

Key Error Rates in Physical Layer Key Generation: Theoretical Analysis and Measurement-Based Verification

Ozan Alp Topal, Gunes Karabulut Kurt, and Berna Özbek

Abstract—Channel gains are frequently used to obtain a secret key that can be used for encryption in physical layer security systems. However, the channel gains captured by the nodes may not always be the same due to channel estimation errors. This would result in a non-zero key error rate (KER). In this letter, we obtain theoretical expressions for KER in orthogonal frequency division multiplexing systems. Tight KER approximations are provided based on Gauss–Laguerre quadrature. A measurement-based study is conducted by using software defined radio nodes to demonstrate the validity and the practicality of the provided results.

Index Terms—Key error rate (KER), physical layer security, key extraction, physical layer key generation.

I. INTRODUCTION

WIRELESS medium has security vulnerabilities towards passive attacks such as eavesdropping, due to its broadcast nature. A security solution against such attacks is to use encryption/decryption mechanisms that rely on a bit sequence referred to as the secret key [1]. This key needs to be available at the legitimate transmitter node and receiver node. Even though encryption solutions can provide a reliable protection, generation of the key and maintaining its randomness comes at a high cost of computational complexity. Key-sharing algorithms such as Diffie-Hellman are frequently used for this purpose [2]. Enabled by the channel reciprocity property [3], physical layer (PHY) security techniques provide a new perspective for key generation. This property relies on the reversible nature of electromagnetic waves, ensuring observation of similar channel gains (the amplitudes of channel coefficients) between two nodes. Spatially distinct users encounter distinct wireless channel gains (or other characteristics), that can be used to generate secret keys. This approach is referred to as PHY key generation. The randomness of wireless channel provides continuous and unpredictably changing key characteristics. PHY key generation is significantly less complex than the classical key-sharing algorithms. In fact, it merely introduces a quantization process as the channel needs to be tracked for data communication in any case.

Manuscript received July 25, 2017; accepted August 5, 2017. Date of publication August 15, 2017; date of current version December 15, 2017. This work was supported by TUBITAK under Project 114E626. The associate editor coordinating the review of this paper and approving it for publication was J. Xu. (Corresponding author: Gunes Karabulut Kurt.)

O. A. Topal and G. K. Kurt are with the Department of Electronics and Communication Engineering, Istanbul Technical University, 34469 Istanbul, Turkey (e-mail: ozan.topal@itu.edu.tr; gkurt@itu.edu.tr).

B. Özbek is with the Department of Electric and Electronics Engineering, Izmir Institute of Technology, 35430 Izmir, Turkey (e-mail: bernaobzbek@iyte.edu.tr).

Digital Object Identifier 10.1109/LWC.2017.2740290

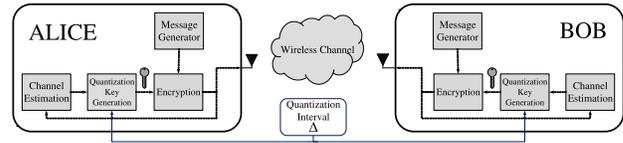


Fig. 1. The system model. Using time division duplexing, Alice and Bob generate the secret key based on the channel estimates. Estimation errors may cause non-identical keys.

The basic secret key generation model is set between two nodes, Alice and Bob. Each node uses PHY information to generate a secret key. If their keys match, then they can securely communicate. However, even one bit key mismatch would disrupt communication. Therefore, bit error rate (BER) becomes insufficient to measure the system performance and the key error rate (KER) needs to be considered. Even in the presence of reciprocal channels, the channel estimates in real-life can be erroneous due to channel estimation error, leading to non-zero KER values.

