Second Order and Moderate Deviation Analysis of a Block Fading Channel with Deterministic and Energy Harvesting Power Constraints

Deekshith P K, K Gautam Shenoy and Vinod Sharma
Department of ECE, IISc, Bangalore, India.
Email: {deekshithp, konchady, vinod}@iisc.ac.in

Abstract

We consider a block fading additive white Gaussian noise (AWGN) channel with perfect channel state information (CSI) at the transmitter and the receiver. First, for a given codeword length and non-vanishing average probability of error, we obtain lower and upper bounds on the maximum transmission rate. We derive bounds for three kinds of power constraints inherent to a wireless transmitter. These include the canonical peak power constraint and average power constraint, and a time varying peak power constraint imposed by an energy harvesting device—a mechanism that powers many modern-day wireless transmitters. The bounds characterize second order deviation of finite blocklength coding rates from the channel capacity, which is in turn achieved by water-filling power allocation across time. The bounds obtained also indicate the rate enhancement possible due to CSI at the transmitter in the finite blocklength regime. Next, we provide bounds on the optimal exponent with which error probability drops to zero when channel coding rate is simultaneously allowed to approach capacity at a certain rate, as the codeword length increases. These bounds identify what is known as the moderate deviation regime of the block fading channel. We compare the bounds numerically to bring out the efficacy of our results.

Index Terms

Finite block length regime, block fading, channel state information, water filling, energy harvesting, moderate deviation.
I. Introduction

In a wireless system, knowledge of instantaneous channel gains at the transmitter can be used to increase overall data rate by adaptive modulation and coding, and power control [1], [2]. For instance, depending on the fading dynamics, capacity of a flat fading AWGN channel with CSI at the transmitter (CSIT) and the receiver (CSIR) is enhanced using water-filling power control [3]. However, to capture rate enhancement with CSIT, using metrics like the water-filling capacity is fully justified only when there is no delay constraint in the system. Delay requirements are increasingly becoming stringent in 5G applications (see, for instance, [4]), and more refined metrics are required to characterize the performance of wireless systems under such circumstances. In this work, we consider a wireless system with perfect CSIT and CSIR subject to a delay requirement and certain power constraints. For such a system, we characterize the data rate enhancement with CSIT, and how quickly the error probability drops, when coding rate under power adaptation approaches the channel capacity, as delay tends to infinity.

Delay requirements in a wireless network can arise due to quality of service (QoS) constraints arising from the nature of data transmitted, readiness of the users to pay for extra resources, network architecture, storage limitations etc. All these factors determine at various levels the overall system performance. However, in this work we restrict to delay incurred at the physical layer in sending a codeword to the receiver.

With delay constraints imposed at the physical layer, the traditional approach to study rate enhancement due to CSIT (either perfect or imperfect) is to first characterize delay limited capacity, outage capacity or average capacity of the channel (see [5], Chapter 23 for details on these capacity notions) as per requirement. Once the characterization is in place, the approach is to obtain a power allocation strategy that maximizes the required capacity expression. In this regard, [6] obtains the optimal power allocation that maximizes the required capacity expression under the assumption of non-causal CSIT. In [7], the authors obtain the optimal power allocation scheme maximizing the average capacity with causal CSIT. Nonetheless, the above mentioned notions of capacity may not be realistic metrics to evaluate performance of delay sensitive systems with CSIT. This is because, the capacity notions are inherently asymptotic.

In this work, we provide bounds to characterize the rate enhancement due to CSIT over a block fading AWGN channel, at a given codeword length and average probability of error upto second order, using finite block length analyses pioneered in [8], [9], [10]. Further, we
also characterize rate at which error probability drops to zero when the channel coding rate approaches the capacity at a pre-specified rate, by conducting a moderate deviation analysis in the spirit of [11], [12]. We derive the bounds under three power constraints that are relevant to a wireless transmitter. Our assumption of perfect CSIT is idealistic. However, a wireless transmitter can acquire instantaneous channel gains either using channel reciprocity if the wireless system is time duplex with moderate mobility, or using a dedicated feedback link if the system is frequency duplex with a reasonable round trip delay. In addition, rates obtained under perfect CSIT assumption provide upper bounds for rates achievable without CSIT or with imperfect CSIT and many such characterizations are available in literature [3], [6], [7]. Also, knowledge of the power control strategies suitable for delay constrained systems sheds insights into how system energy is to be used in such systems. Efficient usage of system energy is essential for energy constrained transmitters which will become prominent in future wireless networks [13].

Using finite block length analysis, for instance in the spirit of [10], to characterize second order back off from the capacity of a fading channel has attracted a lot of research interest in the past. We will briefly review some of those results and discuss our contributions in that context. In [14], authors consider a scalar coherent fading channel with stationary fading (generalization of block fading) without CSIT and characterize dispersion of that channel. In [15], authors show that the second order optimal power allocation scheme over a quasi static fading channel with CSIT and CSIR is truncated channel inversion. Quasi static fading corresponds to block fading with a single block. The second order backoff therein is in terms of the number of channel uses within a block. Non-ergodic setting of communication within a single coherence interval is considered also in [16] and authors therein derive a finite block length version of outage capacity. A MIMO Rayleigh block fading channel with no CSIT and CSIR is considered in [17] and achievability and converse bounds are derived for the short packet communication regime. A high-SNR normal approximation of the maximal coding rate over a block fading Rayleigh channel without CSIT and CSIR is obtained in [18].

With regard to the moderate deviation analysis of communication channels, results for a discrete memoryless channel is obtained in [11] and for an AWGN channel in [12]. Exact moderate deviation constants are characterized for the corresponding channels in these works.

As mentioned, one of the power constraints under which we derive second order rate bounds is imposed by an energy harvesting device. In this context, capacity of a point-to-point real AWGN channel and a fast fading (i.e., a single channel use within a block) channel with energy
harvesting transmitter is obtained in [19]. In the spirit of [9], [10], the problem of characterizing the second order back-off from the capacity of a real AWGN channel with an energy harvesting transmitter is undertaken in [20], [21]. Even though, exact second order back-off coefficient of the channel model is still unknown, bounds on the same can be found in these works. However, such a finite codeword length analysis for a fading channel with an energy harvesting transmitter does not seem to be available.

Now we summarize the key contributions of this work.

- We provide lower and upper bounds on the maximal channel coding rate at a codeword length $n$ and average probability of error $\epsilon$, over a complex AWGN channel subjected to block fading, with CSIT and CSIR. This is a refinement and extension of our work on finite state fading block channel in [22]. Ours is a second order characterization in terms of the number of coherence blocks over which the communication spans. Such a characterization, in terms of the number of coherence blocks, is hitherto not available. Our characterization is in stark contrast to the non-ergodic setting of communication spanning a single block in [15], wherein the characterization is in terms of number of channel uses within a block. Further, our rate bounds characterize the back-off from the water-filling channel capacity in the finite block length regime. We also demonstrate numerically in Section VII that water filling power adaptation at finite blocklength can provide higher rates (upto second order) compared notions like to optimal delay limited power control [23]. In addition, we derive bounds separately for three kinds of power constraints wireless transmitters are normally subjected to namely, peak power constraint, average power constraint and a time varying peak power constraint inherent to an energy harvesting transmitter.

- In deriving the bounds, the CSIT assumption makes the analysis involved and non-trivial. In particular, we derive the bounds for block fading channel with general, complex-valued fading gains and hence the type based approach in [24] for finite state discrete memoryless renders infeasible. Further, in obtaining the upper bounds, the dependence of the channel input on the fading states makes the corresponding optimization problems difficult to solve. To circumvent this, we derive alternate bounds utilizing the properties of asymptotically optimal power allocation scheme, viz, the water filling scheme, McDiarmid’s inequality and strong approximation of partial sums of i.i.d. random variables [25]. We make use of the strong approximation result to bound the probability term in the relaxed meta converse (Lemma 2). This is in contrast to the Berry-Esseen approach for the finite state channel...
in [24]. We find that using strong approximation results is conceptually simpler and yields tighter upper bounds for the complex fading channel. An alternate approach is to discretize the state space, make use of the finite state approach and apply the limits to obtain the result for the general fading case. However, such an approach yields loose upper bounds. The use of strong approximation results to obtain finite block length bounds is novel in such a setting and can be of independent interest.

- We derive moderate deviation bounds in the spirit of [11], [12], for the block fading channel under consideration. These bounds characterize the optimal rate at which the probability of error converges to zero when the channel coding rate is allowed to scale at a particular rate as the codeword length increases. Such an analysis has not been previously undertaken for the canonical or the energy harvesting version of a block fading channel.

This paper is organized as follows. In Section II, we introduce the system model and notation. We provide lower and upper bounds on the maximal channel coding rate with peak codeword power constraint in Section III. Next, in Section IV we provide lower and upper bounds on the maximal channel coding rate with average codeword power constraint. Similar bounds for the case when the transmitter is harvesting energy from the environment are provided in Section V. Moderate deviation rate is obtained in Section VI. In Section VII we compare the bounds numerically and demonstrate the utility of the bounds derived. We conclude in Section VIII. Proofs are delegated to the appendices.

II. Model and Notation

A. Channel Model

We consider a point-to-point, discrete time, frequency non-selective block fading channel subject to additive, circularly symmetric complex Gaussian noise at the receiver. The probability density function (pdf) of the additive noise random variable with covariance matrix

\[
\begin{bmatrix}
\frac{\sigma_N^2}{2} & 0 \\
0 & \frac{\sigma_N^2}{2}
\end{bmatrix}
\]

is denoted as \(CN(0, \sigma_N^2)\). The additive noise process is independent and identically distributed (i.i.d.) across channel uses. A slot refers to the time period between successive channel uses; the period between \(i\)th and \((i + 1)\)th channel use is slot \(i\). A block refers to a time period of duration \(T_c\) (channel coherence time) over which the gain of the underlying wireless channel remains constant. Let the delay constraint imposed, by some application, e.g., voice, on the communication at the physical layer be \(D\) time units. In this work, we restrict exclusively to the time for information transmission. Thus, let \(D\) be the time left for transmission after taking care of
the transmission of training symbols for channel state estimation, synchronization, transmission of control information etc. We assume the transmitter has perfect knowledge of the channel state. Considering for convenience $D$ as an integer multiple of $T_c$, $B = D/T_c$ is the number of blocks over which the communication spans. Let $n_c$ denote the number of times the in-phase and quadrature channels are used within a block. Then, the number of complex channel uses for the whole of communication, or equivalently, the codeword length $n = B n_c$.

The channel gain or the fading coefficient in block $b$ is denoted as $H_b \in \mathbb{C}$, the set of complex numbers, such that $\mathbb{E} [ |H_b|^2] = \sigma_H^2 < \infty$. Here, $\mathbb{E} [\cdot]$ denotes the expectation operator and $|\cdot|$ denotes absolute value. We assume the channel gains are i.i.d. across blocks and is independent of the additive noise process. Let $F_H$ (known to both the transmitter and the receiver) denote the cumulative distribution function (cdf) common to all $H_b, b \in [1 : B]$. (Here, for natural number $m$, $[m : m + \ell - 1]$ denotes the set of $\ell$ consecutive natural numbers from $m$.) The instantaneous channel gains are assumed to be known to the transmitter as well as the receiver and the transmitter gets to know them only causally. We refer to this as the full CSIT and CSIR assumption. Even though, in practice, a wireless transmitter will not have perfect channel state information, rates obtained under such an assumption provide upper bounds for rates with partial CSIT. Further, our analysis can be considered as the first step in understanding the possible reduction in rate under various partial CSIT assumptions like quantized state, minimum mean square error estimate of the state etc. Also, for the energy harvesting case considered later, our analysis yields results for no CSIT and full CSIT cases either of which are otherwise unknown.

Let $X_{(b-1)n_c+k}$ denote the channel input corresponding to the $k$th channel use in the $b$th block, where, $k \in [1 : n_c], b \in [1 : B]$. For convenience, from here on, $[b, k] \triangleq (b-1)n_c+k$. Let $Z_{[b,k]}$ and $Y_{[b,k]}$ denote the corresponding noise variable and the channel output, respectively. Then, $Y_{[b,k]} = H_b X_{[b,k]} + Z_{[b,k]}$. For this channel model, if the number of blocks $B$ tends to infinity (and hence $D \to \infty$), it is well known that the channel capacity is given by

$$ C(P) \triangleq \mathbb{E} \left[ \log \left( 1 + \frac{|H_1|^2 P_{WF}(|H_1|)}{\sigma_N^2} \right) \right], \quad (1) $$

where, $P_{WF}(|H_1|) \triangleq \left( \lambda - \frac{\sigma_N^2}{|H_1|^2} \right)^+, (\cdot)^+ = \max(0, \cdot)$ (see [5], Section 23.2 wherein the result is given for a real block fading channel and [26], Section 5.4.6 for a complex block fading channel with $n_c = 1$). Further, $C(P)$ is the capacity under an average power constraint and $\lambda$ is obtained by solving the equation $\mathbb{E} [P_{WF}(|H_1|)] = \bar{P}$. Here, $P_{WF}(\cdot)$ is called the water-filling power allocation with average power $\bar{P}$. 


