Excited Markov Chain Monte Carlo MIMO Detector with 8-Antenna 802.11ac Testbed Demonstration

Jonathan C. Hedstrom, Chung Him (George) Yuen, Rong-Rong Chen, Behrouz Farhang-Boroujeny
ECE department, University of Utah, USA

***Abstract*—The development of low complexity, high performance spatial-multiplexing MIMO detectors continues to be an important area of research capable of increasing the spectral efficiency and capacity of wireless networks. The Markov Chain Monte Carlo (MCMC) detector has shown promise as a high performance method with low complexity growth. We present a solution to the high SNR stalling problems of previous MCMC detectors. Near-MAP performance is verified in simulation and in real-world measurements on an 8-antenna MIMO testbed using the 802.11ac WiFi protocol. This demonstration shows that the channel models predominantly used in the MCMC literature are too well-conditioned to provide an understanding of performance and complexity for indoor channels. Additional information is provided on the methods and techniques to match simulation to measurement and to construct a low cost and effective 8-antenna MIMO testbed.**

I. Introduction

The use of spatial-multiplexing multiple-input multiple-output (MIMO) is increasingly being adopted in wireless protocols as higher spectral efficiency is needed to meet the capacity requirements of modern wireless networks [1], [2]. It has the potential to linearly increase spectral reuse and capacity as the number of streams increases. Up to 8-stream MIMO is defined in 802.11ac WiFi and LTE Advanced Release 10 protocols. With these large sizes, the performance and complexity scaling of the MIMO detector is extremely important.

The sphere-decoding (SD) class of MIMO detectors are known to have near maximum-a-posteriori (MAP) performance [3]. Specifically the K-Best variations of sphere-decoding have been demonstrated in effective VLSI designs, but their complexity increases quickly with the number of antennas, the number of transmitted bits per channel use, and the list size [4]. Therefore, the search for lower complexity MIMO detectors is ongoing.

Markov Chain Monte Carlo (MCMC) has been shown to have near optimal performance at low signal-to-noise-ratio (SNR) and to have efficient hardware implementations [5], [6]. It has recently been demonstrated on a 4-antenna MIMO testbed limited to 4 QAM constellations operating at low SNR [7]. Counterintuitively, MCMC performance decreases with SNR due to a characteristic stalling behavior [8].

We have solved the high SNR stalling problem in our excited MCMC (X-MCMC) detector. Its implementation is similar to static, fixed value temperature scaling methods [9], but is able to dynamically calculate the scaling as needed. This enhancement does not require hybridization with another detector, such as initialization with MMSE or SD solutions [8], [10].

The X-MCMC detector is shown to have near-MAP performance at 4- and 8-antenna 64 QAM sizes. Simulations are matched with real-world measurements using 802.11ac WiFi packets. We show that most MCMC results in the literature use an insufficiently difficult, uncorrelated channel model to draw conclusions on their potential real-world indoor performance and complexity. Details on constructing a low cost and effective 8-antenna testbed are also provided.

This paper is structured as follows. There is a brief statement of the MIMO system model in Section II. The MAP and K-Best MIMO detectors used for comparison purposes are explained in Section III. A basic MCMC detector is outlined in Section IV and then expanded upon with the proposed X-MCMC enhancements in V. Sections VI and VII describe the models and metrics needed in simulating and measuring the channel and interference characteristics. Finally, Section VIII covers the hardware and software details of the testbed used in collecting the measurements matched to simulation results in Section IX. Concluding remarks are made in Section X.

II. System Model

The goal of a spatial-multiplexing MIMO transceiver is to exploit the multipath of the environment to support overlapping data streams. This concept allows reuse of the spectrum and therefore has the potential to dramatically increase the data rates and capacity of wireless networks.

To perform the spatial-multiplexing, the transmitter sends independent data streams simultaneously on the N_t transmit antennas and is received at the N_r receiver antennas. For the 802.11ac WiFi protocol used in this paper, orthogonal frequency division multiplexing (OFDM) is specified with 4, 16, 64, or 256 quadrature amplitude modulation (QAM) on each of the active frequency-domain subcarriers. This leads to the frequency-domain system model

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}. \tag{1}$$

Here, $\mathbf{y}$ is the received signal vector, $\mathbf{H}$ is a slow flat fading complex channel matrix containing the pairwise gain and phase between antennas, $\mathbf{s}$ is the vector of transmitted QAM constellation symbols, and $\mathbf{n}$ is a noise vector. The noise elements are assumed to be independent and identically distributed (i.i.d.) complex Gaussian random variables with variance of σ_n^2 per each real and imaginary dimension. Assuming the transmit and receive side have the same number of antennas N, the dimensions of the vectors and matrices are $N \times 1$ and $N \times N$.

The work by J. C. Hedstrom is supported in part by the National Science Foundation under grants 1449033 and 1632569 and in collaboration with Farhang Wireless Inc. We would also like to thank the Linux, Python, Numpy, CFFI, Matplotlib, and Seaborn open-source teams for the excellent software used in this work.