In order to properly assess the performance of PHY key generation systems, KER needs to be accurately modeled. With this objective, we consider the system model illustrated in Fig. 1. Both Alice and Bob will capture channel coefficients possibly with some variations due to channel estimation errors. We feed channel gains to a uniform quantizer, as frequently used in [4]. Both nodes utilize the same quantizer, with the quantization interval Δ . Quantized gains can be used to generate secret keys. Our proposed system considers raw key extraction which constitute a baseline for KER, and can be reduced through information reconciliation. Our contributions are:

- Theoretical KER expressions in presence of channel estimation error is derived considering complex Gaussian channels. The results are extended to orthogonal frequency division multiplexing (OFDM) scheme that is frequently used in current communication standards.
- Tight approximations to KER expressions are obtained by using Marcum-Q functions and Gauss-Laguerre quadrature (GLQ).
- Numerical results are given to verify the theoretical expressions. The tightness of the provided approximations are shown.
- Software defined radio (SDR) based tests are conducted to demonstrate the validity and the real-life applicability of the obtained KER expressions.

Related Works: As received signal strength (RSS) measurements are already available in wireless devices, they are frequently used to generate a secret key [5]. Another related source for key generation is the channel gain [6]. Channel gain is advantageous especially due to its simple acquisition (as in

RSS) and rapid variations, improving the randomness of secret keys. But the correlation between channel gains needs to be carefully considered. In [7], the correlation between reciprocal channels are theoretically modeled. KER expressions are not yet derived in the literature, and the majority of the available results depend on empirical measurements [8]. Phases of the channel coefficients also show reciprocal characteristics, but their acquisition requires complex systems that are tightly synchronized.

Notation: $X \sim \mathcal{CN}(\nu, \sigma^2)$ refers to a random variable X , which follows a complex Gaussian distribution with mean ν and variance σ^2 . $\lfloor \cdot \rfloor$ represents the floor function. $\text{erfc}(\cdot)$ expresses the complementary error function. $R \sim \mathcal{R}(\nu, \sigma^2)$ represents the Rice distribution for statistically independent $Y \sim N(\nu \cos \theta, \sigma^2)$ and $Z \sim N(\nu \sin \theta, \sigma^2)$, where $R = \sqrt{Y^2 + Z^2}$. When $\nu = 0$ we obtain the Rayleigh distribution $R \sim \mathcal{R}(0, \sigma^2)$.

II. KEY ERROR RATE

We consider a two-way OFDM system composed of N subcarriers. The channel estimates are obtained in a time division duplex mode. As indicated in [7], in order to obtain channel reciprocity, we assume that sampling time difference between two nodes is smaller than the coherence time. As in [9], we denote the parallel channel corresponding to the n^{th} OFDM subcarrier by H_n , for $n = 0, \dots, N-1$. We consider complex Gaussian channels, where $H_n \sim \mathcal{CN}(\nu, \sigma^2)$. Despite the channel reciprocity assumption, Alice and Bob may estimate channel coefficients in erroneous forms. Let the estimated channel coefficient at Alice and Bob be respectively represented by $\hat{H}_{n,1}$ and $\hat{H}_{n,2}$, where

$$\hat{H}_{n,k} = H_n + e_{n,k}, \quad (1)$$

for $k = 1, 2$. In the literature, the estimation error is modeled as complex normal distributed, where $e_{n,k} \sim \mathcal{CN}(0, \sigma_{e,k}^2)$ [10]. Assuming a uniform quantizer with a step size Δ , an estimated channel gain $|\hat{H}_{n,k}|$ is mapped to the l^{th} quantization interval I_l , where $l_n = \lfloor \frac{|\hat{H}_{n,1}|}{\Delta} \rfloor$, and $I_l = [l_n \Delta, (l_n + 1) \Delta]$. If both estimates are located in the same interval, their decisions and eventually key bits would be the same.

