseudo-random sequences than this exposes a security vulnerability [4]. So while promising sources for digital randomness have been demonstrated, the widespread usage of pseudo-random sequences continues to be a vulnerability in many communications systems.
Figure 2.1: Symmetric encryption where the common session key is used for both encryption and decryption

2.2 Key Distribution

The fundamental dilemma in key exchanges is that secure key distribution is needed to establish secure communication. One possible way to ensure this security is to predistribute keys between nodes that might need to communicate so that a secure key is available whenever communication is required. This prearranged code can be part of a code book allowing each node to securely communicate with all other nodes. A primary difficulty of systems using prearranged keys between all node pairs is that the need for two nodes to securely communicate must be known and accommodated well in advance. While this assumption is reasonable for smaller networks, this requirement is prohibitive in many situations.

If prearranging communication between every possible set of nodes is prohibitive, one alternative approach is to set up a network where every node has a unique key that can securely communicate with special key distribution centers [1]. When two nodes desire to securely communicate, then each node securely obtains the session key from a key distribution center that is aware of the nodes’ need to communicate. The downside of this approach is that a trusted third party must be present to organize communication between any two nodes. This requirement is a severe limitation in most wireless communication configurations.
2.2.1 Asymmetric Encryption

The primary shortcoming of the previous algorithms is that they did not allow two nodes to securely identify a session key without the aid of prearranged secure symmetric encryption. To overcome this drawback, one widely adopted protocol for key agreement is asymmetric encryption shown in Figure 2.1. In this approach, the encryption is performed with a public key which is openly communicated to all nodes. Once data is encrypted into a cypher text, it can only be converted into plain text again with the use of the private key that is only known by the target recipient. The private key and the public key are naturally related, but by design the relationship between the two makes it very difficult to extract the private key from a known public key. Two popular approaches for this encryption are the Diffie Hellman key exchange [3] and RSA [28]. The security of these techniques is derived from the non reciprocal computational burden of discrete logarithms in the case of Diffie Hellman or prime factorization in the case of RSA. For discrete logarithms, if $y = \mod (\alpha^x, p)$, where $p$ and $\alpha$ are given, then computing $y$ from $x$ can be done in polynomial time. The reverse operation of computing $x$ from $y$ can only be solved in exponential time using existing hardware and algorithms.

Figure 2.2 shows a diagram of how the Diffie Hellman key exchange is performed. In the exchange, $\alpha$ and $p$ are known publicly and Alice and Bob respectively possess the
numbers $X_a$ and $X_b$ which should be random and secret. From these, Alice computes $Y_a = \text{mod} \left( \alpha^{X_a}, p \right)$, Bob computes $Y_b = \text{mod} \left( \alpha^{X_b}, p \right)$, and each transmits the computed value to the other node. The exponential complexity of a discrete logarithm prevents anyone including Alice or Bob from computing an unknown $X_a$ or $X_b$ from the transmitted information. Following transmission, Alice computes the key denoted $K$ as $K = \text{mod} \left( Y_a^{X_b}, p \right)$ and Bob similarly computes $K$ as $K = \text{mod} \left( Y_b^{X_a}, p \right)$. Since

$$\text{mod} \left( (np+z)^x, p \right) = \text{mod} \left( \text{mod} \left( np+z, p \right)^x, p \right) = \text{mod} \left( z^x, p \right) \quad (2.2)$$

for integers $n$, $z$, and $x$, the equivalence of the two keys can be shown from

$$K = \text{mod} \left( Y_a^{X_b}, p \right) \quad (2.3)$$

$$= \text{mod} \left( \text{mod} \left( \alpha^{X_a}, p \right)^{X_b}, p \right) \quad (2.4)$$

$$= \text{mod} \left( (\alpha^{X_a})^{X_b}, p \right) \quad (2.5)$$

$$= \text{mod} \left( (\alpha^{X_b})^{X_a}, p \right) \quad (2.6)$$

$$= \text{mod} \left( \text{mod} \left( \alpha^{X_b}, p \right)^{X_a}, p \right) \quad (2.7)$$

$$K = \text{mod} \left( Y_b^{X_a}, p \right). \quad (2.8)$$

With $p$, $\alpha$, $Y_a$, and $Y_b$ transmitted without encryption, Eve possesses all of the information required to uniquely identify $X_a$ and $X_b$. This makes the security of the exchange inseparable from the computational burden of the discrete logarithm. If an efficient method for computing a discrete logarithm is ever produced, the security of the exchange will be completely compromised. While currently not yet practical, early results in quantum computing suggest that the discrete logarithm problem is solvable in polynomial time [29]. While the details of RSA will not be discussed here, the work in [29] shows that the security achieved through the computational burden of prime factorization would also be marginalized with practical quantum computing. The possibility of this advancement makes finding an alternative to public key encryption for real time key exchanges an important topic of research.
For existing hardware and algorithms, asymmetric encryption suffers from drawbacks. This is because Eve possesses all of the information necessary to systematically explore the relationship between input plain texts and output cypher texts to gain information about the private key. This strategy, called a chosen text attack, is a major vulnerability of public key encryption techniques [30]. The possibility of Eve leveraging such an attack necessitates that asymmetric encryption techniques use much larger keys than symmetric encryption to achieve the same security. The work in [21] asserts that 1024-bit RSA encryption achieves the same secrecy as 80-bit symmetric encryption. With keys of that length, public key encryption techniques are much more computationally intense than comparable symmetric approaches. So while asymmetric encryption is an independent alternative to symmetric encryption, the computational complexity of the asymmetric encryption and decryption often motivates hybrid techniques where only a session key is exchanged through asymmetric encryption. A symmetric encryption technique then uses this session key to secure the subsequent data transfer.
2.2.2 Key Generation From Common Randomness

One technique for potentially overcoming the vulnerabilities associated with traditional key exchange is to establish the key based on commonly-observed randomness [5, 6], particularly when the common randomness is derived from a physically-observed phenomenon such as reciprocal multipath electromagnetic propagation. Reciprocal electromagnetic propagation [31] provides that the complex gain observed between two antennas with no active components is independent of which antenna is transmitting and which antenna is receiving. With proper calibration, this allows two nodes to observe the same channel response that can be quantized for key bit generation.

Keys formed from reciprocal channel response measurements can provide robust security because the channel observed between the two legitimate nodes is uncorrelated with that observed by a malicious eavesdropper only a few wavelengths away from either legitimate node [16]. With the nodes themselves forming an important component of the scattering environment, it is also impractical for an eavesdropper to accurately measure the channel response previously observed by the legitimate nodes once either has relocated to a new position. Furthermore this response will be determined by the interaction of the propagating waves with the entire scattering environment. The quantity and complexity of these interactions make producing an accurate estimate of the channel response well out of reach of any existing forward model. The complexity of these interactions also provides this key generation scheme an inherent randomness derived from the inherent randomness in nature.

2.3 Mutual Information between Channel Estimates

In wireless propagation, the actual channel response depends on the constructive and destructive superposition of a large number of impinging waves. The exact measurement of this response evolves over the course of time in a fading environment, with an ensemble of measurements characterized by some probability density function (pdf). When the receiving node collects a measurement, the resulting observation is a measure of the true channel response corrupted by noise. The characterization of the channel response as well as the probabilistic characterization of the estimation error at each node determines the key rate, or the maximum number of bits that can be extracted from channel estimates.
In this work, the channel responses are assumed to be represented by a circularly symmetric complex Gaussian distribution. If a node possesses several antennas, the channel measurements observed by the different receiving antennas will be correlated and thus characterized by a multivariate pdf. For this case, the response is assumed to be characterized by a complex jointly Gaussian pdf.

For a SISO channel, let \( w \sim \mathcal{CN}(0, \mathcal{E}^2) \) represent that the quantity is the complex channel gain between Alice and Bob, where \( \sim \mathcal{CN}(\mu, \sigma^2) \) denotes a complex Gaussian random variable with mean \( \mu \) and variance \( \sigma^2 \). The receiver noise is assumed to be complex additive white Gaussian noise (AWGN) with Alice’s error \( \eta_a \) characterized as \( \eta_a \sim \mathcal{CN}(0, \sigma_a^2) \) and Bob’s error similarly defined as \( \eta_b \sim \mathcal{CN}(0, \sigma_b^2) \). Letting Alice’s estimate of \( w \) be \( \hat{w}_a = w + \eta_a \) and Bob’s estimate defined similarly as \( \hat{w}_b = w + \eta_b \), the estimates can be jointly characterized as

\[
\begin{bmatrix}
\hat{w}_a \\
\hat{w}_b
\end{bmatrix} = \mathcal{CN}
\begin{bmatrix}
0 \\
0
\end{bmatrix}, K_{A,B}
\]

where

\[
K_{A,B} = \begin{bmatrix}
\mathcal{E}^2 + \sigma_a^2 & \mathcal{E}^2 \\
\mathcal{E}^2 & \mathcal{E}^2 + \sigma_b^2
\end{bmatrix}
\]

The key rate is given as

\[
I_k = I(\hat{w}_a; \hat{w}_b) = H(\hat{w}_a) + H(\hat{w}_b) - H(\hat{w}_a, \hat{w}_b)
\]

where \( I(\cdot;\cdot) \) is the mutual information of the two arguments. As the sum of two independent Gaussian random variables, \( \hat{w}_a \) is defined as \( \hat{w}_a \sim \mathcal{CN}(0, \mathcal{E}^2 + \sigma_a^2) \) with a corresponding entropy in bits given as

\[
H(\hat{w}_a) = 2 \log_2(\pi e) + \log_2(\mathcal{E}^2 + \sigma_a^2)
\]

The entropy of \( H(\hat{w}_a, \hat{w}_b) \) is similarly given as

\[
H(\hat{w}_a, \hat{w}_b) = 4 \log_2(\pi e) + \log_2(|K_{A,B}|)
\]
where $|\cdot|$ represents a determinant. Substituting (2.14) and the Alice and Bob forms of (2.13) into (2.12) and then canceling terms, the mutual information expression becomes

$$I_k = \log_2 \left( \mathcal{E}^2 + \sigma_a^2 \right) + \log_2 \left( \mathcal{E}^2 + \sigma_b^2 \right) - \log_2 \left( |K_{A,B}| \right)$$

$$= \log_2 \left( \frac{\left( \mathcal{E}^2 + \sigma_a^2 \right) \left( \mathcal{E}^2 + \sigma_b^2 \right)}{|K_{A,B}|} \right).$$

When multiple antennas are used with a MIMO or beamforming system, the channel response becomes a *channel response vector* $\mathbf{w}$ with Alice’s estimate given as $\hat{\mathbf{w}}_a = \mathbf{w} + \mathbf{\eta}_a$ where $\mathbf{\eta}_a$ is the a vector representing the estimation error. Bob’s estimate and estimation error are denoted similarly with $\hat{\mathbf{w}}_b = \mathbf{w} + \mathbf{\eta}_b$. The channel response correlation matrix is given as $\mathbf{W} = \mathbb{E}\{\mathbf{ww}^\dagger\}$ where $\{\cdot\}^\dagger$ denotes a conjugate transpose, and the full covariance of both channel estimates therefore becomes

$$K_{AB} = \mathbb{E} \left\{ \begin{bmatrix} \hat{\mathbf{w}}_a \\ \hat{\mathbf{w}}_b \end{bmatrix} \begin{bmatrix} \hat{\mathbf{w}}_a^\dagger \\ \hat{\mathbf{w}}_b^\dagger \end{bmatrix} \right\} = \begin{bmatrix} \hat{\mathbf{W}}_{aa} & \hat{\mathbf{W}}_{ab} \\ \hat{\mathbf{W}}_{ba} & \hat{\mathbf{W}}_{bb} \end{bmatrix}.$$

The mutual information between $\hat{\mathbf{w}}_a$ and $\hat{\mathbf{w}}_b$ is

$$I_k = \log_2 \frac{|\hat{\mathbf{W}}_{aa}| |\hat{\mathbf{W}}_{bb}|}{|K_{A,B}|}$$

which is simply the sum of the mutual information between independent combinations of channel response estimates. Applying the determinant identities

$$\left| \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \right| = |C_{11}| |C_{22} - C_{21}C_{11}^{-1}C_{12}|$$

and

$$|\mathbf{D}_1 + \mathbf{D}_2\mathbf{D}_3| = |\mathbf{D}_1| |\mathbf{I} + \mathbf{D}_3\mathbf{D}_1^{-1}\mathbf{D}_2|$$

leads to

$$|K_{A,B}| = |\hat{\mathbf{W}}_{aa}| |\hat{\mathbf{W}}_{bb}| |\mathbf{I} - \hat{\mathbf{W}}\hat{\mathbf{W}}_{aa}^{-1}\hat{\mathbf{W}}\hat{\mathbf{W}}_{bb}^{-1}|.$$
When (2.21) is substituted into (2.18), the resulting expression for mutual information is

\[ I_k = -\log_2 |I - W\hat{W}^{-1}W\hat{W}^{-1}|. \]  

(2.22)

### 2.4 Related Research

The expressions for mutual information in (2.22) upper bounds what is achievable with given channel realizations. When two nodes attempt to extract the maximum amount of mutual information from correlated random variables, a fundamental result observed by Stepian and Wolf in [32] is that feedback between the two nodes is required to achieve the key rate from (2.11). Specifically this work shows that the feedback from Alice to Bob must be greater than \( H(\hat{w}_a|\hat{w}_b) \) if the bounding key rate is to be achieved. This result is interesting from a theoretical perspective because it differs from what is seen in traditional communications where feedback does not change the theoretically achievable data rate.

While the idea of key generation from a wireless channel has been known for nearly twenty years [7], much of the work on this topic in the last decade focuses on understanding the expression for available mutual information and on practical approaches for improving the bit generation rate relative to the bound. In [33], Wilson et. al. consider the independent channel measurements that can be made with an ultrawideband probing system. These independent measurements arise from frequency dependent fading present in a wideband multipath channel. Leveraging the orthogonal frequency division multiplexing (OFDM) notion of orthogonal channels called frequency sub-carriers [34], Wilson’s work provides some bounds on the mutual information in a wideband channel and then provides some different approaches for feedback to ensure key agreement. In [9], Bloch et. al. provide a comprehensive key establishment procedure that offers a method for channel estimation, key generation and reconciliation, privacy amplification, and then message protection. The work in [8] suggests a more suitable low density parity check (LDPC) code that achieves better key agreement than previous coding methods.

One difficulty that a number of the works in key bit extraction do not address is the need for relatively short codes to ensure key agreement. Specifically, reciprocal probing may generate only 10 or 20 key bits per second which is much smaller than the typical code
lengths used to achieve near capacity performance. To address this issue, the authors in [10] propose a modified coding scheme that has a small code length and is designed to account for the errors associated with the non-reciprocal error sources in realistic channel probing configurations.

When considering channel probing for secret key generation using multiple antenna systems, the body of research is much smaller. Work by Wallace et. al. in [13] first discussed the information theoretic bounds for MIMO channel estimation and then proposed a feedback strategy that improved key agreement. The work in [16] extends these ideas with additional analysis of performance and improved insight into the impact of an eavesdropper on the secret key rate. In [15] Chen and Jensen provide a strategy for multiple antenna key generation in the presence of temporal and spatial correlation, where the concept exploits the fact that just as the channel response from closely spaced antenna elements are correlated, so too are responses observed by the same antenna over a short temporal window. With spatial and temporal correlation, any key extraction algorithm must decorrelate in space and time simultaneously in order to ensure that the observed key bits are uncorrelated.

While each of these prior efforts provides a valuable contribution to the general problem of reciprocal probing key establishment, they each address a different question than that investigated in this work. This is because each considers the channel estimate quality as predetermined and so the expression for key rate in (2.22) simply provides a maximum number of bits for the presumed channel response and estimation error covariances. The key rate expression in and of itself does not provide any insight into whether or not the resources available for estimating the channel and determining the corresponding covariance matrices have been intelligently allocated. To address this in a practical way, the work of Wei et. al. in [14] proposes a probing scheme that uses a control loop with feedback about the entropy in the channel to balance the tradeoff between the bit generation rate and the available resource consumption. The contribution of that work is important from a practical perspective, but it does not address the issue of what is theoretically possible with optimal resource allocation.

The maximum possible key rate, called the key capacity in [35], is investigated by Chou et. al. in [12]. This work considers optimal probing for an ultrawideband SISO system
with a fixed number of independent channels with the same SNR. For this case, this work
determines the amount of energy to allocate to probing each sub-carrier to maximize the key
rate or to minimize the alternative performance metric, the energy per key bit. The amount
of energy allocated to probe each sub-carrier is modulated by adjusting the fraction of the
time in which the channel response of each sub-carrier is probed with full energy.

The contributions of Chou parallel some of what is provided in the optimal beam-
former probing development presented here with the key simplifying assumption that inde-
pendent channels are of the same quality. Such an assumption cannot be made with the
spatial channels present in multiple antenna systems because the relative channel qualities
will have a large dynamic range. For this more general case, our work provides an optimal
beamformer probing strategy for maximizing the key rate that can be extracted from the
available spatial channels. Building on the beamforming case, this work also considers im-
proved probing with MIMO systems, although the asymmetry between the channel estimate
quality for each node makes a provably optimal energy allocation strategy out of reach in
the MIMO case.
Chapter 3

Optimal Channel Estimation in Beamformed Systems

The previous chapter provides a derivation of the key rate for a given channel correlation matrix and error model and suggest that the key rate could be maximized through an optimal allocation of the resources available for channel probing. In this chapter, we consider optimal channel estimation for multi-antenna systems where the participating nodes engage in SISO communication using array beamforming. Specifically, we formulate the multi-dimensional channel estimation procedure as a sequence of transmissions, each of which allows estimation of a unique coefficient representing one dimension of the multi-dimensional channel. We first demonstrate that to maximize the key rate, each transmission should be along an eigenvector of the spatial covariance of the channel matrix. Next, we develop a simple procedure for determining the energy that should be allocated to each transmission to maximize the key rate that accounts for the fact that the expression for the key rate is non-concave. The resulting approach for the first time allows computation of the absolute upper bound on the key rate. Because the optimization framework applies to an abstract system without consideration of the actual transmit and receive beamformers that must be applied during the channel estimation procedure, the resulting theoretical key rate is an upper bound on key rate rather than a key capacity. However, we demonstrate application of the technique to a real system by assuming that the channel spatial covariance is separable into transmit and receive contributions. This discussion on practical application also demonstrates how the technique can be applied to practical signal models, such as those that include antenna array mutual coupling. Computational results demonstrate both the nature of the solution as well as the performance benefit realized.
3.1 Channel Estimation

3.1.1 Achievable Key Rate

In the system under consideration, Alice and Bob are each equipped with an antenna array consisting of $N_a$ and $N_b$ antenna elements, respectively. In complex baseband notation, $H$ is the $N_b \times N_a$ matrix of narrowband channel transfer coefficients between the elements of the two arrays, and the $N \times 1$ vector $h$ represents $H$ stacked columnwise so that $N = N_a N_b$. In a traditional key establishment system, Alice and Bob in turn send training data to each other from which each estimates the channel coefficients $h$. For this case, $w = h$ from the treatment in Section 2.3.

More generally, Alice and Bob can each estimate a channel response vector $w$ whose elements represent unique linear combinations of the channel coefficients and use these values for key establishment. To formulate the problem mathematically, we consider an abstract representation of the channel estimation process. This abstraction makes no consideration for practical implementation of the required channel estimation, but it allows us to mathematically develop an upper bound on the key rate that in turn is used in Section 3.4.1 to formulate a practical implementation procedure.

Let $V$ represent an $N \times M$ matrix of channel probing vectors with $i$th column $v_i$, where $M \leq N$ is the rank of $V$. In our abstract system, we assume that Alice and Bob can each estimate the $M \times 1$ response vector $w$, with the estimate given by

$$\hat{w}_\xi = V^\dagger h + \eta_\xi = w + \eta_\xi$$  \hspace{1cm} (3.1)

where $\{\}^\dagger$ denotes a conjugate transpose, $w = V^\dagger h$ is the true channel response vector, $\xi \in \{a, b\}$ denotes Alice or Bob respectively, and $\eta_\xi$ is the estimation error whose elements are modeled as zero-mean circularly-symmetric complex Gaussian random variables. With this framework, the energy used to estimate the $i$th element of $w$ is given by $p_i = v_i^\dagger v_i$, and letting $\text{tr}(\cdot)$ represents the trace, the total energy used for channel estimation can be expressed as

$$P_T = \text{tr}(V^\dagger V) = \sum_{i=1}^{M} p_i.$$  \hspace{1cm} (3.2)
Alice and Bob will use their respective estimates of the channel response vector \( w \) to generate bits that will contribute to the established secret key, and we therefore must formulate the mutual information between these estimates. We assume that the errors on different channel response vector elements are uncorrelated so that \( \mathbb{E}\{ \eta_\zeta \eta_\zeta^\dagger \} = \sigma_\zeta^2 I \) where \( \sigma_\zeta^2 \) is the variance of the estimation error. We also assume that \( \eta_a \) and \( \eta_b \) are uncorrelated with each other and with \( w \). The covariances of the estimates are then

\[
\hat{W}_{\xi\xi} = \mathbb{E}\{ \hat{w}_\xi \hat{w}_\xi^\dagger \} = V^\dagger R V + \sigma_\zeta^2 I = W + \sigma_\zeta^2 I, \tag{3.3}
\]

\[
\hat{W}_{ab} = \mathbb{E}\{ \hat{w}_a \hat{w}_b^\dagger \} = V^\dagger R V = \hat{W}_{ba} = W \tag{3.4}
\]

where \( W = V^\dagger R V = \mathbb{E}\{ \underbar{w} \underbar{w}^\dagger \} \) and

\[
R = \mathbb{E}\{ \underbar{h} \underbar{h}^\dagger \}. \tag{3.5}
\]

Substituting the expressions for \( \hat{W}_{aa} \) and \( \hat{W}_{bb} \) into (2.22) yield the key rate expression

\[
I_k = -\log_2 \left| I - W (W + \sigma_a^2 I)^{-1} W (W + \sigma_b^2 I)^{-1} \right|. \tag{3.6}
\]

This mutual information represents the key rate, or the maximum number of key bits that can be established by Alice and Bob based on the observed channel estimates.

