Diversity versus Training Overhead Trade-off for Low Complexity Switched Transceivers

Vishnu V. Ratnam*, *Student Member, IEEE*, Andreas F. Molisch*, *Fellow, IEEE*,
Naif Rabeah†, *Member, IEEE*, Faisal Alawwad† and Hatim Behairy†
*Ming Hsieh Dept. of Electrical Engineering, University of Southern California
Los Angeles, California, USA. Email: {ratnam, molisch}@usc.edu
†National Center for Electronics and Photonics Technology, King Abdulaziz City for Science and Technology
Riyadh, Saudi Arabia. Email: {nbnrabeah, falouwad, hbehairy}@kacst.edu.sa

***Abstract*—Many wireless transceivers, such as massive-MIMO and Rake, are designed to provide high diversity order. In order to reduce cost, low-complexity switched transceivers, e.g., hybrid antenna preprocessing with selection, or S-Rake, which employ a limited number of RF chains or correlators, are popular. This paper shows that the presence of limited number of RF chains or correlators in such transceivers introduces an inherent trade-off between system performance and the channel estimation overhead. We prove that to maximize achievable rate, it is optimal to perform channel estimation for only a subset of diversity branches with the highest second moments, if the diversity branches have the same fading parameters but different mean powers. We also prove that the achievable data-rate is a unimodal function of the size of this subset L, which ensures that any locally optimum L is also globally optimum. A computationally efficient approximation for the ergodic capacity is introduced, which reduces the cost of finding the optimal L significantly. Simulation results for some practically important settings suggest that the optimal choice of L can improve the data rate by a factor of $20-30\%$.**

I. INTRODUCTION

Diversity, e.g., in the form of spatial (antenna) diversity or delay diversity, is the key method to reduce the negative impact of fading in wireless communications systems. The general trend towards larger number of antennas and/or wider bandwidths means that the number of diversity branches is increasing. For example, a massive multiple-input-multiple-output (MIMO) base station might have 100 antenna elements, or an ultrawideband (UWB) receiver might have ≈ 50 Rake fingers.

While such a high number of diversity branches brings performance advantages, the associated commensurate increase in the number of up/down-conversion chains[1] drastically increases the implementation cost and energy consumption in the transceivers. As a solution, low complexity switched transceivers have been proposed in the literature, such as hybrid antenna preprocessing with selection [1], [2] and S-Rake [3], [4]. In such systems, the number of up/down-conversion chains (K) are fewer than the number of diversity branches the channel offers and an array of switches select K out of the N diversity branches for down-conversion. For such a low-complexity receiver (assuming a single antenna transmitter and in the absence of interference), the capacity optimal way for combining the received signals is generalized selection combining (GSC) [5], also known as hybrid selection / maximum ratio combining [6]. In GSC, the instantaneously strongest K diversity branches are down-converted and maximum-ratio combined. The performance of a GSC system has been studied in great detail for independent Rayleigh fading [7], Nakagami-m fading [8] and arbitrary fading [9]–[11] channels. Similarly, the correlated channel scenarios have been discussed in [8], [12].

For implementing GSC, the receiver (RX) needs the channel state information (CSI) for all the diversity paths. This CSI can be acquired by transmitting a known pilot sequence from the transmitter (TX) during each channel coherence time interval. However, for low complexity systems, since the RX has only $K < N$ down-conversion chains, for each transmitted pilot sequence the RX can acquire CSI for only K diversity paths. Therefore, the pilot sequence has to be re-transmitted $\lceil N/K \rceil$ times to acquire the CSI for all the diversity paths.[2] This overhead for training can be especially large when $N \gg K$, the pilot sequence is long and/or the channel coherence time is short.

In this paper we suggest and analyze a novel method to minimize the impact of this training overhead.[3] It is motivated by the fact that in practice, not all diversity branches have the same average power: for example, in massive MIMO with preprocessing, the different effective beams that are available at the output of the preprocessor for selection carry different powers. Consequently, some diversity branches might make only a minor contribution to boosting the system capacity. Therefore, it is better to acquire CSI for only a subset of L paths, where $K \leq L \leq N$, and L is a trade-off between the training overhead the performance gain from increased diversity. Obviously, the determination of this subset has to be based on second-order statistics of the diversity paths only;

Financial support by KACST project under grant number 33-878 is gratefully acknowledged.

[1]We henceforth only say "up/down-conversion chains", though this always means "up/down-conversion chains for massive MIMO or correlators for Rake receivers."

[2]Fewer re-transmissions may be required with channel sparsity constraints. In this work we assume no such sparsity constraints.

[3]A preliminary study of this trade-off for a UWB system with Rayleigh faded diversity paths has been considered by us in [13].

those second-order statistics change very slowly with time, and thus can be easily tracked at the RX.

The main contributions of this paper are as follows:

1) We formulate the problem of finding the data-rate maximizing L and corresponding set of paths $\mathcal{L}^*(L)$.
2) We prove that for a given value of L, a simple decision rule can be used to find the corresponding optimal set of paths $\mathcal{L}^*(L)$.
3) We prove that the achievable data-rate is a unimodal function of L, which ensures that any locally optimum L is also globally optimum.
4) We propose a computationally efficient approximation for the ergodic capacity of a GSC system, which significantly reduces the cost of finding optimal L.
5) We study the diversity versus training overhead trade-off in two practically important low complexity switched RXs.