For a given estimate of the n^{th} subcarrier's channel at Alice ($\hat{H}_{n,1}$), the corresponding channel estimate at Bob becomes $\hat{H}_{n,2} = \hat{H}_{n,1} - e_{n,1} + e_{n,2}$. In this conditional case, the amplitude of $\hat{H}_{n,2}$ follows Rice distribution with parameters $|\hat{H}_{n,2}| \sim \mathcal{R}(|\hat{H}_{n,1}|, \sigma_e^2)$, where $\sigma_e^2 = \sigma_{e,1}^2 + \sigma_{e,2}^2$. Hence, we can obtain the probability that $|\hat{H}_{n,2}|$ is located in I_l interval as

$$P_l = \int_{l_n \Delta}^{(l_n+1)\Delta} \frac{x}{\sigma_e^2} e^{-\frac{x^2 + |\hat{H}_{n,1}|^2}{2\sigma_e^2}} I_0\left(\frac{x|\hat{H}_{n,1}|}{\sigma_e^2}\right) dx,$$

$$= Q_1\left(\frac{|\hat{H}_{n,1}|}{\sigma_e}, \frac{l_n \Delta}{\sigma_e}\right) - Q_1\left(\frac{|\hat{H}_{n,1}|}{\sigma_e}, \frac{(l_n + 1)\Delta}{\sigma_e}\right). \quad (2)$$

where, P_l represents the probability that the quantizer outputs of the two estimated channel gains are the same for a given $\hat{H}_{n,1}$. $I_0(\cdot)$ denotes the modified Bessel function of the first kind and the first order Marcum-Q function is represented by $Q_1(\cdot, \cdot)$. If we denote the first estimated channel gain $|\hat{H}_{n,1}|$ as y , the key matching probability for a single subcarrier, P_c , becomes

$$P_c = \int_0^\infty \left[Q_1\left(\frac{y}{\sigma_e}, \frac{l_n \Delta}{\sigma_e}\right) - Q_1\left(\frac{y}{\sigma_e}, \frac{(l_n + 1)\Delta}{\sigma_e}\right) \right] f_Y(y) dy, \quad (3)$$

where $f_Y(y)$ denotes the probability density function (pdf) of the corresponding channel model. Considering the complete set of independently fading subcarriers, the key matching probability becomes $P_o = P_c^N$, and the KER becomes $\text{KER} = 1 - P_o$. For Rice distributed channel gains, where $H_n \sim \mathcal{CN}(\nu, \sigma^2)$, the first estimated channel is modeled as $|\hat{H}_{n,1}| \sim \mathcal{R}(\nu, \sigma_s^2)$, where $\sigma_s^2 = \sigma^2 + \sigma_{e,1}^2$. We obtain P_c as given in Eq. (4), as shown at the bottom of this page. For Rayleigh fading channel gains, we set $\nu = 0$, where $I_0(0) = 1$.

For notational convenience, let us define,

$$g(\alpha, \beta_1, \beta_2, \nu) = [Q_1(\alpha, \beta_1) - Q_1(\alpha, \beta_2)] e^{-\frac{\nu^2}{2\sigma_s^2}} I_0\left(\frac{\nu\alpha\sigma_e}{\sigma_s^2}\right), \quad (5)$$

where $\alpha = \frac{y}{\sigma_e}$, $\beta_1 = \frac{l_n \Delta}{\sigma_e}$, $\beta_2 = \frac{l_n \Delta + \Delta}{\sigma_e}$. The difference in Eq. (2) corresponds to $g(\alpha, \beta_1, \beta_2, 0)$ when $\nu = 0$. In order to provide a closed form approximation to the exact KER expression in Eq. (4), we can use a tight approximation to $g(\alpha, \beta_1, \beta_2, 0)$ as

$$\tilde{g}(\alpha, \beta_1, \beta_2, 0) \approx \begin{cases} 1 - e^{-\frac{(\Delta/\sigma_e)^2}{2}} & \alpha = 0, \beta_1 = 0 \\ 1 - Q_1(\alpha, \beta_2) & \alpha \neq 0, \beta_1 = 0 \\ Q_1(\alpha, \beta_1) - Q_1(\alpha, \beta_2) & \alpha \neq 0, \beta_1 \neq 0, \end{cases} \quad (6)$$

