Techniki MIMO komórkowe i bezkomórkowe w sieci 6G

MU-MIMO (Multi-User Multiple-Input Multiple-Output) odnosi się do scenariusza wdrożenia, w którym stacja bazowa wyposażona w wiele anten obsługuje wiele terminali za pomocą jednej lub kilku anten. Poprzez rozdzielenie strumieni multipleksowanych przestrzennie pomiędzy wiele terminali, MU-MIMO zyskuje przewagę nad SU-MIMO dzięki trzem podstawowym zaletom. Po pierwsze, ułatwia korzystanie z terminali o niskiej złożoności, niskich kosztach i energooszczędnych. Po drugie, jest mniej podatny na środowiska propagacyjne ze względu na przestrzenne rozmieszczenie terminali, nawet w warunkach widoczności. Po trzecie, analizy teorii informacji ujawniają, że sumaryczna szybkość transmisji wielu użytkowników jest wyższa niż przepustowość kanału komunikacji pojedynczego użytkownika. Niemniej jednak konwencjonalny MU-MIMO jest nadal trudny do skalowania w celu multipleksowania przestrzennego wyższego rzędu, ponieważ kodowanie wstępne (np. kodowanie brudnego papieru) i dekodowanie osiągające pojemność narzucają wykładniczo rosnącą złożoność. Co najważniejsze, nadajnik wymaga znajomości kanału łącza w dół, a zasoby wydane na pozyskiwanie informacji o stanie kanału rosną wraz z liczbą anten usługowych. Rewolucyjna technika zwana massive MIMO przełamuje tę barierę skalowalności, nie próbując osiągnąć pełnego limitu Shannona i paradoksalnie zwiększając rozmiar systemu. W massive MIMO tylko stacja bazowa uczy się wiedzy o kanale do kodowania wstępnego w łączu w dół i dekodowania w łączu w górę, podczas gdy terminale nie muszą tego wiedzieć. Korzystając z wzajemności kanałów operacji Time Division Duplex (TDD), narzut wymagany do pozyskania CSI zależy od liczby terminali. Dlatego po stronie stacji bazowej można zainstalować dużą tablicę antenową, tak aby liczba anten usługowych była zazwyczaj kilkakrotnie większa od liczby aktywnych użytkowników, podczas gdy skala użytkowników pozostaje niewielka, co sprawia, że złożoność implementacji jest niska. Jedną z zalet korzystania z nieograniczonej liczby anten usługowych jest wzmocnienie kanału, w którym efekty nieskorelowanego szumu odbiornika i szybkiego zanikania są całkowicie eliminowane. Jednak wysoką wydajność w kolokowanym MIMO osiągają przede wszystkim użytkownicy, którzy pozostają w pobliżu centrum komórki. Większość użytkowników na skraju komórki ogranicza się do znacznie gorszej jakości usług z powodu zakłóceń międzykomórkowych. Dlatego zaproponowano rozproszony system Massive MIMO zwany bezkomórkowym Massive MIMO, w którym duża liczba anten usługowych jest losowo rozproszona na dużym obszarze. Wszystkie anteny współpracują fazowo spójnie za pośrednictwem sieci fronthaul i obsługują wszystkich użytkowników w tym samym zasobie czasowo-częstotliwościowym. Nie ma komórek ani granic komórek. Ponieważ ta konfiguracja łączy koncepcje distributedMIMO i massive MIMO, oczekuje się, że odniesie wszystkie korzyści z tych dwóch systemów. Rozdział ten składa się głównie z:

● Teoretycznego wprowadzenia do MU-MIMO, tj. analiz sumarycznej pojemności kanałów rozgłoszeniowych MIMO i kanałów wielodostępowych MIMO.

● Podstaw znanego kodowania brudnego papieru, które może osiągnąć pełną pojemność, oraz zasad jego suboptymalnych odpowiedników o niskiej złożoności, zwanych prekodowaniem zerowym i diagonalizacją bloków.

● Podstawowej konfiguracji massive MIMO, w tym pozyskiwania wiedzy o kanale, liniowego prekodowania w łączu w dół i liniowego wykrywania w łączu w górę.

● Zanieczyszczenia pilota w wielokomórkowych systemach massive MIMO oraz modeli systemowych transmisji danych w łączu w dół i w górę.

● Układu sieci massive MIMO bez komórek, pozyskiwania CSI za pośrednictwem szkolenia w łączu w górę i transmisji danych w łączu w górę.

● Sprzężonego kształtowania wiązki i prekodowania zerowego w sieci massive MIMO bez komórek oraz wpływu starzenia się kanału na wydajność.