Notation: Scalars are represented by light-case letters; vectors by bold-case letters; matrices by capitalized bold-case letters; and sets by calligraphic letters. Additionally, a_i represents the i-th element of a vector $\mathbf{a}$ and $|\mathcal{A}|$ the cardinality of a set $\mathcal{A}$. Also, $\mathbb{E}\{\}$ represents the expectation operator, $\mathbb{I}_A$ the $A \times A$ identity matrix, $\mathbb{O}_{A\times B}$ the $A \times B$ all-zero matrix, $\lceil a \rceil$ the smallest integer larger than a and f_x,F_x the probability density and cumulative distribution for a random variable x, respectively.

II. General Assumptions and Channel model

We consider a GSC system with a single antenna TX. The channel offers N diversity paths and the RX only picks K diversity paths for down-conversion. Without loss of generality, we assume N is a multiple of K.[4] Under these assumptions, the base-band equivalent received signal vector during any symbol duration can be represented as:

$$\mathbf{y} = \sqrt{\rho}\mathbf{S}\mathbf{h}x + \mathbf{S}\mathbf{n} \tag{1}$$

where $\mathbf{y}$ is the $K \times 1$ received signal vector corresponding to the K down-conversion chains, ρ is the average signal-to-noise ratio (SNR), $\mathbf{S}$ is a $K \times N$ sub-matrix of $\mathbb{I}_N$ that picks the best K branches for down-conversion, $\mathbf{h}$ is the $N \times 1$ normalized channel vector corresponding to the N diversity paths, x is the transmit data symbol and $\mathbf{n} \sim \mathcal{CN}(\mathbb{O}_{N\times N}, \mathbb{I}_N)$ is the $N \times 1$ normalized additive white gaussian noise vector. We assume that the channel diversity paths h_i are independent but not identically distributed (i.n.i.d)[5] and their amplitudes follow a Nakagami-m distribution with the probability density function given by:

$$f_{|h_i|}(x) = \frac{2m^m}{\Gamma(m)\Omega_i^m} x^{2m-1} \exp\{-\frac{mx^2}{\Omega_i}\} \tag{2}$$

[4]If this is not true, dummy diversity paths of zero magnitude can be appended

[5]In a Single Input Multiple Output (SIMO) channel, the signals at the antenna elements are often correlated. However, in many common hybrid preprocessing systems [1], [14], [15], the effective channel can be modeled as an i.n.i.d vector, since those systems effectively perform beam selection, and the fading in different beams is independent.

where the shape parameter (m) is fixed but the spread parameter (Ω_i) is different for each diversity path i. Without loss of generality, the channel is normalized such that $\sum_{i=1}^{N} \Omega_i = N$. Some relevant distribution parameters of h_i are enlisted in Table I. We assume the RX has knowledge of the average power $\mathbb{E}\{|h_i|^2\} = \Omega_i$ for all the N paths.[6] Since the average power changes very slowly, it can be easily tracked for all the N paths (see [16] and references therein).

TABLE I
Distribution parameters of diversity path h_i

CDF: $F_{\|h_i\|}(x)$	$\frac{\Gamma_{\text{lower,inc}}(m, mx^2/\Omega_i)}{\Gamma(m)}$
Avg. power: $\mathbb{E}\{\|h_i\|^2\}$	Ω_i

The channel is assumed to be block fading, wherein the channel stays constant for a coherence time interval and then changes to another random realization with distribution as in (2). During each coherence time interval, the pilot sequence is re-transmitted $\lceil L/K \rceil$ times to acquire the CSI for L diversity paths ($K \le L \le N$). Let us define the CSI acquisition set $\mathcal{L} \subseteq \{1, .., N\}$ as the set of indices of diversity paths whose CSI is acquired at the RX. Assuming perfect CSI estimation, it can be shown that the instantaneous SNR for GSC can be expressed as:

$$\gamma_{\text{GSC}}(\mathcal{L}) = \max_{\mathcal{S}\subseteq\mathcal{L}, |\mathcal{S}|=K} \left\{ \rho \sum_{i\in\mathcal{S}} |h_i|^2 \right\} \tag{3}$$

Correspondingly the achievable data rate can be expressed as:

$$R(\mathcal{L}) = (1 - \lceil L/K \rceil \theta_p)\, C(\mathcal{L}) \tag{4}$$

where, $C(\mathcal{L}) = \mathbb{E}\{\log(1 + \gamma_{\text{GSC}}(\mathcal{L}))\}$ is the ergodic capacity and θ_p is the fraction of time-frequency resources consumed by the pilot sequence.[7] From (4), there is a trade-off between the number of diversity branches used L and the amount of CSI training required $\lceil L/K \rceil \theta_p$. In this paper we find the CSI acquisition set $\mathcal{L}_{\text{opt}}$ and its size $L_{\text{opt}} = |\mathcal{L}_{\text{opt}}|$ that maximizes the achievable data rate.

III. Problem Formulation

Consider the following family of optimization problems for $K \le L \le N$:

$$\mathcal{L}^*(L) = \underset{\mathcal{L}\subset\{1,..,N\} \big| |\mathcal{L}|=L}{\operatorname{argmax}} R(\mathcal{L}) \tag{5}$$

Then the rate maximizing CSI acquisition set can be expressed as $\mathcal{L}_{\text{opt}} = \mathcal{L}^*(L_{\text{opt}})$, where:

$$L_{\text{opt}} = \underset{K\le L\le N}{\operatorname{argmax}} \{R(\mathcal{L}^*(L))\} \tag{6}$$

Theorem III.1. *An optimal solution $\mathcal{L}^*(L)$ to* (5) *is given by:*

$$\mathcal{L}^*(L) = \{\ \eta_1 \quad \eta_2 \quad \dots \quad \eta_L\ \} \tag{7}$$