where $\beta_1 \leq \alpha < \beta_2$, due to the definition of the interval I_l . This closed form function depends on the condition of whether one of the input values are zero [11]. Tight approximations of $Q_1(\alpha, \beta)$ can be conditioned on $\beta \leq \alpha$ or $\beta \geq \alpha$ [12]. Therefore, we use two different tight approximations for $Q_1(\alpha, \beta_1)$ and $Q_1(\alpha, \beta_2)$ as shown in Eq. (7a) and Eq. (7b), as shown at the bottom of this page, and replace the $Q_1(\cdot, \cdot)$ functions in $\tilde{g}(\alpha, \beta_1, \beta_2, \nu)$ to obtain $\hat{g}(\alpha, \beta_1, \beta_2, \nu)$. The key matching probability for a single subcarrier can now be expressed as

$$P_c = \int_0^\infty \hat{g}(\alpha, \beta_1, \beta_2, \nu) \frac{y}{\sigma_s^2} e^{-\frac{y^2}{2\sigma_s^2}} dy. \quad (8)$$

$$P_c = \int_0^\infty \left[Q_1\left(\frac{y}{\sigma_e}, \frac{l_n \Delta}{\sigma_e}\right) - Q_1\left(\frac{y}{\sigma_e}, \frac{(l_n + 1)\Delta}{\sigma_e}\right) \right] \frac{y}{\sigma_s^2} I_0\left(\frac{\nu y}{\sigma_s^2}\right) e^{-\frac{(y^2 + \nu^2)}{2\sigma_s^2}} dy \quad (4)$$

$$Q_1(\alpha, \beta) \approx \begin{cases} 1 - \frac{\sqrt{2\pi} \beta I_0(\alpha\beta)}{e^{\alpha\beta} - e^{-\alpha\beta}} \left[\frac{1}{2} \text{erfc}\left(\frac{\alpha - \beta}{\sqrt{2}}\right) + \frac{1}{2} \text{erfc}\left(\frac{\alpha + \beta}{\sqrt{2}}\right) - \text{erfc}\left(\frac{\alpha}{\sqrt{2}}\right) \right] & \beta \leq \alpha \\ \sqrt{\frac{\pi}{2}} \frac{\beta I_0(\alpha\beta)}{e^{\alpha\beta} - e^{-\alpha\beta}} \left[\text{erfc}\left(\frac{\beta - \alpha}{\sqrt{2}}\right) - \text{erfc}\left(\frac{\beta + \alpha}{\sqrt{2}}\right) \right] & \beta > \alpha \end{cases} \quad (7a)$$

$$\left[\text{erfc}\left(\frac{\beta - \alpha}{\sqrt{2}}\right) - \text{erfc}\left(\frac{\beta + \alpha}{\sqrt{2}}\right) \right] \quad \beta > \alpha \quad (7b)$$

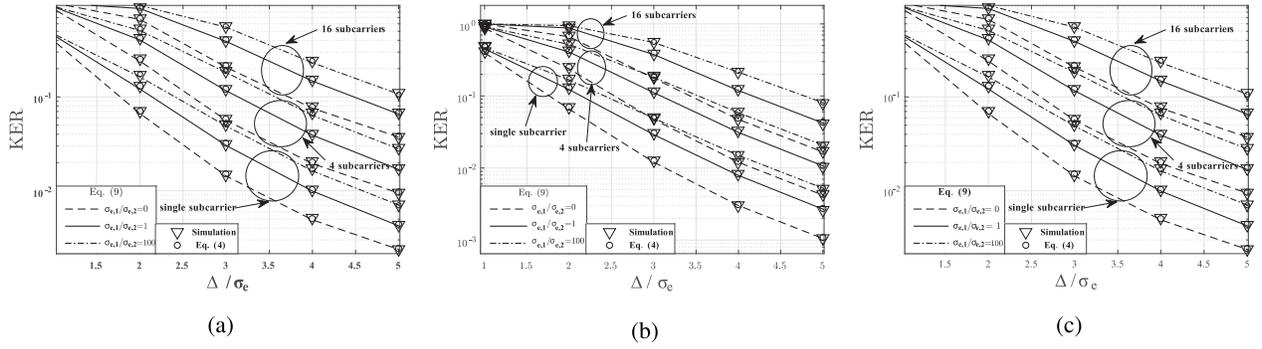


Fig. 2. KER for different carrier numbers with respect to different K factors, (a) $K = 0$, (b) $K = 1$, (c) $K = 20$.