Multi-User MIMO

Przestrzenne multipleksowanie omówione w poprzedniej sekcji jest często określane jako Multiple-Input Multiple-Output (MIMO), co odzwierciedla fakt, że wiele równoległych strumieni danych jest jednocześnie przesyłanych na tej samej częstotliwości do jednego odbiornika. Wykorzystanie wielu anten zarówno w nadajniku, jak i odbiorniku w połączeniu z przetwarzaniem prekodowania i wykrywania ma na celu oddzielenie przestrzennie multipleksowanych sygnałów i stłumienie zakłóceń między różnymi warstwami transmisyjnymi. Ta technika ma bardziej szczegółowy termin Single-User MIMO lub SU-MIMO z powodów, które staną się jasne w poniższym tekście. Jako bezpośrednie rozszerzenie przestrzennego multipleksowania, równoległe warstwy transmisyjne utworzone przez wiele anten nadawczych mogą być przeznaczone dla różnych odbiorników z jedną lub kilkoma antenami odbiorczymi i odwrotnie. W kontekście systemów komunikacji mobilnej lub bezprzewodowych sieci lokalnych termin Multi-User MIMO (MU-MIMO) odnosi się do scenariusza wdrożenia, w którym stacja bazowa lub punkt dostępowy wyposażony w wiele anten nadawczych komunikuje się z wieloma terminalami. Zestaw terminali z jedną lub kilkoma antenami może utworzyć wirtualną tablicę do kultywowania multipleksowania przestrzennego wraz ze stacją bazową z wieloma antenami. Ze względu na stosunkowo dużą zdolność przetwarzania sygnału i wystarczające zasilanie, strona stacji bazowej ponosi ciężar przestrzennego rozdzielenia równoległych strumieni. W ten sposób stacja bazowa wykonuje prekodowanie lub przesyła formowanie wiązki w kierunku wielu użytkowników w łączu w dół i wykrywanie wielu użytkowników w łączu w górę. W konsekwencji, niezwykłą zaletą MU-MIMO w porównaniu z SU-MIMO jest to, że zysk multipleksowania przestrzennego jest zachowany nawet w przypadku tanich terminali z małą liczbą anten. Jest to konieczny wymóg osiągnięcia ekonomii skali w branży mobilnej. Inna podstawowa różnica między MU-MIMO i SU-MIMO wynika z różnicy w kanale bazowym. Osiągnięcie zysku multipleksowania przestrzennego w dużym stopniu zależy od dobrze uwarunkowanych kanałów. W systemie SU-MIMO dekorelacja między sygnaturami przestrzennymi anten wymaga bogatych środowisk rozpraszania z dużym odstępem między antenami lub stosowania polaryzacji anteny. W systemie MU-MIMO dekorelacja między sygnaturami przestrzennymi różnych terminali występuje naturalnie, ponieważ te terminale są bardzo rozproszone. Użytkownicy są rozdzieleni geograficznie, sygnał rozprzestrzenia się w różnych kierunkach, nawet jeśli rozpraszanie w środowisku jest ograniczone. Ponadto w takim układzie dostępny jest zysk różnorodności wielu użytkowników oprócz zysku multipleksowania przestrzennego. Niemniej jednak osiągnięcie potencjału MU-MIMO zależy od dokładnych informacji o stanie kanału (CSI) w nadajniku. Wykazano, że niewielka ilość sprzężenia zwrotnego może być bardzo korzystna w kierowaniu mocy w stronę anten odbiornika. Dokładniej rzecz biorąc, dokładność CSI w SU-MIMO powoduje jedynie karę stosunku sygnału do szumu (SNR), ale nie wpływa na zysk multipleksowania. Jednakże dokładność CSI dostępna w nadajniku ma wpływ na wzmocnienie multipleksowania systemu MU-MIMO. Dlatego też istotne jest dokładne i terminowe dostarczanie CSI do nadajnika, co zawsze jest wyzwaniem ze względu na ograniczenia zasobów sprzężenia zwrotnego lub powagę kanałów bezprzewodowych.

Kanały rozgłoszeniowe i wielodostępowe

SU-MIMO to symetryczny system typu punkt-punkt, który można zatem opisać za pomocą nadajnika i odbiornika, i nie trzeba w nim rozróżniać łącza w dół i łącza w górę. Natomiast MU-MIMO to system asymetryczny, w którym transmisja w łączu w dół ze stacji bazowej do kilku terminali jest określana jako kanał rozgłoszeniowy Gaussa MIMO, podczas gdy transmisja w łączu w górę z kilku terminali do stacji bazowej jest określana jako wielodostępowy kanał Gaussa MIMO. W systemie MU-MIMO

wybiera się K terminali do jednoczesnej komunikacji ze stacją bazową w tym samym zasobie czasowo-częstotliwościowym. Typowy terminal k jest wyposażony w Nk, k = 1, 2,…, K anten, a zatem terminale te mają łącznie $N_u = \Sigma^K_{k=1} N_k$ anten po stronie terminala. Załóżmy, że stacja bazowa ma Nb anten, system tworzy kanał $N_u \times N_b$ w łączu w dół systemu komórkowego, podczas gdy kanał $N_b \times N_u$ w łączu w górę. Strona stacji bazowej może obsługiwać do $N_m = \min(N_b, N_u)$ równoległych strumieni, podczas gdy typowy terminal k jest przypisany do L_k strumieni, spełniając $L_k \leqslant \min(N_k, N_b)$ i równoważnie $\Sigma^K_{k=1} L_k \leqslant N_m$. Takie wielożytkownikowe ustawienie MIMO można oznaczyć jako system $([N_1, N_2, \ldots, N_K], N_b)$ dla łącza w dół lub $(N_b, [N_1, N_2, \ldots, N_K])$ dla łącza w górę. Najpierw przyjrzyjmy się transmisji łącza w górę, gdzie K terminali jednocześnie nadaje w kierunku stacji bazowej. Piszemy $H^{(k)}_{ul} \in \mathbb{C}^{Nb \times Nk}$, aby modelować macierz kanałów od k-tego użytkownika do stacji bazowej. Każdy wpis oznacza wzmocnienie kanału od anteny nadawczej w terminalu do anteny odbiorczej w stacji bazowej. Można założyć, że aktywne terminale są losowo rozmieszczone w komórce, a anteny terminala są wystarczająco rozstawione lub spolaryzowane, co skutkuje niezależnymi kanałami zanikającymi. W płaskich kanałach zanikających Rayleigha współczynnik kanału jest oznaczany przez losową wariancję zespoloną Gaussa o symetrii kołowej z zerową średnią i wariancją jednostkową, mianowicie $h \sim \mathcal{CN}(0,1)$. Kanał MIMO z wieloma dostępami można modelować jako