### 3.1.2 Optimal Basis

The key rate achieved in (3.6) depends on the choice of \( V \). While in conventional channel probing \( V \) would be a scaled orthonormal matrix, by properly selecting \( V \) under the constraint that the total energy expressed in (3.2) is limited, we can potentially increase the key rate. In this section, we demonstrate that maximization of the key rate is achieved if each vector \( v_i \) represents a scaled eigenvector of \( R \), while in Section 3.2 we develop a strategy for selecting the energy \( p_i = v_i^\dagger v_i \). Together, this allows computation of the achievable upper bound on the key rate under constrained total probing energy.

Since we bi-directionally probe the channel so that Alice and Bob can each generate the same key from the reciprocal channel responses, we assume that they probe the channel
using identical conditions so that $\sigma_a^2 = \sigma_b^2 = \sigma_0^2$. We can therefore rearrange (3.6) to the form

$$I_k = -\log_2 \left| I - \left( I + \sigma_0^2 W^{-1} \right)^{-2} \right|. \quad (3.7)$$

Suppose that we have an arbitrary covariance matrix $R$ and a probing scheme represented by $V$ that together uniquely define the channel response covariance matrix $W$. Simple manipulation of (3.7) reveals that the key rate depends on the eigenvalues of $W$. Therefore, suppose further that we identify an alternate probing scheme represented by a different matrix $\tilde{V}$ that results in a response covariance matrix $\tilde{W}$ such that $W$ and $\tilde{W}$ have identical eigenvalues so that the two probing schemes achieve the same key rate. If $\text{tr}(V^\dagger V) \leq \text{tr}(\tilde{V}^\dagger \tilde{V})$, then the original probing scheme achieves the key rate with a reduced total probing energy and can therefore be considered superior.

Let $R = U\Lambda U^\dagger$ represent the eigenvalue decomposition of $R$ where $\Lambda$ is the diagonal matrix of real, non-negative eigenvalues and $U$ is the unitary matrix of eigenvectors. By choosing $V = UP^{1/2}$ where $P^{1/2}$ is diagonal with $i$th diagonal element $\sqrt{\lambda_i}$, we obtain $W = \Lambda P$. Without loss of generality, we arrange $\Lambda$ and $W = \Lambda P$ so that the diagonal elements are in decreasing order. We also emphasize that the values of $P$ must be chosen so that $W = \Lambda P$ and $\tilde{W}$ have identical eigenvalues. Using the fact that $AB$ and $BA$ have the same eigenvalues [36, p. 51], we see that $R^{1/2} \tilde{V} \tilde{V}^\dagger R^{1/2}$ and $\tilde{W} = \tilde{V}^\dagger \tilde{R} \tilde{V}$ have identical eigenvalues, where since $R$ is Hermitian we can write $R = R^{1/2} R^{1/2}$ with $R^{1/2}$ formed using the Cholesky factorization. Since our problem statement forces $W$ and $\tilde{W}$ to have the same eigenvalues, $R^{1/2} \tilde{V} \tilde{V}^\dagger R^{1/2}$ is Hermitian with eigenvalue matrix $\Lambda P$. Therefore, there exists a unitary matrix $\Theta$ such that

$$R^{1/2} \tilde{V} \tilde{V}^\dagger R^{1/2} = \Theta^\dagger \Lambda P \Theta \quad (3.8)$$

or

$$\tilde{V} \tilde{V}^\dagger = R^{-1/2} \Theta \Lambda P \Theta^\dagger R^{-1/2}. \quad (3.9)$$
Therefore

\[
\text{tr}(\tilde{V}\tilde{V}^\dagger) = \sum_{i=1}^{N} \lambda_i(R^{-1/2}\Theta\Lambda P\Theta^\dagger R^{-1/2}),
\]

(3.10)

\[
\text{tr}(\tilde{V}^\dagger\tilde{V}) = \sum_{i=1}^{N} \lambda_i(R^{-1}\Theta\Lambda P\Theta^\dagger)
\]

(3.11)

where \(\lambda_i(\cdot)\) represents the \(i\)th eigenvalue of the matrix argument ordered in decreasing value and (3.11) is possible since 1) the arguments of \(\lambda_i(\cdot)\) in both (3.10) and (3.11) have the same eigenvalues [36, p. 51] and 2) commuting the two matrices on the left hand side is allowable under the trace.

Now, for any positive semidefinite matrices \(A\) and \(B\) [37]

\[
\sum_{i=1}^{N} \lambda_i(AB) \geq \sum_{i=1}^{N} \lambda_{N-i+1}(A)\lambda_i(B).
\]

(3.12)

Combining (3.11) and (3.12) results in

\[
\text{tr}(\tilde{V}^\dagger\tilde{V}) \geq \sum_{i=1}^{N} \lambda_{N-i+1}(R^{-1})\lambda_i(\Theta\Lambda P\Theta^\dagger).
\]

(3.13)

From (3.8), it follows that \(\lambda_i(\Theta\Lambda P\Theta^\dagger) = \Lambda_{ii}p_i\), where \(\Lambda_{ii}\) is the \(i\)th diagonal element of \(\Lambda\). Likewise, since \(\Lambda\) is arranged in order of decreasing values, \(\lambda_{N-i+1}(R^{-1}) = \Lambda_{ii}^{-1}\). These observations lead to

\[
\text{tr}(\tilde{V}^\dagger\tilde{V}) \geq \sum_{i=1}^{N} p_i = \text{tr}(P) = \text{tr}(V^\dagger V),
\]

(3.14)

proving our hypothesis that letting \(V = UP^{1/2}\) allows maximization of the key rate.

Given the optimality of this eigenvector basis, we can use that \(W = \Lambda P\) to express (3.6) as

\[
I_k = -\log_2 \left| I - (I + \sigma_0^2 P^{-1}\Lambda^{-1})^{-2} \right| \quad \text{(3.15)}
\]

\[
= \log_2 \left| \Lambda P\sigma_0^{-2} (2I + \sigma_0^2 P^{-1}\Lambda^{-1})^{-1} + I \right| \quad \text{(3.16)}
\]
which can be further reduced to

\[ I_k = \sum_{i=1}^{N} \log_2 \left[ 1 + \frac{p_i \Lambda_{ii} / \sigma_0^2}{2 + \sigma_0^2 / p_i \Lambda_{ii}} \right]. \tag{3.17} \]

We pause here to emphasize that since the basis consists of the eigenvectors of the covariance, in a practical scenario we must first estimate this covariance to enable optimal channel estimation. If the channel coefficients represent wide-sense stationary random variables, then the covariance can be estimated once and then used for all subsequent channel estimation exchanges.

### 3.2 Energy Allocation

While probing with the eigenvectors of \( \mathbf{R} \) enables maximization of the mutual information, (3.17) explicitly shows that we need to determine the optimal energy allocation variables \( p_i \) that maximize the key rate. Finding this optimal allocation requires an understanding of the first and second derivatives of \( I_k \) in (3.17) with respect to the energy allocations. Since (3.17) indicates that the key quantity is \( \rho_i = p_i / \sigma_0^2 \), we work with the derivatives

\[ A_i(\rho_i) = \ln 2 \frac{\partial I_k}{\partial \rho_i} = \frac{2 \Lambda_{ii}}{\Lambda_{ii} \rho_i + 1} - \frac{2 \Lambda_{ii}}{2 \Lambda_{ii} \rho_i + 1} \]

\[ = \frac{2 \Lambda_{ii}^2 \rho_i}{(\Lambda_{ii} \rho_i + 1)(2 \Lambda_{ii} \rho_i + 1)}, \tag{3.18} \]

\[ A_i'(\rho_i) = \ln 2 \frac{\partial^2 I_k}{\partial \rho_i^2} = \frac{4 \Lambda_{ii}^2}{(2 \Lambda_{ii} \rho_i + 1)^2} - \frac{2 \Lambda_{ii}^2}{(\Lambda_{ii} \rho_i + 1)^2} \]

\[ = \frac{-4 \Lambda_{ii}^4 \rho_i^2 + 2 \Lambda_{ii}^2}{(\Lambda_{ii} \rho_i + 1)^2(2 \Lambda_{ii} \rho_i + 1)^2}. \tag{3.20} \]

We note that \( \partial^2 I_k / \partial \rho_i \partial \rho_j = 0 \) for \( i \neq j \).

#### 3.2.1 Concavity

One standard approach for finding the global extremum of the mutual information in (3.17) subject to the constraint in (3.2) is to use a Lagrange multiplier solution. While
such a solution always produces a local extremum, it is guaranteed to be a global maximum only for a concave function over a convex solution set [38]. This observation motivates an examination of the concavity of our problem.

If a multivariate equation is concave then the Hessian matrix $\nabla^2 I_k$ is negative semidefinite. Since $\partial^2 I_k/\partial \rho_i \partial \rho_j = 0$ for $i \neq j$, the Hessian matrix is diagonal, and its eigenvalues are equal to the matrix diagonal elements. Therefore, the equation is concave if $\partial^2 I_k/\partial \rho_i^2 \leq 0$ for all $i$. The bottom plot in Fig. 3.0 plots (3.20) as a function of $\rho_i$ for three values of $\Lambda_{ii}$. These curves reveal that the second derivative is positive for small $\rho_i$ but monotonically decreases to zero with increasing $\rho_i$, reaching the value of zero at $\rho_i = \hat{\rho}_i = 1/\sqrt{2}\Lambda_{ii}$. On the concave interval $\rho_i > \hat{\rho}_i$, the expression is strictly negative, and therefore the mutual information is a concave function of $\rho_i$ on this interval. The closed interval $[0, \hat{\rho}_i]$ represents a convex complement to the concave interval, referred to as the convex interval. Note that each mode only has one concave interval and one convex interval.
3.2.2 Lagrange Multiplier Optimization

Since each mode has both a convex and concave interval, application of an optimization strategy requires particular care. However, to be able to explore the particular nuances associated with optimization of the mutual information considered in this work, we must first formulate our optimization strategy. Using a Lagrange multiplier formulation is a convenient approach given that we are trying to optimize the mutual information in (3.17) subject to the constraint in (3.2).

When formulating this solution, we first must decide how many modes should be used. We refer to a mode as active when energy is allocated to that mode \((p_i > 0)\) and use \(N_{\text{act}}\) to refer to the number of active modes. Examination of (3.17) reveals that if \(\Lambda_{ij} < \Lambda_{ii}\) but \(p_j > p_i\), then the mutual information could be increased simply by swapping the energy allocations to each mode. Therefore, if the eigenvalues of the \(N \times N\) covariance matrix are ordered such that \(\Lambda_{11} \geq \Lambda_{22} \geq \cdots \geq \Lambda_{NN}\), modes must be activated in the order of their indices. This does not indicate the proper value of \(N_{\text{act}}\) for a given scenario, and therefore this will be considered in the following discussion. For now, the value of \(N_{\text{act}}\) is assumed to be given.

Given \(N_{\text{act}}\) active modes, the optimization is performed through solution of

\[
0 = \frac{\partial}{\partial \rho_i} \left[ I_k + \frac{1}{\gamma} \left( \frac{P_T}{\sigma_0^2} - \sum_{n=1}^{N_{\text{act}}} \rho_n \right) \right]_{\rho_i = X_i} = \frac{1}{\ln 2} A_i(X_i) - \frac{1}{\gamma} \tag{3.22}
\]

for \(1 \leq i \leq N_{\text{act}}\), where \(1/\gamma\) represents the Lagrange multiplier and \(A_i(\rho_i)\) is given in (3.18). Letting \(\alpha = \gamma/\ln 2\), the solution to the Lagrange multiplier expression is

\[
X_i = \frac{2\Lambda_{ii} \alpha - 3 \pm \sqrt{1 - 12\Lambda_{ii} \alpha + 4\Lambda_{ii}^2 \alpha^2}}{4\Lambda_{ii}}. \tag{3.23}
\]

The two possible values for \(X_i\) need to be interpreted carefully. The top plot in Fig. 3.0 shows the first derivative from (3.18), with the straight line representing the value \(1/\alpha\). With reference to (3.23), the two solutions for \(X_i\) correspond to the two points at which
the line for $1/\alpha$ in Fig. 3.0 crosses the curve for the $i$th derivative. Clearly, the upper and lower signs in (3.24) correspond to solutions in the concave and convex intervals, respectively. Of course, because $X_i$ must be non-negative, depending on the value of $\alpha$, it is possible that the number of valid solutions is one or zero, with the latter case indicating that the mode certainly will not be activated.

Naturally, if we constrain the solution so that all active modes lie within their concave intervals, the Lagrange multiplier solution represents a global maximum under the constraints. Similarly, if all active modes lie within their convex intervals, the solution represents a global minimum, which naturally is not of interest in this problem. We are therefore left to explore the optimality of a solution with at least one mode in its convex interval. To this end, we consider the situation of two modes with indices $r$ and $s$ and $\Lambda_{rr} > \Lambda_{ss}$. Given our argument above that modes should be activated in the order of decreasing eigenvalues, we consider that mode $r$ is active, and we wish to understand the implications of taking energy from mode $r$ to activate mode $s$.

For the following discussion, we ignore the scale factor $\ln 2$ that appears in the derivatives of (3.18) and (3.20), as it has no impact on the comparative analysis that we undertake and its pervasive presence complicates the presentation. This simply means that changes in mutual information below actually represent mutual information scaled by $\ln 2$. Figure 3.1 plots representative forms of $A_r(\rho)$ and $A_s(\rho)$ as a function of $\rho$ assuming $\Lambda_{rr} = 2 \Lambda_{ss} = 10$. If $X_s$ represents the energy allocated to mode $s$, then the increase in mutual information as a result of this mode is given by

$$
\Delta I_{k,s} = \int_0^{X_s} A_s(\rho) d\rho. \tag{3.25}
$$

Similarly, the decrease in mutual information as a result of the energy taken from mode $r$ and allocated to mode $s$ is given by

$$
\Delta I_{k,r} = \int_{X_r}^{X_r + X_s} A_r(\rho) d\rho. \tag{3.26}
$$
Figure 3.2: The functions $A_r(\rho)$, $A_s(\rho)$ and $C(\rho)$. The shaded portions represent integrals that show the impact of diverting energy from the higher-order mode $A_r(\rho)$ to the lower-order mode $A_s(\rho)$ in its convex interval.

The shaded areas in the top plot of Fig. 3.1 show the areas represented by these integrals. The graphical interpretation of the Lagrange multiplier value from Fig. 3.1 helps us to recognize that $A_s(X_s) = A_r(X_r) = 1/\alpha$.

We need to show that (3.25) is always less than (3.26) for mode $r$ in its concave interval and mode $s$ in its convex interval. To accomplish this, we introduce an auxiliary function $C(\rho) = 1/\rho$ and use the notation

$$\frac{dC(\rho)}{d\rho}\bigg|_{\rho=\rho_C} \quad A'_r(\rho_r) = \frac{dA_r(\rho)}{d\rho}\bigg|_{\rho=\rho_r}. \quad (3.27)$$

As a first step in our proof, we wish to show that

$$\int_{X_C}^{X_s} C(\rho)d\rho < \Delta I_{k,r} \quad (3.28)$$

where $C(\rho) = 1/\rho$ and $X_C$ is shown in the bottom plot of Fig. 3.1 as the solution to $C(X_C) = 1/\alpha$. 

27
To explore (3.28), we first generally examine the slopes of our functions when \( A_r(\rho_r) = C(\rho_C) = 1/\rho_C \), which leads to

\[
C'(\rho_C) = -\frac{1}{\rho_C^2} = -A_r^2(\rho_r). \tag{3.29}
\]

Using (3.19) and (3.21), we can express \( A_r(\rho_r) \) and \( A'_r(\rho_r) \) in terms of the physical variables \( \Lambda_{rr} \) and \( \rho_r \), leading to the relationship

\[
A'_r(\rho_r) - C'(\rho_C) = \frac{2\Lambda_{rr}^2}{(\Lambda_{rr}\rho_r + 1)^2(2\Lambda_{rr}\rho_r + 1)^2}. \tag{3.30}
\]

Now, we rewrite (3.28) as

\[
\int_0^{X_s} A_r(\rho + X_r) - C(\rho + X_C) d\rho. \tag{3.31}
\]

Based on the result in (3.30), whenever the integrand of (3.31) is zero, the slope of the integrand must be positive. This occurs at \( \rho = 0 \), and may occur at other points along the integration. What this means is that the integrand starts at a value of zero with positive slope and then can never become negative. The mean value theorem tells us that a function that is nonnegative cannot have a negative integral, meaning that (3.28) must be true.

While the top plot of Fig. 3.1 suggests that \( X_r \) lies in the concave interval of mode \( r \), the only requirement in this development is that \( A_r(\rho_r) = C(\rho_C) \), which occurs in the Lagrange multiplier solution when \( \rho_C = X_C \) and \( \rho_r = X_r \) where \( X_r \) can be either of the two possible roots in (3.24). In other words, (3.28) is satisfied regardless of which of the two possible values of \( X_r \) is selected.

As a second step in our proof, we wish to show that

\[
\Delta I_{k,s} < \int_{X_C}^{X_C + X_s} C(\rho) d\rho. \tag{3.32}
\]
In this case, we know that $A_s(X_s) = C(X_C) = 1/X_C$ or that $X_C = 1/A_s(X_s)$. Therefore, the right hand side of (3.32) can be written as

$$\int_{1/A_s(X_s)}^{1/A_s(X_s)+X_s} \frac{1}{\rho} d\rho = \ln [1 + X_s A_s(X_s)]. \quad (3.33)$$

Using (3.18) in (3.25), the left hand side of (3.32) becomes

$$\Delta I_{k,s} = \ln \left[ 1 + \frac{\Lambda_{ss}^2 X_s^2}{(2\Lambda_{ss} X_s + 1)} \right]. \quad (3.34)$$

Given (3.33) and (3.34) and since $\ln(1+y)$ is monotonic in $y$, proof of (3.32) reduces to proving that

$$\frac{\Lambda_{ss}^2 X_s^2}{(2\Lambda_{ss} X_s + 1)} < X_s A_s(X_s). \quad (3.35)$$

By using (3.19) to express $A_s(X_s)$ in terms of the physical parameters, we can simplify (3.35) to

$$X_s < 1/\Lambda_{ss} = \bar{X}_s. \quad (3.36)$$

In other words, provided that $X_s < \bar{X}_s$, (3.35) and therefore (3.32) are true. Based on the analysis of Section 3.2.1, we know that the boundary between the convex and concave intervals for mode $s$ is at the point $\hat{\rho}_s = 1/\sqrt{2}\Lambda_{ss}$, and therefore $\bar{X}_s > \hat{\rho}_s$ which means that $\bar{X}_s$ lies in the concave interval for mode $s$.

With both (3.28) and (3.32) satisfied, it follows that $\Delta I_{k,s} < \Delta I_{k,r}$ for $\Lambda_{ss} < \Lambda_{rr}$ and $X_s < \bar{X}_s$. This means that the increase in mutual information due to allocation of energy to mode $s$ is smaller than the corresponding decrease in mutual information caused by taking that energy from mode $r$. A few points regarding this finding are of significance for directing the optimization strategy.

1. If mode $r$ is active, it is suboptimal to activate mode $s$ unless we can ensure that $\rho_s > \bar{X}_s = 1/\Lambda_{ss}$. Note that $\bar{X}_s > \hat{\rho}_s$ where $\hat{\rho}_s$ is the boundary between the convex and concave intervals for mode $s$, which means that the energy allocation to mode $s$ is within its concave interval.
2. Because it is suboptimal to have the energy allocated to mode \( r \) be smaller than that allocated to mode \( s \), activation of mode \( s \) should only occur when the energy allocated to both modes \( r \) and \( s \) is larger than \( \bar{X}_s \). We recognize that since \( \bar{X}_r < \bar{X}_s \), this means that the energy allocated to mode \( r \) must be above its threshold \( \bar{X}_r \) and will be within the mode’s concave interval.

