Power-Performance Analysis of a Simple One-Bit Transceiver

Kang Gao, N. J. Estes, Bertrand Hochwald, Jonathan Chisum, J. Nicholas Laneman
University of Notre Dame
Email: {kgao,nestes,bhochwald,jchisum,jnl}@nd.edu

***Abstract*—We analyze a one-bit wireless transceiver whose architecture is simple enough that its power versus performance profile can be modeled analytically. We then utilize multiple such transceivers in a communication system operating at millimeter-wave carrier frequencies. Various aspects of the system are analyzed, including the optimum achievable throughput for a given amount of total consumed power. An analogy is drawn between the "transceiver cell" proposed herein and a "computational cell" commonly used in neural networks that allows us to apply neural-network type algorithms to aid in difficult tasks such as channel estimation for a large number of transceivers.**

I. Introduction

Future generations of wireless systems, such as the fifth-generation (5G) systems currently being developed, are looking at carrier frequencies and bandwidths significantly higher than existing services. The FCC has recently proposed the use of spectrum at 28 GHz and 39 GHz for mobile services [1], with the hope that bandwidths in excess of 2 GHz and corresponding high data rates can be supported.

As carrier frequencies rise into the millimeter-wave and sub-millimeter-wave bands, and bandwidths rise into the GHz range, however, the design of low-cost, low-power, highly linear radio-frequency (RF) circuits becomes increasingly challenging. Standard CMOS circuits struggle at carrier frequencies in the tens of GHz to obtain good noise performance and low leakage with small process nodes. High-speed linear amplifiers, mixers, analog-to-digital converters (ADC's), and digital-to-analog converters (DAC's) in the transmitter and receiver chains become power hungry and expensive. For example, a 12-bit 4 Gsample/second ADC (Texas Instruments ADC12J4000) consumes two Watts (2 W) [2].

At the same time, link-budget calculations suggest that multiple transceiver chains and streaming or beamforming techniques are generally needed to communicate reliably in the millimeter-wave band. Indeed, so-called "massive MIMO" (multiple-input multiple-output), where dozens of transceiver chains are used, is being considered as a possible method to overcome fading, shadowing, and pathloss [3]. Although a large number of antennas and transceiver chains are conceivable for a base station, for a portable cell phone battery with 10 W-hr of stored energy, a single ADC could, by iteself, drain the battery in several hours of use. We cannot rely on radical advances in battery technology over the next five years, and for a phone that is expected to maintain power for 24 hours, we therefore have an allowable budget of only 0.5 W on average of consumption throughout the day for the entire phone, including display, computer, transceiver, and interfaces. The radio should have at most half of this budget, therefore giving an average power consumption of 250 mW or less for the combined transmitter and receiver, including sleep and idle modes.

A. Concerns on power consumption lead to simple transceivers

The wireless community is acknowledging that the combined forces of high cost and power per RF chain, multiplied by the requirement of having many chains, conspire to make high-performance, high-bandwidth, portable wireless devices impractical unless the transceiver chains are critically re-examined. A simple trade-off analysis suggests that relaxing some of the linearity requirements in a communication system can yield a beneficial outcome in a wireless system operating in the millimeter and sub-millimeter-wave bands.

The trade-off comes from exchanging bandwidth and the number of RF chains for linearity and effective number of bits. As carrier frequencies rise into the tens of GHz and above, bandwidth becomes more plentiful while linearity becomes more precious. At a carrier frequency of 100 GHz, having a bandwidth of even 5% of the carrier frequency is 5 GHz, which is an enormous playground to create a wireless system operating at very high data rates. To get 2.5 Gbits/second data rate, a single-carrier system operating at near 5 Gsymbols/second, with 0.5 bits/symbol is all that is needed. A single-carrier system transmitting less than one bit per symbol could potentially be achieved with transceivers that operate on the basic principle of π phase shifts. Such systems require minimal linearity and effective number of bits.

If the power consumption of each one-bit-per-symbol (1-bit) circuit is small enough, high spectral and power efficiency in bits/(second-Hz-Watt) or bits/(Hz-Joule) can be attained by employing multiple 1-bit circuits. This measure of efficiency, which we utilize, is similar to a traditional measure, bits/Joule, [4] that is often cited as a measure of energy efficiency. However, the bits/Joule measure typically assumes that bandwidth is not constrained (is unlimited) in attaining maximum efficiency. Although we utilize a large bandwidth, it is not allowed to grow without limit, and we find that the unit bits/(Hz-Joule) allows us to easily compare the energy efficiency of different systems with constrained bandwidth.

Since power consumption of an ADC scales approximately linearly in sampling rate and exponentially in the number of bits per sample [5], we can readily compensate in cost and power for high sampling rates by lowering the resolution of the quantizer. Specifically, a 1-bit quantizer, being itself just a simple comparator, can be made to consume only 100's of microWatts (μW) of power, even at GHz sampling rates [6]. Utilizing many 1-bit chains can yield large bits/(Hz-Joule).

B. Background in simple transceivers

Some early efforts in analyzing the effects of a non-linearity in wireline communication systems include [7] and [8], where information-theoretic arguments are used to derive the capacity of a system where a 1-bit ADC at the receiver is used. Early work in the analysis of a wireless channel includes [9]–[11], where the capacity for 1-bit ADC's and MIMO systems is examined, with known information about the channel.

Early works on the analysis of wireless multi-carrier systems such as orthogonal frequency-division multiplexing (OFDM) with quantization effects are [12], [13]. More recently, OFDM and single-carrier systems with 1-bit receiver ADC's in a massive MIMO system are studied in [14] and [15]. In [16], the information-theoretic capacity is derived for a receiver with an oversampling 1-bit ADC in a wireline communication system. A MIMO system with 1-bit ADC in each receive antenna is considered in [17] and [18]. In [19], the capacity of a flat-fading MIMO channel with 1-bit ADCs is analyzed. In [20] the capacity of a MIMO system with 1-bit ADCs at the receiver is analyzed and the effect of the number of paths, the number of transmit antennas and the number of receive antennas is analyzed. Other similar studies include [21]–[26], where various assumptions are made about the downlink and uplink and which end of the link has a 1-bit ADC, and how much is known about the channel. In [27], a multi-antenna base station system with single-antenna 1-bit receivers is analyzed.

While many of the above efforts consider low-resolution quantization effects at the transmitter or receiver, a few consider low resolution quantizers at both. For example, [28], designs a linear minimum-mean-squared-error precoder to mitigate quantization distortion. In [29], channel estimation with 1-bit ADC's and DAC's is considered.

The wireless transceiver we analyze is simple enough that the end-to-end power-performance profile can be modeled. We examine the use of multiple such transceivers in a communication system, and examine various performance aspects of the system. Of particular interest are optimizing the energy efficiency, as measured in bits/(Hz-Joule), and attaining high throughputs using simple transmitter and receiver processing.

C. One-bit transceiver cell

The sections of this paper, each of which deals with a separate aspect of the transceiver analysis, are summarized as follows.

