Secret Key Generation from Sparse Wireless Channels: Ergodic Capacity and Secrecy Outage

Tzu-Han Chou, *Member, IEEE,* Stark C. Draper, *Member, IEEE,*
and Akbar M. Sayeed, *Fellow, IEEE.*

***Abstract*—This paper investigates generation of a secret key from a reciprocal wireless channel. In particular we consider wireless channels that exhibit sparse structure in the wideband regime and study the impact of sparsity on the secret key capacity. We explore this problem in two steps. First, we study key generation from a *state-dependent discrete memoryless multiple source*. The state of the source captures the effect of channel sparsity. Secondly, we consider a wireless channel model that captures channel sparsity and correlation between the legitimate users' channel and the eavesdropper's channel. Such dependency can significantly reduce the secret key capacity.**

According to system delay requirements, two performance measures are considered: (i) ergodic secret key capacity and (ii) outage probability. We show that in the wideband regime when a white sounding sequence is adopted, a sparser channel can achieve a higher ergodic secret key rate than a richer channel can. For outage performance, we show that if the users generate secret keys at a fraction of the ergodic capacity, the outage probability will decay exponentially in signal bandwidth. Moreover, a larger exponent is achieved by a richer channel.

***Index Terms*—Secret key generation, public discussion, reciprocal wireless channel, channel sounding, ergodic capacity, secrecy outage.**

I. Introduction

The fundamental limit of secret key generation from discrete memoryless multiple source (DMMS) is developed by Ahlswede, Csiszár [1] and Maurer [2]. Their results show that if X, Y, Z (respectively observed by Alice, Bob and Eve) are correlated with a known distribution, it is possible to generate a secret key between Alice and Bob at a positive rate through public discussion. The resulting information rate leaked to Eve can be made arbitrarily small. The supremum of achievable secret key rates is called the *secret key capacity*.

Since their work, there have been many extensions to explore the secret key capacity of more complicated models. In [3, 4], users observe DMMS and also transmit information via wiretap channel [5], but there is no access to public channel for discussion. The authors in [6–9] consider a wiretap channel influenced by a random channel state, known by one (or both) of the legitimate users. In such models, the random channel state can be viewed as a kind of correlated source shared by receiver/transmitter which also influences the transmission medium and thus the secrecy capacity.

In [10, 11], key generation from DMMS is considered where, in [10], the DMMS is excited by a deterministic source and, in [11], by a random source. This sender-excited model is motivated by an application in which key generation is based on the inherent randomness of a reciprocal wireless channel. Consider a situation where Alice and Bob transmit a sounding signal to each other over a reciprocal wireless channel. Due to channel reciprocity, Alice and Bob observe a pair of dependent sources of randomness. Such a source turns out to be a good one for secret key generation. The first reason for this is that the medium is ubiquitous and wireless hardware is nowadays common. Thus, this general setting received much attention in terms of both theory and implementation [12–19]. A second reason for the strong interest in this modality is that such a source can be difficult to eavesdrop upon because the wireless channel can vary quickly in both spatial and temporal dimensions. Indeed, much prior work makes the assumption that the eavesdropper channel is statistically independent of the main channel (the channel between Alice and Bob). This assumption will be roughly correct when key generation occurs in a rich scattering environment. However, there is growing experimental evidence, e.g., [20–23], and physical arguments, e.g., [24–26], that show that realistic wireless channels are sparse at large bandwidths. The effect of channel sparsity on secret key capacity is twofold: first, it reduces the number of degrees-of-freedom[1] (DoF) of the main (Alice-Bob) channel and, second, it induces spatial correlation [27]. The first reduces the secret key capacity in the absence of an eavesdropper, while the second increases spatial correlation and therefore Eve's ability to observe the main channel.

We revisit the key generation problem when the channel exhibits sparsity in the wideband regime. We model such channel characteristics through a *sparsity pattern* that defines the support of the non-zero channel coefficients. Depending on the environment, the sparsity pattern can experience slow or fast time variation. The channel model we introduce also captures the dependence between the main channel and Eve's observations. To study secret key generation in this context define a *state-dependent* discrete multiple memoryless source (SD-DMMS). We particularize the model for the statistical characterization of sparse wireless channels where the sparsity pattern plays the role of the channel state and develop results on ergodic capacity and secrecy outage.

Manuscript received August 18, 2012; revised March 10, 2013. This work has been supported in part by the National Science Foundation under CAREER Grant No. CF-0844539 and Grant No. CNS-0627589.

The authors are with the Dept. of Electrical and Computer Engineering, University of Wisconsin, Madison, WI 53706. E-mail: thchou.th@gmail.com, sdraper@ece.wisc.edu, akbar@engr.wisc.edu. Stark Draper is also with the University of Toronto, ON, Canada.

[1] We define the number of degrees-of-freedom of a channel as the number of independent (non-zero) channel coefficients. These can be in the time, frequency, or spatial domains. For details see the channel model introduced in Section II-A.

In analogy to communication over a fading channel, two regimes are studied. The regime of interest will be a function of the system delay constraint:

- *Ergodic regime (the delay tolerant regime)*: If the key is generated based on observations across a large number of independently realized sparsity patterns, we will find that the *ergodic* secret key capacity is well-defined in the Shannon sense. A main technical constraint is that the system must be able to accommodate long delays.
- *Non-ergodic regime (the delay stringent regime)*: If the source sequence is not observed for a sufficiently long time or if the channel state chances so slowly that the key generation is forced to occur within a period of constant state, not known to the transmitter, the capacity in the Shannon sense is not defined. In this setting, we consider the *secrecy outage probability* which measures the probability that the instantaneous state condition cannot support the key rate while fulfilling the secrecy condition (we will later define this formally).

While both ergodic and non-ergodic secrecy have been studied before (the latter in the state-dependent, fading, wiretap channel (e.g., [28, 29]), the impact of channel sparsity has not been studied. Regarding the ergodic regime, we show that when a white sounding sequence is adopted in the wideband (low power) regime, a sparser channel can achieve a higher secret key rate than a richer channel can. This is analogous to capacity behavior in sparse multi-antenna channels in [30]. Furthermore, at each signal-to-noise ratio (SNR), there is an adequate bandwidth that maximizes the secret key rate. Regarding the non-ergodic regime, we show that the system can achieve an exponential decaying outage probability by using an α-*backoff* scheme ($0 < \alpha \leq 1$) in which secret key rate is a fraction α of the ergodic capacity. Unlike in the ergodic case, a richer channel always has a larger exponent characterizing the decay of the outage probability. In a similar vein as communication over a fading channel, this demonstrates that a large number of DoF helps to smooth out the effect of the unknown state.

The paper is organized as following. In Section II we give some definitions and describe the system model. This includes the correlated sparse wireless channel model, the definition of the SD-DMMS, and the one-way discussion key generation protocol. In Section III, we investigate the ergodic secret key capacity of SD-DMMS and apply this to key generation from a sparse wireless channel. Outage is defined in Section IV. We give a necessary and sufficient condition for an outage event and explore the outage probability when an α-backoff scheme is used. Detailed proofs are deferred to the Appendix.

II. Definitions and System Model

In this paper we are motivated by key generation based on wireless channel that exhibits sparsity in the delay domain. In Section II-A we first specify our model of sparse wireless channels. While earlier works on sparse wireless channels, e.g., see [20, 31–35], require the modeling of only a single channel, in Section II-A we need to model a pair of channels and their interaction: the main (Alice-to-Bob) channel and Eve's correlated observation of that main channel. Motivated by reciprocal wireless channels, in Section II-B we develop an abstracted *state-dependent discrete multiple source* (SD-DMMS) model. In this model the "state" captures the effect of the time-varying sparsity pattern while the key itself is extracted from the conditionally-generated (conditioned on the sparsity pattern) channel fades. Finally, in Section II-C the *one-way public discussion* key generation protocol is formally presented.

A. Sparse reciprocal wireless channel

Consider a wireless communication system with bandwidth W. Say that the channel exhibits sparsity in the delay domain[2] where $\tau_{\max}$ is the maximum delay spread of the channel. Then, $L_{\max} = \lceil \tau_{\max} W \rceil$ is the maximum number of resolvable paths. A sounding sequence

$$\mathbf{d} = [d_1, d_2, \cdots, d_{N_d}]^T \tag{1}$$

is transmitted over time period T_0, where $N_d = \lceil T_0 W \rceil$. The sounding sequence is a known sequence with power $\mathbf{d}^H \mathbf{d} = P$. We assume that each two-way (Alice $\leftrightarrows$ Bob) sounding is done within a channel coherence period (i.e., $T_{coh} \gg 2T_0$). Further multiple channel soundings (indexed by t) are performed within non-overlapping coherence periods. This means that the sets of soundings are jointly statistically independent.

The channel outputs in sounding interval t are

$$\boldsymbol{X}[t] = \mathbf{D}\boldsymbol{H}_{ab}[t] + \boldsymbol{W}_1[t] \qquad \text{(Alice)}\ , \tag{2a}$$

$$\boldsymbol{Y}[t] = \mathbf{D}\boldsymbol{H}_{ab}[t] + \boldsymbol{W}_2[t] \qquad \text{(Bob)}\ , \tag{2b}$$

where $\boldsymbol{H}_{ab}[t] = (H_{ab,1}[t], \cdots, H_{ab,L_{\max}}[t])^T$ is the vector of sampled (virtual) channel coefficients [24, 36], and $\mathbf{D}$ is an $N \times L_{\max}$ Toeplitz matrix where $N = N_d + L_{\max} - 1$:

$$\mathbf{D} = \left[\ \mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_{L_{\max}}\ \right]$$
$$= \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ d_2 & d_1 & \cdots & 0 \\ \vdots & d_2 & \cdots & d_1 \\ d_{N_d} & \vdots & & d_2 \\ \vdots & d_{N_d} & \ddots & \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & 0 & d_{N_d} \end{bmatrix} .$$

A widely used sounding signal is a sequence whose spectrum is asymptotically white in N_d. In this case $\mathbf{D}$ is a full column-rank matrix such that[3]

$$\mathbf{D}^H \mathbf{D} \doteq P \mathbf{I}_{L_{\max}} \tag{3}$$

when N_d is sufficiently large. One such example is $\mathbf{d} = \sqrt{P}\mathbf{e} = \sqrt{P}(1, 0, \cdots, 0)^T$. Another such example is a

[2] In this paper, we consider channel sparsity in the delay domain. It is not difficult to extend the result to the sparsity in either the Doppler or spatial domains, see, e.g., [24–26, 35].

[3] Here and in the following, we say $g(x^n) \doteq g$ if $g(x^n) \to g$ when n is sufficiently large.

pseudo-random (PN) sequence such as is used in spread spectrum system [37]. The noise terms $\boldsymbol{W}_1[t]$ and $\boldsymbol{W}_2[t]$ in (2) are, respectively, independent $\mathcal{CN}\left(\mathbf{0}, \sigma_a^2 \mathbf{I}_N\right)$ and $\mathcal{CN}\left(\mathbf{0}, \sigma_b^2 \mathbf{I}_N\right)$ vectors.

1) Sparse channel model: Most channels that have a small number of physical paths will exhibit sparsity in the delay domain as the signal bandwidth W increases. In particular, in some sets of delay bins, indexed by ℓ, the corresponding channel coefficients, $H_{ab,\ell}[t]$, will be zero. In this paper, we adopt the *sub-linear* law model considered in previous work [34, 35] to capture the sparse channel characteristic. In this model, the channel is called *δ-sparse* if the average number of non-zero channel coefficients scales as

$$L = (\tau_{\max} W)^{\delta} = L_{\max}^{\delta}, \quad \text{where} \quad \delta \in (0,1) \ . \tag{4}$$

The *channel sparsity pattern* of the main channel in sounding interval t is

$$\boldsymbol{S}_{ab}[t] = \left(S_{ab,1}[t], \cdots, S_{ab,L_{\max}}[t]\right) \in \mathcal{S}^{L_{\max}},$$

where the set $\mathcal{S} = \{0,1\}$ and $E\left[\sum_{\ell=1}^{L_{\max}} S_{ab,\ell}[t]\right] = L$. This pattern defines the support of the channel vector

$$\boldsymbol{H}_{ab}[t] = \left(H_{ab,1}[t],\ H_{ab,2}[t] \ \ldots\ H_{ab,L_{\max}}[t]\right).$$

In other words, $H_{ab,\ell}[t] = 0$ if and only if $S_{ab,\ell}[t] = 0$. The channel coefficients $H_{ab,\ell}[t]$ are independent $\mathcal{CN}\left(0, \nu_\ell^2\right)$ variables where the variance $\nu_\ell^2 = 0$ if $S_{ab,\ell}[t] = 0$. The channel has *unit* power, i.e., $\sum_\ell \nu_\ell^2 = 1$. Later, we use channel "degrees-of-freedom" (DoF) to refer to the *weight* of the realization of the vector $\boldsymbol{S}_{ab}[t]$. Thus, L is the expected number of DoF of the channel. We term $\boldsymbol{S}_{ab}[t]$ the *state* of $\boldsymbol{H}_{ab}[t]$. A *rich* multipath channel corresponds to $\delta \to 1$.