[6]The average path power corresponds to the Power Delay Profile in a UWB system and the Power Angle Spectrum in a SIMO system

[7]In general, θ_p is a function of not only the symbol duration but also of the number of active TXs with orthogonal pilots (see Sec V).

where $\boldsymbol{\eta}$ is a permutation of the vector $[1,...,N]$ such that $\Omega_{\eta_i} \geq \Omega_{\eta_j}$ for all $i \leq j$.[8]

Proof. Suppose $\{\eta_1, \eta_2, ..., \eta_L\}$ is not an optimal solution to (5). Consider any optimal solution $\mathcal{L}^*(L) \neq \{\eta_1, \eta_2, ..., \eta_L\}$. Then there exist distinct numbers $a_1, ..a_p, b_1, .., b_p$ $(1 \leq a_1, ..a_p \leq L < b_1, .., b_p \leq N)$ such that:

$$\left(\mathcal{L}^*(L) \cup \{\eta_{a_1}, ..\eta_{a_p}\}\right) \backslash \{\eta_{b_1}, .., \eta_{b_p}\} = \{\eta_1, ..., \eta_L\}$$

Note that, from the definition of η, we have $\Omega_{\eta_{a_j}} \geq \Omega_{\eta_{b_j}}$ for all $1 \leq j \leq p$. From (3), we then have:

$$\begin{aligned} \gamma_{\text{GSC}}(\mathcal{L}^*(L)) &= \max_{\mathcal{S} \subseteq \mathcal{L}^*(L), |\mathcal{S}|=K} \left\{ \rho \sum_{i \in \mathcal{S}} |h_i|^2 \right\} \\ &\leq \max_{\mathcal{S} \subseteq \mathcal{L}^*(L), |\mathcal{S}|=K} \left\{ \rho \sum_{i \in \mathcal{S}} \alpha_i |h_i|^2 \right\} \end{aligned} \quad (8)$$

where, we define constants:

$$\alpha_i = \begin{cases} \frac{\Omega_{\eta_{a_j}}}{\Omega_{\eta_{b_j}}} & \text{for } i = \eta_{b_j}, 1 \leq j \leq p \\ 1 & \text{otherwise} \end{cases} \quad (9)$$

It can be easily verified from (2), that $|h_{\eta_{a_j}}| \overset{d}{=} \sqrt{\alpha_{\eta_{b_j}}} |h_{\eta_{b_j}}| \ \ \forall j$, where $\overset{d}{=}$ denotes equality in distribution. Now since h_i are independently distributed for $1 \leq i \leq N$, from (8) we have:

$$\begin{aligned} \gamma_{\text{GSC}}(\{\eta_1, ..., \eta_L\}) &\overset{d}{=} \max_{\mathcal{S} \subseteq \mathcal{L}^*(L), |\mathcal{S}|=K} \left\{ \rho \sum_{i \in \mathcal{S}} \alpha_i |h_i|^2 \right\} \\ \Rightarrow \gamma_{\text{GSC}}(\{\eta_1, ..., \eta_L\}) &\overset{d}{\geq} \gamma_{\text{GSC}}(\mathcal{L}^*(L)) \end{aligned} \quad (10)$$

where $\overset{d}{\geq}$ represents first order stochastic dominance of the left hand side over the right hand side. Using (4) and (10) we further have:

$$R(\mathcal{L}^*(L)) \leq R(\{\eta_1, ..., \eta_L\}) \quad (11)$$

which is in contradiction to our initial assumption. This concludes the proof. □

Though it seems intuitively correct to only acquire the CSI for the L diversity paths with the highest average power, it is worth mentioning that Theorem III.1 is non-trivial. For example, the above does not hold if the shape parameter m were different for each diversity path.

Since we now know how to find $\mathcal{L}^*(L)$, the problem of finding $\mathcal{L}_{\text{opt}}$ is reduced to finding optimal size L_{opt} in (6).

Theorem III.2. *$C(\mathcal{L}^*(L))$ is a non-negative, non-decreasing and a concave function*[9] *of L i.e.,*

$$\Delta C_{L+1} \leq \Delta C_L \ \ \textit{for } K \leq L \leq N-1 \quad (12)$$

where $\Delta C_L \triangleq C(\mathcal{L}^*(L)) - C(\mathcal{L}^*(L-1))$.

[8]A version of this theorem for Rayleigh fading was proved by us in [13]

[9]By concavity, we mean that the piece-wise linear function obtained by interpolating $C(\mathcal{L}^*(L))$ for non-integer L is concave

Proof. Since $\mathcal{L}^*(L) \subset \mathcal{L}^*(L+1)$, from (3) and (4), $C(\mathcal{L}^*(L))$ is a non-negative, non-decreasing function of L.

For any L, let us define a new random vector $\hat{\mathbf{h}}$ such that: $\hat{h}_{\eta_i} = h_{\eta_i}$ for $i \notin \{L, L+1\}$, $\hat{h}_{\eta_L} \overset{\text{d}}{=} h_{\eta_L}$ but is independent of $\mathbf{h}$, and $\hat{h}_{\eta_{L+1}} = h_{\eta_L} \sqrt{\Omega_{\eta_{L+1}} / \Omega_{\eta_L}}$. It can be easily verified from (2) that $\hat{\mathbf{h}} \overset{\text{d}}{=} \mathbf{h}$. Let $|h^L_{(i)}|, |\hat{h}^L_{(i)}|$ represent magnitude of the i-th largest diversity paths (in magnitude) from the sets $\{|h_j| \big| j \in \mathcal{L}^*(L)\}$ and $\{|\hat{h}_j| \big| j \in \mathcal{L}^*(L)\}$, respectively.[10] Then from (3) we have:

$$\gamma_{\text{GSC}}(\mathcal{L}^*(L)) = \sum_{i=1}^{K} \rho |h^L_{(i)}|^2 \quad (13)$$