3. We recognize from the previous proof that (3.25) will also be less than (3.26) when \( X_r \) is in its convex interval (smaller root in (3.24)). Therefore, if the available energy \( P_T \) is small, it must all be allocated to mode \( r \), which may place the solution within the convex interval of mode \( r \). However, based on the arguments above, once multiple modes are active, the solution must lie within their concave intervals and therefore represents the global maximum.

3.2.3 Implementation

These results can be combined to formulate a relatively straightforward algorithm for determining the energy allocation that maximizes the mutual information. For \( N_{\text{act}} \) active modes and given our ordering of the eigenvalues, we know that all active modes must receive an energy allocation larger than the threshold corresponding to the active mode with the smallest eigenvalue \( \bar{X}_{N_{\text{act}}} = 1/\Lambda_{N_{\text{act}}N_{\text{act}}} \). Therefore, our Lagrange multiplier solution must find the values of \( \rho_i \) that maximize \( I_k \) subject to the constraint \( \rho_i > \bar{X}_{N_{\text{act}}} \) for \( i \leq N_{\text{act}} \) and \( \sum_{i=1}^{N_{\text{act}}} \rho_i = P_T/\sigma_0^2 \). Since this constraint leads to a convex solution set, the Lagrange multiplier solution will be a global maximum. Furthermore, the constraint means that we must have \( P_T/\sigma_0^2 \geq N_{\text{act}} \bar{X}_{N_{\text{act}}} \), an observation that allows us to quickly determine the maximum number of active modes \( (N_{\text{act,m}}) \) that can be supported.

Given this upper bound on \( N_{\text{act}} \), we now construct the Lagrange multiplier solution for all possible values of \( N_{\text{act}} \) in the range \( 1 \leq N_{\text{act}} \leq N_{\text{act,m}} \). For each possible value of \( N_{\text{act}} \), a numerical search must be used to find the value of \( \alpha \) in (3.24) such that the energy constraint \( \sum_{i=1}^{N_{\text{act}}} X_i = P_T/\sigma_0^2 \) is satisfied. In this work, we begin by explicitly defining a cost
function in the variable $\alpha$, or

$$f(\alpha) = \sum_{i=1}^{N_{\text{act}}} X_i(\alpha) - P_T/\sigma_0^2.$$  \hfill (3.37)

Then, according to the Newton-Raphson method [39], we initialize $\alpha$ using $\alpha_0 = 0$ and iteratively compute new values using $\alpha_{k+1} = \alpha_k - f(\alpha_k)/f'(\alpha_k)$, where $f'(\alpha) = df(\alpha)/d\alpha$, until $|f(\alpha)| < 10^{-8}$. The final value of $\alpha_k$ is then used to compute the energy allocation and the corresponding mutual information $I_k$ for the specified value of $N_{\text{act}}$. Finally, we select the value of $N_{\text{act}}$ that achieves the highest value of $I_k$.

### 3.2.4 Interpretation

To interpret this optimal sounding energy allocation (OSEA) upper bound, consider a correlation matrix with two eigenvalues, $\Lambda_{11} = 10$ and $\Lambda_{22} = 1$. The energy allocated to each mode using a “waterfilling” solution, equal energy allocation (traditional sounding), and OSEA is plotted as a function of the normalized total energy $P_T/\sigma_0^2$ in the top plot in Fig. 3.2. In the waterfilling solution used, the energy allocated to an active mode is given as

$$\rho_i = (\zeta - 1/\Lambda_{ii})^+$$ \hfill (3.38)

where $(x)^+ = x$ if $x > 0$ and $(x)^+ = 0$ if $x \leq 0$ and $\zeta$ is chosen so that the total energy constraint is satisfied. Therefore, for waterfilling the energy allocated to each active mode always increases with the available energy. In contrast, for OSEA the solution suddenly allocates a large amount of energy ($\rho_2 > 1/\Lambda_{22}$) to the second mode by taking the same amount of energy from the already active mode once enough energy is available for both modes to be active.

The bottom plot in Fig. 3.2 shows the difference $I_k - I_{k,E}$, where $I_{k,E}$ represents the mutual information when the modes receive equal energy (traditional probing). The performance when all energy is allocated to a single mode is also shown. As can be seen, OSEA and waterfilling achieve the same performance (equal to that for the single mode) until waterfilling prematurely allocates energy to the second mode. When OSEA abruptly allocates
energy to the second mode, the slope of the mutual information curve also changes abruptly. At high SNR (large values of $P_T$), OSEA, waterfilling, and equal energy approach the same performance, as at this point the differences in energy allocation become a small portion of the total energy allocated to each mode. While the differences in mutual information for this scenario are small, this example illustrates the operation of OSEA.

3.3 Results

Some computational examples illustrate the impact of OSEA on the potential performance of key establishment techniques. In these scenarios, Bob’s linear array consists of $N_b = 3$ vertically-oriented dipoles with half-wavelength element spacing. Alice’s linear array similarly consists of $N_a$ vertically-oriented dipoles equally spaced over a total aperture of two wavelengths, where $2 \leq N_a \leq 10$.

Each channel realization is generated by assuming three clusters of multipaths propagating in the horizontal plane between Alice and Bob. The average propagation environment
is defined by its power angular spectrum (PAS) representing the average power per unit angle in the horizontal plane. The contribution of each cluster to the PAS satisfies a truncated Laplacian functional form with a variance of $30^\circ$, with the departure angle $\phi_{t,q}$ and arrival angle $\phi_{r,q}$ of the $q$th cluster generated as a realization of a uniformly distributed random variable on $[0, 2\pi)$. The $q$th cluster has a magnitude $\beta_q$ generated as a Rayleigh random variable normalized such that $\sum_q \beta_q^2 = 1$.

The covariance for each random channel realization is then computed using the closed-form integration technique in [40] that uses the radiation patterns for the antenna elements and the PAS describing the propagation environment. All results represent averages computed over 60 random channel realizations.

Fig. 3.3 plots the achieved average mutual information as a function of $N_a$ for three different values of $P_T/\sigma_0^2$ for both OSEA and equal energy allocation. As the number of elements in Alice’s array increases, the new modes created tend to have reduced eigenvalues, since the array elements are packed within the same total array aperture and therefore the resulting channel coefficients will be increasingly similar. There will, however, be a slight increase in the values of the dominant eigenvalues. Because OSEA considers the relative values of these eigenvalues, it allocates the probing energy to properly exploit the strong modes and de-emphasize or altogether ignore the weak modes. As a result, the bound on the key length for OSEA increases with $N_a$. In contrast, equally allocating the energy across all modes (traditional probing) wastes resources on poor modes, leading to a reduction in performance with $N_a$ for this strategy.

Fig. 3.4 plots the ratio $I_k/I_{k,E}$ as a function of $P_T/\sigma_0^2$ with $I_k$ computed using OSEA for three different values of $N_a$. This plot reaffirms that the benefit of using optimal energy allocation is substantial when the total energy available for probing is limited. As $P_T$ gets large, however, the performance benefit decreases since the difference in energy allocated to each mode becomes only a small fraction of the total allocated energy, as discussed previously.

### 3.4 Practical Implementation

#### 3.4.1 Transmit and Receive Beamformers

As discussed in Section 3.1.1, the optimization developed in this paper is for an abstract system that disregards how actual radios could estimate the required channels.
Figure 3.4: The mutual information achieved using OSEA compared to that achieved using equal energy allocation as a function of the number of antennas in Alice’s array.

More specifically, estimating the channel response vectors in (3.1) requires that we apply transmit and receive beamformers that implement the probing vector \(\mathbf{v}_i\), but in general \(\mathbf{v}_i\) cannot be separated into distinct contributions at the transmitting and receiving nodes.

To modify the framework for practical implementation, consider now the real system shown in Fig. 3.5 in which Alice’s and Bob’s arrays are connected to a beamformer. Alice’s multi-antenna transmission to Bob represents the \(i\)th symbol \(x_i\) from the information sequence weighted by the \(N_a \times 1\) beamforming vector \(\mathbf{a}_i\), and the signals received on Bob’s antennas are weighted by the \(N_b \times 1\) beamforming vector \(\mathbf{b}_i\). Without loss of generality, it is assumed that each \(\mathbf{a}_i\) and \(\mathbf{b}_i\) are unit length and that the energy allocated to probing the beamformer weighting pair is specified through the magnitude of \(x_i\). The received symbol can therefore be expressed as

\[
\hat{w}_{b,i} = \mathbf{b}_i^T \mathbf{H} \mathbf{a}_i x_i + \eta_{b,i} = (\mathbf{a}_i^T \otimes \mathbf{b}_i^T) \mathbf{h} x_i + \eta_{b,i}
\]  

(3.39)
Figure 3.5: The ratio of the mutual information achieved using OSEA to that achieved with equal energy allocation as a function of the normalized total available energy $P_T/\sigma_0^2$.

Figure 3.6: Block diagram of the beamforming system used for channel estimation and establishment of secret encryption keys.

where $\{\cdot\}^T$ indicates a transpose and $\otimes$ represents a Kronecker product. It is possible that the transmit and receive beamformer vectors remain constant over a block of symbols. A similar description applies when Bob transmits to Alice.
Comparing (3.39) to (3.1) and given the development in Section 3.1.2, we see that if \((a_i^T \otimes b_i^T)x_i = v_i^\dagger\), with \(V\) chosen as the scaled eigenvectors of \(R = E \{ hh^\dagger \}\), then the practical system of Fig. 3.5 is able to implement the probing specified in (3.1). The challenge is finding the transmit and receive beamformers from this requirement, since for general channels there is no guarantee that the unitary eigenvectors of \(R\) satisfy this Kronecker product form.

To approach this problem, we enlist the commonly-used assumption in MIMO channel modeling and communication system analysis that the covariance matrix is separable [41]. Given our definition of \(h\), this separability means we can write the covariance as \(\bar{R} = R_a \otimes R_b\) where \(R_a\) and \(R_b\) represent one-sided covariance matrices for Alice and Bob, respectively. Physically, this model is known to have limitations, as it assumes that the multipath structure at one end of the link is independent of the structure at the other end [42]. However, given the Kronecker form in (3.39), this represents a likely candidate for allowing direct determination of the transmit and receive beamforming vectors from the OSEA framework.

To estimate the one-sided covariance matrices, we use the Rank-1 Approximation (which refers to the algorithm used, not to the rank of the final matrices constructed) to find the matrices \(R_a\) and \(R_b\) that minimize \(\|R - R_a \otimes R_b\|_F\), where \(\|\cdot\|_F\) represents the Frobenius norm [43]. Based on these matrices, we apply the OSEA algorithm to the matrix \(\bar{R} = R_a \otimes R_b\) to establish the optimal probing energy \(p_i\) for the \(i\)th basis vector \(v_i = \tilde{u}_i\), where \(\tilde{u}_i\) is the \(i\)th eigenvector of \(\bar{R}\). Given the Kronecker form, the eigenvectors of \(\bar{R}\) can be written as \(\tilde{U} = U_a \otimes U_b\), where \(U_a\) and \(U_b\) represent the eigenvectors of \(R_a\) and \(R_b\), respectively. Therefore, suppose that the \(i\)th eigenvector is \(\tilde{u}_i = u_{a,m} \otimes u_{b,n}\), where \(u_{\xi,m}\) represents the \(m\)th column of \(U_{\xi}\), \(\xi \in [a,b]\). This means that the transmit and receive beamformers are \(a_i = u_{a,m}^*\) and \(b_i = u_{b,n}^*\).

Naturally, the covariance matrix for an actual channel generally is not separable, and therefore constraining the covariance to the Kronecker product form represents an approximation. Figure 3.6 compares the performance achieved when applying OSEA with the Kronecker form \(\bar{R}\) to the performance achieved when using the non-separable covariance \(R\) for the abstract system as a function of the number of elements \(N_a\) in Alice’s linear array for the same simulation scenario used in Fig. 3.3. These results demonstrate that use of
the Kronecker approximation results in very little performance degradation, at least for the channel structure considered here. This suggests that OSEA is potentially applicable to practical scenarios.

3.4.2 Antenna Array Mutual Coupling

While the development used to demonstrate the OSEA approach has assumed a simple model for the signals transmitted or received by Alice and Bob, the framework developed can naturally be applied to models that incorporate more complicated system descriptions. To illustrate application of the framework to such a scenario, we formulate a signal model that incorporates antenna array mutual coupling.

In this model, let \( \bar{e}_{a,n}(\Omega_a) \) and \( \bar{e}_{b,m}(\Omega_b) \) respectively represent the vector radiation patterns for the \( n \)th element in Alice’s array and the \( m \)th element in Bob’s array with all other elements in the arrays terminated in an open circuit. Note that \( \Omega_a \) and \( \Omega_b \) generically represent the angular coordinates in the coordinate frames designated for Alice and Bob, respectively. If Alice is transmitting and \( \bar{P}_{ba}(\Omega_b, \Omega_a) \) represents the dyadic (i.e. including...
polarization effects) complex gain function for transmission from Alice to Bob, then the
vector of open-circuit signal voltages (neglecting noise) on Bob’s array is given as
\[ v_{0,b} = H_0 i_a, \] (3.40)
\[ H_{0,mm} = \int \bar{e}_{b,m}(\Omega_b) \cdot \bar{P}_{ba}(\Omega_b, \Omega_a) \cdot \bar{e}_{a,n}(\Omega_a) d\Omega_b d\Omega_a \] (3.41)
where each entry of the vector \( i_a \) is the current driving an antenna element of Alice’s array.

Suppose that Alice’s array has a mutual impedance described by the matrix \( Z_a \) and is driven by a set of generators with voltages represented by the vector \( v_a \) and impedances represented by the impedance matrix \( Z_{L,a} \). The driving voltages and antenna currents are then related through the expression \( i_a = (Z_{L,a} + Z_a)^{-1}v_a \). Similarly, if Bob’s array has a mutual impedance described by the matrix \( Z_b \) and is terminated with an impedance represented by the matrix \( Z_{L,b} \), then the vector of currents through this termination is \( i_b = (Z_{L,b} + Z_b)^{-1}v_{0,b} \). Therefore, we can write
\[ i_b = (Z_{L,b} + Z_b)^{-1}H_0(Z_{L,a} + Z_a)^{-1}v_a = H_{ba}v_a. \] (3.42)

Notice that if we analyze the system in reverse with Bob transmitting the voltage vector \( v_b \), the vector of currents through Alice’s terminations is
\[ i_a = (Z_{L,a} + Z_a)^{-1}H_0^T(Z_{L,b} + Z_b)^{-1}v_b = H_{ab}v_b. \] (3.43)

As long as all impedance matrices are reciprocal \((Z = Z^T)\), then \( H_{ab} = H_{ba}^T \), which indicates that the channel including the antenna and termination impedances is reciprocal as required for key establishment.

Because our computations construct the covariance matrix based on the radiation patterns and the PAS of the environment, we must now use our development to determine the effective radiation patterns of the array as seen at the terminals where the transmit voltages and received currents are observed. With the help of (3.41) and (3.42), we can
construct effective radiation patterns for Alice and Bob that assume the form

\[
\hat{e}_{a,q}(\Omega_a) = \sum_{n=1}^{N_a} \left[(Z_{L,a} + Z_a)^{-1}\right]_{qn} \bar{e}_{a,n}(\Omega_a),
\]

(3.44)

\[
\hat{e}_{b,p}(\Omega_b) = \sum_{m=1}^{N_b} \left[(Z_{L,b} + Z_b)^{-1}\right]_{pm} \bar{e}_{b,m}(\Omega_b)
\]

(3.45)

where \([\cdot]_{mn}\) represents the \(mn\)th element of the matrix within the brackets and we have used that \(Z_a\) and \(Z_{L,a}\) are reciprocal (symmetric). These effective patterns can be used with the PAS to construct the covariance, and the computations can then proceed as discussed in Section 3.3.

One subtle challenge with this practical formulation is that as the coupling becomes stronger in the transmit array, for certain terminations the matrices \(Z_{L,\xi} + Z_\xi\), \(\xi \in [a, b]\) can develop small eigenvalues that can lead to very high gain subspaces in the effective channel corresponding to impractical supergain solutions. To ensure that the OSEA optimization does not use these solutions, we simply introduce loss in the transmit array elements. Specifically, we assume each element has an efficiency of 95%, which means that each diagonal element of \(R_\xi\) is modified according to \(R_{\xi,nn} \leftarrow R_{\xi,nn}/0.95\) where \(R_\xi\) is the real part of \(Z_\xi\).

For the half-wave dipoles used in this study, the radiation pattern of an element with all other elements open-circuited is well approximated by its isolated radiation pattern, which is known in closed form [44]. The mutual impedance matrix can also be approximated analytically [44]. We assume that the terminations are chosen to represent a \textit{self-impedance match} or \(Z_{L,\xi} = \hat{Z}_\xi^\dagger\) where \(\hat{Z}_\xi\) represents a diagonal matrix containing the diagonal elements of \(Z_\xi\).

The PAS for this simulation consists of three clusters in three-dimensional space, with each cluster shaped as a circularly symmetric Laplacian distribution, with a variance of 30°, centered at randomly realized (uniformly distributed) departure and arrival angles on the unit sphere. The magnitude of each cluster is chosen as discussed in Section 3.3. The results shown represent average performance over 360 random PAS realizations. Figure 3.7 plots the key rate as a function of number of elements in Alice’s array for different values of \(P_T/\sigma_0^2\). Because coupling tends to increase the similarity between channel coefficients
on adjacent antennas, at close element spacing the newly introduced eigenvalues are small. Therefore, when equally allocating energy across the entire probing bases, the probing energy is inefficiently used to estimate weak channel subspaces. This accounts for the sharp decrease in key rate for equal energy allocation once $N_a$ grows large. While the redistribution of the eigenvalues in the covariance with an increasing number of antenna elements can also result in some reduction in performance for OSEA, the basic behavior is that the performance remains constant after it reaches its maximum. This result demonstrates how practical signal models can dramatically impact the predicted performance and further illustrates that OSEA can be effectively applied even with such complicated and realistic models.

3.5 Chapter Summary

This chapter formulates the upper bound on the key rate that can be realized when a beamforming system establishes secret keys based on reciprocal wireless channel estimates. This upper bound arises when using the eigenvectors of the channel spatial covariance matrix
as beamforming weights and then optimally allocating energy for each weight vector. Simulations demonstrate that the achieved key rate is substantially higher than that achieved using traditional probing, particularly at low SNR. Because the eigenvectors used as beamformer weights cannot be easily separated into actual transmit and receive weights, this chapter proposes a suboptimal approach that achieves the required separability using a Kronecker approximation of the full covariance matrix. Simulations reveal that the performance of this suboptimal implementation can approach that achieved using the non-separable covariance.
Chapter 4

The Impact of the Propagation Environment on OSEA Key Generation

The results presented alongside OSEA development Chapter 3 are a limited demonstration of the technique in a Rayleigh-fading channel. The purpose of this chapter is to provide a detailed analysis of its behavior in a variety of environments. This is particularly relevant since the basic algorithm cannot be applied to realistic transmit and receive beamforming systems, and therefore an approximation must be used to allow this application. As this approximation is based on a Kronecker product representation of the multi-antenna spatial covariance and since this representation is known to suffer from deficiencies under certain propagation conditions [41, 42], exploration of the technique over a range of propagation conditions helps to define the accuracy and applicability of the technique. The analysis provided demonstrates that the practical channel estimation procedure based on the Kronecker approximation suffers when a moderate line-of-sight (LOS) component is present. To overcome this limitation, we propose a modification to the Kronecker representation that separately models the LOS and multipath contributions to the spatial covariance matrix. Simulation results show that the performance of this modified approach is close to that of the optimal upper bound over a broad range of propagation conditions.

4.1 LOS Propagation

4.1.1 Propagation Description

We use a simulation to illustrate the impact of propagation conditions on OSEA performance. In this analysis, we describe the propagation, assumed to be confined to the horizontal plane, using the power angular spectrum (PAS) $\Gamma(\phi_b, \phi_a)$ representing the expected power impinging on Bob’s array from the direction $\phi_b$ due to a signal component departing
Alice’s array from the angular direction $\phi_a$. The contribution of the LOS component to the PAS can be expressed as

$$\Gamma_0(\phi_b, \phi_a) = \beta_0 \delta(\phi_a - \phi_{a,0}, \phi_b - \phi_{b,0})$$ (4.1)

where $\delta(\cdot)$ denotes a delta function and $\phi_{a,0}$ and $\phi_{b,0}$ are the departure and arrival angles of the LOS component, respectively (assuming transmission from Alice to Bob). The $q$th cluster multipath, $q > 0$, is described by a truncated Laplacian function with an angular spread of $30^\circ$ in each angular dimension and centered at the transmit and receive angles $\phi_{a,q}$ and $\phi_{b,q}$, respectively. For simulation purposes, $\phi_{a,q}$ and $\phi_{b,q}$ for $q \geq 0$ are computed as random variables drawn from a uniform distribution on $[0, 2\pi)$. The magnitude $\beta_q$ representing the peak of the $q$th cluster for $q > 0$ is also a random variable drawn from the Rayleigh distribution and scaled so that the average power in the multipath clusters is

$$\frac{1 + K}{4\pi^2} \sum_{q=1}^{Q} \int \int \Gamma_q(\phi_b, \phi_a) \, d\phi_b \, d\phi_a = 1$$ (4.2)

where $\Gamma_q(\phi_r, \phi_t)$ is the PAS contribution from the $q$th multipath cluster, $Q$ is the total number of multipath clusters, and $K$ is the Rician $K$-factor [45]. The gain $\beta_0$ of the LOS component is then set to achieve this $K$-factor, or simply $\beta_0 = K$. The covariance for each random channel realization is then computed using the closed-form integration technique in [40]. To realize a Rayleigh channel, we simply set $\beta_0 = K = 0$.