B. Transmitter Model

Let $S$ be the message to be transmitted, chosen randomly uniformly from $[1 : M]$. The message is encoded into a codeword $X' = (X'_1, \ldots, X'_n)$, possibly randomly chosen. Let $X = (X_1, \ldots, X_n)$ be the corresponding channel input vector. As we explain below, $X'$ and $X$ can possibly be different owing to power control, energy unavailability etc. Also, in a wireless transmitter, there can be various kinds of constraints on $X$. Next, we explain these constraints.

In practice, a wireless transmitter is subjected to various kinds of power constraints. Inherently, there is a restriction on the maximum power that can be expended, due to circuitry limitations, regulatory requirements etc. We refer to such a constraint as the peak power constraint. If the transmitter has knowledge about the gain of the wireless channel, it can adjust the transmit power according to the channel gain. However, the average power that can be expended over all channel gains is usually bounded. This constraint, referred to as the average power constraint, corresponds to the long term power utilization efficiency of the system. Though in reality these constraints are simultaneously present, in this work, as is the usual practice (see [14] for an instance of a peak power constrained transmitter and [15] for an average power constrained transmitter), we study them only in isolation. In addition, we also consider an energy harvesting transmitter motivated by the fact that such communication systems are increasingly becoming popular [27]. We will consider this in more detail later.

1) Peak Codeword Power Constrained Transmitter: Given the message $S$ to be transmitted and the fading gains $(H_1, \ldots, H_b) \equiv H^{(b)}$ till block $b \in [1 : B]$ (we denote the corresponding realizations as $h^{(b)}$), at $[b, k]$th channel use, the encoder chooses the codeword symbol $X'_{[b,k]} \equiv X'_{[b,k]}$. In certain circumstances, especially while proving lower bounds on the maximal achievable rate, it is convenient to consider that the transmitter encodes the message independent of channel gain realizations and later adapt the power of codeword symbols using a power controller. Such a scheme is attractive from an implementation perspective as well, as it eliminates the need for variable rate coding by multiplexing several codebooks (see [28], Section IV). In such a case, $X'_{[b,k]}$ will be different from $X_{[b,k]} \equiv X_{[b,k]}$, the corresponding channel input symbol. When the encoding is dependent on the causally available channel gains and hence, there is no separate power control, $X = X'$. We say that a transmitter has peak codeword power constraint (abbreviated as PP), if, for each $m \in [1 : M]$ and every realization
\( \mathbf{h} \triangleq (h_1, \ldots, h_B) \) of the fading vector \( \mathbf{H} \triangleq (H_1, \ldots, H_B) \), the channel input vector \( \mathbf{X} \) satisfies
\[
\sum_{b=1}^{B} \sum_{k=1}^{n_c} |X_{[b,k]}(m, h^{(b)})|^2 \leq Bn_c \bar{P}.
\] (2)
The above inequality holds almost surely (a.s.) with respect to the underlying joint distribution.

2) Average Codeword Power Constrained Transmitter: We say that a transmitter has average codeword power constraint (AP), if for each \( m \in [1 : M] \), the channel input vector \( \mathbf{X} \) satisfies
\[
\mathbb{E} \left[ \sum_{b=1}^{B} \sum_{k=1}^{n_c} |X_{[b,k]}(m, H^{(b)})|^2 \right] \leq Bn_c \bar{P}.
\] (3)
Here the expectation is with respect to the fading gain vector \( \mathbf{H} \).

3) Energy Harvesting Transmitter: In this case, the transmitter is connected to an energy harvesting device that provides power to transmit codeword symbols. Prior to \( [b, k]^{th} \) channel use, the device harvests energy \( E_{[b,k]} \geq 0 \) from its ambient environment. The harvested energy \( E_{[b,k]} \) is made available for transmission in the same slot itself. We assume the amount of energy harvested is i.i.d. across slots, with mean \( \bar{E} \), variance \( \sigma^2_E \) and \( \mathbb{E}[E_{[1,1]}^4] < \infty \).

The harvesting device is equipped with an energy buffer (equivalently, an accumulator or a battery) which we simply refer to as a buffer. The buffer stores the residual energy after \( [b, k]^{th} \) transmission for future use. This model is often referred to as the harvest-use-store model [19]. We assume the buffer has infinite energy storage capacity. This is a simplifying assumption and is usually made in literature [19], [29]. Prior to \( [b, k]^{th} \) transmission, given \( S, H^{(b)} \) and energy realizations \( (E_{[1,1]}, \ldots, E_{[b,k]}) \) \( \equiv E^{([b,k])} \), the encoder chooses \( X'_{[b,k]}(S, E^{([b,k])}, H^{(b)}) \equiv X'_{[b,k]} \) as the \( [b, k]^{th} \) codeword symbol. The exact symbol transmitted in \( [b, k]^{th} \) slot could be different from \( X'_{[b,k]} \). This is due to the fact that the transmitter is harvesting energy and hence, may not have enough energy to transmit \( X'_{[b,k]} \) in that slot. The corresponding channel input is \( X_{[b,k]}, X'_{[b,k]} \in \mathbb{C} \).

Let \( A_{[b,k]} \) denote the energy available in the buffer at the beginning of \( [b, k]^{th} \) transmission. We assume that the buffer is initially empty, i.e. \( A_{[1,1]} = 0 \). Since we consider the harvest-use-store model, the total energy available to the transmitter at the beginning of \( [b, k]^{th} \) transmission is \( E_{[b,k]} + A_{[b,k]} \equiv \hat{A}_{[b,k]} \). The buffer state \( A_{[b,k]} \) evolves according to \( A_{[b,k+1]} = \hat{A}_{[b,k]} - |X_{[b,k]}|^2, \ k \in [1 : n_c - 1], \ n_c \neq 1, \) and \( A_{[b+1,1]} = \hat{A}_{[b,n_c]} - |X_{[b,n_c]}|^2, \ k = n_c, \) where \( |X_{[b,k]}|^2 \leq \hat{A}_{[b,k]} \). This recursive relation and \( A_{[b,k]} \geq 0 \) a.s. for all \( b, k \) yields the energy harvesting constraint (EH)
\[
\sum_{\ell=1}^{b} \sum_{i=1}^{k} |X_{[\ell,i]}|^2 \leq \sum_{\ell=1}^{b} \sum_{i=1}^{k} E_{[\ell,i]}, \ \text{a.s.}
\] (4)
The decoder used in obtaining our lower bounds is same as that used in the $\beta\beta$ achievability bound [30]. We require the average probability of error $\mathbb{P}[\psi(Y, H) \neq S] \leq \epsilon$, $0 < \epsilon < \tfrac{1}{2}$.

**Goal:** Let the maximum size of the codebook with codeword length $n$ and average probability of error $\epsilon$, with PP constraint be denoted as $M_p^*(n, \epsilon, \bar{P}) \equiv M_p^*$. The corresponding maximal coding rate (in bits per channel use) $R_p^*(n, \epsilon, \bar{P}) \equiv R_p^* = n^{-1} \log M_p^*$. Similarly, for transmitters with AP and EH constraint we define $M_a^*$ and $R_a^*$, and $M_e^*$ and $R_e^*$, respectively. Our primary goal is to obtain tractable lower and upper bounds for $R_p^*$, $R_a^*$ and $R_e^*$. In addition, we also conduct a moderate deviation analysis to characterize the exponential rate at which error probability drops to zero, when $R_p^*$, $R_a^*$ and $R_e^*$ are allowed to converge to capacity at a pre-specified rate.

We conclude this section by making note of the notation that we use in the rest of the paper. We fix all logarithms to the base 2. However, $\exp(\cdot)$ denotes the exponent $e$. Let $C(x) \triangleq \log (1 + x)$, $\mathcal{L}(x) \triangleq x/(1 + x)$, $V(x) \triangleq x(2 + x)/(1 + x)^2$. The set of integers, positive integers, real and positive real numbers are denoted as $\mathbb{Z}$, $\mathbb{Z}_+$, $\mathbb{R}$ and $\mathbb{R}_+$, respectively. The indicator function of an event $A$ is $1_A$ and $A^{\ell} \equiv A_1 \times \ldots \times A_{\ell}$. For $p = 1, 2$, given $a, c$ belonging to $\mathbb{C}^\ell$, $||a||_p$ denotes the $p$–norm of $a$ and $\langle a, c \rangle$ the inner product of $a$ and $c$. In particular, $||a||_2$ is simply denoted as $||a||$. For $\alpha \in \mathbb{R}$, $a^\alpha \triangleq (a_1^\alpha, \ldots, a_\ell^\alpha)$ and $\alpha \cdot a = (\alpha a_1, \ldots, \alpha a_n)$. For a function $f$ defined on $\mathbb{C}$, $(f(a^1), \ldots, f(a^\ell))$ is sometimes compactly expressed as $f(a^\alpha)$. If $\ell = Bn_c$, $a_b = (a_{b,1}, \ldots, a_{b,nc})$ and $a_{b,k} = a_{b,k,1} + ja_{b,k,2}$. For $q \in \mathbb{R}_+$, $\ell \in \mathbb{Z}_+$, let $X_{\ell}^{c}(a_{b}) = \{x : ||x||_2 = \sqrt{q}\}$. At times, we denote the sum of $a_1, \ldots, a_{\ell}$, $\sum_{k=1}^{\ell} a_{k}$ as $S_{\ell}(a_{1})$. The notation $U \perp V$ denotes that the random vectors are independent. The variance of a random variable $U$ is denoted as $\mathbb{V}[U]$ and the expectation with respect to $U$ as $\mathbb{E}_U[\cdot]$. Also, $U \sim F$ denotes $U$ is distributed according to the distribution $F$. The function $\Phi(\cdot)$ denotes the cdf of a standard Gaussian random variable, $\phi(\cdot)$ denotes the corresponding pdf, $\Phi^{-1}(\cdot)$ denotes the inverse cdf and $X \sim \mathcal{N}(a, b)$ denotes that the random variable $X$ has normal distribution with mean $a$ and variance $b$. We use the standard Bachman-Landau asymptotic notation $o(\cdot)$, $O(\cdot)$. The notation $\lim \equiv \lim \sup$, $\lim \equiv \lim \inf$. The notation $\overset{D}{=} \equiv$ means equivalence in distribution.

### III. PEAK POWER CONSTRAINED TRANSMITTER

In this section, we provide lower and upper bounds on $R_p^*$, the maximal coding rate under peak codeword power constraint mentioned in (2). With $\mathcal{P}_{WF}(\cdot)$ and $\lambda$ defined as in (1), let $G_b \triangleq |H_b|/\sqrt{\mathcal{P}_{WF}(|H_b|)}/\sigma_N$. For $\alpha \in (0, 1)$, $c_\alpha \triangleq \sqrt{2(n_c + 1) \log (1/\alpha\epsilon)}/\log \epsilon$ and $V_{BF}(\bar{P}) \triangleq \mathbb{E} \mathbb{[} V \left( G_1^2 \right)] + n_c \mathbb{V} \mathbb{[} C \left( G_1^2 \right)] + \mathbb{V} \mathbb{[} \mathcal{L} \left( G_1^2 \right)] . \quad(5)$
With the above notation in place, we have the following result that provides lower and upper bounds on \( R^*_p \).

