Technical Report 158

Project Title:

Online Matching, Black-box Optimization and Hyper-parameter Tuning

Research Supervisor: Sanjay Shakkottai
Wireless Networking and Communications Group

August 2020

Data-Supported Transportation Operations & Planning Center (D-STOP)

A Tier 1 USDOT University Transportation Center at The University of Texas at Austin

D-STOP is a collaborative initiative by researchers at the Center for Transportation Research and the Wireless Networking and Communications Group at The University of Texas at Austin.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
D-STOP/2020/158		

4. Title and Subtitle	5. Report Date
MmWave Codebook Selection in Rapidly-Varying Channels via Multinomial Thompson Sampling	August 2020
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Yi Zhang, Soumya Basu, Sanjay Shakkottai, and Robert W. Heath Jr.	Report 158

9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Data-Supported Transportation Operations & Planning Center (D-STOP) The University of Texas at Austin 3925 W. Braker Lane, 4th Floor Austin, TX 78759	11. Contract or Grant No. DTRT13-G-UTC58

12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
United States Department of Transportation University Transportation Centers 1200 New Jersey Avenue, SE Washington, DC 20590	14. Sponsoring Agency Code

15. Supplementary Notes

Supported by a grant from the U.S. Department of Transportation, University Transportation Centers Program.

16. Abstract

Millimeter-wave (mmWave) communications, using directional beams, is a key enabler for high-throughput mobile ad hoc networks. These directional beams are organized into multiple codebooks according to beam resolution, with each codebook consisting of a set of equal width beams that cover the whole angular space. The codebook with narrow beams delivers high throughput, at the expense of scanning time. Therefore overall throughput maximization is achieved by selecting a mmWave codebook that balances between beamwidth (beamforming gain) and beam alignment overhead. Further, these codebooks have some potential natural structures such as the non-decreasing instantaneous rate or the unimodal throughput as one traverses from the codebook with wide beams to the one with narrow beams. We study the codebook selection problem through a multi-armed bandit (MAB) formulation in mmWave networks with rapidly-varying channels. We develop multiple novel Thompson Sampling-based algorithms for our setting given different codebook structures with theoretical guarantees on regret. We further collect real-world (60 GHz) measurements with 12-antenna phased arrays, and show the performance benefits of our approaches in an IEEE 802.11ad/ay emulation setting..

17. Key Words	18. Distribution Statement
millimeter-wave, codebook optimization, rapidly-varying channel, multi-armed bandit, Thompson sampling, experimental measurements	No restrictions. This document is available to the public through NTIS (http://www.ntis.gov): National Technical Information Service 5285 Port Royal Road Springfield, Virginia 22161

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified		

Form DOT F 1700.7 (8-72) **Reproduction of completed page authorized**

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation's University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Acknowledgements

The authors recognize that support for this research was provided by a grant from the U.S. Department of Transportation, University Transportation Centers.

MmWave Codebook Selection in Rapidly-Varying Channels via Multinomial Thompson Sampling

Yi Zhang
The University of Texas at Austin
Austin, TX, USA
yi.zhang.cn@utexas.edu

Soumya Basu
Google, LLC
Mountain View, CA, USA
basusoumya@google.com

Sanjay Shakkottai
The University of Texas at Austin
Austin, TX, USA
sanjay.shakkottai@utexas.edu

Robert W. Heath Jr.
North Carolina State University
Raleigh, NC, USA
rwheathjr@ncsu.edu

ABSTRACT

Millimeter-wave (mmWave) communications, using directional beams, is a key enabler for high-throughput mobile ad hoc networks. These directional beams are organized into multiple codebooks according to beam resolution, with each codebook consisting of a set of equal width beams that cover the whole angular space. The codebook with narrow beams delivers high throughput, at the expense of scanning time. Therefore overall throughput maximization is achieved by selecting a mmWave codebook that balances between beamwidth (beamforming gain) and beam alignment overhead. Further, these codebooks have some potential natural structures such as the non-decreasing instantaneous rate or the unimodal throughput as one traverses from the codebook with wide beams to the one with narrow beams. We study the codebook selection problem through a multi-armed bandit (MAB) formulation in mmWave networks with rapidly-varying channels. We develop multiple novel Thompson Sampling-based algorithms for our setting given different codebook structures with theoretical guarantees on regret. We further collect real-world (60 GHz) measurements with 12-antenna phased arrays, and show the performance benefits of our approaches in an IEEE 802.11ad/ay emulation setting.

CCS CONCEPTS

• **Networks** → **Mobile networks**; • **Computing methodologies** → **Machine learning algorithms**; • **Mathematics of computing** → **Bayesian computation**.

KEYWORDS

millimeter-wave, codebook optimization, rapidly-varying channel, multi-armed bandit, Thompson sampling, experimental measurements

1 INTRODUCTION

Large antenna arrays are key to the success of millimeter-wave (mmWave) networks because of their high directional gain. However, to get the benefits of this directionality, transmitters (TX) and receivers (RX) need to align their respective beams to maximize throughput. Each radio has a codebook – a collection of beams with a predefined beam resolution (indicated by beamwidth), and covering the whole angular space (see Figure 1) – the radios exhaustively sweep over the beams in a codebook to establish the optimal beam-pair link [28]. Such sweep-based techniques have been incorporated into standards such as IEEE 802.11ad/ay [4] and 5G NR [5], because of robustness and good coverage [27].

While a codebook consisting of beams with a narrow beamwidth is beneficial as these beams provide higher beamforming gain (and thus a higher signal-to-noise-ratio (SNR)), it comes at a price. Such a codebook correspondingly contains a large number of beams to cover angular space, with the time taken to sweep over them being linear in the number of beams [16]. Indeed with emerging standards such as IEEE 802.11ay, the number of beams can scale to as much as 2048 [4, 25]. Furthermore, a beam-pair link needs to be frequently re-established in mobile and rapidly varying channel settings (see [9]), thus resulting in significant overheads.

To resolve this tension between high throughput and large sweep times, a promising and practical solution is to have multiple codebooks of different beam resolutions (each codebook spanning the whole angular space, see Figure 1 and **Remark** 1), and choose a specific codebook in a *scenario-specific* manner. Depending on the device location and frequency of link realignment (which is driven by scenario-specific device location/mobility, and channel variability), the radio might choose to use a codebook of wide beams (low beamforming gain but fast sweep, beneficial to devices that either require frequent realignment or can tolerate low beamforming gain due to their central location), or at the other extreme, a codebook of narrow beams (high beamforming gain but slow sweep, beneficial to devices requiring infrequent realignment or located far-away from the base station). Indeed the experiments in [36] have shown that the optimal beam resolution is scenario-specific, and unsuitable choices could severely degrade the overall throughput. This intuition has propagated into standards, where a *family of codebooks* has been first standardized in IEEE 802.15.3c millimeter-wave WPANs [1] and further proposed in the ongoing standardization of IEEE 802.11ay by [25].

In this paper, we focus on the codebook selection problem given a set of mmWave codebooks ranging from low to high beam resolution (see Figure 1). Our goal is to learn the optimal codebook by dynamically exploring the trade-off between the high instantaneous throughput provided by the codebook of narrow beams and the low overhead associated with the codebook of wide beams. We exploit online learning techniques to design codebook selection algorithms for rapidly-varying mmWave networks. **The major contributions are summarized below:**

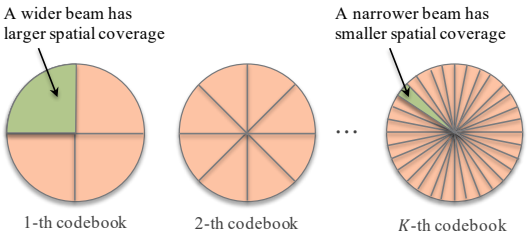

Figure 1: Example codebooks of directional beams

(1) **Algorithm Design:** Using a multi-armed bandit (MAB) framework, we propose multiple novel Thompson Sampling (TS)-based bandit algorithms using *Dirichlet priors* for the codebook selection problem. In particular, we first propose a generic TS algorithm without requiring any structure among codebooks. Second, we propose a constrained TS algorithm that exploits the known *general structure* among codebooks to further improve the system performance. Most importantly, we propose a Unimodal TS (UTS) algorithm to deal with a well-observed natural structure among a family of codebooks ranging from low to high resolution – the effective throughputs of codebooks often have a unimodal property.

(2) **Theoretical and Empirical Results:** We provide theoretical guarantees for the proposed algorithms by deriving upper bounds for their regrets (expected loss in cumulative throughput) with respect to a genie algorithm that always uses the optimal codebook. In particular, our proofs provide the theoretical guarantee for the UTS with *Dirichlet priors*, which is an important missing part of the state-of-the-art TS algorithms. Next, we collect real-world channel measurements at 60 GHz with two 12-antenna phased arrays, and use them to validate the proposed algorithms by emulating a realistic IEEE 802.11ad system. Our results show that the proposed TS-based algorithms are superior to state-of-the-art bandit algorithms.

2 SYSTEM MODEL

We consider a slot-based mobile ad hoc mmWave system, in which a TX establishes the wireless link with an RX by doing the codebook-based beam scanning. Specifically, a codebook is a set of directional beams of the same beam resolution (indicated by beamwidth) that covers the whole angular space. There are multiple codebooks available at the TX while the RX only has one fixed codebook (antenna array size and power consumption are generally limited at the RX, i.e. mobile devices). Different codebooks have directional beams of different beamwidth, which helps balancing high beamforming gain (by delivering high SNR using narrow beams) and low training overhead (by avoiding mass sweeps using wide beams). See Figure 1 for a pictorial representation of the set of codebooks.