The sparsity pattern $\boldsymbol{S}_{ab}[t]$ will, in general, be time-varying. However, in most case of interest, $\boldsymbol{S}_{ab}[t]$ will change much more slowly than the channel coefficients $\boldsymbol{H}_{ab}[t]$. This is because the main reflectors, by which paths are resolved by different delay bins, move more slowly than the phase changes that induces the time-variations in fading coefficients [36, 38, 39]. The result of this disparity is that the secret key will mostly be derived from the randomness in channel coefficients rather than from that in the sparsity pattern (which has a much lower entropy rate). Furthermore, there exists good techniques to estimate the sparsity pattern reliably based on a few observations, e.g., [40]. Thus, we consider $\boldsymbol{S}_{ab}[t]$ to be known to Alice and Bob. Let T be the number of channel sounding periods during which the sparsity pattern remains constant. We term this the *sparsity coherence period.* Thus, the m-th sparsity coherent period extends from $t = (m-1)T + 1$ to $t = mT$. In this interval $\boldsymbol{S}_{ab}[t]$ remains constant, i.e., we assume that for all t, $(m-1)T + 1 \leq t \leq mT$, $\boldsymbol{S}_{ab}[t] = \boldsymbol{S}_{ab}[mT]$ (but $H_{ab,\ell}[t]$ changes for $S_{ab,\ell}[t] \neq 0$). We further assume that $\boldsymbol{S}_{ab}[t]$ is independent across periods.

Modeling the distribution of the state itself is a difficult task, so we consider a simple model

$$\Pr(S_{ab,\ell} = 1) = \frac{L}{L_{\max}} = (\tau_{\max} W)^{-(1-\delta)} \triangleq \rho \tag{5}$$

for all ℓ. In other words, the $S_{ab,\ell}$ is a binary random variable according to a Bernoulli distribution with parameter ρ (denoted $\mathrm{Bern}\left(\rho\right)$).

2) Eavesdropper's correlation model: Eve's channel output is similar to (2)[4]:

$$\boldsymbol{Z}[t] = \mathbf{D}\boldsymbol{H}_e[t] + \boldsymbol{W}_3[t] \qquad \text{(Eve)}\ , \tag{6}$$

where the noise is $\mathcal{CN}\left(\mathbf{0}, \sigma_e^2 \mathbf{I}_N\right)$. The channel coefficient vector $\boldsymbol{H}_e[t] = (H_{e,1}[t], \cdots, H_{e,L_{\max}}[t])^T$ is δ_e-sparse[5]. In general δ_e could be different from δ. The state of $\boldsymbol{H}_e[t]$ is denoted by $\boldsymbol{S}_e[t]$, and each element is distributed according to $\mathcal{CN}\left(0, \upsilon_\ell^2\right)$. Similar to (5), Eve's probability that $S_{e,\ell} \neq 0$ for any particular ℓ is denoted as ρ_e, i.e.,

$$\Pr(S_{e,\ell} = 1) = (\tau_{\max} W)^{-(1-\delta_e)} \triangleq \rho_e \tag{7}$$

We model the correlation between $\boldsymbol{H}_e[t]$ and $\boldsymbol{H}_{ab}[t]$ in a two-step process as follows:

- *Correlation between $\boldsymbol{S}_e$ and $\boldsymbol{S}_{ab}$*: For each delay bin ℓ, the probability that Eve has non-zero channel gain given $S_{ab,\ell}$ is
$$\Pr(S_{e,\ell} = 1 | S_{ab,\ell} = 1) = \theta \tag{8a}$$
$$\Pr(S_{e,\ell} = 1 | S_{ab,\ell} = 0) = \beta \tag{8b}$$
(see Fig. 1) for all $1 \leq \ell \leq L_{\max}$.
- *Correlation between individual channel coefficient*: For those channel coefficients in the "common support" delay bins, i.e., in the set $\{\ell : S_{ab,\ell} = S_{e,\ell} = 1\}$, the correlation coefficients are
$$\eta(H_{ab,\ell}, H_{e,\ell}) \triangleq \frac{E[H_{ab,\ell} H_{e,\ell}^*]}{\sqrt{E[|H_{ab,\ell}|^2] E[|H_{e,\ell}|^2]}} = \eta\ .$$

The parameter θ captures the expected fraction of coefficients that the main channel and Eve's channel have in common. One can think of the relationship between the state (i.e., the sparsity pattern) of the main channel and that of the eavesdropper's observation as a binary memoryless channel. However, because Eve's marginal channel is δ_e-sparse, the channel need not be symmetric so, in general, $\Pr(S_{e,\ell} = 1 | S_{ab,\ell} = 0) \neq \Pr(S_{e,\ell} = 0 | S_{ab,\ell} = 1)$. This is illustrated in Figure 1. Finally, the parameter η captures the effect that the paths (of both channel) located in the common delay bin shares the same physical scattering.

Remark 1: The model parameter space $\{(\theta, \eta), \delta, \delta_e\}$ captures many scenarios of interest. One important aspect to capture is the the correlation between Eve and main channels. There are two main considerations that factor into modeling

[4] In order to get meaningful observations, we assume Eve is located close to one of the users. So only one of the two Eve's channel outputs during the two-way sounding correlates with the main channel. The other output is independent of the main channel due to fast spatial decorrelation. The model could, of course, be extended to the situation where Eve gets two looks. Physically, Eve would have to control two monitors, one located near Alice, the other near Bob, effectively a scenario with two cooperating eavesdroppers.

[5] In the signal model, we assume Eve's channel has the same maximum delay spread $\tau_{\max}$ as the main channel for convenience. In sparse channels, a subset of coefficients from the set $\{1, 2, ..., L_{max}\}$ will be dominantly non-vanishing and the size of this non-vanishing support is more important. Effectively this difference in the two channels is reflected in the different sparsity parameters δ, δ_e.

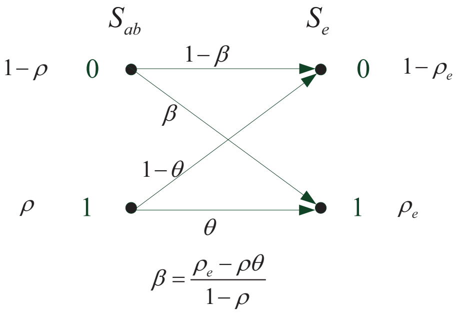

Fig. 1. Transition probability $\Pr(S_{e,\ell}|S_{ab,\ell})$.

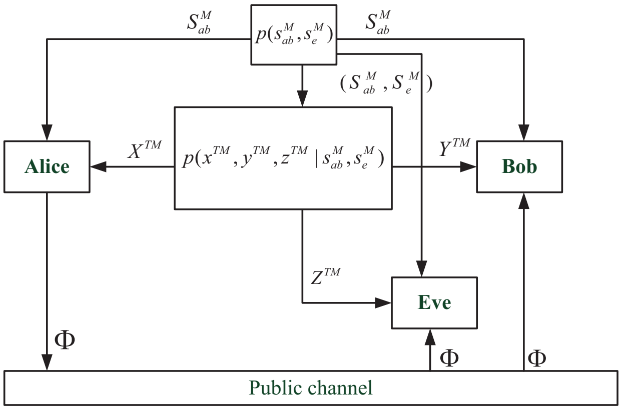

Fig. 2. State-dependent DMMS model.

this correlation. The first is the distance between Eve and Bob, which mainly affects the leakage to Eve. The second is the richness/sparseness of the multipath, which affects both the amount of leakage to Eve and the level of common randomness between Alice and Bob. When, for example, Eve gets close to Bob we would expect $\delta_e \to \delta$ and both θ and η to increase, increasing leakage. The parameter pair (δ, δ_e), and thus (ρ, ρ_e), controls the maximum number of DoF. As a second example, say that the separation between Eve and Bob and Alice is fixed. Then, for sparser channels, the common channel coefficients between Eve's channel and the main channel will generally exhibit a stronger correlation (larger η). The underlying reasons for this are a little subtle and are based on the idea of path partitioning, as discussed in [24, 36]. Essentially, in a sparse multipath environment, a common channel coefficient (with the same index in both channels) is likely to be generated by the same underlying physical propagation paths, leading to a higher correlation. Whereas, in a rich environment, with a very large number of propagation paths, there is more mixing of physical propagation paths across channel coefficients, which would generally result in a lower correlation between the common non-zero coefficients of Eve's channel and the main channel. We note that in the limit of extremely rich multipath, all channel coefficients will be non-zero and hence common to both Eve's and main channel and at the same time all coefficients will be uncorrelated due to a rich mixing of propagation paths across channel coefficients. The exact dynamics of these changes in θ and η as Eve moves would depend on the particulars of the physical environment.

To generate a secret key, users repeat the two-way channel sounding (2) (and (6)) TM times. The key is distilled from the super-block pair $\{(\boldsymbol{X}[t], \boldsymbol{S}_{ab}[t]), (\boldsymbol{Y}[t], \boldsymbol{S}_{ab}[t])\}_{t=1}^{TM}$. We concentrate on situations in which Eve also has access to $\boldsymbol{S}_{ab}[t]$. This is depicted in Fig. 2. This means our main results will be conservative in the sense that users cannot take advantage of the common randomness due to $\boldsymbol{S}_{ab}[t]$. However, as we will discuss after Corollary 2, this potential loss in secret key capacity will become negligible when the state changes sufficiently slowly. Furthermore, we consider the situation that given a power constraint, users can adapt their sounding power to prior knowledge of statistic of channel sparsity ρ. However, we don't consider a scenario in which the users' sounding policy can adapt to an instantaneous sparsity vector $\boldsymbol{S}_{ab}[t]$, since the $\boldsymbol{S}_{ab}[t]$ would be obtained after channel sounding.

In the following section we abstract away the actual sounding process and specify a state-dependent source model where the state varies more slowly than the underlying source-realization process from which the key is generated. When we study the ergodic case, we will let M grow arbitrarily large while we will hold T constant, a constant that indicates the number of source realizations when the state remains the same. When we study the outage case, $M = 1$, and T can be large.

B. *State-dependent discrete memoryless multiple source*

To leverage results on information theoretic security, we consider the state-dependent (SD) DMMS model depicted in Figure 2. The observation triple $(X^{TM}, Y^{TM}, Z^{TM}) \in \mathcal{X}^{TM} \times \mathcal{Y}^{TM} \times \mathcal{Z}^{TM}$ is generated according to $p(x^{TM}, y^{TM}, z^{TM}|s_{ab}^M, s_e^M)$, conditioning on the pair of length-M sequences: $(s_{ab}^M, s_e^M) \in \mathcal{S}^M \times \mathcal{S}^M$.

As discussed in Section II-A, S_{ab}^M is the state sequence of Alice and Bob's correlated source X^{TM}, Y^{TM} and S_e^M is the state sequence of Eve's observation Z^{TM}. The states have joint distribution $p(s_{ab}^M, s_e^M)$. Since the states vary more slowly than the conditionally-generated sources, there is a length of time T during which the states remain constant. This corresponds to the sparsity coherence period discussed earlier. A large T means that the states are changing slowly. As mentioned earlier, we assume that states are available to the corresponding observers and that, in addition, Eve also knows the state of the legitimate users. In other words, Alice and Bob both know S_{ab} but not S_e while Eve knows S_e and S_{ab}. This is depicted in Figure 2. Note that this is a conservative assumption while it let us to derive tight information leakage bounds for any state realization s_{ab} (see Section IV).

We call the state *memoryless* if

$$p(s_{ab}^M, s_e^M) = \prod_{m=1}^{M} p(s_{ab,m}, s_{e,m}) \ . \tag{9}$$

Similarly, the source is *memoryless* if

$$\begin{aligned} &p(x^{TM}, y^{TM}, z^{TM}|s_{ab}^M, s_e^M) \\ &\quad = \prod_{m=1}^{M} \prod_{i=(m-1)T+1}^{mT} p(x_i, y_i, z_i|s_{ab,m}, s_{e,m}) \ . \end{aligned} \tag{10}$$

Note that in (10) one sees the effect of the sparsity coherence period. The triplet of source samples (X_i, Y_i, Z_i) is conditionally and independently generated from the same state pair $(S_{ab,m}, S_{e,m})$ for all i, $(m-1)T < i \leq mT$. Each of (X_i, Y_i, Z_i) stands for the vector of channel output in II-A.