$$\mathbf{r} = \sum_{k=1}^{K} \mathbf{H}_{ul}^{(k)} \mathbf{s}_k + \mathbf{n}, \quad (1)$$

gdzie $r \in \mathbb{C}^{Nb \times 1}$, $s_k \in \mathbb{C}^{Nk \times 1}$ i $n \in \mathbb{C}^{Nb \times 1}$ oznaczają odpowiednio wektor odebranych symboli, wektor przesłanych symboli na terminalu k i wektor szumu. Moc transmisji na k-tym terminalu jest ograniczona przez $\mathbb{E}[s^H_k s_k] \leqslant P_k$ lub równoważnie $\mathrm{tr}(\mathbb{E}[s_k s^H_k]) \leqslant P_k$, podczas gdy szum na antenę odbiornika jest niezależnym zespolonym szumem Gaussa o zerowej średniej i wariancji σ^2_n, mianowicie $\mathbf{n} \sim \mathcal{CN}(0, \sigma_n^2 \mathbf{I}_{N_b})$. Oznaczamy przez $u_k \in \mathbb{C}^{Lk \times 1}$ wektor symboli informacyjnych od użytkownika k, który jest przekształcany do sk zgodnie ze wzorem

$s_k = T_k u_k$, (2),

gdzie $T_k \in \mathbb{C}^{Nk \times Lk}$ oznacza macierz prekodowania użytkownika k. Używając $H_{ul} \in \mathbb{C}^{Nb \times Nu}$ do oznaczenia ogólnego kanału wielodostępu MIMO w systemie MU-MIMO, mamy

$$\mathbf{H}_{ul} = [\mathbf{H}_{ul}^{(1)}, \mathbf{H}_{ul}^{(2)}, \ldots, \mathbf{H}_{ul}^{(K)}]. \quad (3)$$

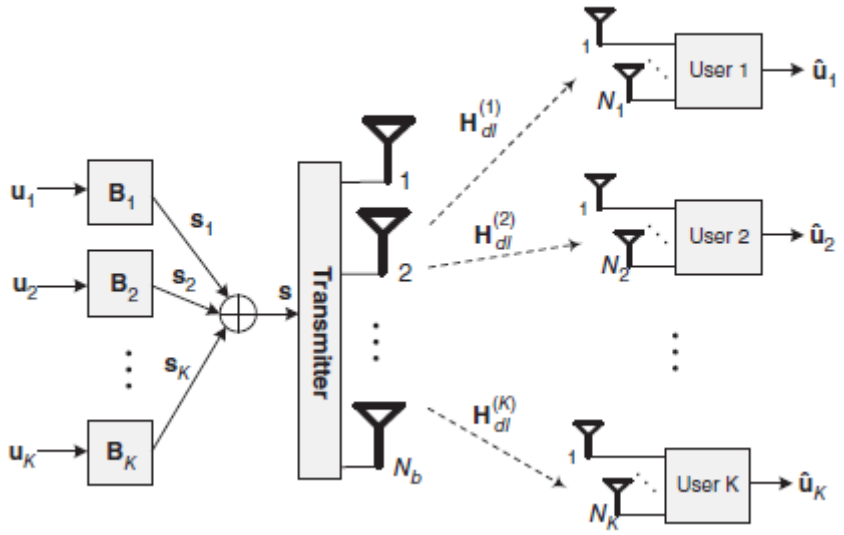
Budowanie wektora zawierającego wszystkie symbole transmisji z terminali K

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}_1 \\ \vdots \\ \mathbf{s}_K \end{bmatrix} \quad (4)$$

Równanie (1) można zapisać w postaci

$$\mathbf{r} = \sum_{k=1}^{K} \mathbf{H}_{ul}^{(k)} \mathbf{s}_k + \mathbf{n} = \sum_{k=1}^{K} \mathbf{H}_{ul}^{(k)} \mathbf{T}_k \mathbf{u}_k + \mathbf{n} = \mathbf{H}_{ul} \mathbf{s} + \mathbf{n}, \quad (5)$$

co jest równoważne modelowi systemu SU-MIMO z antenami nadawczymi N_u i antenami odbiorczymi N_b. Innymi słowy, typowy terminal k nadaje s_k do stacji bazowej, która generuje odebrany składnik $r_k = H^{(k)}_{ul}s_k$, co prowadzi do całkowitego odebranego sygnału $r = \Sigma^K_{k=1} r_k$. Następnie skupmy się na transmisji downlink, gdzie stacja bazowa wysyła sygnały w kierunku K terminali przez ten sam zasób czasowo-częstotliwościowy, jak pokazano na rysunku.



Cały system można modelować jako

$r = H_{dl}s + n$, (6),

gdzie $H_{dl} \in \mathbb{C}^{Nu \times Nb}$ oznacza macierz kanału pomiędzy Nb antenami nadawczymi w stacji bazowej i N_u antenami odbiorczymi rozłożonymi na K terminalach, $r \in \mathbb{C}^{Nu \times 1}$, $s \in \mathbb{C}^{Nb \times 1}$ i $n \in \mathbb{C}^{Nu \times 1}$ oznaczają odpowiednio wektor wszystkich odebranych symboli, wektor symboli transmitowanych w stacji bazowej i wektor szumu. Ograniczenie mocy w stacji bazowej jest wyrażone przez $\mathbb{E}[s^H s] \leqslant P$ lub równoważnie $tr(\mathbb{E}[ss^H]) \leqslant P$.