The GLQ can now be used to obtain a tight approximation to Eq. (8) as $\int_0^\infty e^{-\psi} f(\psi) d\psi \approx \sum_{m=1}^M w_m f(\psi_m)$, where $w_m = \frac{\psi_m}{(M+1)^2 [L_{M+1}^2(\psi_m)]}$. The degree of the polynomial, M , denotes the number of considered terms. As $M \rightarrow \infty$, the approximation becomes an equality [13]. ψ_m represents the m^{th} root of the Laguerre polynomial $L_M(\psi)$. A change of variables can be applied by defining $\frac{y^2}{2\sigma_s^2} = \psi$ and $\frac{y}{\sigma_s^2} = d\psi$. Hence KER for N subcarriers can be obtained as

$$\text{KER} \approx 1 - \left(\sum_{m=1}^M w_m \hat{g} \left(\frac{\sqrt{2\sigma_s^2} \psi_m}{\sigma_e}, \beta_1, \beta_2, \nu \right) \right)^N. \quad (9)$$

Here, Δ , ν , σ , $\sigma_{e,1}$ and $\sigma_{e,2}$ are the critical parameters affecting the KER value. A larger Δ would reduce KER, whereas larger standard deviations of the channel estimation errors ($\sigma_{e,1}$, $\sigma_{e,2}$), would increase KER.

III. NUMERICAL RESULTS

10^8 Monte Carlo iterations are run based on the theoretical model expressed in Eq. (1). First, the non-zero mean complex Gaussian reciprocal channel coefficients are generated. Their gains follow Rice distribution. Different K-factor values are considered, where $K = \frac{\nu^2}{2\sigma^2}$. Then, estimation errors are generated as zero mean complex Gaussian random variables with standard deviations $\sigma_{e,1}$, $\sigma_{e,2}$. Erroneous channel estimates are then quantized. Finally, assigned quantization levels are considered as keys and the KER values are obtained.

As the channel estimate error variance could be more dynamic in real-life, we consider a fixed Δ with varying σ_e values, and observe the impact of $\frac{\Delta}{\sigma_e}$. It can be seen from Fig. 2 that KER decreases as this ratio increases. In addition to joint estimation error, the $\frac{\sigma_{e,1}}{\sigma_{e,2}}$ ratio is an important factor. Nodes of different capabilities (such as base stations or user equipments) may have different channel estimation error variances due to practical impairments, such as the noise figure in RF amplifiers. To investigate a diverse set of use-cases, we consider three different scenarios. In the first scenario, we consider perfect channel knowledge at Alice while Bob estimates the channel erroneously ($\frac{\sigma_{e,1}}{\sigma_{e,2}} = 0$). In the second scenario identical nodes are considered ($\frac{\sigma_{e,1}}{\sigma_{e,2}} = 1$), where, both Alice and Bob estimate channel of equal quality. On the final scenario ($\frac{\sigma_{e,1}}{\sigma_{e,2}} = 100$), channel estimation error on Alice is more severe than that of Bob. It can be seen that increasing $\frac{\sigma_{e,1}}{\sigma_{e,2}}$ negatively affects KER. Another critical parameter is the K-factor of the

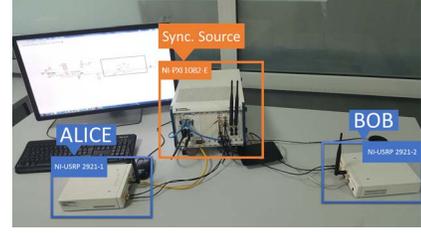
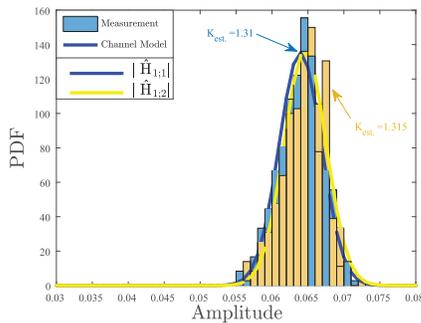


Fig. 3. SDR testbed for channel estimation and key extraction.