1) Transceiver cell hardware architecture: We analyze a low-power one-bit BPSK transceiver cell whose design is simple enough that a power-performance profile may be modeled. The architecture of the transceiver is generic enough that its model is representative of a class of similar transceivers that can be designed for a variety of carrier frequencies and bandwidths. We show how the SNR at the receiver varies with the total power consumption at both the transmitter and the receiver.

2) Achievable rate: When there are M transmitter/receiver pairs, we show that the channel capacity (in bits/channel-use) increases linearly with M under certain conditions; however, it cannot generally achieve M bits/channel-use for $M \geq 2$. In a block-fading channel in which the channel is constant for T_b samples and unknown, and changes independently every interval T_b, the capacity is 0 for $T_b = 1$, no matter how many transceivers there are. The capacity of a single transmitter/receiver pair as a function of T_b and SNR is provided.

3) Machine learning and channel estimation: Classical channel estimation algorithms like the MMSE estimator have complexity that is generally prohibitive, especially for large number of transceivers. By drawing an analogy between a "computational cell" commonly used in neural networks and the proposed transceiver cell, we may apply sub-optimal but efficient neural-network-based algorithms for communication-related functions such as channel estimation. For example, the soft-support vector machine (SVM) algorithms used in machine-learning can be applied to estimate the channel with low complexity, even for a large number of transceivers.

4) Beamforming: The 1-bit quantized nature of the transceiver cell does not permit traditional high-

resolution processing techniques such as singular value decomposition based beamforming. We provide a framework for beamforming and show that, under certain circumstances, the sign of the channel is used to form the optimal beamforming vector at transmitter, and bit-error probabilities are used as beamforming weights at the receiver.

5) Energy efficiency: We calculate bits/(Hz-Joule) as a performance metric that allows us to profile the achievable rate (in bits/channel-use) versus system power consumption (in Watts). This guides the selection of the operational power of the system that achieves optimum energy efficiency.

II. Transceiver cell hardware architecture

Many components and circuits in a phone such as the power amplifier (PA), low-noise amplifier (LNA), voltage-controlled oscillators and phase-locked loops (VCO's and PLL's), and mixers, when operating at frequencies above 20 GHz and bandwidths above 2 GHz show markedly higher power consumption than their microwave narrowband counterparts. Multi-bit ADC's and DAC's, in particular, consume power on the order of Watts to digitize at GHz sample rates. Rather than selectively removing components of existing transceivers in an effort to save power, we propose to strip a radio transceiver into a basic cell that has a minimum number of subsystems. This allows us to map the consumed power at the transmitter and receiver to the performance of the overall system. The transceiver architecture is generic enough that its model is representative of similar one-bit transceivers that can be designed for a variety of carrier frequencies and bandwidths.

A quick way to get very low power consumption is to employ a simple comparator at the receiver to act as a 1-bit ADC, and equivalent switch as a 1-bit DAC at the transmitter. Recent research [6] shows that very high-speed 1-bit comparators can be made to operate at 100's of microWatts (μW) at sample rates extending to the GHz range. We therefore concentrate our modeling efforts on the radio-frequency portion of the hardware, which generally consumes considerably more.

To convey 1 bit per symbol, π phase modulation (BPSK) or on/off keying (OOK) structures can be used. In this paper, we focus on a BPSK transceiver.

A. BPSK Transmitter

A simple BPSK transmitter contains a 1-bit DAC with low quality-factor filter, a passive mixer for up-conversion to RF, and an RF power amplifier operating as shown in Fig. 1. The 1-bit nonlinear architecture permits the transmitter to operate in saturation for maximum efficiency. The mixer and DAC can operate in the sub-milliWatt range, as long as the PA has enough gain

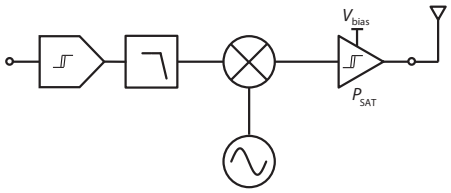

Fig. 1. Simple transmitter for BPSK showing 1-bit DAC, mixer, and PA

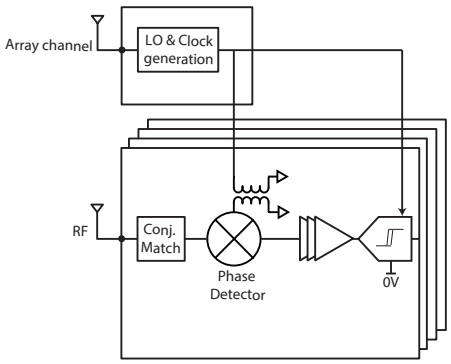

Fig. 2. Simple receiver requires a local oscillator phase reference driving a phase-detecting mixer

to amplify the signal for transmission; then the PA dominates the power consumption of the transmitter. At millimeter-wave frequencies, regardless of architecture (either saturated or harmonic tuned) a high efficiency power amplifier design is limited by device parasitics and the current state of the art yields maximum power-added efficiency (PAE) of between 35–45% [30]. We consider a PAE of $\eta_t = 45\%$, and therefore

$$P_{\rm tx}(W_{\rm tx}) = \eta_t W_{\rm tx}, \tag{1}$$

where $P_{\rm tx}$ is the over-the-air transmitter power ranging from sub-mW to 10 mW, and $W_{\rm tx}$ is the consumed transmitter power. This is a first-order representation of $P_{\rm tx}$ as a function of $W_{\rm tx}$, and represents the effect of transmitter performance (as measured by output power) as a function of transmitter consumed power.

B. BPSK Receiver

At the receiver, a PLL/VCO drives a mixer for down-conversion, and we assume, for simplicity, that there is no LNA, and therefore accept a correspondingly high noise figure. A block-diagram design of a simple BPSK receiver is shown in Fig. 2.

We consider the design of the PLL/VCO and frequency converter with the objective of minimizing consumed power. Frequency conversion mixers often include multiple nonlinear devices in a balanced configurations to provide port-to-port isolation, increase linearity, and suppress spurious products. However, local oscillator (LO) power is linearly proportional to the number of

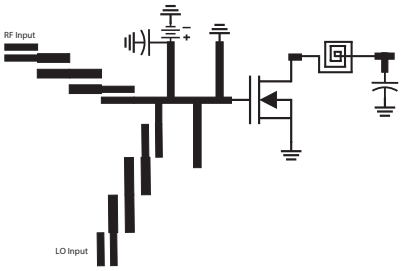

Fig. 3. A single FET nonlinear transconductance mixer was designed for low power in the Qorvo 0.15μm GaAs process. The FET is biased for low LO and RF power levels. The LO and RF combine at the gate, and IF is taken from a lumped-element low-pass filter on the drain.

nonlinear devices in the mixer, so the simple receiver uses a single nonlinear device in an effort to keep power consumption as low as possible. As a result, LO and RF feed-through to the IF port may be large and there will be spurious products. We accept these by-products of low power consumption.

