Abstract—This paper considers wireless downlink transmission when multiple antennas are employed at both the base station (BS) and the mobile terminals. Assorted dynamic transmission schemes based on TDMA and SDMA are described and their associated system throughput limits are compared, under different assumptions on the availability of channel side information (CSI) at the BS. The trade-off between scheme complexity and system throughput is discussed.

I. INTRODUCTION

This paper studies wireless downlink transmission schemes and their achievable system throughput for a time-varying MIMO broadcast channel (MIMO-BC). The situation of interest is a single cell system with $M$ transmit antennas at the BS and $N$ receive antennas at each of $K$ mobile terminals. The expected throughput, which is the statistical expectation of the total data rate delivered by the BS to mobile terminals with respect to the random MIMO-BC, is used as the limit measure of the system long-term average throughput. If $M = 1$, the BC is degraded [1] and the transmission scheme for maximizing the expected throughput is Dynamic TDMA, i.e., only one user associated with the best channel quality is scheduled for transmission at one time. If $M > 1$, the BC becomes non-degraded and the scheme for maximizing the expected throughput becomes Dynamic SDMA, i.e., an optimal set of multiple users should be selected for transmission at one time [2]. Both Dynamic TDMA and Dynamic SDMA can explore a multiuser diversity gain of the system throughput because the BS dynamically assigns data rates to the selected user(s) based on the quality of the multiuser channel. Dynamic TDMA supports a single user at one time and therefore is less complex for implementation than Dynamic SDMA. The multiuser diversity available in Dynamic TDMA is limited to be the selection diversity across users while Dynamic SDMA can fully explore the multiuser diversity by simultaneously transmitting to the optimal set of users. It is shown in [3] [4] that when the number of users, $K$, becomes very large, the ratio of the system throughput with Dynamic SDMA to that with Dynamic TDMA converges to be $M / \min(M, N)$. This can be explained by the fact that the spatial multiplexing gain, which is the multiplicative gain of the system throughput when $K$ becomes very large, is $M$ for Dynamic SDMA but $\min(M, N)$ for Dynamic TDMA. Therefore, when $M$ is much larger than $N$, Dynamic SDMA can exhibit a substantial throughput gain than Dynamic TDMA.

In order to explore the multiuser diversity gain of the throughput, both Dynamic TDMA and Dynamic SDMA select one user for transmission only when its channel quality is sufficiently good and transmitting to this user contributes to maximizing the expected throughput. However, other possible factors on each user’s data traffic like traffic arrival process and delay requirement are not considered in the expected throughput. Under the assumption that the network is homogeneous (i.e., the time-varying channel of each user is statistically independent but has identical distribution), the expected throughput is a relevant performance indicator because the time-varying nature of the channel ensures that each user is assigned with an equal portion of throughput in the long term, provided that the data traffic is delay-tolerant. Furthermore, the expected throughput provides a performance limit for more practical transmission schemes that incorporate other data transmission requirements into the user selection process.

The optimal Dynamic SDMA solution to maximizing the expected throughput of the MIMO-BC is based on so-called “Dirty-Paper-Coding (DPC)” technique [5][6][7][8]. DPC requires complex non-linear processing at both encoder and decoder. A suboptimal SDMA solution, usually referred as Beamforming (BF), supports simultaneous transmission to multiple users by assigning a different beam pattern to each active user. Only linear processing is required for BF at the encoder and therefore a BF-based SDMA system is considered less complex than a DPC-based one. The design problems related to both DPC and BF for the downlink BC are often more conveniently solved by considering the dual uplink multiple-access channel (MAC), (see [6][7] and references therein).

Optimal Dynamic SDMA and TDMA solutions assume that the complete channel side information (CSI) is available at both the BS and the mobile terminals. In many systems, CSI is unknown at the BS and needs to be measured at each mobile terminal and then fed back to the BS. The overhead for complete CSI feedback may be too costly to implement and therefore partial CSI feedback with a tolerable system throughput degradation becomes an important issue for such systems. In [9], a technique so-called “Opportunistic Beamforming (OBF)” that combines random linear precoding and effective signal-to-noise ratio (ESNR) feedback is proposed for a Dynamic TDMA system with $N = 1$. OBF avoids complete CSI feedback and still approaches the performance of coherent
transmission that requires complete CSI at the BS, when \( K \) becomes very large. The concept of OBF is also applied to Dynamic TDMA with \( N > 1 \) [10] and BF-based Dynamic SDMA [11].