In the one-way discussion protocol (which will be detailed next in II-C), Alice sends a message Φ over a public channel. Bob recovers Alice's key based on his observation (Y^{TM}, S_{ab}^M) and Φ. Using Bayes' Rule we factor the conditional probability as

$$p(x,y,z|s_{ab},s_e) = p(x,y|s_{ab},s_e)p(z|x,y,s_{ab},s_e). \quad (11)$$

There are two types of degradedness we now define for the eavesdropper: state degradation and observation degradation. We will say that the an eavesdropper is degraded if *both* hold.

Definition 1 (Degradedness)**.** *We define the following notions of degradedness:*

1) *State degradedness: State S_e is a degraded version of S_{ab} with respect to (X, Y) if*

$$p(x,y|s_{ab},s_e) = p(x,y|s_{ab}) \ \forall \ x,y,s_{ab},s_e. \quad (12)$$

2) *Observation degradedness: The eavesdropper observation Z is a degraded version of Y if*

$$p(z|x,y,s_{ab},s_e) = p(z|y,s_{ab},s_e) \ \forall \ x,y,z,s_{ab},s_e. \quad (13)$$

3) *Degraded eavesdropper: An eavesdropper is said to be degraded if both (12) and (13) hold.*

If S_e is a degraded version of S_{ab} then we have the Markov relation $(X,Y) - S_{ab} - S_e$. We note that this is a natural condition for the sparse wireless channel model: S_e might give us some idea of which delay bins the non-zero reflectors are in, i.e., S_{ab}, but it is not likely to give us a sense of the small-scale fading effects that yield the non-zero values of the fading coefficients in those bins, i.e., X and Y. If Eve's observation Z is a degraded version of Bob's observation Y then, for any given states (s_{ab}, s_e), Eve's output is a cascade of Bob's output and a channel represented by $p(z|y, s_{ab}, s_e)$. This is natural due to the motivation of reciprocal wireless channel – anyone other than Alice and Bob will have a worse estimate of the channel between them. Combining these two aspects into the condition for a degraded eavesdropper means that we can rewrite (11) as

$$p(x,y,z|s_{ab},s_e) = p(x,y|s_{ab})p(z|y,s_{ab},s_e) \ . \quad (14)$$

C. One-way discussion key generation protocol

Let $\mathcal{K} = \{1, \ldots, 2^{TMR}\}$ be the key space. There is an authenticated public channel available to users to exchange error-free public messages in the set $\{1, \ldots, 2^{TMR_\phi}\}$. The one-way public discussion secret key generation protocol consists of three functions:

$$f_1 : \mathcal{X}^{TM} \times \mathcal{S}^M \to \mathcal{K} \ , \quad (15a)$$
$$g : \mathcal{X}^{TM} \times \mathcal{S}^M \to \{1, \ldots, 2^{TMR_\phi}\} \ , \quad (15b)$$
$$f_2 : \mathcal{Y}^{TM} \times \mathcal{S}^M \times \{1, \ldots, 2^{TMR_\phi}\} \to \mathcal{K} \ , \quad (15c)$$

which define Alice's key, public message, and Bob's key, respectively. Namely,

$$K = f_1(X^{TM}, S_{ab}^M) \ , \quad (16a)$$
$$\Phi = g(X^{TM}, S_{ab}^M) \ , \quad (16b)$$
$$\hat{K} = f_2(Y^{TM}, S_{ab}^M, \Phi) \ . \quad (16c)$$

Definition 2 (Achievability)**.** *For any positive integer T, a secret key rate R is (weakly) achievable if for any $\epsilon > 0$, there is a secret key generation system defined in* (15) *such that for sufficiently large M,*

$$\Pr(K \neq \hat{K}) < \epsilon \ , \quad (17)$$
$$\frac{1}{TM} I(K; Z^{TM}, \Phi, S_e^M, S_{ab}^M) < \epsilon \ , \quad (18)$$
$$\frac{1}{TM} H(K) > R - \epsilon \ . \quad (19)$$

Condition (19) means the key is almost uniformly distributed over the set $\mathcal{K}$. System secrecy is measured in terms of the mutual information defined in (18) which says that the information about the key leaked to eavesdropper is negligible. We note that in the above definition the units of key rate, and of leakage rate, are bits (or nats) *per channel usage*, i.e., $H(K)$ is normalized by TM. This will prove convenient in our discussions of both the ergodic and outage situations where we let M or T be arbitrarily large. The supremum of achievable secret key rates is called the *secret key capacity*[6].

III. Bounds on Ergodic Secret Key Capacity

For applications that can tolerate long delays, the key generation protocol can operate across a large number of independent state realizations. In this setting the parameter T, which models the number of channel uses across which each state realization remains constant, is kept fixed while M is allow to grow arbitrarily large. In this setting the secret key capacity in the Shannon sense is well-defined and is termed the *ergodic secret key capacity*.

Definition 3. *The ergodic secret key capacity C_{er} is the supremum of achievable secret key rates, i.e.,*

$$C_{\text{er}} = \sup\{R : R \textit{ is (weakly) achievable.}\} \quad (20)$$

A. Bounds on Ergodic Capacity of SD-DMMS

The theorems developed by Ahlswede, Csiszár [1, Theorem 1] and Maurer [2, Theorem 1,2] can be applied to the ergodic case of the source model in Figure 2 to get the following lemma.

Lemma 1.

$$C_{\text{er}}^- \leq C_{\text{er}} \leq C_{\text{er}}^+ \ , \quad (21)$$

where

$$C_{\text{er}}^+ = I(X;Y|Z,S_{ab},S_e) \ , \quad (22)$$
$$C_{\text{er}}^- = I(X;Y|S_{ab}) - I(X;Z,S_e|S_{ab}) \ . \quad (23)$$

[6] In contrast to the focus on weak secrecy in this paper, in [41, 42] results are developed on the strong secret key capacity, in which the error probability (17) and information leakage rate (18) decrease exponentially in block length.

Proof: The proof is given in Appendix A. ■

The following corollary says that the upper and lower bound equal one another when the eavesdropper's observation is degraded.

Corollary 2. *When the eavesdropper's source is degraded* (14)*, the ergodic secret key capacity is*

$$C_{\mathrm{er}} = I(X;Y|S_{ab}) - I(X;Z,S_e|S_{ab}) \ . \tag{24}$$

Proof: It can be verified by examining (22) that

$$\begin{aligned} C_{\mathrm{er}}^{+} &= I(X;Y,Z|S_{ab},S_e) - I(X;Z|S_{ab},S_e) \\ &= I(X;Y|S_{ab},S_e) + I(X;Z|Y,S_{ab},S_e) \\ &\quad - I(X;Z|S_{ab},S_e) \\ &\overset{(a)}{=} I(X;Y|S_{ab}) - I(X;Z|S_{ab},S_e) \\ &= I(X;Y|S_{ab}) - I(X;Z,S_e|S_{ab}) + I(X;S_e|S_{ab}) \\ &\overset{(b)}{=} I(X;Y|S_{ab}) - I(X;Z,S_e|S_{ab}) \\ &= C_{\mathrm{er}}^{-} \ , \end{aligned}$$

where the equality (a) is due to the degradedness conditions defined in (12) and (13) so that, respectively, $I(X;Y|S_{ab},S_e) = I(X;Y|S_{ab})$ and $I(X;Z|Y,S_{ab},S_e) = 0$. Equality (b) is also due to (12) so that S_e is conditionally independent of X given S_{ab}. ■

Remark 2: While in this paper we analyze the case when Eve knows the sparsity pattern of the main channel S_{ab}, for the ergodic case it is a simple exercise to extend the results to the case were Eve knows only S_e (and not S_{ab}). In the proofs of the appendix one simply does not include S_{ab} in Eve's observation set. The results come out similarly except that there is an additional, common, term in C_{er}, C_{er}^{+}, and C_{er}^{-}. This term is $\frac{1}{T}H(S_{ab}|Z^T,S_e)$ which is the entropy contained in the sparsity pattern observed by Alice and Bob, but not (in this setting) by Eve. Note that when the state changes slowly (or equivalently when T is large) $\frac{1}{T}H(S_{ab}|Z^T,S_e) \to 0$. In this situation the contribution to the secret key capacity due to the sparsity pattern S_{ab} becomes negligible, and the capacity approaches (from above) the capacity result of this section. As discussed in Section II-A this is a common situation.

B. Ergodic secret key rate of sparse wireless channel

We now apply Corollary 2 to the sparse channel model specified in II-A. In Section III-B1 we examine the expressions for mutual information $I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S}_{ab})$ and $I(\boldsymbol{X};\boldsymbol{Z},\boldsymbol{S}_e|\boldsymbol{S}_{ab})$ for the vector channel described by (2) and (6). Then, in Section III-B2 we identify conditions under which the eavesdropper's observation is degraded. Finally, in Sections III-B3 and III-B4 we focus on the randomness due to the sparsity patterns and analyze the wideband limit.

1) Mutual information: Define Q_ℓ to be the product $S_{ab,\ell} \times S_{e,\ell}$ so $Q_\ell \in \{0,1\}$. Thus $Q_\ell = 1$ if and only if the support (the sparsity pattern) of $\boldsymbol{H}_{ab}$ and of $\boldsymbol{H}_e$ are both non-zero in the ℓ-th delay bin. Also define two functions:

$$I_{ab}(\gamma_a,\gamma_b) = \log\left(\frac{(1+\gamma_a)(1+\gamma_b)}{1+\gamma_a+\gamma_b}\right) \ , \tag{25a}$$

$$I_e(\gamma_a,\gamma_e) = \log\left(\frac{(1+\gamma_a)(1+\gamma_e)}{1+\gamma_a\gamma_e(1-|\eta|^2)+\gamma_a+\gamma_e}\right) \ . \tag{25b}$$

where $\gamma_a = \frac{P}{\sigma_a^2}$, $\gamma_b = \frac{P}{\sigma_b^2}$ and $\gamma_e = \frac{P}{\sigma_e^2}$. We show in Appendix B that

$$I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S}_{ab}) = E\left[\sum_{\ell=1}^{L_{\max}} S_{ab,\ell} I_{ab}(\nu_\ell^2\gamma_a, \nu_\ell^2\gamma_b)\right] \ , \tag{26a}$$

$$I(\boldsymbol{X};\boldsymbol{Z},\boldsymbol{S}_e|\boldsymbol{S}_{ab}) = E\left[\sum_{\ell=1}^{L_{\max}} Q_\ell I_e(\nu_\ell^2\gamma_a, \upsilon_\ell^2\gamma_e)\right] \ . \tag{26b}$$

In the above expressions, the expectation is taken over the random length-$L_{\max}$ sparsity patterns $\boldsymbol{S}_{ab}$ and $\boldsymbol{S}_e$. Note the factor Q_ℓ in (26b). When $S_{ab,\ell} = 1$ but $S_{e,\ell} = 0$ the eavesdropper has no measurement of that channel coefficient ($Q_\ell = 0$). Thus, the eavesdropper has no observation of that common randomness and the negative mutual information term in (23) is zero.

It is clear in (26) that the channel sparsity patterns ($\boldsymbol{S}_{ab}$ and $\boldsymbol{S}_e$) affect the mutual information in two ways. The first is via the number of terms in the summation, i.e., the number of channel DoF. The second is via the correlation coefficient η which affects the information leakage through the $I_e(\cdot)$ in each delay bin observed by the eavesdropper.

2) Degraded condition: Because Eve's channel is correlated to the main channel, she may get a good estimation of $\boldsymbol{H}_{ab}$ if she has a higher SNR than Alice or Bob. To guarantee the positivity of the secret key rate, we need to characterize the conditions under which Eve's observation is of lower quality than either Alice's or Bob's. To develop such a condition we concentrate on delay bins ℓ where $Q_\ell = 1$. We note that in delay bins such that $Q_\ell = 0$ either $S_{ab,\ell} = 0$ or Eve has no observation of $H_{ab,\ell}$, so there is no need to consider those delay bins. Projecting the channel outputs onto $\mathbf{d}_\ell$, the ℓ-th column of $\mathbf{D}$, we get

$$X_\ell = \mathbf{d}_\ell^H \boldsymbol{X} \doteq P H_{ab,\ell} + W_{1,\ell}, \tag{27a}$$

$$Y_\ell = \mathbf{d}_\ell^H \boldsymbol{Y} \doteq P H_{ab,\ell} + W_{2,\ell}. \tag{27b}$$

Because the sounding signal is an (asymptotically) white sequence, X_ℓ (and Y_ℓ) form a sufficient statistic for estimating $H_{ab,\ell}$. The noise $W_{1,\ell}$ (resp. $W_{2,\ell}$) is a zero mean complex Gaussian with variance $P\sigma_a^2$ (resp. $P\sigma_b^2$). Similarly, Eve's sufficient statistic is

$$\begin{aligned} Z_\ell &= \mathbf{d}_\ell^H \boldsymbol{Z} \doteq P H_{e,\ell} + W_{3,\ell} \\ &\equiv P\left(\frac{\upsilon_\ell}{\nu_\ell}\eta H_{ab,\ell} + \sqrt{1-|\eta|^2} H'_\ell\right) + W_{3,\ell} \ . \end{aligned} \tag{28}$$

Because of $\eta(H_{ab,\ell}, H_{e,\ell}) = \eta$, we have equivalently written $H_{e,\ell}$ as a sum of two terms. The first term is a scaled version of $H_{ab,\ell}$. The second, H'_ℓ, is a $\mathcal{CN}\left(0,\upsilon_\ell^2\right)$ random variable that is independent of $H_{ab,\ell}$. We see from (28) that Eve's observation Z_ℓ contains two types of noise. The first is the receiver noise $W_{3,\ell}$. The second is due to the uncorrelated H'_ℓ.

