PERFORMANCE OF INTERLEAVED TRAINING FOR SINGLE-USER HYBRID MASSIVE ANTENNA DOWNLINK

Cheng Zhang[§], *Yindi Jing*[‡], *Yongming Huang*[§], *and Luxi Yang*[§]

[§]Southeast University, P. R. China, email: {zhangcheng1988, huangym, lxyang}@seu.edu.cn.
[‡]University of Alberta, Canada, email: yindi@ualberta.ca.

ABSTRACT

In this paper, we study the beam-based training design for the single-user (SU) hybrid massive antenna system based on outage probability performance. First, an interleaved training design is proposed where the feedback is concatenated with the training procedure to monitor the training status and to have the training length adaptive to the channel realization. Then, the average training length and outage probability are derived for the proposed interleaved training and SU transmission. Analytical results and simulations show that the proposed interleaved scheme achieves the same outage performance as the traditional full-training scheme but with significant saving in the training overhead.

***Index Terms*—** Massive antenna, beam-training, outage probability, training overhead.

1. INTRODUCTION

For massive antenna systems [1–4], the full-digital implementation with one full radio-frequency (RF) chain for each antenna tends to be impractical due to its high hardware cost and power consumptions [5, 6]. Recently, enabled by the cost-effective phase shifters, a hybrid analog-digital structure has been applied for massive antenna systems which can effectively reduce the hardware complexity and cost [7, 8]. One crucial practical issue for hybrid massive antenna downlink is the acquisition of channel state information (CSI) at the base station (BS). One popular method for the downlink training is to combine the codebook based beam training with the traditional channel estimation [8, 9] in which a finite codebook is used for the analog precoders, called beams. Then the channel estimation problem is transformed to the estimation of the beam-domain effective channels. This can reduce the dimension of the channel estimation problem if the number of desired beams is limited.

Several typical beam-based training schemes have been proposed for single-user (SU) hybrid massive antenna downlink. One straightforward method is to exhaustively train all beams in the codebook, then to find the best beam or beam combination for transmission. Another typical method uses hierarchical search [8, 9], where beams are iteratively trained with shrinking beamwidth until the optimal beam with acceptable beamwidth is found. While hierarchical search has lower training overhead than exhaustive search, its beam alignment quality is very sensitive to the pre-beamforming signal-to-noise-ratio (SNR) [10, 11]. Meanwhile, its advantage of lower training overhead diminishes as the number of channel paths increases [8]. In [12], the tradeoff of transmission beamwidth and training overhead was studied.

The basic idea of existing schemes is to obtain the complete effective CSI for the beam codebook. Then based on the obtained effective CSI values, data transmission designs are proposed. In such schemes, the training design and data transmission design are decoupled. For satisfactory performance, the size of the training beam codebook increases linearly with the BS antenna number, leading to heavy training overhead for massive antenna systems. The decoupled nature of existing schemes also imposes limitation on the tradeoff of training overhead and performance. Further, only the throughput or diversity gain has been considered in existing work. In [13, 14], an interleaved training scheme was proposed for the downlink of SU full-digital massive antenna systems and independent and identically distributed (i.i.d.) channels, where the BS trains its channels sequentially and receives CSI or indicator feedback immediately after each training. With each feedback, the BS decides to train another channel or terminate training based on whether an outage occurs. The scheme was shown to achieve significant reduction in training overhead with the same outage performance as traditional schemes.

In this paper, we study the beam-based training design jointly with the data transmission design for SU hybrid massive antenna systems. We adopt the ideas in [13, 14] to propose interleaved training designs for such systems that are dynamic and adaptive, in the sense that the length of the training interval depends on the channel realization and the termination of the training interval depends on the previous training results, to achieve a better tradeoff between the outage performance and the training overhead. Our work is different in the two aspects compared with [13, 14]. First, we consider hybrid massive antenna systems with beam-based training. Second, a more general channel model is used that incorporates channel correlation and limited scattering. Given these differences in system and channel models, the proposed interleaved training scheme and performance analysis are largely different. In summary, for SU hybrid massive antenna systems with limited scattering channels, a beam-based interleaved training scheme is proposed. Then the average training length and the outage probability of the proposed scheme are studied. Analytical results and simulations show that the proposed scheme achieves the same outage probability as the traditional non-interleaved full-training scheme but with significant saving in the training overhead. Moreover, useful insights are obtained on the effect of important system parameters on the average training length and the outage performance.