In mmWave systems, each communication time slot includes a beam alignment phase and a data transmission phase. The evolution of a time slot is described as follows. At the beam alignment phase, the TX selects one of the available codebooks to perform the beam alignment with the RX by testing all the beams in this codebook. At the end of this phase, the index of the beam with the highest received signal strength (RSS) will be sent back to the TX. Subsequently, the TX will use this best beam to transmit data for the remaining time resources in this slot, which is referred to as a data transmission phase. In particular, the TX will transmit the data with the highest supportable modulation and coding scheme (MCS), which is obtained by referring to a *predefined* RSS-MCS table. This is a typical mmWave system and the adopted beam alignment process is similar to the sector level sweep (SLS) used in IEEE 802.11ad/ay [3, 4] and 5G NR [5]. *Our objective* is to identify the optimal codebook that maximizes the expected system throughput.

The codebook generation is out of the scope of this work. A simple way to generate multiple codebooks of different beamwidths, shown in Figure 1, is to exploit antenna on/off techniques [39], which is also used in our experimental evaluation.

Remark 1. *Compared to gathering all the beams of different resolutions into a giant codebook, organizing the beams into multiple codebooks by their width has the following practical advantages: (1) It facilitates the beam management in the context that the size of the mmWave antenna array is scaling up [42]. (2) It enables the codebook optimization in a scenario-specific manner (see experimental results in [36]), leading to greatly improved performance. (3) From the perspective of practical implementation, using one codebook of equal-width beams for a single link establishment can avoid numerous antenna on/off operations (required by changing beamwidth [39]), which could reduce the operation overhead and simplify the antenna hardware designs. As mentioned earlier, standard bodies are recognizing the benefits of a family of codebooks, e.g. IEEE 802.15.3c millimeter-wave WPANs [1] and proposals in IEEE 802.11ay by [25].*

3 PROBLEM STATEMENT

In this section, we mathematically characterize the beam alignment and the data transmission phases described in Section 2. We study the codebook selection problem through a multi-armed bandit (MAB) framework. At each time-slot, one of K possible codebooks (aka actions) is chosen by the learning algorithm (aka player), and the corresponding effective data rate (aka reward) is observed. By learning the choice of the best codebook, the goal is to minimize the cumulative loss with respect to an omniscient genie [8].

3.1 RSS-MCS table

As mentioned in Section 2, there exists a predefined RSS-MCS table used by the TX to decide which is the highest supportable MCS given the best RSS feedback by the RX. We suppose this RSS-MCS table has $(M+1)$ levels of MCS. The data rate associated with MCS m is the m-th element of a rate vector $\tilde{\mathbf{r}} = [\tilde{r}_0, \tilde{r}_1, \ldots, \tilde{r}_M]^{\mathrm{T}}$, where $\tilde{r}_0 < \tilde{r}_1 < \ldots < \tilde{r}_M$, and the minimum required RSS for supporting MCS m is denoted as rss_m, which yields a RSS vector $\mathbf{rss} = [\mathrm{rss}_0, \mathrm{rss}_1, \ldots, \mathrm{rss}_M]^{\mathrm{T}}$. In particular, MCS 0 represents the data rate of 0 ($\tilde{r}_0 = 0$ and $\mathrm{rss}_0 = -\inf$), namely that the RSS is too low to support any data transmission (failed link connection). Without loss of generality, we define a normalized rate vector by dividing $\tilde{\mathbf{r}}$ by $\tilde{r}_M$, which is denoted as $\mathbf{r} = [r_0, r_1, \ldots, r_M]^{\mathrm{T}}$, where $r_m = \frac{\tilde{r}_m}{\tilde{r}_M}$. Thus, r_m is bounded by $[0, 1]$ and we will use this normalized rate vector $\mathbf{r}$ in the following. We denote $[K]^+ \triangleq \{1, 2, \ldots, K\}$, $[K] \triangleq \{0, 1, \ldots, K\}$ and $\mathbf{1}\{\cdot\}$ as the indicator function for later use.

3.2 Channel distribution and evolution of a time slot

We consider a discrete-time setting, where $t = 1, 2, ..., T$ is a finite time horizon and each time step represents a communication time slot. We denote K as the number of codebooks at the TX and S_k as the number of beams in the k-th codebook. We denote the random mmWave channel at time slot t as $h(t)$ following a discrete state channel distribution $\mathcal{H}$ over some (possibly) continuous state-space. As the channels are rapidly varying in mmWave MANETs, we suppose that the channel state realizations of different time slots are independent of each other [19].

In each time slot, at the beam alignment phase, the TX chooses a codebook $k \in [K]^+$ and sequentially tests each beam in this codebook (beam alignment for the specified codebook). Denoting by rss(t,k) the maximum RSS obtained by sweeping over all the beams in the k-th codebook, we then have

$$\mathrm{rss}(t,k) = \max_{j\in[S_k]^+} f(h(t),k,j), \tag{1}$$

where f is an unknown function that reflects the *overall physical layer impact* on the received signals, which includes channel gain, sidelobe effects, RF impairments, beam pattern imperfection, thermal noise, etc.

Given the maximum RSS, the TX uses a predefined RSS-MCS table to determine the highest supportable MCS for the data transmission phase, which can be mathematically expressed as

$$r(t) = \max_{m\in[M]} \mathbf{1}\left\{\mathrm{rss}(t,I(t)) \geq \mathrm{rss}_m\right\} r_m, \tag{2}$$

where $I(t)$ denotes the index of codebook selected at the t-th time slot and $r(t)$ is the determined data rate, which is termed as *instantaneous data rate*. As a result, we can see that given a selected codebook $I(t) \in [K]^+$ by a certain policy, the instantaneous data rate $r(t)$ follows a one-trial **multinomial distribution** with the support $\{r_0, r_1, \ldots, r_M\}$ and the parameter $\mathbf{p}_k = [p_{0,k}, p_{1,k}, \ldots, p_{M,k}]^{\mathrm{T}}$, where $p_{m,k} = \mathbf{P}\left\{r(t) = r_m | I(t) = k\right\}$, $m \in [M]$ and $k \in [K]^+$.

3.3 Reward of codebooks and cumulative regret of the system

We adopt a model-free framework to formulate our codebook selection problem, which directly characterizes the performance of codebooks by their multinomial distributions, i.e. parameters $\{\mathbf{p}_k\}_{k=1}^K$. This allows us to bypass the complex assumptions on the channel distribution $\mathcal{H}$ and the unknown function f in (1). The performance metric of the k-th codebook (the mean reward of k-th arm) is the effective data rate of the codebook, $r_k^{\mathrm{eff}}(t)$ (defined shortly). We first denote $r_k^{\mathrm{ins}}(t)$ as the instantaneous data rate of codebook k, whose expectation can be given as $\mathbb{E}\left[r_k^{\mathrm{ins}}(t)\right] = \mathbf{r}^{\mathrm{T}}\mathbf{p}_k$. As described before, only part of the total time slot is used for data transmission, which motivates us to define a variable, termed as *effective coefficient*, to present the ratio of time that is allocated for the data transmission phase, which is given as $C_k^{\mathrm{eff}} = \left(T^{\mathrm{slot}} - T_k^{\mathrm{train}}\right)/T^{\mathrm{slot}}$, where T_k^{train} is a codebook-dependent constant representing the total beam alignment time including getting feedback and T^{slot} is the fixed time-slot duration.

With C_k^{eff}, we can now define the *effective data rate*, denoted by $r_k^{\mathrm{eff}}(t)$, to represent the average data rate over the whole time slot, which is given as $r_k^{\mathrm{eff}}(t) = r_k^{\mathrm{ins}}(t)C_k^{\mathrm{eff}}$. Note that $r_k^{\mathrm{eff}}(t)$ determines the real system throughput when the k-th codebook is chosen. Therefore, the reward of k-th arm follows a multinomial distribution with the support $\{r_0 C_k^{\mathrm{eff}}, r_1 C_k^{\mathrm{eff}}, \ldots, r_M C_k^{\mathrm{eff}}\}$ and the parameter $\{p_{0,k}, p_{1,k}, \ldots, p_{M,k}$, which gives its expectation μ_k as

$$\mu_k = \mathbb{E}\left[r_k^{\mathrm{eff}}(t)\right] = C_k^{\mathrm{eff}}\mathbf{r}^{\mathrm{T}}\mathbf{p}_k. \tag{3}$$

The *optimal codebook* $k^* = \arg_{k\in[K]^+}\max \mu_k$ is the one that provides the maximum expected effective data rate.

In this work, we consider minimizing the expected cumulative regret/loss over the T slots. The expected cumulative regret of a codebook selection algorithm is defined as the difference between the total expected reward of the optimal codebook and the total expected reward obtained by the algorithm, which can be given as

$$R(T) = \sum_{t=1}^{T}\mathbb{E}\left[r_{k^*}^{\mathrm{eff}}(t)\right] - \mathbb{E}\left[r_{I(t)}^{\mathrm{eff}}(t)\right] = T\mu_{k^*} - \sum_{t=1}^{T}\mu_{I(t)}. \tag{4}$$

3.4 Natural structure among codebooks and discussions

In this subsection, we incorporate the physical layer structural aspects of the codebooks as model assumptions. The following **Assumption 2** leverages the fact that aligned narrower beams provide higher beamforming gain, hence larger RSS as compared to their wider counterparts. Without loss of generality, we assume that the codebooks are numbered in terms of decreasing beamwidth (widest beamwidth numbered 1).