We also define:

$$\begin{aligned} \Delta\gamma_{\text{GSC}}(L) &\triangleq \gamma_{\text{GSC}}(\mathcal{L}^*(L)) - \gamma_{\text{GSC}}(\mathcal{L}^*(L-1)) \\ &= \max\{\rho |h^{L-1}_{(K)}|^2, \rho |h_{\eta_L}|^2\} - \rho |h^{L-1}_{(K)}|^2 \end{aligned} \quad (14)$$

Now the incremental capacity can be expressed as:

$$\begin{aligned} \Delta C_L &= C(\mathcal{L}^*(L)) - C(\mathcal{L}^*(L-1)) \\ &= \mathbb{E}\left\{ \int_0^{\Delta\gamma_{\text{GSC}}(L)} \frac{1}{1 + \gamma_{\text{GSC}}(\mathcal{L}^*(L-1)) + x} dx \right\} \end{aligned} \quad (15)$$

Since $\hat{\mathbf{h}} \overset{\text{d}}{=} \mathbf{h}$, we can also write:

$$\Delta C_{L+1} = \mathbb{E}\left\{ \int_0^{\Delta\hat{\gamma}_{\text{GSC}}(L+1)} \frac{1}{1 + \hat{\gamma}_{\text{GSC}}(\mathcal{L}^*(L)) + x} dx \right\} \quad (16)$$

where $\hat{\gamma}_{\text{GSC}}(\mathcal{L}^*(L)), \Delta\hat{\gamma}_{\text{GSC}}(L+1)$ are as in (13)–(14) with terms of $\mathbf{h}$ replaced by corresponding terms of $\hat{\mathbf{h}}$. As $\mathcal{L}^*(L-1) \subset \mathcal{L}^*(L)$, from the definition of $\hat{\mathbf{h}}$ it can be easily verified that $|\hat{h}^L_{(K)}| \geq |h^{L-1}_{(K)}|$ and $\hat{\gamma}_{\text{GSC}}(\mathcal{L}^*(L)) \geq \gamma_{\text{GSC}}(\mathcal{L}^*(L-1))$, for all channel realizations. Additionally, using theorem III.1, $\hat{h}_{\eta_{L+1}} \leq h_{\eta_L}$ and so $\Delta\hat{\gamma}_{\text{GSC}}(L+1) \leq \Delta\gamma_{\text{GSC}}(L)$. Using these results and from (15)–(16), the theorem follows. □

Since $C(\mathcal{L}^*(L))$ is non-decreasing function of L, from (4) we have: $R(\mathcal{L}^*(L)) \leq R(\mathcal{L}^*(K \lceil \frac{L}{K} \rceil))$ for all $K \leq L \leq N$. Therefore we can simplify the optimization problem in (6) to:

$$L_{\text{opt}} = \underset{L \in \{K, 2K, .., N\}}{\text{argmax}} \{R(\mathcal{L}^*(L))\} \quad (17)$$

For ease of notation, let $C(\mathcal{L}^*(0)) = C(\mathcal{L}^*(N+K)) = 0$. Then any $L^* \in \{K, 2K, .., N\}$ is a local maximum of (17) iff:

$$\begin{aligned} R(\mathcal{L}^*(L^*-K)) \leq R(\mathcal{L}^*(L^*)) \geq R(\mathcal{L}^*(L^*+K)) \\ \equiv \Delta C^K_{L^*} \geq g(L^*) \text{ and } \Delta C^K_{L^*+K} \leq g(L^*+K) \end{aligned} \quad (18)$$

where $\Delta C^K_{L^*} = C(\mathcal{L}^*(L^*)) - C(\mathcal{L}^*(L^*-K))$ and:

$$g(L) \triangleq \frac{C(\mathcal{L}^*(L-K))}{\frac{1}{\theta_p} - \frac{L}{K}} \quad (19a)$$

Since from theorem III.2, ΔC^K_L is a non-increasing function of L and $g(L)$ is a non-decreasing function of L, any locally

[10]We set $|h^L_{(i)}| = 0$ if $i > |\mathcal{L}^*(L)|$

optimum L^* for (17) is also a globally optimum solution. Therefore, instead of a brute-force search, we use the following linear search algorithm to find L_{opt}:[11]

Algorithm 1: Find L_{opt}

$N, K, m, \rho, \mathbf{\Omega}$ - inputs
Initialize $L = K$, $C(\mathcal{L}^*(0)) = 0$;
while $L < N$ **do**
  Compute $C(\mathcal{L}^*(L))$;
  Compute $C(\mathcal{L}^*(L+K))$;
  if $\Delta C_L^K \geq g(L)$ and $\Delta C_{L+K}^K \leq g(L+K)$ **then**
    return L;
  end if
  $L = L + K$;
end while
return N