**Theorem 1.** Consider the block fading channel described in Section II-A with a peak power constrained transmitter as in [II-B1]. For \( n \) large enough and any \( \alpha \in (0, 1) \), the maximal rate \( R^*_p \) is lower bounded as

\[
R^*_p \geq C(\bar{P}) - \frac{c_\epsilon}{\sqrt{n}} + \sqrt{\frac{V_{BF}(\bar{P})}{n}} \Phi^{-1} \left( \frac{(1 - \alpha)\epsilon}{2} \right) + O \left( \frac{\log n}{n} \right).
\]

(6)

Further, \( R^*_p \) is upper bounded as

\[
R^*_p \leq C(\bar{P}) + \sqrt{\frac{V_{BF}(\bar{P})}{n}} \Phi^{-1}(\epsilon) + o \left( \frac{1}{\sqrt{n}} \right).
\]

(7)

**Proof.** See Appendix A. \( \square \)

A. Discussion of results

First, note that, with no CSIT and PP constraint, the optimal coefficient of \( 1/\sqrt{n} \) term, i.e., the channel dispersion or optimal second order coefficient, is obtained by replacing \( G_1 \) in the definition of \( V_{BF}(\bar{P}) \) in [5] with \( |H_1| \) [14]. If we restrict to the second order approximation of the bounds on \( R^*_p \) (i.e., discarding the terms whose order is higher than \( 1/\sqrt{n} \) on the RHS of (6) and (7)), in Section VII-A we will numerically evaluate the upper bound and show that the bound could be significantly lower than that predicted by the capacity expression in (1), depending on the value of system parameters. Similarly, we will also show that the rate predicted by the second order approximation of the lower bound in (6) obtained by means of power control, can be higher than that without power control in [14]. This will also point to the rate enhancement possible, depending on the value of various system parameters, with CSIT in place. However, we also note that second order coefficients in our lower and upper bounds do not match. Hence we also study the proximity of the bounds numerically in VII-A.

Next, we consider the AP constraint case.

IV. AVERAGE POWER CONSTRAINED TRANSMITTER

The benefit of power control is more pronounced when the system has an average power power constraint rather than a peak power constraint. In this section, we provide lower and upper bounds on the achievable rate \( R^*_a \) subject to AP constraint. Towards that, we have the
following definition. A power allocation function is a mapping $P : C \mapsto \mathbb{R}_+$ that defines the amount of power expended on channel states. With other notation as in the previous section, we have the following result.

**Theorem 2.** Consider the block fading channel described in Section II-A with an average power constrained transmitter as in II-B2. For $n$ large enough, the maximal rate $R^*_a$ is lower bounded as

$$R^*_a \geq C(\bar{P}) + \sqrt{\frac{V_{BF}(\bar{P})}{n}} \Phi^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right).$$

(8)

Further, if we restrict to the class of power allocation policies with $E[P^{2+\delta}(H_1)] < \infty$, for some $\delta > 0$, $R^*_a$ is upper bounded as

$$R^*_a \leq C(\bar{P}) + \sqrt{\frac{V_{BF}(\bar{P})}{n}} \Phi^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right).$$

(9)

**Proof.** See Appendix B.

A. **Discussion of results**

The above bounds illustrate that water filling power allocation is second order optimal (under a mild additional assumption on the power allocation function) for the case of AP constraint under the finite block length and non-vanishing error probability setting. From the proof in Appendix B, we note that for a block fading channel model with finite number of fading states, the upper bound in (9) holds without the moment restriction. We also note that the capacity achieving water filling power allocation with average power $\bar{P}$ is peak power constrained (see (1)) and hence the additional moment constraint is naturally met. In addition, our bounds also establishes that, as in the capacity case, non-causal CSIT is not helpful in improving the rate upto second order (again, under a mild assumption on power allocation function).

We illustrate the rate enhancement possible due to CSIT under AP constraint in Section VII-A.

V. **Energy Harvesting Transmitter**

In this section, our goal is to obtain tractable lower and upper bounds on the achievable rate $R^*_e$, with codeword length $n$, average probability of error $\epsilon$, satisfying energy harvesting constraints with average harvested energy $\bar{E}$. We begin with explaining an encoding-decoding scheme that we use in the proof of our lower bound.
A. Save and Transmit Scheme

Corresponding to each \( m \in [1 : M] \), the encoder generates an \( n \) length codeword \( X'(m) \triangleq (X'_1(m), \ldots, X'_n(m)) \) of i.i.d. random variables \( X'_i(m) \) distributed according to \( \mathcal{CN}(0, 1) \), where \( n = Bn_c \) and \( i \in [1 : n] \). This encoding scheme is used in conjunction with a power controller and a transmission scheme called the Save and Transmit scheme. The transmission scheme is known to achieve the capacity of a real AWGN channel \(^{[31]}\).

Save and Transmit scheme consists of two phases. To transmit a codeword of length \( n \), the transmitter does not transmit in the first \( Nn \) harvesting slots, where \( N \) is appropriately chosen. This phase is called the energy saving phase. After the first \( Nn \) slots, the buffer contains some energy and the codeword \( X'(S) \) is transmitted in the next \( n \) slots, where \( S \) is the message selected. This is the transmission phase. If \( [b, k]^{th} \) symbol to be transmitted requires more energy than what is available i.e., \( |X'_{[b,k]}|^2 > \hat{\Delta}_{[b,k]} \), we refer to the slot as being in an energy outage. The channel output in the transmission phase is
\[
Y_{[b,k]} = H_b X'_{[b,k]} \mathbb{1}\{\hat{\Delta}_{[b,k]} \geq |X'_{[b,k]}|^2\} + Z_{[b,k]}.
\]

The decoder \( \psi : \mathbb{C}^{(n)} \times \mathbb{C}^{(B)} \mapsto [1 : M] \) upon receiving \( Y \triangleq (Y_{[1,1]}, \ldots, Y_{[B,n_c]}) \), obtains \( \psi(Y, H) \equiv \hat{S} \), an estimate of the message \( S \) transmitted. The decoder used in obtaining our lower bound is same as that used in obtaining the lower bounds in Theorem 1 and Theorem 2.

We recall certain prior results that we use in deriving the bounds. The capacity of a fast fading channel (i.e., \( n_c = 1 \)) with an energy harvesting transmitter was previously characterized in \(^{[19]}\) (see Theorem 3 therein). Using the same analysis as in \(^{[19]}\), with \( \mathbb{E}[\mathcal{P}_{WF}(|H_1|)] = \bar{E} \), it is easy to see that the capacity of the block fading channel under consideration is given by
\[
\mathcal{C}(\bar{E}) = \mathbb{E}\left[ \log \left( 1 + \frac{|H_1|^2 \mathcal{P}_{WF}(|H_1|)}{\sigma_N^2} \right) \right]. \tag{10}
\]

The following lemma from \(^{[20]}\) will aid in tackling the proof of the lower bound by decoupling the channel statistics from the analysis of the energy outage event described in Section V-A.

**Lemma 1.** Consider a communication channel with an energy harvesting transmitter. Suppose the input to the channel is generated i.i.d. with mean zero and variance \( \bar{E} \), and the energy arrivals are i.i.d. with mean \( \bar{E} \) and variance \( \sigma_E^2 \). Let \( \epsilon \) be the allowed probability of error (maximal or average) for the system. Then, given \( 0 < \alpha < 1 \), there exists a Save and Transmit scheme, with the duration of transmission phase \( n \) slots and the duration of the saving phase \( N_n = K_{e,\alpha} \sqrt{n} \) slots. Further, with \( K_{e,\alpha} = \sqrt{\frac{4(2\bar{E}^2 + \sigma_E^2)}{(1-\epsilon)^2 \bar{E}^2}} \), the probability of error in decoding
\[ P[\hat{S} \neq S] \equiv \epsilon_n \text{ referred to as the non-energy harvesting error is bounded as} \]
\[ \epsilon_n \leq \alpha \epsilon - \frac{4\sigma_E^2}{K_{e,\alpha} E^2 \sqrt{n}}. \]  

Next lemma is a version of the meta-converse (derived from [10], [21]) that is suitably adjusted to incorporate the effects of the stochastic nature of harvested energy and fading gains. We use this result in our proof of the upper bound.

**Lemma 2.** Consider a block fading channel as in Section II-A with an energy harvesting transmitter as in Section II-B3. Let the codewords be generated according to \( P_{X|E,H} \in \mathcal{P}(E,H) \), where \( \mathcal{P}(E,H) \) is the set of input distributions that satisfy the energy harvesting constraints described in Section II-B3. Then, any code with \( M \) codewords of codeword length \( n \) and average probability of error \( \epsilon \), satisfies
\[ M \leq \sup_{P_{X|E,H} \in \mathcal{P}(E,H)} \mathbb{P} \left[ \log \frac{d_{Y|X,H}}{d_{Y|H}} \leq \log \gamma_n - \epsilon \right] \]  
for any \( \gamma_n > 0 \) and any auxiliary channel \( Q_{Y|H} \), whenever the RHS of (12) is non-negative. Here, the probability in the denominator is with respect to the distribution \( P_{E,P_H,P_{X|E,H}P_{Y|X,H}} \).

Now, we use the above lemmas to obtain the following result.

**Theorem 3.** Consider the channel in Section II-A and energy harvesting model in Section II-B3. With \( C(E) \) as in (10), \( K_{e,\alpha} \) as in (11), \( V_{EF}^*(E) \triangleq \mathbb{E}[\mathcal{L}(G_1^2)] + n_c \sqrt{\mathbb{V}[C(G_1^2)]} + \sqrt{\mathbb{V}[\mathcal{L}(G_1^2)]} \) and
\[ V_{e,\alpha}(E) \triangleq \sqrt{V_{EF}^*(E)} \Phi^{-1}(\alpha \epsilon) - K_{e,\alpha} C(E). \]  
Then, the maximal coding rate \( R^*_e \) satisfies the following bounds.

1) For any \( 0 < \alpha < 1 \) and sufficiently large \( n \),
\[ R^*_e \geq C(E) + V_{e,\alpha}(E)/\sqrt{n} + o(1/\sqrt{n}). \]  
2) For sufficiently large \( n \), with \( V_1(x) \triangleq \mathcal{L}^2(x) + \sigma_E^2/\lambda^2 \), for \( x \in \mathbb{R}_+ \), \( \lambda \) as in (10) and
\[ V_{EF}^*(E) \triangleq \mathbb{E}[V_1(G_1^2)] + n_c \sqrt{\mathbb{V}[C(G_1^2)]} + \sqrt{\mathbb{V}[\mathcal{L}(G_1^2)]}, \]
\[ R^*_e \leq C(E) + \sqrt{V_{EF}^*[E]/n\Phi^{-1}(\epsilon)} + o(1/\sqrt{n}). \]  

**Proof.** We provide a sketch of arguments involved in the proof of the lower bound in (14). As in the case of PP and AP constraints, we decouple coding and power control. The encoder generates \( M \) i.i.d. codewords with the symbols distributed independently \( \mathcal{CN}(0,1) \). The codebook is shared
with the receiver. The transmitter adopts the Save and Transmit scheme described in Section V-A. Thus, the transmitter waits for first $N_n$ slots where, $N_n$ is fixed as in Lemma 1. After $N_n$ slots, the transmitter starts sending the codeword corresponding to the message chosen. Codeword symbol $X'_{[b,k]}$ in slot $[b,k]$ is a complex number with two components $X'_{[b,k],1}$ and $X'_{[b,k],2}$, to be sent across the in-phase and quadrature channels, respectively. Each component is passed through a power controller and the output of the power controller is $X_{[b,k],i} = \sqrt{P_{WF}(|H_b|)}X'_{[b,k],i}$, for $i = 1, 2$. If $X^2_{[b,k],1} + X^2_{[b,k],2} \leq \hat{A}_{[b,k]}$, both symbols are transmitted, else no transmission happens in that slot. We choose $N_n$ as in Lemma 1 so that, the non-energy harvesting error is upper bounded as in (11). We use this estimate to perform a similar analysis as in Appendix A and Appendix B, we obtain (14). Proof of (15) is deferred to the Appendix D.

B. Discussion of Results

As mentioned earlier, this is the first attempt at providing a finite blocklength analysis of a fading channel powered by an energy harvesting transmitter. We note that the second order coefficients, i.e., coefficients of $\sqrt{n}$, of lower and upper bounds of the new result do not match. In this regard, we note the unavailability of a matching second order coefficient even for an AWGN channel with an energy harvesting transmitter [20], [21].

Note that an energy harvesting wireless transmitter appropriately regulates the transmission power based on its CSI and energy availability. However, in our proof of the inner bound, we decouple the channel coding from the power control. Similarly, the energy harvesting mechanism is also decoupled from the channel coding in the Save and Transmit scheme. This decoupling suggests that our proof for the full CSIT case can be readily adapted to the case of no CSIT and full CSIR. Specifically, replacing $G_1^2$ with $|H_1|^2/\sigma_N^2$ in the expressions in (13) and using it in (14) yields a valid lower bound for the case of no CSIT and full CSIR. Similarly, the analysis in the Appendix holds for the no CSIT, full CSIR case as well. In fact, the analysis can be considerably simplified owing to the fact that the channel input does not depend on the fading states when CSIT is not available. Thus, replacing $G_1^2$ with $|H_1|^2/\sigma_N^2$ to compute the terms in the upper bound in (15) results in a valid upper bound for the no CSIT, full CSIR case.