Assumption 2 (Nondecreasing instantaneous data rate). *For any two codebooks with indexes k_1 and k_2, such that $k_1 < k_2$, for all time $t \geq 1$, $rss(t,k_1) \leq rss(t,k_2)$ holds.*

Assumption 2 implies that a higher (non-lower) MCS can be supported by the codebook with larger index (finer beamwidth), which is mathematically given as

$$\mathbf{r}^{\mathrm{T}}\mathbf{p}_1 \leq \mathbf{r}^{\mathrm{T}}\mathbf{p}_2 \leq \ldots \leq \mathbf{r}^{\mathrm{T}}\mathbf{p}_K. \tag{5}$$

Training time for codebooks with wider beams is less, assuming training time per beam is constant, and thus we need to train fewer beams when using wider codebooks. This implies,

$$C_1^{\mathrm{eff}} > C_2^{\mathrm{eff}} > \ldots > C_K^{\mathrm{eff}}. \tag{6}$$

When the codebooks are efficiently designed, the following assumption is suitable for our system (see **Remark** 5).

Assumption 3 (Unimodal effective data rate). *The expected rewards of codebooks, i.e. $\{\mu_k\}_{k=1}^K$ (with, $\mu_k = C_k^{\mathrm{eff}}\mathbf{r}^{\mathrm{T}}\mathbf{p}_k$) follows a unimodal pattern, i.e. there exists a unique $k^* \in \{1, \ldots, K\}$ such that μ_k is increasing with k for all $k \leq k^*$, and μ_k is decreasing with k for all $k \geq k^*$:*

$$\mu_1 \leq \ldots \leq \mu_{k^*} \geq \ldots \geq \mu_K. \tag{7}$$

Thus, we have mathematically modeled the codebook selection problem in rapidly-varying mmWave channels as a MAB problem. In the next section, we will design efficient bandit algorithms to

solve it. A few remarks on the proposed framework are further listed below for completeness.

Remark 4. *We note that **Assumption** 2 and the equation (6) does not necessarily provide the unimodality described by (7). For example,* $\left\{C_k^{eff}\right\} = (0.8, 0.7, 0.4, 0.35)$ *and* $\{\mathbf{r}^T\mathbf{p}_k = (0.1, 0.2, 0.3, 0.4)$. *Similarly, **Assumption** 2 is not implied by **Assumption** 3.*

Remark 5. ***Assumption** 3 is motivated by the fact that the system Shannon capacity is a unimodal function of beamwidth when doing a 2D beam scanning, as discussed below. We use* b_k *to represent the width of beams in the k-th codebook. Suppose the size of the beam scanning area is* θ *(e.g.* $\theta = 360^\circ$ *for 2D-scanning), then we have* $T_k^{train} = \frac{\theta}{b_k} T^{mer}$*, where* T^{mer} *is the time duration for testing a single beam. Further, the beamforming gain can be roughly approximated as* $\frac{C_0}{b_k}$ *[2], where* C_0 *is a constant parameter related to the used antenna array. Thus, the Shannon capacity* r_k^{cap} *can be given as*

$$r_k^{cap} = B\left(1 - \frac{\theta T^{mer}}{b_k T^{slot}}\right)\log_2\left(1 + h\frac{C_0 P_{TX}}{b_k P_N}\right), \tag{8}$$

where B is the bandwidth, h is the channel effect, P_{TX} *is the transmit power and* P_N *is the noise power. By denoting* $C_1 \triangleq \frac{\theta T^{mer}}{T^{slot}}$ *and* $C_2 \triangleq \frac{C_0 P_{TX}}{P_N}$*,* r_k^{cap} *is sampled from the function* $r^{cap}(b)$ *given as*

$$r^{cap}(b) = B\left(1 - \frac{C_1}{b}\right)\log_2\left(1 + h\frac{C_2}{b}\right). \tag{9}$$

It can be shown that the function in (9) is unimodal with respect to b [33]. The throughput (mean reward of arm), however, is an expectation of this expression over the channel effect. Our assumption essentially states that even after taking an expectation, unimodality holds. Our numerical evaluation with the 3GPP NR outdoor channel model and real-world measurements both confirm this observation.

Remark 6. *We note that unimodality has been previously exploited in beam alignment [21]. Essentially, their notion of unimodality is that for a **single codebook** of beams, the performance of these beams have a unimodal pattern. Our notion of unimodality given in **Assumption** 3 is different. When we have multiple codebooks, each consisting of beams of the same resolution, the **performance of these codebooks** exhibit the unimodal structure. Our notion of codebook unimodality hinges on the trade-off between the increased scanning time for codebooks with a large number of narrow beams versus the increased instantaneous rate from the high directional gains.*

4 ALGORITHMS AND REGRET GUARANTEES

In this section, we design four online learning algorithms for different structural constraints on the set of codebooks. Our objective is to design algorithms that will maximize the use of the optimal codebook. An ideal algorithmic choice for this task is Thompson Sampling (TS) which is a popular Bayesian approach to solving MAB problems because of its efficient implementation and excellent empirical performance [10, 24]. The core of TS is to use the observations to dynamically update the posterior of a predefined prior distribution. The classic TS algorithms like [7, 19, 20, 24] are designed for MAB problems with Bernoulli arms and thus cannot be directly applied to our problem which has weighted multinomial distribution. For our case, we adapt the recently proposed Multinomial TS (MTS) [32] which can deal with the multinomial arms. However, in our case, there are multiple differences for which appropriate adaptations are necessary.

1) First, in **Algorithm 1** we design weighted MTS (WMTS) that handles the multinomial rewards $\{\mathbf{r}^T\mathbf{p}_k\}$ weighted by the coefficients $\{C_k^{\text{eff}}\}$. A similar weighted generalization is done for Bernoulli rewards in [19].

2) Second, when the weights are time varying and stochastic, i.e. $\{C_k^{\text{eff}}(t)\}$ are i.i.d. vectors with mean $\{E[C_k^{\text{eff}}]\}$, we design **Algorithm 2**, general MTS (GMTS), which modulates the prior update with observations $\{C_k^{\text{eff}}(t)\}$ after codebook selection.

3) In **Algorithm 1** and **2**, we have not incorporated the structural assumptions, i.e. **Assumption 2** and **3**, into our designs. We next design **Algorithm 3**, constrained WMTS (CWMTS), that is based on [20] which can incorporate either **Assumption 2** or **3** or both.

4) Even though CWMTS can handle general constraints, its implementation has high complexity due to the posterior sampling from a constrained set. In order to move to a more practical algorithm under **Assumption 3** (unimodality of the rewards), we propose unimodal WMTS (UWMTS) in **Algorithm 4**. This algorithm carefully combines the techniques in [32] to handle multinomial rewards, with the leader-tracking based procedure of [30, 35] to present the improved regret guarantees.

In all the above settings, we provide theoretical guarantees on the upper bounds of the cumulative regrets.

4.1 Notations

We present the following notations for later use in this section: $\boldsymbol{\mu} = [\mu_1, \ldots, \mu_K]^{\mathrm{T}}$, $\boldsymbol{\alpha}_k = [\alpha_{0,k}, \ldots, \alpha_{M,k}]^{\mathrm{T}}$ and $\mathbf{1}_M$ denotes a vector of M ones. $Dir\,(\boldsymbol{\alpha}_k)$ denotes the Dirichlet distribution with parameter vector $\boldsymbol{\alpha}_k$. We use $Bernoulli(p)$ to represent a Bernoulli pmf with success probability of p. We use $\mathrm{KL}\,(\mathbf{p}, \mathbf{g})$ to represent the Kullback-Leibler divergence between two one-trial multinomial distributions parameterized by probability vector $\mathbf{p}$ and $\mathbf{g}$, i.e. two categorical distribution, and we define that $\mathcal{K}_{\inf}\,(\mathbf{p}, \mu|\mathbf{s}) = \inf\left\{\mathrm{KL}(\mathbf{p}, \mathbf{g})\middle|\mathbf{s}^{\mathrm{T}}\mathbf{g} > \mu\right.$. We use scalar a_k to represent the k-th element of a vector which is denoted by a bold font $\mathbf{a}$, where k could start with 0 or 1, depending on the context. We denote $\mathcal{P}$ as a problem parameter set that contains all information of our codebook selection problem, i.e. $\mathcal{P} = \left\{\mathbf{r}, \mathbf{p}_k, C_k^{\text{eff}}, \forall k \in [K]^+\right\}$.

4.2 Algorithm without prior knowledge of structural properties

In this subsection, we propose the Weighted Multinomial Thompson Sampling (WMTS) algorithm, which does not require any prior knowledge of the structure among the performance of arms. We maintain K Dirichlet priors, which are conjugate priors for the multinomial reward distributions $\{\mathbf{p}_k\}_{k=1}^K$, for the K arms individually. The details of WMTS is given in **Algorithm 1**. The term *Weighted* emphasizes that different effective coefficient C_k^{eff} scales the support of each arm differently. The performance guarantee of WMTS is given by the following **Theorem** 7.