The following notations are used throughout the paper. Let \( T, \mbox{ Tr } \) and \( \dagger \) denote the matrix transpose, trace, and conjugate transpose, respectively, \( \| \cdot \| \) be the vector Euclidean norm, \( \mathbb{E}_X \) be the expectation with respect to the random variable \( X \), \( \mbox{Diag}(x) \) be the diagonal matrix with the diagonal elements to be \( x \), \( h_{w} (H_{w}) \) be a vector (matrix) with each of its elements to be independent complex Gaussian random variables each having independent real and imaginary parts with zero-mean and variance of half, \( I \) be the identity matrix, \( B_x \) be a random unitary matrix, \( B_x^\dagger B_x = I \), generated by taking the eigenvectors of a random matrix \( H_x^\dagger H_x \).

II. SYSTEM MODEL

A downlink MIMO-BC with \( M(>1) \) transmit antennas and \( K \) mobile users each with \( N \) receive antennas is considered. The BS schedules transmission on a time-slot basis. The channel of each user is assumed to be constant during one slot time and vary from one slot time to another. Denote \( x^\dagger \) as the downlink \( M \times 1 \) transmit signal vector, and \( H_k \) as the \( N \times M \) multiuser channel gain matrix. The multiuser channel is homogeneous, i.e., \( H \in \mathcal{H}_w \).

The average receiver antenna signal-to-noise ratio is defined as \( SNR = P \), where \( P \) is the power constraint of the BS, i.e., \( \mbox{Tr}(\mathbb{E}[x^\dagger x]) = P \). During one slot time, the total delivered rate in bits/channel use from the BS is denoted as \( R^\dagger \), which is the sum rate of all active users if Dynamic SDMA is used or is the rate of one single active user if Dynamic TDMA is used, where the word “Dynamic” refers that the active user(s) may change for different slot times. The expected throughput of the system, \( C^\dagger \), is taken as the expected value of \( R^\dagger \) with respect to \( H \), i.e., \( C^\dagger = \mathbb{E}[H]R^\dagger \).

Let \( x^\dagger = \sum_{i=1}^{S} b_i u^\dagger_i \), where \( b_i \) and \( u^\dagger_i \) denote the downlink precoding vector and data symbol, respectively, for the \( i \)th data stream and \( S \) denotes the total number of independent data streams. Assume \( \| b_i \| = 1, \forall i \). Let \( u^\dagger = [u^\dagger_1, \ldots, u^\dagger_S]^T \), and \( \Lambda^\dagger \triangleq \mathbb{E}[u^\dagger u^\dagger] = \mbox{Diag}(p_1, \ldots, p_S) \), \( \sum_{i=1}^{S} p_i = P \). At the receiver, the decoding vector for the \( i \)th data stream is denoted by \( a^\dagger_i \), i.e., \( \hat{u}^\dagger_i = a^\dagger_i y^{\dagger}_{(k)} \), where \( k(i) \) denotes the index of the user to which the \( i \)th data stream is assigned.

The dual uplink MAC channel of the original BC in (1) can be written as

\[
y = \sum_{i=1}^{S} H^\dagger_{k(i)} x^\dagger_i + v \tag{2}
\]

where \( x^\dagger_i = a_i u^\dagger_i \) and \( v \in h_{w} \). Let \( \| a_i \| = 1, \forall i \), and \( u^\dagger = [u^\dagger_1, \ldots, u^\dagger_S]^T \), then \( \Lambda^\dagger \triangleq \mathbb{E}[u^\dagger u^\dagger] = \mbox{Diag}(q_1, \ldots, q_S) \).

The receiver decodes the signal of the \( i \)th data stream as \( \hat{u}^\dagger_i = b^\dagger_i y \). Therefore, the precoding vector (\( b_i \)) and decoding vector (\( a_i \)) of the original downlink BC becomes now the decoding vector (\( b^\dagger_i \)) and precoding vector (\( a_i \)) of the dual uplink channel, respectively, for the \( i \)th data stream.