2. SYSTEM MODEL AND PROBLEM STATEMENT

Consider the downlink of a hybrid massive antenna system, where the BS employs $N_t \gg 1$ antennas with $N_{RF} \in [1, N_t)$ RF chains and serves one single-antenna user. We consider the typical uniform array at the BS and high dimensionality. Thus the beamspace representation of the channel is used [15]. Denote the discrete Fourier transform (DFT) matrix as $\mathbf{D} \in \mathbb{C}^{N_t \times N_t}$ where the i-th column is $\mathbf{d}_i = [1, e^{-j2\pi(i-1)/N_t}, ..., e^{-j2\pi(i-1)(N_t-1)/N_t}]^T, \forall i$. Assume there are $L \in [1, N_t]$ distinguishable paths [15] in the user's channel, and define the set of their direction indices as $\mathcal{I} = \{I_1, ..., I_L\}$. The channel vector can be written as

$$\mathbf{h} = \mathbf{D}\bar{\mathbf{h}} = [\mathbf{d}_1, ..., \mathbf{d}_{N_t}][\bar{h}_1, ..., \bar{h}_{N_t}]^T, \quad (1)$$

where $\bar{h}_i \sim \mathcal{CN}(0, 1/L)$, i.e., circularly symmetric complex Gaussian distribution with the mean being 0 and the variance being $1/L$, for $i \in \mathcal{I}$ and $\bar{h}_i = 0$ for $i \notin \mathcal{I}$, and the L-combination $(I_1, \cdots, I_L)$ is assumed to follow discrete uniform distribution with each element on $[1, N_t]$.

While generally speaking, L can be arbitrary, we consider the typical scenario that L is a constant with respect to N_t, i.e., $L = \mathcal{O}(1)$. This corresponds to channels with limited scattering where having more antennas does not result in more distinguishable paths. One application is the outdoor environment with a few dominant clusters, especially at mmWave band [15]. Similarly, $N_{RF} = \mathcal{O}(1)$ is considered from the perspective of low hardware cost.

2.1. Hybrid Precoding and Outage Probability

The hybrid structure at the BS allows for an analog RF precoding $\mathbf{F}_{RF} \in \mathbb{C}^{N_t \times L_s}$ followed with a baseband beamforming $\mathbf{f}_{BB} \in \mathbb{C}^{L_s \times 1}$ where L_s is the number of beams used for the transmission and $L_s \leq N_{RF}$. All elements of $\mathbf{F}_{RF}$ have the same constant norm, i.e., $\|[\mathbf{F}_{RF}]_{i,j}\|^2 = 1/N_t, \forall i, j$. The codebook-based beamforming scheme is used, where columns of $\mathbf{F}_{RF}$ are chosen from a codebook of vectors $\mathcal{F}_{RF}$ [8], [9]. Naturally, with the channel model in (1), the DFT codebook is used [16], where $\mathcal{F}_{RF} = \{\mathbf{d}_1^*/\sqrt{N_t}, \cdots, \mathbf{d}_{N_t}^*/\sqrt{N_t}\}$ with $\mathbf{d}_i^*$ being the complex conjugate of $\mathbf{d}_i$. Each element in the codebook is also called a beam and there are N_t beams in total. With given $\mathbf{F}_{RF}$, the effective channel matrix for the baseband is $\mathbf{h}^T\mathbf{F}_{RF}$. More specifically, $\mathbf{h}^T\mathbf{d}_i^*/\sqrt{N_t} = \sqrt{N_t}\bar{h}_i$ is the effective channel on Beam i. If $\bar{h}_i \neq 0$, Beam i is a non-zero beam for the user.