Theorem 7. *For the codebook selection problem with the access to* $\{C_k^{\text{eff}}\}_{k=1}^K$*, WMTS has the following problem-dependent regret bound*

Algorithm 1 Weighted Multinomial Thompson Sampling

1: Input: Horizon $T \geq 1$, number of codebooks $K \geq 1$, number of non-zero MCSs $M \geq 1$, effective coefficients $\left\{C_k^{\text{eff}}\right\}_{k=1}^{K}$, normalized rate vector $\mathbf{r} = [r_0, r_1, \ldots, r_M]^{\mathrm{T}}$.
2: Initialize: $\alpha_{m,k} = 1$ for $\forall m \in [M]$ and $k \in [K]^+$.
3: **for** $t = 1, \ldots, T$ **do**
4: **for** $k = 1, \ldots, K$ **do**
5: Sample $\mathbf{d}_k(t) \sim Dir\left(\boldsymbol{\alpha}_k\right)$.
6: **end for**
7: $I(t) = \arg\max_{k \in [K]^+} C_k^{\text{eff}} \mathbf{r}^{\mathrm{T}} \mathbf{d}_k(t)$.
8: Select $I(t)$-th codebook to perform the beam alignment and collect RSS feedback.
9: Lookup the RSS-MCS table and obtain the maximum admissible rate for data transmission phase, yielding that $r(t) = r_{m(t)}$ and $m(t) \in [M]$.
10: Prior update: $\alpha_{m(t),I(t)} := \alpha_{m(t),I(t)} + 1$.
11: **end for**

for any $\epsilon_0 > 0$:

$$R(T) \leq \sum_{k=1, k \neq k^*}^{K} \frac{(1+\epsilon_0)\left(\mu_{k^*} - \mu_k\right)}{\mathcal{K}_{\inf}\left(\mathbf{p}_k, \mu_{k^*} \middle| C_k^{\text{eff}} \mathbf{r}\right)} \log T + W(\mathcal{P}, \epsilon_0), \quad (10)$$

where $W(\mathcal{P}, \epsilon_0)$ is a problem-dependent constant that does not depend on T.

Proof. The proof directly follows [32] by generalizing it to that different arms can have different supports for their respective multinomial distributions. □

4.2.1 Discussion of further generalization. In this part, we briefly discuss a further generalization of **Algorithm 1** when $\{C_k^{\text{eff}}\}_{k=1}^{K}$ are inaccessible. For the codebook-based beam training adopted in our studied system, T_k^{train} can be easily calculated, as detailed in the evaluation section. However, if other specially designed beam alignment algorithms were used, e.g. an algorithm that terminates with a good enough beam (see the Section of related work for more examples), T_k^{train} could be random variables whose realizations are only accessible after completing the beam alignment. This is indeed an example of generalizations of our proposed MAB framework. Motivated by this, we also derive General Multinomial Thompson Sampling (GMTS) algorithm, which is denoted as **Algorithm 2** (the detailed algorithm description is omitted due to space limitation). The key step in GMTS is to randomize the reward of arm after observing the sample-path-dependent $C_k^{\text{eff}}(t)$, where $k = I(t)$. To be specific, we generate a Bernoulli random variable X with parameter $C_k^{\text{eff}}(t)$, namely $X \sim Bernoulli(C_k^{\text{eff}}(t))$. If X is zero, then we randomize the reward to be zero, i.e. $m(t) = 0$.

The performance comparison between WMTS and GMTS is shown in the evaluation results. The performance guarantee of GMTS is given by the following **Theorem** 8.

Theorem 8. *For a general codebook selection problem without the access to the sample-path-dependent $\{C_k^{\text{eff}}(t)\}_{k=1}^{K}$, GMTS has the following problem-dependent regret bound for any $\epsilon_0 > 0$:*

$$R(T) \leq \sum_{k=1, k \neq k^*}^{K} \frac{(1+\epsilon_0)\left(\tilde{\mu}_{k^*} - \tilde{\mu}_k\right)}{\mathcal{K}_{\inf}\left(\tilde{\mathbf{p}}_k, \mu^* \middle| \mathbf{r}\right)} \log T + W(\tilde{\mathcal{P}}, \epsilon_0), \quad (11)$$

where $W(\tilde{\mathcal{P}}, \epsilon_0)$ is a problem-dependent constant that does not depend on T, $\tilde{\mathcal{P}} = \left\{\mathbf{r}, \mathbf{p}_k, \mathbf{E}[C_k^{\text{eff}}(t)], \forall k \in [K]^+\right\}$, $\tilde{\boldsymbol{\mu}} = [\tilde{\mu}_1, \ldots, \tilde{\mu}_K]^T$, $\tilde{\mu}_k = \mathbf{r}^T \tilde{\mathbf{p}}_k$, $\tilde{\mathbf{p}}_{m,k} = \mathbf{p}_{m,k} \mathbf{E}[C_k^{\text{eff}}(t)]$ for $m \in [M]^+$, $\tilde{\mathbf{p}}_{0,k} = 1 - \sum_{m=1}^{M} \tilde{\mathbf{p}}_{m,k}$ and $k^* = \arg_{k \in [K]^+} \max \tilde{\mu}_k$.

Proof. With the above described randomization, all the arms follow their own multinomial distribution with a *transformed parameter* $\tilde{\mathbf{p}}_k$ but a *common support* $\mathbf{r}$. We can then directly apply **Theorem** 7 to get the regret bound given in (11). □

4.3 Algorithm using general structural properties

In this subsection, we propose the Constrained Weighted Multinomial Thompson Sampling (CWMTS) algorithm, which leverages the prior knowledge of structural properties among codebooks summarized in Section 3.4. CWMTS is indeed an extension of WMTS, which is inspired by the constrained *Bernoulli* Thompson Sampling (CoTS) proposed in [20]. Its procedure is summarized as follows.

Instead of sampling $\mathbf{D}(t) \triangleq \{\mathbf{d}_1(t), \ldots, \mathbf{d}_K(t)\}$ from the product of those K independent Dirichlet priors, we sample $\mathbf{D}(t)$ in the following way:

$$\mathbf{D}(t) \propto \mathbf{1}\left\{\mathbf{D}(t) \in \boldsymbol{\Phi}\right\} \prod_{k=1}^{K} Dir\left(\boldsymbol{\alpha}_k\right)\left(\mathbf{d}_k(t)\right), \quad (12)$$

where $\boldsymbol{\Phi}$ denotes the *parameter space* that is the set of all possible estimates of $\{\mathbf{p}_k\}_{k=1}^{K}$, and $Dir\left(\boldsymbol{\alpha}_k\right)\left(\mathbf{d}_k(t)\right)$ is the probability density function (PDF) of $Dir\left(\boldsymbol{\alpha}_k\right)$ for $\mathbf{d}_k(t)$. In particular, by omitting the time index t and denoting $\mathbf{D} \triangleq \{\mathbf{d}_1, \ldots, \mathbf{d}_k\}$, under **Assumption** 2, we have

$$\boldsymbol{\Phi} \triangleq \left\{\mathbf{D} \middle| \mathbf{r}^{\mathrm{T}} \mathbf{d}_1 \leq \mathbf{r}^{\mathrm{T}} \mathbf{d}_2 \leq \ldots \leq \mathbf{r}^{\mathrm{T}} \mathbf{d}_K\right\}, \quad (13)$$

and under **Assumption** 3, we have

$$\boldsymbol{\Phi} \triangleq \left\{\mathbf{D} \middle| C_1^{\text{eff}} \mathbf{r}^{\mathrm{T}} \mathbf{d}_1 \leq \ldots \leq C_{k^*}^{\text{eff}} \mathbf{r}^{\mathrm{T}} \mathbf{d}_{k^*} \geq \ldots \geq C_K^{\text{eff}} \mathbf{r}^{\mathrm{T}} \mathbf{d}_K\right\}. \quad (14)$$

Given that $I(t)$-th codebook is used and the observed reward is $r(t) = r_{m(t)}$, the prior of $\mathbf{D}(t+1)$ after Bayesian update is

$$\mathbf{D}(t+1) \propto \mathbf{1}\left\{\mathbf{D}(t) \in \boldsymbol{\Phi}\right\} \times \prod_{k=1, k \neq I(t)}^{K} Dir\left(\boldsymbol{\alpha}_k\right)\left(\mathbf{d}_k\right) \times Dir\left(\boldsymbol{\alpha}_{I(t)} + \mathbf{e}_{m(t)}\right)\left(\mathbf{d}_{I(t)}\right), \quad (15)$$

where $\mathbf{e}_{m(t)}$ is a unit vector where the $m(t)$-th element is one. (15) shows that the update rules of priors is the same as that in the WMTS algorithm but we control the estimation of the distributions of arms in a more specific parameter space. We summarize CWMTS in **Algorithm** 3.