VI. MODERATE DEVIATION BOUNDS

In this section, we characterize the optimal rate of convergence, viz., the moderate deviation constant, of average probability of error, to zero when the channel coding rate is converging to
the capacity at a rate between $\sqrt{n}$ and $n$. We derive the moderate deviation constant for the block fading channel in Section II with transmitters subjected to different kinds of power constraints mentioned therein. Let $p_{e,\text{avg}}^*(n, M_n)$ denote the minimum average probability of error over a given channel at codeword length $n$ and number of messages $M_n$. We begin with the following definition in [12].

**Definition 1.** A channel with capacity $C$ is said to satisfy the moderate deviation property with constant $\mu$ if for any sequence of integers $M_n$ such that $\log M_n = nC - na_n$ where, the sequence \( \{a_n, n \geq 1\} \) is strictly positive, $a_n \to 0$, $na_n^2 \to \infty$, we have $\lim_{n \to \infty} \frac{1}{na_n^2} \log p_{e,\text{avg}}^*(n, M_n) = -\frac{1}{2\mu}$.

We have the following result.

**Theorem 4.** For the block fading channel in Section II-A with transmitters subjected to PP and AP constraints, the moderate deviation constant $\mu$ is such that $V_{\text{BF}}'(\bar{P}) \leq -1/2\mu \leq V_{\text{BF}}(\bar{P})$, where $V_{\text{BF}}(\bar{P})$ is defined as (5) and $V_{\text{BF}}'(\bar{P}) \triangleq E[L(G_{1}^{2})] + n_c V[C(G_{1}^{2})] + V[L(G_{1}^{2})]$. Also, under EH constraint, $\mu$ is such that $V_{\text{EF}}'(\bar{E}) \leq -1/2\mu \leq V_{\text{EF}}''(\bar{E})$.

**Proof.** See Appendix D.

**VII. NUMERICAL EXAMPLES**

**A. Non-energy harvesting transmitter**

In this section, we compare numerically, the bounds obtained for the non-energy harvesting case (Theorem 1 and Theorem 2). We will compare the bounds up to the second order term in the rate expression. We assume fading distribution to be $\mathcal{CN}(0, \sigma^2_H)$. Also, we assume $\sigma^2_N = 4$ and $\epsilon = 0.05$ and $n_c = 10$. By fixing $\bar{P} = 5$dB, we plot the convergence of various bounds to the channel capacity $C(\bar{P}) = 0.6892$ bits/channel use (equation (1)), as $B$ increases, in Figure 1. The acronyms LB, UB and TIC refer to the lower bound, the upper bound and truncated channel inversion, respectively. We have also plotted the rate with no CSIT under PP constraint and with the rate corresponding to truncated channel inversion. Truncated channel inversion provides the delay limited capacity for point-to-point fading channels ([23], Section III.B.1). We observe that power allocation that achieves the delay limited capacity achieves only an inferior rate to that achieved via the water filling power allocation scheme in the finite block length regime.

In Figure 2 we fix $n_c = 10$, $B = 1000$ and provide comparison of the various bounds for different values of $\bar{P}$. By comparing the lower bound under the PP constraint with the rate for no
Fig. 1: Comparison of rate versus $B$ for the non-energy harvesting case. We fix $n_c = 10$, $\bar{P} = 5$dB, $\epsilon = 0.05$, $\sigma^2_N = 4$ and $\sigma^2_H = 0.1$.

CSIT case under PP constraint, we observe the rate enhancement possible due to the knowledge of CSIT (via power control). In Figure 3 we compare the effect of fading parameter $\sigma_H$ on the second order approximation of rates. We observe that our second order approximation of LB under PP constraint provides better gains for lower values of $\sigma_H$ with other parameters fixed.
In this section, we compare our lower and upper bounds on the maximal coding rate, numerically. State-of-the-art ambient electric field energy harvesting technology promises a harvesting energy density of 17 micro Joules (µJ) per unit time and unit volume (see, for instance [32] and Table I therein). We consider $E$ in the range 1 to 17 µJ. For simplicity, we normalize and refer to 1µJ as a unit. We fix the average probability of error $\epsilon = 0.1$ which is typically the targeted probability of error in massive machine-type transmission [33]. Further, we fix $n_c = 20$, $B = 400$, $\sigma_N^2 = 0.4$, $\sigma_H^2 = 0.9$ and $\sigma_E^2 = 0.1$. We assume fading distribution to be $\mathcal{CN}(0,\sigma_H^2)$. We plot the second order approximation of the lower and upper bounds (terms excluding $o(\sqrt{n})$ term on the right hand side of (14) and (15) with an additional scaling of $1/n$) as a function of $E$ in Figure 4. In obtaining the plot, we have optimized over the parameter $\alpha$ appearing in (14), for $\alpha$ between 0.1 and 0.9. For the chosen set of parameters, we observe that the second order approximation of the upper bound differs from that of the lower bound by a maximum of $\approx 15\%$. Also, the duration of the saving phase $N_n$ (of the Save and Transmit scheme) in this example is approximately 846 slots, the optimum $\alpha = 0.104$ and $K_{\epsilon,\alpha} \approx 9.5$. Next, we illustrate the convergence of the second order approximation of the bounds to the capacity as the number of blocks $B$ increases, in Figure 5. Here, we have chosen $E = 17$ units and other parameters are fixed as before. The bounds in this case differ by $\approx 11\%$.

**B. Energy harvesting transmitter**

Fig. 4: Comparison of rate versus $E$ for the energy harvesting case. We fix $\epsilon = 0.1$, $n_c = 20$, $B = 400$, $\sigma_N^2 = 0.4$, $\sigma_H^2 = 0.9$ and $\sigma_E^2 = 0.1$.

Fig. 5: Comparison of rate versus $B$ for the energy harvesting case. We fix $\epsilon = 0.1$, $n_c = 20$, $E = 17$, $\sigma_N^2 = 0.4$, $\sigma_H^2 = 0.9$ and $\sigma_E^2 = 0.1$. 
VIII. Conclusion

In this paper, we have obtained upper and lower bounds for the maximal coding rate over a block fading channel with peak, average and energy harvesting power constraints on the transmitted codeword. The bounds obtained shed light on the rate enhancement possible due to the availability of CSIT. The bounds also characterize the performance of water-filling power allocation in the finite block length regime. Further, we have derived moderate deviation bounds. These bounds characterizes the rate of convergence of probability of error to zero when the channel coding rate is allowed to scale with codeword length in a pre-specified way.

Appendix A

Proof of Theorem 1

A. Lower bound on rate with PP constraint

We prove the lower bound for a real block fading channel with gain $G_b = |H_b| \sqrt{P_{WF}(|H_b|)}/\sigma_N$ in block $b$. The scheme presented can be readily adapted to the actual model in Section II-A by independently coding over in-phase and quadrature channels to get the lower bounds we use. Let $G \triangleq (G_1, \ldots, G_B)$ and the additive noise in slot $[b, k]$ be $\tilde{Z}_{[b,k]}$, where $\tilde{Z}_{[b,k]} \sim \mathcal{N}(0, 1/2)$. In a block, the transmitter uses the channel $2n_c$ times ($n_c$ is defined as in Section II-A).

1) Codebook Generation: Given codeword length $n$, fix $\delta_n \in (0, 1)$; for our choice of $\delta_n$, see Section A-A5. Corresponding to each message $m \in [1 : M]$, generate i.i.d., PP constrained codewords uniformly randomly from $\mathcal{X}_n(1 - \delta_n) \subset \mathbb{R}^n$ (where, as defined in Section II, $\mathcal{X}_\ell(q)$ denotes the surface of a sphere in $\mathbb{R}^\ell$ with radius $\sqrt{\ell q}$) and share it with the receiver. The codeword corresponding to message $m$, $\tilde{X}(m)$ is such that $\tilde{X}(m) \perp G$ for all $m$. For each $m$, the channel input symbol in slot $[b, k]$ is $\tilde{X}_{[b,k]}(m)$.

2) Power Control: To send message $m$, at the beginning of block $b$, if $G_b = 0$ the transmitter does not transmit any symbols in that block and waits to observe $G_{b+1}$ (unless it is the last block). Else, if $G_b > 0$, transmitter computes $|H_b|$ from $G_b$ and checks if

$$\sum_{\ell=1}^b \sum_{k=1}^{2n_c} \tilde{X}_{[\ell,k]}^2(m) P_{WF}(|H_{\ell}|) \leq 2Bn_c \bar{P}$$

(16)

is met. If the constraint is not met, the PP constraint in (2) is violated and the channel input is 0 for blocks $\ell \in [b : B]$. That is, the transmission is halted. Else, if the constraint is met, the corresponding codeword symbols are transmitted in the block. If the transmitter sends codeword
symbols successfully in all $B$ blocks without violating the constraint in (16), the constraint in (2) is met. Then, the channel output $\hat{Y}_{[b,k]} = G_b \hat{X}_{[b,k]}(m) + \tilde{Z}_{[b,k]}$, $b \in [1 : B]$ and $k \in [1 : 2n_c]$.

3) Decoding: The decoder declares $\hat{S}$ as the transmitted message using the Neyman-Pearson decoder employed in obtaining the $\beta\beta$ achievability bound in [30] (see Theorem 1 therein). We choose the test subject to a detection probability $1 - \epsilon'_n + \tau$, where we fix $\epsilon'_n \in (0, \epsilon)$ and $\tau \in (0, \epsilon_n')$ (our choices of $\epsilon'_n$ and $\tau$ are provided in Section A-A5). Further, the auxiliary channel in using the decoder is $Q_{Y|G} = \prod_{b=1}^B \prod_{k=1}^{2n_c} \mathcal{N}(0, (1 - \delta_n) G_b^2 \sigma_N^2 + \sigma_N^2)$. Then, using the $\beta\beta$ achievability bound, under an average probability of error in decoding $\epsilon_n'$,

$$M \geq \frac{\beta_\tau \left( P_{\hat{Y},G} \left( \begin{array}{c} Y \\ X \end{array} \right) \right)}{\beta_1 - \epsilon'_n + \tau \left( P_{\hat{X}(1),G} \left( \begin{array}{c} X \\ \tilde{Y} \end{array} \right) \right) P_{\hat{X}(1),G} \left( \begin{array}{c} X \end{array} \right) Q_{Y|G}},$$

(17)

where the $\beta$ notation is as in [10].

4) Analysis of $\beta\beta$ bound in (17): For convenience, denote the inequality in (17) as $M \geq \beta_\tau \left( P_{\hat{Y},G} \left( \begin{array}{c} Y \\ X \end{array} \right) \right) / \beta_1 - \epsilon'_n + \tau \left( P_{\hat{X}(1),G} \left( \begin{array}{c} X \\ \tilde{Y} \end{array} \right) \right) P_{\hat{X}(1),G} \left( \begin{array}{c} X \end{array} \right) Q_{Y|G}$. Here, $P_{\hat{Y},G}$ and other probability terms are similarly defined. To obtain a tractable lower bound for the RHS of (17) we upper bound $\beta_1 - \epsilon'_n + \tau \left( P_{\hat{X}(1),G} \left( \begin{array}{c} X \\ \tilde{Y} \end{array} \right) \right)$ as

$$\beta_1 - \epsilon'_n + \tau \left( P_{\hat{X}(1),G} \left( \begin{array}{c} X \\ \tilde{Y} \end{array} \right) \right) \leq Q_{\bar{f}} \left[ \frac{dP_{\hat{Y}}}{dQ_{\bar{f}}} \geq \gamma_0 \right] \leq \frac{c_2}{\gamma_0 \sqrt{n}}.$$  

(18)

Here, $\log \gamma_0 = \mu_n + \sqrt{\nu_n} \Phi^{-1}(\epsilon''_n)$ and, $\mu_n$ and $\nu_n$ are the mean and variance of $\log \frac{dP_{\hat{Y}}}{dQ_{\bar{f}}}$ under $P_{\hat{Y}}$, respectively. Also, $\epsilon''_n = \epsilon'_n - \tau - (2c_1/\sqrt{n})$ and $c_1 > 0$. In addition, $\gamma_0$ is such that $P_{\hat{Y}} \left( \frac{dP_{\hat{Y}}}{dQ_{\bar{f}}} \geq \gamma_0 \right) \geq 1 - \epsilon'_n + \tau$. This follows from applying Berry-Esseen theorem ([34], Chapter 16) to $P_{\hat{Y}} \left( \frac{dP_{\hat{Y}}}{dQ_{\bar{f}}} \geq \gamma_0 \right) | \tilde{X}(1) = \tilde{x}(1)$ ($c_1$ is the Berry-Esseen correction term). Also, (b) follows from Lemma 20 in [35] and $c_2 > 0$.