978-1-4673-8999-0/17/$31.00 ©2017 IEEE

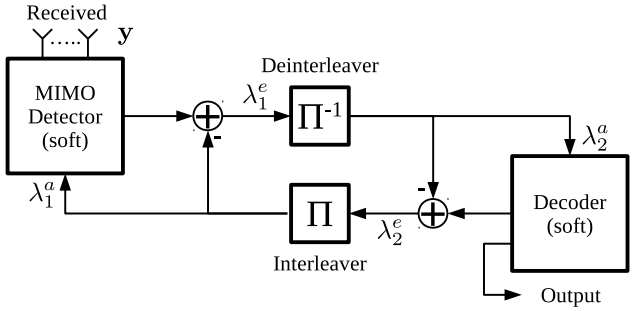


Fig. 1. Turbo loop structure with a MAP, MCMC, X-MCMC, or K-Best MIMO detector iteratively exchanging soft information with a forward error correction decoder to perform joint detection.

Note that depending on context, the bit vector $\mathbf{x}$, which is the unmapped version of the symbols $\mathbf{s}$, may be described as comprising of 1's and 0's or equivalently +1's and −1's.

The primary challenge in a MIMO transceiver is to accurately estimate the $K = N\log_2(N_{\text{QAM}})$ bits in $\mathbf{x}$, where N_{QAM} is the QAM constellation size.

Notation

In the equations that follow, some specialized notation is used for compactness and clarity. Vectors and matrices are expressed with bold fonts and the latter are capitalized. The removal of the k^{th} element of a vector is shown with set notation as $\{\cdot\}^{\setminus k}$. A variable or vector in which the k^{th} transmitted bit has been forced to a one or zero is shown with $\{\cdot\}^{k+}$ and $\{\cdot\}^{k-}$, respectively. When two nearly identical equations are needed differing only in use of $k+$ or $k-$, $k\pm$ is used to represent both versions.

III. MIMO Detector Introduction

The MIMO detector uses the received signal $\mathbf{y}$ and estimated channel $\mathbf{H}$ to calculate an estimate of the transmitted bit sequence $\mathbf{x}$. There are many different MIMO detector approaches explored in the literature, but here we only touch on MAP, K-Best, MCMC, and X-MCMC. MAP is needed in our analysis as it provides a theoretical bound against which the other methods can be compared. K-Best is a useful reference as it has been shown to have near-MAP performance while being realizable. Yet, there is still an interest in lower complexity near-MAP MIMO detectors as K-Best complexity becomes large as the number of antennas and QAM size increases.

The above mentioned MIMO detectors are all compatible with turbo iterations, the exchange of soft-information between the detector and decoder as in Fig. 1. This allows for the iterative joint detection of the signal for enhanced performance. Here and in the equations that follow, the soft-information is in the form of log likelihood ratios (LLRs) and represented by λ^a and λ^e for the *a priori* and extrinsic probability versions.

The maximum-a-posteriori (MAP) detector is the theoretically optimal soft-output MIMO detector utilizing prior-information input. It tests every 2^K permutation of the sequence $\mathbf{x}$ against the received signal and calculate the probability of each bit being a one or zero. These probabilities are typically calculated as log likelihood ratios (LLR) to improve stability and lower complexity. We use MAP with the max-log approximation as in [3] to provide a fair comparison with MCMC, X-MCMC, and K-Best which also use this approximation. Unfortunately, the complexity of such a detector increases exponentially with the number of transmitted bits, making it unrealizable except for the most trivial of cases. For comparison, we have implemented max-log MAP (Max-MAP) on a graphics processing unit (GPU), allowing up to 4-antenna 64 QAM Max-MAP results to be presented.

For comparison purposes, we have implemented the K-Best variation of the sphere-decoder described in [4]. This was chosen because efficient VLSI designs have been demonstrated that will allow for future performance and complexity comparisons.

IV. MCMC Detector

The Markov Chain Monte Carlo (MCMC) detector estimates the output LLR statistics by means of Monte Carlo sampling [5]. This can be thought of as a method to identify a list of important bit permutation samples to approximate the full list of permutations used in the MAP detector. The challenge is to make the process computationally efficient by using an easy to calculate short list that accurately captures the statistics of the full permutation list. We select the bitwise MCMC algorithm described in [6] as our foundational method because it is shown to be efficiently implementable in hardware. The two main components of the MCMC detector is the Gibbs sampler, which identifies the short permutation list, and the LLR output calculation, which uses the sample list.

The Gibbs sampler starts with an initial estimate of the transmitted bit sequence, either randomly selected or initialized with prior information. It then cycles through the bits, making a weighted random decision to change the k^{th} bit where each cycle through the bits is an iteration. The weighting is based on the probability P_{gibbs} of the bit being a +1 as in

$$P_{\text{gibbs}} = P(x_k = +1 \,|\, \mathbf{y}, \mathbf{x}^{\setminus k}, \boldsymbol{\lambda}^a) = \frac{1}{1+e^{-\gamma_k}} \tag{2}$$

with

$$\gamma_k = \frac{1}{2\sigma_n^2}\left(\left\|\mathbf{y}-\mathbf{H}\mathbf{s}^{k-}\right\|^2 - \left\|\mathbf{y}-\mathbf{H}\mathbf{s}^{k+}\right\|^2\right) + \lambda_k^a. \tag{3}$$

The chain of bit flips can be thought of as a probabilistically guided random walk where each calculation of $\left\|\mathbf{y}-\mathbf{H}\mathbf{s}^{k\pm}\right\|^2$ can be used as a sample.