Before stating the theoretical regret bound of the CWMTS algorithm, we present the following notations. We denote $\mathcal{A}$ as the action space, namely that $\mathcal{A} = [K]^+$ as we have K codebooks. We denote $\mathcal{Y}$ as the observation space, i.e. the possible values of reward. Then we have $\mathcal{Y} = \left\{r_m C_k^{\text{eff}}, k \in [K]^+, m \in [M]\right\}$. We denote π_t as the Dirichlet prior used in the t-th time slot, and denote π_0 is the initial prior, i.e. $Dir\left(\mathbf{1}_{M+1}\right)$, as initialized in line 2 of **Algorithm** 3. In addition, we make one following assumption:

Algorithm 3 Constrained Weighted Multinomial TS

Input: Horizon $T \geq 1$, number of codebooks $K \geq 1$, number of non-zero MCSs $M \geq 1$, effective coefficients $\left\{C_k^{\text{eff}}\right\}_{k=1}^{K}$, normalized rate vector $\mathbf{r} = [r_0, r_1, \ldots, r_M]^{\mathrm{T}}$.
Initialize: $\alpha_{m,k} = 1$ for $\forall m \in [M]$ and $k \in [K]^+$.
for $t = 1, \ldots, T$ **do**
  Sample $\mathbf{D}(\mathbf{t}) \sim \mathbf{1}\{\mathbf{D} \in \mathbf{\Phi}\} \prod_{k=1}^{K} Dir(\boldsymbol{\alpha}_k)(\mathbf{d}_k(t))$.
  $I(t) = \arg\max_{k \in [K]^+} C_k^{\text{eff}} \mathbf{r}^{\mathrm{T}} \mathbf{d}_k(t)$.
  Select $I(t)$-th codebook to perform the beam alignment and collect RSS feedback.
  Lookup the RSS-MCS table and obtain the maximum admissible rate for data transmission phase, yielding that $r(t) = r_{m(t)}$ and $m(t) \in [M]$.
  Prior update: $\alpha_{m(t),I(t)} := \alpha_{m(t),I(t)} + 1$.
end for

Assumption 9. *(Unique optimal codebook) The optimal codebook is assumed to be unique, i.e., $\mu_{k^*} > \mu_k, \forall k \neq k^*$.*

With the above notation and **Assumption** 9, the following Theorem now holds.

Theorem 10. *Suppose that **Assumption** 9 holds, then a regret bound for the CWMTS algorithm is given as follows: For any $\epsilon, \delta \in (0,1)$, there exists $T^* \geq 0$ such that for all time horizon $T \geq T^*$, with probability at least $1-\delta$, CWMTS has the following problem-dependent regret bound:*

$$R(T) \leq \left(\mu_{k^*} - \min_{k \in [K]^+} \mu_k\right)\left(\frac{1+\epsilon}{1-\epsilon}\right) \sum_{k=1, k \neq k^*}^{K} \frac{\log T}{\mathcal{K}_{\inf}\left(\mathbf{p}_k, \mu_{k^*} \middle| C_k^{\text{eff}} \mathbf{r}\right)} + E(\epsilon, \delta, \mathcal{A}, \mathcal{Y}, \mathbf{\Phi}, \pi_0), \tag{16}$$

where $E(\epsilon, \delta, \mathcal{A}, \mathcal{Y}, \mathbf{\Phi}, \pi_0)$ is a problem-dependent constant that does not depend on T.

Proof. The proof immediately follows (with minor changes to account for multinomial instead of Bernoulli random variables) from [18, 20]. □

The above theorem shows that the regret associated with CWMTS also scales logarithmically with time as WMTS and GMTS do. **Assumption** 9 is made only for notational ease in the proof and it does not significantly affect the result given in **Theorem** 10, as pointed out in [18].

4.3.1 Discussion on the limitation of CWMTS. The straightforward way to implement CWMTS is to use rejection sampling, namely that we sample $\mathbf{D}$ from $\prod_{k=1}^{K} Dir(\boldsymbol{\alpha}_k)$ until $\mathbf{D} \in \mathbf{\Phi}$. As the authors note in [20], a disadvantage of this approach is that it can be slow when the probability of getting a valid $\mathbf{D}$ is small. In [20], the authors proposed a heuristic Sequential Inverse Transform Sampling (SITS) approach by sampling $\mathbf{d}_k$ sequentially with individual constraint $\mathbf{r}^{\mathrm{T}}\mathbf{d}_k \leq \mathbf{r}^{\mathrm{T}}\mathbf{d}_{k+1}$. Note however that $\mathbf{d}_k$ are correlated with each other; thus while the heuristic SITS returns a valid sample in $\mathbf{\Phi}$, it may not be from the correct distribution. Thereby, designing an efficient implementation of CWMTS (that results in samples from the correct distribution) is also an interesting future direction.

4.4 Unimodal Thompson Sampling

In this part, we present a novel algorithm exploiting the property that the effective data rates have a unimodal pattern, as stated in **Assumption** 3. We term it as Unimodal Weighted Multinomial Thompson Sampling (UWMTS). This is a novel combination of the Multinomial TS [32] and the Unimodal Bernoulli TS [30, 35]. The key element of this combination would be highlighted later.

To explain UWMTS, we set the following notations. We denote $N_k(t) \triangleq \sum_{i=1}^{t} \mathbf{1}\{I(i) = k\}$ as the number of times that k-th codebook is used up to t-th time slot, and the estimated expected reward of the k-th codebook as $\hat{\mu}_k(t) \triangleq \frac{\sum_{i=1}^{t} \mathbf{1}\{I(i)=k\} r(i) C_k^{\text{eff}}}{N_k(t)}$. In particular, we define an empirical leader $L(t) = \arg_{k \in [K]^+} \hat{\mu}_k(t)$ and the number of times arm k has been leader up to time t as $l_k(t) = \sum_{i=1}^{t} \mathbf{1}\{l(i) = k\}$.

The core of UWMTS is to restrict WMTS to the neighborhood of the leader and meanwhile add a leader exploration mechanism to detect the optimal arm with high probability. Specifically, UWMTS chooses the arm at time t by following policy:

$$I(t) = \begin{cases} L(t) & Mod\left(l_{L(t)}(t), \gamma\right) = 0, \\ \text{Run WMTS in } \mathcal{N}_{L(t)}^{+}, & \text{otherwise}, \end{cases} \tag{17}$$

where *Mod* is the modulo function, γ is the frequency that the leader is exploited, $\mathcal{N}_k^+ = \mathcal{N}_k \cup \{k\}$ with that $\mathcal{N}_k$ is the set of neighboring arms of arm k, i.e. $\mathcal{N}_k = \{k-1, k+1\} \cap [K]^+$ in our case. It is worth pointing out that there is no leader exploration when $\gamma = \infty$ and there is no theoretical guide on how to choose its value. It is empirically found by our simulation and [35] that choosing a smaller value ($2 \leq \gamma \leq K$) results in a relatively good performance. The description of UWMTS is given in **Algorithm** 4.

UTS was proposed with *Bernoulli arms* and unimodal reward structure in [30], and it is proved to have asymptotically optimal regret in [35]. We adapt the framework in [35] and generalize the proofs therein from *Bernoulli arms* to *multinomial arms*. Such generalization, even in standard MAB (see, [32]), is known to be non-trivial as connecting the posterior of the reward (which follows Dirichlet distribution), to the observed rewards (which follows multinomial distribution) is difficult due to the absence of a closed form expression, unlike the Bernoulli case where the Beta-Binomial transform is used [6]. We leverage the tail bounds of Dirichlet distribution in [32], and derive the posterior concentration for the arms in the neighborhood of the optimal arm, which in our case includes two suboptimal arms and the optimal arm due to unimodality. This allows us to show each of these two suboptimal arm is played $O(\log(T))$ times in expectation, where the constant associated with the $\log(T)$ term is asymptotically optimal. Similar to [35], the other $(K-3)$ suboptimal arms are shown to be rarely played, i.e. $O(1)$ times in expectation, as the leader election method concentrates fast. Thus, we provide the first regret upper bound for UTS with multinomial arms, summarized in **Theorem** 11 below:

Theorem 11. *For codebook selection problem with the access to $\{C_k^{\text{eff}}\}_{k=1}^{K}$, under **Assumption 3**, UWMTS has the following problem-dependent regret bound for any $\gamma \geq 2$ and any $\epsilon_0 > 0$:*

$$R(T) \leq \sum_{k \in \mathcal{N}_{k^*}} \frac{(1+\epsilon_0)(\mu_{k^*} - \mu_k)}{\mathcal{K}_{\inf}\left(\mathbf{p}_k, \mu^* \middle| C_k^{\text{eff}} \mathbf{r}\right)} \log T + U(\mathcal{P}, \epsilon_0, \gamma), \tag{18}$$

where $U(\mathcal{P}, \epsilon_0, \gamma)$ is a constant that does not depend on T.