III. DYNAMIC SDMA AND TDMA: COMPLETE CSI AT BS

A. Dynamic SDMA

In Dynamic SDMA, mobile terminals are physically separated and receiver cooperation between terminals is not possible. Therefore, precoding at the BS is considered. Both non-linear precoding, including optimal DPC and suboptimal “zero-forcing” DPC (ZF-DPC)[5], and linear precoding (also referred as BF), including optimal “minimum mean-square-error” BF (MMSE-BF) and suboptimal ZF-BF, are possible precoding techniques for Dynamic SDMA. This paper studies the precoder design and the associated expected throughput for the downlink MIMO-BC from the combined viewpoints of the generalized decision-feedback-equalization (GDFE) theory[13] and the generalized BC-MAC duality [6][7]. The correspondence between downlink precoder structures and their corresponding uplink GDFE decoder structures is established as follows:

- **MMSE-DFE decoder** assumes that in decoding the \( i \)th data stream, the interference signal from data streams \( 1 \) to \( i-1 \) are correctly decoded and subtracted from the received signal \( y \). Therefore, only interference signal from data streams \( i+1 \) to \( S \) are present. The unbiased MMSE-DFE decoding vector for the \( i \)th data stream, designed with the MMSE criteria (to minimize the MSE between the decoded symbol and the transmit symbol), becomes

\[
b_i = (I + \sum_{j=1}^{S} q_j H^\dagger_{k(j)} a_j a_j^\dagger H_{k(j)})^{-1} H^\dagger_{k(i)} a_i. \tag{3}
\]

The \( b_i \) obtained in (3) is also the downlink optimal **DPC-based precoder** [6].

- **ZF-DPC decoder** estimates the \( i \)th data stream by first subtracting the interference signal from decoded data streams \( 1 \) to \( i-1 \) (as in MMSE-DFE decoding) and then filtering out the interference signal from data streams \( i+1 \) to \( S \) by the ZF criteria, i.e.,

\[
b_i = \arg \max_{x \in \mathcal{A}_i} x^\dagger H^\dagger_{k(i)} a_i \tag{4}
\]

where \( \mathcal{A}_i = \{ x : \| x \| = 1, x^\dagger H^\dagger_{k(j)} a_j = 0, \forall j = i + 1, \ldots, S \} \). The \( b_i \) in (4) is also the **ZF-DPC precoder**[5].

- **MMSE decoder** in stead of using successive decoding in MMSE-DFE, decodes each data stream by treating the signal from all the other data streams as interference. The unbiased decoding vector for the \( i \)th data stream designed with the MMSE criteria can be written as

\[
b_i = (I + \sum_{j=1,j \neq i}^{S} q_j H^\dagger_{k(j)} a_j a_j^\dagger H_{k(j)})^{-1} H^\dagger_{k(i)} a_i. \tag{5}
\]
The $b_i$ obtained in (5) is also known to be the optimal downlink linear precoder, i.e., **MMSE precoder**[6].

- **ZF decoder** in stead of using successive decoding in ZF-DFE, simultaneous decodes each data stream by filtering out the interference from all the other data streams using the ZF criteria, i.e.,

\[ b_i = \arg \max_{x \in B_i} x^T H_{k(i)}^T a_i, \tag{6} \]

where $B_i = \{ x : ||x|| = 1, x^T H_{k(j)}^T a_j = 0, \forall j = 1, \ldots, S, j \neq i \}$. The $b_i$ in (6) is also the well-known downlink **ZF precoder**.