After a sufficient quantity of important samples have been taken in this manner, the list Z of all sampled bit permutations can be used to calculate the output LLR

$$\begin{aligned}\lambda_k^e \approx &\frac{1}{2}\min_{\mathbf{x}\in Z^{k-}}\left(\frac{1}{\sigma_n^2}\left\|\mathbf{y}-\mathbf{H}\mathbf{s}^{k-}\right\|^2 - \mathbf{x}^{\setminus k}\cdot\boldsymbol{\lambda}^{a,\setminus k}\right)\\ &-\frac{1}{2}\min_{\mathbf{x}\in Z^{k+}}\left(\frac{1}{\sigma_n^2}\left\|\mathbf{y}-\mathbf{H}\mathbf{s}^{k+}\right\|^2 - \mathbf{x}^{\setminus k}\cdot\boldsymbol{\lambda}^{a,\setminus k}\right).\end{aligned} \tag{4}$$

One of the convenient features of this algorithm is that a designer can trade between performance, implementation complexity, and latency by adjusting N_{gibbs}, the number of parallel Gibbs samplers, and N_{iter}, the number of cycles performed across the bit sequence.

The primary issue with the MCMC detector is that its performance is known to degrade as SNR increases [5]. This is due to a high SNR stalling problem which we will attempt to solve in the following section.

In this paper, we have chosen to use the MMSE hybrid version from [8] for comparison purposes. This method suggests initializing one of the parallel Gibbs samplers with the hard-decision from an MMSE detector solution. The initialization is meant to help mitigate the high SNR stalling problem, but, as will be seen in the results of Section IX, the MMSE initialized method fails on real-world correlated channels.

V. X-MCMC Detector

Here we propose a dynamic method to solve the high SNR stalling problem of MCMC we refer to as excited MCMC (X-MCMC). It is based on the observation that the root cause of the stalling problem is that γ_k in (3) reaches extreme values when sampling far from a correct solution region. This results in the probability values in (2) saturating, creating a nearly deterministic walk that quickly stalls.

When the γ_k values have a small magnitude, P_{gibbs} probabilities are moderate and the desired guided random walk behavior is seen. This occurs when the Euclidean squared distance $\|\mathbf{y}-\mathbf{Hs}\|^2$ is small, and thus the Gibbs sampler is likely close to a correct solution. This distance metric is defined for convenience as

$$d^{k\pm} = \left\|\mathbf{y}-\mathbf{H}\mathbf{s}^{k\pm}\right\|^2. \tag{5}$$

Based on these observations, it is apparent that a scaling variable is needed. It should have no effect near the correct solution, and it must keep the sampler active with moderated values of γ_k when far away from the correct solution region. This can be accomplished by applying a linear scaling factor to (3). Thus the P_{gibbs} calculation should use the modified value

$$\gamma_k' = \frac{1}{r_\gamma}\frac{(d^{k^-}-d^{k^+})}{2\sigma_n^2} + \lambda_k^a \tag{6}$$

where r_γ is a time varying, confidence adjusting factor. This factor should be unity when close the the solution region, and should become larger as the distance from the correct solution becomes greater. This behavior can be accomplished by taking the ratio between the distance in the current state space region and the expectation of correct solution distances, $2N\sigma_n^2$. This suggests that r_γ should take the form

$$r_\gamma = \frac{\min_{d\in \mathrm{D}_\gamma}(d)}{2N\sigma_n^2} \tag{7}$$

where the minimum is taken over the list D_γ, the $d^{k\pm}$ distances observed since the last change in the Gibbs state $\mathbf{x}$. By using the recent history of distances, the scaling factor takes into account the error in the current region instead of just the most recently calculated k^{th} distance. This heuristic approach has the desired scaling behavior and has proven to work well in practice.

After applying the dynamic excitation coefficient r_γ to solve the high SNR stalling problem, one finds that the Gibbs sampler will converge quickly to the region of a potentially valid solution. Once this happens, r_γ will be approximately one and the Gibbs sampler will stop exploring new solution regions, repeatedly sampling the same bit permutations and collect no new samples. This is referred to as pseudo-convergence and appears to be less understood in MIMO communications applications, but has been noted in the wider MCMC literature [11]. This is problematic because it is insufficient to simply find the transmitted bit sequence. Solution diversity is needed to calculate quality output LLR information.

Although symptomatically similar, pseudo-convergence is different from the stalling problem discussed previously. It occurs when the posterior probability distribution is multi-modal with weak connections between modes, i.e. important solution regions are weakly connected to other important regions through very low transitional probabilities. Thus, the Gibbs sampler may stay in one mode, i.e. similar bit permutations connected with likely transition probability, and collect a large number of highly correlated samples. When encountered, pseudo-convergence decreases the sampling efficiency of MCMC, resulting in a need for a large number of parallel Gibbs samplers.

We propose a simple detection and excitation strategy to allow a Gibbs sampler to exit a pseudo-convergence mode and visit other important regions of the solution space. When the Gibbs sampler has not changed its state $\mathbf{x}$ in K steps, then pseudo-convergence is detected. After detection, the next bit is forced to change. We have found that this 1-bit forced-flip strategy is effective in practice and has low implementation complexity since a re-initialization is not necessary, unlike a full Gibbs sampler restart.

Since X-MCMC has a variable convergence time, it is still possible that for a small percentage of cases it does not converge to a stationary posterior distribution in the predetermined fixed number of iterations. With a large enough number of samples, the probability of poor convergence is vanishingly small, but since it is desirable to minimize the complexity of the design and therefore the number of iterations, it should be considered likely in practice. This can be detected and mitigated by similar metrics and methods as in (7) and (6), by expanding it to use the best distance observed across all Gibbs samplers. Therefore we specify the metric

$$r_\lambda = \frac{\min_{d\in \mathrm{D}_\lambda}(d)}{2N\sigma_n^2} \tag{8}$$

where the minimum is taken over the list D_λ, of all $d^{k\pm}$ distances calculated in all parallel Gibbs samplers.