Rozkładając równanie (6) jako

$$\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_K \end{bmatrix} = \begin{bmatrix} \mathbf{H}^{(1)}_{dl} \\ \mathbf{H}^{(2)}_{dl} \\ \vdots \\ \mathbf{H}^{(K)}_{dl} \end{bmatrix} \mathbf{s} + \begin{bmatrix} \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_K \end{bmatrix}, \quad (7)$$

gdzie $r_k \in \mathbb{C}^{Nk \times 1}$ i $n_k \in \mathbb{C}^{Nk \times 1}$ reprezentują wektor odebranych symboli i wektor szumu, odpowiednio, w terminalu k, a $H^{(k)}_{dl} \in \mathbb{C}^{Nk \times Nb}$ modeluje kanał od stacji bazowej do k-tego użytkownika, który jest podmacierzą składającą się z N_k wierszy H_{dl}. Wtedy wiemy, że indywidualny model dedykowany typowemu terminalowi k jest dany przez

$$\mathbf{r}_k = \mathbf{H}^{(k)}_{dl}\mathbf{s} + \mathbf{n}_k, \quad k = 1, 2, \ldots, K, \quad (8)$$

stosując $r = [r^T_1, r^T_2, \ldots, r^T_K]^T$, $n = [n^T_1, n^T_2, \ldots, n^T_K]^T$, i

$$\mathbf{H}_{dl} = \begin{bmatrix} \mathbf{H}_{dl}^{(1)} \\ \vdots \\ \mathbf{H}_{dl}^{(K)} \end{bmatrix}. \quad (9)$$

Dla uproszczenia używamy również $u_k \in \mathbb{C}^{Lk\times 1}$ do oznaczenia wektora symboli informacyjnych przeznaczonych dla użytkownika k, które mogą być wstępnie kodowane indywidualnie zgodnie z

$s_k = B_k u_k$, (10)

gdzie $Bk \in \mathbb{C}^{Nb\times Lk}$ oznacza macierz wstępnego kodowania dedykowaną użytkownikowi k w stacji bazowej, a s_k jest składową całkowitego sygnału przesyłanego s przez użytkownika k, spełniającą

$$\mathbf{s} = \sum_{k=1}^{K} \mathbf{s}_k = \sum_{k=1}^{K} \mathbf{B}_k \mathbf{u}_k. \quad (11)$$

Alternatywnie, przesyłany sygnał może być generowany przez wspólne kodowanie wstępne

s= Bu (12)

gdzie $B \in \mathbb{C}^{Nb\times L}$ jest macierzą prekodowania dla wszystkich symboli informacyjnych $u \in \mathbb{C}^{L\times 1} = [u^T_1 , u^T_2 ,\ldots, u^T_K]^T$ przy całkowitej liczbie strumieni danych $L = \Sigma^K_{k=1} L_k$. Łatwo wywnioskować, że

$$\mathbf{B} = [\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_K]. \quad (13)$$

Wówczas równanie (8) można wyrazić także za pomocą

$$\mathbf{r}_k = \mathbf{H}_{dl}^{(k)}\mathbf{s} + \mathbf{n}_k = \mathbf{H}_{dl}^{(k)}\mathbf{B}\mathbf{u} + \mathbf{n}_k = \mathbf{H}_{dl}^{(k)} \sum_{k=1}^{K} \mathbf{B}_k \mathbf{u}_k + \mathbf{n}_k. \quad (14)$$

Podobnie równanie (6) można zapisać w postaci

$$\mathbf{r} = \mathbf{H}_{dl}\mathbf{s} + \mathbf{n} = \mathbf{H}_{dl}\mathbf{B}\mathbf{u} + \mathbf{n} = \mathbf{H}_{dl} \sum_{k=1}^{K} \mathbf{B}_k \mathbf{u}_k + \mathbf{n} \quad (15)$$

Suma pojemności wielu użytkowników

W systemie typu punkt-punkt pojemność kanału stanowi miarę limitu wydajności: niezawodna komunikacja z dowolnie małym prawdopodobieństwem błędu może być osiągnięta przy dowolnej szybkości R < C, podczas gdy niezawodna komunikacja jest niemożliwa, gdy R > C. W przypadku systemu wielożytkownikowego składającego się ze stacji bazowej i K terminali, koncepcja ta jest rozszerzana na podobną metrykę wydajności zwaną regionem pojemności. Charakteryzuje się ona przestrzenią K-wymiarową $\mathfrak{C} \in \mathbb{R}^K_+$, gdzie $\mathbb{R}_+$ oznacza zbiór nieujemnych liczb rzeczywistych, a $\mathfrak{C}$ jest zbiorem wszystkich K-krotek (R_1, R_2,…, R_K) tak, że ogólny użytkownik k może niezawodnie komunikować się z szybkością R_k jednocześnie z innymi. Ze względu na współdzielone zasoby transmisji istnieje kompromis: jeśli ktoś chce wyższej szybkości, niektórzy inni użytkownicy muszą obniżyć swoje szybkości. Z tego regionu pojemności można wyprowadzić metrykę wydajności, tj. sumę pojemności