It can be verified that the ESNR for the decoded $i$th data stream in the dual MAC is equal to

\[ ESNR_{i}^{ul} = q_i ||b_i^T H_{k(i)}^T a_i||^2. \tag{7} \]

And the following result hold:

**Proposition 1:** There exists a pair of matrix, $\Lambda^{ul}$ and $\Lambda^{dl}$,

\[ \text{Tr}(\Lambda^{ul}) = \text{Tr}(\Lambda^{dl}) = P, \] such that $ESNR_{i}^{ul} = ESNR_{i}^{dl} = ESNR_i$, for $i = 1, \ldots, S$, for any of GDFE precoder and its dual decoder structures. Furthermore, the sum of transmit rates from all mobile terminals in the dual uplink MAC, $R^{ul}$, is equal to the total rate delivered by the BS in the downlink BC, $R^{dl}$, i.e.,

\[ R^{ul} = R^{dl} = \sum_{i=1}^{S} \log_2(1 + ESNR_i). \tag{8} \]

**Proof:** For ZF-DFE decoder / ZF-DPC precoder and ZF decoder / ZF precoder, it can be easily verified if $\Lambda^{ul} = \Lambda^{dl}$, $ESNR_{i}^{ul} = ESNR_{i}^{dl}$, $\forall i$. For MMSE-DFE decoder / optimal DPC precoder and MMSE decoder / MMSE precoder, it is shown in [6] that $ESNR_{i}^{ul} = ESNR_{i}^{dl}$, $\forall i$, but $\Lambda^{ul} \neq \Lambda^{dl}$ in general. \( \square \)

**Remark 1:** In the dual uplink MAC, the achievable throughput, $R^{ul}$, in (8), associated with different decoder structures can be easily compared. For example, given other parameters identical, the optimal MMSE-DFE decoder has the largest $R^{ul}$ than the other three decoder structures. And MMSE decoder has a larger $R^{ul}$ than ZF decoder and so does ZF-DFE decoder than ZF decoder. As a result, by **Proposition 1**, the throughput of the original downlink BC, $R^{dl}$, associated with corresponding dual decoder structures can be compared.

**Remark 2:** $R^{dl}$ can be maximized through jointly optimizing parameters including $a_i$ (decoding vectors), $b_i$ (precoding vectors), $\Lambda^{dl}$ (power allocation), $k(i)$ (user selection and ordering) and $S$ (number of data streams). For example, in [7], $R^{dl}$ associated with the optimal DPC precoder is maximized and the optimal number of data streams, $S$, is known to be at most $M^2$, but in general greater than $M$ [2].

**Remark 3:** $R^{dl}$ associated with MMSE decoder, given $P_i$ and $S$, can be easily maximized by modifying the iterative algorithm in [12]. However, the optimization of $P_i$ and $S$ still remains unknown.

**Remark 4:** $R^{dl}$ associated with corresponding ZF-DFE decoder or ZF decoder, given $a_i$, can be easily maximized by standard water-filling algorithm [1]. Due to the dimensionality constraint, the optimal number of data streams, $S$, is at most $M$, and therefore can be obtained through $O(\frac{M}{k})$ numbers of searching. However, the optimization of $a_i$ still remains unknown for $N > 1$.

### B. Dynamic TDMA

Dynamic TDMA selects only one user for transmission at one time. Consider now the $k$th user, $k = 1, \ldots, K$, for transmission and because the user’s channel, $H_k$, is assumed to be known at both the BS and the terminal, the optimal precoder/decoder structures for the downlink transmission are determined by the singular value decomposition (SVD) of $H_k$ [1], i.e., $H_k = V_k D \text{diag}(\gamma_{k,1}, \ldots, \gamma_{k,S}) U_k^T$, where $S = \min(M, N)$ and $\gamma_{k,1} \geq \ldots \geq \gamma_{k,S} > 0$. Let $B \triangleq [b_1, \ldots, b_S] = U_k$ and $A \triangleq [a_1, \ldots, a_S] = V_k$. The decoded signal at the terminal can be written as $\hat{a}_i^k = a_i^k y_k = \gamma_{k,i} x_i^k + z_k,i$ and the downlink rate delivered by the BS to the $k$th user becomes

\[ R^{dl}(k) = \sum_{i=1}^{S} \log_2(1 + p_i \gamma_{k,i}^2). \tag{9} \]

And the power allocation to each data streams, $[p_i]$, can be further optimized by standard water-filling algorithm to maximize $R^{dl}(k)$ in (9). The BS selects the $k$th user for transmission if it has the highest $R^{dl}(k)$ among all the users.