IV. Computing the capacities

In this section, we find a method to calculate $C(\mathcal{L}^*(L))$. Most of the prior works on the performance analysis of GSC, rely on finding the moment generating function (MGF) of the SNR. Finding the MGF is in itself a computationally intensive exercise involving $\approx K\binom{L}{K}$ one-dimensional integrals in general [10]. Therefore techniques to find the capacity from the MGF such as [17], [18] become computationally infeasible if K and/or L are large. Instead, in this work we rely on the upper bound on capacity:

$$C_{\text{UB}}(\mathcal{L}) \triangleq \log(1 + \mathbb{E}\{\gamma_{\text{GSC}}(\mathcal{L})\}) \geq C(\mathcal{L}) \tag{20}$$

to find a near-optimal L_{opt} in Algorithm 1. It can be easily verified that Theorems III.1 and III.2 are also applicable if we replace $C(\mathcal{L})$ by $C_{\text{UB}}(\mathcal{L})$. Computing $C_{\text{UB}}(\mathcal{L})$ (which is a function of the mean SNR) is also an involved exercise involving $\approx K^2\binom{L}{K}$ one-dimensional integrals [10]. Though some works also find closed form results [7], they involve a larger number of iterations and thus do not necessarily reduce the computational complexity. This computational load can be extremely large especially if K and/or L are large and therefore alternate approaches are required. Observing that we know $C_{\text{UB}}(\mathcal{L}^*(L-K))$ while finding $C_{\text{UB}}(\mathcal{L}^*(L))$ in Algorithm 1, we can recursively define:

$$e^{C_{\text{UB}}(\mathcal{L}^*(L))} = e^{C_{\text{UB}}(\mathcal{L}^*(L-1))} + \mathbb{E}\{\Delta\gamma_{\text{GSC}}(L)\} \tag{21}$$

Using (14), we can write:

$$\mathbb{E}\{\Delta\gamma_{\text{GSC}}(L)\} = \int_{x=0}^{\infty} x^2 \Big[F_{|h_{(K)}^{L-1}|}(x) f_{|h_{\eta_L}|}(x) - f_{|h_{(K)}^{L-1}|}(x)(1 - F_{|h_{\eta_L}|}(x))\Big] dx \tag{22}$$

[11] Though a binary search for a local optimum may have fewer iterations, the linear search can be used to recursively compute $C(\mathcal{L}^*(L))$, as shall be illustrated in section IV

where:

$$f_{|h_{(K)}^{L-1}|}(x) = \sum_{\mathbf{b}\in\mathcal{P}_{L-1}^{(1:K-1,K+1:L-1)}} f_{|h_{b_K}|}(x) \left[\prod_{i=1}^{K-1}\Big(1 - F_{|h_{b_i}|}(x)\Big)\right] \times \left[\prod_{j=K+1}^{L-1} F_{|h_{b_j}|}(x)\right] \tag{23}$$

$$F_{|h_{(K)}^{L-1}|}(x) = \sum_{k=0}^{K-1} \sum_{\mathbf{b}\in\mathcal{P}_{L-1}^{(1:k,k+1:L)}} \left[\prod_{i=1}^{k}\Big(1 - F_{|h_{b_i}|}(x)\Big)\right] \times \left[\prod_{j=k+1}^{L-1} F_{|h_{b_j}|}(x)\right] \tag{24}$$

and $\mathcal{P}_{L-1}^{(a:b,c:d)}$ is set of all permutations of the vector $[\eta_1, .., \eta_{L-1}]$ such that $\forall \mathbf{b} \in \mathcal{P}_{L-1}^{(a:b,c:d)}, b_a < b_{a+1} < .. < b_b$ and $b_c < b_{c+1} < .. < b_d$. In general, this recursive definition does not lead to any significant savings in computing $C_{\text{UB}}(\mathcal{L}^*(L))$. However, in the special case where $\mathcal{L}^*(L-1)$ has independent and identically distributed (i.i.d.) diversity paths, we have:

$$f_{|h_{(K)}^{L-1}|}^{\text{iid}}(x) = K\binom{L-1}{K} f_{|h_{L-1}^{\text{iid}}|}(x)\Big(1 - F_{|h_{L-1}^{\text{iid}}|}(x)\Big)^{K-1} \times \Big(F_{|h_{L-1}^{\text{iid}}|}(x)\Big)^{L-K-1} \tag{25}$$

$$F_{|h_{(K)}^{L-1}|}^{\text{iid}}(x) = \sum_{k=0}^{K-1}\binom{L-1}{k}\Big(1 - F_{|h_{L-1}^{\text{iid}}|}(x)\Big)^{k} \times \Big(F_{|h_{L-1}^{\text{iid}}|}(x)\Big)^{L-k-1} \tag{26}$$

$$\mathbb{E}\{\gamma_{\text{GSC}}^{\text{iid}}(\mathcal{L}^*(L-1))\} = \sum_{k=1}^{K} k\binom{L-1}{k}\int_{x=0}^{\infty}\Big[x^2 f_{|h_{L-1}^{\text{iid}}|}(x) \times \Big(1 - F_{|h_{L-1}^{\text{iid}}|}(x)\Big)^{k-1}\Big(F_{|h_{L-1}^{\text{iid}}|}(x)\Big)^{L-k-1}\Big] dx \tag{27}$$

where $f_{|h_{L-1}^{\text{iid}}|}(x)$ and $F_{|h_{L-1}^{\text{iid}}|}(x)$ are the marginal PDF and CDF, respectively, of $|h_i|\ \forall i \in \mathcal{L}^*(L-1)$. In this special case, computing $C_{\text{UB}}(\mathcal{L}^*(L))$ from $C_{\text{UB}}(\mathcal{L}^*(L-1))$ only involves computing K one-dimensional integrals.