This distance ratio has a desirable characteristic that it becomes approximately unity when the MCMC detector has sampled regions near the transmitted bit sequence. If the ratio r_λ is much larger than one, it can be directly used to scale down the overly confident extrinsic LLR values that otherwise could confuse the forward-error-correcting (FEC) decoder with poor quality statistics. Thus we propose applying (8) to (4) to condition the output LLRs as in

$$\begin{aligned}\lambda_k^{e\prime} =& \frac{1}{2}\min_{\mathbf{x}\in \mathrm{Z}^{k^-}}\left(\frac{1}{r_\lambda\sigma_n^2}\left\|\mathbf{y}-\mathbf{H}\mathbf{s}^{k^-}\right\|^2 - \mathbf{x}^{\setminus k}\cdot\boldsymbol{\lambda}^{a,\setminus k}\right)\\ &-\frac{1}{2}\min_{\mathbf{x}\in \mathrm{Z}^{k^+}}\left(\frac{1}{r_\lambda\sigma_n^2}\left\|\mathbf{y}-\mathbf{H}\mathbf{s}^{k^+}\right\|^2 - \mathbf{x}^{\setminus k}\cdot\boldsymbol{\lambda}^{a,\setminus k}\right).\end{aligned} \tag{9}$$

Here, we have presented a brief heuristic explanation of the X-MCMC detector. For a more complete derivation and mathematical analysis see [12].

VI. Channel Models

Using the correct channel model in simulation has proven to be an important aspect of understanding the performance, complexity, and overall behavior of the MCMC and X-MCMC detectors. If the channel model used is less ill-conditioned than real channels, the MCMC detector converges easily with a small number of Gibbs samples. This can lead to an overoptimistic assessment of performance and complexity.

The condition of the channel matrix depends on how correlated its columns are. When the columns are correlated, it is more difficult to separate the spatial streams and noise can more easily interfere with the signal. Thus, in addition to the signal-to-noise ratio (SNR) of the channel, the near-orthogonality of the columns of $\mathbf{H}$ is an important parameter affecting the capacity of a MIMO channel. It has been shown that the condition number (CN) of the matrix, the ratio between the strongest and weakest singular values, is a good metric to quantify the quality of a channel $\mathbf{H}$ [13]. This is defined as

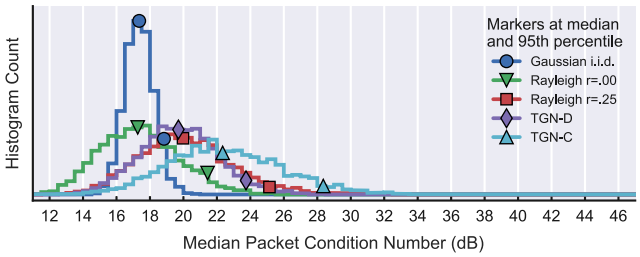
$$\mathrm{CN}(\mathbf{H}) = \frac{v_{\max}}{v_{\min}} = \frac{\sqrt{\mu_{\max}}}{\sqrt{\mu_{\min}}} \tag{10}$$

where $v_{\max}$ and $v_{\min}$ are the maximum and minimum singular values of $\mathbf{H}$ and $\mu_{\max}$ and $\mu_{\min}$ are the maximum and minimum eigenvalues of $\mathbf{H}^{\dagger}\mathbf{H}$. It can also be expressed in dB by applying $20\log_{10}(\cdot)$ to the ratio.

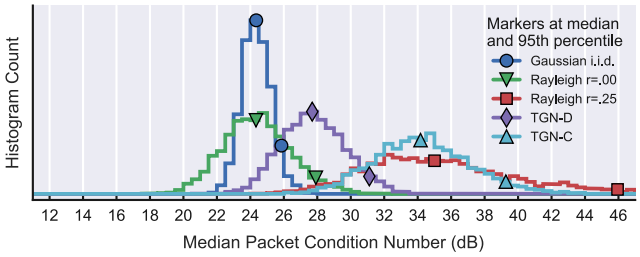
To compare the orthogonality and therefore difficulty of several common channel models, see the CN histograms in Fig. 2. Each data point is the median CN of 52 frequency-domain channels corresponding to the active OFDM subcarriers in 20 MHz bandwidth 802.11ac. For the Gaussian i.i.d. model, all channels were independently generated from i.i.d. complex Gaussian variables of zero mean and 0.5 variance per dimension. In the Rayleigh and TGn models, a single time-domain realization is used to produce the 52 frequency-domain channels per data point. The Rayleigh model with non-zero r is generated as in [7] where r adjusts the strength of the added correlation. The TGn models are from the 802.11n WiFi specification in [14] and is based on defined clusters of scatterers. The TGn model letter designates differing clusters and delay profiles which affects the degree of correlation and severity of fading.

The Rayleigh $r = 0.25$ model was confirmed as a valid model for 4-antenna indoor environments with measurements in our prior work [7], but to allow better comparisons with other papers in the literature, we now use the standard 802.11n WiFi TGn specified channel models. As can be seen in Fig. 2, the previous model matches well with the statistics of the TGn Model-D for 4-antenna situations.