After the BS chooses L_s beams for analog precoding, the transceiver equation can be written as

$$y = \sqrt{P}\mathbf{h}^T\mathbf{F}_{RF}\mathbf{f}_{BB}s + n = \sqrt{P}\bar{\mathbf{h}}^T\mathbf{D}\mathbf{F}_{RF}\mathbf{f}_{BB}s + n, \quad (2)$$

where P is the short-term total transmit power, s denotes the data stream with unit power, and n is the additive noise following $\mathcal{CN}(0, 1)$. For a given transmission rate R_{th}, an outage event occurs if $\|\bar{\mathbf{h}}^T\mathbf{D}\mathbf{F}_{RF}\mathbf{f}_{BB}\|^2 \leq \alpha = (2^{R_{th}} - 1)/P$ where α is called the target normalized received SNR. Thus, for random $\mathbf{h}$, the outage probability is

$$\text{out}(\mathbf{F}_{RF}, \mathbf{f}_{BB}) \triangleq \Pr(\|\bar{\mathbf{h}}^T\mathbf{D}\mathbf{F}_{RF}\mathbf{f}_{BB}\|^2 \leq \alpha). \quad (3)$$

If the effective channel vector is perfectly known at the BS, the baseband design for minimizing the outage probability is $\mathbf{f}_{BB} = (\bar{\mathbf{h}}^T\mathbf{D}\mathbf{F}_{RF})^H/\|\bar{\mathbf{h}}^T\mathbf{D}\mathbf{F}_{RF}\|$. In this case, the outage probability becomes $\Pr(\|\bar{\mathbf{h}}^T\mathbf{D}_t\mathbf{F}_{RF}\|^2 \leq \alpha)$.

If more than L_s non-zero beams are available, beam selection is needed. Define the set of available non-zero beam indices as $\mathcal{A} = \{a_1, ..., a_j\}$. The optimal beam-selection is to find the strongest L_s ones within the set. By ordering the magnitudes of the effective channels as $\|\bar{h}_{s_1}\| \geq \|\bar{h}_{s_2}\| \geq ... \geq \|\bar{h}_{s_{L_s}}\| \geq \cdots \geq \|\bar{h}_{s_j}\|$. The set of indices of the selected beams is $\mathcal{S} = \{s_1, ..., s_{L_s}\}$. Thus the beamforming matrices are:

$$\mathbf{F}_{RF} = \left[\frac{\mathbf{d}_{s_1}^*}{\sqrt{N_t}}, ..., \frac{\mathbf{d}_{s_{L_s}}^*}{\sqrt{N_t}}\right], \mathbf{f}_{BB} = \frac{[\bar{h}_{s_1}, ..., \bar{h}_{s_{L_s}}]^H}{\sqrt{\|\bar{h}_{s_1}\|^2 + ... + \|\bar{h}_{s_{L_s}}\|^2}}. \quad (4)$$

The outage probability reduces to $\Pr\left(\sum_{i=1}^{L_s} \|\bar{h}_{s_i}\|^2 \leq \alpha/N_t\right)$.

2.2. Existing Beam-Based Training and Motivations of Interleaved Training

To implement hybrid precoding, CSI is needed at the BS, thus downlink training and CSI feedback must be conducted. For the considered hybrid massive antenna system, beam-based training is a more economical choice than traditional antenna-domian training [17]. One typical beam-based training scheme operates as follows: for each channel realization, the BS sequentially transmits along all N_t beams for the user to estimate the corresponding effective channels. The effective channel values are then sent back to the BS. Other than this full training, partial training can be conducted by choosing L_t out of the N_t beams [18].

In the aforementioned schemes, the training length is fixed regardless of the channel realization and further the effective CSI feedback is separated from the training procedure. Thus we refer such designs as *non-interleaved training (NIT)*. The combination of NIT and data transmission for SU systems is referred to as NIT-SU transmission scheme. To save training time while still having the best outage probability performance, *interleaved training* idea is used, where the training length is adaptive to the channel realizations. Further, the effective CSI or indicator feedback is concatenated with the pilot transmissions to monitor the training status and guide the action of the next symbol period.