Proof. See Appendix A in the full version of this work [40]. □

Remark 12. *We note that UWMTS can significantly reduce the regret as the coefficient of logarithmic term is restricted to the neighborhood of the optimal arm, i.e. $\mathcal{N}_{k^*}$ with $|\mathcal{N}_{k^*}| \leq 2$. This reduces the regret from $O(K \log T)$ to $O(2 \log T)$.*

Algorithm 4 Unimodal Weighted Multinomial TS

Input: Horizon $T \geq 1$, number of codebooks $K \geq 1$, number of non-zero MCSs $M \geq 1$, effective coefficients $\left\{C_k^{\text{eff}}\right\}_{k=1}^{K}$, normalized rate vector $\mathbf{r} = [r_0, r_1, \ldots, r_M]^{\mathrm{T}}$, and leader exploration parameter γ.
Initialize: $\alpha_{m,k} = 1$, $\hat{\mu}_k(t) = 0$, $N_k(t)=0$, $l_k(t) = 0$ for $\forall m \in [M]$ and $k \in [K]^+$. We omit time index t of $\hat{\mu}_k(t)$, $N_k(t)$, $l_k(t)$ in the following.
for $t = 1, \ldots, K$ **do**
  $I(t) = t$.
  Select $I(t)$-th codebook to perform the beam alignment and collect RSS feedback.
  Lookup the RSS-MCS table and obtain the maximum admissible rate for data transmission phase, yielding that $r(t) = r_{m(t)}$ and $m(t) \in [M]$.
  Prior update: $\alpha_{m(t),I(t)} := \alpha_{m(t),I(t)} + 1$.
  Mean update: $\hat{\mu}_{I(t)} := \frac{\hat{\mu}_{I(t)} N_{I(t)} + r(t) C_k^{\text{eff}}}{N_{I(t)}+1}$.
  Arm counter update: $N_{I(t)} := N_{I(t)} + 1$.
end for
for $t = K+1, \ldots, T$ **do**
  $L(t) = \arg\max_{k \in [K]^+} \hat{\mu}_k$.
  Leader counter update: $l_{L(t)} := l_{L(t)} + 1$.
  if $Mod(l_{L(t)}, \gamma) == 0$ **then**
    $I(t) = L(t)$.
  else
    for $k \in \mathcal{N}_{L(t)}^+$ **do**
      Sample $\mathbf{d}_k \sim Dir(\boldsymbol{\alpha}_k)$.
    end for
    $I(t) = \arg\max_{k \in \mathcal{N}_{L(t)}^+} C_k^{\text{eff}} \mathbf{r}^{\mathrm{T}} \mathbf{d}_k$.
  end if
  Select $I(t)$-th codebook to perform the beam alignment and collect RSS feedback.
  Lookup the RSS-MCS table and obtain the maximum admissible rate for data transmission phase, yielding that $r(t) = r_{m(t)}$ and $m(t) \in [M]$.
  Prior update: $\alpha_{m(t),I(t)} := \alpha_{m(t),I(t)} + 1$.
  Mean update: $\hat{\mu}_{I(t)} := \frac{\hat{\mu}_{I(t)} N_{I(t)} + r(t) C_k^{\text{eff}}}{N_{I(t)}+1}$.
  Arm counter update: $N_{I(t)} := N_{I(t)} + 1$.
end for

5 EVALUATION RESULTS

In this section, we evaluate the proposed algorithms in comparison with the following state-of-the-art bandit algorithms: (1) Bernoulli Thompson Sampling (BTS) [24]: we randomize the codebook rewards to be Bernoulli random variables such that this primitive TS algorithm is applicable. (2) Weighted Bernoulli Thompson Sampling (WBTS) [19]: a modified version of BTS. (3) KL-UCB [15]: as the reward of arms are bounded by $[0, 1]$, the classic KL-UCB can be directly applied. (4) Optimal Sampling Unimodal Bandit (OSBU) [12]: OSUB is developed based on KL-UCB by further adding the leader mechanism to exploit the structural property that the rewards are unimodal. (5) Unimodal Weighted Bernoulli Thompson Sampling (UWBTS) [35]: UWBTS is a straightforward extension of WBTS by using the structural property that the rewards are unimodal.

In the following, we perform a trace-driven simulation. The simulated system adopts IEEE 802.11ad Standard, with carrier frequency of $f_c = 60$ GHz and with a bandwidth of $B = 1.76$ GHz [3, 38]. We incorporate the **real-world channel measurements**, captured at 60 GHz and in terms of SNR, into the simulated system.

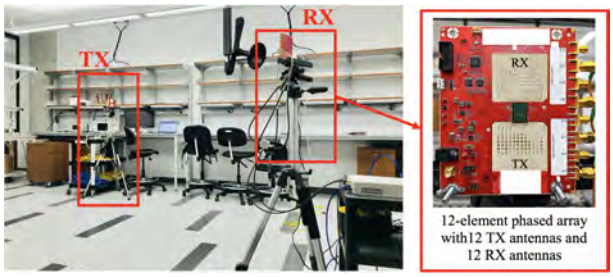

Figure 2: Experimental setup

5.1 System parameters

In this part, we summarize the system parameters for the simulation. The duration of testing each beam T^{mer} is 17 μs [3] and the duration per time slot T_{slot} is set as 50 ms. We adopt the RSS-MCS table provided by IEEE 802.11ad Standards for single-carrier transmission mode [3]. Accordingly, the unnormalized rate vector $\tilde{\mathbf{r}}$ is [0, 27.5, 385, 770, 962.5, 1155, 1251.25, 1540, 1925, 2310, 2502.5, 2695, 3080, 3850, 4620, 5005, 5390, 5775, 6390, 7507.5, 8085]$^{\mathrm{T}}$ Mbps and the RSS vector **rss** is [-inf, -78, -68, -66, -65, -64, -63, -62, -61, -60, -59, -57, -55, -54, -53, -51, -50, -48, -46, -44, -42]$^{\mathrm{T}}$ dBm. By considering a noise power level of -78 dBm, we could further compute the corresponding SNR values to get a SNR-MCS table for reference as our collected channel measurements are in terms of SNR.

5.2 Real-world measurement collection

In this part, we present our experimental setup and the collected real-world channel measurements. The testbed used for capturing the SNR measurements consists of two 12-antenna SiBEAM Sil6342 phased arrays that up/down convert the signal to/from 60 GHz, and two N210 USRPs with a bandwidth of 5 MHz, as shown in Figure 2. By controlling the number of activated antennas N_{Ant} and using phased array calibration techniques proposed by [41], we can generate directional beams of different widths. Since our antenna array has only 12 elements, there is no major gain in having too many codebooks (as their resolutions will be too close); thus, we generate 6 representative codebooks as shown in Figure 3.

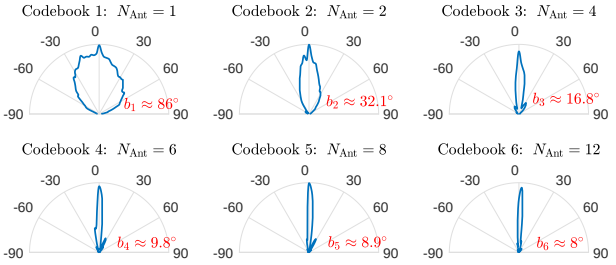

Figure 3: Example beam patterns of the 6 codebooks generated by the SiBEAM Sil6342 60 GHz phased arrays.

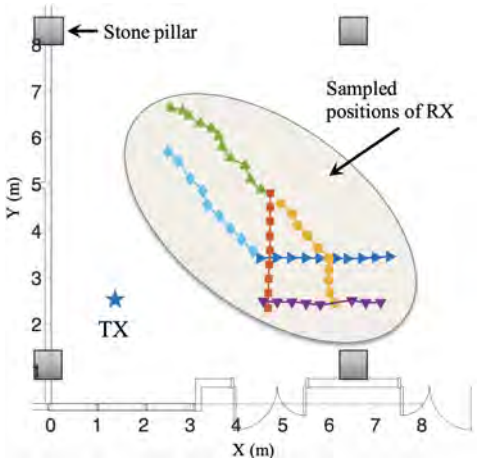

Figure 4: Sketch map of a spacious lab in which the mmWave channel measurements are taken. Different markers form different potential trajectories of RX.

In our evaluation, we consider that $K = 6$ codebooks, given in Figure 3 are available at the TX, and the RX uses the fixed Codebook 6. The size of codebook $\{S_k\}_{k=1}^K$ can be calculated with $S_k = \frac{360}{b_k}$ by considering a 2D beam scanning. Due to the limited bandwidth of USRP and the overhead/challenges of implementing a real-time system with *user mobility*, we use the testbed to measure the SNRs along certain predefined trajectories of RX and interpolate the values SNR with respect to the distance between TX and RX given a target velocity (4 m/s). The sampled positions of the RX are shown in Figure 4. At each position, the SNR is measured 4 times for each codebook at the TX. Implementing a real-time system for performance evaluation would be a promising future direction but out of the scope of this work. For simplicity, we did not collect measurements for non-line-of-sight (NLOS) scenarios since we perform the beam sweeping with directional beams and the NLOS scenarios will simply result in higher path loss, which is handled by our developed MAB framework.

Based on the above setting, we further compute the values of key parameters as follows. The effective coefficients $(C_1^{\text{eff}}, \ldots, C_K^{\text{eff}})$ is computed by $T_k^{\text{train}} = S_k S_K T^{\text{mer}}$ and $C_k^{\text{eff}} = (T^{\text{slot}} - T_k^{\text{train}})/T^{\text{slot}}$, and they are (0.9235, 0.8164, 0.6634, 0.4339, 0.3727, 0.3115). To compute the ground truth distribution $\{\mathbf{p}_k\}_{k=1}^K$, we use the distribution statistics of the interpolated SNRs. We omit the exact values of $\{\mathbf{p}_k\}_{k=1}^K$ due to the space limitation. The expected instantaneous data rate $\left(\mathbf{r}^{\mathrm{T}}\mathbf{p}_1, \ldots, \mathbf{r}^{\mathrm{T}}\mathbf{p}_K\right)$ can be calculated as (0.1397, 0.2940, 0.4390, 0.5879, 0.6626, 0.7507). The eventual expected rewards of the K codebooks $(\mu_1, \ldots, \mu_K)$ are (0.1290, 0.2400, 0.2912, 0.2551, 0.2469, 0.2338). It can be verified that the above setting satisfies both **Assumption 2** and **3**. We run the evaluation for $T = 10000$ time slots and average the results by 100 realizations.

5.3 Discussions on performance comparison

In Figure 5a, we show the performance of the proposed WMTS when there is no prior knowledge of any problem structure. First, it can be seen that WMTS outperforms the state-of-the-art bandit algorithms and has a much smaller cumulative regret. Moreover, WMTS converges much faster than the other algorithms, this implies that our proposed algorithm can provide more flexibility and robustness in non-stationary environments, in which the channel distribution is time-varying. Further, we can observe that GMTS also provides a competitive performance.