### 4.1.2 OSEA Performance

For consistency, the same simulation parameters from Chapter 3 are used throughout the simulations presented in this chapter. Each simulated data point represents the average of 60 random realizations of the propagation environment. Both Alice and Bob have linear arrays of dipoles with Alice’s array spanning a total aperture of $2\lambda$, where $\lambda$ is the free-space wavelength. Bob’s array spans a total aperture of $1\lambda$ and consists of $N_b = 3$ dipoles. Unless otherwise specified, Alice’s array consists of $N_a = 6$ dipoles, the number of multipath clusters used to model the environment is $Q = 5$, and Rician channels are realized with a $K$-factor
of $K = 1$. All results are computed for various values of the signal-to-noise ratio (SNR) specified as $P_T/\sigma_0^2$.

Consider the case where the number of dipoles in Alice’s array is swept over the range $2 \leq N_a \leq 10$. Figure 4.0 plots the key rate $I_k$ assuming the propagation represents a Rayleigh channel ($K = 0$) for three different values of SNR. Curves are provided for OSEA with the full covariance, OSEA with the Kronecker approximation (Kron Approx), and traditional probing (Equal Allocation). Figure 4.1 provides the same results for a Rician channel. The general trend observed for OSEA and equal allocation are similar for both types of fading channels. At low SNR, however, the key rate achieved in the Rician propagation environment is greater than that achieved in the Rayleigh environment. This occurs because for low available energy $P_T$, the OSEA algorithm places most if not all of the energy into the beamformers associated with the dominant mode whose relative importance is higher for the Rician channel as a result of the LOS component. At high SNR, the algorithm benefits from more modes of similar quality, and therefore the Rayleigh environment tends to yield slightly

Figure 4.1: Key rate achieved using OSEA with the full covariance, OSEA using a Kronecker approximation to the covariance, and equal energy allocation as a function of the number of antennas in Alice’s array for several values of SNR with Rayleigh fading.
Figure 4.2: Key rate achieved using OSEA with the full covariance, OSEA using a Kronecker approximation to the covariance, and equal energy allocation as a function of the number of antennas in Alice’s array for several values of SNR with Rician fading.

higher values of $I_k$. Finally, the results reveal the benefit achieved using OSEA relative to that achieved using traditional probing.

Perhaps more importantly, Figs. 4.0 and 4.1 reveal important observations about the Kronecker approximation. Specifically, for Rayleigh fading, the Kronecker approximation appears to produce very little degradation in performance. Similarly, for low SNR and Rician fading where the dominant mode created by the LOS component dominates the contribution to $I_k$, the Kronecker approximation works well, indicating that it models the contribution of the LOS component with relatively high accuracy. However, at high SNR with Rician fading, the performance of OSEA with the Kronecker approximation can be worse than that achieved using traditional channel estimation. Because the Kronecker approximation assumes independent scattering at the transmitter and receiver, it tends to create high-gain modes that couple the LOS component at the transmitter with multipath energy at the receiver (and vice versa). Since real propagation does not include this coupling, allocation of energy to estimation of such non-physical modes degrades performance.
4.2 Modified Kronecker Approximation

The observed performance degradation resulting from application of OSEA with the Kronecker approximation in the presence of a LOS signal path motivates an alternative channel estimation strategy that performs well in all propagation environments. Since the LOS component is independent of the multipath clusters, the covariance matrix can be separated into the sum of a unit-rank LOS covariance matrix \( R_{\text{LOS}} \) and a full-rank multipath covariance matrix \( R_{\text{MP}} \). To model \( R_{\text{LOS}} \), we form the array steering vectors \( s_a \) and \( s_b \) for a plane wave departing from Alice’s array and impinging on Bob’s array. If \( \Lambda_{\text{LOS}} \) represents the channel gain for this plane wave propagation, then

\[
R_{\text{LOS}} = \Lambda_{\text{LOS}} (s_a \otimes s_b) \dagger (s_a \otimes s_b). \quad (4.3)
\]

When a strong LOS path is present, this LOS path is the main contribution to the dominant eigenvector/eigenvalue of \( R \). If \( u_1 \) is the dominant eigenvector of \( R \), we can compute estimates \( \tilde{s}_a \) and \( \tilde{s}_b \) to the LOS steering vectors using the Rank-1 approximation that minimizes \( \| u_1 - \tilde{s}_a \otimes \tilde{s}_b \|_F \). We then form the approximation to the LOS covariance matrix using

\[
\tilde{R}_{\text{LOS}} = \Lambda_{11} (\tilde{s}_a \otimes \tilde{s}_b) \dagger (\tilde{s}_a \otimes \tilde{s}_b) \quad (4.4)
\]

where \( \Lambda_{11} \) is the dominant eigenvalue of \( R \). Finally, we write the estimate of the multipath covariance matrix using the Kronecker form \( \tilde{R}_{\text{MP}} = \tilde{R}_{a,\text{MP}} \otimes \tilde{R}_{b,\text{MP}} \) where the component matrices are computed using the Rank-1 approximation that minimizes \( \| (R - \tilde{R}_{\text{LOS}}) - \tilde{R}_{\text{MP}} \|_F \). This procedure ensures that the approximations to the LOS and multipath covariance matrices have the Kronecker structure required for establishing transmit and receive beamformers.

With these approximations, let \( \Lambda_{ii} \) represent the \( i \)th eigenvalue of \( \tilde{R}_{\text{MP}} \). We apply the energy optimization scheme to the set of eigenvalues \( \left[ \Lambda_{11}, \Lambda_{12}, \Lambda_{22}, ..., \Lambda_{NN} \right] \) where \( N = N_a N_b \). The solution for the energy associated with \( \Lambda_{11} \) is then applied to transmission of the beamformer combination \( \tilde{s}_a \) and \( \tilde{s}_b \) while the energy associated with \( \Lambda_{ii} \) is applied to transmission of the beamformers constructed from the \( i \)th eigenvector of \( \tilde{R}_{\text{MP}} \). It is important to understand that this approach is actually probing the \( N \) dimensional space containing...
**Figure 4.3:** Key rate achieved using OSEA with the full covariance, Kronecker approximation, and modified Kronecker approximation as a function of the number of antennas in Alice’s array for several values of SNR with Rician fading.

We refer to this approach as the *modified Kronecker approximation*. Figure 4.2 compares the performance achieved with the modified Kronecker approximation to that obtained using OSEA and the original Kronecker approximation for Rician propagation under the conditions used for Fig. 4.1. These results demonstrate that in contrast to the behavior observed with the original Kronecker approximation, the performance achieved with the modified Kronecker approximation is very close to that obtained with OSEA applied to the full covariance matrix, even for high SNR. Figure 4.3 plots similar results in the Rayleigh environment under the conditions used for Fig. 4.0. In this case, the performance achieved using the modified Kronecker approximation is slightly worse than that obtained with the original Kronecker approximation. This degradation occurs because
the modified Kronecker approximation assumes a LOS component that is not present in the channel. However, the degradation created by this modeling error is relatively small since the result of first extracting the dominant eigenvector and then modeling the remaining dimensions of the covariance preserves in large part the dominant covariance eigenstructure.

While Figs. 4.2 and 4.3 suggest that the modified Kronecker approximation has the potential to allow accurate implementation of the OSEA algorithm for different propagation conditions, more detailed simulations are necessary to show that the technique is applicable over a range of propagation conditions. Several representative studies are therefore provided in the following sections.

### 4.2.1 Impact of Scatterer Density

Because the number of multipath clusters in the PAS generally impacts the number of significant modes available in the channel, we expect this parameter to directly influence the value of $I_k$. In this analysis, we sweep the number of multipath clusters in the PAS from
1 to 100. Figures 4.4 and 4.5 plot the results of this computation for Rician and Rayleigh fading, respectively. In interpreting these plots, we emphasize that because the PAS functions are normalized using (4.2), as the number of multipath clusters increases, weak eigenvalues become larger at the expense of the dominant eigenvalues. Therefore, when $P_T$ is large enough to exploit a large number of modes, increasing the number of clusters is beneficial. This explains the observation that $I_k$ increases with the number of clusters for both Rician and Rayleigh fading at an SNR of 25 dB. However, if $P_T$ is small and can only effectively use a small number of modes, then moving gain from the larger eigenvalues into the smaller ones decreases performance. This is readily observed in the curves for Rayleigh fading at an SNR of 5 dB. On the other hand, for Rician fading, because the gain of the LOS component does not change with the number of clusters and most of the energy is allocated to this dominant contribution to the propagation, the overall impact of decreased dominant mode quality is reduced.

It is interesting that this behavior at low SNR differs from that observed for the waterfilling solution used in multi-antenna communication capacity analysis. Specifically, in waterfilling, it is optimal to equalize all modes and assign a little communication energy to each. In OSEA, however, even if modes are nearly equal in quality, the solution will allocate energy only to a single mode until the total energy grows large enough for two modes to operate with relatively high energy allocations.

In contrast, at 15 dB SNR, the decrease observed for the OSEA solution in Rayleigh fading is less pronounced than that for Rician fading. At this SNR level, OSEA is able to better exploit a large number of modes, and therefore for Rayleigh fading we essentially observe a balance between using more modes and degradation in the dominant modes. For Rician fading, however, the presence of the dominant mode detracts from the growth enabled by exploiting additional multipath modes, since such a large fraction of the probing energy will be allocated to the single dominant mode.

Finally, it is interesting to observe that the performance for the Kronecker approximation in Rician fading remains essentially constant with the number of multipath clusters for all SNR values. On the other hand, the trend of the modified Kronecker approximation is similar to that of the OSEA solution based on the full covariance. For Rayleigh fading,
Figure 4.5: Key rate achieved using OSEA with the full covariance, Kronecker approximation, and modified Kronecker approximation as well as with equal energy allocation as a function of the number of multipath clusters for several values of SNR with Rician fading.

we again observe that the Kronecker approximation achieves a performance close to that of OSEA.
4.2.2 Impact of Rician $K$-Factor

The preceding results demonstrate that the presence of an LOS component in the propagation channel influences the key rate obtained using OSEA with or without the Kronecker approximations, motivating a study of how the relative strength of the LOS component as measured by the Rician $K$-factor impacts the achieved value of $I_k$. Through
our numerical simulations we have observed generally that as $K$ gets larger, the dominant eigenvalue of $\mathbf{R}$ increases almost linearly with $K$ and, given our normalization, that increase occurs at the expense of the remaining eigenvalues. As this occurs, OSEA with or without the Kronecker approximations devotes an increased fraction of the available energy to channel estimation when using the dominant eigenvector at the expense of channel estimation when using other eigenvectors. Stated another way, as the LOS component increases in relative importance, it becomes the single source of significant information between the two channel estimates, reducing the relative importance of the multipath contributions for key establishment.

This analysis helps us to explain the behavior of the curves in Fig. 4.6 that plots $I_k$ as a function of the Rician $K$-factor. Specifically, at low SNR significant energy is allocated to the dominant mode, the relative importance of which increases with $K$. As a result, for $P_T/\sigma_0^2 = 5$ dB, $I_k$ increases with $K$. In contrast, when the SNR is high, OSEA allocates energy to estimation of a significant number of modes when $K$ is small. As $K$ increases, the solution allocates additional energy to the dominant mode, which results in a slight increase in the contribution to $I_k$ due to this mode. However, higher-order modes experience decreasing eigenvalues with increasing $K$, and less energy is available to estimation of these modes due to the increased allocation to the dominant mode. The resulting decrease in the number of active modes dramatically reduces $I_k$, and the overall result is a net reduction in the key rate.

To reinforce these observations, we consider Fig. 4.7 that plots the average number of active modes $N_{\text{act}}$ as a function of $K$ for the simulations used in Fig. 4.6. These results show that at 5 dB SNR, only the dominant mode is active for $K \geq 6$, confirming the importance of the growth of this mode with $K$ for small SNR. For high SNR, we observe a dramatic reduction in the number of active modes with increasing $K$ corresponding to the decrease in $I_k$ shown in Fig. 4.6.

The results in Fig. 4.6 also provide insight into the impact of the LOS component on the performance of OSEA with the Kronecker approximations. In general, the accuracy of the Kronecker approximation initially decreases with increasing $K$ but then becomes more accurate as $K$ moves beyond a certain value. This trend is more clearly illustrated in the
Figure 4.7: Key rate achieved using OSEA with the full covariance, Kronecker approximation, and modified Kronecker approximation as well as with equal energy allocation as a function of the Rician $K$-factor for several values of SNR.

top plot of Fig. 4.8 which shows the normalized difference between the performance achieved using OSEA and that obtained using the Kronecker approximation. As $K$ becomes large, the covariance matrix $\mathbf{R}$ has a rank that steadily decreases and ultimately reduces to unity. At this point, the Kronecker approximation is able to accurately model the unit-rank matrix, and therefore the performance improves. For the modified Kronecker approximation, the performance tracks the OSEA upper bound closely for all values of $K$. This is shown in
Figure 4.8: The number of active modes used in OSEA as a function of the Rician $K$-factor for several values of SNR.

The bottom plot in Figure 4.8, which shows that the relative error between the modified Kronecker approximation and the OSEA upper bound is less than 5% for all $K$.

The preceding results demonstrate that when a notable LOS component is present, the modified Kronecker approximation outperforms the original Kronecker approximation. However, we have yet to quantify the relative performance of the two algorithms with no LOS component. Figure 4.9 plots the normalized performance difference between each Kronecker approach and OSEA in a Rayleigh environment. These results demonstrate that for SNR values above a few dB, the original Kronecker approximation outperforms the modified Kronecker approximation. This is expected since the modified Kronecker approximation assumes the presence of a non-existent LOS component.

4.3 Chapter Summary

The results presented in this chapter explore the upper bound on the number of key bits that can be established using reciprocal channel estimation between beamformed multi-antenna nodes. Specifically, the performance of the upper bound as well as practical
implementations are explored as a function of the propagation conditions. The results show that for Rayleigh fading, the Kronecker approximation to OSEA works well, but that this approximation suffers from performance difficulties when a LOS component is considered. To overcome this limitation, a modified Kronecker approximation is proposed which performs well in both Rician and Rayleigh environments. Detailed simulation results demonstrate the impact of multipath richness, strength of the LOS component, SNR, and array size on the performance.

Figure 4.9: Normalized difference between the number of bits enabled by OSEA and by the Kronecker approximation (top) or by the modified Kronecker approximation (bottom) as a function of the Rician $K$-factor.
Figure 4.10: Normalized difference between the number of bits enabled by OSEA and by the Kronecker approximation or the modified Kronecker approximation with Rayleigh propagation as a function of SNR.
Chapter 5

A Spatial Bound on Key Rate for Beamformer Channel Probing

Given the channel coefficient matrix $H$ – which is specified by the electromagnetic propagation and the antenna arrays – and a constraint on the total energy available for channel estimation, the OSEA solution constructs the beamforming weights (including energy applied to each combination of beamformers) that maximize the achievable key rate. This is the upper bound achievable when the antenna array (and propagation channel) are specified, and does not consider the optimal design of the antennas to further maximize the achievable key rate for a given propagation channel. The objective of this chapter is to apply the OSEA algorithm to determine the optimal array of currents that should exist within constrained apertures at the transmitter and receiver to optimally exploit the degrees of freedom in a multipath channel for the purposes of maximizing the achievable key rate. The development builds on prior studies of the multi-antenna communication capacity [46], but is fundamentally unique due to the differences between capacity and key rate mutual information expressions.

The chapter proceeds by applying OSEA to a generic set of currents represented by a set of basis functions that is complete over the finite aperture and determining the optimal weighting coefficients for the basis expansion that maximize the key rate given a stochastic description of the channel. The development demonstrates that the key rate converges to a limiting value as the number of Fourier basis functions becomes large. This is followed by a proof that this limiting key rate, called the key rate spatial bound, upper bounds the key rate achievable by any set of square integrable currents in the same volume satisfying a common energy constraint. The formulation also incorporates a modification to the OSEA algorithm to accommodate the mutual coupling resulting from closely-arranged currents as well as spatially-correlated estimation errors produced by interference. The treatment then
provides a practical mechanism for limiting the number of basis functions that should be used for computations. Simulations are used to explore the implications of the algorithm modifications, the convergence properties of the technique, and the general behaviors of the key rate spatial bound.

5.1 Antenna Representation

For simplicity in presentation, this work limits the discussion to two dimensions which allows use of scalar currents. While the approach can naturally be extended to three-dimensional vector currents, the notational complexity detracts from the presentation of the core ideas. We define the mathematical representation for Alice transmitting to Bob, recognizing that this implicitly defines the representation for the reverse link given the reciprocal nature of electromagnetic propagation and assuming reciprocal antennas. The currents representing Alice’s antennas flow in the $z$ direction and are confined to a rectangular region $V_a$ of dimensions $L_{a,x}$ and $L_{a,y}$ in the $x$ and $y$ dimensions, respectively. Similarly, the currents representing Bob’s antennas flow in the $z$ direction and are confined to the rectangular region $V_b$ of dimensions $L_{b,x}$ and $L_{b,y}$ in the $x$ and $y$ dimensions, respectively. We use the vector notation $\mathbf{r}_a = (x_a, y_a)$ to denote the coordinates relative to Alice’s coordinate frame, with a similar definition $\mathbf{r}_b$ for Bob’s coordinate frame.

Given this notation, the $n$th Fourier current basis function for Alice’s antenna is defined as

$$f_{a,n} (\mathbf{r}_a) = \frac{1}{\sqrt{L_x L_y}} \exp \left\{ j \left( \frac{2\pi}{L_{a,x}} n_x x_a + \frac{2\pi}{L_{a,y}} n_y y_a \right) \right\}$$

where $n$ is a two-dimensional index specifying the basis function orders $n_x$ and $n_y$ that respectively represent the number of periods in $x$ and $y$. Ignoring the usual cylindrical wave behavior of the far-field radiation from this antenna, the current distribution can be converted into a vertically polarized far-field radiation pattern $e_{a,n} (\phi_a)$ using the integral [31]

$$e_{a,n} (\phi_a) = \int_{V_a} G (\phi_a, \mathbf{r}_a') f_{a,n} (\mathbf{r}_a') \, d\mathbf{r}_a'$$

58
where $\phi_a$ is the angle in polar coordinates within Alice’s coordinate frame and $G(\phi_a, r'_a)$ is the two dimensional Green’s function given as

$$G(\phi_a, r'_a) = \exp\left\{ \frac{2\pi}{\lambda} (x'_a \cos \phi_a + y'_a \sin \phi_a) \right\} \quad (5.3)$$

where $\lambda$ is the free space wavelength. The far-field radiation pattern becomes

$$e_{a,n}(\phi_a) = \sqrt{L_{a,x} L_{a,y}} \text{sinc} \left( \frac{L_{a,x}}{\lambda} \cos \phi_a + n_x \right) \text{sinc} \left( \frac{L_{a,y}}{\lambda} \sin \phi_a + n_y \right) \quad (5.4)$$

where $\text{sinc}(x) = \sin(\pi x)/\pi x$.

With this representation, we can formulate the elements of the spatial covariance matrix for the channel coefficient matrix $H$. Let $\beta(\phi_b, \phi_a)$ represent the complex gain of the vertically-polarized electric field departing Alice’s antenna at angle $\phi_a$ and impinging on Bob’s antenna from angle $\phi_b$. The complex gain observed between Alice’s $n$th basis functions and Bob’s $m$th basis function is

$$H_{mn} = \int \int e_{b,m}(\phi_b) \beta(\phi_b, \phi_a) e_{a,n}(\phi_a) \, d\phi_a \, d\phi_b. \quad (5.5)$$

Assuming that $\beta(\phi_b, \phi_a)$ represents a zero-mean Gaussian random process, if the field departing into (arriving from) one angle is uncorrelated with that departing into (arriving from) another angle, then the multipath gain function satisfies

$$E\{\beta(\phi_b, \phi_a) \beta(\phi'_b, \phi'_a)^*\} = B(\phi_b, \phi_a) \delta(\phi_a - \phi'_a) \delta(\phi_b - \phi'_b) \quad (5.6)$$

where $B(\phi_b, \phi_a)$ is the power azimuth spectrum (PAS). The covariance matrix element corresponding to $H_{mn}$ and $H_{uv}$ is

$$R_{(mn),(uv)} = E\{H_{mn}H_{uv}^*\} \quad (5.7)$$

$$= \int \int e_{b,m}(\phi_b) e_{b,u}^*(\phi_b) B(\phi_b, \phi_a) e_{a,n}(\phi_a) e_{a,v}^*(\phi_a) \, d\phi_a \, d\phi_b. \quad (5.8)$$
5.2 Convergence

The OSEA solution applies to a finite set of antenna currents. Therefore, extending the solution to a countably infinite basis function expansion of the antenna currents requires a proof that the key rate converges as the number of basis functions becomes large. This is accomplished here by first demonstrating that the mode gains converge as the number of basis functions becomes large. It is then shown that when the mode gains converge, the OSEA key rate also converges.