Next, from [14] (see equation (59) therein), choosing $\tau$ to be a constant (i.e., not depending on $n$), $\beta_\tau \left( P_{\hat{Y},G} \left( \begin{array}{c} Y \\ X \end{array} \right) \right) = O(1)$. Combining this fact with (17) and (18), we obtain

$$\log M_p^* \geq \mu_n + \sqrt{\nu_n} \Phi^{-1}(\epsilon''_n) + O(\log n).$$  

(19)

By Taylor’s theorem, $\mu_n \geq nC(\bar{P}) - n\delta''_n$, where $\delta''_n = (\bar{P}/(2\lambda)) \delta_n + \delta_n^2$. Further, from Lemma 3 $\nu_{BF}(\bar{P}) \geq \nu_n$. Combing these inequalities with the fact $\Phi^{-1}(\epsilon''_n) < 0$ to lower bound the RHS of (19), we obtain $R_p^* \geq C(\bar{P}) - \delta''_n + \sqrt{\nu_{BF}(\bar{P})/n} \Phi^{-1}(\epsilon''_n) + O\left(\frac{\log n}{n}\right)$. What remains is a characterization of $\delta''_n$ and $\epsilon''_n$ appearing in the above equation. Next, we do that by analysing the constraint in (16).
5) **Analysis of the constraint in (16):** Let $\mathcal{E}_T$ be the event that the constraint in (16) is violated for some $b \in [1 : B]$. Then,

$$\mathbb{P} \left[ \mathcal{E}_T \cup \left\{ S \neq \hat{S} \right\} \right] \leq \mathbb{P} \left[ \mathcal{E}_T \right] + \epsilon',$$

In order to upper bound $\mathbb{P} [\mathcal{E}_T]$, we choose $\delta_n$ (mentioned in Section A-A1) to be $(2\lambda c_n) / (\bar{P} \sqrt{n})$, where for any $\alpha \in (0, 1),$

$$c_n = \sqrt{(2n_c + 2) \log(1/\alpha e) / \log e}.$$ 

Then, from Lemma 4 we obtain the bound $\mathbb{P} [\mathcal{E}_T] \leq \alpha e + c_3 / \sqrt{n}$, where $c_3 = 64(2n_c + 3)^4$. Based on this bound, we choose $\epsilon_n' = (1 - \alpha)\epsilon - (c_3 / \sqrt{n})$ so that $\mathbb{P}[\mathcal{E}_T \cup \{S \neq \hat{S}\}] \leq \epsilon$. Here, we require $n$ such that $\epsilon_n' > 0$. Further, as required in obtaining (19), we choose $\tau = (1 - \alpha)\epsilon / 2$, i.e., not depending on $n$. Thus, we obtain $\epsilon_n''$ in (19) to be $(1 - \alpha)\epsilon / 2 - c_4 / \sqrt{n}$, where $c_4 = 2c_2 + c_3$ and $n$ is such that $\epsilon_n'' > 0$. Also, by virtue of choice of $\delta_n$, $\delta_n'' = c_n / \sqrt{n} + (4\lambda^2 c_n^2) / (\bar{P}^2 n)$. Plugging in the expressions for $\epsilon_n''$ and $\delta_n''$ in the lower bound on $R^*_p$, denoting $\epsilon_\alpha = (1 - \alpha)\epsilon / 2$ and applying Taylor’s theorem, we obtain the required result.

**Lemma 3.** With the notation as in Section II and $\alpha \in [0, 1]$, let $V_{BF}(\bar{P}, \alpha) \triangleq \mathbb{E} [V(\alpha G^2_1)] + n_c \mathbb{V} [C(\alpha G^2_1)] + \mathbb{V} [L(\alpha G^2_1)].$ The function $V_{BF}(\bar{P}, \alpha)$ is monotonically decreasing in $\alpha.$

**Proof.** Observe that $V_{BF}(\bar{P}, \alpha)$ can be written as $[1 - \mathbb{E}^2 [1 - L(\alpha G^2_1)]] + n_c \mathbb{V} [C(\alpha G^2_1)].$ Since $L(\alpha G^2_1)$ is monotonically increasing in $\alpha,$

$$[1 - \mathbb{E}^2 [1 - L(G^2_1)]] \geq [1 - \mathbb{E}^2 [1 - L(\alpha G^2_1)]].$$

Thus, the result follows if we show $\mathbb{V}[C(G^2_1)] \geq \mathbb{V}[C(\alpha G^2_1)].$ Towards that, first we invoke Taylor’s theorem and express

$$C(G^2_1) = C(\alpha G^2_1) + (1 - \alpha)(L(\alpha G^2_1)/u),$$

for some $u \in (\alpha, 1).$ Next, for convenience we denote $f_1(G_1) = C(\alpha G^2_1), \theta = (1 - \alpha)/u$ and $f_2(G_1) = \theta L(\alpha G^2_1)$ and define $\gamma = \mathbb{E} [f_1(G_1)f_2(G_1)] - \mathbb{E} [f_1(G_1)] \mathbb{E} [f_2(G_1)].$ Then, using Chebyshev’s association inequality ([36], Theorem 2.14), we observe $\gamma \geq 0.$ Using this observation and the Taylor’s expansion for $C(G^2_1),$

$$\mathbb{V}[C(G^2_1)] = \mathbb{V} [f_1(G_1)] + \mathbb{V} [f_2(G_1)] + 2\gamma \geq \mathbb{V}[C(\alpha G^2_1)].$$

Combining this with the monotonicity of $L(\alpha G^2_1)$ in $\alpha$, we obtain the required result. 

\[\square\]
Lemma 4. Fix the channel described in Section II-A. With \( \mathcal{X}_n(\cdot) \) defined as in Section II let \( \tilde{X} \) be uniformly distributed on \( \mathcal{X}_n(1 - \delta_n) \subset \mathbb{R}^n \), for \( \delta_n > 0 \). Let \( \mathcal{P}_{WF}(\cdot) \) be the water filling power allocation function with average power \( \bar{P} \), as in (I). Further, the event \( \mathcal{E}_T \) corresponding to violating the transmission constraint in (16) in some block is

\[
\mathcal{E}_T = \bigcup_{b=1}^{B} \left\{ \sum_{\ell=1}^{b} \| \tilde{X}_\ell \|^2 \mathcal{P}_{WF}(|H_\ell|) > n\bar{P} \right\},
\]

where the notation \( \tilde{X}_\ell \) is as mentioned in Section II. Then, with \( \lambda \) as in the definition of \( \mathcal{P}_{WF}(\cdot) \),

\[
\mathbb{P}[\mathcal{E}_T] \leq \exp \left( -\frac{n\bar{P}^2\delta_n^2}{8(n_c+1)\lambda^2} \right) + \frac{64(2n_c+3)^4}{\sqrt{n}}.
\]

Proof. Let \( U_b \triangleq \| \tilde{x}_b \|^2 \mathcal{P}_{WF}(|H_\ell|) - \| \tilde{x}_b \|^2 \bar{P} \). Making use of the notation for summation \( S_B(\cdot) \) defined in Section II, observe \( \mathbb{P}\left[ \mathcal{E}_T \mid \tilde{X} = \tilde{x} \right] = \mathbb{P}[S_B(U_1) > n\delta_n \bar{P}] \). Since \( \mathcal{P}_{WF}(\cdot) \) is such that \( 0 \leq \mathcal{P}_{WF}(\cdot) \leq \lambda, U_b \) is a zero mean, bounded random variable whose support set is contained in \( [-\| \tilde{x}_b \|^2 \lambda, \| \tilde{x}_b \|^2 \lambda] \). From Hoeffding’s inequality ([36], Theorem 2.8),

\[
\mathbb{P}\left[ \mathcal{E}_T \mid \tilde{X} = \tilde{x} \right] \leq \exp \left( -\frac{n\bar{P}^2\delta_n^2}{S_B(\| \tilde{x}_b \|^4)} \right),
\]

where \( \delta_n = (\delta_n \bar{P}/(2\lambda)) \). The above inequality yields

\[
\mathbb{P}[\mathcal{E}_T] \leq \mathbb{E} \left[ \exp \left( -\frac{n\bar{P}^2\delta_n^2}{S_B(\| \tilde{x}_1 \|^4)} \right) \right].
\]

To further bound \( \mathbb{P}[\mathcal{E}_T] \), let \( \tilde{X} \) be an i.i.d. \( \mathcal{N}(0, 1) \) vector. Then, \( \tilde{X} \overset{D}{=} (\sqrt{n(1 - \delta_n)/\| \tilde{x} \|}) \cdot \tilde{X} \).

From Chebyshev’s inequality \( \mathbb{P}\left[ \| \tilde{X} \|^2 > (n + n^{3/4}) \right] \leq (2/\sqrt{n}) \),

\[
\mathbb{P}[\mathcal{E}_T] \leq \mathbb{E} \left[ \exp \left( -\frac{n\bar{P}^2\delta_n^2}{S_B(\| \tilde{x}_1 \|^4)} \right) \right] + 2/\sqrt{n}, \tag{20}
\]

where we denote \( \hat{\delta}_n = (\delta_n/(1 - \delta_n)) (1 - (1/n^{1/4})) \). Next, we tackle the expectation term on the RHS of the above equation. Towards that, we apply Chebyshev’s inequality to obtain

\[
\mathbb{P}\left[ S_B(\| \tilde{x}_1 \|^4) > B \mathbb{E}[\| \tilde{x}_1 \|^4] + B^{3/4} \right] \leq \frac{32(2n_c+3)^4}{\sqrt{n}}. \tag{21}
\]

Combining (20) and (21) with the fact \( \mathbb{E}[\| \tilde{x}_1 \|^4] = 4(n_c + 1)n_c \), we obtain the result. \( \square \)

B. Upper bound on rate with PP constraint

Without loss of optimality (upto second order), we assume that PP constraint in (2) is satisfied with equality (see for instance, [35], Lemma 65). Also, we assume that CSIT is known non-causally to the transmitter, as it can only improve the rate. Let \( M \) denote the size of an arbitrary codebook \( C \) with codewords of length \( n \) satisfying the PP with equality, and average probability of error \( \epsilon \) if used over the block fading channel described in Section II. We divide the proof into various steps. Next, we explain those steps.
1) Choice of auxiliary channel: In order to invoke the meta converse ([10], Theorem 26), we choose the auxiliary channel $Q_{Y|H} = \prod_{b=1}^{B} \prod_{k=1}^{2m_c} \mathcal{N}\left(0, |H_b|^2 \mathcal{P}_{WF}(|H_b|) + \sigma_N^2 \right)$. Note that, in $[b,k]^{th}$ slot, output $Y_{[bk]}$ depends only on $H_b$ and is independent of the channel input.

2) Meta-converse bound and its relaxation: Let $\beta$ notation be as in [10], $\mathbb{P}_{Y|H,X}$ be the channel in Section II-A and $\epsilon'$ be the average probability of error if $\mathcal{C}$ is used over $Q_{Y|H}$. Invoking meta-converse [10] and Claim [1],

$$\beta \leq \epsilon'(\mathbb{P}_{X,H,Y}, Q_{X,H,Y}) \leq 1/M.$$ 

We lower bound $\beta \leq \epsilon'(\mathbb{P}_{X,H,Y}, Q_{X,H,Y})$ (see [10], equation (102)) as

$$((\mathbb{P}[\mathcal{L}] - \epsilon) / \gamma) \leq \beta \leq \epsilon'(\mathbb{P}_{X,H,Y}, Q_{X,H,Y}),$$

where $\gamma$ is an arbitrary positive real number, $\mathcal{L} \triangleq \{ i(X, H, Y) \leq \log \gamma \}$ and $i(X, H, Y) \triangleq \log \frac{dQ_{X,H,Y}}{dQ_{X,H,Y}}$. Considering the relaxation of meta-converse, and optimizing over the set $\mathcal{F}_n^{(p)}$ of input distributions with support on $\mathcal{X}_n(\bar{P})$,

$$\log M_n^{\ast} \leq \log \gamma - \inf_{\mathcal{F}_n^{(p)}} \log \mathbb{P}[\mathcal{L}] - \epsilon)^{+}.$$ 

3) Lower bounding $\mathbb{P}[\mathcal{L}]$ and the final bound: From Lemma 5, for some positive constants $c_4$, $c_6$ depending only on the fading distribution and $\sigma_N^2$ and an appropriate choice of $\gamma$, $\mathbb{P}[\mathcal{L}] \geq \epsilon + c_6/n^c_4$. Plugging this in the bound on $\log M_n^{\ast}$, we obtain

$$\log M_n^{\ast} \leq \log \gamma + c_4 \log n - \log c_4.$$ 

With $V_{BF}(\bar{P})$ as in (5), choose

$$\log \gamma = nC(\bar{P}) + \Phi^{-1}(\epsilon) \sqrt{nV_{BF}(\bar{P})} + o(\sqrt{n}).$$

Plugging this in the bound on $\log M_n^{\ast}$, we obtain the final bound in (7). \qed

Claim 1. For the auxiliary channel defined in Section A-B1, given an arbitrary codebook $\mathcal{C}$ of size $M$, the average probability of decoding error $\epsilon' = 1/M$. 