Eve's observation Z_ℓ will be a degraded version of Y_ℓ if Eve has a lower SNR than Bob. This occurs if

$$\frac{\nu_\ell^2 P}{\sigma_b^2} > \frac{|\eta|^2 \upsilon_\ell^2 P}{(1-|\eta|^2)\upsilon_\ell^2 P + \sigma_e^2} \ . \tag{29}$$

Otherwise, Y_ℓ is a degraded version of Z_ℓ. If the sounding signal power is small and Eve has a suitably smaller noise variance σ_e^2, in particular, when

$$(1-|\eta|^2)\upsilon_\ell^2 P < |\eta|^2 \frac{\upsilon_\ell^2}{\nu_\ell^2}\sigma_b^2 - \sigma_e^2 \ , \tag{30}$$

then Bob's output is noisier and no secret key can be extracted at a positive rate from the ℓ-th delay bin. This is because when P is small, Eve's independent noise (due to H'_ℓ) is decreased. It is observed in [43] that there is a cutoff SNR below which the secret key capacity is zero. If $\nu_\ell^2 = \upsilon_\ell^2$ and all the users (Alice, Bob and Eve) are operating at the same SNR, i.e., $\sigma_a^2 = \sigma_b^2 = \sigma_e^2 = \sigma^2$, the secret key capacity will always be positive because Eve experiences additional noise (due to the uncorrelated H'_ℓ).

3) Achievable secret key rate: In order to see the effect of channel sparsity when the bandwidth is large (but finite), we focus on the equal-SNR case and consider a uniform delay profile. Per the above discussion this is a degraded eavesdropper setting. Define the random number of non-zero channel coefficients in the main Alice-to-Bob and in Eve's channel to be, respectively,

$$B_{ab} = \sum_{\ell=1}^{L_{\max}} S_{ab,\ell}, \tag{31a}$$

$$B_e = \sum_{\ell=1}^{L_{\max}} S_{e,\ell}. \tag{31b}$$

Note that B_{ab} and B_e are binomial distributions, respectively, $\text{Bino}\left(L_{\max},\rho\right)$ and $\text{Bino}\left(L_{\max},\rho_e\right)$. Consider a uniform delay profile, i.e., $\nu_\ell^2 = \frac{1}{B_{ab}}$ for all ℓ for which $S_{ab,\ell}=1$; similarly let $\upsilon_\ell^2 = \frac{1}{B_e}$ for all ℓ for which $S_{e,\ell}=1$.

Let $I_{\mathrm{s}}(P)$ be the instantaneous key rate $I(\boldsymbol{X};\boldsymbol{Y}|\boldsymbol{S}_{ab}) - I(\boldsymbol{X};\boldsymbol{Z},\boldsymbol{S}_e|\boldsymbol{S}_{ab})$ for fixed $\boldsymbol{S}_{ab}$ and $\boldsymbol{S}_e$. i.e.,

$$I_{\mathrm{s}}(\gamma) = B_{ab} I_{ab}\left(\frac{\gamma}{B_{ab}},\frac{\gamma}{B_{ab}}\right) - B_q I_e\left(\frac{\gamma}{B_{ab}},\frac{\gamma}{B_e}\right) \ , \tag{32}$$

where $\gamma \triangleq \frac{P}{\sigma^2}$ and

$$B_q = \sum_{\ell=1}^{L_{\max}} Q_\ell, \tag{33}$$

is the number of overlapping delay bins. From (26) and Corollary 2 the achievable secret key rate is

$$I_{\mathrm{er}}(\gamma) = E\left[I_{\mathrm{s}}(\gamma)\right] \ . \tag{34}$$

As we will see later in Section III-B4, $I_{\mathrm{s}}(\gamma)$ is convex in low SNR (as also is $I_{\mathrm{er}}(\gamma)$). From Corollary 2, $I_{\mathrm{er}}(\gamma)$ is the ergodic capacity when users use a uniform sounding signal with constant power P. If users are allowed to adapt their sounding scheme to the channel sparsity level, a uniform sounding strategy is not optimal. Let $\mathcal{P}$ denote all sounding policies that satisfy average power constraint $E[\mathbf{d}^H\mathbf{d}] \leq P$, we can achieve

$$C_{\mathrm{er}}(\gamma) = \max_{\mathcal{P}} I_{\mathrm{er}}(\gamma) \ . \tag{35}$$

Theorem 3 (Time-sharing sounding achieves capacity)**.** *Let the time-sharing parameters set*

$$\Omega = \left\{ (\lambda,\gamma_1,\gamma_2): \begin{array}{r} 0<\lambda\leq 1, \\ \gamma_1 \geq 0, \gamma_2 \geq 0 \\ \lambda\gamma_1 + \lambda\gamma_2 = \gamma \end{array} \right\} \ , \tag{36}$$

then the ergodic secret key capacity subject to average power constraint is

$$C_{\mathrm{er}}(\gamma) = \max_{(\lambda,\gamma_1,\gamma_2)\in\Omega} \lambda I_{\mathrm{er}}\left(\gamma_1\right) + (1-\lambda) I_{\mathrm{er}}\left(\gamma_2\right) \ . \tag{37}$$

Proof: Provided in Appendix C. ■

The physical interpretation of the auxiliary variable λ is that the key rate $\lambda I_{\mathrm{er}}\left(\gamma_1\right) + (1-\lambda) I_{\mathrm{er}}\left(\gamma_2\right)$ can be achieved by a time-sharing strategy that sounds the channel during some fraction λ, $0<\lambda\leq 1$, of the time at SNR γ_1 and the remainder the time at SNR γ_2. Theorem 3 says that the ergodic secret key capacity can be achieved by a λ^* time-sharing sounding strategy where $(\lambda^*,\gamma_1^*,\gamma_2^*)$ is the triplet that maximizes (37). As we discuss later in Section III-B4, in the wideband regime $I_{\mathrm{er}}(\gamma)$ is convex when γ is small, so a good choice is to set $\gamma_2 = 0$ and $\gamma_1 = \frac{\gamma}{\lambda_*}$. This corresponds to an *on-off* sounding strategy. One finds that at low SNRs (as $\gamma \to 0$) a sparser sounding pattern (i.e., $\lambda^* \to 0$) is better, while at high SNRs, a denser sounding pattern is better (i.e., $\lambda^* \to 1$).

4) Wideband regime: One way to increase the secret key capacity is to increase the bandwidth W of the wireless channel. However, in most real-world settings the channel DoF do not increase linearly in W. To see how W affects the secret key rate, we examine $C_{\mathrm{er}}(P)$ from (37) in the wideband regime.

In the wideband case, each channel DoF is sounded at a low SNR. At low SNR we can approximate (25) as

$$I_{ab}(x,x) \approx \frac{x^2}{\ln 2} \ , \tag{38a}$$

$$I_e(x,y) \approx \frac{|\eta|^2 xy}{\ln 2} \ . \tag{38b}$$

for x and y small. The ergodic key rate

$$\begin{aligned} I_{\mathrm{er}}(\gamma) &\approx \frac{1}{\ln 2} E\left[B_{ab}\left(\frac{\gamma}{B_{ab}}\right)^2 - B_q|\eta|^2 \frac{\gamma}{B_{ab}}\frac{\gamma}{B_e}\right] \\ &= \frac{\gamma^2}{\ln 2} E\left[\frac{1}{B_{ab}} - |\eta|^2 \frac{B_q}{B_{ab}}\frac{1}{B_e}\right] \\ &\overset{(a)}{\approx} \frac{\gamma^2}{\ln 2}\frac{(1-\theta|\eta|^2)}{L} = \frac{\gamma^2}{\ln 2}\frac{(1-\theta|\eta|^2)}{(\tau_{\max}W)^\delta} \ . \end{aligned} \tag{39}$$

The approximation (a) is accurate when $L \gg 1$ (cf. [44, eq.(5)]). The right hand side of (39) is a quadratic function of γ, thus $I_{\mathrm{er}}(\gamma)$ is convex in γ at low SNRs.

Figure 3 plots $I_{\mathrm{er}}(\gamma)$ versus γ, for a bandwidth of $W=100$ MHz, $\tau_{\max} = 10\mu$s, and for various values of the sparsity parameter $\delta \in [0.5,1]$. We see that a sparser channel (small δ) achieves a higher key rate at low SNRs. We can also observe this from (39). According to Theorem 3 and noticing that (39) is quadratic in γ, a sparser signal in time (an on-off signal) results in a higher secret key rate. In other words, in the wideband (power-limited) regime, fewer DoF (either in channel or in time domain) can achieve a higher key rate. This occurs

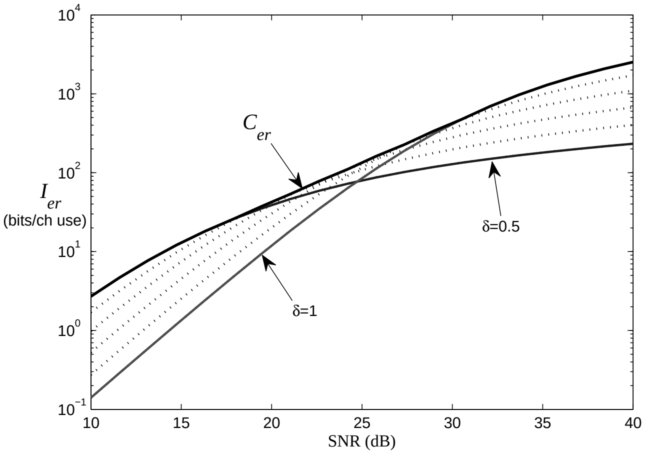

Fig. 3. The achievable secret key rate $I_{\mathrm{er}}(\gamma)$ plotted versus γ, the signal-to-noise ratio (SNR). The bandwidth used in the example is $W = 100$ MHz, the maximum delay spread is set to $\tau_{\max} = 10\mu$s, the conditional probability of overlap in $\boldsymbol{S}_{ab}$ and $\boldsymbol{S}_e$ is $\theta = 0.5$, and the correlation between channel coefficients is $\eta = 0.1$. The value of $I_{\mathrm{er}}(\gamma)$ is plotted for various sparsity levels $\delta \in \{0.5, 0.6, 0.7, 0.8, 0.9, 1\}$.

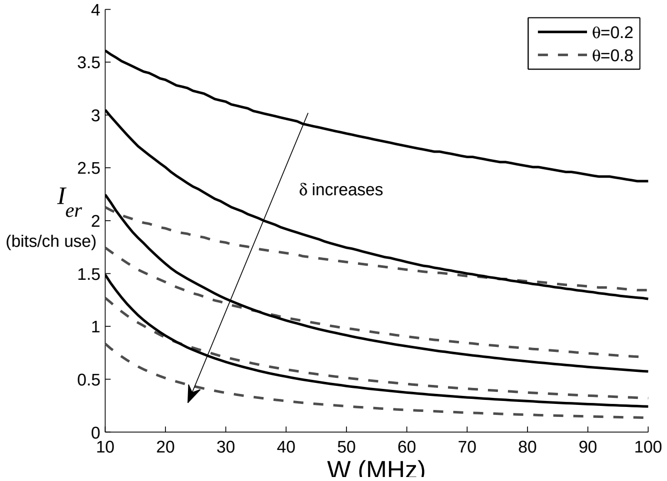

Fig. 4. Achievable secret key rate I_{er} plotted versus bandwidth W. This figure plots the tradeoff at a fixed SNR of 10 dB, $\eta = 0.8$, $\theta \in \{0.2, 0.8\}$, and for four values of the sparsity parameter $\delta \in \{0.5, 0.63, 0.76, 0.9\}$.