Rice channel. We select 3 different K-factor values as 0, 1 and 20 to test KER, as shown in Fig. 2 (a), (b) and (c). As the K-factor increases, lower Δ/σ_e ratios enable the same KER values. Simulation results are compared with the exact KER expression (as $\text{KER} = 1 - P_c^N$) in Eq. (4) and the approximate closed form expression in Eq. (9) for varying K , $\sigma_{e,1}$, $\sigma_{e,2}$ and Δ values. For Eq. (9) we observe that $M = 10$ can tightly approximate the KER expression. Lines show the approximate KER by Eq. (9) while the circles indicate numerical estimation from the exact expression in Eq. (4). The tightness of the proposed approximations can be observed from the figures. While higher $\sigma_{e,1}$, $\sigma_{e,2}$ and σ values degrade the KER performance, higher Δ and ν values reduce KER. The ratio of the quantization interval to the joint channel estimation error (σ_e) determines the robustness of the quantization against estimation errors.

IV. TESTBED DESCRIPTION AND MEASUREMENT RESULTS

USRP NI-2921 kits are used as SDR nodes in the test environment, along with LabVIEW software. The testbed of the implementation of point to point OFDM system is shown in Fig. 3. 360 subcarriers are transmitted by using a discrete Fourier transform of size of 480. The residual positions are zero padded. Quadrature phase shift keying (QPSK) is selected in tests, the carrier frequency is selected as 2.45 GHz. Transmission rate is 1 MS/sec and the cyclic prefix length is set to 120 samples. As the synchronization source NI PXI-6683 module is used. This module provides 10 MHz synchronization clock source from a GPS receiver that is connected to the master clock. The external clock signal is shared with the transmitter and the receiver nodes over the cable. At the receiver node, zero forcing approach is used for channel estimation. The overall transmission of a packet is completed in approximately 0.012 ms, consistent with the measurement

Fig. 4. Distributions of $|\hat{H}_{1,1}|$ and $|\hat{H}_{1,2}|$.TABLE I
MEASUREMENT RESULTS

Δ	Measurement	Eq. (9)	Relative error
0.0132	0.2436	0.2694	10.5%
0.0133	0.2344	0.2555	9%
0.0134	0.2248	0.2388	6.2%
0.0717	0.081	0.0866	6.9%
0.0718	0.07271	0.0766	5.3%
0.073	0.0577	0.0511	11.4%

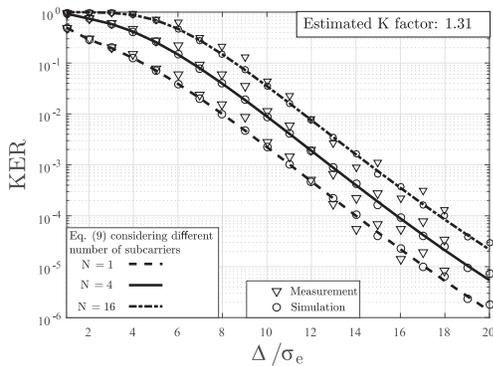


Fig. 5. Comparison of the obtained KER values.

studies in the literature to observe channel reciprocity [7], [14]. Keys are extracted every second. All subcarriers are used as pilot symbols in order to avoid any additional channel estimation error through the use of interpolation. As shown in Fig. 3, USRP node Alice is used as transmitter and USRP node Bob is used as receiver. Then, their roles are reversed. Transmission on both ways is repeated 1000 times. The distance between nodes is set to 50 cm.

Fig. 4 shows reciprocal channel gain histograms corresponding to first subcarriers and their estimated pdfs. Their distributions and K-factors slightly differ. The main reason for this variation is the estimation errors on both sides, along with the receiver front-end imperfections. The amplitudes of the estimated channel gain coefficients are quantized and KER values are measured for different Δ values. KER measurements are compared with theoretical KER values from Eq. (9) as given in Table I (where the channel estimation error variances are estimated as $\sigma_e = 0.005$ and $(\sigma_{e,1}/\sigma_{e,2} = 1)$). Due to the distribution of the channel coefficients, the slightest change in Δ affects KER significantly. The results show that KER can be accurately estimated through the provided expression.