To reduce the cost of computation in the general i.n.i.d. case, while computing $\mathbb{E}\{\Delta\tilde{\gamma}_{\text{GSC}}(L)\}$[12] (from (22)), we approximate $\mathcal{L}^*(L-1)$ to be composed of i.i.d. components. In other words, we approximate $f_{|h_{(K)}^{L-1}|}(x)$ and $F_{|h_{(K)}^{L-1}|}(x)$ by $f_{|h_{(K)}^{L-1}|}^{\text{iid}}(x)$ and $F_{|h_{(K)}^{L-1}|}^{\text{iid}}(x)$, respectively, where the i.i.d. spreading parameter $\Omega_{L-1}^{\text{iid}}$ is such that: $\mathbb{E}\{\gamma_{\text{GSC}}^{\text{iid}}(\mathcal{L}^*(L-1))\} = \mathbb{E}\{\tilde{\gamma}_{\text{GSC}}(\mathcal{L}^*(L-1))\}$. Note that from (21), $\mathbb{E}\{\tilde{\gamma}_{\text{GSC}}(\mathcal{L}^*(L-1))\}$ is already available when computing $\mathbb{E}\{\Delta\tilde{\gamma}_{\text{GSC}}(L)\}$.

[12] We use $\tilde{X}$ to denote an approximation for X ($X = \gamma_{\text{GSC}}, C_{\text{UB}}$)

This procedure is detailed in Algorithm 2 and shall be referred to as 'RecursiveIID Approx'. The proposed approximation is accurate when either the spreading parameters Ω_i are equal for some i and negligible for others or are skewed such that $\sum_{i=1}^{K}\Omega_i \gg \sum_{i=K+1}^{N}\Omega_i$. A study of the accuracy of the approximation for several practically relevant power spectra ($\mathbf{\Omega}$) are studied in Fig 1. Here the ergodic capacity $C(\mathcal{L}^*(L))$ is compared to both the capacity upper bound $C_{\text{UB}}(\mathcal{L}^*(L))$ and the Recursive-IID Approx. $\tilde{C}_{\text{UB}}(\mathcal{L}^*(L))$ (as obtained via Algorithm 2). Since the exact computations of $C(\mathcal{L}^*(L))$, $C_{\text{UB}}(\mathcal{L}^*(L))$ are infeasible, we rely on Monte-Carlo simulations. The results show that Recursive-IID Approx provides a very good approximation to $C_{\text{UB}}(\mathcal{L}^*(L))$. Though there is a some gap between $C(\mathcal{L}^*(L))$ and $\tilde{C}_{\text{UB}}(\mathcal{L}^*(L))$, the gap is more or less constant. As shall be shown in Sec V, the impact of this gap on L_{opt} is minimal. Similar results are observed for other power spectra, barring a few heavy tail functions like the Zipf probability mass function.

Algorithm 2: Compute $\tilde{C}_{\text{UB}}(\mathcal{L}^*(L))$ recursively[12]

$\mathbb{E}\{\tilde{\gamma}_{\text{GSC}}(\mathcal{L}^*(L-1))\}, L, K, m, \rho, \Omega_{\eta_L}$ - inputs
if $L \le K$ **then**
  return
  $\tilde{C}_{\text{UB}}(\mathcal{L}^*(L)) = \log(1 + \mathbb{E}\{\tilde{\gamma}_{\text{GSC}}(\mathcal{L}^*(L-1))\} + \rho\Omega_{\eta_L})$
end if
Find $\Omega^{\text{iid}}_{L-1}$ s.t.

$$\mathbb{E}\{\gamma^{\text{iid}}_{\text{GSC}}(\mathcal{L}^*(L-1))\} = \mathbb{E}\{\tilde{\gamma}_{\text{GSC}}(\mathcal{L}^*(L-1))\}$$

where $\mathbb{E}\{\gamma^{\text{iid}}_{\text{GSC}}(\mathcal{L}^*(L-1))\}$ is as defined in (27) and:

$$f_{|h^{\text{iid}}_{L-1}|}(x) \triangleq \frac{2}{\Gamma(m)}\left[\frac{m}{\Omega^{\text{iid}}_{L-1}}\right]^m x^{2m-1}\exp\{-\frac{mx^2}{\Omega^{\text{iid}}_{L-1}}\}$$

$$f_{|h^{\text{iid}}_{L-1}|}(x) \triangleq \frac{\Gamma_{\text{lower,inc}}(m, mx^2/\Omega^{\text{iid}}_{L-1})}{\Gamma(m)}$$

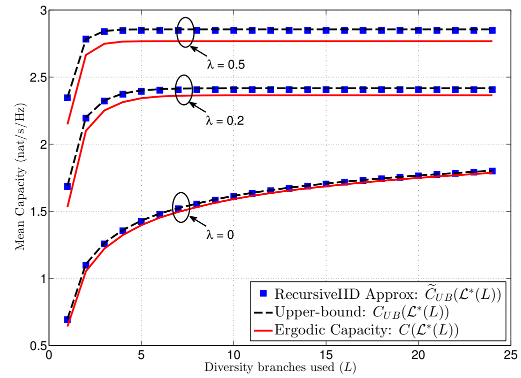
{For example, using FSOLVE in MATLAB}
Compute $\mathbb{E}\{\Delta\tilde{\gamma}_{\text{GSC}}(L)\}$ from (22) with $\tilde{f}_{|h^{L-1}_{(K)}|}(x), \tilde{F}_{|h^{L-1}_{(K)}|}(x)$ as given by (25)–(26).
$\mathbb{E}\{\tilde{\gamma}_{\text{GSC}}(\mathcal{L}^*(L))\} = \mathbb{E}\{\tilde{\gamma}_{\text{GSC}}(\mathcal{L}^*(L-1))\} + \mathbb{E}\{\Delta\tilde{\gamma}_{\text{GSC}}(L)\}$
return $\tilde{C}_{\text{UB}}(\mathcal{L}^*(L)) = \log(1 + \mathbb{E}\{\tilde{\gamma}_{\text{GSC}}(\mathcal{L}^*(L))\})$