The most important observation to note in Fig. 2 is that the Gaussian i.i.d. channels, the most commonly used model in the MCMC literature [5], [6], [9], have a much lower condition number and therefore is much easier than the other models. This means that any performance and complexity analysis analysis performed using an i.i.d. Gaussian channel model should be considered overoptimistic compared to real indoor environments. This will be revisited with the results in Section IX.



(a) 4-antenna channels generated.



(b) 8-antenna channels generated.

Fig. 2. Comparison of channel distributions using various models. Each data point is the median of a set of 52 frequency-domain channels corresponding to the active OFDM subcarriers in 802.11ac.

VII. Noise and Interference Metrics

One of the challenging aspects of comparing communications system simulations to testbed measurements is quantifying interference. During simulations, often the only form of external interference is additive white Gaussian noise (AWGN) with constant variance over a test. This type of interference is quantified well with a typical signal-to-noise-ratio (SNR) metric comprising the average signal energy over the average noise energy. In real-world systems, there are many types of interference including noise, distortion, channel estimation error, and other active transmissions which all vary over both time and frequency. This variability is not effectively represented by an SNR measurement only made with training fields or pilot symbols. Physical methods to control a test such as cabling the receiver to transmitter or using an anechoic chamber are inappropriate for spatial-multiplexing MIMO because it removes the multipath environment needed for spatial reuse.

A metric often used in lab testing is the error-vector-magnitude (EVM) [15]. It uses knowledge of the transmitted data to calculate the ratio of signal energy to interference directly using the payload. This makes it useful in capturing time varying conditions. The inverse-EVM (IEVM) can be extended to MIMO with

$$\mathrm{IEVM}^2_{\mathrm{MIMO}} = \frac{\sum_{i}^{N_{\mathrm{subc}}} \|\mathbf{s}_i\|^2}{\sum_{i}^{N_{\mathrm{subc}}} \|\hat{\mathbf{s}}_i - \mathbf{s}_i\|^2} = \frac{\sum_{i}^{N_{\mathrm{subc}}} \|\mathbf{s}_i\|^2}{\sum_{i}^{N_{\mathrm{subc}}} \left\|\hat{\mathbf{H}}_i^{-1}\mathbf{y}_i - \mathbf{s}_i)\right\|^2} \tag{11}$$

where N_{subc} is the number of active OFDM subcarriers within a frame and $\{\hat{\cdot}\}$ designates an estimate. The problem with this method is that when the channel is ill-conditioned, the matrix inverse $\mathbf{H}_i^{-1}$ can create a large amount of noise enhancement. This results in an overestimate of the interference which

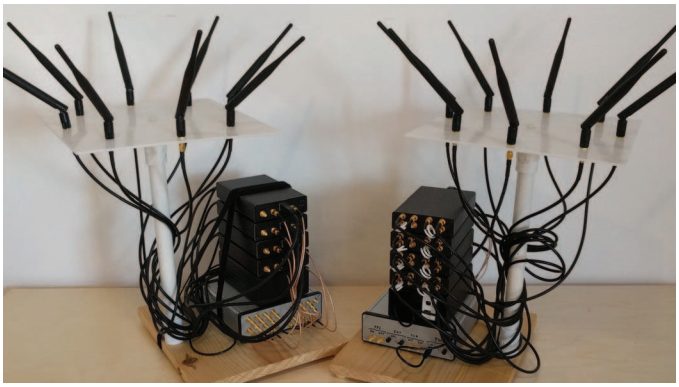


Fig. 3. 8-antenna experimental testbed based on Ettus USRP B210 software defined radios.

near maximum-likelihood detectors such as MAP, MCMC, X-MCMC, and K-Best do not experience. During our testing, this overestimate was commonly in the order of 10-20 dB which made it misleading and unreliable. Improvements could be made by replacing the channel inverse based estimate of $\hat{\mathbf{s}}$ with an alternative, such as MMSE, which creating less noise enhancement, but instead we propose removing the need for the estimate completely.

We have developed the harmonic-mean-signal to arithmetic-mean-distortion ratio (HSADR) metric as an SNR measure that uses the payload and avoids the noise enhancement of IEVM. It is defined as

$$\text{HSADR} = \frac{\left(\frac{1}{N_{\text{subc}}}\sum_{i}^{N_{\text{subc}}}\left(\left\|\hat{\mathbf{H}}_i\mathbf{s}_i\right\|^2\right)^{-1}\right)^{-1}}{\frac{1}{N_{\text{subc}}}\sum_{i}^{N_{\text{subc}}}\left\|\mathbf{y}_i-\hat{\mathbf{H}}_i\mathbf{s}_i\right\|^2}. \tag{12}$$

There are two important features of this definition. First, the channel in IEVM is moved from the bottom of the ratio to the top, thus removing the problematic noise enhancement. Next, the arithmetic-mean-signal, which emphasizes large values, is replaced with the harmonic-mean, which emphasizes small values [16]. In a single-input single-output (SISO) EVM calculation, taking the arithmetic-mean of the noise over channel gain (n/h) results in emphasizing small, weak channel conditions, i.e. deep fades. When the channel is moved to the top of the ratio and no longer inverted, poor channels have little effect on the metric. By taking the harmonic-mean of $\hat{\mathbf{H}}\mathbf{s}$, weak channels have a larger impact on the mean which is desirable.

Using the HSADR metric of (12) on testbed measurements with irregularly interfered and distorted packet payloads results in a more consistent analysis with smooth BER curves. Note that the HSADR metric requires prior knowledge of the transmitted payload and so is only used for analysis and plotting purposes and not by the signal processing blocks.