In order to test the validity of the proposed approach, recorded coefficients are used as a reference for the system model given in Section II. The K-factor shown in Fig. 5 indicates that the reference channel gain estimates follow Rice distribution ($K = 1.31$). We generate complex Gaussian estimation errors and add them to recorded channel coefficients. Theoretical KER values are compared with the obtained KER values for the same set of parameters in Fig. 5. The slight difference between measurement and theoretical results are caused due to the erroneous channel coefficients, captured during transmissions. These results demonstrate the accuracy and the practicality of the obtained KER expressions. The obtained expression is critical to analyze the impact of the system parameters. KER can be further reduced by using information reconciliation techniques.

V. CONCLUSION

KER expressions are derived for complex Gaussian channels in presence of estimation errors. Theoretical expressions are verified by simulations. SDR based measurements demonstrate the applicability of the results.

REFERENCES

- [1] Y. Liu, H.-H. Chen, and L. Wang, "Physical layer security for next generation wireless networks: Theories, technologies, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 347–376, 1st Quart., 2017.
- [2] W. Diffie and M. Hellman, "New directions in cryptography," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 6, pp. 644–654, Nov. 1976.
- [3] G. S. Smith, "A direct derivation of a single-antenna reciprocity relation for the time domain," *IEEE Trans. Antennas Propag.*, vol. 52, no. 6, pp. 1568–1577, Jun. 2004.
- [4] J. Zhang, T. Q. Duong, A. Marshall, and R. Woods, "Key generation from wireless channels: A review," *IEEE Access*, vol. 4, pp. 614–626, 2016.
- [5] H. Liu, J. Yang, Y. Wang, and Y. Chen, "Collaborative secret key extraction leveraging received signal strength in mobile wireless networks," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, 2012, pp. 927–935.
- [6] B. Azimi-Sadjadi, A. Kiayias, A. Mercado, and B. Yener, "Robust key generation from signal envelopes in wireless networks," in *Proc. 14th ACM Conf. Comput. Commun. Security*, Alexandria, VA, USA, 2007, pp. 401–410.
- [7] J. Zhang, A. Marshall, R. Woods, and T. Q. Duong, "Efficient key generation by exploiting randomness from channel responses of individual OFDM subcarriers," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2578–2588, Jun. 2016.
- [8] J. W. Wallace and R. K. Sharma, "Automatic secret keys from reciprocal MIMO wireless channels: Measurement and analysis," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 381–392, Sep. 2010.
- [9] M. Guillaud, D. T. M. Slock, and R. Knopp, "A practical method for wireless channel reciprocity exploitation through relative calibration," in *Proc. ISSPA*, Sydney, NSW, Australia, 2005, pp. 403–406.
- [10] L. Berriche, K. Abed-Meraim, and J.-C. Belfiore, "Effect of imperfect channel knowledge on the MIMO channel outage capacity," in *Proc. IEEE SPAWC*, Cannes, France, 2006, pp. 1–5.
- [11] Y. Sun, A. Baricz, and S. Zhou, "On the monotonicity, log-concavity, and tight bounds of the generalized Marcum and Nuttall Q -functions," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1166–1186, Mar. 2010.
- [12] A. Baricz, "Tight bounds for the generalized Marcum Q -function," *J. Math. Anal. Appl.*, vol. 360, no. 1, pp. 265–277, 2009.
- [13] M. Abramowitz and I. A. Stegun, "Handbook of mathematical functions," *Appl. Math. Series*, vol. 55, no. 62, p. 39, 1966.
- [14] J. Zhang, R. Woods, T. Q. Duong, A. Marshall, and Y. Ding, "Experimental study on channel reciprocity in wireless key generation," in *Proc. IEEE SPAWC*, Edinburgh, U.K., 2016, pp. 1–5.