V. Simulation Results

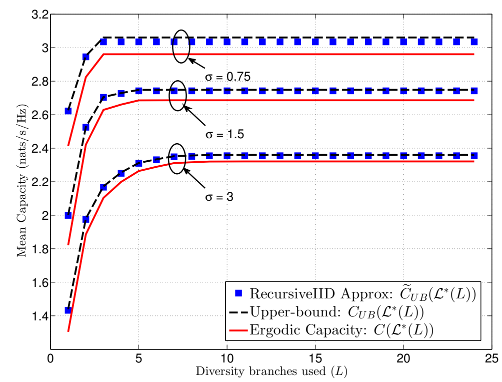
For simulations, we consider a system with a single antenna TX and a low complexity switched RX. We consider two relevant scenarios: 1) A UWB system with impulse radio signalling and an S-Rake RX 2) An Orthogonal Frequency Division Multiplexed SIMO system with a multi-antenna RX. For finding the pilot overhead, we assume that there are U such single antenna TXs in the system.[13] Orthogonal pilots are assigned to the TXs to prevent pilot contamination.

[12]The constant κ is chosen such that $\sum_{i=1}^{N}\Omega_i = N$

[13]The U TXs may have dedicated RXs, such as in a peer-to-peer network, or may have a common RX such as in a multiple access channel.



(a) Truncated exponential



(b) Truncated gaussian

Fig. 1. Capacity as a function of L for different diversity power spectra: (a) Considers an exponential power spectrum $\Omega_i = \kappa\exp\{-\lambda i\}$ (b) Considers a Gaussian power spectrum $\Omega_i = \kappa\exp\left\{-\frac{(i-\lfloor N/2\rfloor)^2}{2\sigma^2}\right\}$ (system parameters: $m = 2, N = 24, K = 2, \rho = 1$)[12]

The fractional pilot overhead in the two cases is computed as: $\theta_p^{\text{UWB}} \approx \frac{T_{\text{symb}}U}{T_{\text{coh}}}$ and $\theta_p^{\text{MIMO}} \approx \frac{\tau_{\text{rms}}U}{T_{\text{coh}}}$.[14] The simulation parameters are summarized in Table II and are similar to the parameters in IEEE 802.15.4a PAN (Personal Area Network) standard [19] and the cellular LTE (Long Term Evolution) standard [20], respectively.

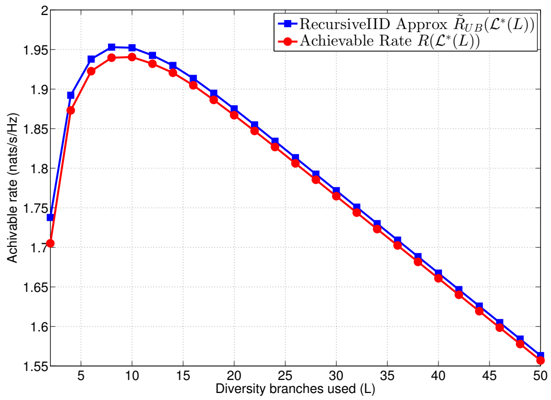
TABLE II
Simulation Parameters

Cellular Layout	**UWB**	**SIMO**
No. of diversity paths (N)	50	100
No. of down-conversion chains (K)	2	5
Carrier freq	6 GHz	2 GHz
Coherence time (T_{coh})	10 ms	10 ms
Delay spread (τ_{rms})	100 ns	500 ns
Symbol duration (T_{symb})	$8\mu s$	$100\mu s$
No. of Users (U)	25	200
Fractional pilot overhead (θ_p)	0.02	0.01

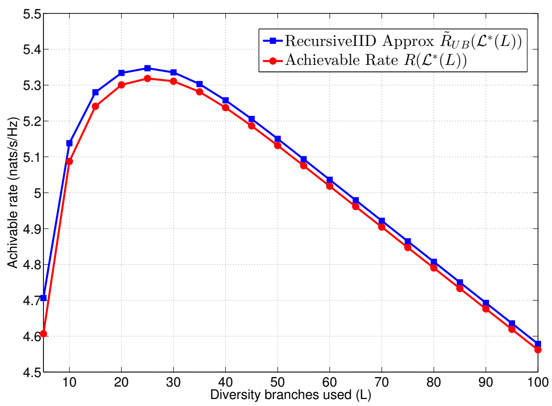
Assuming that the multiple TXs have orthogonal access in

[14]This result is obtained considering that we need one pilot per TX in each coherence time and coherence bandwidth for OFDMA and that the TXs have orthogonal time access in a UWB Personal Area Network [19].

time, frequency or space (no interference), we can restrict to a single TX-RX analysis, as is considered in this work. The achievable rates for the two scenarios as a function of L is presented in Fig 2. Here, the achievable rate $R(\mathcal{L}^*(L))$ (as obtained via Monte-Carlo simulations) are compared to the Recursive-IID Approx $\tilde{R}_{\text{UB}}(\mathcal{L}^*(L))$ (as obtained via Algorithm 2 and (4)). The results suggest that the proposed



(a) UWB: $m = 5, \rho = 1$



(b) SIMO: $m = 1, \rho = 10$

Fig. 2. Achievable rate rate as a function of L for two practical scenarios: (a) A UWB system with $\Omega_i = \sum_{j=1}^{6} \kappa \exp\{-\frac{j}{25} - \frac{2i-10j}{15}\} u[2i-10j]$ (where $u[i] = 1$ for $i \geq 0$ and $u[i] = 0$ otherwise) (b) Considers a SIMO system with $\Omega_i = \kappa \sum_{j=1}^{20} (j/20)^2 \exp\left\{\frac{[1.8(i-50)-\phi_j]^2}{50}\right\}$ $\left(\text{where } \phi_j = 36(-1)^j \sqrt{-2\log(j/20)}\right)$ [12]

recursive IID algorithm predicts the value of L_{opt} very accurately. Also, as expected, $L_{\text{opt}} \ll N$ and this choice leads to a significant increase in achievable data rate($\approx 20 - 30\%$).