5.2.1 Mode Gain Convergence

Proving convergence of the eigenvalues of the covariance as the number of basis functions grows requires formulation of the diagonal elements of the covariance matrix. From (5.8), we have

\[ R_{(m), (n)} = \int \int |e_{b, (\phi_b)}|^2 B(\phi_b, \phi_a) |e_{a, (\phi_a)}|^2 d\phi_a d\phi_b. \]  

(5.9)

To help make the notation explicit, we denote the full covariance matrix corresponding to an infinite set of basis functions as countably infinite dimensional matrix \( R_\infty \). To demonstrate convergence, we first form the covariance matrix \( R^{(M)} \) involving the lowest-order basis functions with indices that satisfy \( \{m_x, m_y, n_x, n_y\} \leq M \), where \( M \) is a positive integer. We then define the covariance \( R_e \) as the covariance matrix for the channel coefficients corresponding to basis functions excluded from \( R^{(M)} \). The full covariance \( R_\infty \) can then be written as a block matrix representation with \( R^{(M)} \) and \( R_e \) representing the matrices on the diagonal. Naturally, there will be additional entries in \( R_\infty \) represented by blocks off the diagonal, but these are not important for the purposes of demonstrating convergence.

With this representation,

\[ \text{tr} (R_\infty) = \text{tr} (R^{(M)}) + \text{tr} (R_e). \]  

(5.10)

Since both \( \text{tr} (R_\infty) \) and \( \text{tr} (R_e) \) sum over an infinite number of elements, it is possible that they can diverge. However, because they differ only by the constant \( \text{tr} (R^{(M)}) \), if one con-
verges, then both will. For now, we assume that the trace of each exists and is finite. Since the matrices are covariances, each is positive semi-definite. With $R^{(M)}$ forming a sub-matrix along the diagonal of the Hermitian matrix $R_\infty$, it is shown in [47, p. 311] that

$$\lambda_i(R_\infty) \geq \lambda_i(R^{(M)})$$

(5.11)

where $\lambda_i(\cdot)$ denotes the $i$th largest eigenvalue of the matrix argument. Each $\lambda_i(R_\infty)$ is finite since the trace is assumed finite and $\text{tr}(R_\infty) = \sum_i \lambda_i(R_\infty)$ where each $\lambda_i(R_\infty)$ is non-negative. Recognizing that $\text{tr}(R^{(M)}) = \sum_i \lambda_i(R^{(M)})$ allows (5.10) to be expressed as

$$\sum_{i=1}^{\infty} \lambda_i(R_\infty) = \text{tr}(R_e) + \sum_{i=1}^{\infty} \lambda_i(R^{(M)})$$

(5.12)

where $\lambda_i(R^{(M)}) = 0$ if $i$ is larger than the dimension of $R^{(M)}$. Combining (5.11) with (5.12) and recognizing that all the quantities in both are non-negative, it can be shown for each $i$ that

$$\lambda_i(R_\infty) \leq \lambda_i(R^{(M)}) + \text{tr}(R_e).$$

(5.13)

Therefore, if $\text{tr}(R_e)$ is finite, it is possible to bound the eigenvalues of the correlation matrix of the infinite basis function expansion. Furthermore, if it can be shown that increasing the number of basis function included in $R^{(M)}$ can make $\text{tr}(R_e)$ arbitrarily small, then $\lambda_i(R^{(M)})$ converges to $\lambda_i(R_\infty)$ as the number of basis functions becomes large.

To bound the diagonal elements of $R_e$, consider now the expression

$$\sqrt{L_{a,x}}\text{sinc}[(L_{a,x}/\lambda) \cos \phi_a + n_x]$$

(5.14)

that forms part of the radiation pattern in (5.4). Defining $D_{a,x}(n_x)$ as

$$D_{a,x}(n_x) = \begin{cases} \sqrt{L_{a,x}} & |n_x| \leq \frac{L_{a,x}}{\lambda} + 1 \\ \sqrt{L_{a,x}} & |n_x| > \frac{L_{a,x}}{\lambda} + 1 \end{cases}$$

(5.15)
then for all \( n_x \),
\[
D_{a,x}(n_x) \geq \max_{\phi_a \in [0,2\pi]} \left| \sqrt{L_{a,x}} \text{sinc} \left( \frac{L_{a,x} \cos \phi_a + n_x}{\lambda} \right) \right|.
\]  
(5.16)

Using this definition along with a similar definition for \( D_{a,y}(n_y) \), it follows that for all \( n \) and for all \( \phi_a \),
\[
|e_{a,n}(\phi_a)| < D_{a,x}(n_x) D_{a,y}(n_y).
\]  
(5.17)

Since real propagation channels have finite gain, the PAS is uniformly bounded by
\[
\Gamma = \max_{\phi_a,\phi_b \in [0,2\pi]} B(\phi_b, \phi_a)
\]  
(5.18)

where \( \Gamma \) is a real, non-negative number. With these definitions, the variance of the channel gain from Alice’s \( n \)th basis function and Bob’s \( m \)th basis function is upper bounded by
\[
R_{(mn),(mn)} < 4\pi^2\Gamma D_{a,x}^2(n_x) D_{a,y}^2(n_y) D_{b,x}^2(m_x) D_{b,y}^2(m_y).
\]  
(5.19)

Computations of \( \text{tr}(R^{(M)}) \) and \( \text{tr}(R_e) \) involve sums of \( R_{(mn),(mn)} \), and we therefore explore the convergence of sums that, based on (5.19), help us to establish the convergence of these trace computations. First, defining \( \tau_{a,x} = \text{ceil}(L_{a,x}/\lambda) \), where \( \text{ceil}(x) \) is the smallest integer greater than \( x \), we can write
\[
\sum_{|n_x| > M} D_{a,x}^2(n_x) < 2L_{a,x} \int_{M}^{\infty} \frac{1}{(n_x - \tau_{a,x})^2} dn_x
\]  
\[
= \frac{2L_{a,x}}{M - \tau_{a,x}}.
\]  
(5.20)

Similarly, if \( \tau_{a,y} = \text{ceil}(L_{a,y}/\lambda) \), then
\[
\sum_{n_y = -\infty}^{\infty} D_{a,y}^2(n_y) < \sum_{n_y = -1 - \tau_{a,y}}^{1+\tau_{a,y}} D_{a,y}^2(n_y) + \sum_{|n_y| > 1+\tau_{a,y}} D_{a,y}^2(n_y)
\]  
\[
< 2L_{a,y}\tau_{a,y} + 3L_{a,y} + 2L_{a,y} \int_{1+\tau_{a,y}}^{\infty} \frac{1}{(n_y - \tau_{a,y})^2} dn_y
\]  
\[
= 2L_{a,y}\tau_{a,y} + 5L_{a,y}.
\]  
(5.22)
Let $R_{e,a}^{(M_x)}$ denote the covariance matrix for basis functions in Alice’s currents satisfying $|n_x| > M$. Using (5.19), (5.24), and (5.21) and letting $\varphi = 8\pi^2 L_{a,x} L_{a,y} L_{b,x} L_{b,y}$, we can write

$$\text{tr} \left( R_{e,a}^{(M_x)} \right) < \varphi \frac{(2\tau_{a,y} + 5)(2\tau_{b,x} + 5)(2\tau_{b,y} + 5)}{M - \tau_{a,x}}$$  \tag{5.25}$$

for $M > \tau_{a,x}$, which tends to zero as $1/M$. By the same argument, $\text{tr} \left( R_{e,a}^{(M_y)} \right)$, $\text{tr} \left( R_{e,b}^{(M_x)} \right)$, and $\text{tr} \left( R_{e,b}^{(M_y)} \right)$, which have similar definitions, also decay as $1/M$. Given our specification of $R^{(M)}$, every diagonal element of $R_e$ involves basis functions whose integer orders satisfy

$$\max \{|n_x|, |n_y|, |m_x|, |m_y|\} > M.$$  \tag{5.26}$$

Consequently, any diagonal element of $R_e$ will also appear as a diagonal element in one or more of $R_{e,a}^{(M_x)}$, $R_{e,a}^{(M_y)}$, $R_{e,b}^{(M_x)}$, $R_{e,b}^{(M_y)}$. Therefore

$$\text{tr} \left( R_e \right) < \text{tr} \left( R_{e,a}^{(M_x)} \right) + \text{tr} \left( R_{e,a}^{(M_y)} \right) + \text{tr} \left( R_{e,b}^{(M_x)} \right) + \text{tr} \left( R_{e,b}^{(M_y)} \right)$$

$$< \frac{\Phi}{M}$$  \tag{5.27}$$

for some constant $\Phi$ and for $M > \{\tau_{a,x}, \tau_{a,y}, \tau_{b,x}, \tau_{b,y}\}$, meaning that the trace of $R_e$ also decays as $1/M$. This means that $\text{tr} \left( R_e \right)$ and $\text{tr} \left( R_\infty \right)$ exist and $\lambda_i \left( R^{(M)} \right)$ converges to $\lambda_i \left( R_\infty \right)$ because

$$\lambda_i \left( R_\infty \right) - \lambda_i \left( R^{(M)} \right) < \frac{\Phi}{M}.$$  \tag{5.29}$$

for $M > \{\tau_{a,x}, \tau_{a,y}, \tau_{b,x}, \tau_{b,y}\}$. In other words, we have proven that the eigenvalues of the finite covariance matrix converge to those of the infinite dimensional covariance matrix as the number of basis functions becomes large.

### 5.2.2 Key Rate Convergence

We now must show that since the covariance eigenvalues converge as the number of basis functions increases, the resulting key rate converges as well. Since continuous functions map convergent sequences to convergent sequences, proving that the key rate expression is a continuous function of the covariance eigenvalues is sufficient to prove that the key rate...
converges. The proof of continuity must account for the fact that in the OSEA solution, the energy allocated to estimation of each mode changes as a function of the changing eigenvalues, also referred to as mode gains.

For a finite set of basis functions, the OSEA solution computes the optimal energy allocation for each possible number of active modes and then chooses the number of active modes that achieves the best performance. As the number of basis functions becomes large, the number of modes available for activation also increases. However, the lower bound on the required energy to activate a mode is given as

\[ p_i \geq \max_{j \in \{N_{\text{act}}\}} \frac{\sigma^2_j}{\Lambda_{jj}} \]  

(5.30)

where \( \{N_{\text{act}}\} \) is the set of indices of the active modes. This indicates that for finite total available energy and finite mode gain, the number of modes that can be activated by the OSEA solution is bounded. This means that even if an infinite number of modes is available, the OSEA solution can be computed by considering a finite set of possible values of \( N_{\text{act}} \). Therefore, if the key rate for each value of \( N_{\text{act}} \) is a continuous function of mode gain, then the maximum of the finite set of potentially optimal functions is also a continuous function of mode gain.

We are therefore left to prove that for a fixed set of active modes, the key rate from optimal energy allocation is a continuous function of mode gain. The optimal energy allocated to estimation of the \( i \)th mode is given by (3.24) subject to the constraint in (5.30). Suppose that we now increase the number of available modes leading to new eigenvalues \( \Lambda'_{ii} \), but we maintain the same number of active modes \( N_{\text{act}} \). Furthermore, let \( p'_i \) be computed from (3.24) using the new eigenvalues \( \Lambda'_{ii} \) but with the value of \( \alpha \) computed for the original solution with eigenvalues \( \Lambda_{ii} \). The total difference in energy allocated to probing is \( \Delta P = -\sum_{i=1}^{N_{\text{act}}} (p_i - p'_i) \). Since for fixed \( \alpha \) the expression for \( p_i \) in (3.24) is differentiable with respect to \( \Lambda_{ii} \) and therefore is a continuous function of \( \Lambda_{ii} \), then for every \( \epsilon > 0 \) there exists a \( \delta > 0 \) such that when \( \sum_{i=1}^{N_{\text{act}}} |\Lambda_{ii} - \Lambda'_{ii}| < \delta \) we have

\[ \sum_{i=1}^{N_{\text{act}}} |p_i - p'_i| < \frac{\epsilon}{2}. \]  

(5.31)
Now, let $p_i$ represent the optimal probing energy when $\alpha$ has been updated to enforce the total energy constraint $\sum_{i=1}^{N_{\text{act}}} p_i = P_T$ for the changed eigenvalues $\Lambda'_{ii}$. Because the total available energy $P_T$ remains constant throughout this analysis, $\sum_{i=1}^{N_{\text{act}}} (p_i - p'_i) = \sum_{i=1}^{N_{\text{act}}} (p_i - p'_i)$. However, since each value of $p'_i$ satisfies a Lagrange multiplier, any change in $\alpha$ to change the solution from $p'_i$ to $\overline{p}_i$ either adds energy to all active modes or removes energy from all active modes, depending on the sign of $\Delta P$, meaning that the sign of $p_i - p'_i$ is the same for all $i \leq N_{\text{act}}$. This means that

$$\sum_{i=1}^{N_{\text{act}}} |p_i - p'_i| = \sum_{i=1}^{N_{\text{act}}} |p_i - p'_i| \leq \sum_{i=1}^{N_{\text{act}}} |p_i - p'_i|.$$  (5.32)

The total change in energy across all modes is given as

$$\sum_{i=1}^{N_{\text{act}}} |p_i - \overline{p}_i| = \sum_{i=1}^{N_{\text{act}}} |p_i - p'_i + p'_i - \overline{p}_i| \leq \sum_{i=1}^{N_{\text{act}}} |p_i - p'_i| + |p'_i - \overline{p}_i| \leq \sum_{i=1}^{N_{\text{act}}} 2 |p_i - p'_i|$$  (5.33)

where the last inequality stems from (5.32). From (5.31) it follows that whenever

$$\sum_{i=1}^{N_{\text{act}}} |\Lambda_{ii} - \Lambda'_{ii}| < \delta$$  (5.34)

we have

$$\sum_{i=1}^{N_{\text{act}}} |p_i - p'_i| < \epsilon$$  (5.35)

meaning that the optimal energy allocations are a continuous function of mode gain. Therefore, the optimal energy $p_i$ allocated to estimation of the $i$th mode for a fixed number of active modes converges as the number of elements in the basis function expansion becomes large.

Since the actual mode gains $\Lambda_{ii}$ similarly converge, the elementwise product $p_i \Lambda_{ii}$ converges as the number of basis functions becomes large for fixed $N_{\text{act}}$. Also, the expression
for key rate in (3.17) is differentiable with respect to and therefore a continuous function of \( p_i \Lambda_{ii} \), meaning that the convergent sequence of \( p_i \Lambda_{ii} \) leads to convergence in the key rate for fixed \( N_{\text{act}} \). Since the OSEA solution simply chooses the maximum from a finite number of key rates each computed for a different number of active modes \( N_{\text{act}} \), the OSEA key rate from the finite basis function expansion converges as the number of Fourier basis functions becomes large.

### 5.3 Current Approximation

Demonstrating that the key rate converges as the number of basis functions becomes large does not guarantee that the key rates achieved with two different sets of basis functions are the same, or equivalently that the resulting key rate for the Fourier basis set is optimal. The objective of this section is to demonstrate that the spatial bound on the key rate upper bounds what is achievable for any finite set of square integrable currents satisfying a total energy constraint. To accomplish this, we consider the key rate observed for a finite set of square-integrable current functions at Alice and Bob, with each current function representing an antenna at each node. We first demonstrate that the eigenvalues of the covariance achieved using a finite basis expansion of the currents converge to the eigenvalues of the actual covariance as the number of basis functions becomes large. We then show that the transmission energy achieved using the basis expansion matches that of the actual currents. As these proofs require understanding of the convergence of the radiation pattern associated with each current distribution, we first study this convergence.

Let \( g_{a,\ell}(r_a) \) represent Alice’s current distribution from the \( \ell \)th current pair. Because of its convergence properties, we approximate this current distribution function as a finite Cesaro sum which is simply the arithmetic mean of partial Fourier series that converges to the Fourier series as the number of terms increases [48, p. 154]. For \( 2M + 1 \) basis functions, this series can be expressed as

\[
g_{a,\ell}^{(M)}(r_a) = \sum_{n_x, n_y = -M}^{M} \left( 1 - \frac{|n_x|}{M + 1} \right) \left( 1 - \frac{|n_y|}{M + 1} \right) c_{a,n}^{(\ell)} f_{a,n}(r_a) \tag{5.36}
\]
where the coefficients $c^{(\ell)}_{a,n}$ represent the standard Fourier series coefficients

$$c^{(\ell)}_{a,n} = \int_{V_a} f_{a,n}(r_a) g_{a,\ell}(r_a) \, dr_a.$$  \hfill (5.37)

Under this approximation, whenever $g_{a,\ell}(r_a)$ has a finite $L_1$ norm, we have [49, Lemma 3]

$$\lim_{M \to \infty} \int_{V_a} \left| g_{a,\ell}(r_a) - g_{a,\ell}^{(M)}(r_a) \right| \, dr_a = 0.$$  \hfill (5.38)

Since any function with a bounded $L_2$ norm on a compact set also has a bounded $L_1$ norm [50, p 53], it follows that (5.38) holds for square integrable $g_{a,\ell}(r_a)$.

Convergence of the Fourier representation of the current distribution also leads to convergence of its associated radiation pattern. Let $s_{a,\ell}(\phi_a)$ and $s_{a,\ell}^{(M)}(\phi_a)$ respectively represent the radiation patterns for the current distributions $g_{a,\ell}(r_a)$ and $g_{a,\ell}^{(M)}(r_a)$. The $L_1$ norm of the difference between these two radiation patterns is

$$\int_{\phi_a} \left| s_{a,\ell}(\phi_a) - s_{a,\ell}^{(M)}(\phi_a) \right| \, d\phi_a = \int_{\phi_a} \left| \int_{V_a} G(\phi_a, r'_a) \left[ g_{a,\ell}(r'_a) - g_{a,\ell}^{(M)}(r'_a) \right] \, dr'_a \right| \, d\phi_a$$

$$\leq \int_{\phi_a} \int_{V_a} \left| G(\phi_a, r'_a) \left[ g_{a,\ell}(r'_a) - g_{a,\ell}^{(M)}(r'_a) \right] \right| \, dr'_a \, d\phi_a$$

$$\leq 2\pi \int_{V_a} \left| g_{a,\ell}(r'_a) - g_{a,\ell}^{(M)}(r'_a) \right| \, dr'_a$$

where we have used the fact that $|G(\phi_a, r'_a)| = 1$. The result of (5.38) means that the $L_1$ norm of the difference between radiation patterns also goes to zero as $M \to \infty$. This fact will be used to prove convergence of the covariance matrix elements as well as the energy radiated by the current distributions.

Finally, the radiation pattern for the $q$th current distribution can be upperbounded as

$$s_{a,\ell}(\phi_a) \leq \int_{V_a} |G(\phi_a, r'_a)| \left| g_{a,\ell}(r'_a) \right| \, dr'_a$$

$$\leq \int_{V_a} \left| g_{a,\ell}(r'_a) \right| \, dr'_a.$$  \hfill (5.42)
Since the $L_1$ norm of $g_{a,\ell}(r'_a)$ is assumed finite, the radiation patterns associated with the current distributions are uniformly bounded.

### 5.3.1 Covariance Matrix Eigenvalues

Let $X$ and $X^{(M)}$ represent the covariance matrix observed with the original current distributions and the finite Fourier approximation, respectively. We can compute $X_{(mn),(uv)}$ using (5.8) with the basis function radiation pattern $e_{\xi,\zeta}(\phi_\xi)$ replaced with the pattern $s_{\xi,\zeta}(\phi_\xi)$ for $\xi \in [a, b]$ and $\zeta \in [m, n, u, v]$. We similarly define

$$X^{(M)}_{(\tilde{m}n),(\tilde{u}v)} = \int \int s^{(M)}_{b,m}(\phi_b) s^{(M)*}_{b,u}(\phi_b) B(\phi_b, \phi_a) s^{(M)}_{a,n}(\phi_a) s^{(M)*}_{a,v}(\phi_a) \ d\phi_a d\phi_b$$

(5.44)

where the subscript $\tilde{m}$ indicates that the radiation pattern inside the integral is the approximate pattern $s^{(M)}_{b,m}(\phi_b)$ rather than the actual pattern $s_{b,m}(\phi_b)$. We can then express the difference between the covariance elements for the actual and approximate patterns as

$$\left| X_{(mn),(uv)} - X^{(M)}_{(\tilde{m}n),(\tilde{u}v)} \right| \leq \left| X_{(mn),(uv)} - X^{(M)}_{(\tilde{m}n),(\tilde{u}v)} \right| + \kappa \int \left| s_{b,m}(\phi_b) - s^{(M)}_{b,m}(\phi_b) \right| d\phi_b$$

(5.45)

which arranges the expression so that the two covariance matrix elements within each term of the form $|\mu - \nu|$ on the right hand side differ only by use of one radiation pattern in the integration. Each of these differences can be bounded by an expression similar to

$$\left| X_{(mn),(uv)} - X^{(M)}_{(\tilde{m}n),(\tilde{u}v)} \right| < \kappa \int \left| s_{b,m}(\phi_b) - s^{(M)}_{b,m}(\phi_b) \right| d\phi_b$$

(5.46)

where

$$\kappa = \max_{\phi_a} \left| \int s^{*}_{b,u}(\phi_a) B(\phi_b, \phi_a) s_{a,n}(\phi_a) s^{*}_{a,v}(\phi_b) \ d\phi_a \right|.$$ 

(5.47)

Since $B(\phi_b, \phi_a)$ is bounded through (5.18) and $s^{*}_{b,u}(\phi_b)$, $s_{a,n}(\phi_a)$, and $s^{*}_{a,v}(\phi_a)$ are uniformly bounded through (5.43) for all values of $\phi_a$ and $\phi_b$, $\kappa$ is a finite constant. Equation (5.41) coupled with (5.46) therefore means that the first term on the right hand side of (5.45) goes
to zero as $M$ increases. Since the same analysis applies to each of the terms on the right hand side of (5.45), the approximation $X^{(M)}_{(mn), (\tilde{m}\tilde{n})}$ converges to $X_{(mn), (uv)}$ as $M$ becomes large.