Proof. Let $S$ be the message, $\hat{S}$ the output of the decoder and $S \triangleq \mathbb{C}^B \times \mathcal{X}_n(\bar{P}) \times \mathbb{C}^n$. Then

$$\epsilon' \leq \sum_{m=1}^{M} \frac{1}{M} dF_H Q_{[y|h]} \mathbb{P} [\hat{S} = m|h, y] dy dF_{X|h} = \frac{1}{M},$$

where $(a)$ follows by noting $S \perp H$, definition of $Q$ channel and the Markov relation $S - (H, Y) \rightarrow \hat{S}$. \qed

Claim 2. For $h \in \mathbb{C}$, $\lambda$ and $\mathcal{P}_{WF}(\cdot)$ as defined in (1), $|h|^2/(\sigma_N^2 + |h|^2 \mathcal{P}_{WF}(|h|)) \leq 1/\lambda$. Also, with $L(\cdot)$ as in Section II, $L(\mathcal{P}_{WF}(|h|)) = \frac{\mathcal{P}_{WF}(|h|)}{\lambda}$. 

Proof. Since $\mathcal{P}_W(|h|) \geq \left( \lambda - \frac{\sigma^2}{|h|^2} \right)$, we have $\lambda|h|^2 \leq \sigma^2 + |h|^2 \mathcal{P}_W(|h|)$ and the first claim follows. Now, let $\mathcal{A} \triangleq \{ \lambda|h|^2 > \sigma^2 \}$. Observe $\mathcal{P}_W(|h|) = \mathbb{1}_\mathcal{A} \mathcal{P}_W(|h|)$. Hence,

$$
\mathcal{L}(|h|^2 \mathcal{P}_W(|h|)) = \mathbb{1}_\mathcal{A} \mathbb{1}_{|h|^2 \mathcal{P}_W(|h|)} \left( \sigma^2 + |h|^2 \mathcal{P}_W(|h|) \right).
$$

However, $\mathbb{1}_\mathcal{A} \frac{|h|^2 \mathcal{P}_W(|h|)}{\sigma^2 + |h|^2 \mathcal{P}_W(|h|)} = \mathbb{1}_\mathcal{A} \mathbb{1}_{|h|^2 \mathcal{P}_W(|h|)} \left( \frac{\lambda}{\sigma^2 + |h|^2 \mathcal{P}_W(|h|)} \right)$ which is equal to $\mathcal{P}_W(|h|)/\lambda$. 

\[\Box\]

Lemma 5. For $\gamma > 0$, let $\mathcal{I}_\gamma$ be defined as in Section A-B2 ε be the average probability of error for the channel model in Section II-A. There exist positive constants $c_4, c_6$ depending only on the channel parameters such that $\mathbb{P}[\mathcal{I}_\gamma] \geq \epsilon + c_6/n^{c_4}$.

Proof. With $Y$ being the output vector corresponding to the original channel $\mathbb{P}_{Y|X',H}$, $\mathcal{I}_\gamma$ in Section A-B2 can be equivalently written as $\mathcal{I}_\gamma = \{ \log \frac{d\mathbb{P}_{Y|X',H}}{d\mathbb{P}_{Y|H}} \leq \log \gamma \}$. To analyse $\mathbb{P}[\mathcal{I}_\gamma]$, first we consider a random variable with the same distribution as the log likelihood ratio in the definition of $\mathcal{I}_\gamma$. Towards that, we define

$$
L_b^{(1)} = \sum_{i=1}^{2} \frac{H_{b,i}^2}{\sigma_N^2} \frac{||X_{b,i}||^2}{(1 + G_b^2)} , \quad L_b^{(2)} = \sum_{i=1}^{2} \frac{H_{b,i} \langle X_{b,i}, Z_{b,i} \rangle}{\sigma_N^2} \frac{1}{(1 + G_b^2)} ,
$$

$$
L_b^{(3)} = -\sum_{i=1}^{2} \frac{G_b^2}{\sigma_N^2} \frac{||Z_{b,i}||^2}{(1 + G_b^2)} , \quad L_b = n_c C(G_b^2) + \sum_{\ell=1}^{3} L_b^{(\ell)} .
$$

Using the notation from Section II, define $S_B(L_1) = \sum_{b=1}^{B} L_b$ and $\mathcal{I}_\gamma' = \{ S_B(L_1) \leq \log \gamma_n \}$. Then, by direct verification, we have $S_B(L_1) \overset{(D)}{=} \log \frac{d\mathbb{P}_{Y|X',H}}{d\mathbb{P}_{Y|H}}$, Using this, we obtain $\mathbb{P}[\mathcal{I}_\gamma] = \mathbb{P}[\mathcal{I}_\gamma']$. Next, we lower bound $\mathbb{P}[\mathcal{I}_\gamma']$. Define

$$
L_b^{(4)} = \sum_{i=1}^{2} \frac{(n_c \bar{P} \sigma_N^2 - \mathcal{P}_W(||H_b||)||Z_{b,i}||^2)}{\lambda \sigma_N^2} .
$$

Denoting $L_b = n_c C(G_b^2) + L_b^{(2)} + L_b^{(4)}$ and $S_B(L_1') = \sum_{b=1}^{B} L_b'$, define $\mathcal{I}_\gamma'' = \{ S_B(L_1') \leq \log \gamma \}$. Using Claim 2 in Appendix A, we obtain $S_B(L_1) \leq S_B(L_1')$. From this inequality, we have $\mathbb{P}[\mathcal{I}_\gamma'] \geq \mathbb{P}[\mathcal{I}_\gamma'']$.

Next, we intersect the event $\mathcal{I}_\gamma''$ with certain high probability events so as to further lower bound $\mathbb{P}[\mathcal{I}_\gamma'']$ appropriately. Towards defining the events, consider mutually independent collection of i.i.d. $\mathcal{N}(0, 1)$ random variables $\{ U_b^{(i)}, b \in [1 : B] \}$, for $i \in \{1, 2\}$. Also, recall the definition of the notation $C(G^2)$ from Section II and denote $W_B^{(1)} = (||C(G^2)||_1 - B n_c \bar{P}) / \sqrt{\lambda [n_c C(G_b^2)]}$. Then, for some positive constant $c_1$ (depending on the parameters fixed in this analysis), define
the event \( \mathcal{E}_2 = \left\{ W_B^{(1)} \leq \sum_{b=1}^{B} U_{b}^{(1)} + 2c_1 \log B \right\} \). Further, let \( W_B^{(2)} = \sum_{b=1}^{B} \left( L_{b}^{(4)} / \sqrt{\mathbb{V}[L_{1}^{(4)}]} \right) \).

Then, for \( c_2 > 0 \) (depending on the other parameters), let

\[
\mathcal{E}_3 = \left\{ W_B^{(2)} \leq \sum_{b=1}^{B} U_{b}^{(2)} + 2c_2 \log B \right\}.
\]

In order to show that the event \( \mathcal{E}_2 \cap \mathcal{E}_3 \) is indeed a high probability event, we make use of the strong approximation principle of partial sum of i.i.d. random variables. Towards that, since \( \mathbb{E} [ |H_1|^2 ] < \infty \), observe that there exists \( t_1 > 0 \) such that \( \mathbb{E} [ \exp (tC(G_1^2))] < \infty \), for \( |t| \leq t_1 \).

Similarly, since \( \mathcal{P}_{WF}(\cdot) \leq \lambda \), there exists \( t_2 > 0 \) such that \( \mathbb{E} [ \exp (tL_1^{(4)})] < \infty \), for \( |t| \leq t_2 \). Hence, using strong approximation of partial sum of i.i.d. random variables (25), Theorem 2.6.2), there exist positive constants \( c_3, \ c_4 \) such that \( \mathbb{P} [ \mathcal{E}_2^c \cup \mathcal{E}_3^c ] \leq c_3 / B^{c_4} \).

Next, we note that

\[
\nu_1 = \mathbb{V} [ n_c C(G_1^2)] + \mathbb{V}[L_1^{(4)}].
\]

Also, let

\[
\nu_2 (X', H) \equiv \nu_2 = \frac{1}{B} \sum_{b=1}^{B} \sum_{i=1}^{2} \frac{2H_{b,i}^2 \| X_{b,i}' \|^2}{\sigma_N^2 (1 + G_{b}^2)^2}.
\]

Finally, let \( c_5 = \max\{c_1, c_2\} \), where \( c_1 \) and \( c_2 \) are as in the definitions of \( \mathcal{E}_2 \) and \( \mathcal{E}_3 \). Then, using the definition of events \( \mathcal{E}_1 \) and \( \mathcal{E}_2 \), the strong approximation bound mentioned above, and the fact that sum of independent standard normal random variables is a Gaussian random variable with appropriate mean and variance, we lower bound \( \mathbb{P} [ T_n'' ] \) as

\[
\mathbb{P} [ T_n''] \geq \mathbb{E} \left[ \Phi \left( \frac{\log \gamma - Bn_c C(\bar{P}) - c_5 \log B}{\sqrt{B (\nu_1 + \nu_2 (X', H))}} \right) \right] - c_3 / B^{c_4}. \tag{22}
\]

Next, conditioned on \( H = h, \ X' = x' \), we further lower bound \( \nu_2 \) appearing on the RHS of (22) by optimizing over choice of \( x' \in \mathcal{X}_n(\bar{P}) \). Towards that, we note that

\[
\min_{x' \in \mathcal{X}_n(\bar{P})} \nu_2 = \min_{\mathcal{P} \in \mathcal{P}_B(\bar{P})} \mathbb{E}^{B}_{V_1, V_2} \left[ \frac{2V_1^2 \mathcal{P}(V_1)}{\sigma_N^2 (1 + f(V_1, V_2)^2)} \right], \tag{23}
\]

where \( V_1, V_2 \) are independent random variables with common distribution being the empirical distribution of \( h, \ \mathbb{E}^{B}_{V_1, V_2} [\cdot] \) denotes the expectation with respect to the empirical distribution of \( h, \ f(V_1, V_2) = (V_1^2 + V_2^2) \sqrt{\mathcal{P}_{WF}(V_1^2 + V_2^2)}/\sigma_N \) and

\[
\mathcal{P}_B(\bar{P}) \triangleq \{ \mathcal{P} : \mathbb{R} \rightarrow \mathbb{R}_+, \mathbb{E}^B_{V_1}[\mathcal{P}(V_1)] = \bar{P} \}. \]

Next, we note that \( \mathcal{P} \) that attains the above minimum has the property \( \mathcal{P}_B^*(h) \leq \lambda \), for \( h \in \{h_{b,i}, b \in [1 : B], i = 1, 2\} \). This is justified by noting that for \( \mathcal{P}_B^* \) that attains the above lower bound, replacing \( \mathcal{P}_B^* \) with \( \tilde{\mathcal{P}}(h) = \mathcal{P}_B^*(h) \) if \( h > 1/\sqrt{\lambda} \), \( \mathcal{P}_B(\bar{P}) \geq \tilde{\mathcal{P}}(h) \), and \( \tilde{\mathcal{P}}(h) = \mathcal{P}_B(\bar{P}) \), else. With this, we further lower bound \( \nu_2 \) if we choose \( \log \gamma < Bn_c C(\bar{P}) + c_5 \log B \).
We will ensure that the choice of $\gamma$ is subject to this constraint. Using $\mathcal{P}_s(h) \leq \lambda$, next we apply McDiarmid’s inequality ([36], Theorem 6.2) to replace $\nu_2$ with $\mathbb{E}[\nu_2] - \delta'_B$, for some $\delta'_B = o(\sqrt{B})$ to further lower bound the RHS of (22). Thus, we obtain $\mathbb{P}[\mathcal{I}_B'] \geq \epsilon + c_6/n^{c_4}$ by choosing $\log \gamma = Bn_c C(\bar{P}) + \sqrt{V_{\text{BF}}(\bar{P})} \Phi^{-1}(\epsilon + \sqrt{n}) + c_5 \log B$ and $c_6 > 0$. Here, $V_{\text{BF}}(\bar{P})$ is as in (5) and is obtained as a lower bound to $\nu_1 + \mathbb{E}[\nu_2]$. Also, condition on $\gamma$ is ensured as $\epsilon < 1/2$. This concludes the proof.