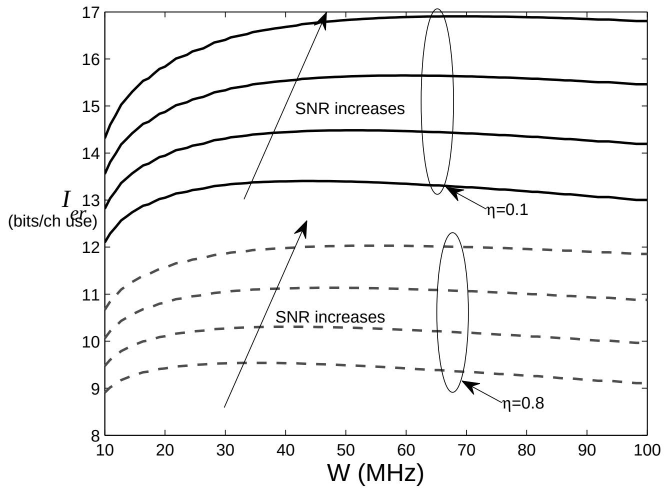

Fig. 5. Achievable secret key rate I_{er} plotted versus bandwidth W. This figure plots the tradeoff at a fixed sparsity $\delta = 0.5$ and for four SNRs $\gamma \in \{15\mathrm{dB}, 15.3\mathrm{dB}, 15.6\mathrm{dB}, 16\mathrm{dB}\}$, $\eta \in \{0.1, 0.8\}$.

because the key generation problem is a combined channel sounding and channel coding problem: channel sounding due to the channel excitation, and channel coding due to the de-noising by public messaging. By focusing energy on fewer DoF we raise their SNR, enabling key generation to occur at a higher rate. In contrast, a richer channel (large δ) results in a higher key rate at a high SNR since that is a DoF-limited (and not a power-limited) regime.

Figures 4 and 5 plot I_{er} as a function of W. This provides another view of the tradeoff between power and DoF. In Figure 4 we fix γ at 10 dB, $\eta = 0.8$, $\theta \in \{0.2, 0.8\}$ and plot I_{er} for different values of channel sparsity δ in the range $[0.5, 1)$. We note that in the wideband regime, i.e., at low SNRs, a larger δ results in a smaller key rate. Thus, in this regime a sparser channel is better. In Figure 5 we fix the channel sparsity $\delta = 0.5$ and plot I_{er} as a function of γ for SNRs ranging from 10dB to 30dB and $\eta \in \{0.1, 0.8\}$. We see that at each SNR, there is a unique channel bandwidth W^* that maximizes the key rate.

To close this section we make some notes on practical aspects of our setting. First, the capacity results assume knowledge of Eve's statistical correlations (θ, η). In the real world, users may not be able to know those information. However, (θ, η) can be considered as a part of threat model in a system to prioritize the system security level. The actual sounding signal (e.g., SNR, time-sharing) and key rate are determined according to the operating system security level.

IV. Secrecy Outage

In this section we consider an analysis appropriate to situations in which the communication system is subject to a stringent delay requirement. In particular say that the allowable delay is of the same order as the time-scale at which the sparsity pattern changes. In this situation the pattern is roughly constant during the key generation process. Since users know only their own state, and not Eve's, they cannot adapt their key generation process to Eve's state. Thus, satisfying the secrecy condition (18) is problematic and the secret key capacity in the Shannon sense is not well defined.

In this section we consider an outage formulation. To study this setting we set $M = 1$ while allowing T to be arbitrarily large. For any realization $(S_{ab}, S_e) = (s_{ab}, s_e)$ we say that a rate-R_e secrecy outage occurs if

$$\frac{1}{T} I(K; Z^n, \Phi | S_{ab} = s_{ab}, S_e = s_e) > R_e \tag{40}$$

for some $R_e > 0$. This means that there is a non-vanishing information rate leaked to Eve. For compactness in the rest of the paper, we adopt the notation $I(X; Y | S = s) = I(X; Y | s)$ and $H(X | S = s) = H(X | s)$, i.e., replacing $S = s$ with s.

Define two quantities,

$$C_s^+(s_{ab}, s_e) = I(X; Y|Z, s_{ab}, s_e), \tag{41a}$$

and

$$\begin{aligned} C_s^-(s_{ab}, s_e) =& I(X; Y|s_{ab}) - I(X; Z|s_e, s_{ab}) \\ &+ H(X|s_{ab}, s_e) - H(X|s_{ab}), \end{aligned} \tag{41b}$$

the conditional key rate bounds given state-pair (s_{ab}, s_e). The difference of conditional entropies $H(X|s_{ab}, s_e) - H(X|s_{ab})$ has to do with the information leakage given Eve's additional knowledge of s_e (over Alice and Bob's knowledge of only s_{ab}). We note the difference of conditional entropies cannot be combined into a single conditional mutual information term because they are conditioned on specific state realizations and thus taken with respect to different conditional probabilities ($p_{X|S_{ab}}(x|s_{ab})$ vs. $p_{X|S_{ab},S_e}(x|s_{ab}, s_e)$. This makes the lower bound slightly different from the ergodic expressions (23). However, when the state is degraded (12), the entropy terms cancel.

Theorem 4 provides upper and lower bounds on the rate of information leakage. Since both Alice and Bob know the realization of S_{ab} they can choose the key rate and the public message rate to be a function of s_{ab}. While this was also true in the ergodic case, there we were able to set the key and public messaging rate to satisfy the expected channel state conditions. In contrast, in the outage setting we experience only a single realization of the state. It will thus be helpful to indicate the dependence explicitly as we will study various choices for the rates. We denote the key rate as $R(s_{ab})$ and the public message rate as $R_\phi(s_{ab})$.

Theorem 4. *Consider any rate-$R(s_{ab})$ secret key generation system for which Bob can reliably recover X^T. This means that for any state $S_{ab} = s_{ab}$ known to both Alice and Bob, $\lim_{T\to\infty} \Pr(X^T \neq f_2(Y^T, s_{ab}, \Phi)) = 0$ for some sequence of functions $f_2(\cdot)$. Then, for any $\epsilon > 0$ and any realization $S_e = s_e$ of Eve's state:*

(i) The information leaked to Eve is lower bounded as

$$\begin{aligned} &\frac{1}{T} I(K; Z^T, \Phi|S_{ab} = s_{ab}, S_e = s_e) \\ &\quad \geq \frac{1}{T} H(K|s_{ab}, s_e) - C_s^+(s_{ab}, s_e) - \epsilon, . \end{aligned} \tag{42}$$

and

(ii) There exists a coding scheme that satisfies (17), (19) *and*

$$\begin{aligned} &\frac{1}{T} I(K; Z^T, \Phi|S_{ab} = s_{ab}, S_e = s_e) \\ &\quad \leq \frac{1}{T} H(K|s_{ab}, s_e) - C_s^-(s_{ab}, s_e) + \epsilon. \end{aligned} \tag{43}$$

Proof: The proof can be found in Appendix D. ■

When the eavesdropper is degraded, $H(K|s_{ab}, s_e) = H(K|s_{ab})$. Furthermore, $\frac{1}{T} H(K|s_{ab}) = R(s_{ab})$ since s_{ab} is used to set the key rate $R(s_{ab})$ but does not reveal any entropy about K itself (as it should not because it is also known to Eve). Furthermore, as we show below, $C_s^+(s_{ab}, s_e) = C_s^-(s_{ab}, s_e)$ for all pairs (s_{ab}, s_e). This means that, as is encapsulated in the next corollary, Theorem 4 gives a necessary and sufficient condition for outage.

Corollary 5. *In a secret key generation system with a degraded eavesdropper, if the users generate the secret key at rate $R(s_{ab})$, a rate-R_e outage occurs if and only if $R(s_{ab}) - C_s(s_{ab}, s_e) > R_e$ where*

$$C_s(s_{ab}, s_e) = I(X; Y|s_{ab}) - I(X; Z|s_{ab}, s_e). \tag{44}$$

We conclude the section by showing equality between $C_s^+(s_{ab}, s_e)$ and $C_s^-(s_{ab}, s_e)$ for degraded eavesdroppers.

$$\begin{aligned} &C_s^-(s_{ab}, s_e) \\ &= I(X; Y|s_{ab}) - I(X; Z|s_{ab}, s_e) + H(X|s_{ab}, s_e) - H(X|s_{ab}) \\ &\overset{(a)}{=} I(X; Y|s_{ab}, s_e) - I(X; Z|s_{ab}, s_e) \\ &\overset{(b)}{=} I(X; Y|s_{ab}, s_e) - I(X; Z|s_{ab}, s_e) + I(X; Z|Y, s_{ab}, s_e) \\ &= I(X; Y, Z|s_{ab}, s_e) - I(X; Z|s_{ab}, s_e) = C_s^+(s_{ab}, s_e). \end{aligned}$$

Equality (a) follows from the condition for state degradedness (12) and (b) from that for observation degradedness (13).

A. Wideband sparse channel

In the reciprocal wireless channel case, we know from Section III-B that $C_s(S_{ab}, S_e) = I_s(\gamma)$, as is given in (32). Using the approximations from (38), we have in the wideband regime that

$$C_s(\boldsymbol{S}_{ab}, \boldsymbol{S}_e) \approx \frac{\gamma^2}{\ln 2} \left(\frac{1}{B_{ab}} - |\eta|^2 \frac{B_q}{B_{ab}} \frac{1}{B_e} \right). \tag{45}$$

Recall that B_{ab} (resp. B_e) is the weight of the vector $\boldsymbol{S}_{ab}$ (resp. $\boldsymbol{S}_e$), cf. (31), and B_q is the weight of support common to $\boldsymbol{S}_{ab}$ and $\boldsymbol{S}_e$, cf. (33). Also note that B_{ab}, B_e, B_q are random variables, meaning that the overall quantity in (45) is also random. Unfortunately, there is no simple expression for the distribution of $C_s(\boldsymbol{S}_{ab}, \boldsymbol{S}_e)$ in (45). Since the users are assumed to know $\boldsymbol{S}_{ab}$ (and, therefore, B_{ab}), in order to understand how the channel sparsity affects the probability of outage, we consider the case where $B_{ab} = L$ (recall that in our model $L = E[B_{ab}]$ which is also the most likely value of B_{ab}), similarly we assume that $B_e = L$. Thus, the only uncertainty the users have is B_q. This is the number of delay bins from which Eve can learn the key. Conditioned on $B_{ab} = L$, and according to the correlation model of (8), B_q has a Binomial distribution $\text{Bino}(L, \theta)$.

Applying these assumptions to Theorem 4, the outage probability is

$$\begin{aligned} \mathsf{P}_{\text{out}} &= \Pr\left(R > C_s(\boldsymbol{S}_{ab}, \boldsymbol{S}_e)\right) \\ &\approx \Pr\left(R > \frac{\gamma^2}{L \ln 2}\left(1 - |\eta|^2 \frac{B_q}{L}\right)\right) \\ &= \Pr\left(B_q > \frac{1}{|\eta|^2}\left(1 - \ln 2 \frac{LR}{\gamma^2}\right)\right) \ , \end{aligned} \tag{46}$$

where $L = (W\tau_{\max})^\delta$. We see that a larger signal-to-noise ratio γ, a larger W, or a smaller η will decrease the outage probability. However, the sparsity δ changes both the distribution of B_q and the right hand side of the argument in (46) via L. Thus, it is not immediately clear how δ impact P_{out}. When users don't know the instantaneous secret key capacity, a conservative strategy is to generate a key at a smaller rate.

We now consider a strategy in which the key is generated at rate $R = \alpha I_{\rm er}(\gamma)$ where $0 < \alpha < 1$. This strategy backs off from the ergodic key rate (39) by a constant factor. We refer to this strategy as the "α-backoff" strategy. The outage probability (46) can now be simplified to be

$$\mathsf{P}_{\rm out} \approx \Pr\left(B_q \geq aL\right) \tag{47}$$
$$\text{where} \quad a = (1-\alpha)\frac{1}{|\eta|^2} + \alpha\theta \ .$$

Since B_q is approximately $\text{Bino}\,(L, \theta)$, the sparsity δ (and therefore the actual number of degrees-of-freedom L) determines the distribution of B_q, which characterizes Eve's ability to observe the main channel. From (47), we can see that when the α-backoff strategy is used, the correlation coefficient η determines how fast the threshold aL deviates from θL (the mean of B_q) as α is decreased. Note that in (47), the SNR (or, equivalently, the power P) does not appear in the formula. This is because the key rate is proportional to the ergodic key rate $I_{\rm er}(\gamma)$, which is a quadratic function of γ in the wideband regime, cf. (39), and thus cancels the γ^2 that appears in (46).

We next use results from large deviation theory [45] to upper bound the tail probability of a binomial distribution.

Lemma 6. *[45, Theorem 1] Let S_n be a binomial random variable $\text{Bino}\,(n,p)$. For $p < a < 1$, and for $n = 1, 2, 3, \cdots$, then*

$$\Pr(S_n \geq an) \leq 2^{-nD(a\|p)} \tag{48}$$

where

$$D(a\|p) \equiv a \log\frac{a}{p} + (1-a)\log\frac{(1-a)}{1-p} \tag{49}$$

is the Kullback-Leibler divergence between the probability distributions $\text{Bern}\,(a)$ and $\text{Bern}\,(p)$.