The noise figure (decrease in SNR from the input to the output of a two-port network) of the entire receiver is then dominated by the conversion loss of the mixer, which decreases as the LO power increases. From [31], conversion loss (or NF) of a FET mixer is

$$\mathrm{NF}(W_{\mathrm{rx}}) = \frac{(2\pi f_{RF}\overline{C_{gs}}V_p^2)^2}{8R\eta_r W_{\mathrm{rx}} I_{DSS}^2}, \tag{2}$$

where f_{RF} is the RF frequency, R is the characteristic impedance of the mixer (usually $50\,\Omega$), $\overline{C_{gs}}$ is the gate source capacitance, V_p is the pinch-off voltage, I_{DSS} is the drain current for zero bias, η_r is the efficiency of the LO.

We design a simple FET mixer operating at 39 GHz in a Qorvo $0.15\,\mu$m GaAs foundry process, whose schematic is shown in Fig. 3. The corresponding parameters are $f_{RF} = 39$ GHz, $I_{DSS} = 10$ mA, $\overline{C_{gs}} = 10^{-13}$F, $V_p = 1$ V. We further consider the LO efficiency to be $\eta_r = 25\%$. Then, (2) can be simplified as

$$\mathrm{NF}(W_{\mathrm{rx}}) = \frac{\alpha}{W_{\mathrm{rx}}}, \tag{3}$$

where $\alpha = 60$mW. This equation, to first order, characterizes the performance (as measured by noise figure) of the receiver as a function of its consumed power.

III. System model

Using the transceiver cell, we can model a typical wireless channel and the corresponding signal-to-noise ratio (SNR) at the receiver as a function of the total consumed power at the transmitter and receiver.

A distance of 40 meters between transmitter and receiver is reasonable for systems operating in the millimeter-wave band. The pathloss at 39 GHz carrier frequency [32] is given by the formula $20\log_{10}\left(\frac{4\pi}{\lambda}\right) + 20\log_{10}(40) \approx 93$ dB, where λ is the wavelength at 39 GHz. We are targeting 2 GHz of bandwidth, and therefore the noise power is $N_0 + 10\log_{10}(2\times 10^9) \approx -81$ dBm where $N_0 = -174$ dBm/Hz is the thermal noise level. With $G_{\mathrm{tx}} = 5$ dB of antenna gain at the transmitter and $G_{\mathrm{rx}} = 5$ dB at the receiver, we have a receive signal strength of -73 dBm when the transmitter power is at its maximum of 10 dBm. Our signal-to-noise-ratio (SNR) at a receiver situated 40 m from the transmitter is then

$$\begin{aligned}\rho(W_{\mathrm{tx}}, W_{\mathrm{rx}}) &= P_{\mathrm{tx}}(W_{\mathrm{tx}}) - 93 + 81 + 5 + 5 - \mathrm{NF}(W_{\mathrm{rx}}) \\ &= P_{\mathrm{tx}}(W_{\mathrm{tx}}) - \mathrm{NF}(W_{\mathrm{rx}}) - 2,\end{aligned} \tag{4}$$

where $P_{\mathrm{tx}}(W_{\mathrm{tx}})$ and $\mathrm{NF}(W_{\mathrm{rx}})$ are shown in (1) and (3).

The total power consumption of a transceiver cell is $W_{\mathrm{rx}} + W_{\mathrm{tx}}$. Clearly, increasing either W_{rx} or W_{tx} improves $\rho(W_{\mathrm{tx}}, W_{\mathrm{rx}})$ by increasing P_{tx} and decreasing NF.

A. Received signal model

We use a standard multi-transceiver set-up where there are M transmitters and N receivers. The baseband representation of the transmitted signal is the $M \times 1$ vector $\mathbf{x}$. BPSK modulation is used at each transmitter and the baseband transmitted signals are represented as $\mathbf{x} \in \{\pm 1\}^M$.

The baseband-equivalent of the channel matrix, denoted H, represents the wireless environment between the transmitter and receiver. We assume a relatively flat channel since the millimeter-wave frequencies we are employing tend to have limited delay-spread [32].

The received signal is represented by the baseband $N \times 1$ vector $\mathbf{y}$. Depending on the nature of the transceiver cell, the entries of this vector either represent downconverter quantized I or Q phase information. In this paper, we assume only I branch is used at the receiver. The channel, which is traditionally modeled as complex-valued because of the possible quadrature (Q) component of received signal, can instead simply be modeled as real-valued since the imaginary part is not demodulated. The received signal is then

$$\mathbf{y} = \mathrm{sign}(\sqrt{\frac{\rho}{M}}H\mathbf{x} + \mathbf{v}), \tag{5}$$

where $\mathbf{x} \in \{\pm 1\}^M$, $\mathbf{y} \in \{\pm 1\}^N$, H is a $N \times M$ matrix, $\mathbf{v}$ is a $N \times 1$ vector, ρ is the expected received SNR at each receive antenna. The additive noise $\mathbf{v} \sim \mathcal{N}(0, I)$ is independent of H and $\mathbf{x}$. The function $\mathrm{sign}(\cdot)$ provides the sign of the input as its output. The channel matrix H is scaled so that

$$\mathbb{E}[(H\mathbf{x})^H(H\mathbf{x})] = M\mathbb{E}[\mathbf{v}^H\mathbf{v}] = MN \tag{6}$$

and one example of H is Rayleigh independent fading channel where elements of H are i.i.d. Gaussian $\mathcal{N}(0, 1)$.

B. Transition probability matrix vs channel matrix:

Because of the single-bit nature of these transceivers, (5) describes a discrete memoryless channel (DMC) with transition probability

$$p\left(\mathbf{y}|\mathbf{x}\right)=\prod_{k=1}^{N}Q\left(-y_k[\sqrt{\frac{\rho}{M}}H\mathbf{x}]_k\right) \tag{7}$$

where y_k is the kth element of $\mathbf{y}$, $[\cdot]$ is the kth element of the enclosed vector, and $Q(\cdot)$ is the well-known Q-function.

The channel of a DMC with 2^M possible inputs and 2^N possible outputs can be described by a $2^M \times 2^N$ transition probability matrix (TPM), whose elements are $p\left(\mathbf{y}|\mathbf{x}\right)$ with $\mathbf{x}\in\{\pm1\}^M$ and $\mathbf{y}\in\{\pm1\}^N$, determined by ρ and H using (7). Therefore, the TPM is fully determined by H (for known ρ). On the other hand, when the $2^M \times 2^N$ TPM is given, H is uniquely determined. Therefore, the TPM and channel matrix H are in a one-to-one mapping.

One example with $M=N=1$ is actually a binary symmetric channel (BSC) with crossover probability

$$p_b=Q\left(-\sqrt{\rho}h\right), \tag{8}$$

where h is the channel scalar.

C. Comments on modulation and synchronization

Because of the constrained nature of these transceivers, we are confined to single-carrier modulation methods. In general, we assume timing and frequency synchronization are available at the receiver. Fortunately, the burdens of synchronization, while not trivial, do not scale with the number of transceivers.

IV. Achievable Rate

When there are M transmitter/receiver pairs, we show that uniform input is optimal when receiver knows the channel and the channel capacity (in bits/channel-use) increases linearly with M; however, it cannot generally achieve M bits/channel-use for $M\geq 2$ even at infinite SNR. A Rayleigh independent fading channel is considered, where elements of H are i.i.d. $\mathcal{N}(0,1)$, and ρ is known.