3. PROPOSED INTERLEAVED TRAINING AND SU TRANSMISSION

This section is on the interleaved beam-based training and the corresponding SU transmission design. For interleaved training, instead of training all beams and finding the best combination as in NIT, the training should stop right after enough beams have been trained to avoid outage. Since the set of L non-zero beams for the user $\mathcal{I}$ is random with uniform distribution on the set $\{1, \cdots, N_t\}$, and the channel coefficients along the non-zero beams are i.i.d., the priorities of the training for all beams are the same. Therefore, the natural order is used for beam training, i.e., the BS trains from the first beam to the N_t-th beam sequentially. The training stops when the outage can be avoided based on the already trained beams or no more beam is available for training.

Let $\mathcal{B}$ contain the indices of the non-zero beams that have been trained. Thus $L_B = \min(N_{RF}, |\mathcal{B}|)$ is the maximum number of non-zero beams that can be used for data transmissions. Let $\mathcal{S}$ contain the indices of the L_B known non-zero beams with the largest norm. The proposed interleaved training and joint SU transmission (IT-SU) scheme is shown in Algorithm 1.

In the proposed scheme, at the ith training interval where $i \leq N_t$, the BS sends a pilot for the user to estimate the ith beam value: $\bar{h}_i$ (The scalar $\sqrt{N_t}$ is omitted for brief notation). If it is a non-zero beam (i.e., $\|\bar{h}_i\| > 0$), the user compares the received SNR provided by the L_B strongest beams known from the finished i training intervals with the target value α to see if an outage event will happen. If $\|\bar{h}_i\| = 0$ or $\sum_{l \in \mathcal{S}} \|\bar{h}_l\|^2 \leq \alpha/N_t$ and $i < N_t$, the already trained beams cannot provide a beam combination to avoid outage. Thus the user feeds back the indicator "0" to request the BS to continue training the next beam. For $i = N_t$, all beams have been trained and the outage is unavoidable with any beam combination. If $\sum_{l \in \mathcal{S}} \|\bar{h}_l\|^2 > \alpha/N_t$, enough beams have been trained to avoid outage. Thus the user feeds back the L_B non-zero effective channels $\bar{h}_l, l \in \mathcal{S}$ and their indices and the BS aligns the L_B beams with $\mathbf{F}_{RF}$ and matches the effective channel vector with $\mathbf{f}_{BB}$ as in (4) to conduct data transmission.

Algorithm 1 Proposed Interleaved Training and Joint SU Transmission (IT-SU) scheme

```
B = ∅;
for i = 1, ..., N_t do
    The BS trains the i-th beam; The user estimates h̄_i;
    If ||h̄_i|| > 0, B = B ∪ {i} and the user finds S, which contains
    the indices of the L_B non-zero beams with the largest norm,
    then calculates Σ_{l∈S} ||h̄_l||^2;
    if ||h̄_i|| = 0 or Σ_{l∈S} ||h̄_l||^2 ≤ α/N_t then
        The user feeds back "0"; Continue;
    else
        The user feeds back h̄_l, ∀l ∈ S and their indices;
        The BS constructs F_RF and f_BB as in (4); Break;
    end if
end for
```

4. AVERAGE TRAINING LENGTH ANALYSIS

Since for different channel realizations, the number of beams being trained in our proposed IT-SU scheme can vary due to the randomness in the path profile and path gains, in this section we study the average training length measured in the number of training intervals per channel realization where the average is over channel distribution. For any given channel realization and at any step of the training process, larger N_{RF} means that the same or more beam combinations are available. Thus the same or larger received SNR can be provided, which results in the same or an earlier termination of the training period. Meanwhile, since there are at most L non-zero paths in the channel, having a larger N_{RF} than L cannot provide better beam combination compared with that of when $N_{RF} = L$. Therefore, the average training length of the IT-SU scheme over all channel realizations is a decreasing function of N_{RF} for $N_{RF} \le L$, then keeps unchanged for $N_{RF} \ge L$. Thus, in the following analysis, we only need to consider the scenario of $1 \le N_{RF} \le L$. Recall that L is a finite value while $N_t \to \infty$. The asymptotic result on the average training length of the proposed IT-SU scheme is provided.