In Figure 5b, we present the performance gain achieved by the CWMTS algorithm when the nondecreasing property (i.e. **Assumption 2**) is known to hold. As we can see, CWMTS does not provide a better regret performance than WMTS until 10,000 slots, but it converges much faster than WMTS.

In Figure 5c, we further show the performance of CWMTS and UWMTS ($\gamma = 3$) given that the unimodality property (i.e. **Assumption 3**) is known to hold. Some interesting observations can be drawn: (1) CWMTS outperforms OSUB ($\gamma = 3$) and UWBTS when it uses the property that the rewards have the unimodal pattern. (2) It is clear that UWMTS outperforms all the other algorithms given the unimodality, and the performance improvement is significant, which is consistent with **Remark 12**. (3) All the algorithms using multinomial distribution converge faster than the other algorithms.

Finally, if a random selection policy is adopted (instead of a learning-based policy), the average normalized throughput would remain at $\frac{1}{K}\sum_{k=1}^K \mu_k = 0.2327$. In contrast, our online learning framework can learn the optimal codebook quickly, and the normalized throughput would be almost $\mu_{k^*} = 0.2912$, which implies a throughput improvement by more than 25%.

6 RELATED WORK

(1) Model-driven beamwidth optimization: One of the most related lines of work is beamwidth optimization. In [33], the authors initially modeled and derived the trade-off caused by beamwidth in a multi-user mmWave network. Similar optimizations that balance the beamforming gain and the beam training overhead were also investigated in [13, 23, 26]. However, their solutions heavily depend on the physical layer assumptions or prior knowledge such as channel model, beam pattern model, and network topology, which restricts their flexibility in practical deployments in MANETs where the channel is rapidly changing. In contrast with these prior work, our proposed MAB-based solutions are model-free, and thus do not rely on the assumption of channel or user mobility.

(2) Data-driven codebook construction: Some recent work has used offline data-driven machine learning methodologies to perform beam alignment and beamwidth selection simultaneously.

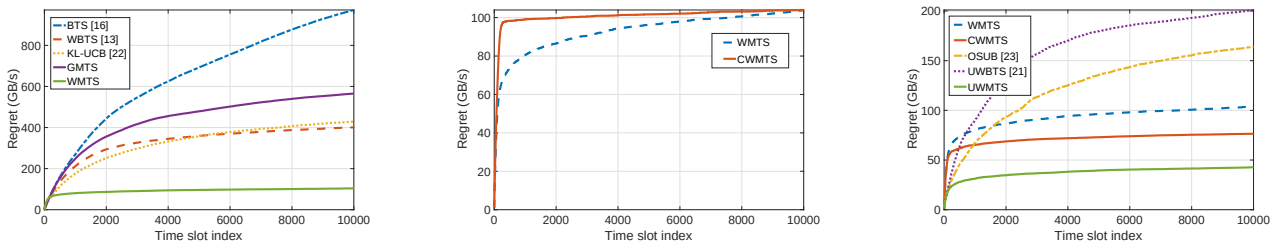

Figure 5: Regret performance based on real-world measurements: (a). No knowledge of problem structure. (b). With problem structure that instantaneous data rates are nondecreasing. (c). With problem structure that effective data rates are unimodal.

In [37], a deep learning technique was exploited to learn an optimal set of beam pairs by considering the environment information as feature spaces. Similarly, in [11], a large amount of experimental data were gathered to build a beamforming codebook of minimum size and subject to a guaranteed gain. Besides, a geo-located context database was built in [14] to assist the beam width/directions selection. [11, 14, 37] all showed that significant system improvement was achieved over conventional beam alignment strategies. These offline data-driven approaches however require a large amount of historical data for a given deployment site, which limits its fast implementation. Further, since they only focused on the successful connection probability of the eventually learned codebook, the trade-off between beam alignment quality and data transmission efficiency was not exploited therein. Finally, no theoretical guarantee of performance was provided.

(3) Beam alignment (including hierarchical search): Other than codebook optimization, much of the prior work focuses on selecting the best beam from a single codebook without considering the effect of beam resolution, for example, [22] proposed *Agile-Link* which finds the best beam by a random hashing and voting mechanism. Some work exploits a priori knowledge of the channel to avoid exhausted beam search [14, 17, 28, 34, 41]. However, prior information would require additional sensors or statistics. Moreover, adaptive approaches were also investigated: *ACO* was proposed in [29] to estimate the full channel, whereas four probes per antenna element are required, which results in poor scalability. Another approach – hierarchical search – starts (in each time slot) from a coarse beam and progressively uses finer beams to shorten the training time [27, 38]. However, it has several drawbacks: limited coverage due to the initial use of wide beams [27]; zooming in wrong directions due to beam imperfectness and interference [22]; and large feedback overhead (per measurement) in asymmetric links where devices have to respond by directional beams due to power limitation [31]. In contrast, we have focused on the mmWave codebook selection by dynamically learning a site-specific or device-specific codebook over time. Indeed, the above algorithms could be also incorporated into our framework by regarding different algorithms (or an algorithm with different parameters) as different "abstract codebooks".

(4) Related bandit algorithms Thompson Sampling (TS) is a widely used method for solving MAB problems. In [24], a regret bound was shown for TS with *Bernoulli arms*. In [19], the weighted binary TS was derived based on [24] to deal with the case where the reward of each Bernoulli arm was multiplied by a different constant. One of the most related work is [32], in which the authors provided the regret bound for TS with *multinomial arms* of the same support. The above algorithms, however, do not exploit structure across arms and satisfy asymptotic optimality for unstructured bandit problems. In [20], the constrained weighted binary TS was proposed to allow incorporating general structural properties among arms. An improved performance was achieved, but an efficient implementation is still lacking (see also Section 4.3.1). To exploit reward unimodality, the OSUB algorithm was proposed in [12] based on KL-UCB. A very recent work [35] derived a theoretical guarantee for UTS with Bernoulli arms. Our proposed algorithms augment these prior studies. We highlight that we provide the first theoretical guarantees for UTS with weighted multinomial rewards.

7 CONCLUSIONS

In this work, we have considered the codebook selection problem in mmWave MANETs with rapidly-varying wireless channels. We have modeled it as a MAB problem and have proposed novel TS-based algorithms with/without knowing the structures among codebooks. We have derived the theoretical regret upper bounds for the proposed algorithms. The real-world mmWave measurements based evaluation has validated the benefits of our algorithms.

ACKNOWLEDGMENTS

This work was partially supported by the U.S. Army Research Labs grant W911NF-19-1-0221, NSF grant CNS-1731658 and the US DoT supported D-STOP Tier 1 University Transportation Center.

REFERENCES

[1] 2009. IEEE Standard for Information technology– Local and metropolitan area networks– Specific requirements– Part 15.3: Amendment 2: Millimeter-wave-based Alternative Physical Layer Extension. *IEEE Std 802.15.3c-2009 (Amendment to IEEE Std 802.15.3-2003)* (Oct. 2009), 1–200.

[2] 2014. *WP5: Propagation, Antenna, and Multi-Antenna Techniques: D5.1 - Channel Modeling and Characterization.* Technical Report. EU and Japanese Government.

[3] 2016. IEEE Standard for Information technology–Telecommunications and information exchange between systems Local and metropolitan area networks–Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. *IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012)* (Dec. 2016), 1–3534.

[4] 2019. IEEE Draft Standard for Information Technology–Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks–Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications–Amendment: Enhanced Throughput for

Operation in License-Exempt Bands Above 45 GHz. *IEEE P802.11ay/D4.0, June 2019* (Jul. 2019), 1–791.

[5] 2019. System Architecture for the 5G System. *document TS 23.501 V16.1.0, 3GPP, Jun. 2019* (Jun. 2019), 1–219.

[6] Shipra Agrawal and Navin Goyal. 2012. Analysis of Thompson Sampling for the Multi-armed Bandit Problem. In *Proc. of the 25th Annual Conference on Learning Theory (COLT'12)*. Edinburgh, Scotland, 39.1–39.26.

[7] Shipra Agrawal and Navin Goyal. 2013. Further Optimal Regret Bounds for Thompson Sampling. In *Proc. of the Sixteenth International Conference on Artificial Intelligence and Statistics (AISTATS'13)*. Scottsdale, Arizona, USA, 99–107.

[8] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. 2002. Finite-Time Analysis of the Multiarmed Bandit Problem. *Machine Learning* 47, 2-3 (May 2002), 235–256.

[9] Irmak Aykin, Berk Akgun, Mingjie Feng, and Marwan Krunz. 2020. MAMBA: A Multi-armed Bandit Framework for Beam Tracking in Millimeter-wave Systems. In *Proc. of 2020 IEEE International Conference on Computer Communications (INFOCOM 2020)*. Shanghai, China, 1469–1478.

[10] Olivier Chapelle and Lihong Li. 2011. An Empirical Evaluation of Thompson Sampling. In *Proc. of the 24th International Conference on Neural Information Processing Systems (NeurIPS'11)*. Granada, Spain, 2249–2257.

[11] Mohaned Chraiti, Dmitry Chizhik, Jinfeng Du, Reinaldo A. Valenzuela, Ali Ghrayeb, and Chadi Assi. 2019. Beamforming Learning for mmWave Communication: Theory and Experimental Validation. arXiv ePrint 1912.12406.