A. Receiver knows the channel

We assume the channel is known perfectly to the receiver. Then, the capacity of the system is

$$C_R=\max_{p_{\mathbf{x}}(\cdot),\mathbf{x}\in\{\pm1\}^M}\mathbf{I}(\mathbf{x};\mathbf{y},H) \tag{9}$$

where $p_{\mathbf{x}}(\cdot)$ is the input distribution independent of H, and $\mathbf{I}(\cdot;\cdot)$ is the mutual information notation. We show that the uniform input distribution on $\{\pm1\}^M$ achieves capacity.

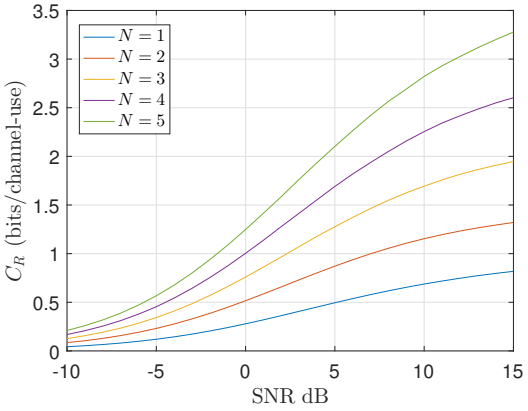

Fig. 4. Capacity (13) versus SNR with $M=N=1,\ldots,5$, showing saturation at approximately 10-15 dB.

For any input distribution $p_{\mathbf{x}}(\mathbf{x})$, the achievable rate is

$$R(p_{\mathbf{x}}(\mathbf{x}))=\mathbf{I}(\mathbf{x};\mathbf{y},H),\mathbf{x}\sim p_{\mathbf{x}}(\mathbf{x}) \tag{10}$$

We have

$$R(p_{\mathbf{x}'}(\mathbf{x}))=R(p_{\mathbf{x}}(\mathbf{x})) \tag{11}$$

where $p_{\mathbf{x}'}(\mathbf{x})$ is the distribution of $\mathbf{x}'$ with

$$\mathbf{x}'=U\mathbf{x},U=\text{diag}(\mathbf{u}),\mathbf{u}\in\{\pm1\}^M, \tag{12}$$

$\text{diag}(\mathbf{u})$ is a diagonal matrix with $\mathbf{u}$ as diagonal elements, since HU^H and H have the same distribution. Since the mutual information is a concave function of the input distribution $p_{\mathbf{x}}(\mathbf{x})$,

$$\begin{aligned}R(p_{\mathbf{x}}(\mathbf{x}))&=\frac{1}{2^M}\sum_{\mathbf{x}'=U\mathbf{x},U=\text{diag}(\mathbf{u}),\mathbf{u}\in\{\pm1\}^M}R(p_{\mathbf{x}'}(\mathbf{x}))\\&\leq R\left(\frac{1}{2^M}\sum_{\mathbf{x}'=U\mathbf{x},U=\text{diag}(\mathbf{u}),\mathbf{u}\in\{\pm1\}^M}p_{\mathbf{x}'}(\mathbf{x})\right)\\&=R(p_{\bar{\mathbf{x}}}(\mathbf{x})),\end{aligned}$$

where $\bar{\mathbf{x}}$ is uniformly distributed on $\{\pm1\}^M$. This shows that uniform input achieves capacity.

With a uniformly distributed input, we obtain

$$C_R=\mathbb{E}_H\left[\mathbb{E}_{\mathbf{x}}\left[\sum_{\mathbf{y}\in\{\pm1\}^N}p(\mathbf{y}|\mathbf{x})\log_2\frac{p(\mathbf{y}|\mathbf{x})}{\mathbb{E}_{\mathbf{x}}[p(\mathbf{y}|\mathbf{x})]}\right]\right]. \tag{13}$$

The simulated results of (13) with $M=N$ as a function of SNR and $M=N=1,\ldots 5$ are shown in Fig. 4. The results show that the capacity saturates at approximately $10-15$ dB.

B. High SNR

We analyze the performance limit of the system by computing the capacity of the system at high SNR, and

where the TPM is known to the receiver. As $\mathrm{SNR} \to \infty$, the TPM comprises 1's and 0's since $\mathbf{y}$ is determined from $\mathbf{x}$ for a given H. The transition probability becomes

$$p\left(\mathbf{y}|\mathbf{x}\right) = \mathbb{1}\left(\mathbf{y} = \mathrm{sign}(H\mathbf{x})\right), \tag{14}$$

where $\mathbb{1}(\cdot)$ outputs 1 when the input is true and outputs 0 otherwise.

For a single transmitter/receiver pair, we have $p_b = 0$ or $p_b = 1$. Then, the capacity of the BSC is $C_b = 1$ bit/channel-use.

However, in the multi-cell case, the capacity is not necessarily M bits/channel-use, because the receiver may not be able to distinguish all the 2^M transmitted symbols. For example, there may exist different vectors $\mathbf{x}_1, \mathbf{x}_2 \in \{\pm 1\}^M$ that satisfy

$$\mathrm{sign}(H\mathbf{x}_1) = \mathrm{sign}(H\mathbf{x}_2) \tag{15}$$

and therefore cannot be distinguished at the receiver for any SNR.

To show this, consider the signal set at the receiver defined as

$$\mathcal{A}(H) = \{\mathbf{y} | \mathbf{y} = \mathrm{sign}(H\mathbf{x}), \mathbf{x} \in \{\pm 1\}^M\}. \tag{16}$$

Clearly, we have

$$|\mathcal{A}(H)| \leq \min\{2^M, 2^N\}. \tag{17}$$

Therefore, the capacity is upper bounded by

$$C_R \leq \mathbb{E}_H[|\mathcal{A}(H)|] = C_T \leq \min(M, N), \tag{18}$$

where C_T is the capacity of the system when the transmitter and receiver know the channel, and C_R is the capacity when only receiver knows the channel. Equality is achieved with one transmitter/receiver pair. At high SNR,

$$\begin{aligned}\mathbf{I}(\mathbf{x}; \mathbf{y}, H) &= \underbrace{\mathbf{I}(\mathbf{x}; H)}_{=0} + \mathbf{I}(\mathbf{x}; \mathbf{y}|H) \\ &= \mathbf{H}(\mathbf{y}|H) - \underbrace{\mathbf{H}(\mathbf{y}|\mathbf{x}, H)}_{=0} = \mathbf{H}(\mathbf{y}|H),\end{aligned} \tag{19}$$

where $\mathbf{H}(\cdot)$ is the entropy.