### IV. DYNAMIC SDMA AND TDMA: PARTIAL CSI AT BS

In this section, it is assumed that the knowledge of each user’s channel is completely known at the terminal, but unknown at the BS. It is also assumed that there exists a reliable uplink from each terminal to the BS such that each terminal is able to feed back its own CSI to the BS. In [9], a technique so-called “Opportunistic Beamforming (OFB)” that combines random linear precoding and ESNR feedback is proposed for a Dynamic TDMA system with $N = 1$. In this Section, this technique is studied for a general MIMO BC with $M \geq 1$ and $N \geq 1$. A general framework of OFB can be described as follows. Each time slot of transmission consists of two modes: **Pilot Transmission Mode** and **Data Transmission Mode**. In the Pilot Transmission Mode, the BS first determines the number of data streams, $S$, and then selects a random precoding matrix $B(\in B_r)$ to modulate the pilot signal, $u_{pilot}^{dl}$, which is assumed to be known at all mobile terminals. Each user then estimates its equivalent composite channel, $H_{eq,k}$, which is the product of its own physical channel and the random precoding matrix, i.e., $H_{eq,k} = H_k B$, and then determines the ESNR of each data stream and feeds back them to the BS. In the Data Transmission Mode, the BS then selects the user(s) for transmission based on the ESNR feedback from users. There are two important consequences associated with selecting a random precoding matrix, $B$, for each time slot: (1) The distribution of $H_{eq,k}$ can be different from that of the original physical channel $H_k$ and this would have an effect on the user selection. (2) If the original physical channel, $H_k$, is slowly time-varying, i.e., each of its elements is highly...
correlated from one time slot to another, the equivalent channel $\mathbf{H}_{eq,k}$ generated by a random precoding matrix is different for each time slot and therefore prevents one user from persistently capturing the channel for transmission, i.e., random precoding has an inherent fair scheduling mechanism to ensure that each user is assigned with some time slots for transmission even in the short-term (relative to the channel coherence time).

A. Dynamic SDMA

For Dynamic SDMA, the scheme presented in this subsection is similar to the one in [11]. In the Pilot Transmission Mode, the BS first broadcasts the pilot signal modulated by the random precoding matrix with $\mathbf{S}(\leq M)$ independent data streams. The $k$th user then determines the ESNR of each transmitted data stream with the unbiased MMSE decoding vector, i.e.,

$$a_{k,i} = (\mathbf{I} + \sum_{j=1, j\neq i}^{S} p_j \mathbf{H}_k^j \mathbf{b}_j \mathbf{H}_k)^{-1} \mathbf{H}_k \mathbf{b}_i. \quad (10)$$

and the ESNR for the $i$th data stream becomes

$$ESNR_{k,i} = p_i ||a_{k,i}^* \mathbf{H}_k \mathbf{b}_i||^2. \quad (11)$$

The BS, after receiving the ESNR feedback from all users, assigns the $i$th data stream to the user $k(i)$, where $k(i) = \{k : ESNR_{k,i} > ESNR_{k',i}, k' = 1, \ldots, K \}$. Let $ESNR_i \triangleq ESNR_{k(i)}$. The total delivered rate of the BS can be also written as in (8).

B. Dynamic TDMA

For Dynamic TDMA, each user determines the set of ESNRs for all data streams and then feeds back to the BS. The BS then determines the achievable sum-rate for each user and selects the user that has the largest sum-rate for transmission. The set of ESNRs for each user are determined by the detector structure used at the mobile terminal. If the equivalent channel of each user, $\mathbf{H}_{eq,k}$, is considered, the single-user MIMO channel in Dynamic TDMA with random precoding can be considered equivalently as a MAC in (2) with $N = 1$, $a_i = 1 \forall i$, $S = M$ and $\mathbf{H}_{k(i)} \rightarrow \mathbf{H}_{eq,k}$. Therefore, the detector structure described in Section III-A can be directly employed at each mobile terminal and the set of ESNRs in (7) and the sum-rate in (8), associated with each detector structure in a MAC, are the same for the single-user MIMO transmission in Dynamic TDMA with independent data streams each being considered as one virtual use as in the MAC.