$$C_{\text{sum}} = \max_{(R_1,R_2,\ldots,R_K)\in\mathfrak{C}} \left(\sum_{k=1}^{K} R_k \right) \quad (16)$$

wskazując maksymalną całkowitą przepustowość, jaką można osiągnąć. Użyjmy najprostszego systemu wielodostępnego składającego się z odbiornika z pojedynczą anteną i dwóch użytkowników wyposażonych w pojedynczą antenę nadawczą. Użytkownicy 1 i 2 wysyłają transmitowane symbole s_1 i s_2 do odbiornika w kanale AWGN łącza w górę. Odebrany symbol to

$$r = s_1 + s_2 + n, \quad (17)$$

gdzie ograniczenia mocy dla s_1 i s2 wynoszą odpowiednio P_1 i P_2, a $n \sim \mathcal{CN}(0, \sigma_n^2)$ jest zespolonym szumem Gaussa. Współczynniki użytkowników 1 i 2 wynoszą odpowiednio R_1 i R_2, tworząc następujący obszar pojemności

$$\mathfrak{C} = \left\{ (R_1, R_2) \in \mathbb{R}_+^2 \;\middle|\; \begin{array}{c} R_1 < \log_2\left(1 + \frac{P_1}{\sigma_n^2}\right) \\ R_2 < \log_2\left(1 + \frac{P_2}{\sigma_n^2}\right) \\ R_1 + R_2 < \log_2\left(1 + \frac{P_1 + P_2}{\sigma_n^2}\right) \end{array} \right\} \quad (18)$$

Pierwsze ograniczenie w równaniu (18) oznacza, że osiągalna szybkość użytkownika 1 jest ograniczona przez system pojedynczego użytkownika, w którym użytkownik 2 jest nieobecny. Podobnie drugie ograniczenie wskazuje ograniczenie pojedynczego użytkownika dla użytkownika 2. Trzecie ograniczenie mówi, że suma szybkości nie może przekroczyć pojemności systemu typu punkt-punkt z mocą sygnału odebranego równą sumie mocy sygnału odebranego tych dwóch użytkowników. Szczególnie interesujące jest to, że jeden użytkownik może osiągnąć ograniczenie pojedynczego użytkownika, podczas gdy inny użytkownik może jednocześnie nadawać z szybkością różną od zera. Jest to zaleta systemu wielożytkownikowego w porównaniu z systemem pojedynczego użytkownika. Osiąga się ją poprzez kolejne usuwanie zakłóceń lub SIC wykonane w dwóch krokach. W pierwszym kroku odbiornik wykrywa symbol użytkownika 1, traktując sygnał od użytkownika 2 jako kolorowy szum. Osiągnięta szybkość dla użytkownika 1 wynosi

$$R_1 = \log_2\left(1 + \frac{P_1}{P_2 + \sigma_n^2}\right) \quad (19)$$

Odejmując s_1 od r, odbiornik może wykryć s_2 tylko za pomocą szumu białego addytywnego. W tym przypadku

$$R_2 = \log_2\left(1 + \frac{P_2}{\sigma_n^2}\right) \quad (20)$$

Suma stawek jest równa

$$C_{\text{sum}} = R_1 + R_2 = \log_2\left(1 + \frac{P_1}{P_2 + \sigma_n^2}\right) + \log_2\left(1 + \frac{P_2}{\sigma_n^2}\right) \quad (21)$$

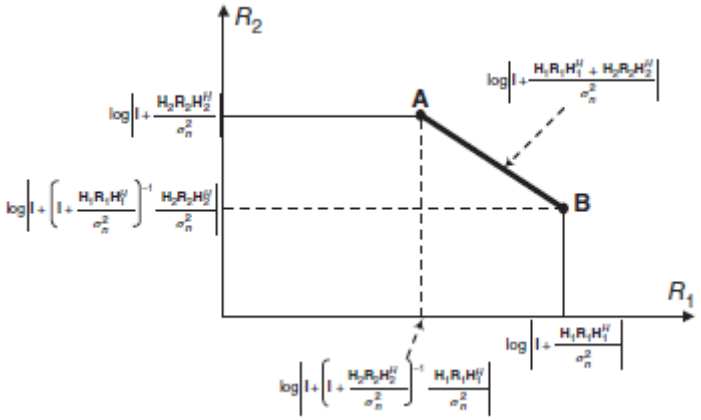
Można go rozszerzyć do systemu dwuużytkownikowego, w którym stacja bazowa z antenami Nb komunikuje się z dwoma użytkownikami wieloantenowymi k = 1, 2. Użytkownik k jest wyposażony w anteny N_k i jednocześnie wysyła $s_k \in \mathbb{C}^{Nk\times 1}$ w kanałach łącza w górę o płaskim zaniku, z ograniczeniem mocy nadawania

$$tr(\mathbf{R}_k) \leqslant P_k \quad (22)$$

gdzie $R_k = \mathbb{E}[s_k s^H_k]$ jest macierzą kowariancji sk. Oznaczamy przez $H_k \in \mathbb{C}^{Nb\times Nk}$ macierz kanału od użytkownika k do stacji bazowej. Obszar pojemności staje się wtedy

$$\mathfrak{C} = \left\{ (R_1, R_2) \in \mathbb{R}_+^2 \;\middle|\; \begin{array}{l} R_1 < \log_2 \det \left[\mathbf{I}_{N_b} + \dfrac{\mathbf{H}_1 \mathbf{R}_1 \mathbf{H}_1^H}{\sigma_n^2} \right] \\ R_2 < \log_2 \det \left[\mathbf{I}_{N_b} + \dfrac{\mathbf{H}_2 \mathbf{R}_2 \mathbf{H}_2^H}{\sigma_n^2} \right] \\ C_{sum} < \log_2 \det \left[\mathbf{I}_{N_b} + \dfrac{\mathbf{H}_1 \mathbf{R}_1 \mathbf{H}_1^H + \mathbf{H}_2 \mathbf{R}_2 \mathbf{H}_2^H}{\sigma_n^2} \right] \end{array} \right\}. \quad (23)$$