\[ \square \]

**Appendix B**

**Proof of Theorem 2**

**A. Lower bound on rate with AP constraint**

The proof is similar to that of the lower bound with PP constraint in Section A-A. With the same setting as therein and decoupling coding and power control policy, it is immediate to verify that $\mathbb{E}[H](\sum_{b=1}^{n} \frac{2^m}{n} e^{2} \mathcal{X}^2_{[b]}(m)) = 2Bn_c \bar{P}$ is satisfied for all $m$. Hence, using the same analysis as in the PP case (excluding the analysis of PP constraint violation event), $R_a^* \geq C(\bar{P}) + \sqrt{\frac{V_{\text{BF}}(\bar{P})}{n}} \Phi^{-1}(\epsilon) + O\left(\frac{\log n}{n}\right)$. \[ \square \]

**Upper bound on rate with AP constraint**

The proof follows exactly along the same lines of the proof in Section A-B till lower bounding the event $\mathbb{P}[\mathcal{I}_1]$ in Section A-B3. In case of AP case, we make use of Lemma 6 to lower bound this probability term. From Lemma 6, for some positive constants $c_9$, $c_{10}$ and an appropriate choice of $\gamma$, $\mathbb{P}[\mathcal{I}_1] \geq \epsilon + c_9/n^{c_{10}}$. Using this estimate on the relaxed meta converse as in PP case, $\log M_a^* \leq \log \gamma + c_{10} \log n - \log c_9$. With $V_{\text{BF}}(\bar{P})$ as in (5), choose $\log \gamma = n C(\bar{P}) + \Phi^{-1}(\epsilon) \sqrt{n V_{\text{BF}}(\bar{P})} + o(\sqrt{n})$. Plugging this in the above bound, $R_a^* \leq C(\bar{P}) + \Phi^{-1}(\epsilon) \sqrt{\frac{V_{\text{BF}}(\bar{P})}{n}} + o\left(\frac{1}{\sqrt{n}}\right)$. \[ \square \]

**Lemma 6.** Let $\zeta_1$ be defined as in Section A-B2 and $\epsilon$ be the average probability of error for the channel in Section (II-A). There exists positive constants $c_9 > 0$ and $c_{10} > 0$ such that $\mathbb{P}[\mathcal{I}_1] \geq \epsilon + c_9/n^{c_{10}}$.

**Proof.** In proving the lemma, we make use of the method and notation in Lemma 5. With AP constraint as well, using the arguments therein, it is easy to observe $\mathbb{P}[\mathcal{I}_1] = \mathbb{P}[\mathcal{I}_1']$. Next, to lower bound $\mathbb{P}[\mathcal{I}_1']$, define $L_{b}^{(2)} = \sum_{i=1}^{2} \frac{n_{i} P(H_{i}) \sigma^2_{i}}{\lambda \sigma^2_{i}}$, where $\mathbb{E}[P(H_1)] = \bar{P}$. Also, define $L_{b}^{(6)} = \sum_{i=1}^{2} \frac{-P_{\text{BF}}(H_1) ||Z_{b,i}||^2}{\lambda \sigma^2_{i}}$. Denoting $\tilde{L}_{b} = n_c C(G_b^2) + L_{b}^{(2)} + L_{b}^{(5)} + L_{b}^{(6)}$ and $S_B(\tilde{L}) = \sum_{b=1}^{B} \tilde{L}_{b}$, define
\( \tilde{\mathcal{I}}_\gamma = \{ S_B(\tilde{L}) \leq \log \gamma \} \). Using Claim 2 in Appendix A, we obtain \( S_B(L) \leq S_B(\tilde{L}) \). From this inequality, we have \( \mathbb{P}[\tilde{\mathcal{I}}'_\gamma] \geq \mathbb{P}[\tilde{\mathcal{I}}_\gamma] \).

Next, as in the PP case, we intersect the event \( \tilde{\mathcal{I}}_\gamma \) with certain high probability events so as to further lower bound \( \mathbb{P}[\tilde{\mathcal{I}}_\gamma] \) appropriately. Towards defining the events, consider mutually independent collection of i.i.d. \( \mathcal{N}(0,1) \) random variables \( \{U_b^{(i)}, b \in [1 : B]\}, \) for \( i \in \{1, 2, 3\} \). Define \( W_B^{(1)} \) and \( \mathcal{E}_2 \) as in Lemma 5. Further, let

\[
W_B^{(3)} = \sum_{b=1}^{B} \left( \left( L_b^{(5)} - \mathbb{E} \left[ L_1^{(5)} \right] \right) / \sqrt{\mathbb{V} \left[ L_1^{(5)} \right]} \right).
\]

For \( c_3 > 0 \) (depending on the other parameters), denote \( \mathcal{E}_4 = \left\{ W_B^{(3)} \leq \sum_{b=1}^{B} U_b^{(2)} + 2c_3 \log B \right\} \). Similarly, for \( L_b^{(6)} \) as defined previously, define

\[
W_B^{(4)} = \sum_{b=1}^{B} \left( \left( L_b^{(6)} - \mathbb{E} \left[ L_1^{(6)} \right] \right) / \sqrt{\mathbb{V} \left[ L_1^{(6)} \right]} \right).
\]

For some \( c_4 > 0 \) (depending on the other parameters), let \( \mathcal{E}_5 = \left\{ W_B^{(4)} \leq \sum_{b=1}^{B} U_b^{(3)} + 2c_4 \log B \right\} \).

In order to show that the event \( \mathcal{E}_2 \cap \mathcal{E}_4 \cap \mathcal{E}_5 \) is indeed a high probability event, we make use of the strong approximation principle of partial sum of i.i.d. random variables. Here, we make use of the additional assumption that \( \mathbb{E}[\mathcal{P}^{(2+\delta)}(\|H_1\|))] < \infty \), for any small \( \delta > 0 \), in the statement of Theorem 2. Then, using the arguments invoked in Lemma 5 using strong approximation of partial sum of i.i.d. random variables ([37], Theorem 12.7), there exist positive constants \( c_7, c_8 \) such that \( \mathbb{P}[\mathcal{E}_2 \cup \mathcal{E}_4^c \cup \mathcal{E}_5^c] \leq c_7 / B^{c_8} \). Now, invoking the same lines of argument as in Lemma (5), we obtain the required result.

\section*{Appendix C}

\textbf{Proof of Theorem 3}

We will make use of Lemma 2 in our proof. Furthermore, we assume that the energy harvesting constraints hold with equality. This will not impact the second order term (see [10], [21] for this trick known as the Yaglom map trick). We also assume that the energy harvesting constraint holds for \([B, n_c]\)th slot only, i.e., \( \|X'\|_2 \leq \|E\|_1 \) a.s.. This is a relaxation of the constraint in Section II-B3 and hence can only improve the rate. To proceed, we need the following definitions. Crucially, we rely on the notation introduced in Section II. Let

\[
L_b^{(1)} = \sum_{i=1}^{2} \frac{H_{b,i}^2 \|X'_{b,i}\|_2^2}{\sigma_N^2 (1 + G_b^{2})}, \quad L_b^{(2)} = \sum_{i=1}^{2} \frac{H_{b,i} \langle X'_{b,i}, Z_{b,i} \rangle}{\sigma_N^2 (1 + G_b^{2})},
\]
\[ L_b^{(3)} = - \sum_{i=1}^{2} \frac{G_b^2 \|Z_{b,i}\|^2}{\sigma^2_N (1 + G_b^2)} \cdot L_b = n_c C(G_b^2) + \frac{3}{\ell} L_b^{(\ell)}. \]

Using the notation from Section II, \( S_B(L) = \sum_{b=1}^{B} L_b \) and \( S_B(L') = \sum_{b=1}^{B} L'_b \), where,

\[ L_b' = n_c C(G_b^2) + \frac{\sigma^2_N \|E_b\|_1 - \mathcal{P}_W(|H_b|)||Z_b||^2}{\lambda \sigma^2_N} + L_b^{(2)}. \]

Let \( \mathcal{E} = \{ L_B \leq \log \gamma_n \} \) and \( \mathcal{E}_0 = \{ L_B \leq \log \gamma_n \} \). For \( i \in \{1, 2, 3\} \), let \( \{ U_b^{(i)}, b \in [1 : B] \} \) be mutually independent collection of i.i.d. \( \mathcal{N}(0, 1) \) random variables.

Define \( W_B^{(1)} = [||C(G^2)||_1 - Bn_c C(\bar{E})] / \sqrt{\lambda \sigma^2_E} \), where the notation \( C(G^2) \) is as mentioned in Section II. For some positive constant \( c_1 \) (depending on other parameters), define

\[ \mathcal{E}_1 = \{ W_B^{(1)} \leq \sum_{b=1}^{B} U_b^{(1)} + 2c_1 \log B \}. \]

Let \( W_B^{(2)} = [||E||_1 - Bn_c \bar{E}] / \sqrt{\lambda \sigma^2_E} \). For \( c_2 > 0 \) (depending on the other fixed parameters), define

\[ \mathcal{E}_2 = \{ W_B^{(2)} \leq \sum_{b=1}^{B} U_b^{(2)} + 2c_2 B^{1/4} \}. \]

Let \( L_b^{(4)} = \mathcal{P}_W(|H_b|)||Z_b||^2 - n_c \bar{E} \) and \( W_B^{(3)} = \sum_{b=1}^{B} L_b^{(4)} / [\mathbb{V}[L_1^{(4)}]]^{1/2} \). For \( c_3 > 0 \) (depending on other parameters), let

\[ \mathcal{E}_3 = \{ W_B^{(3)} \geq \sum_{b=1}^{B} U_b^{(3)} - 2c_3 \log B \}. \]

From the fact that \( \mathbb{E}[|H_1|^2] < \infty \), observe there exists \( t_1 > 0 \) such that \( \mathbb{E}[\exp(t C(G_1^2))] < \infty \), for \( |t| \leq t_1 \). Similarly, since \( \mathcal{P}_W(\cdot) \leq \lambda \), there exists \( t_2 > 0 \) such that \( \mathbb{E}[\exp(t L_1^{(4)})] < \infty \), for \( |t| \leq t_2 \). By assumption, \( \mathbb{E}[E_{1,1}^4] < \infty \). Hence, using strong approximation of partial sum of i.i.d. random variables (see [25], Theorem 2.6.2 and [38], Theorem 4), there exist positive constants \( c_4, c_5 \) such that \( \mathbb{P} \left[ \bigcup_{i=1}^{3} \mathcal{E}_i^c \right] \leq c_4 / B^{c_5} \).

Finally, let \( \nu_1 \triangleq n c \mathbb{V}[C(G_1^2)] + n \sigma^2_E / \lambda^2 + B \mathbb{V}[L_1^{(4)}] \) and \( \nu_2 \triangleq \nu_1 + \sum_{b=1}^{B} 2L_b^{(1)} / (1 + G_b^2) \).