While convergence of the covariance matrix elements suggests convergence of the covariance matrix eigenvalues, a simple proof demonstrates this latter observation. Let $X = UA_XU^\dagger$ be the eigenvector decomposition of $X$ where $U$ is the unitary matrix of eigenvectors and $A_X$ is the diagonal matrix of real, non-negative eigenvalues of $X$ arranged in decreasing order. We consider the matrix $U^\dagger X^{(M)} U$ that has the same eigenvalues as $X^{(M)}$. This matrix can be arranged as

$$U^\dagger X^{(M)} U = \Lambda_X + U^\dagger (X - X^{(M)}) U. \quad (5.48)$$

Because $X$ and $X^{(M)}$ are of finite dimensions and since $X^{(M)}$ converges to $X$ on an elementwise basis, it follows that for every $\epsilon > 0$ there exists a value $M'$ such that if $M > M'$ then for every $m, n, u, v$

$$\left| X_{(mn), (uv)} - X^{(M)}_{(\tilde{m}\tilde{n}), (\tilde{u}\tilde{v})} \right| < \frac{\epsilon}{L^3} \quad (5.49)$$

where $L$ is the total number of current distribution pairs. If $\|\cdot\|_{\text{max}}$, $\|\cdot\|_2$, and $\|\cdot\|_F$ respectively represent the maximum (maximum element), $L_2$, and Frobenius norm of the matrix argument, then for matrices $C$ and $D$ we have [51, p. 56], [52, p. 279]

$$\|C\|_{\text{max}} \leq \|C\|_2 \leq \|C\|_F, \quad (5.50)$$

$$\|CD\|_F \leq \|C\|_F \|D\|_F \quad (5.51)$$

and $\|C\|_F^2 = \text{tr} (C^\dagger C)$. Using these expressions, it is possible to upper bound the maximum element of $\Delta^{(M)}$ as

$$\|\Delta^{(M)}\|_{\text{max}} \leq \|U^\dagger\|_F \|X - X^{(M)}\|_F \|U\|_F$$

$$\leq L \|X - X^{(M)}\|_F \quad (5.52)$$

$$< \frac{\epsilon}{L}. \quad (5.53)$$
Without loss of generality, let $\epsilon$ be chosen so that for all $k, \hat{k}$ if $\lambda_k(X) \neq \lambda_{\hat{k}}(X)$ then $|\lambda_k(X) - \lambda_{\hat{k}}(X)| > 2\epsilon$. Further, define a set of closed disks in the complex plane with the $k$th disk $T_k$ defined as

$$
T_k(X^{(M)}) = \left\{ x \in \mathbb{C} : \left| \lambda_k(X) + \Delta^{(M)}_{kk} - x \right| \leq \frac{(L-1)\epsilon}{L} \right\}.
$$

(5.55)

The Gershgorin circle theorem states that all of the eigenvalues of $X^{(M)}$ lie in the union of these disks and that if there are $K$ disks that are disjoint from all other disks then the union of those disks contains $K$ eigenvalues [53, p. 325]. With $\left| \Delta^{(M)}_{kk} \right| < \epsilon/L$ for all $k$, each $T_k$ can be contained in the disk $T'_k$ defined as

$$
T'_k(X^{(M)}) = \left\{ x \in \mathbb{C} : |\lambda_k(X) - x| \leq \epsilon \right\}.
$$

(5.56)

Having chosen $\epsilon$ so that the difference between any non-repeated eigenvalues is greater than $2\epsilon$ and since each circle is of radius $\epsilon$, for all $k$ and $\hat{k}$, if $T'_k(X^{(M)}) \neq T'_\hat{k}(X^{(M)})$ then $T'_k(X^{(M)})$ and $T'_\hat{k}(X^{(M)})$ are disjoint. This means that each $\lambda_k(X^{(M)})$ is contained in a circle of radius $\epsilon$ centered at $\lambda_k(X)$ which indicates that $|\lambda_k(X) - \lambda_k(X^{(M)})| < \epsilon$ or that the eigenvalues of $X^{(M)}$ converge to the eigenvalues of $X$ as the number of basis functions becomes large. In Section 5.2 we show that when the eigenvalues of a covariance matrix converge to a set of limiting values, the key rate converges to the key rate of those limiting eigenvalues.

### 5.3.2 Energy

This work considers the constrained energy as either $L_2$ norm of the currents that radiate the signal or the $L_2$ norm of the radiation pattern, the latter of which actually represents the radiated energy. We therefore examine the convergence behavior of the $L_2$ norm for both of these quantities. The difference between the $L_2$ norms of the actual and
approximate radiation patterns is bounded as
\[
\int \left| s_{a,\ell}(\phi_a) \right|^2 - \left| s_{a,\ell}^{(M)}(\phi_a) \right|^2 d\phi_a = \int \left( \left| s_{a,\ell}(\phi_a) \right| + \left| s_{a,\ell}^{(M)}(\phi_a) \right| \right) \left( \left| s_{a,\ell}(\phi_a) \right| - \left| s_{a,\ell}^{(M)}(\phi_a) \right| \right) d\phi_a
\]
\[= \Xi \int \left( \left| s_{a,\ell}(\phi_a) \right| - \left| s_{a,\ell}^{(M)}(\phi_a) \right| \right) d\phi_a \]
\[\leq \Xi \int \left| s_{a,\ell}(\phi_a) - s_{a,\ell}^{(M)}(\phi_a) \right| d\phi_a \]
\[\text{(5.57)}
\]
\[\leq \Xi \int \left( \left| s_{a,\ell}(\phi_a) \right| + \left| s_{a,\ell}^{(M)}(\phi_a) \right| \right) \left( \left| s_{a,\ell}(\phi_a) \right| - \left| s_{a,\ell}^{(M)}(\phi_a) \right| \right) d\phi_a
\]
\[\leq \Xi \int \left( \left| s_{a,\ell}(\phi_a) \right| + \left| s_{a,\ell}^{(M)}(\phi_a) \right| \right) d\phi_a \]
\[= \Xi \int \left( \left| s_{a,\ell}(\phi_a) \right| + \left| s_{a,\ell}^{(M)}(\phi_a) \right| \right) d\phi_a \]
\[\text{(5.58)}
\]
\[\text{where}
\]
\[\Xi = \max_{\phi_a} \left( \left| s_{a,\ell}(\phi_a) \right| + \left| s_{a,\ell}^{(M)}(\phi_a) \right| \right). \]
\[\text{(5.60)}
\]
Equation (5.41) provides that as \( M \) gets large that (5.59) goes to zero since \( \Xi \) is guaranteed finite from (5.43). This means that the energy radiated by the approximation to the current converges to that transmitted by the actual current distribution.

If we more loosely define the energy as the \( L_2 \) norm of the current distribution, then we can use Parseval’s theorem to demonstrate convergence. Specifically, for the approximate current distribution we have
\[
\int_{V_a} \left| g_{a,\ell}^{(M)}(r_a) \right|^2 \, dr_a = \sum_{n_x, n_y = -M}^M \left( 1 - \frac{|n_x|}{M + 1} \right)^2 \left( 1 - \frac{|n_y|}{M + 1} \right)^2 \left| c_{a,n}^{(\ell)} \right|^2 \]
\[\text{(5.61)}
\]
\[\leq \sum_{n_x, n_y = -M}^M \left( 1 - \frac{|n_x|}{M + 1} \right)^2 \left( 1 - \frac{|n_y|}{M + 1} \right)^2 \left| c_{a,n}^{(\ell)} \right|^2 \]
\[\text{(5.62)}
\]
Similarly, for the original current distribution, Parseval’s theorem gives
\[
\int_{V_a} \left| g_{a,\ell}(r_a) \right|^2 \, dr_a = \sum_{n_x, n_y} \left| c_{a,n}^{(\ell)} \right|^2.
\]
\[\text{(5.63)}
\]
By inspection,
\[
\sum_{n_x, n_y = -M}^M \left( 1 - \frac{|n_x|}{M + 1} \right)^2 \left( 1 - \frac{|n_y|}{M + 1} \right)^2 \left| c_{a,n}^{(\ell)} \right|^2 < \sum_{n_x, n_y} \left| c_{a,n}^{(\ell)} \right|^2
\]
for finite $M$ which means that

\[
\int_{V_a} |g_a^{(M)}(r_a)|^2 \, d\mathbf{r}_a \leq \int_{V_a} |g_a,\ell(r_a)|^2 \, d\mathbf{r}_a.
\]

(5.64)

Equality in (5.64) applies as $M$ gets large, which means that the energy represented by the approximate current distribution converges to that of the actual current distribution as $M$ gets large.

This analysis demonstrates that the key rate achieved and the energy represented by the complex exponential approximation converges to the respective quantities for the original current function as the number of basis functions becomes large. Since any arbitrary weighted combinations of basis functions are guaranteed to not outperform OSEA applied to the corresponding basis function expansion, it follows that no set of square integrable current distributions that satisfy a common energy constraint will outperform the OSEA solution for the complex exponential basis function expansion as the number of included basis functions becomes large. This means that the limiting key rate for OSEA applied to a finite Fourier expansion is a spatial bound on the key rate that can be achieved with reciprocal beamformer probing.

5.4 Modified OSEA

Our definition of $\mathbf{H}$ relates the open-circuit voltage at the receiver to the current driving the transmitter with all other transmit elements terminated in an open circuit. The signal $x_i$ in our model therefore represents driving current, and our constraint $\sum_{i=1}^{N_{\text{act}}} x_i^2 = P_T$ actually constrains the currents rather than the radiated energy. If the antenna elements are widely separated so that the mutual impedance is small, then the energy radiated is proportional to the square of the current, making this constraint reasonable. While this assumption is widely adopted in the signal processing literature [54], it is not valid for closely spaced array elements or for the case studied here where the antennas are abstracted as currents that can coexist within a single aperture.

Similarly, the original OSEA solution assumes that the estimation errors $\eta_a$ and $\eta_b$ at Alice and Bob are spatially white (covariance matrix being a scaled identity), which
is consistent with the noise creating the error arising due to the radio frequency receiver. However, since radio frequency front-end noise can be reduced through careful design and since the performance of most communication systems is limited by interference, we choose to consider interference, which will generally result in spatially correlated estimation errors, in this analysis. It is necessary for us to modify the formulation of OSEA to accommodate these changes. For consistency, we consider currents that vary in the $x$-$y$ plane, although the derivation extends to three dimensions.

### 5.4.1 Spatially Correlated Estimation Error

With a beamforming system, let the vector of open-circuit noise voltages observed on Bob’s array during the $i$th transmission be denoted $\nu_{b,i}$ so that the estimation error after receive beamforming is $\eta_{b,i} = \mathbf{b}_i^T \nu_{b,i}$ where $\eta_{b,i}$ is the $i$ element of $\eta_b$. Let the interference signal as a function of angle during the $i$th transmission be represented as $\psi_i(\phi_b)$. The resulting open-circuit noise voltage observed by the $m$th array element is

$$\nu_{b,im} = \int e_{b,m}(\phi_b) \psi_i(\phi_b) \, d\phi_b. \quad (5.65)$$

Consistent with our signal model, we further assume that $\mathbb{E}\{\psi_i(\phi_b)\psi_i^*(\phi'_b)\} = \sigma_I^2 B_I(\phi_b) \delta(\phi_b - \phi'_b)$ where $B_I(\phi_b)$ is the PAS of the interference. The error covariance matrix $K_b$ then has elements

$$K_{b,nu} = \mathbb{E}\{\nu_{b,im} \nu_{b,iu}^*\} \quad (5.66)$$

$$= \int e_{b,m}(\phi_b) B_I(\phi_b) e_{b,u}^*(\phi_b) \, d\phi_b \quad (5.67)$$

where we have used that the expectation is over the transmission index $i$ which is a discrete representation of time. If each antenna also contributes a thermal noise voltage that is modeled as a zero-mean complex Gaussian random variable with variance $\sigma_{\text{Loss}}^2$ and assuming that the thermal noise on each element is independent of that on the other elements as well as of the interference, we can write that the total error covariance is $\bar{K}_b = K_b + \sigma_{\text{Loss}}^2 \mathbf{I}$. 

73
To illustrate the impact of this spatially correlated estimation error, we reformulate
the system model from Chapter 3. Specifically, we write the vector of open-circuit voltages
on Bob’s antennas (before application of the receive beamformer) as
\[
y_{b,i} = H_{i} x_{i} + \nu_{b,i}.
\] (5.68)

To be able to apply OSEA to this system, the receiver must first whiten the error estimate,
or
\[
\tilde{y}_{b,i} = \tilde{K}_{b}^{-1/2} y_{b,i} = \tilde{K}_{b}^{-1/2} H_{i} x_{i} + \tilde{K}_{b}^{-1/2} \nu_{b,i}
\] (5.69)

where \( \tilde{K}_{b}^{1/2} \) is the matrix square root defined such that \( \tilde{K}_{b}^{1/2} (\tilde{K}_{b}^{1/2})^\dagger = \tilde{K}_{b} \). With this
representation, \( \tilde{\nu}_{b,i} \) has covariance \( I \). If we now apply the receive beamformer \( b_i \), we obtain
\[
\tilde{w}_{b,i} = b_i^T \tilde{K}_{b}^{-1/2} H_{i} x_{i} + b_i^T \tilde{\nu}_{b,i}.
\] (5.70)

The actual receive beamformer is therefore the combination of the whitening operation and
the post-whitened beamformer, or \( b_i^T = b_i^T \tilde{K}_{b}^{-1/2} \).

5.4.2 Radiated Energy Constraint

For transmission with \( a_i x_i \), the far field radiation pattern of the array is given by
\[
z_{a,i} (\phi_a) = \sum_{n=1}^{N_a} a_{i,n} x_i e_{a,n} (\phi_a).
\] (5.71)

The radiated energy \( p_{a,i} \) associated with \( z_{a,i} (\phi_a) \) is
\[
p_{a,i} = \frac{1}{Z_0} \int_{\Theta_{a,nv}} z_{a,i}^* (\phi_a) z_{a,i} (\phi_a) \ d\phi_a
\] (5.72)
\[
= x_i^2 \sum_{n=1}^{N_a} \sum_{v=1}^{N_a} a_{i,n}^* a_{i,v} \left( \frac{1}{Z_0} \int_{\Theta_{a,nv}} e_{a,n}^* (\phi_a) e_{a,v} (\phi_a) \ d\phi_a \right) a_{i,v}
\] (5.73)
\[
= x_i^2 a_i^\dagger \Theta_a a_i
\] (5.74)
where $Z_0$ is the free-space wave impedance. The total energy radiated by Alice over all training transmissions becomes

$$P_{a,\text{rad}} = \sum_{i=1}^{N_{\text{act}}} x_i^2 a_i^\dagger \Theta_a a_i.$$  \hfill (5.75)

For the total radiated energy to be proportional to the square of current for all $a_i$, we must have $\Theta_a \propto I$, which with reference to (5.73) means that the radiation pattern of each element in Alice’s array is normalized and orthogonal to the radiation patterns of all other array elements. Clearly this will not hold for an arbitrary array, although it is possible to reformulate the problem to constrain the radiated energy rather than the currents.

We recognize from (5.73) that $\Theta_a$ represents the mutual resistance matrix for Alice’s array. If each antenna element has the same resistive loss $R_{\text{Loss}}$, then we can define $\tilde{\Theta}_a = \Theta_a + R_{\text{Loss}} I$ as an effective resistance matrix for the array. The energy delivered to the array for the $i$th excitation is then $\tilde{p}_{a,i} = a_i^\dagger \tilde{\Theta}_a a_i$. We define the beamforming weights of a virtual array

$$\tilde{a}_i = \tilde{\Theta}_a^{1/2} a_i$$  \hfill (5.76)

such that $\tilde{p}_{a,i} = x_i^2 \tilde{a}_i^\dagger \tilde{a}_i$. Our signal model therefore becomes

$$\tilde{w}_{b,i} = \tilde{b}_i^T \tilde{H} \tilde{a}_i x_i + \tilde{\eta}_{b,i}.$$  \hfill (5.77)

### 5.4.3 OSEA Application

From (5.77), we can form the virtual channel $\tilde{H} = \tilde{K}_b^{-1/2} \tilde{H} \tilde{\Theta}_a^{-1/2}$ and write

$$\tilde{w}_{b,i} = \tilde{b}_i^T \tilde{H} \tilde{a}_i x_i + \tilde{\eta}_{b,i}.$$  \hfill (5.78)

$$= \left( \tilde{a}_i \otimes \tilde{b}_i \right)^T \tilde{h} + \tilde{\eta}_{b,i}.$$  \hfill (5.79)

where $\tilde{h}$ represents $\tilde{H}$ stacked columnwise into a vector. If we apply OSEA to this equation, the beamforming vectors $\tilde{b}_i$ will be unitary, and therefore $\mathbb{E} \left\{ \tilde{\eta}_b \tilde{\eta}_b^\dagger \right\} = I$, consistent with the original OSEA formulation with $\sigma_0^2 = 1$. Similarly, the energy delivered to the array for the $i$th excitation becomes $\tilde{p}_{a,i} = x_i^2$, which is also consistent with the original OSEA
formulation. Of course, after application of OSEA, we construct the beamformers for the actual arrays as \( a_i = \tilde{\Theta}_a^{-1/2} \tilde{a}_i \) and \( b_i^T = \tilde{b}_i^T \tilde{K}_b^{-1/2} \).

The form of \( \tilde{H} \) demonstrates our motivation for including loss resistance and antenna thermal noise in the formulation. Specifically, for tightly coupled antennas (as might occur for overlapping currents), small eigenvalues of \( \Theta_a \) and \( K_b \) can lead to problems when the inverse is taken, problems that physically represent impractical supergain beamformer weightings [55]. Including loss and thermal noise physically removes the possibility of these supergain solutions, which mathematically represents itself as a regularization of the inverse [55].

One limitation of this formulation is that it can destroy the reciprocity of the channel that is required for key establishment. Specifically, because the effective channel is transformed, the transformation must be identical for transmission from Alice to Bob and for transmission from Bob to Alice. Furthermore, the effective beamformers used by Alice and Bob must be identical for channel estimation in both directions. For this to be the case, we must have \( \tilde{\Theta}_a \propto \tilde{K}_b \) and \( \tilde{\Theta}_b \propto \tilde{K}_a \).

To resolve this issue, we note that if \( B_1(\phi_b) = 1 \), then \( K_b = \sigma_l^2 Z_0 \Theta_b \) where \( \Theta_b \) is obtained from (5.73) simply by changing the subscript a to b. Physically, this means that the interference must be assumed to arrive from all directions. Furthermore, if we set \( \sigma_{Loss}^2 = \sigma_l^2 Z_0 R_{Loss} \), then our requirement for reciprocal channel estimation is satisfied. We assume these conditions are satisfied in the remainder of this work.

### 5.5 Finite Basis Function Expansion

The proof of convergence in Section 5.2 demonstrates that as the number of basis functions becomes large, the eigenvalues of the correlation matrix converge to a limiting value. While this is important mathematically, we must remember that application of OSEA to currents represented by the basis functions requires computing the eigenvalues of matrices whose dimensions scale as the product of the number of basis used at Alice and Bob. Therefore, practical considerations require us to efficiently limit the number of basis functions used. Naturally, this could be done by identifying the number of basis functions required
eigenvalue convergence to within a specified tolerance, but experience shows that this results in a large number of basis functions and high computational complexity.

We therefore resort to a more practical two-stage method for determining the number of basis functions to include. In the first stage, we recognize that as the frequency of oscillation of the basis function representing the current increases, the radiated power associated with that current decreases (for a fixed current peak magnitude), as such currents represent supergain excitation. The power radiated by Alice’s $n$th individual basis function is specified by $\Theta_{a,nn}$, and therefore we can limit the range of Fourier basis functions to be those that satisfy $\Theta_{a,nn} > 1/Q_1$, where $Q_1$ represents a threshold (an identical thresholding is used for Bob).