Theorem 1. *For the hybrid massive antenna system with $N_t \gg 1$ BS antennas, finite constant number of channel paths L, and target normalized received SNR α, when the number of RF chains is one, i.e., $N_{RF} = 1$, or is the same as the path number, i.e., $N_{RF} = L$, the average training length of the proposed IT-SU scheme can be written as follows*

$$T_{\textit{IT-SU}} = \frac{N_t}{L+1} + \mathcal{O}(1). \tag{5}$$

Proof. Due to space limit, please refer to the journal version [19, Appendix B]. □

For the two special cases of N_{RF} values, the average training length of the IT-SU scheme increases linearly with N_t, but the slope decreases linearly with L. The traditional NIT-SU scheme with full training has a fixed training length N_t. Thus the proposed IT-SU scheme has significant saving in training time as N_t is very large. It is noteworthy that this gain in time is obtained with no cost in outage probability performance (more details will be explained in the next section). Moreover, the average training length is independent of the threshold α.

From the discussion at the beginning of this section, we know that for any value of N_{RF}, the average training length is lower bounded by its value for $N_{RF} = L$ and upper bounded by its value for $N_{RF} = 1$. Thus the analytical results in Theorem 1 lead to the following corollary.

Corollary 1. *For the hybrid massive antenna system with $N_t \gg 1$ BS antennas, finite constant number of channel paths L, and target normalized received SNR α, the average training length of the proposed IT-SU scheme can be written as* (5) *for any number of RF chains N_{RF}.*

5. OUTAGE PERFORMANCE ANALYSIS

In this section, we will first analyze the outage probability of the NIT-SU scheme with arbitrary fixed training length L_t. Then, the outage performance of the proposed IT-SU scheme is shown to be the same as the NIT-SU scheme with full training (where $L_t = N_t$).

Define j as the number of non-zero channel paths aligned by the L_t training beams. Let $L_s = \min(N_{RF}, j)$ and $\eta_j = \binom{L_t}{j}\binom{N_t-L_t}{L-j}/\binom{N_t}{L}$ which is the probability that j paths are aligned by the L_t training beams. Further, define

$$J_1 = \max(0, L - N_t + L_t), \quad J_2 = \min(L, L_t). \tag{6}$$

Theorem 2. *For the hybrid massive antenna system with N_t BS antennas, the L-path channel, N_{RF} RF chains and the target normalized received SNR α, the outage probability of the NIT-SU scheme with training length $L_t \le N_t$ is*

$$\mathrm{out}(\textit{NIT-SU}, L_t) = \sum\nolimits_{j=J_1}^{J_2} \eta_j \bar{P}_j, \tag{7}$$

where $\bar{P}_0 = 1$,

$$\bar{P}_j = \binom{j}{L_s}\Bigg[\frac{\Upsilon(L_s, \frac{\alpha L}{N_t})}{(L_s-1)!} + \sum_{l=1}^{j-L_s}(-1)^{L_s+l-1}\binom{j-L_s}{l} \times \left(\frac{L_s}{l}\right)^{L_s-1}\left(\frac{1-e^{(-\frac{\alpha L}{N_t})(1+l/L_s)}}{1+l/L_s} - B(l)\right)\Bigg], \quad j > 0, \tag{8}$$

$$B(l) = \begin{cases} \sum_{m=0}^{L_s-2} \frac{1}{m!}\left(-\frac{l}{L_s}\right)^m \Upsilon(m+1, \frac{\alpha L}{N_t}) & L_s \ge 2 \\ 0 & \textit{otherwise} \end{cases},$$

and $\Upsilon(s, x) = \int_0^x t^{s-1}e^{-t}dt$ is the lower incomplete gamma function.