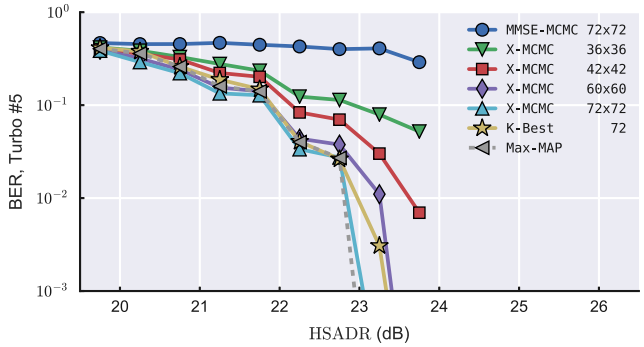
VIII. 8-ANTENNA MIMO TESTBED

The testbed shown in Fig. 3 is in its third generation and is a useful contribution to the research community. Based on the Ettus B210 software defined radio and Analog Devices AD9361 chipset, it cost $10,000 US dollars to build, making it affordable for many wireless researchers. This is in contrast to most commercially available alternatives which cost $60,000 to $250,000 US dollars, usually with less frequency coverage. The main features that these expensive options additionally include are increased bandwidth, improved noise figure, *official* >2-stream MIMO support, and larger capture capabilities. Here, *official* means that a company supports an intended functionality whereas *unofficial* means it is possible but without the support of the manufacturer. Our B210 solution is capable of 30.7 MHz instantaneous bandwidth, 70MHz-6GHz frequency range, and *unofficial* 16-stream MIMO in bursts (largest number of streams verified, larger may be possible). This provides the ability to work with all of the 802.11ac and worldwide LTE bands as well as produce the 20MHz bandwidth needed for portions of the WiFi and LTE protocols.

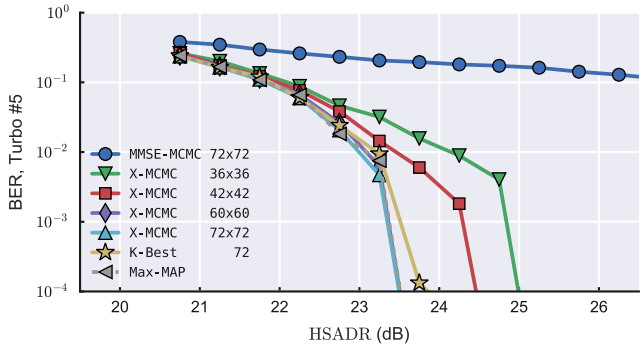
The main hardware components of each side of the testbed are four Ettus B210 radios (each containing two transceivers), one synchronized source of four 10MHz clock and pulse-per-second (PPS) signals, a powered 4-port USB3 hub, and a linux computer. Optionally, we have added the Mini-circuits ZX60-83LN-S+ 0.5-7GHz, 21dB gain, broadband amplifiers to the transmit side. The powered hub is capable of powering both a B210 and two amplifiers off of one USB3 cable, reducing system cabling and complexity. A potential issue in this design is avoiding low quality USB3 chipsets which are known to create problems with the B210 radios. We have seen dropped data, re-ordered data, and overheating chips that require hard resets. Using a quality hub with good linux drivers avoids this irregular behavior.

A MIMO synchronization source is available from Ettus and others. We elected to build our own lower cost, higher performance one based on a 5ppb 10 MHz ovenized oscillator, an Atmel ATTiny85 microcontroller for PPS generation, and SN74AC logic inverters as buffers. In testing, we achieved <10ns synchronization offset between radios which was at the limit of our testing procedures and a small fraction of the shortest 400ns guard interval of 802.11ac WiFi.

Overall, the custom software was the most time consuming and difficult portion of building the testbed. We based our design on Python and C++. This provides flexibility and a great degree of control of the hardware while still being developer friendly. As Python is being used to produce all of our simulation results, it was natural to use it to synthesize and process 802.11ac packets. The packets are sent and received through a socket to a separate C++ program controlling the radios with the Ettus USRP hardware driver (UHD) API. Using the burst mode is essential in doing large MIMO sizes as we have found that the USB3 connection is limited to 1.1 Gbps in practice, much less than the needed 6.4 Gbps for 8-antennas streaming at 25Msps and 32 bits per raw RF sample. These radios readily buffer 10k samples, allowing individual high bandwidth MIMO packets to be sent and received using timed commands despite the USB3 bottleneck. Finally, it should be noted that the Ettus UHD library does not officially support greater than 2x MIMO on the B210 as of UHD library version 3.9. Since we were using C++ it was straightforward to make a custom wrapper which synchronizes and controls multiple

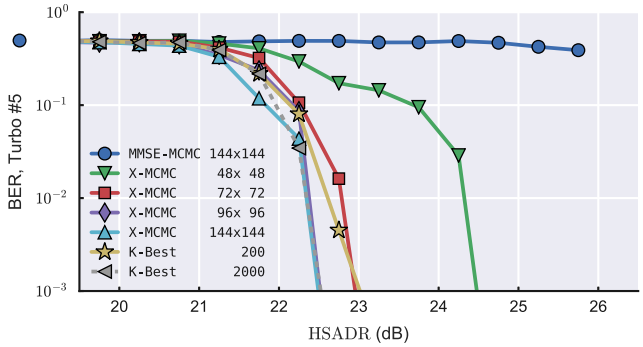


(a) Testbed transmitted data using 802.11ac.