Remark 1: Comparing Dynamic TDMA with random precoding and optimal precoding when complete CSI is known at the BS in Section III-B, it is noted that if the random precoding matrix $\mathbf{B}$ occurs to be very “similar” to $\mathbf{U}_k$ obtained from the SVD decomposition of $\mathbf{H}_k$, for the $k$th user, the sum-rate associated with random precoding approaches to that with the optimal precoding. This is the rational behind random precoding, i.e., when the number of users, $K$, becomes very large, there is a high probability that the selected random precoding matrix matches the true optimal precoding matrix for some user (though this user might not be the true user selected in optimal precoding). Ideally, if $\mathbf{B} = \mathbf{U}_k$, it can be verified that all GDFE decoder structures in Section III-A becomes identical, i.e., $\mathbf{A} = \mathbf{V}_k$.

Remark 2: If the number of users, $K$, is finite, Dynamic TDMA with random precoding suffers from a throughput loss compared with optimal precoding. In fact, the maximum sum-rate achievable with random precoding can be expressed as

$$R_{eq}(k) = \max_{\Lambda_{eq}} \log_2 |\mathbf{I} + \mathbf{H}_{eq,k} \Lambda_{eq} \mathbf{H}_{eq,k}^\dagger| \quad (12)$$

$R_{eq}(k)$ in (12) is achieved by the MMSE-DFE decoder in Section III-A with the optimal power allocation, $\Lambda_{eq}$. Unfortunately, with only ESNR feedback, this sum-rate is not obtainable. The reason is as follows: At the Pilot Transmission Mode, the optimal $\Lambda_{eq}$ is unknown at the BS, therefore, equal power allocation is assumed, i.e., $\Lambda_{eq} = \frac{1}{K} \mathbf{I}$. Each user then measures the set of ESNRs for each data streams, based on the GDFE decoder structure employed and feeds back them to the BS. If the decoder structure is MMSE-DFE, it is not possible for the BS to solve for the optimal power allocation in (12) for the Data Transmission Mode, given only the set of ESNRs. In contrast, if ZF-DFE or linear ZF decoder is used, it is possible for the BS to re-allocate the power to data streams based on the set of feedback ESNRs and by using the standard water-filling algorithm. Therefore, $R_{eq}(k)$ in (12) can only serve as a throughput upper-bound for Dynamic TDMA with random precoding and ESNR feedback.

V. NUMERICAL EXAMPLES

Monte Carlo simulation is used to compare the expected throughput, $C_{eq}$, associated with different transmission schemes described in the paper. For all schemes, the optimal set of user(s) for maximizing $C_{eq}$ is selected. Fig.1 shows $C_{eq}$ versus the number of users for 4 schemes, including Dynamic SDMA with DPC precoder and random precoder, Dynamic TDMA with SVD-based precoder and random precoder (using Eq.12). For Dynamic SDMA with random precoder, it is assumed that $\mathbf{S} = \mathbf{M}$ and $\Lambda^{eq} = \frac{1}{M} \mathbf{I}$. For the other three schemes, $\mathbf{S}$ and $\Lambda^{eq}$ are optimized. It is observed that for the case of complete CSI, Dynamic SDMA with DPC precoder achieves a larger expected throughput than Dynamic TDMA because Dynamic SDMA can fully exploit the multiuser diversity by transmitting to the optimal set of users. The throughput gain is more substantial when $M > N$. In contrast, for the case of partial CSI and random precoding, it is possible that Dynamic SDMA performs worse than Dynamic TDMA. The reason is that when random precoding is used, the throughput of Dynamic SDMA is limited by the interference between data streams and it is not possible for active users to jointly mitigate the interference because they are physically separated. In Dynamic TDMA, a single user at one time is scheduled for transmission and therefore the BS is not able to fully utilize the multiuser diversity. However, a joint decoding of different data streams is possible at each user’s terminal in order to mitigate the interference among data streams. Fig.2 shows the expected throughput of Dynamic SDMA with different GDFE precoder structures. For ZF-DPC, linear MMSE and linear ZF
strategy for both dynamic schemes tends to be transmitting to one user at one time. The channel model assumed in the paper does not consider the possible receiver antenna correlation at each terminal, which can result in a further throughput gain of Dynamic SDMA than TDMA because Dynamic SDMA can transmit to users that are physically separated and therefore can explore the multiuser diversity gain even with receiver antenna correlation at each terminal.

**REFERENCES**