Now, we outline the key steps of the proof. Consider a scheme, i.e. distribution \( P_{X'_t | Y, H} \), that satisfies the relaxed energy harvesting constraint (mentioned at the beginning of the proof) and attains an average probability of error no greater than \( \epsilon \). We fix the auxiliary channel as described in Section A. With this choice, we consider lower bounding the probability term in (12). That is, \( \mathbb{P} \left[ \log \frac{dP_{X'_t | Y, H}}{dP_{X'_t | Y, H}} \leq \log \gamma_n \right] \overset{(a)}{=} \mathbb{P} [\mathcal{E}] \overset{(b)}{=} \mathbb{P} [\mathcal{E}_0] \). Here, (a) follows from the fact that the corresponding random variables (in the definition of the events) are equal in distribution. Next, (b) follows from Claim 2 in Appendix A and the assumption of relaxed energy harvesting.
constraint. Now, we further lower bound $\mathbb{P}[\mathcal{E}_0]$ using the strong approximation result mentioned above. Specifically, note that $\mathbb{P}[\mathcal{E}_0] \geq \mathbb{P}[\bigcap_{i=0}^{3} \mathcal{E}_i]$. However,

$$\mathbb{P} \left[ \bigcap_{i=0}^{3} \mathcal{E}_i \right] \geq \Phi \left( \frac{\log \gamma_n - n \mathbb{C}(\mathcal{E})}{\sqrt{\nu_2}} \right) - \delta_n.$$ 

Here, with $\delta_n = (c_4 n^{c_5})/n^{c_5}$, the bound follows from strong approximation bound $\mathbb{P} \left[ \bigcup_{i=1}^{3} \mathcal{E}_i^c \right] \leq c_4/B^{c_5}$. Next, we observe

$$\Phi \left( \frac{\log \gamma_n - n \mathbb{C}(\mathcal{E})}{\sqrt{\nu_2}} \right) \geq \Phi \left( \frac{\log \gamma_n - n \mathbb{C}(\mathcal{E})}{\sqrt{n \mathbb{V}_{\mathbb{E}}(\mathcal{E})}} \right)$$

if we ensure $\log \gamma_n < n \mathbb{C}(\mathcal{E})$ and noting that $\nu_2 \geq n \mathbb{V}_{\mathbb{E}}(\mathcal{E}) (V_{\mathbb{E}}(\mathcal{E}))$ as defined in the statement of Theorem 3. Finally, $\Phi \left( \frac{\log \gamma_n - n \mathbb{C}(\mathcal{E})}{\sqrt{n \mathbb{V}_{\mathbb{E}}(\mathcal{E})}} \right) - \delta_n \geq \epsilon + \delta_n$ is obtained by choosing $\log \gamma_n = n \mathbb{C}(\mathcal{E}) + \sqrt{n \mathbb{V}_{\mathbb{E}}(\mathcal{E})} \Phi^{-1}(\epsilon + 2\delta_n)$. Note that, as required, $\log \gamma_n < n \mathbb{C}(\mathcal{E})$ can be ensured for appropriately chosen $n$, as $\epsilon < 1/2$. Thus, we have shown $\mathbb{P}[\mathcal{E}_0] \geq \epsilon + \delta_n$. The result follows by plugging in the above estimate and $\gamma_n$ in (12), and applying Taylor’s theorem. 

\section*{Appendix D}

\subsection*{Proof of Theorem 4}

First, with $a_n$ as in Definition [1], we consider lower bounding $\lim_{n \to \infty} \frac{1}{na_n^2} \log p_{e,av}^*(n, M_n)$ under PP constraint. Consider any $(n, M_n, \varepsilon_n, \bar{P})$ code with $M_n = \exp(n(C(\bar{P}) - a_n))$. Fix $\theta > 1$ and let $\gamma = \exp(n(C(\bar{P}) - \theta a_n))$. Consider the auxiliary channel in Section A-B1. Invoking the meta converse bound and applying the relaxations as in Appendix (A) (see Section A-B2 therein), we obtain $p_{e,av}^*(n, M_n) \geq \mathbb{P}[^{\mathcal{I}_3}] - \exp(na_n(\theta - 1))$. Using a similar analysis as in Appendix A-B that involves lower bounding the probability term $\mathbb{P}[^{\mathcal{I}_3}]$ in terms of the probability of sum of i.i.d. standard Gaussian random variables, and invoking the result from [39] (Theorem 3.7.1), we obtain the required lower bound as in [12]. A similar analysis yields $V_{\mathbb{E}}^n(\mathcal{E})$ as lower bound on the moderate deviation constant under EH constraint. With an additional moment constraint on the power allocation function (as in the statement of Theorem 2), we obtain $V_{\mathbb{E}}(\bar{P})$ as the lower bound for the AP constraint as well.

For the lower bound for PP, AP and EH constraint, the $\beta \beta$ bound and the power control policies explained in Appendix A-A, B-A and V-A respectively. We fix the same auxiliary channel as in the analysis therein. For appropriate $\delta_n > 0$, with i.i.d. $\mathcal{N}(0, 1 - \delta_n)$ input in the achievability proof for PP, we observe that there exists $(n, M_n, \varepsilon_n, \bar{P})$ codes such that $\log M_n^*(n, \varepsilon_n, \bar{P}) \geq$
\[- \log \gamma_0 + O(1) \text{ (see the proof in Appendix A-A)}. \]

Choose \( \log \gamma_0 = nC(\bar{P}) - \theta n a_n \), \( 0 < \theta < 1 \) so that we obtain the existence of codes with \( \log M^*_{n}(n, \epsilon_n, \bar{P}) \geq nC(\bar{P}) - \theta n a_n \), for \( n \) large, by [39], Theorem 3.7.1, \( \limsup_{n \to \infty} \frac{\log p_{\text{avg}}^*(n,M_n)}{n} \leq -\frac{\theta^2}{2V_{BF}(\bar{P})} \). Taking \( \theta \to 1 \), we obtain the required result. A similar analysis yields the result for the AP and EH constraint case as well. \( \square \)

REFERENCES

[1] M.-S. Alouini and A. J. Goldsmith, “Capacity of rayleigh fading channels under different adaptive transmission and diversity-combining techniques,” IEEE Transactions on Vehicular Technology, vol. 48, no. 4, pp. 1165–1181, 1999.

[2] J. Tang and X. Zhang, “Quality-of-service driven power and rate adaptation for multichannel communications over wireless links,” IEEE Transactions on Wireless Communications, vol. 6, no. 12, 2007.

[3] A. J. Goldsmith and P. P. Varaiya, “Capacity of Fading Channels with Channel Side Information,” IEEE Transactions on Information Theory, vol. 43, no. 6, pp. 1986–1992, 1997.

[4] H. Ji, S. Park, J. Yeo, Y. Kim, J. Lee, and B. Shim, “Ultra-reliable and low-latency communications in 5g downlink: Physical layer aspects,” IEEE Wireless Communications, vol. 25, no. 3, pp. 124–130, 2018.

[5] A. El Gamal and Y.-H. Kim, Network Information Theory. Cambridge university press, 2011.

[6] G. Caire, G. Taricco, and E. Biglieri, “Optimum Power Control over Fading Channels,” IEEE Transactions on Information Theory, vol. 45, no. 5, pp. 1468–1489, 1999.

[7] R. Negi and J. M. Cioffi, “Delay-Constrained Capacity with Causal Feedback,” IEEE Transactions on Information Theory, vol. 48, no. 9, pp. 2478–2494, 2002.

[8] V. Strassen, “Asymptotische Abschätzungen in Shannon’s Informationstheorie,” in Trans. Third Prague Conf. Inf. Theory, 1962, pp. 689–723.

[9] M. Hayashi, “Information Spectrum Approach to Second-Order Coding Rate in Channel Coding,” IEEE Transactions on Information Theory, vol. 55, no. 11, pp. 4947–4966, 2009.

[10] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel Coding Rate in the Finite Blocklength Regime,” IEEE Transactions on Information Theory, vol. 56, no. 5, pp. 2307–2359, 2010.

[11] Y. Altuğ and A. B. Wagner, “Moderate deviations in channel coding,” IEEE Transactions on Information Theory, vol. 60, no. 8, pp. 4417–4426, 2014.

[12] Y. Polyanskiy and S. Verdú, “Channel dispersion and moderate deviations limits for memoryless channels,” in Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on. IEEE, 2010, pp. 1334–1339.

[13] R. Mahapatra, Y. Nijsure, G. Kaddoum, N. U. Hassan, and C. Yuen, “Energy Efficiency Tradeoff Mechanism Towards Wireless Green Communication: A Survey,” IEEE Communications Surveys and Tutorials, vol. 18, no. 1, pp. 686–705, 2016.

[14] Y. Polyanskiy and S. Verdú, “Scalar Coherent Fading Channel: Dispersion Analysis,” in Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on. IEEE, 2011, pp. 2959–2963.

[15] W. Yang, G. Caire, G. Durisi, and Y. Polyanskiy, “Optimum Power Control at Finite Blocklength,” IEEE Transactions on Information Theory, vol. 61, no. 9, pp. 4598–4615, 2015.

[16] E. MolavianJazi and J. N. Laneman, “On the second-order coding rate of non-ergodic fading channels,” in Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on. IEEE, 2013, pp. 583–587.

[17] G. Durisi, T. Koch, J. Östman, Y. Polyanskiy, and W. Yang, “Short-Packet Communications over Multiple-Antenna Rayleigh-Fading Channels,” IEEE Transactions on Communications, vol. 64, no. 2, pp. 618–629, 2016.
[18] A. Lancho, T. Koch, and G. Durisi, “A High-SNR Normal Approximation for Single-Antenna Rayleigh Block-Fading Channels,” in Information Theory (ISIT), 2017 IEEE International Symposium on. IEEE, 2017, pp. 1773–1777.

[19] R. Rajesh, V. Sharma, and P. Viswanath, “Capacity of Gaussian Channels with Energy Harvesting and Processing Cost,” Information Theory, IEEE Transactions on, vol. 60, no. 5, pp. 2563–2575, 2014.

[20] K. G. Shenoy and V. Sharma, “Finite blocklength achievable rates for energy harvesting awgn channels with infinite buffer,” in Information Theory (ISIT), 2016 IEEE International Symposium on. IEEE, 2016, pp. 465–469.

[21] S. L. Fong, V. Y. Tan, and A. Özgür, “On achievable rates of awgn energy-harvesting channels with block energy arrival and non-vanishing error probabilities,” IEEE Transactions on Information Theory, vol. 64, no. 3, pp. 2038–2064, 2018.

[22] P. Deekshith and V. Sharma, “Finite blocklength rates over a fading channel with csit and csir,” in 2018 IEEE International Conference on Communications (ICC). IEEE, 2018, pp. 1–7.

[23] S. V. Hanly and D. N. C. Tse, “Multiaccess fading channels. ii. delay-limited capacities,” IEEE Transactions on Information Theory, vol. 44, no. 7, pp. 2816–2831, 1998.

[24] M. Tomamichel and V. Y. Tan, “Second-order Coding Rates for Channels with State,” IEEE Transactions on Information Theory, vol. 60, no. 8, pp. 4427–4448, 2014.

[25] M. Csörgő and P. Révész, Strong approximations in probability and statistics. Academic Press, 2014.

[26] D. Tse and P. Viswanath, Fundamentals of wireless communication. Cambridge university press, 2005.

[27] M.-L. Ku, W. Li, Y. Chen, and K. R. Liu, “Advances in energy harvesting communications: Past, present, and future challenges,” IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1384–1412, 2016.

[28] G. Caire and S. Shamai, “On the capacity of some channels with channel state information,” IEEE Transactions on Information Theory, vol. 45, no. 6, pp. 2007–2019, 1999.

[29] S. L. Fong, V. Y. Tan, and J. Yang, “Non-asymptotic achievable rates for energy-harvesting channels using save-and-transmit,” IEEE Journal on Selected Areas in Communications, no. 99, 2015.

[30] W. Yang, A. Collins, G. Durisi, Y. Polyanskiy, and H. V. Poor, “Beta-beta bounds: Finite-blocklength analog of the golden formula,” IEEE Transactions on Information Theory, 2018.

[31] O. Ozel and S. Ulukus, “Achieving awgn capacity under stochastic energy harvesting,” IEEE Transactions on Information Theory, vol. 58, no. 10, pp. 6471–6483, 2012.

[32] O. B. Akan, O. Cetinkaya, C. Koca, and M. Ozger, “Internet of hybrid energy harvesting things,” IEEE Internet of Things Journal, vol. 5, no. 2, pp. 736–746, 2018.

[33] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, “5g wireless network slicing for embb, urllc, and mmtc: A communication-theoretic view,” IEEE Access, 2018.

[34] W. Feller, Introduction to Probability Theory and Its Applications. John Wyley, 1971, vol. 2.

[35] Y. Polyanskiy, “Channel Coding: Non-Asymptotic Fundamental Limits,” Ph.D. dissertation, Princeton Univ., Princeton, NJ, USA, 2010.

[36] S. Boucheron, G. Lugosi, and P. Massart, Concentration Inequalities: A Nonasymptotic Theory of Independence. Oxford university press, 2013.

[37] A. DasGupta, Asymptotic theory of statistics and probability. Springer Science & Business Media, 2008.

[38] J. Komlós, P. Major, and G. Tusnády, “An approximation of partial sums of independent rv’s, and the sample df. ii,” Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, vol. 34, no. 1, pp. 33–58, 1976.

[39] A. Dembo and O. Zeitouni, Large deviations techniques and applications, volume 38 of Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, 2010.