Pierwsze dwa ograniczenia wskazują, że osiągalna szybkość transmisji każdego użytkownika jest ograniczona przez pojemność systemu SU-MIMO z antenami nadawczymi N_k i antenami odbiorczymi N_b, gdzie usuwany jest inny użytkownik. Trzecie ograniczenie oznacza, że suma pojemności dwóch użytkowników jest równa systemowi typu punkt-punkt, w którym dwóch aktywnych użytkowników działa jako pojedynczy użytkownik z antenami nadawczymi $N_1 + N_2$, wysyłając niezależne sygnały i poddając się różnym ograniczeniom mocy. Przykład obszaru pojemności zilustrowano na rysunku 9.2.



Punkt A odpowiada optymalnemu przypadkowi, w którym odbiornik wykrywa najpierw s1, traktując zakłócenia między użytkownikami od użytkownika 2 jako kolorowy szum. Maksymalna szybkość transmisji użytkownika 1 jest ograniczona przez Tse i Viswanatha .

$$R_1 = \log_2 \det \left[\mathbf{I}_{N_b} + \left(\mathbf{I}_{N_b} + \frac{\mathbf{H}_2 \mathbf{R}_2 \mathbf{H}_2^H}{\sigma_n^2} \right)^{-1} \frac{\mathbf{H}_1 \mathbf{R}_1 \mathbf{H}_1^H}{\sigma_n^2} \right]. \quad (24)$$

a następnie odbiornik może wykryć s_2 z szybkością

$$R_2 = \log_2 \det \left[\mathbf{I}_{N_b} + \frac{\mathbf{H}_2 \mathbf{R}_2 \mathbf{H}_2^H}{\sigma_n^2} \right] \quad (25)$$

tak samo jak system SU-MIMO składający się ze stacji bazowej i tylko użytkownika 2. Jeśli odwrócimy kolejność usuwania zakłóceń, to punkt B jest osiągnięty. Pozostałe punkty na segmencie AB zawierają wszystkie optymalne punkty operacyjne, aby zmaksymalizować sumaryczną pojemność. Każdy punkt na tym segmencie można uzyskać poprzez podział czasu między dwoma priorytetami usuwania w punkcie A i punkcie B. Dla celów ilustracji na Rysunku, dyskusja została ograniczona do systemu z dwoma użytkownikami, ale uogólnienie na K użytkowników jest naturalne. Obszar pojemności jest teraz wielościanem K-wymiarowym, który można opisać atematycznie za pomocą

$$\mathfrak{C} = \left\{ (R_1, \ldots, R_K) \in \mathbb{R}_+^K \;\middle|\; \begin{array}{c} R_k < \log_2 \det \left[\mathbf{I}_{N_b} + \frac{\mathbf{H}_k \mathbf{R}_k \mathbf{H}_k^H}{\sigma_n^2} \right], \forall\, k \\ C_{\text{sum}} < \log_2 \det \left[\mathbf{I}_{N_b} + \frac{\sum_{k=1}^K \mathbf{H}_k \mathbf{R}_k \mathbf{H}_k^H}{\sigma_n^2} \right] \end{array} \right\} \quad (26)$$

Wnioskuje się, że sumaryczna pojemność systemu MU-MIMO jest na ogół ściśle większa niż pojemność pojedynczego użytkownika dowolnego z użytkowników w tym systemie. Jest to szczególna zaleta stosowania transmisji wieloużytkownikowej.

Kodowanie Dirty Paper

W łączu w górę systemu MU-MIMO odebrane sygnały powodują nieortogonalną superpozycję strumieni danych od różnych użytkowników, co oznacza, że proces wykrywania danych można przeprowadzić przy użyciu dobrze znanych technik wykrywania wielu użytkowników. W szczególności optymalne architektury odbiorników są wyprowadzane w oparciu o zasadę maksymalnego prawdopodobieństwa lub kryterium prawdopodobieństwa Maximum A Posteriori (MAP) [Sanguinetti i Poor, 2009]. Oferują one wydajność zbliżoną do wydajności systemu wolnego od zakłóceń, ale za cenę zaporowej złożoności, która rośnie wykładniczo wraz z liczbą użytkowników i strumieni danych. Algorytmy wykrywania suboptymalnego stosują transformację liniową w odbiorniku w postaci dopasowanego filtra, wymuszania zera lub detektorów minimalnego średniego błędu kwadratowego (MSE), które osiągają rozsądny kompromis między wydajnością a złożonością. Alternatywnie, lepszą wydajność można uzyskać stosując nieliniową detekcję z eliminacją zakłóceń (np. wymuszanie zera (ZF)-SIC, dopasowany filtr (MF)-SIC i minimalny średni błąd kwadratowy (MMSE)-SIC). Dlatego pomijamy łącze w górę i skupiamy się tylko na algorytmach prekodowania wieloużytkownikowego lub kształtowania wiązki w transmisji łącza w dół, gdzie zostaną zbadane Dirty Paper Coding (DPC) i dwa liniowe przetwarzanie, tj. prekodowanie zerowego wymuszania i prekodowanie Block Diagonalization (BD). Nazwa DPC pochodzi od „Writing on Dirty Paper", tytułu artykułu Costy na temat pojemności kanału gaussowskiego, przedstawionego wzorem