By this lemma, the outage probability is upper bounded as

$$\mathsf{P}_{\rm out} \leq 2^{-LD(a\|\theta)} \ . \tag{50}$$

Figure 6 plots the numerical results of the secrecy outage exponent $LD(a\|\theta)$ in the wideband regime. It shows that when the α-backoff strategy is used, the mechanism through which channel sparsity impacts the outage probability differs from how the channel sparsity impacts the ergodic secret key rate. A richer channel (larger δ) always has larger exponent than a sparser channel. In contrast, Figure 3 demonstrates a sparser channel yields a higher ergodic secret key rate in the wideband regime.

V. Conclusions and Future work

In this paper we study a setting in which two users desire to distill a common secret key based on the inherent randomness of a reciprocal wireless channel. Our particular interest is in the effect of channel sparsity, e.g., in the delay domain, on the reliable rate of secret key generation. Channel density often scales sub-linearly in signal bandwidth. This scaling affects the inherent randomness of the main channel and increases the eavesdropper's ability to observe the main channel. Since channel sparsity is an important characteristic of many real-world wireless channels, and since it has such a large impact on secret key capacity, it is crucial to understand better the interplay between channel sparsity and the achievable rates of secret key generation. Such understanding is crucial in being able to deliver secure communication systems with robust guarantees.

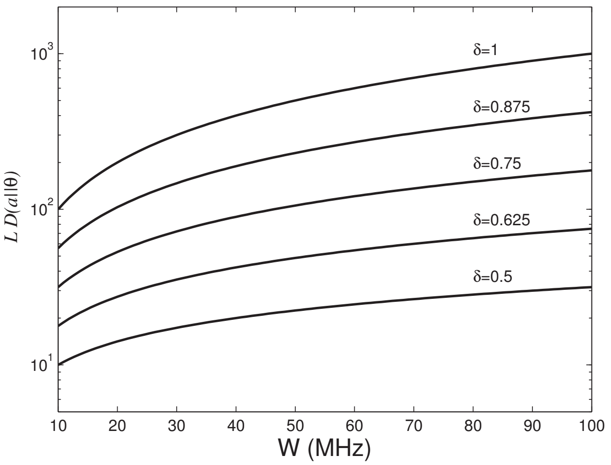

Fig. 6. Plot of outage exponent $LD(a\|\theta)$ vs. bandwidth W. Simulation parameters include the maximum delay spread $\tau_{\max} = 100$ns, the conditional probability of overlap in $\boldsymbol{S}_{ab}$ and $\boldsymbol{S}_e$, $\theta = 0.5$ and the correlation between channel coefficients, $\eta = 0.1$. The sparsity parameter δ is plotted for $\delta \in \{0.5, 0.625, 0.75, 0.875, 1\}$.

In this paper we consider two settings: the ergodic and non-ergodic regimes. In the ergodic setting, at each SNR there is an adequate bandwidth to maximizes the secret key rate. Moreover, we find that when a white sounding sequence is adopted in the wideband (low-SNR) regime, a higher secret key rate is achieved by a sparser channel. This is due to the focusing of the transmitted energy on the fewer non-zero channel coefficients. In the non-ergodic setting, for channels whose sparsity changes relatively slowly, we adopt a secrecy outage measure of performance. If the key rate is a fraction α of the ergodic capacity, we show that richer channels always have larger exponents which characterize the decay of the probability of outage. This results indicates that a larger number of degrees-of-freedom can smooth out the detrimental effects of an unknown eavesdropper state.

Interesting future directions include the characterization of two-way capacity. In this setting Bob is able to send public side information to Alice to help her recover Y^{TM}, in addition to Alice's message to Bob. The capacity of this model will, in general, be different from the one-way capacity characterized in this paper. A second extension is to random sounding. In the model considered herein the signal used to excite the reciprocal wireless channel is a deterministic sequence. If a randomized sounding signal were used (based on private sources of randomness) the randomization could be used to embed additional confidential. In [42] an initial investigation of this setting is made in the context of ergodic channels.

Appendix

A. Proof of Lemma 1

We first prove the lower bound (23). We start by applying the results of Theorem 1 in [1] where in the following we use $(\mathscr{U}, \mathscr{T}, \mathscr{Y}, \mathscr{Z})$ to correspond, respectively, to the random variables (U, T, Y, Z) used in [1]. We make the following choice of (auxiliary) random variables $\mathscr{U} = \emptyset$, $\mathscr{T} = (X^T, S_{ab})$, $\mathscr{Y} = (Y^T, S_{ab})$ and $\mathscr{Z} = (Z^T, S_e, S_{ab})$. We note that these choices include the state and the length-T channel outputs for each user during a single state coherent time. With these choices the achievable key rate *per length-T coherence time* is

$$\begin{aligned} & I(\mathscr{T}; \mathscr{Y}) - I(\mathscr{T}; \mathscr{Z}) \\ & \quad = I(X^T, S_{ab}; Y^T, S_{ab}) - I(X^T, S_{ab}; Z^T, S_e, S_{ab}). \end{aligned} \tag{51}$$

Using the chain rule and, in step (a) below, the conditionally memoryless nature of the source model from (10), the first term can be simplified as

$$\begin{aligned} & I(X^T, S_{ab}; Y^T, S_{ab}) \\ & = I(S_{ab}; Y^T, S_{ab}) + I(X^T; Y^T, S_{ab}|S_{ab}) \\ & = H(S_{ab}) + \sum_{i=1}^{T} I(X_i; Y^T|X^{i-1}, S_{ab}) \\ & \overset{(a)}{=} H(S_{ab}) + \sum_{i=1}^{T} I(X_i; Y_i|S_{ab}) \\ & = H(S_{ab}) + T \cdot I(X; Y|S_{ab}). \end{aligned}$$

Similarly, the second term can be simplified as

$$\begin{aligned} & I(X^T, S_{ab}; Z^T, S_e, S_{ab}) \\ & = I(S_{ab}; Z^T, S_e, S_{ab}) + I(X^T; Z^T, S_e|S_{ab}) \\ & = H(S_{ab}) + \sum_{i=1}^{T} I(X_i; Z^T, S_e|X^{i-1}, S_{ab}) \\ & \overset{(b)}{=} H(S_{ab}) + \sum_{i=1}^{T} I(X_i; Z_i, S_e|S_{ab}) \\ & = H(S_{ab}) + T \cdot I(X; Z, S_e|S_{ab}) \ , \end{aligned}$$

where (b) is again due to the memoryless property. Substituting into (51) and dividing it by T, we get the lower bound *per channel usage* to be

$$C_{\text{er}}^{-} = I(X; Y|S_{ab}) - I(X; Z, S_e|S_{ab}). \tag{52}$$

We next prove upper bound (22) by a sequence of (in)equalities. For any length-M secret key generation system where $P_e^{(M)} = \Pr[K \neq \hat{K}]$ we upper bound the entropy of K as

$$\begin{aligned} & H(K) = I(K; \hat{K}) + H(K|\hat{K}) \\ & \overset{(a)}{\leq} I(K; \hat{K}, \Phi, Z^{TM}, S_e^M, S_{ab}^M) + \epsilon_M \\ & = I(K; \hat{K}|\Phi, Z^{TM}, S_e^M, S_{ab}^M) + I(K; \Phi, Z^{TM}, S_e^M, S_{ab}^M) + \epsilon_M \\ & \overset{(b)}{\leq} I(K; \hat{K}|\Phi, Z^{TM}, S_e^M, S_{ab}^M) + TM\epsilon + \epsilon_M \\ & \overset{(c)}{\leq} I(K; Y^{TM}, S_{ab}^M|\Phi, Z^{TM}, S_e^M, S_{ab}^M) + TM\epsilon + \epsilon_M \\ & \leq I(K, \Phi; Y^{TM}, S_{ab}^M|Z^{TM}, S_e^M, S_{ab}^M) + TM\epsilon + \epsilon_M \\ & \overset{(d)}{\leq} I(X^{TM}; Y^{TM}|Z^{TM}, S_e^M, S_{ab}^M) + TM\epsilon + \epsilon_M \\ & = \sum_{i=1}^{TM} I(X_i; Y^{TM}|X^{i-1}, Z^{TM}, S_{ab}^M, S_e^M) + TM\epsilon + \epsilon_M \\ & \leq TM \cdot I(X; Y|Z, S_{ab}, S_e) + TM\epsilon + \epsilon_M. \end{aligned} \tag{53}$$

In (a), by Fano's Inequality $\epsilon_M < 1 + P_\epsilon^{(M)} TMR$. In (b) we apply the secrecy condition (18). In (c), $\hat{K}$ is a function of (Y^{TM}, S_{ab}^M) given Φ. In (d), (K, Φ) is a function of X^{TM} given S_{ab}^M. In (53) we apply the memoryless property of the state (9) and the conditional memoryless property of the source given the state (10). Dividing the result by TM on both sides, we get C_{er}^{+} in (22).

B. Derivation of mutual information (26)

First consider $I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{S}_{ab})$:

$$\begin{aligned} I(\boldsymbol{X}; \boldsymbol{Y}|\boldsymbol{S}_{ab}) & = E[h(\boldsymbol{X}|\boldsymbol{S}_{ab}) + h(\boldsymbol{Y}|\boldsymbol{S}_{ab}) - h(\boldsymbol{X}, \boldsymbol{Y}|\boldsymbol{S}_{ab})] \\ & = E\left[\log\left(\frac{\det(\mathbf{R}_{\boldsymbol{X}}) \cdot \det(\mathbf{R}_{\boldsymbol{Y}})}{\det(\mathbf{R}_{\boldsymbol{XY}})}\right)\right], \end{aligned} \tag{54}$$

where $h(\boldsymbol{X})$ is the differential entropy [46] of $\boldsymbol{X}$, and the expectation is taken over the distribution of $\boldsymbol{S}_{ab}$. Let $\mathbf{R}_{\boldsymbol{X}}$ denote the covariance matrix of $\boldsymbol{X}$ when the input $\boldsymbol{S}_{ab}$ is fixed to the sample vector $\mathbf{S}$, i.e., $\boldsymbol{S}_{ab} = \mathbf{S}$. We have

$$\mathbf{R}_{\boldsymbol{X}} = E[\boldsymbol{X}\boldsymbol{X}^H|\mathbf{S}] = \mathbf{D}\mathbf{R}_{\mathbf{h}}\mathbf{D}^H + \sigma_a^2 \mathbf{I}_N \ ,$$

where $\mathbf{R}_{\mathbf{h}} = \text{diag}(\nu_1^2, \cdots, \nu_{L_{\max}}^2)$ and $\nu_\ell^2 = 0$ if $S_\ell = 0$. Similarly we can calculate

$$\begin{aligned} \mathbf{R}_{\boldsymbol{Y}} & = \mathbf{D}\mathbf{R}_{\mathbf{h}}\mathbf{D}^H + \sigma_b^2 \mathbf{I}_N \\ \mathbf{R}_{\boldsymbol{XY}} & = \left[\begin{array}{c|c} \mathbf{R}_{\boldsymbol{X}} & \mathbf{D}\mathbf{R}_{\mathbf{h}}\mathbf{D}^H \\ \hline \mathbf{D}\mathbf{R}_{\mathbf{h}}\mathbf{D}^H & \mathbf{R}_{\boldsymbol{Y}} \end{array}\right] . \end{aligned}$$

We simplify the determinants as follows:

$$\begin{aligned} \det(\mathbf{R}_{\boldsymbol{X}}) & = \det(\mathbf{D}\mathbf{R}_{\mathbf{h}}\mathbf{D}^H + \sigma_a^2 \mathbf{I}_N) \\ & = (\sigma_a^2)^N \det\left(I_N + \frac{\mathbf{D}\mathbf{R}_{\mathbf{h}}\mathbf{D}^H}{\sigma_a^2}\right) \\ & \overset{(a)}{=} (\sigma_a^2)^N \det\left(I_{L_{\max}} + \frac{\Lambda\mathbf{D}^H\mathbf{D}\Lambda}{\sigma_a^2}\right) \\ & \overset{(b)}{=} (\sigma_a^2)^N \prod_{\ell=1}^{L_{\max}} \left(1 + \frac{P}{\sigma_a^2}\nu_\ell^2\right) \\ & = (\sigma_a^2)^N \prod_{\ell:S_\ell=1} \left(1 + \frac{P}{\sigma_a^2}\nu_\ell^2\right), \end{aligned}$$

where (a) follows by defining $\Lambda = \sqrt{\mathbf{R_h}}$ and by applying Sylvester's determinant formula: $\det(\mathbf{I}_m + \mathbf{AB}) = \det(\mathbf{I}_n + \mathbf{BA})$ where $\mathbf{A}$ and $\mathbf{B}$ are $m \times n$ and $n \times m$ matrices, respectively. Step (b) is due to (3). Similarly, we find that

$$\det(\mathbf{R}_{\boldsymbol{Y}}) \doteq (\sigma_b^2)^N \prod_{\ell:S_\ell=1} \left(1 + \frac{P}{\sigma_b^2}\nu_\ell^2\right)$$

$$\det(\mathbf{R}_{\boldsymbol{XY}}) \doteq (\sigma_a^2\sigma_b^2)^N \prod_{\ell:S_\ell=1} \left(1 + \frac{(\sigma_a^2+\sigma_b^2)P}{\sigma_a^2\sigma_b^2}\nu_\ell^2\right) .$$

Substituting into (54), we get (26a).