Proof. See Appendix A. □

Theorem 2 provides an analytical expression for the outage probability of the NIT-SU scheme with any fixed training length L_t. The expression is in closed-form other than the special function Υ. Although the effect of N_{RF} and L_t on the outage performance of the NIT-SU scheme is implicit in (7), since for each channel realization, increasing N_{RF} from 1 to $\min(L, L_t)$ and/or having larger L_t can provide the same or more available strongest non-zero beams for transmission, the outage probability is a decreasing function of $N_{RF} \in [1, \min(L, L_t)]$ and $L_t \in [1, N_t]$.

For the IT-SU scheme in Algorithm 1, an outage happens only when all beams have been trained and no beam combination can satisfy the target SNR requirement. Thus the outage probability is the same as NIT-SU when $L_t = N_t$. The following corollary is obtained.

Corollary 2. *For the hybrid massive antenna system with N_t BS antennas, the L-path channel, N_{RF} RF chains and the target normalized received SNR α, the outage probability of the IT-SU scheme is*

$$\text{out}(\textit{IT-SU}) = \binom{L}{\tilde{N}_{RF}}\left[\frac{\Upsilon\left(\tilde{N}_{RF}, \frac{\alpha L}{N_t}\right)}{(\tilde{N}_{RF}-1)!} + \sum_{l=1}^{L-\tilde{N}_{RF}} (-1)^{\tilde{N}_{RF}+l-1}\binom{L-\tilde{N}_{RF}}{l} \times \left(\frac{\tilde{N}_{RF}}{l}\right)^{\tilde{N}_{RF}-1}\left(\frac{1-e^{\left(-\frac{\alpha L}{N_t}\right)\left(1+l/\tilde{N}_{RF}\right)}}{1+l/\tilde{N}_{RF}} - C(l)\right)\right], \tag{9}$$

where $\tilde{N}_{RF} = \min(N_{RF}, L)$ and

$$C(l)=\begin{cases}\sum_{m=0}^{\tilde{N}_{RF}-2}\frac{1}{m!}\left(-\frac{l}{\tilde{N}_{RF}}\right)^m \Upsilon\left(m+1, \frac{\alpha L}{N_t}\right) & \tilde{N}_{RF} \geq 2 \\ 0 & \textit{otherwise}\end{cases}. \tag{10}$$

From Theorem 1, the average training length of the proposed IT-SU scheme is smaller than N_t. Moreover, since smaller training time results in larger outage probability for the NIT-SU scheme, the IT-SU scheme is superior in terms of outage probability compared with the NIT-SU scheme with partial training at the same training length.

For the special case with $N_{RF} = 1$, a simplified outage probability expression is obtained.

Corollary 3. *For the hybrid massive antenna system with $N_t \gg 1$ BS antennas, finite constant number of channel paths L, single RF chain, and the target normalized received SNR α, the outage probability of the IT-SU scheme is as follows:*

$$\text{out}(\textit{IT-SU}) = (\alpha L/N_t)^L + \mathcal{O}\left(N_t^{-(L+1)}\right) \tag{11}$$

Proof. See Appendix B. □

This means that the outage probability scales as $\mathcal{O}\left(N_t^{-L}\right)$ for large N_t. Therefore, for the SU massive antenna system with single RF chain, arbitrarily small outage probability can be achieved for any desired date rate and any fixed power consumption P as long as N_t is large enough. For the case of multiple RF chains where $1 < N_{RF} \leq L$, since larger N_{RF} results in smaller outage probability with the IT-SU scheme, arbitrarily small outage probability can also achieved for an arbitrary date rate with any fixed power consumption P when N_t is large enough.