We use the uniform distribution on $\mathbf{x}$ in (9) and (19) to obtain

$$\begin{aligned}C_R &= -\mathbb{E}_H \sum_{\mathbf{y}} \mathbb{E}_{\mathbf{x}}[\mathbb{1}(\mathrm{sign}(H\mathbf{x}) = \mathbf{y})] \log_2 \mathbb{E}_{\mathbf{x}}[\mathbb{1}(\mathrm{sign}(H\mathbf{x}) = \mathbf{y})] \\ &\overset{(a)}{=} -2^N \mathbb{E}_H\left[\mathbb{E}_{\mathbf{x}}\left[\mathbb{1}(\mathrm{sign}(H\mathbf{x}) = \underline{1}) \log_2 \mathbb{E}_{\mathbf{x}}[\mathbb{1}(\mathrm{sign}(H\mathbf{x}) = \underline{1})]\right]\right] \\ &\overset{(b)}{=} -2^N \mathbb{E}_H\left[\mathbb{1}(\mathrm{sign}(H\underline{1}) = \underline{1}) \log_2 \mathbb{E}_{\mathbf{x}}[\mathbb{1}(\mathrm{sign}(H\mathbf{x}) = \underline{1})]\right] \\ &= M - \mathbb{E}_H\left(\log_2 \sum_{\mathbf{x}} [\mathbb{1}(\mathrm{sign}(H\mathbf{x}) = \underline{1} | \mathrm{sign}(H\underline{1}) = \underline{1})]\right) \\ &\overset{(c)}{\geq} M - \log_2\left(\sum_{\mathbf{x}} \mathbb{E}_H[\mathbb{1}(\mathrm{sign}(H\mathbf{x}) = \underline{1} | \mathrm{sign}(H\underline{1}) = \underline{1})]\right)\end{aligned} \tag{20}$$

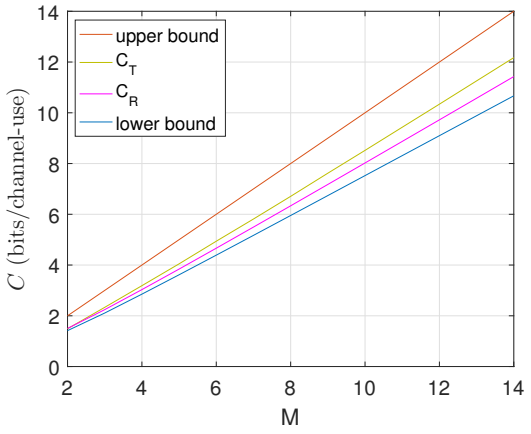

Fig. 5. Capacity C_T (18), C_R (20), and lower bound (21) for $M = N$, along with trivial upper bound M. The capacity increases linearly with M.

$$\overset{(d)}{=} M - \log_2\left(\sum_{k=0}^{M} \binom{M}{k} \left(1 - \frac{1}{\pi} \arccos\left(\frac{M-2k}{M}\right)\right)^N\right), \tag{21}$$

where $\underline{1}$ is a vector of all 1's, (a) and (b) are achieved by only considering $\mathbf{y} = \underline{1}$, and $\mathbf{x} = \underline{1}$ based on the symmetry of $\mathbf{x}$ and H, (c) uses Jensen's inequality, and (d) is derived according to a similar calculation made in the context of geometry [33]. From (20),

$$C_R < M \tag{22}$$

for $M \geq 2$.

For any non-uniform distribution on $\mathbf{x}$, we have

$$\mathbf{I}(\mathbf{x}; \mathbf{y}, H) \leq \mathbf{H}(\mathbf{x}) < M. \tag{23}$$

Therefore, when $M > 2$, we also have

$$C_T < M. \tag{24}$$

Fig. 5 shows the capacity in high SNR and its bounds, including C_T (18), C_R (20), the lower bound (21), and the upper bound M for $M = N$. It is readily seen from (21) that the capacity increases linearly with $M = N$.

C. Receiver does not know the channel

For finite SNR ρ, and H unknown to both the transmitter and the receiver, we consider a block fading model in which the channel is constant for some discrete time block T_b, and changes every interval T_b independently. Within one block of T_b symbols, (5) yields

$$Y_b = \mathrm{sign}(\sqrt{\frac{\rho}{M}} H X_b + V_b), \tag{25}$$

where $X_b \in \{\pm 1\}^{M \times T_b}$ is the transmitted signal matrix, $Y_b \in \{\pm 1\}^{N \times T_b}$ is the received signal matrix, V_b is the $N \times T_b$ matrix of additive Gaussian noise independent

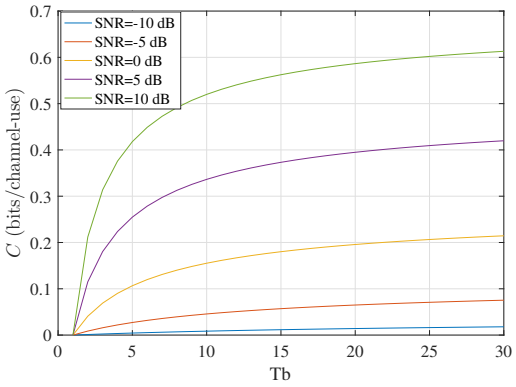

Fig. 6. Capacity C (28) versus block length T_b with a single transmitter/receiver pair for various SNR.

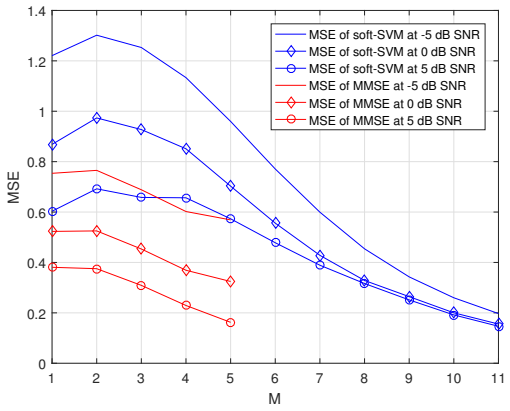

Fig. 7. Mean-square error (MSE) of soft-SVM (Algorithm 1) and MMSE estimator (29). All 2^M possible signals are transmitted for training. Because of computational difficulties, we only show $M \leq 5$ for the MMSE estimator.

of H and X_b, and sign$(\cdot)$ provides the sign of the input as the output.

The capacity in bits per channel-use becomes

$$C = \frac{1}{T_b} \max_{p_{X_b}(\cdot), X_b \in \{\pm 1\}^{M \times T_b}} \mathbf{I}(X_b; Y_b), \qquad (26)$$

where the transition probability between X_b and Y_b is unknown, but its distribution can be derived through the distribution of H.

Since H is random, when $T_b = 1$, then $Y_b \in \{\pm 1\}^N$ is uniformly-distributed, independently of X_b. Therefore $C = 0$.

For $T_b > 1$, we consider a single transmitter/receiver in which the model (25) can be simplified as

$$\mathbf{y}_b = \text{sign}(\sqrt{\rho} h \mathbf{x}_b + \mathbf{v}_b), \qquad (27)$$

where $\mathbf{x}_b, \mathbf{y}_b \in \{\pm 1\}^{T_b}$, $h \sim \mathcal{N}(0,1)$, $\mathbf{v}_b \sim \mathcal{N}(0, I)$, h and $\mathbf{v}_b$ are independent. The kth element of $\mathbf{x}_b$ and $\mathbf{y}_b$ are denoted as $x_b(k)$ and $y_b(k)$, respectively.