With the number of basis functions limited, we can construct the matrix $\Theta_a$ and compute its eigendecomposition. Given our regularization to form $\tilde{\Theta}_a$, it is clear that if the $v$th eigenvalue $\Lambda_{a,vv}$ of $\Theta_a$ is very small, the radiation associated with that excitation will be small compared to the loss represented by the loss resistance $R_{\text{Loss}}$. Therefore, we can limit the number of virtual array elements by constructing $\Theta_a$ using only those eigenvalues that satisfy $\Lambda_{a,vv} > 1/Q_2$, where again $Q_2$ represents a threshold. We emphasize that $\Theta_a$ has been computed from, and therefore includes the influence of, the large set of basis functions. The dimensionality reduction is completed in the eigenspace to maintain computational efficiency. Simulations demonstrating the impact of this dimension reduction are provided in Section 5.6.3.

5.6 Results

We now use simulations to illuminate some fundamental behaviors of the key rate spatial bound. The simulated channels are defined using the model described in Section 3.3. The optimal antennas considered are defined over a square of side length $L$ at both Alice and Bob. Unless otherwise specified the aperture dimension is $L = 1\lambda$, with the virtual elements for that area defined with an energy constraint. The value of $R_{\text{Loss}}$ in each is chosen so that an antenna with radiation pattern $e(\phi) = 1$ radiates 95% of the energy delivered to the antenna. Furthermore, unless otherwise specified, the thresholds are chosen as $Q_1 = Q_2 = 40,000$. 
5.6.1 Constrained Energy vs. Current

As a first case, we compare the key rate $I_k$ achievable when the energy is constrained to that achievable when the current is constrained. Figure 5.0 plots the performance for both constraints as a function of aperture size for different values of the signal-to-noise ratio (SNR) $P_T/\sigma_0^2$. As expected, the performance increases with both array size and SNR, although the dependence on aperture size is more pronounced under the current constraint. This is because the energy radiated from the far field patterns in (5.0) is scaled by the term $L^2$ which means increased radiated energy as the aperture size increases.

5.6.2 Pulse Basis Functions

The key rate spatial bound represents the upper limit on performance for antennas occupying the same volume (or area). Therefore, it is interesting to compare the performance of the spatial bound based on Fourier basis function to the key rate for pulse basis function. If we have $M$ such basis functions per dimension, the $n$th basis function has a magnitude...
Figure 5.2: Ratio of the key rate achieved using OSEA for an $M \times M$ array of square current pulses and the key rate spatial bound as a function of aperture size for an SNR of 5 dB.

For both the Fourier and pulse basis expansions, the radiated energy constraint is used to compute the performance.

Figure 5.1 plots the ratio of the key rate achieved from OSEA using pulse basis functions ($I_{k,\text{PUL}}$) to that achieved using Fourier basis functions ($I_{k,\text{FOU}}$) as a function of aperture dimension for an SNR of 5 dB and three values of $M$. For all values of $M$, the key rate for the pulse basis functions is close to the bound for small apertures. However, as the aperture dimensions increase, the relative performance degrades because the relatively small number of pulse functions are unable to fully exploit the spatial modes available in the propagation environment.
Figure 5.3: The OSEA key rate of an $M \times M$ dipole array modeled by square current pulses relative to key rate spatial bound as a function of SNR.

Figure 5.2 plots the same performance measure as a function of SNR for several different values of $M$ for an aperture dimension of $L = 1\lambda$. For each value of $M$, the relative performance increases with SNR until $P_T/\sigma_0^2 = 1$ after which the relative performance levels off. Figure 5.3 plots the number of active modes used by the OSEA solution as a function of SNR for the grid of pulse functions as well as for the key rate spatial bound. This result shows that in the region where performance in Fig. 5.2 is most sensitive to SNR, there is only one active mode. Manipulation of the key rate expression in (3.17) shows that when $N_{\text{act}} = 1$ and $s_1 = P_T A_{11}/\sigma_0^2$ is small, the key rate grows as $\log_2(1 + s_1^2)$, which is strongly dependent on the SNR. As SNR increases, the multiple terms used in the sum of (3.17) grow more slowly with SNR, resulting in the behavior observed in Fig. 5.2.

5.6.3 Simulated Convergence

While the convergence of the key rate spatial bound has been analytically proven, this proof does not demonstrate the convergence behavior of the algorithm as the number of basis
functions grows. This is particularly relevant given our discussion in Section 5.5 regarding reduction of computational complexity by limiting the number of basis functions used in the expansion. We recall from that discussion that our practical approach to computational complexity reduction involves the thresholds $Q_1$ and $Q_2$, and therefore this study focuses on the impact of these threshold values. For purposes of this analysis, the thresholds are swept between the values $Q_{\text{min}} = 100$ and $Q_{\text{max}} = 1,000,000$. In the following simulations, key rate convergence is quantified as the relative difference between the key rate $I_{k,Q_{\text{max}}}$ obtained when $Q_1 = Q_2 = Q_{\text{max}}$ and the key rate $I_{k,Q}$ achieved with the designated values of $Q_1$ and $Q_2$.

Figure 5.4 plots the key rate convergence for $Q_1 = Q_2 = Q$ for several different values of SNR. These results show that there is only a 0.5% improvement observed when increasing $Q$ from 10,000 to 1,000,000, suggesting that the performance achieved with any threshold in this interval represents a good approximation to the performance achieved with a much larger value of $Q$. 

Figure 5.4: The number of active modes obtained from the OSEA solution for an $M \times M$ array of pulse functions as well as for the key rate spatial bound as a function of SNR.
Figure 5.5: Normalized difference between the key rate $I_{k,Q_{\text{max}}}$ achieved using $Q_1 = Q_2 = Q_{\text{max}}$ and the key rate $I_{k,Q}$ obtained using $Q_1 = Q_2 = Q$.

Figure 5.5 plots the convergence for an SNR of 5 dB when either $Q_1$ or $Q_2$ is swept while the other is fixed at $Q_{\text{max}}$. When $Q_1$ is fixed, there is negligible change in performance as a function of $Q_2$. Conversely, the convergence when only $Q_1$ changes is almost identical to that observed when both $Q_1$ and $Q_2$ change. These observations can be better understood by exploring the number of basis functions and virtual elements (see discussion in Section 5.5) used for the different thresholds. Figure 5.6 plots the number of Fourier basis functions that satisfy the condition specified for the $Q_1$ threshold, while Fig. 5.7 shows the number of virtual elements that satisfy the constraints for the thresholds $Q_1$ and $Q_2$, where once again when one is swept the other is fixed at $Q_{\text{max}}$. Figure 5.6 shows that for $Q_2 = Q_{\text{max}}$, the 3% improvement achieved over the range $Q_{\text{min}} \leq Q_1 \leq Q_{\text{max}}$ requires increasing the number of Fourier basis functions from 25 to 4261. On the other hand, Fig. 5.7 reveals that increasing $Q_2$ when $Q_1 = Q_{\text{max}}$ only results in addition of two virtual elements that have $\Lambda_{a,v} \approx 1/5000$, which is very small. These low quality excitations have negligible impact on system performance.
Figure 5.6: Normalized difference between the key rate $I_{k,Q_{\text{max}}}$ achieved using $Q_1 = Q_2 = Q_{\text{max}}$ and the key rate $I_{k,Q}$ obtained when sweeping $Q_1$ or $Q_2$ (or both) at an SNR of 5 dB.

5.7 Chapter Summary

This chapter applies the optimal beamformer channel estimation technique for secret key establishment from Chapter 3 to beamforming arrays with elements defined by arbitrary current distributions represented by an expansion using Fourier basis functions. The optimal channel estimation scheme upper bounds the key rate achievable with any square-integrable current distribution, and the key rate achieved with the basis function expansion converges as the number of basis functions becomes large. This upper limit, referred to as the key rate spatial bound, serves as a number against which the performance of practical arrays can be compared. Simulations demonstrate the numerical convergence properties of the algorithm and illustrate key behaviors of the key rate spatial bound in practical channels.
Figure 5.7: The number of basis functions satisfying the constraint $\Theta_{a,nn} > 1/Q_1$ when $Q_2 = Q_{\text{max}}$.

Figure 5.8: The number of virtual elements satisfying the constraints $\Theta_{a,nn} > 1/Q_1$ and $\Lambda_{a,\nu\nu} > 1/Q_2$ thresholds.
Chapter 6

Improved Channel Estimation Resource Allocation for MIMO Systems

The existence of an optimal strategy for beamformer channel probing and the corresponding key rate increases raise a question of whether similar results can be obtained when probing with MIMO systems. In reformulating the problem to account for the enhanced capabilities in MIMO systems, our work to date has been unable to define a globally optimal MIMO channel probing strategy. Instead we propose an iterative technique that determines an optimized allocation of energy used for each transmission in the channel estimation process. Numerical simulations are used to explore the convergence properties of the iterative algorithm as well as the impact of different system characteristics such as array size, signal-to-noise ratio (SNR), and propagation characteristics on the achieved key rate. In all cases, the key rate achieved using the algorithm is compared to that obtained when the energy is equally allocated to estimation of all channel dimensions. The results show that at low SNR, the proposed channel estimation approach provides dramatic improvement in the achieved key rate. However as SNR becomes large, the performance of the proposed algorithm degrades relative to that achieved using equal allocation. This is most pronounced in a Ricean propagation environment where equal allocation outperforms the iterative optimization by a small margin at high SNR.

6.1 Channel Estimation

6.1.1 System Model

With the receiving nodes in the system under consideration capable of estimating multiple modes simultaneously, we are required to carefully redefine how each node estimates each element of the channel response vector. As before, Alice’s and Bob’s arrays consist of
Figure 6.1: System diagram showing Alice transmitting the vector $\mathbf{a}_n x_{a,n}$ from her $N_a$ antennas and Bob applying the beamforming matrix $\mathbf{B}$ to the received signal to produce $N_b$ channel coefficient estimates $\hat{w}_{b,mn}$ for $1 \leq m \leq N_b$.

$N_a$ and $N_b$ radiating elements, respectively. Suppose that Alice transmits the $N_a \times N_a$ matrix $\mathbf{A} \mathbf{X}_a$ to Bob, where $\mathbf{A}$ is a unitary matrix with $n$th column $\mathbf{a}_n$, as depicted in Fig. 6.0. The diagonal matrix $\mathbf{X}_a$ is real with $n$th diagonal element $x_{a,n} = \sqrt{p_{a,n}}$, where $p_{a,n}$ represents the energy allocated to transmission of vector $\mathbf{a}_n$. After receiving the $n$th transmission, Bob applies a set of weighting vectors represented by the $N_b \times N_b$ matrix $\mathbf{B}/x_{a,n}$, where $\mathbf{B}$ is a unitary matrix with $m$th column $\mathbf{b}_m$. We refer to $\mathbf{A}$ and $\mathbf{B}$ as beamformer matrices and their column vectors as beamformers.

Let $\mathbf{H}$ represent the $N_b \times N_a$ matrix of narrow band channel coefficients between the elements of the two arrays. Application of the $m$th receive beamformer to the received signal for the $n$th transmit beamformer leads to the result

$$\hat{w}_{b,mn} = \mathbf{b}_m^T \mathbf{H} x_{a,n} + \frac{1}{x_{a,n}} \mathbf{b}_m^T \mathbf{\nu}_{b,n}$$

(6.1)

where $\{\cdot\}^T$ is a transpose. The vector $\mathbf{\nu}_{b,n}$ represents zero-mean Gaussian noise with covariance $\sigma_b^2 \mathbf{I}$, where $\mathbf{I}$ is the identity matrix, observed during reception of the $n$th transmission.

If we let the $N_aN_b \times 1$ vector $\mathbf{h}$ represent $\mathbf{H}$ stacked columnwise, then the received signal
can be expressed as
\[ \hat{w}_{b, mn} = (a^T_n \otimes b^T_m) h + \frac{1}{x_{a,n}} \eta_{b, mn} \] (6.2)
where \( \otimes \) indicates a Kronecker product and \( \eta_{b, mn} = b^T_m \nu_{b,n} \). Arranging the coefficients \( \hat{w}_{b, mn} \) as well as the noise \( \eta_{b, mn} \) into column vectors leads to
\[ \hat{w}_b = (A^T \otimes B^T) h + (X^{-1}_a \otimes I) \eta_b \] (6.3)
where because \( B \) is unitary, the zero-mean noise vector \( \eta_b \) has covariance \( \sigma^2_b I \). When Bob transmits to Alice, he does so using the beamformers \( B \) and Alice receives with the beamformers \( A \). Application of an analysis similar to that shown above for transmission from Alice to Bob and ensuring that the ordering of Alice’s estimates of the channel response vector matches that of Bob’s leads to Alice’s estimate
\[ \hat{w}_a = (A^T \otimes B^T) h + (I \otimes X^{-1}_b) \eta_a \] (6.4)
where \( \eta_a \) is zero-mean estimation error with covariance \( \sigma^2_a I \).

We pause here to emphasize a subtle complexity associated with MIMO channel estimation. Suppose that we wish to directly estimate the channel coefficients \( h \) by letting \( A = I \) and \( B = I \). When Alice transmits a single vector (corresponding here to excitation from a single antenna element), Bob is able to estimate a column of the channel matrix \( H \). In contrast, when Bob transmits a vector to Alice, Alice is able to estimate a row of the channel matrix. Unless \( X_a \) and \( X_b \) are scaled identity matrices, the energy allocated to enable Bob’s estimate of a specific channel coefficient (element of \( H \)) likely differs from that allocated to enable Alice’s estimate of that same coefficient. In other words, the accuracy of Alice’s and Bob’s estimates of each channel coefficient will differ. This asymmetry is represented by the different matrix forms of the noise vectors in (6.3) and (6.4).

### 6.1.2 Key Rate

The key rate, which represents the maximum number of key bits that can be generated when Alice and Bob respectively possess the estimates \( \hat{w}_a \) and \( \hat{w}_b \), is computed from the mu-
tual information between $\hat{w}_a$ and $\hat{w}_b$. Assuming that the elements of the channel coefficient matrix $H$ are zero-mean complex Gaussian random variables, computation of the mutual information requires construction of the covariance of the channel response estimates [16]. Let $R = E \{ h h^\dagger \}$ represent the covariance of the channel coefficients and $V = A^* \otimes B^*$. Relevant covariance matrices are then constructed using

$$\hat{W}_{aa} = E \{ \hat{w}_a \hat{w}_a^\dagger \} = V^\dagger RV + \Upsilon_a = W + \Upsilon_a, \quad (6.5)$$
$$\hat{W}_{bb} = E \{ \hat{w}_b \hat{w}_b^\dagger \} = W + \Upsilon_b, \quad (6.6)$$
$$\hat{W}_{ab} = E \{ \hat{w}_a \hat{w}_b^\dagger \} = W = \hat{W}_{ba} \quad (6.7)$$

where $W = E \{ ww^\dagger \}$ and

$$\Upsilon_a = E \{ (I \otimes X_b^{-1}) \eta_a \eta_a^\dagger (I \otimes X_b^{-1})^\dagger \} \quad (6.8)$$
$$= \sigma_a^2 (I \otimes P_b^{-1}), \quad (6.9)$$
$$\Upsilon_b = \sigma_b^2 (P_a^{-1} \otimes I) \quad (6.10)$$

with $P_\xi = X_\xi X_\xi^\dagger, \xi \in [a, b]$.

Substituting the expressions for $\hat{W}_{aa}$ and $\hat{W}_{bb}$ into (2.22) yield the key rate expression

$$I_k = -\log_2 |I - W (W + \Upsilon_a)^{-1} W (W + \Upsilon_b)^{-1}|. \quad (6.11)$$

Note that this equation is identical to the key rate expression in (3.6) but the structure of $W, \Upsilon_a$ and $\Upsilon_b$ have changed to account for the parallel channel estimates that can be made by a MIMO system.

6.1.3 Beamformer Matrices

Because $W$ depends on the channel estimation beamformer matrices $A$ and $B$ and because $\Upsilon_a$ and $\Upsilon_b$ depend on the energy allocated to transmission of each beamforming vector through the matrices $P_a$ and $P_b$, the key rate in (6.11) directly depends on these channel estimation quantities. The objective of this chapter is to determine the beamformer
matrices and energy allocations that maximize the key rate when the total energy used for channel estimation is constrained.

When the transmitter and receiver are constrained to use beamforming for communication rather than full MIMO processing, the beamforming work in Chapter 3 demonstrated that the optimal choice for the beamforming vectors are the eigenvectors of the covariance matrix $R$. Let $R = U\Lambda U^\dagger$ represent the eigenvalue decomposition of $R$, where $U$ is the unitary matrix of eigenvectors and $\Lambda$ is the diagonal matrix of real eigenvalues. If $V = U$ we obtain $W = \Lambda$, which is diagonal. Since $\Upsilon_a$ and $\Upsilon_b$ are also diagonal, this will diagonalize the matrix inside the determinant in (6.11) and provide a result that is dependent only on the eigenvalues of $R$ along with the noise variances and the channel estimation energy allocations.

However, our formulation requires that $V = A^* \otimes B^*$, meaning that the eigenvectors of $R$ must have this Kronecker structure. If we apply the common assumption that the full covariance can be expressed as $R = R_a \otimes R_b$ where $R_a$ and $R_b$ represent one-sided covariance matrices observed at Alice and Bob, respectively [41], then we have

$$U = U_a \otimes U_b,$$

(6.12)

$$\Lambda = \Lambda_a \otimes \Lambda_b$$

(6.13)

where $U_\xi$ and $\Lambda_\xi$ are respectively the eigenvectors and eigenvalues of $R_\xi$. In this case, choosing $A = U_a^*$ and $B = U_b^*$ achieves $V = U = A^* \otimes B^*$.

More generally, real channels do not conform to this separable Kronecker form [42]. However, the Kronecker approximation for beamformed systems demonstrates that if we construct the one-sided covariance matrices to minimize $\|R - R_a \otimes R_b\|_F$ where $\|\cdot\|_F$ is a Frobenius norm [43], then the difference between the key rate achieved using the actual covariance and that obtained using the Kronecker approximation to the covariance is very small for Rayleigh fading with the errors becoming larger for Ricean fading. In Sec. 6.3.4 we compare the performance of the proposed MIMO algorithm in the presence of both Ricean and Rayleigh fading. We emphasize that choosing the one-sided covariances in this
fashion produces a different result from that obtained by estimating the one-sided covariances directly from the channel coefficients in $H$.

Under the assumption of separability of the covariance matrix, the key rate in (6.11) becomes

$$I_k = \sum_{m=1}^{N_b} \sum_{n=1}^{N_a} \log_2 \left( \frac{\rho_{a,n} \Lambda_{mn} + 1}{\rho_{a,n} \Lambda_{mn} + \rho_{b,m} \Lambda_{mn} + 1} \right) \rho_{a,n} \Lambda_{mn} + \rho_{b,m} \Lambda_{mn} + 1$$

(6.14)

where $\rho_{a,n} = p_{a,n}/\sigma^2_{b}$, $\rho_{b,m} = p_{b,m}/\sigma^2_{a}$, and $\Lambda_{mn} = \Lambda_{a,n} \Lambda_{b,m}$. The total energy used by each node to transmit all beamforming vectors is limited to $P_T$ according to

$$\sum_{n=1}^{N_a} p_{a,n} = \sum_{m=1}^{N_b} p_{b,m} = P_T.$$  

(6.15)

Each term in the sum in (6.14) represents the contribution to the key rate associated with transmission/reception with the beamformers $a_n$ and $b_m$ that have been formed from the eigenvectors of $R$. This form explicitly shows that the corresponding eigenvalues of $R$ and the energy allocated to transmission using these beamformers determine the number of bits contributed through channel estimation with each beamformer pair.

While diagonalizing the matrix $W$ and achieving the relatively simple result for the key rate in (6.14) is convenient, this does not prove that deriving the beamformers from the eigenvectors of the covariance is optimal. Unfortunately, we have been unable to generate a rigorous proof demonstrating the optimality of this choice for MIMO systems. However, the fact that this choice can be proven optimal for beamformed systems is a strong motivation to use this approach here. Furthermore, the results in Section 6.3 demonstrate that this approach leads to significant increases in the achieved key rate, reinforcing that the technique has merit. Finally, the eigenvector decomposition identifies the dimensions of $R$ with the largest possible variance (largest eigenvalues), which means that the largest eigenvalue corresponds to the largest variance for any combination of transmit and receive beamformers, and the second largest eigenvalue corresponds to the largest variance for any beamformers orthogonal to the first. It is therefore reasonable that we should allocate transmission energy based on this efficient representation of the variances associated with the different dimensions of $R$. 