[12] Richard Combes and Alexandre Proutiere. 2014. Unimodal Bandits: Regret Lower Bounds and Optimal Algorithms. In *Proc. of the 31st International Conference on Machine Learning (ICML '14)*. Beijing, China, 521–529.

[13] Jiancun Fan, Liyuan Han, Xinmin Luo, Ying Zhang, and Jingon Joung. 2020. Beamwidth Design for Beam Scanning in Millimeter-Wave Cellular Networks. *IEEE Transactions on Vehicular Technology* 69, 1 (Jan. 2020), 1111–1116.

[14] Ilario Filippini, Vincenzo Sciancalepore, Francesco Devoti, and Antonio Capone. 2018. Fast Cell Discovery in Mm-Wave 5G Networks With Context Information. *IEEE Transactions on Mobile Computing* 17, 7 (Jul. 2018), 1538–1552.

[15] Aurélien Garivier and Olivier Cappé. 2011. The KL-UCB Algorithm for Bounded Stochastic Bandits and Beyond. In *Proc. of the 24th Annual Conference on Learning Theory (COLT '11)*. Budapest, Hungary, 359–376.

[16] Yasaman Ghasempour, Muhammad K. Haider, Carlos Cordeiro, Dimitrios Koutsonikolas, and Edward Knightly. 2018. Multi-Stream Beam-Training for MmWave MIMO Networks. In *Proc. of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom '18)*. New Delhi, India, 225–239.

[17] Nuria González-Prelcic, Anum Ali, Vutha Va, and Robert W. Heath. 2017. Millimeter-Wave Communication with Out-of-Band Information. *IEEE Commun. Mag.* 55, 12 (Dec. 2017), 140–146.

[18] Aditya Gopalan, Shie Mannor, and Yishay Mansour. 2014. Thompson Sampling for Complex Online Problems. In *Proc. of the 31st International Conference on on Machine Learning (ICML '14)*. Beijing, China, 100–108.

[19] Harsh Gupta, Atilla Eryilmaz, and R. Srikant. 2018. Low-Complexity, Low-Regret Link Rate Selection in Rapidly-Varying Wireless Channels. In *Proc. of 2018 IEEE International Conference on Computer Communications (INFOCOM 2018)*. Honolulu, HI, USA, 540–548.

[20] Harsh Gupta, Atilla Eryilmaz, and R. Srikant. 2019. Link Rate Selection using Constrained Thompson Sampling. In *Proc. of 2019 IEEE International Conference on Computer Communications (INFOCOM 2019)*. Paris, France, 739–747.

[21] Morteza Hashemi, Ashutosh Sabharwal, C. Emre Koksal, and Ness B. Shroff. 2018. Efficient Beam Alignment in Millimeter Wave Systems Using Contextual Bandits. In *Proc. of 2018 IEEE International Conference on Computer Communications (INFOCOM 2018)*. Honolulu, HI, USA, 2393–2401.

[22] Haitham Hassanieh, Omid Abari, Michael Rodriguez, Mohammed Abdelghany, Dina Katabi, and Piotr Indyk. 2018. Fast Millimeter Wave Beam Alignment. In *Proc. of the 2018 Conference of the ACM Special Interest Group on Data Communication (SIGCOMM '18)*. Budapest, Hungary, 432–445.

[23] Kishor Chandra Joshi, Solmaz Niknam, R. Venkatesha Prasad, and Balasubramaniam Natarajan. 2020. Analyzing the Tradeoffs in Using Millimeter Wave Directional Links for High Data-Rate Tactile Internet Applications. *IEEE Transactions on Industrial Informatics* 16, 3 (Mar. 2020), 1924–1932.

[24] Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. 2012. Thompson Sampling: An Asymptotically Optimal Finite-Time Analysis. In *Proc. of the 23rd International Conference on Algorithmic Learning Theory (ALT '12)*. Lyon, France, 199–213.

[25] Oteri Kome, Lin Cen, Lou Hanqing, and Yang Rui. 2016. *Further Details on Multi-Stage, Multi-Resolution Beamforming Training in 802.11ay, doc.: IEEE 802.11-16/1447r1.* Retrieved Dec. 12, 2020 from https://mentor.ieee.org/802.11/dcn/16/11-16-1447-01-00ay-further-details-on-multi-stage-multi-resolution-beamforming-training-in-802-11ay.pptx

[26] Jia Liu and Elizabeth S. Bentley. 2019. Hybrid-Beamforming-Based Millimeter-Wave Cellular Network Optimization. *IEEE Journal on Selected Areas in Communications* 37, 12 (Dec. 2019), 2799–2813.

[27] Giordani Marco, Mezzavilla Marco, and Zorzi Michele. 2016. Initial Access in 5G mmWave Cellular Networks. *IEEE Commun. Mag.* 54, 11 (Nov. 2016), 40–47.

[28] Thomas Nitsche, Adriana B. Flores, Edward W. Knightly, and Joerg Widmer. 2015. Steering With Eyes Closed: Mm-Wave Beam Steering Without In-Band Measurement. In *Proc. of 2015 IEEE International Conference on Computer Communications (INFOCOM 2015)*. Kowloon, Hong Kong, China, 2416–2424.

[29] Joan Palacios, Daniel Steinmetzer, Adrian Loch, Matthias Hollick, and Joerg Widmer. 2018. Adaptive Codebook Optimization for Beam Training on Off-the-Shelf IEEE 802.11Ad Devices. In *Proc. of the 24th Annual International Conference on Mobile Computing and Networking (MobiCom '18)*. New Delhi, India, 241–255.

[30] Stefano Paladino, Francesco Trovò, Marcello Restelli, and Nicola Gatti. 2017. Unimodal Thompson Sampling for Graph-Structured Arms. In *Proc. of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI '17)*. San Francisco, CA, USA, 2457–2463.

[31] Zhou Pei, Cheng Kaijun, Han Xiao, Fang Xuming, Fang Yuguang, He Rong, Long Yan, and Liu Yanping. 2018. IEEE 802.11ay-Based mmWave WLANs: Design Challenges and Solutions. *IEEE Communications Surveys & Tutorials* 20, 3 (Thirdquarter 2018), 1654–1681.

[32] Charles Riou and Junya Honda. 2020. Bandit Algorithms Based on Thompson Sampling for Bounded Reward Distributions. In *Proc. of the 31st International Conference on Algorithmic Learning Theory (ALT '20)*, Vol. 117. San Diego, CA, USA, 777–826.

[33] Hossein Shokri-Ghadikolaei, Lazaros Gkatzikis, and Carlo Fischione. 2015. Beam-searching and Transmission Scheduling in Millimeter Wave Communications. In *2015 IEEE International Conference on Communications*. London, UK, 1292–1297.

[34] Gek Hong Sim, Sabrina Klos, Arash Asadi, Anja Klein, and Matthias Hollick. 2018. An Online Context-Aware Machine Learning Algorithm for 5G mmWave Vehicular Communications. *IEEE/ACM Transactions on Networking* 26, 6 (Dec. 2018), 2487–2500.

[35] Cindy Trinh, Emilie Kaufmann, Claire Vernade, and Richard Combes. 2020. Solving Bernoulli Rank-One Bandits with Unimodal Thompson Sampling. In *Proc. of the 31st International Conference on Algorithmic Learning Theory (ALT '20)*, Vol. 117. San Diego, CA, USA, 862–889.

[36] Song Wang, Jingqi Huang, and Xinyu Zhang. 2020. Demystifying Millimeter-Wave V2X: Towards Robust and Efficient Directional Connectivity under High Mobility. In *Proc. of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom '20)*. London, United Kingdom, Article 51, 14 pages.

[37] Yuyang Wang, Aldebaro Klautau, Monica Ribero, Anthony C. K. Soong, and Robert. W. Heath. 2019. MmWave Vehicular Beam Selection With Situational Awareness Using Machine Learning. *IEEE Access* 7 (2019), 87479–87493.

[38] Wen Wu, Nan Cheng, Ning Zhang, Peng Yang, Weihua Zhuang, and Xuemin Shen. 2019. Fast mmwave Beam Alignment via Correlated Bandit Learning. *IEEE Transactions on Wireless Communications* 18, 12 (Dec. 2019), 5894–5908.

[39] Zhenyu Xiao, Pengfei Xia, and Xiang-Gen Xia. 2017. Codebook Design for Millimeter-Wave Channel Estimation With Hybrid Precoding Structure. *IEEE Transactions on Wireless Communications* 16, 1 (Jan. 2017), 141–153.

[40] Yi Zhang, Soumya Basu, Sanjay Shakkottai, and Robert W. Heath Jr. 2020. Supplementary Materials to Paper MmWave Codebook Selection in Rapidly-Varying Channels via Multinomial Thompson Sampling. https://www.dropbox.com/s/12gppf7am2qdlw4/mmwave-extended.pdf?dl=0

[41] Yi Zhang, Kartik Patel, Sanjay Shakkottai, and Robert W. Heath Jr. 2019. Side-Information-Aided Noncoherent Beam Alignment Design for Millimeter Wave Systems. In *Proc. of the 20th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '19)*. Catania, Italy, 341–350.

[42] Renjie Zhao, Timothy Woodford, Teng Wei, Kun Qian, and Xinyu Zhang. 2020. M-Cube: A Millimeter-Wave Massive MIMO Software Radio. In *Proc. of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom '20)*. London, United Kingdom, Article 15, 14 pages.