This model is equivalent to a BSC with crossover probability p_b shown in (8), and where p_b is changed independently every time block T_b when h changes. We have

$$\begin{aligned} C &= \frac{1}{T_b}(\mathbf{H}(\mathbf{y}_b) - \mathbf{H}(\mathbf{y}_b|\mathbf{x}_b)) \\ &= 1 - \frac{1}{T_b}\mathbb{E}_{\mathbf{x}_b}\left(\mathbb{E}_{\mathbf{y}_b|\mathbf{x}_b}\left(-\log_2 \mathbb{P}(\mathbf{y}_b|\mathbf{x}_b)\right)\right) \\ &= 1 + \frac{1}{T_b}\mathbb{E}_{\mathbf{y}_b|\mathbf{x}_b=\underline{1}}\left(\log_2 \mathbb{P}(\mathbf{y}_b|\mathbf{x}_b = \underline{1})\right) \\ &= 1 + \frac{1}{T_b}\sum_{k=0}^{T_b}\binom{T_b}{k}\Big(\mathbb{E}_h\left[p_b^k(1-p_b)^{T_b-k}\right] \\ &\qquad \log_2\left(\mathbb{E}_h\left[p_b^k(1-p_b)^{T_b-k}\right]\right)\Big). \qquad (28) \end{aligned}$$

Fig.6 shows the results for a variety of block lengths T_b and SNR's.

For multiple transceiver chains, the optimal input distribution is more complicated, but a lower bound of the capacity can be obtained by using uniform input. We omit this analysis here.

V. Machine Learning and Channel Estimation

In this section, we draw some analogies between the proposed "transceiver cell" and the classical "computational cell" used in modeling neural networks. As a result, simple machine learning algorithms can brought to bear on aspects of communication, such as channel estimation.

As an example, we consider $N = 1$ and an arbitrary M. Assume T_t training symbols $\mathbf{x}_t(1), \ldots, \mathbf{x}_t(T_t)$ known to the receiver are transmitted for learning the channel and the corresponding received symbols at the receiver are $y_t(1), \ldots, y_t(T_t)$. The transmitted training signals are collected as a matrix $X_t \in \{\pm 1\}^{M \times T_t}$ and the received signals are collected as a vector $\mathbf{y}_t \in \{\pm 1\}^{1 \times T_t}$.

The classical MMSE estimator, defined as the conditional mean $\mathbf{h}$:

$$\hat{\mathbf{h}}_{MMSE} = \mathbb{E}[\mathbf{h}|X_t, \mathbf{y}_t] \qquad (29)$$

minimizes mean-square error (MSE) $e_{MMSE} = \mathbb{E}_{\mathbf{h}}\|\hat{\mathbf{h}}_{MMSE} - \mathbf{h}\|^2/M$. Unfortunately, it does not have a closed-form solution and the complexity increases dramatically as M increases.

However, the estimation problem can be cast into a question often solved in the context of neural networks: what are the "weights" whose linear combination feed a threshold function? A neural network is made up of neural cells connected with each other, and each neural cell is a computation unit with output as a nonlinear operation of the input. The most common nonlinear

operation is taking the sign of the linear combination of the inputs, modeled as

$$y = \text{sign}(\mathbf{w}^T\mathbf{x}), \tag{30}$$

where y is the output, $\mathbf{x}$ is the input vector, and $\mathbf{w}$ are the "weights" of the cell. These weights play the role of the unknown channel in our communication system.

The support vector machine (SVM) algorithm [34] is widely used in classification problems to estimate $\mathbf{w}$ that solves $y_t(k) = \text{sign}(\mathbf{w}^T\mathbf{x}_t(k))$ for all $k = 1, \ldots, T_t$. In particular, a version called soft-SVM handles the classification problem in the presence of noise. The estimate, denoted $\hat{\mathbf{h}}_{SVM}$, is displayed as Algorithm 1. The MSE of the channel estimation using the soft-SVM algorithm is $e_{SVM} = \mathbb{E}_{\mathbf{h}}\|\hat{\mathbf{h}}_{SVM} - \mathbf{h}\|^2/M$. The soft-SVM algorithm is simple to implement, even for a large number of transceiver cells.

Algorithm 1 Soft-SVM Algorithm

Input: A training set $(\mathbf{x}_t(k), y_t(k))(1 \le k \le T_t), N_t$
  $\#\mathbf{x}_t(k)$ the kth transmitted training vector
  $\#y_t(k)$ the received signal from the kth training
  $\#T_t$ the number of trainings
  $\#N_t$ maximum iteration number
  $\theta^{(1)} = 0 \cdot \mathbf{x}_t(1)$
  $\hat{\mathbf{h}} = \theta^{(1)}$
  for $t = 1, \cdots, N_t$ **do**
    $\hat{\mathbf{h}} = \hat{\mathbf{h}} + \frac{1}{t}\theta^{(t)}$;
    if $(y_t(k)(\hat{\mathbf{h}})^T\mathbf{x}_t(k) > 0)$ for all $1 \le k \le T_t$, **then**
      break;
    else
      Choose m uniformly random from set $\mathcal{A}$,
      where $\mathcal{A} = \{k : y_t(k)(\hat{\mathbf{h}})^T\mathbf{x}_t(k) \le 0\}$;
      $\theta^{(t+1)} = \theta^{(t)} + y_t(m)\mathbf{x}_t(m)$;
    end if
  end for
  $\hat{\mathbf{h}}_{SVM} = \sqrt{M}\hat{\mathbf{h}}/\|\hat{\mathbf{h}}\|$;
Output: $\hat{\mathbf{h}}_{SVM}$

We show the performance of soft-SVM algorithm and MMSE estimator in Fig. 7 as a function of M for various SNR's. The soft-SVM algorithm compares favorably to the MMSE through a wide range of SNR's and number of transmitter cells. In this figure, all 2^M possible distinct training signals are used, and we set $N_t = 300$. The difficulty of computing the MMSE estimator for M greater than 5 makes a complete comparison impossible for all M. We can see that e_{SVM} decreases quickly as the number of antennas M increases.

VI. Beamforming

We provide a framework for beamforming and show that, under certain circumstances, the sign of the channel is used to form the optimal beamforming vector at transmitter. The optimum beamformer at the receiver is formed from knowledge of the bit error probabilities.

A. Beamforming at the transmitter

We use $N = 1$ as an example to show beamforming at the transmitter. The classical beamforming solution is compared with the 1-bit transceiver version.