6. NUMERICAL RESULTS

In this section, simulation results are shown to verify the analytical results in this paper. In the left plot of Fig. 1, the average training length of the IT-SU scheme is demonstrated for $\alpha = 4$. It can be seen that 1) for $L = 1, 3, 6$, the average training length for IT-SU increases linearly with N_t with decreasing slopes; 2) the gap between the simulated average training length and the asymptotic result in Theorem 1 is small and shrinks as N_t increases; 3) increasing N_{RF} from 1 to L results in negligible reduction in training overhead; 4) compared to the NIT-SU scheme with full training, the saving in training overhead for IT-SU is significant, especially for larger L. In the right plot of Fig. 1, the outage probabilities of the proposed IT-SU scheme and the NIT-SU scheme with partial training are demonstrated for $\alpha = 4$. The training length of the NIT-SU scheme is set to be the same as that of IT-SU for fair comparison. It can be seen that 1) the analytical result in Theorem 2 matches the simulated value well; 2) IT-SU has much lower outage probability than NIT-SU; 3) increasing N_{RF} from 1 to L can reduce the outage probability of IT-SU; 4) the result in Corollary 3 (the black dotted line) tightly matches the simulation for the $L = 1$ case, while for the $L = 6$ case, a gap is visible but the analytical result well represents the asymptotic behavior.

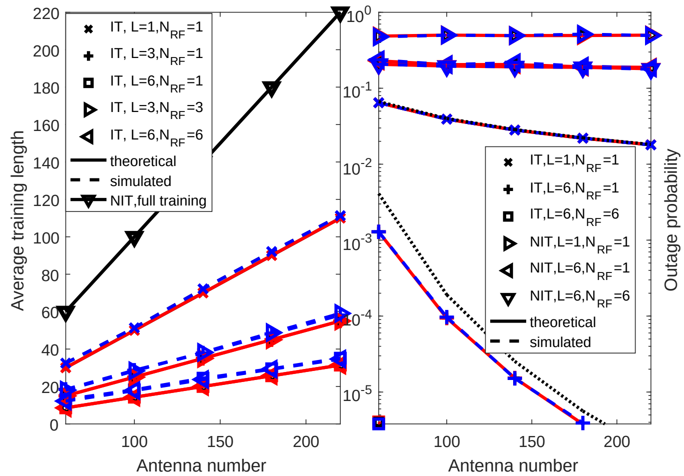

Fig. 1. Average training length and outage probability of IT-SU and NIT-SU with full/partial training.

7. CONCLUSION

For single-user hybrid massive antenna systems with limited scattering channels, we studied the interleaved beam-based training and joint beamforming design with outage probability as the performance measure. Via concatenating the feeback with the training, the training length can be adaptive to channel realizations. Then, analytical expressions of the average training length and the outage probability were derived. Analytical and simulated results show that the proposed scheme achieves the same outage performance as the traditional full-training scheme while saves the training overhead significantly.

8. APPENDIX

A. Proof for Theorem 2: Due to the space limit, only a sketch of the proof is give and details are omitted. For a given channel realization, with the NIT-SU scheme and training length L_t, outage happens when the strongest L_s beams among the j trained non-zero beams cannot avoid outage. Notice that J_1 and J_2 are defined as the bounds of j. Define $\bar{P}_j$ as the probability of outage conditioned on that j non-zero beams are found. Apparently, $\bar{P}_0 = 1$. The outage probability can be represented as (7). Define the indices of the strongest L_s beams as $s_1, ..., s_{L_s}$. According the PDFs of $x = \sum_{l=1}^{L_s} \|\bar{h}_{s_l}\|^2$ for $0 < j \leq N_{RF}, j > N_{RF}$ (for the latter case, the distribution of the partial sum of order statistics [20, Eq. 3.19] should be used), (8) can be obtained.

B. Proof for Corollary 3: By utilizing the power series of the lower incomplete gamma function and the Taylor series of exponential function in (9) with $\tilde{N}_{RF} = 1$, we have

$$\text{out(IT-SU)}=\sum_{l=0}^{L-1}(-1)^{l+1}\binom{L}{l+1}\left[\left(1-\frac{\alpha L}{N_t}+\mathcal{O}(N_t^{-2})\right)^{1+l}-1\right].$$

By replacing $t = l + 1$ and utilizing the equality $(x + y)^n = \sum_{l=0}^{n}\binom{n}{l}x^{n-l}y^l$, (11) can be obtained.

9. ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant 61720106003, the Research Project of Jiangsu Province under Grant BE2015156 and the Scientic Research Foundation of Graduate School of Southeast University.

10. REFERENCES

[1] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: how many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, pp. 160–171, Feb. 2013.

[2] C. Zhang, Y. Jing, Y. Huang, and L. Yang, "Performance scaling law for multicell multiuser massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, pp. 9890–9903, Nov. 2017.

[3] C. Zhang, Y. Huang, Y. Jing, S. Jin, and L. Yang, "Sum-rate analysis for massive MIMO downlink with joint statistical beamforming and user scheduling," *IEEE Trans. Wireless Commun.*, vol. 16, pp. 2181–2194, Apr. 2017.

[4] Y. Long, Z. Chen, J. Fang, "Nonasymptotic analysis of capacity in massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 4, pp. 541–544, Oct. 2015.

[5] M. Xiao, S. Mumtaz, and Y. Huang et al., "Millimeter wave communications for future mobile networks," *IEEE J. Sel. Areas Commun.*, vol. 35, pp. 1909–1935, Sept. 2017.

[6] Q. Ding and Y. Jing, "Receiver energy efficiency and resolution profile design for massive MIMO uplink with mixed-ADC," *IEEE Trans. Veh. Technol.*, DOI: 10.1109/TVT.2017.2763825, online published.

[7] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, pp. 501–513, Apr. 2016.

[8] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, pp. 831–846, Oct. 2014.

[9] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Millimeter wave beamforming for wireless backhaul and access in small cell networks," *IEEE Trans. Commun.*, vol. 61, pp. 4391–4403, Oct. 2013.

[10] C. Liu, M. Li, S. V. Hanly, I. B. Collings, and P. Whiting, "Millimeter wave beam alignment: large deviations analysis and design insights," *IEEE J. Sel. Areas Commun.*, vol. 35, pp. 1619–1631, July 2017.

[11] J. Zhang, Y. Huang, Q. Shi, J. Wang, and L. Yang, "Codebook design for beam alignment in millimeter wave communication systems," *IEEE Trans. on Commun.*, vol. 65, pp. 4980–4995, 2017,

[12] H. Shokri-Ghadikolaei, L. Gkatzikis, and C. Fischione, "Beam-searching and transmission scheduling in millimeter wave communications," in *2015 IEEE Int. Conf. Commun. (ICC)*, 2015, pp. 1292–1297.

[13] E. Koyuncu and H. Jafarkhani, "Interleaving training and limited feedback for point-to-point massive multiple-antenna systems," in *2015 IEEE Int. Symp. Inf. Theory (ISIT)*, 2015, pp. 1242–1246.

[14] E. Koyuncu and H. Jafarkhani, "Interleaving training and limited feedback for point-to-point massive multiple-antenna systems," *Tech. Rep., [Online]:* `https://webfiles.uci.edu/ekoyuncu/intlv.pdf`, Jan. 2015.

[15] R. W. Heath, N. Gonzlez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, pp. 436–453, Apr. 2016.

[16] S. He, J. Wang, Y. Huang, B. Ottersten, and W. Hong, "Codebook-based hybrid precoding for millimeter wave multiuser systems," *IEEE Trans. Signal Process.*, vol. 65, pp. 5289–5304, Oct. 2017.

[17] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, pp. 884–893, Mar. 2006.

[18] G. Lee, J. So, and Y. Sung, "Impact of training on mmWave multi-user MIMO downlink," in *2016 IEEE Global Conf. Signal and Inf. Process. (GlobalSIP)*, 2016, pp. 753–757.

[19] C. Zhang, Y. Jing, Y. Huang, and L. Yang, "Interleaved training and training-based transmission design for hybrid massive antenna downlink," *arXiv preprint arXiv:1710.06949*, Oct. 2017.

[20] H. C. Yang and M. S. Alouini, *Order statistics in wireless communications: diversity, adaptation, and scheduling in MIMO and OFDM systems,* Cambridge Univ. Pr., 2011.