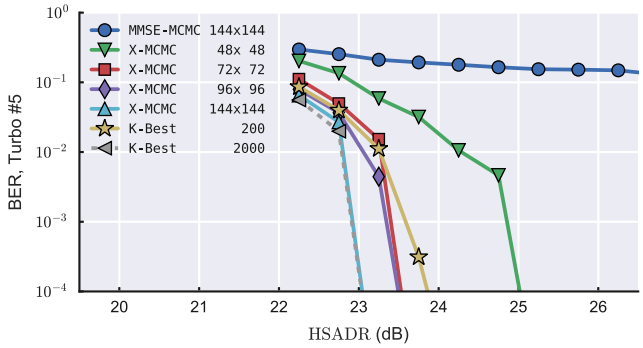


(b) Simulation with WiFi TGn Model-C channel model.

Fig. 4. 4 antenna, 64 QAM, 3/4 LDPC coding rate BER curves. MCMC values are $N_{\text{gibbs}} \times N_{\text{iter}}$ and K-Best values are the list length.



(a) Testbed transmitted data using 802.11ac.



(b) Simulation with WiFi TGn Model-D channel model.

Fig. 5. 8 antenna, 64 QAM, 3/4 LDPC coding rate BER curves. MCMC values are $N_{\text{gibbs}} \times N_{\text{iter}}$ and K-Best values are the list length.

radios.

Several more implementation details specific to our measurements may be helpful to those designing similar systems. To align the receiver's short capture window with the sparsely transmitted packets, on startup it first captures 1-stream of 11 ms of data. By transmitting data at exactly 10ms intervals, the long 1-stream collection can be used to identify a timing offset between the two systems and align future 8-stream MIMO collections which can only be done in short bursts. To make the data in each packet randomized but still known, a heavily coded extra field was added to the packet header containing the random seed used in creating the payload data. This allows the receiver to reconstruct the true transmitted bits for error analysis. On each burst mode transmission of the B210, there appears to be a phase instability for the first 10 micro-seconds of non-zero data, thus a random pad must be placed at the beginning of each packet.

IX. Results

The following testbed measurement results were transmitted using an 802.11ac packet structure over the system described in Section VIII at 2.484 GHz. The low-density parity-check (LDPC) code described in the protocol was used for both measurement and simulation. Simulations were performed in the frequency-domain without use of the WiFi time-domain packet structure.

The most important factor for matching testbed to simulation results is selection of a channel model with sufficient correlation to produce similar distributions of ill-conditioned $\mathbf{H}$ matrices. In Fig. 6, condition number histograms are shown using the same testbed data set as in Fig. 4a and 5a. By comparing these distributions to the channel models in Fig. 2, we see that the WiFi TGn Model-C is a reasonable match for the 4-antenna measurements and Model-D for the 8-antenna measurements. These specific models have been used in the corresponding simulation results of Fig. 4b and 5b.

Secondly, channel realizations simulated over an LDPC block should not be independent. A good procedure is to generate a single time-domain channel realization for each LDPC block, and then extract the same 52 active OFDM subcarrier realizations as used in 802.11ac. If more than 52 realizations are needed for the LDPC block, then they are reused. This results in the likelihood of an LDPC block encountering many deep fades simultaneously much more likely resulting in a significant shift of the BER curves to the right.

In the 4-antenna results of Fig. 4, it can be seen that X-MCMC approaches near Max-MAP performance as the number of Gibbs samplers increases. Similarly, the 8-antenna results of Fig. 5 confirm this, but since MAP is too complex to calculate on the 8-antenna results, a very large K-Best detector is used as an estimate of the Max-MAP performance bound. A moderate-sized K-Best detector is used as a reference point so that the difficulty of this channel can be better appreciated. Notice that the MMSE initialized MCMC detector does not converge whereas the X-MCMC detector converges to near Max-MAP performance. This verifies our claims that X-MCMC has solved the high SNR stalling problem.

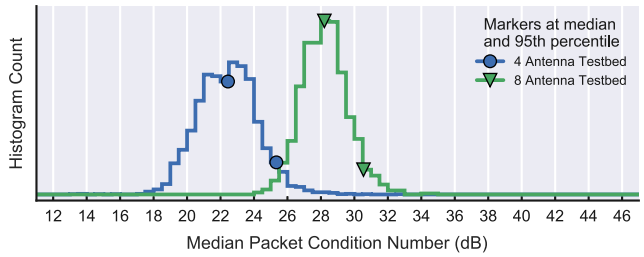


Fig. 6. Distribution of the channels observed in the testbed data used to generate BER curve results. Median is across the N_{subc} channels needed for one LDPC codeword.

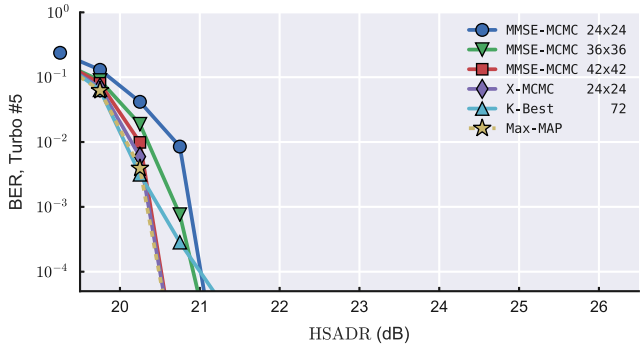


Fig. 7. Simulated BER curves with i.i.d complex Gaussian channel model, 4 antennas, 64 QAM, 3/4 LDPC coding. This channel model is the most common in the MCMC literature and is too well-conditioned. Notice that here all of the detectors easily converge whereas some methods completely fail on real indoor channels, Fig. 4.