Follow a similar calculation, we get $I(\boldsymbol{X};\boldsymbol{Z},\boldsymbol{S}_e|\boldsymbol{S}_{ab})$ in (26b) by noting that

$$\mathbf{R}_{\boldsymbol{XZ}} = \left[\begin{array}{c|c} \mathbf{R}_{\boldsymbol{X}} & \mathbf{D}\mathbf{R}_{\mathbf{h}\tilde{\mathbf{h}}}\mathbf{D}^H \\ \hline \mathbf{D}\mathbf{R}_{\mathbf{h}\tilde{\mathbf{h}}}\mathbf{D}^H & \mathbf{R}_{\boldsymbol{Z}} \end{array}\right] .$$

where $\mathbf{R}_{\mathbf{h}\tilde{\mathbf{h}}}$ is a diagonal matrix and its ℓ-th diagonal element is equal to $\eta\nu_\ell^2$ if $Q_\ell = 1$ or is equal to zero if $Q_\ell = 0$.

C. Proof of Theorem 3

The proof is similar to [10, Theorem 4]. First note that $I_{\mathrm{s}}(\gamma)$ is non-decreasing in γ and so is $I_{\mathrm{er}}(\gamma)$. This can be verified by evaluating $\frac{\partial I_{\mathrm{s}}(\gamma)}{\partial\gamma}$, which is non-negative. Recall from (37) that

$$\bar{I}_{\mathrm{er}}(\gamma) = \max_{(\lambda,\gamma_1,\gamma_2)\in\Omega} \lambda I_{\mathrm{er}}(\gamma_1) + (1-\lambda) I_{\mathrm{er}}(\gamma_2) \ , \tag{55}$$

where Ω is defined in (36). Note that $\bar{I}_{\mathrm{er}}(\gamma)$ is also a non-decreasing function of γ. Since $I_{\mathrm{er}}(\gamma)$ is not a concave function its hypograph [47] **hypo** $I_{er}(\gamma) = \{(\gamma,t)|t \le I_{er}(\gamma)\}$ is not a convex set. Now, by construction we observe that the hypograph of $\bar{I}_{er}(\gamma)$ is the convex hull of **hypo** $I_{er}(\gamma)$. But, since $I_{\mathrm{er}}(\gamma)$ is concave for γ sufficiently large this tells us that, **hypo** $I_{er}(\gamma)$ is a convex set, therefore, $\bar{I}_{er}(\gamma)$ is concave.

We now show that $\bar{I}_{\mathrm{er}}\left(\frac{P}{\sigma^2}\right)$ is equal to $C_{\mathrm{er}}\left(\frac{P}{\sigma^2}\right)$ defined in (35) over the average power constraint P. Let $\mathcal{P}$ be the set of all sounding policies satisfying average power constraint P. Specifically, let the sounding policy in $\mathcal{P}$ allocate power P_s to sounding signals with probability $p(s)$ such that $E[P_s] = \sum_s p(s)P_s \le P$. Note that $C_{\mathrm{er}}\left(\frac{P}{\sigma^2}\right) \ge \bar{I}_{\mathrm{er}}\left(\frac{P}{\sigma^2}\right)$. We upper bound $C_{\mathrm{er}}\left(\frac{P}{\sigma^2}\right)$ as follows:

$$\begin{aligned} C_{\mathrm{er}}\left(\frac{P}{\sigma^2}\right) &= \max_{\mathcal{P}} \sum_s p(s) I_{\mathrm{er}}\left(\frac{P_s}{\sigma^2}\right) \\ &\le \max_{\mathcal{P}} \sum_s p(s) \left[\max_{(\lambda,\gamma_1,\gamma_2)\in\Omega} \lambda I_{\mathrm{er}}(\gamma_1) + (1-\lambda) I_{\mathrm{er}}(\gamma_2)\right] \\ &= \max_{\mathcal{P}} \sum_s p(s) \bar{I}_{\mathrm{er}}\left(\frac{P_s}{\sigma^2}\right) \\ &\overset{(a)}{\le} \max_{\mathcal{P}} \bar{I}_{\mathrm{er}}\left(\frac{\sum_s p(s)P_s}{\sigma^2}\right) \\ &\overset{(b)}{\le} \bar{I}_{\mathrm{er}}\left(\frac{P}{\sigma^2}\right) \ , \end{aligned}$$

where (a) and (b) are due to the concavity of $\bar{I}_{\mathrm{er}}(\gamma)$ and its non-decreasing character.

D. Proof of Theorem 4

We first show the lower bound (42):

$$\begin{aligned} &I(K;Z^T,\Phi|s_{ab},s_e) \\ &= H(K|s_{ab},s_e) - H(K|Z^T,\Phi,s_{ab},s_e) \\ &= H(K|s_{ab},s_e) - [H(K,\Phi|Z^T,s_{ab},s_e) - H(\Phi|Z^T,s_{ab},s_e)] \\ &= H(K|s_{ab},s_e) + H(\Phi|Z^T,s_{ab},s_e) \\ &\quad - [H(X^T,K,\Phi|Z^T,s_{ab},s_e) - H(X^T|Z^T,K,\Phi,s_{ab},s_e)] \\ &\ge H(K|s_{ab},s_e) + H(\Phi|Z^T,s_{ab},s_e) - H(X^T|Z^T,s_{ab},s_e) \ , \end{aligned} \tag{56}$$

The inequality holds since (K,Φ) are functions of (X^T,s_{ab}) and we dropped the non-negative term $H(X^T|Z^T,K,\Phi,s_{ab},s_e)$.

We now bound $H(\Phi|Z^T,s_{ab},s_e)$:

$$\begin{aligned} &H(\Phi|Z^T,s_{ab},s_e) \\ &= H(X^T,Y^T,\Phi|Z^T,s_{ab},s_e) - H(X^T,Y^T|Z^T,\Phi,s_{ab},s_e) \\ &\overset{(a)}{=} H(X^T,Y^T|Z^T,s_{ab},s_e) - H(X^T,Y^T|Z^T,\Phi,s_{ab},s_e) \\ &= H(X^T|Y^T,Z^T,s_{ab},s_e) + H(Y^T|Z^T,s_{ab},s_e) \\ &\quad - H(X^T,Y^T|Z^T,\Phi,s_{ab},s_e) \\ &\overset{(b)}{\ge} H(X^T|Y^T,Z^T,s_{ab},s_e) + H(Y^T|Z^T,\Phi,s_{ab},s_e) \\ &\quad - H(X^T,Y^T|Z^T,\Phi,s_{ab},s_e) \\ &= H(X^T|Y^T,Z^T,s_{ab},s_e) - H(X^T|Y^T,Z^T,\Phi,s_{ab},s_e) \\ &\overset{(c)}{\ge} H(X^T|Y^T,Z^T,s_{ab},s_e) - T\epsilon \ . \end{aligned} \tag{57}$$

Equality (a) holds because Φ is a function of X^T and S_{ab}. Inequality (b) holds because conditioning reduces entropy. Inequality (c) is due to the reliable condition $\lim_{T\to\infty} \Pr(X^T \neq f_2(Y^T,s_{ab},\Phi)) = 0$ and by applying Fano's inequality.

Substituting (57) into (56) we get

$$\begin{aligned} &I(K;Z^T,\Phi|s_{ab},s_e) \\ &\quad \ge H(K|s_{ab},s_e) + H(X^T|Y^T,Z^T,s_{ab},s_e) \\ &\qquad - H(X^T|Z^T,s_{ab},s_e) - T\epsilon \\ &\quad = H(K|s_{ab},s_e) - I(X^T;Y^T|Z^T,s_{ab},s_e) - T\epsilon \\ &\quad = H(K|s_{ab},s_e) - T\cdot I(X;Y|Z,s_{ab},s_e) - T\epsilon \ , \end{aligned}$$

where the memoryless property is applied in the last equality. This shows the lower bound (42).

We now show the upper bound (43):

$$\begin{aligned} &I(K;Z^T,\Phi|s_{ab},s_e) \\ &\quad = H(K|s_{ab},s_e) - H(K|Z^T,\Phi,s_{ab},s_e) \end{aligned}$$

We need to show that there exists a coding scheme such that

$$\frac{1}{T}H(K|Z^T,\Phi,s_{ab},s_e) \ge C_s^-(s_{ab},s_e) - 2\epsilon \ . \tag{58}$$

We will use the following lemma in the proof.

Lemma 7. *([11]) For any state realization of the main channel $S_{ab} = s_{ab}$ and any $\epsilon > 0$, if $R_\phi(s_{ab}) > H(X|Y,s_{ab})$, there exists a coding scheme with a public message of rate-$R_\phi(s_{ab})$ such that for T sufficiently large*

(i) $\Pr(X^T \neq f_2(Y^T, s_{ab}, \Phi)) \to 0$,
(ii) $\frac{1}{T} H(X^T | K, \Phi, Z^T, S_e, s_{ab}) \leq \epsilon$,

where $f_2(\cdot)$ is Bob's decoding function.

The proof of Lemma 7 uses a random coding technique to show existence. The first statement is exactly the Slepian-Wolf theorem [48]. The second statement is the equivocation analysis that says that if Eve knows K and Φ along with her observations (here Z^T and S_e), she can recover the sequence X^T. We refer the reader to [11, eq. (16)] for the details. [49, eq. (25)] also shows a similar result. We note that while S_e is random in the above lemma in the following we will need to make statements for any state realization $S_e = s_e$.

Adopting the coding scheme in Lemma 7 where the public message $\phi \in \{1, \ldots, 2^{TR_\phi(s_{ab})}\}$, we prove (58) through the following sequence of (in)equalities:

$$\begin{aligned}
&H(K|\Phi, Z^T, s_{ab}, s_e) \\
&= H(X^T, K|\Phi, Z^T, s_{ab}, s_e) - H(X^T|K, \Phi, Z^T, s_e, s_{ab}) \\
&\overset{(a)}{\geq} H(X^T, K|\Phi, Z^T, s_{ab}, s_e) \\
&\qquad - \frac{1}{\min_{s_e \in \mathcal{S}_e} p_{S_e}(s_e)} H(X^T|K, \Phi, Z^T, S_e, s_{ab}) \\
&\overset{(b)}{\geq} H(X^T, K|\Phi, Z^T, s_{ab}, s_e) - T\epsilon \\
&= H(X^T, K, \Phi|Z^T, s_{ab}, s_e) - H(\Phi|Z^T, s_{ab}, s_e) - T\epsilon \\
&\overset{(c)}{=} H(X^T|Z^T, s_{ab}, s_e) - H(\Phi|Z^T, s_{ab}, s_e) - T\epsilon \\
&\overset{(d)}{\geq} H(X^T|Z^T, s_{ab}, s_e) - T \cdot R_\phi(s_{ab}) - T\epsilon \\
&\overset{(e)}{=} H(X^T|Z^T, s_{ab}, s_e) - T \cdot H(X|Y, s_{ab}) - 2T\epsilon \\
&= T\big[H(X|Z, s_{ab}, s_e) - H(X|Y, s_{ab}) + H(X|s_{ab}, s_e) \\
&\qquad - H(X|s_{ab}, s_e) + H(X|s_{ab}) - H(X|s_{ab}) - 2\epsilon\big] \\
&= T[I(X;Y|s_{ab}) - I(X;Z|s_{ab}, s_e) + H(X|s_{ab}, s_e) \\
&\qquad - H(X|s_{ab}) - 2\epsilon].
\end{aligned}$$

In the above (a) follows from the following general statement. Consider any pair of random variable (U, V). Then, $H(U|V) = \sum_{v \in \mathcal{V}} p_V(v) H(U|V = v) \geq p_V(v_0) H(U|V = v_0)$ for any particular $v_0 \in \mathcal{V}$. If we assume that $\mathcal{V}$ is finite (or at least that it has a non-zero infimum), then turning the inequality around we have $H(U|V = v_0) \leq (1/\min_{v \in \mathcal{V}} p_V(v)) H(U|V)$. Making the proper mappings: U to X^T, V to S_e, v_0 to s_e and adding the rest of the conditioning results in inequality (a). Inequality (b) is due to part (ii) of Lemma 7 where we have folded the (constant) factor of $(1/\min_{s_e \in \mathcal{S}_e} p_{S_e}(s_e))$ into the ϵ; (c) follows because the pair (K, Φ) is a function of (X^T, s_{ab}); (d) comes from the bound on the alphabet size of ϕ. In (e) we choose $R_\phi(s_{ab}) = H(X|Y, s_{ab}) + \epsilon$. This is the smallest choice of rate that will ensure that the reconstruction constraint (part (i) of the Lemma) is satisfied. But, since the information leakage increases with increasing $R_\phi(s_{ab})$ we want to choose $R_\phi(s_{ab})$ as small as possible. We note that this choice of state-dependent rate, $R_\phi(s_{ab}) = H(X|Y, s_{ab}) + \epsilon$, is valid because Alice and Bob both know the realization $S_{ab} = s_{ab}$.