1) Classical high-resolution system: In a classical high-resolution linear system, beamforming is

$$\mathbf{x} = \mathbf{w}s, \tag{31}$$

where $\mathbf{w}$ is the transmitted beamforming vector with $\|\mathbf{w}\| = \sqrt{M}$ and s is the transmitted symbol. The result is

$$\mathbf{w}_{T,\ell} = \underset{\|\mathbf{u}\|=\sqrt{M}}{\text{argmax}} |\mathbf{h}^T\mathbf{u}|, \tag{32}$$

where one solution is

$$\mathbf{w}_{T,\ell} = \frac{\sqrt{M}\mathbf{h}}{\|\mathbf{h}\|}. \tag{33}$$

2) Single-bit transceiver: In contrast to the high-resolution system, the received signal can only be $y \in \{\pm 1\}$, and therefore the transmitted symbol is effectively $s \in \{\pm 1\}$. We can still use (31) as the beamforming framework with $s \in \{\pm 1\}$, and with $\mathbf{w} \in \{\pm 1\}^M$. The best beamforming vector $\mathbf{w}$ will maximize the SNR at the receiver, whence

$$\mathbf{w}_{T,q} = \underset{\mathbf{u}\in\{\pm 1\}^M}{\text{argmax}} |\mathbf{h}^T\mathbf{u}|, \tag{34}$$

where one solution is

$$\mathbf{w}_{T,q} = \text{sign}(\mathbf{h}), \tag{35}$$

and the corresponding transmitted signal is

$$\mathbf{x} = \mathbf{w}_{T,q}s \tag{36}$$

with $s \in \{\pm 1\}$.

This conclusion does not readily generalize to multiple transceivers because the structure shown in (36) can only transmit two distinct symbols (corresponding to one bit). A generalized beamforming framework is required when it is desired to transmit more than one bit using a multiplicity of transceivers.

B. Beamforming at the receiver

We use a single transmitter to exemplify beamforming at the receiver. The model is

$$\mathbf{y} = \sqrt{\rho}\mathbf{h}x + \mathbf{v}, \tag{37}$$

where $\mathbf{y}$ is the received signal, the channel is a $M \times 1$ vector $\mathbf{h}$, and $\mathbf{v}$ is the additive Gaussian noise.

1) Classical high-resolution system: In this case, maximal-ratio combining (MRC) is applied at the receiver, and

$$\mathbf{w}_{R,l} = \frac{\mathbf{h}}{\sqrt{\rho}\mathbf{h}^T\mathbf{h}} \tag{38}$$

$$\hat{x} = \underset{x\in\mathcal{X}}{\text{argmin}} |\mathbf{w}_{R,l}^T\mathbf{y} - x|^2 \tag{39}$$

where $\mathcal{X}$ is the alphabet of the transmitted symbols, $\mathbf{w}_{R,l}$ is the linear combining beamforming vector, $\hat{x}$ is the estimate of x.

2) Single-bit transceiver: In contrast, the maximum-likelihood detector is

$$\hat{x} = \underset{x\in\{\pm 1\}}{\operatorname{argmax}} \sum_{k=1}^{N} \log \mathbb{P}(y_k|x), \tag{40}$$

where y_k is the kth element of $\mathbf{y}$. We obtain

$$\begin{aligned}\hat{x} &= \operatorname{sign}\left(\sum_{k=1}^{N} \log \frac{\mathbb{P}(y_k|x=1)}{\mathbb{P}(y_k|x=-1)}\right) \\ &= \operatorname{sign}(\sum_{k=1}^{N} y_k \log \frac{q_k}{p_k}) \\ &\triangleq \operatorname{sign}(\mathbf{w}_{R,q}^T \mathbf{y}),\end{aligned} \tag{41}$$

where $p_k = Q(\sqrt{\rho} h_k)$ is the bit-error probability of the kth receiver, $q_k = 1 - p_k$, and $\mathbf{w}_{R,q}$ is the beamforming vector defined as

$$\mathbf{w}_{R,q} = [\log \frac{q_1}{p_1}, \log \frac{q_2}{p_2}, \cdots, \log \frac{q_N}{p_N}]^T. \tag{42}$$

Equation (42) shows the beamforming weights at the receiver as a function of the bit-error probabilities.

For multiple transceivers where more than 2 transmitted symbols are possible,

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathcal{X}}{\operatorname{argmax}} \sum_{k=1}^{N} \log \mathbb{P}(y_k|\mathbf{x}), \tag{43}$$

where $\mathcal{X}$ is the set of possible transmitted symbols. Unfortunately, we cannot readily get a simple beamforming expression such as (41).

VII. Energy Efficiency

As shown in (4), increasing the consumed power at the transceiver increases the SNR ρ in (5). This increases the channel capacity. In this section, we introduce energy efficiency as a performance metric of the system, which can be used to find the optimum operational power of the transceivers in a system setting. The basic premise is that increasing the consumed power increases the capacity, but consuming too much power per bit results in low efficiency.

We express the capacity of the system in bits/(second-Hz), bits/symbol, or bits/(channel-use) equally, where each channel-use corresponds to a new symbol. Then, the energy efficiency with unit bits/(Hz-Joule) is defined as

$$E = \frac{C}{MW_{\text{tx}} + NW_{\text{rx}}}, \tag{44}$$

where C is the capacity in bits/(channel-use) and W_{tx} and W_{rx} are the transmitter and receiver consumed powers per cell.

The corresponding achievable rate in bits/second becomes

$$R = E \cdot B \cdot (MW_{\text{tx}} + NW_{\text{rx}}), \tag{45}$$

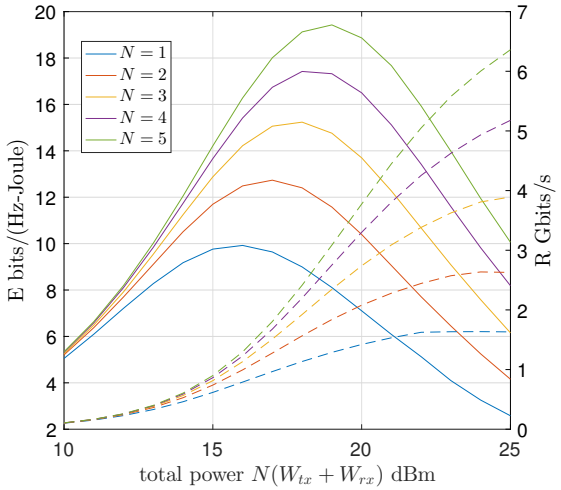

Fig. 8. Energy efficiency E (solid lines, left y-axis, equation (44)) versus total power consumed by all transmitters and receivers, and corresponding rate R in Gbits/second (right y-axis, color-matched dashed lines, equation (45)) for $M = N = 1, \cdots, 5$.

where E is the energy efficiency, and B is the system bandwidth.

For maximum energy efficiency, we wish to operate the system in the SNR region where a significant fraction of the maximum throughput has been attained. The shapes of the curves in Fig. 4 start saturating at 10-15 dB, suggesting that this SNR represents the upper end of our target range.

We use (4), (13), (44), and (45), together to obtain Fig. 8, which displays achievable rate in bits/(Hz-Joule) (solid line, left y-axis) or bits/(second) (dash line, right y-axis) versus total power consumed by all transceivers $N(W_{\text{tx}} + W_{\text{rx}})$ for $M = N = 1, \ldots, 5$. For each value of $W_{\text{rx}} + W_{\text{tx}}$, the optimum pair $(W_{\text{tx}}, W_{\text{rx}})$ is chosen to maximize the bits/(Hz-Joule). The simulation results suggest that the preferred operating point of total consumed power is $N(W_{\text{tx}}+W_{\text{rx}}) \in [15, 20]$ dBm, since the curves show highest energy efficiency.