It is useful to see how much of an impact the wrong channel model can have on analysis conclusions. In Fig. 7, an i.i.d. complex Gaussian channel is used as in most of the MCMC literature [5], [6], [9]. Compared to the more accurate and much more difficulty TGn Model-C in used in Fig. 4b and the measurements in Fig. 4a, a much smaller MCMC Gibbs sampler is needed for the BER to converge to Max-MAP performance. Also, there is a large disagreement over the effectiveness of the MMSE initialized MCMC detector. The Gaussian channel model shows a functioning solution whereas both the real-world and TGn simulated model performance shows strong stalling.

Even after using a well matched channel model and the HSADR interference metric in (12), the waterfall locations of the simulation results do not perfectly match measurement in Fig. 4 and 5. This is acceptable and to be expected as the distribution of condition numbers do not perfectly match and HSADR is an imperfect metric. What is important is that they are close (within 1-2 dB), the convergence behavior of all detector methods agree, and the parameters of MCMC and K-Best necessary for convergence are similar. This means that in the future these simulation techniques and models can be used to perform a much deeper analysis of MCMC and X-MCMC without performing real-world tests. If a closer match between simulation and measurement results is desired, see our previous work in [7] which uses condition number slicing to control channel distributions. Slicing is no longer our preferred method as it removes the contribution of outliers from the data sets.

X. Conclusion

The previous MCMC algorithms introduced in the literature display problematic stalling behavior at high SNR. Here we have proposed the X-MCMC method and shown results demonstrating that it solves this problem. We have also described the construction of a low cost and effective 8-antenna MIMO testbed which others may replicate to do testing at all 802.11ac and worldwide LTE frequency bands. We matched our measurements with simulations and showed that the channel models most often used in the MCMC literature are insufficient, leading to overoptimistic conclusions.

Future work should include a VLSI implementation and complexity analysis of the algorithm. A more thorough mathematical investigation and analysis of X-MCMC is under preparation [12].

References

[1] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "802.11 with multiple antennas for dummies," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 1, pp. 19–25, 2010.

[2] F. Boccardi, B. Clerckx, A. Ghosh, E. Hardouin, G. Jongren, K. Kusume, E. Onggosanusi, and Y. Tang, "Multiple-antenna techniques in LTE-advanced," *IEEE Communications Magazine*, vol. 50, no. 3, pp. 114–121, 2012.

[3] B. M. Hochwald and S. Ten Brink, "Achieving near-capacity on a multiple-antenna channel," *IEEE Transactions on Communications*, vol. 51, no. 3, pp. 389–399, 2003.

[4] Z. Guo and P. Nilsson, "Algorithm and implementation of the K-best sphere decoding for MIMO detection," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 491–503, 2006.

[5] B. Farhang-Boroujeny, H. Zhu, and Z. Shi, "Markov chain Monte Carlo algorithms for CDMA and MIMO communication systems," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1896–1909, 2006.

[6] S. A. Laraway and B. Farhang-Boroujeny, "Implementation of a Markov chain Monte Carlo based multiuser/MIMO detector," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 56, no. 1, pp. 246–255, 2009.

[7] J. C. Hedstrom, C. H. Yuen, and B. Farhang-Boroujeny, "Markov chain Monte Carlo based multiuser/MIMO detector: 802.11ac implementation and measurement," in *IEEE International Conference on Communications (ICC)*, 2015, pp. 4846–4852.

[8] X. Mao, P. Amini, and B. Farhang-Boroujeny, "Markov chain Monte Carlo MIMO detection methods for high signal-to-noise ratio regimes," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2007, pp. 3979–3983.

[9] M. Hansen, B. Hassibi, A. G. Dimakis, and W. Xu, "Near-optimal detection in MIMO systems using gibbs sampling," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2009, pp. 1–6.

[10] R. Peng, K. H. Teo, J. Zhang, and R.-R. Chen, "Low-complexity hybrid QRD-MCMC MIMO detection," in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2008, pp. 1–5.

[11] S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, *Handbook of Markov Chain Monte Carlo*. Chapman and Hall/CRC, 2011.

[12] J. C. Hedstrom, C. H. Yuen, R.-R. Chen, and B. Farhang-Boroujeny, "Achieving near MAP performance with an excited Markov chain Monte Carlo MIMO detector," *Under preparation*, 2017.

[13] H. Artés, D. Seethaler, and F. Hlawatsch, "Efficient detection algorithms for MIMO channels: A geometrical approach to approximate ML detection," *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2808–2820, 2003.

[14] V. Erceg, L. Schumacher, P. Kyritsi, A. Molisch, D. Baum, A. Gorokhov, C. Oestges, Q. Li, K. Yu, N. Tal *et al.*, "IEEE P802.11 wireless LANs - indoor MIMO WLAN TGn channel models," 2004, iEEE 802.11 document 03/940r4.

[15] M. D. McKinley, K. A. Remley, M. Myslinski, J. S. Kenney, D. Schreurs, and B. Nauwelaers, "EVM calculation for broadband modulated signals," in *ARFTG Conf. Dig*, 2004, pp. 45–52.

[16] P. S. Bullen, *Handbook of Means and Their Inequalities*. Springer Science & Business Media, 2013, vol. 560, ch. The Arithmetic, Geometric, and Harmonic Means.