r = s + i + n, (27)

gdzie $i \sim \mathcal{N}(0, I)$ to zakłócenia, $n \sim \mathcal{N}(0, \sigma^2)$ to szum gaussowski, odebrany sygnał to r ∈ ℝ, a nadawany sygnał s ∈ ℝ, który jest używany do przesyłania u i spełnia $s^2 \leqslant P$. Jeśli i nie jest znane ani nadajnikowi, ani odbiornikowi, pojemność wynosi

$$C = \frac{1}{2} \log_2 \left(1 + \frac{P}{\sigma^2 + I} \right) \quad (28)$$

Costa przedstawił zaskakujący wynik, z którego wynika, że jeśli i jest doskonale znane koderowi, pojemność tego systemu jest taka sama jak standardowego kanału Gaussa o SNR równym P/σ^2:

$$C = \frac{1}{2}\log_2\left(1 + \frac{P}{\sigma^2}\right) \quad (29)$$

niezależnie od zakłóceń. Tytuł pracy Costy powstał na podstawie analogii do problemu pisania na brudnym papierze, gdzie czytelnik nominalnie nie potrafi odróżnić brudu od atramentu. Możemy sobie wyobrazić zaprojektowanie przesyłanego sygnału

s- u – i (30)

aby to zrealizować, ale optymalny nadajnik dostosowuje swoje sygnały do zakłóceń, zamiast próbować je anulować. DPC został zastosowany dla kanałów transmisyjnych MIMO w celu zaprojektowania optymalnej strategii transmisji osiągającej pojemność. Poddany doskonale znanemu CSI w nadajniku, może całkowicie złagodzić efekt zakłóceń wielu użytkowników i osiągnąć pojemność kanału AWGN, bez kary za moc i bez wymagania, aby odbiornik znał sygnał zakłócający. Rozważmy system z nadajnikiem wieloantenowym i wieloma odbiornikami z jedną anteną, model systemu w równaniu. (6) można dostosować do

r = Hs + n (31)

gdzie $H \in \mathbb{C}^{K\times Nb}$ oznacza macierz kanału między Nb antenami nadawczymi i K odbiornikami jednoantenowymi, wektor odebrany, wektor nadawany i wektor szumu są wyrażone odpowiednio przez $r \in \mathbb{C}^{K\times 1}$, $s \in \mathbb{C}^{Nb\times 1}$ i $n \in \mathbb{C}^{K\times 1}$. Wiedza o kanale H jest znana nadajnikowi i wszystkim odbiornikom, a dane wejściowe są ograniczone przez $\mathbb{E}[s^H s] \leqslant P$. Niech s = Bu, gdzie $B \in \mathbb{C}^{Nb\times K}$ oznacza macierz prekodowania, a wpisy u są generowane przez kolejne kodowanie brudnego papieru za pomocą książek kodowych Gaussa . Równanie (31) przenosi się do

$$\mathbf{r} = \mathbf{HBu} + \mathbf{n} = \mathbf{Wu} + \mathbf{n}. \quad (32)$$

Wstępnie zakodowany kanał daje zestaw k = 1, 2,…, K kanałów interferencyjnych

$$r_k = w_{kk}u_k + \sum_{k'<k} w_{kk'}u_{k'} + \sum_{k'>k} w_{kk'}u_{k'} + n_k, \quad (33)$$

gdzie $w_{kk'}$ reprezentuje (k, k')-ty element W, r_k, u_k i n_k oznaczają k-ty wpis r, u i n, odpowiednio, a moc nadawania jest ograniczona przez $\mathbb{E}[u_k u^*_k] \leqslant P_k$. Biorąc pod uwagę określoną kolejność kanałów interferencyjnych, koder uważa sygnał interferencyjny $\Sigma_{k'<k} w_{kk'}u_{k'}$ spowodowany przez użytkowników $k' < k$ za znany nieprzyczynowo, a detektor użytkownika k traktuje sygnał wnioskowania $\Sigma_{k'>k} w_{kk'}u_{k'}$ jako dodatkowy szum. Poprzez zastosowanie DPC w nadajniku i minimalnego euklidesowego dekodowania odległości w każdym odbiorniku, osiągalna szybkość sumy wynosi

$$R_{\mathrm{DPC}} = \sum_{k=1}^{K} \log\left(1 + \frac{|w_{kk}|^2 P_k}{\sigma_n^2 + \sum_{k'>k}|w_{kk'}|^2 P_{k'}}\right) \quad (34)$$

Wybór macierzy prekodowania można uzyskać, przeprowadzając rozkład QRH = LQ, gdzie $Q \in \mathbb{C}^{Nm\times Nb}$ ma wiersze ortogonalne, spełniając $QQ^H = I$, $L \in \mathbb{C}^{K\times Nm}$ jest dolnym trójkątem, a $N_m = \min(K, N_b)$. Przyjmując B = QH, otrzymany wektor staje się

$$\mathbf{r} = \mathbf{HBu} + \mathbf{n} = \mathbf{LQQ}^H\mathbf{u} + \mathbf{n} = \mathbf{Lu} + \mathbf{r} \quad (35)$$

odpowiadający zestawowi K kanałów interferencyjnych, jeżeli $N_b \geq K$, tj.