VI. Conclusion

In this paper, we study the trade-off between diversity and training overhead for low complexity switched transceivers. We conclude that if the diversity paths have same fading parameters but different mean powers, it is data rate optimal to only acquire CSI for a subset of paths (of size L) with the highest mean powers. Simulation results for some practically important settings suggest that the optimal choice of L can improve data rate by a factor of ≈ 20–30%. Under typical scenarios $L_{\text{opt}} > K$, suggesting that with judicious pilot training, S-Rake outperforms P-Rake [3] and introducing a selection stage improves performance of MIMO hybrid preprocessing [1], even after accounting for training overhead.

References

[1] P. Sudarshan, N. B. Mehta, A. F. Molisch, and J. Zhang, "Channel statistics-based RF pre-processing with antenna selection," *IEEE Transactions on Wireless Communications*, vol. 5, pp. 3501–3511, December 2006.

[2] A. F. Molisch and X. Zhang, "FFT-based hybrid antenna selection schemes for spatially correlated MIMO channels," *IEEE Communications Letters*, vol. 8, pp. 36–38, Jan 2004.

[3] D. Cassioli, M. Z. Win, F. Vatalaro, and A. F. Molisch, "Performance of low-complexity RAKE reception in a realistic UWB channel," in *Communications, 2002. ICC 2002. IEEE International Conference on*, vol. 2, pp. 763–767 vol.2, 2002.

[4] W. C. Lan, M.-S. Alouini, and M. K. Simon, "Optimum spreading bandwidth for selective rake reception over rayleigh fading channels," *Selected Areas in Communications, IEEE Journal on*, vol. 19, no. 6, pp. 1080–1089, 2001.

[5] T. Eng, N. Kong, and L. B. Milstein, "Comparison of diversity combining techniques for rayleigh-fading channels," *IEEE Transactions on Communications*, vol. 44, pp. 1117–1129, Sep 1996.

[6] M. Z. Win and J. H. Winters, "Analysis of hybrid selection/maximal-ratio combining of diversity branches with unequal snr in rayleigh fading," in *Vehicular Technology Conference, 1999 IEEE 49th*, vol. 1, pp. 215–220 vol.1, Jul 1999.

[7] M. S. Alouini and M. K. Simon, "An MGF-based performance analysis of generalized selection combining over rayleigh fading channels," *IEEE Transactions on Communications*, vol. 48, pp. 401–415, Mar 2000.

[8] R. K. Mallik and M. Z. Win, "Analysis of H-S/MRC in correlated nakagami fading," in *Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE*, vol. 2, pp. 1389–1393 vol.2, Nov 2002.

[9] Y. Ma and S. Pasupathy, "Performance of generalized selection combining on generalized fading channels," in *Communications, 2003. ICC '03. IEEE International Conference on*, vol. 5, pp. 3041–3045 vol.5, May 2003.

[10] Y. Ma and S. Pasupathy, "Efficient performance evaluation for generalized selection combining on generalized fading channels," *IEEE Transactions on Wireless Communications*, vol. 3, pp. 29–34, Jan 2004.

[11] A. Annamalai, G. K. Deora, and C. Tellambura, "Theoretical diversity improvement in GSC(N,L) receiver with nonidentical fading statistics," *IEEE Transactions on Communications*, vol. 53, pp. 1027–1035, June 2005.

[12] X. Zhang and N. C. Beaulieu, "Performance analysis of generalized selection combining in generalized correlated nakagami- m fading," *IEEE Transactions on Communications*, vol. 54, pp. 2103–2112, Nov 2006.

[13] V. V. Ratnam, A. F. Molisch, N. Alrabeah, F. Alawwad, and H. Behairy, "JS-RAKE: Judiciously trained selective rake receiver for UWB systems," in *2016 IEEE International Conference on Ubiquitous Wireless Broadband*, 2016.

[14] A. Alkhateeb, O. El Ayach, G. Leus, and R. Heath, "Hybrid precoding for millimeter wave cellular systems with partial channel knowledge," in *Information Theory and Applications Workshop (ITA), 2013*, pp. 1–5, Feb 2013.

[15] A. Adhikary, J. Nam, J. Y. Ahn, and G. Caire, "Joint spatial division and multiplexing 2014;the large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, pp. 6441–6463, Oct 2013.

[16] S. Haghighatshoar and G. Caire, "Channel vector subspace estimation from low-dimensional projections," *CoRR*, vol. abs/1509.07469, 2015.

[17] F. Yilmaz and M. S. Alouini, "A unified MGF-based capacity analysis of diversity combiners over generalized fading channels," *IEEE Transactions on Communications*, vol. 60, pp. 862–875, March 2012.

[18] F. Yilmaz and M. S. Alouini, "Novel asymptotic results on the high-order statistics of the channel capacity over generalized fading channels," in *Signal Processing Advances in Wireless Communications (SPAWC), 2012 IEEE 13th International Workshop on*, pp. 389–393, June 2012.

[19] "Low-rate wireless personal area networks (LR-WPANs)." IEEE 802.15.4, 2011.

[20] "Further advancements for E-UTRA physical layer aspects." 3GPP TR 36.814, 2010.