90
6.1.4 Equal Energy Allocation

To further reinforce that use of the eigenvectors is a reasonable if not optimal approach, consider the case where the total energy is evenly allocated to each transmission so that \( p_{a,n} = p_a = P_T/N_a \) and \( p_{b,m} = p_b = P_T/N_b \), an approach to which we refer as equal allocation. In this case, \( \Upsilon_a = (\sigma_a^2/p_b)I \) and \( \Upsilon_b = (\sigma_b^2/p_a)I \) are scaled identity matrices and the eigenvectors of \( W \) can be factored out of the matrix inside the determinant in (6.11). The resulting key rate reduces to the form in (6.14) with the substitutions \( p_{a,n} = p_a \) and \( p_{b,m} = p_b \). In other words, we obtain this form for the key rate, that depends only on the eigenvalues of \( R \) and the noise variances at each node, provided only that the chosen beamformers are unitary. The fact that only the eigenvalues of \( R \) are important suggests that in the more complicated case where the energies \( p_{a,n} \) and \( p_{b,m} \) are unequal, the form used in (6.14) wherein these energies are allocated according to the eigenvalues of \( R \) is reasonable.

6.2 Iterative Optimization for Energy Allocation

Our objective is now to identify the values of \( p_{a,n} \) and \( p_{b,m} \) that maximize the key rate in (6.14) subject to the total node energy constraint in (6.15). This optimization can be accomplished using a Lagrange multiplier if the objective function is concave, which is satisfied if the Hessian of the multivariate function is negative semi-definite. Unfortunately, the function in (6.14) is not concave for simultaneous optimization of the energy allocations for both nodes. However, if we fix Bob’s set of energy allocations, then the Hessian of \( I_k \) with respect Alice’s energy allocations is a diagonal matrix with the \( n \)th diagonal element given as

\[
\ln 2 \frac{\partial^2 I_k}{\partial \rho_{a,n}^2} = \sum_{m=1}^{N_b} \left( \frac{1}{\rho_{a,n} + \rho_{b,m} + 1/\Lambda_{mn}} \right)^2 - \frac{1}{(\rho_{a,n} + 1/\Lambda_{mn})^2}. \tag{6.16}
\]

Since the denominator of the first term is greater than or equal to the denominator of the second term, the sum is never positive. Consequently the key rate in (6.14) is a concave function of Alice’s energy allocation when Bob’s energy allocation is fixed.
A similar result applies in determining Bob’s energy allocation when Alice’s energy allocation is fixed, guaranteeing that a Lagrangian multiplier solution will produce the optimal energy allocation at one node when that at the other is fixed. This suggests an iterative optimization technique where the energy allocation at one node is initialized, and a Lagrange multiplier optimization is performed at the second node. The energy allocation at the second node is then fixed while that at the first node is obtained through optimization. Since each optimized function is concave, performance is guaranteed to improve (or at least not degrade) with each iteration.

The optimal energy allocation for Alice given that Bob’s energy allocation is fixed must satisfy the Lagrange multiplier equation

$$\frac{\partial}{\partial \rho_{a,n}} \left[ I_k + \frac{\gamma}{\ln 2} \left( \frac{P_T}{\sigma_b^2} - \sum_{i=1}^{N_a} \rho_{a,i} \right) \right] = 0 \quad (6.17)$$

for all $n$, where $\gamma$ is the Lagrange multiplier and the factor of $\ln 2$ is included for convenience. This expression simplifies to $\partial I_k / \partial \rho_{a,n} = \gamma / \ln 2$ or

$$\gamma = \sum_{m=1}^{N_b} \frac{1}{\rho_{a,n} + 1/\Lambda_{mn}} - \frac{1}{\rho_{a,n} + \rho_{b,m} + 1/\Lambda_{mn}} \quad (6.18)$$

for all $n$. Eq. (6.19) makes it clear that since the sum must produce the same value $\gamma$ for each value of $n$, if $\Lambda_{a,r} > \Lambda_{a,s}$ for two different values $r$ and $s$, then $p_{a,r} > p_{a,s}$ ($\rho_{a,r} > \rho_{a,s}$). In other words, the solution will allocate more energy to transmit beamformer vectors associated with larger eigenvalues of the covariance. A similar expression to that in (6.18) applies when solving for Bob’s energy allocations with Alice’s allocations fixed.

Iterative determination of the energy allocations $p_{a,n}$ and $p_{b,m}$ begins with an initialization of $p_{b,m}$. Then, we apply the Newton-Raphson method [39] to determine the value of $\gamma$ that satisfies the energy constraint in (6.15), where at each step of the computation we also apply the Newton-Raphson method to determine the values of $p_{a,n}$ that satisfy (6.18). Since the expression for the total energy in (6.15) is not a differentiable function of $\gamma$, application of
Newton-Raphson for determination of $\gamma$ requires numerical approximation of the derivative. We next solve the corresponding set of nonlinear equations for $p_{b,m}$ using an identical procedure. This procedure is repeated iteratively until the solutions converge. Let $\rho^{(v)}_{\xi,\ell}$ represent the normalized energy allocation for the $\ell$th mode of node $\xi \in [a, b]$ at the $v$th iteration and let $C^{(v)}_{\xi}$ represent the diagonal matrix with $\ell$th diagonal element $C^{(v)}_{\xi,\ell} = \left(\rho^{(v+1)}_{\xi,\ell} - \rho^{(v)}_{\xi,\ell}\right)^2$. Convergence is achieved for the iteration index $v$ at which

$$
\left\{\text{tr}\left[C^{(v)}_a\right]\right\}^{1/2} + \left\{\text{tr}\left[C^{(v)}_b\right]\right\}^{1/2} < 10^{-8}
$$

where $\text{tr}[-]$ is the trace.

To demonstrate the convergence of the iterative optimization technique, we apply the algorithm to 500 randomly-realized system descriptions. Specifically, the number of elements in Alice’s and Bob’s arrays is drawn from a discrete uniform distribution on $[2, 9]$, and the eigenvalues of the covariance matrices $\Lambda_a$ and $\Lambda_b$ as well as the total allowed transmission energy $P_T$ are randomly drawn from an exponential distribution with a variance of 2. Bob’s energy allocation is initialized to transmit all energy for the beamformer associated with the largest covariance eigenvalue. In all cases, we find that the simulation has converged within 5 iterations. However, we run the simulation for an additional 3 iterations and define the final achieved key rate as $I^{(\text{max})}_k$. Figure 6.1 plots the averaged normalized difference between $I^{(\text{max})}_k$ and the key rate $I^{(v)}_k$ achieved at the $v$th iteration as a function of the iteration number. As can be observed, the iterative algorithm converges quickly.

Our development shows that the cost function for the energy allocation at one node with that at the other node fixed is concave, meaning that the solution to the Lagrange multiplier problem represents a global maximum. However, because we iteratively solve this problem, there is no guarantee that the final solution represents a global maximum of the joint optimization. Proving global optimality requires either demonstrating existence of a single local maximum or proving convergence to the global maximum. Because we have not been able to prove either of these analytically, we resort to numerical analysis to explore this issue. Specifically, we apply a steepest ascent (gradient search) algorithm to numerically maximize the mutual information from a number of different initial energy allocations called
starting points. Because we know that the energy associated with larger eigenvalues should be larger than that associated with smaller eigenvalues, the solution space is restricted to the Cartesian product of an $N_a + 1$ simplex and an $N_b + 1$ simplex. The vertices of these simplexes are the points where all probing vectors receiving any energy receive equal energy. All pairs of vertices from the solution space are used as starting points for the search to guarantee that the search will approach the maximum from different directions. For over 250,000 randomly generated systems, each steepest ascent search converged to the same solution as that generated by the iterative allocation scheme for every starting point. This provides strong evidence that the cost function contains a single global maximum that is properly identified by the iterative optimization.

6.3 Results

To explore the impact of the preceding developments, we compare the performance realized with equal energy allocation to that obtained when estimating the channel using
the eigenvectors of the covariance matrices as the beamformer matrices coupled with the iterative energy optimization. For these simulations, both Alice and Bob use uniform linear arrays of vertically oriented dipoles with a total array length of $2\lambda$, where $\lambda$ is the free space wavelength. For all simulations, we assume $\sigma_a^2 = \sigma_b^2 = \sigma_0^2$ and define the SNR as $P_T/\sigma_0^2$.

For simplicity, we assume that the propagation is confined to the horizontal plane. The channel realizations are based on the work in [40] where the channel covariance matrix is formed directly from the PAS. The PAS description used for Rayleigh propagation is mathematically defined in Section 3.3 and the PAS description for Ricean propagation is defined in Section 4.1.1. Unless otherwise specified, each channel realization consists of $Q = 5$ multipath clusters and the propagation environment is assumed Rayleigh.

6.3.1 Array Size

We first consider how the number of elements in each array impacts performance by sweeping the size of Alice’s array over the interval $2 \leq N_a \leq 10$ and leaving Bob’s array fixed with $N_b = 5$ elements. Figure 6.2 plots the key rate achieved for equal energy allocation and for the energy allocation obtained using iterative optimization as a function of $N_a$ for three different values of SNR. These results show that the achieved key rate increases substantially with SNR for both techniques, although the performance achieved using iterative optimization is superior, particularly for larger array sizes. The relative performance is more clearly observed in Fig. 6.3 which plots the ratio of the key rate achieved with equal allocation ($I_{k,E}$) to that obtained with iterative optimization as a function of SNR for different values of $N_a$. At an SNR of $-10$ dB, the iterative optimization produces a key rate that is roughly five times that achieved using equal energy allocation. As the SNR increases, the relative benefit of the iterative optimization decreases.

6.3.2 Energy per Key Bit

The key rate, representing the maximum number of bits that can be generated using reciprocal channel estimation, is a nonlinear function of the total allocated energy. We can define a measure of efficiency as the energy per key bit generated, which is simply the ratio of the total energy to the key rate, or $P_T/I_k$. Figure 6.4 plots the normalized energy per bit
(\(\frac{P_T}{\sigma_0^2}I_k\)) as a function of SNR for two values of \(N_a\) when in both cases \(N_b = 5\). This result shows that allocating either too little or too much energy reduces the energy efficiency of key bit generation. This is in contrast to the situation in communications where the highest energy efficiency is achieved using low-rate communication with low signal energy. The major reason for this difference is that when the SNR is very low, the resulting poor channel estimates observed at both nodes lead to reduced key generation efficiency, and therefore there is an optimal SNR in terms of maximizing the efficiency. These results further reinforce the improvement in performance enabled by the iterative optimization approach that is most significant at low SNR.

### 6.3.3 Number of PAS Clusters

Next, we explore the impact of the propagation environment on the performance of MIMO channel estimation for key establishment. The number of significant eigenvalues of the covariance, which directly impacts the key rate, is directly tied to the number of clusters present in the PAS description. Therefore, we set \(N_a = N_b = 5\) and sweep the number of
Figure 6.4: Ratio of the key rate achieved with equal energy allocation ($I_{k,E}$) to that achieved using the iterative optimization for different values of $N_a$ when $N_b = 5$.

clusters over the range $1 \leq Q \leq 100$. Figure 6.5 plots the key rate achieved by the iterative optimization and by equal energy allocation as a function of the number of clusters averaged over 180 channel realizations for two values of the SNR. The relative performance is more clearly observed in Fig. 6.6 which shows the ratio of the key rates achieved using the two methods.

For both SNR values, the performance advantage achieved with iterative optimization decreases as the number of clusters increases. This occurs because the iterative optimization allocates the energy to accurately estimate the subspace of the channel associated with the strong covariance eigenvalues, as these are the eigenvalues that contribute most to the key rate. Equal allocation, on the other hand, evenly spreads the resources to estimation of the entire channel. However, as the number of clusters increases, an increasing number of eigenvalues become significant, and therefore accurately estimating the entire channel becomes important. The result is that the iterative optimization equalizes the energy allocated to probing of each vector, and therefore the result converges to that of equal allocation.
One interesting observation from Fig. 6.5 is that at low SNR, the performance degrades with the number of clusters while at higher SNR, an opposite trend is observed. This behavior stems from our normalization on the cluster amplitudes $\beta_q$ which makes it so that as the number of clusters increases, the strength of the dominant scatterers decreases. Mathematically this translates to the energy represented in the dominant eigenvalues for small values of $Q$ being spread over a larger number of weaker eigenvalues as $Q$ increases. The iterative technique at low SNR exploits these large eigenvalues created by the dominant clusters to produce a large key rate. As the number of clusters grows, however, at low SNR the scheme does not have the energy to properly exploit a large number of similarly-valued eigenvalues, and the key rate therefore decreases.

6.3.4 Line of Sight Propagation

In Chapter 4, simulation results show that when a line of sight component is present, the performance of the Kronecker approximation to OSEA is degraded. With that same approximation applied in the iteratively optimized MIMO probing strategy, it is important
to understand how well the iteratively optimized energy allocation performs with Ricean propagation. This analysis is more difficult than in the beamforming case because our work has not generated a bound on the MIMO key rate against which performance can be measured.

To understand the effects of LOS propagation, Fig. 6.7 plots the performance of equal allocation relative to that of the iterative algorithm as a function of SNR for several values of the Ricean $K$-factor. At low SNR, we see that for all propagation descriptions, iterative optimization outperforms equal allocation. As SNR increases however, the performance advantage of the iterative algorithm fades. Finally, when SNR is large and a LOS component is present, the performance of equal allocation eclipses that achieved by the iterative optimization by as much as 8%. When no LOS component is present, the iterative optimization outperforms equal allocation, although the relative difference becomes very small.

For OSEA, the reduced performance observed for the Kronecker approximation in the presence of LOS propagation leads to an alternative energy allocation strategy that isolates the LOS component of the correlation matrix and then allocates energy to that...
component as a separate channel. This is possible with beamformer probing because a practical probing strategy must only be able to separate each composite probing vector into Alice and Bob’s beamformer weightings. In essence, each probing vector must be the Kronecker product of two beamformer weightings. The Kronecker approximation in the iterative MIMO probing strategy relies on the fact that the entire covariance matrix is the Kronecker product of two separate matrices in order to form both the energy allocation and the probing vectors. This is why it is possible to find a bounding key rate with OSEA when no approximation is made while the approximation must be made in order to find the performance of the iteratively optimized algorithm. With this more restrictive limitation, the addition of an isolated probing vector corresponding to line of sight propagation cannot be integrated into the required Kronecker structure. Consequently, the degradation observed with LOS propagation at high SNR is an algorithm limitation which should be considered when adopting the iterative strategy for operation at high SNR when a LOS signal path may be present.
With these results demonstrating that equal allocation outperforms the iterative optimization with the Kronecker approximation under specific conditions, it is important to understand that the optimality of eigenvector probing with a MIMO system having Kronecker structured covariance matrix is a separate issue from the validity of the Kronecker approximation to an arbitrary matrix when implementing our proposed MIMO probing strategy. This subtlety is important when interpreting the results in this section which primarily illustrate the performance as the Kronecker approximation and the associated channel separability break down due to presence of a LOS signal path.

6.4 Chapter Summary

This chapter provides a strategy that increases the achievable number of key bits established by a MIMO system using reciprocal channel estimation. The algorithm uses the eigenvectors of the transmit and receive covariance matrices as beamforming vectors and then uses an optimization to determine the energy allocated to transmission of each vector. The development shows that the cost function for the optimal energy at one node with that
at the other node fixed is concave and uses this notion to implement an iterative optimization strategy, with numerical analysis providing strong evidence that the optimization converges to the single global optimum solution. Simulation results using the technique show significant increases in the achieved key rate at low SNR and further explore the relationship between the minimum energy per key bit and the total available energy for traditional and iteratively optimized energy allocations. Additional results explore the utility of the proposed algorithm with different propagation environment characteristics.
Chapter 7

Conclusion

This dissertation provides new strategies for optimally probing a wireless channel with multiple antenna systems to maximize the key rate. For these systems, certain antenna excitations are more efficient than others at coupling energy between the transmitting and receiving nodes. For MIMO and beamforming systems, the proposed probing strategies preferentially allocate energy to the more efficient excitations at the expense of the less efficient excitations.

For beamforming systems, Chapter 3 develops a theoretically optimal probing strategy called OSEA that upper bounds the key rate achievable by any set of beamformer weightings satisfying a given energy constraint. This technique solves for the optimal composite probing vectors that encapsulate the beamformer weightings applied at each node. The optimal composite probing vectors are the eigenvectors of the spatial covariance matrix, and the energy allocated to each is chosen from one of a few locally optimal energy allocation configurations. To implement OSEA in practice requires converting the composite probing vectors back into beamformer weightings. This cannot be done directly, and so therefore propose a physically realizable modification to OSEA based on a Kronecker approximation to the spatial covariance matrix. Simulations demonstrate the utility of OSEA and the Kronecker approximation when compared to traditional probing strategies.

In Chapter 4, simulations are used to evaluate the performance of OSEA for a variety of propagation environment descriptions. These results show that the performance of the Kronecker approximation to OSEA degrades when a line of sight propagation path is present. This shortcoming is addressed with a modification to the Kronecker approximation, which separately accounts for the effects of the line of sight path. Extensive results compare the modified Kronecker approximation, the original Kronecker approximation, traditional
probing, and OSEA. These results demonstrate that the modified algorithm performs well in all propagation environments.

In essence the OSEA probing strategy bounds the key rate achievable when probing the propagation channel connecting two arrays. This leaves open a question of how well Alice and Bob’s arrays do at extracting all of the information available in the connecting propagation environment. Constraining the physical dimensions of the the antenna arrays, the work in Chapter 5 answers this question by applying OSEA to a finite vector current Fourier basis function expansion of the aperture volume. As the number of basis functions included in the expansion becomes large, the OSEA key rate of the expansion converges to a key rate spatial bound. We prove that the key rate spatial bound upper bounds the key rate achievable by any other antenna arrays within the same volume that satisfy a common energy constraint. Simulation results explore the effect of truncating the number of basis functions and the effect of optimizing with respect to an alternative energy constraint.

Chapter 6 defines a sophisticated resource allocation strategy for MIMO probing. The proposed technique requires a Kronecker approximation to the spatial covariance matrix to determine the energy allocation strategy. The energy allocation is computed through an iterative technique that uses concave maximization to optimize Alice’s energy allocation while Bob’s energy allocation is fixed. A similar procedure is then used to optimize Bob’s energy allocation with Alice’s allocation fixed. This process is iterated until the energy allocations at both nodes converge. The convergence of the algorithm is analyzed as well as the performance as a function of several relevant system parameters.

For the proposed beamforming and MIMO probing strategies, results demonstrate that the relative performance improvement is greatest at low SNR. In the beamforming case, OSEA produces a twenty fold improvement in key rate over traditional probing for very low SNR. As SNR increases, the advantage decreases so that the difference between the key rate achieved using OSEA and that obtained using traditional probing is less than 10% at high SNR for the array parameters simulated. For the MIMO case, the key rate achieved by iterative optimization is more than five times larger than that obtained using traditional probing at low SNR. As the SNR increases, this difference decays to zero when no line of sight signal path is present. For MIMO probing at high SNR when a line of sight
path is present, the results demonstrate that traditional probing can outperform the iterative technique by as much as 8%.

7.1 Future Work

The work presented here is a first step toward understanding how to better use the resources available to a multiple antenna system to maximize the key rate. This first step leaves open several possible directions for future research related to optimal multiple antenna channel probing.

1. The work in this dissertation considers how MIMO or beamforming systems should probe a channel with spatial correlation between antenna elements. In real systems, the channel response is correlated in both time and space. While OSEA applied to the full spatial and temporal correlation matrix upper bounds the key rate that can be achieved by a beamforming system, the practical beamformer implementations such as the Kronecker approximation would not directly extend to this more general case. The iteratively optimized MIMO probing strategy would also require significant modification. One thing that might be addressed by a new probing strategy accounting for spatial and temporal correlation is the rate at which different subspaces decorrelate and if the response between some antenna excitations should be probed more regularly than the response between others.

2. One key limitation of the probing strategies we propose is that each is computed using the full spatial covariance matrix. In practice each node’s estimate of the spatial covariance matrix is produced using actual measurements of the channel response. In other words, probing the channel is required to find the optimal way to probe the channel. A next step in future research would be defining a strategy for increased key rate that does not require an estimate of the full spatial covariance matrix.

3. The work in beamforming arrays is a significant theoretical contribution, but the implications are limited by the cost and complexity of beamforming systems. Antenna pattern reconfigurability is more easily achieved using parasitic array elements with
switchable loads. Because the load does not directly govern the current on each element, arbitrary currents cannot be achieved. The optimality of the OSEA solution is provable because the optimal choice of currents diagonalizes the covariance matrix, a result that is unachievable with the parasitic loads. This fundamental modification makes it nearly impossible to find a tight upper bound on key rate for parasitic arrays. Using insights gained from OSEA, it may be possible to determine a practical probing scheme that performs well over a range of parameters.

4. The results presented in this dissertation are entirely based on simulation. Previous work in [15] demonstrates that it is possible to use reciprocal estimates of the channel response to generate key bits. Implementing the Kronecker approximation for beamformer channel probing or the iterative optimization for the MIMO case would provide insight into each algorithm’s utility in practice and help to uncover additional difficulties associated with practical implementation.
Bibliography


[19] F. Miller, *Telegraphic code to insure privacy and secrecy in the transmission of telegrams*. C.M. Cornwell, 1882. 6