For example, if we have $M = N = 5$, the operating total power will be about $W = 19$ dBm (less than 100 mW) to get the maximum energy efficiency $E = 19.4$ bits/(Hz-Joule), with corresponding rate $R = 3.1$ Gbits/s. This power consumption meets the requirements presented in the Introduction for a portable handheld wireless device.

VIII. Conclusion

The single-bit transceivers we have modeled herein are simple enough to allow an end-to-end system analysis of per-bit energy efficiency as a function of total power consumption. We believe that the model is generic enough that it applies to similar single-bit transceivers

that can be designed for a variety of carrier frequencies and bandwidths. By drawing connections between the transceiver cells and computational cells used in neural network models, we have also shown how machine learning algorithms can be used to simplify algorithmic aspects of the communication system.

We think it is an interesting challenge to drive the power consumption (and cost) of a transceiver cell as low as possible, and make up for their limited individual performance and throughput by having large quantities of them operating simultaneously. We believe that this is an advantageous trade-off that is yet to be explored.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant Numbers CCF-1403458, ECCS-1509188, and IIP-1439682, and the Laboratory for Telecommunications Sciences.

REFERENCES

[1] FCC, *Notice of Proposed Rulemaking: FCC 15-138*, October 2015. [Online]. Available: https://apps.fcc.gov/edocs_public/attachmatch/FCC-15-138A1.pdf

[2] *Texas Instruments ADC12J4000*. [Online]. Available: http://www.ti.com/product/adc12j4000/datasheet

[3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Comm. Mag.*, pp. 186–195, Feb. 2014.

[4] H. Kwon and T. Birdsall, "Channel capacity in bits per Joule," *IEEE Journal of Oceanic Engineering*, vol. 11, no. 1, pp. 97–99, 1986.

[5] H.-S. Lee and C. G. Sodini, "Analog-to-digital converters: Digitizing the analog world," *Proceedings of the IEEE*, vol. 96, no. 2, pp. 323–334, 2008.

[6] W. Singh and R. Sharma, "High resolution, high speed and low power comparator for high speed ADCs," in *Proc. Int. Conf. Soft Comp. Tech. and Imp.*, Oct 2015, pp. 113–116.

[7] J. Singh, O. Dabeer, and U. Madhow, "Communication limits with low precision analog-to-digital conversion at the receiver." in *Proc. ICC*, 2007, pp. 6269–6274.

[8] ——, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Transactions on Communications*, vol. 57, no. 12, pp. 3629–3639, 2009.

[9] A. Mezghani and J. A. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," in *2007 IEEE Int. Symp. on Information Theory*, 2007, pp. 1286–1289.

[10] ——, "Analysis of Rayleigh-fading channels with 1-bit quantized output," in *2008 IEEE Int. Symp. Information Theory*, 2008, pp. 260–264.

[11] ——, "Analysis of 1-bit output noncoherent fading channels in the low SNR regime," in *2009 IEEE Int. Symp. Information Theory*, 2009, pp. 1080–1084.

[12] B. M. Murray, H. Suzuki, and S. Reisenfeld, "Optimal quantization of OFDM at digital IF," in *TENCON 2008-2008 IEEE Region 10 Conference*, 2008, pp. 1–5.

[13] X. Shao and C. H. Slump, "Quantization effects in OFDM systems," in *Proc, 29th Symp. on Info. Theory*, 2008, pp. 93–103.

[14] C. Studer and G. Durisi, "Quantized massive MU-MIMO-OFDM uplink," *IEEE Transactions on Communications*, vol. 64, no. 6, pp. 2387–2399, June 2016.

[15] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath Jr, "Uplink performance of wideband massive MIMO with one-bit ADCs," *arXiv preprint arXiv:1602.07364*, 2016.

[16] S. Krone and G. Fettweis, "Capacity of communications channels with 1-bit quantization and oversampling at the receiver," in *2012 35th IEEE Sarnoff Symposium*, 2012, pp. 1–7.

[17] J. Choi, D. J. Love, D. R. Brown, and M. Boutin, "Quantized distributed reception for MIMO wireless systems using spatial multiplexing," *IEEE Transactions On Signal Processing*, vol. 63, no. 13, pp. 3537–3548, 2015.

[18] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 2005–2018, 2016.

[19] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5498–5512, 2015.

[20] ——, "High SNR capacity of millimeter wave MIMO systems with one-bit quantization," in *2014 Info. Theory and Applications Workshop*, 2014, pp. 1–5.

[21] F. Sun, J. Singh, and U. Madhow, "Automatic gain control for ADC-limited communication," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, 2010, pp. 1–5.

[22] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, 2014.

[23] A. Kant Saxena, I. Fijalkow, and A. L. Swindlehurst, "Analysis of One-Bit Quantized Precoding for the Multiuser Massive MIMO Downlink," *ArXiv e-prints*, Oct. 2016.

[24] Y. Li, C. Tao, A. L. Swindlehurst, A. Mezghani, and L. Liu, "Downlink achievable rate analysis in massive MIMO systems with one-bit DACs," *arXiv preprint arXiv:1610.09630*, 2016.

[25] Y. Li, C. Tao, L. Liu, A. Mezghani, and A. L. Swindlehurst, "How much training is needed in one-bit massive MIMO systems at low SNR?" *arXiv preprint arXiv:1608.05468*, 2016.

[26] Y. Li, C. Tao, L. Liu, G. Seco-Granados, and A. L. Swindlehurst, "Channel estimation and uplink achievable rates in one-bit massive MIMO systems," in *2016 IEEE Sensor Array Multi. Sig. Proc. Work.*, 2016, pp. 1–5.

[27] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "One-bit massive MIMO: Channel estimation and high-order modulations," in *2015 IEEE Int. Communication Workshop*, 2015, pp. 1304–1309.

[28] O. B. Usman, H. Jedda, A. Mezghani, and J. A. Nossek, "MMSE precoder for massive MIMO using 1-bit quantization," in *2016 IEEE Int. Conf on Acoust, Speech and Signal Proc.*, 2016, pp. 3381–3385.

[29] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *arXiv preprint arXiv:1404.7736*, 2014.

[30] A. Sarkar and B. A. Floyd, "A 28-GHz harmonic-tuned power amplifier in 130-nm SiGe BiCMOS," *IEEE Transactions on Microwave Theory and Techniques*, vol. PP, no. 99, pp. 1–14, 2017.

[31] B. Henderson and E. Camargo, *Microwave mixer technology and applications*. Artech House, 2013.

[32] A. I. Sulyman, A. T. Nassar, M. K. Samimi, G. R. Maccartney, T. S. Rappaport, and A. Alsanie, "Radio propagation path loss models for 5G cellular networks in the 28 GHz and 38 GHz millimeter-wave bands," *IEEE Communications Magazine*, vol. 52, no. 9, pp. 78–86, 2014.

[33] N. Calkin, K. James, L. Bowman, R. Myers, E. Riedl, and V. Thomas, "Cuts of the hypercube," *Dept. of Mathematical Sciences, Clemson University (July 3, 2008)(manuscript)*, 2008.

[34] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.