Acknowledgements

The authors would like to thank the anonymous reviewers for their detailed comments and thoughtful suggestions.

References

[1] R. Ahlswede and I. Csiszár, "Common randomness in information theory and cryptography part I: Secret sharing," *Information Theory, IEEE Transactions on*, vol. 39, no. 4, pp. 1121–1132, 1993.

[2] U. M. Maurer, "Secret key agreement by public discussion from common information," *Information Theory, IEEE Transactions on*, vol. 39, no. 3, pp. 733–742, 1993.

[3] A. Khisti, S. Diggavi, and G. Wornell, "Secret-key generation with correlated sources and noisy channels," in *Proc. Int. Symp. Inform. Theory*, pp. 1005–1009, July 2008.

[4] V. Prabhakaran, K. Eswaran, and K. Ramchandran, "Secrecy via sources and channels – a secret key-secret message rate tradeoff region," in *Proc. Int. Symp. Inform. Theory*, pp. 1010–1014, July 2008.

[5] A. Wyner, "The wire-tap channel," *The Bell Systems Technical Journal*, vol. 54, pp. 1355–1387, 1975.

[6] Y. Chen and A. Han Vinck, "Wiretap channel with side information," *Information Theory, IEEE Transactions on*, vol. 54, pp. 395–402, Jan. 2008.

[7] W. Liu and B. Chen, "Wiretap channel with two-sided channel state information," in *Proc. Asilomar Conf. Signals, Systems and Computers, 2007*, pp. 893–897, Nov. 2007.

[8] A. Khisti, S. Diggavi, and G. Wornell, "Secret key agreement using asymmetry in channel state knowledge," in *Proc. Int. Symp. Inform. Theory*, pp. 2286–2290, 2009.

[9] A. Khisti, S. Diggavi, and G. W. Wornell, "Secret-key agreement with channel state information at the transmitter," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 672–681, 2011.

[10] T. Chou, S. Draper, and A. Sayeed, "Key generation using external source excitation: Capacity, reliability, and secrecy exponent," *Information Theory, IEEE Transactions on*, vol. 58, pp. 2455–2474, Apr. 2012.

[11] T. Chou, V. Tan, and S. Draper, "The sender-excited secret key agreement model: Capacity theorems," in *Proc. Allerton Conference on Communication, Control, and Computing*, 2011.

[12] A. A. Hassan, W. E. Stark, J. E. Hershey, and S. Chennakeshu, "Cryptographic key agreement for mobile radio," *Digital Signal Processing*, vol. 6, no. 4, pp. 207–212, 1996.

[13] R. Wilson, D. Tse, and R. A. Scholtz, "Channel identification: Secret sharing using reciprocity in ultrawideband channels," *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 3, pp. 364–375, 2007.

[14] C. Ye, A. Reznik, G. Sternberg, and Y. Shah, "On the secrecy capabilities of ITU channels," pp. 2030–2034, Oct. 2007.

[15] C. Ye, S. Mathur, A. Reznik, Y. Shah, W. Trappe, and N. Mandayam, "Information-theoretically secret key generation for fading wireless channels," *Information Forensics and Security, IEEE Transactions on*, vol. 5, pp. 240–254, June 2010.

[16] M. A. J. Jon W. Wallace, Chan Chen, "Key generation exploiting MIMO channel evolution: Algorithms and theoretical limits," March 2009.

[17] N. Patwari, J. Croft, S. Jana, and S. K. Kasera, "High-rate uncorrelated bit extraction for shared secret key generation from channel measurements," *Mobile Computing, IEEE Transactions on*, vol. 9, pp. 17–30, Jan. 2010.

[18] T. Aono, K. Higuchi, T. Ohira, B. Komiyama, and H. Sasaoka, "Wireless secret key generation exploiting reactance-domain scalar response of multipath fading channels," *Antennas and Propagation, IEEE Transactions on*, vol. 53, pp. 3776–3784, Nov. 2005.

[19] A. Sayeed and A. Perrig, "Secure wireless communications:

Secret keys through multipath," in *ICASSP 2008. IEEE International Conference on*, pp. 3013–3016, 2008.

[20] A. Molisch, "Ultrawideband propagation channels-theory, measurement, and modeling," *Vehicular Technology, IEEE Transactions on*, vol. 54, pp. 1528–1545, sept. 2005.

[21] S. Ghassemzadeh, R. Jana, C. Rice, W. Turin, and V. Tarokh, "Measurement and modeling of an ultra-wide bandwidth indoor channel," *Communications, IEEE Transactions on*, vol. 52, pp. 1786–1796, Oct. 2004.

[22] Z. Yan, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," *University of Wisconsin-Madison, Tech. Rep.*, Feb. 2007.

[23] N. Czink, X. Yin, H. Ozcelik, M. Herdin, E. Bonek, and B. H. Fleury, "Cluster characteristics in a MIMO indoor propagation environment," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 1465–1475, Apr. 2007.

[24] A. Sayeed, "Sparse multipath wireless channels: Modeling and implications," in *Proc. ASAP*, 2006.

[25] W. Bajwa, A. Sayeed, and R. Nowak, "Sparse multipath channels: Modeling and estimation," in *Digital Signal Processing Workshop*, pp. 320–325, Jan. 2009.

[26] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE (special issue on Sparse Signal Processing)*, June 2010.

[27] W. Lee, "Effects on correlation between two mobile radio base-station antennas," *Communications, IEEE Transactions on*, vol. 21, pp. 1214–1224, Nov. 1973.

[28] X. Tang, R. Liu, P. Spasojevic, and H. Poor, "On the throughput of secure hybrid-ARQ protocols for Gaussian block-fading channels," *Information Theory, IEEE Transactions on*, vol. 55, pp. 1575–1591, April 2009.

[29] O. Gungor, J. Tan, C. Koksal, H. El-Gamal, and N. Shroff, "Secrecy outage capacity of fading channels," in *Information Sciences and Systems (CISS), 2012 46th Annual Conference on*, 2012.

[30] A. Sayeed and V. Raghavan, "Maximizing MIMO capacity in sparse multipath with reconfigurable antenna arrays," *IEEE Journal on Special Topics in Signal Processing (special issue on Adaptive Waveform Design for Agile Sensing and Communication)*, pp. 156–166, June 2007.

[31] S. Cotter and B. Rao, "Sparse channel estimation via matching pursuit with application to equalization," *Communications, IEEE Transactions on*, vol. 50, pp. 374–377, Mar 2002.

[32] C. Carbonelli, S. Vedantam, and U. Mitra, "Sparse channel estimation with zero tap detection," *Wireless Communications, IEEE Transactions on*, vol. 6, pp. 1743–1763, May 2007.

[33] W. Li and J. Preisig, "Estimation of rapidly time-varying sparse channels," *Oceanic Engineering, IEEE Journal of*, vol. 32, pp. 927–939, Oct. 2007.

[34] V. Raghavan, G. Hariharan, and A. Sayeed, "Capacity of sparse multipath channels in the ultra-wideband regime," *IEEE Journal on Special Topics in Signal Processing (special issue on Performance Limits of Ultra-Wideband Systems)*, pp. 156–166, June 2007.

[35] V. Raghavan and A. Sayeed, "Sublinear capacity scaling laws for sparse MIMO channels," *Information Theory, IEEE Transactions on*, vol. 57, pp. 345–364, Jan. 2011.

[36] A. M. Sayeed and T. Sivanadyan, "Wireless communication and sensing in multipath environments using multi-antenna transceivers," in *Handbook on Array Processing and Sensor Networks*, Ch.5, Wiley-IEEE Press, 2010.

[37] M. K. Simon, J. K. Omura, R. A. Scholtz, and B. K. Levitt, *Spread Spectrum Communications Handbook*. McGraw-Hill, 1994.

[38] J. Proakis, *Digital Communications*. McGraw-Hill, Aug. 2000.

[39] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[40] A. Fletcher, S. Rangan, and V. Goyal, "Necessary and sufficient conditions for sparsity pattern recovery," *Information Theory, IEEE Transactions on*, vol. 55, pp. 5758–5772, dec. 2009.

[41] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *Information Theory, IEEE Transactions on*, vol. 57, pp. 3989–4001, June 2011.

[42] T. Chou, V. Tan, and S. Draper, "The sender-excited secret key agreement model: Capacity and error exponents," *Submitted to Information Theory, IEEE Transactions on*, May 2012. also available at http://arxiv.org/abs/1107.4148.

[43] T. Chou, S. Draper, and A. Sayeed, "Impact of channel sparsity and correlated eavesdropping on secret key generation from multipath channel randomness," in *Proc. Int. Symp. Inform. Theory*, 2010.

[44] E. L. Grab and I. R. Savage, "Tables of the expected value of $1/X$ for positive Bernoulli and Poisson variables," *Journal of the American Statistical Association*, vol. 49, pp. 169–177, Mar. 1954.

[45] R. Arratia and L. Gordon, "Tutorial on large deviations for the binomial distribution," *Bulletin of Mathematical Biology*, vol. 51, no. 1, pp. 125–131, 1989.

[46] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley series in telecommunications, New York: Wiley, 2nd ed., 2006.

[47] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[48] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *Information Theory, IEEE Transactions on*, vol. 19, pp. 471–480, Jul 1973.

[49] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *Information Theory, IEEE Transactions on*, vol. 24, no. 3, pp. 339–348, 1978.

Tzu-Han Chou (S'06-M'12) was born in Taipei, Taiwan. He received the B.S. degree in electrical engineering and the M.S. degree in communication engineering, respectively, from National Taiwan University, and Ph.D degree in electrical and computer engineering at the University of Wisconsin, Madison.

He joined the Computer and Communications Research Labs, Industrial Technology Research Institute, Hsinchu, Taiwan and joined Sunplus Technology, Hsinchu, Taiwan, where he worked on GPRS/WCDMA baseband design. Currently, he is working in Qualcomm Inc, San Diego, CA. His research interests are in the area of wireless communication, information theory, and physical layer security.

Stark C. Draper (S'99-M'03) is an Associate Professor of Electrical and Computer Engineering at the University of Toronto, ON, currently on leave from the University of Wisconsin, Madison. He received the M.S. and Ph.D. degrees in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology (MIT), and the B.S. and B.A. degrees in Electrical Engineering and History, respectively, from Stanford University.

Before moving to Wisconsin, Dr. Draper worked at the Mitsubishi Electric Research Laboratories (MERL) in Cambridge, MA. He held postdoctoral positions in the Wireless Foundations, University of California, Berkeley, and in the Information Processing Laboratory, University of Toronto, Canada. He has worked at Arraycomm, San Jose, CA, the C. S. Draper Laboratory, Cambridge, MA, and Ktaadn, Newton, MA. His research interests include communication and information theory, error-correction coding, statistical signal processing and optimization, security, and application of these disciplines to computer architecture and semiconductor device design.

Dr. Draper has received an NSF CAREER Award, the 2010 MERL President's Award, the UW ECE Gerald Holdridge Teaching Award, the MIT Carlton E. Tucker Teaching Award, an Intel Graduate Fellowship, Stanford's Frederick E. Terman Engineering Scholastic Award, and a U.S. State Department Fulbright Fellowship.

Akbar M. Sayeed (S'89-M'97-SM'02-F'12) is Professor of Electrical and Computer Engineering at the University of Wisconsin-Madison. He received the B.S. degree from the University of Wisconsin-Madison in 1991, and the M.S. and Ph.D. degrees from the University of Illinois at Urbana-Champaign in 1993 and 1996, all in Electrical and Computer Engineering. He was a post-doctoral fellow at Rice University from 1996 to 1997. His research interests include wireless communications, statistical signal processing, communication theory, information theory, machine learning, time-frequency analysis, and applications in wireless communication and sensor networks.

Dr. Sayeed is a recipient of the Robert T. Chien Memorial Award (1996) for his doctoral work at Illinois, the NSF CAREER Award (1999), the ONR Young Investigator Award (2001), and the UW Grainger Junior Faculty Fellowship (2003). He has served the IEEE in a number of capacities, including as a technical program co-chair for the 2007 IEEE Statistical Signal Processing Workshop and the 2008 IEEE Communication Theory Workshop, as an elected member of the Signal Processing for Communications and Networking technical committee of the Signal Processing Society, and as a Guest Editor for special journal issues. He currently serves as an Associate Editor for the IEEE Transactions on Signal Processing.