$$r_k = l_{kk}u_k + \sum_{k'<k} l_{kk'}u_{k'} + n_k, \quad k = 1, \ldots, K \quad (36)$$

gdzie $l_{kk'}$ reprezentuje (k, k')-ty element L. Sygnały wejściowe u są generowane przez kolejne kodowanie brudnego papieru, gdzie sygnał interferencji $\Sigma_{k'<k}l_{kk'}u_{k'}$ jest nie-przyczynowo znany w nadajniku. Mogą istnieć schematy kodowania, takie jak typowy użytkownik k nie widzi interferencji od użytkowników k' < k. Tymczasem, jak zaobserwowano w równaniu (36), macierz prekodowania jest wybierana w celu wymuszenia sygnału interferencji od użytkowników k' > k do zera. Stąd ten schemat jest również nazywany kodowaniem brudnego papieru wymuszającym zero (ZF-DPC). Następnie system MU-MIMO jest przekształcany w zestaw równoległych podkanałów, tak jak w systemie SU-MIMO, tj.

$$r_k = l_{kk}u_k + n_k, \quad k = 1, \ldots, K \quad (37)$$

Osiągalna suma wynosi

$$R_{\mathrm{DPC}} = \sum_{k=1}^{K} \log\left(1 + \frac{|l_{kk}|^2 P_k}{\sigma_n^2}\right) \quad (38)$$

które można optymalizować łącznie biorąc pod uwagę alokację mocy P_k, k = 1, 2,…, K i kolejność użytkownika.

Prekodowanie zerowego wymuszania

Pomimo jego znaczenia z punktu widzenia teorii informacji, implementacja DPC wymaga ogromnej złożoności zarówno w nadajniku, jak i odbiornikach, podczas gdy praktyczny projekt kodów zbliżonych do pojemności jest nadal otwartą kwestią. Omówiono próby w tym kierunku, takie jak uogólnienie prekodowania Tomlinsona-Harashimy (THP) na wielowymiarowy schemat kwantyzacji wektorowej. Z drugiej strony, suboptymalna, ale prosta strategia transmisji dla wielu użytkowników jest reprezentowana przez liniowe prekodowanie, znane również jako liniowe formowanie wiązki transmisyjnej. W przypadku odbioru z pojedynczej anteny, łagodzenie zakłóceń można osiągnąć tylko w stacji bazowej. Najprostsze podejście stosuje inwersję kanału przed transmisją, określaną jako prekodowanie zerowego wymuszania lub liniowe formowanie wiązki ZF. Wstępnie odwraca macierz kanału w nadajniku, tak że zakłócenia między użytkownikami całkowicie znikają we wszystkich odbiornikach. To podejście można łatwo zastosować, gdy liczba użytkowników jest mniejsza niż liczba anten transmisyjnych ($N_b > K$). Jednocześnie dotyczy to przypadku Nb < K, pod warunkiem, że zostanie zastosowany odpowiedni algorytm wyboru użytkownika. Pomimo prostoty i łatwości implementacji, takie schematy osiągają dobrą wydajność, szczególnie gdy liczba użytkowników jest duża. Pozwalając macierzy prekodowania równej pseudoodwrotności macierzy kanału, tj.

$$\mathbf{B} = \mathbf{H}^H(\mathbf{HH}^H)^{-1}, \quad (39)$$

co skutkuje otrzymanym sygnałem

$$\mathbf{r} = \mathbf{HBu} + \mathbf{n} = \mathbf{HH}^H(\mathbf{HH}^H)^{-1}\mathbf{u} + \mathbf{n} = \mathbf{u} + \mathbf{n}, \quad (40)$$

gdzie zakłócenia między użytkownikami są całkowicie tłumione. W ten sposób odbiornik generyczny k obserwuje

$r_k = u_k + n_k$ (41)

co jest równoważne kanałowi AWGN. Suma przepustowości systemu prekodowania ZF jest zatem podana przez

$$R_{ZF} = \sum_{k=1}^{K} \log\left(1 + \frac{P_k}{\sigma_n^2}\right) \quad (42)$$

Wyniki analityczne i numeryczne ujawniają, że prostota prekodowania ZF wiąże się z ceną niezaniedbywalnej straty w kategoriach szybkości sumy w odniesieniu do optymalnej techniki DPC, zwłaszcza gdy $K \leqslant N_b$. Głównym powodem tej kary jest zasadniczo efekt zwiększenia mocy, który występuje w pseudoodwrotnym obliczeniu źle uwarunkowanych macierzy kanałowych. Podejście do poprawy wydajności prekodowania ZF polega na wykorzystaniu różnorodności wielu użytkowników poprzez zastosowanie wyboru użytkownika, gdy $K \gg Nb$. Piszemy $\mathcal{A} \subset \{1, 2,\ldots, K\}$, aby oznaczyć zbiór wybranych użytkowników. Różne wybory skutkują różną szybkością sumy, a zatem maksymalną szybkość systemu uzyskuje się poprzez rozważenie wszystkich możliwych zbiorów, tj.

$$R_{\max} = \max_{\mathcal{A} \subset \{1,2,\ldots,K\}} R_{ZF} \quad (43)$$

Chciwy algorytm wyboru użytkownika ZF jest przedstawiony następująco:

1. Inicjalizacja:

Ustaw n = 1 i znajdź najlepszego użytkownika, takiego jak

$$k_1 = \arg \max_{k=1,2,\ldots,K} \left(\|\mathbf{h}_k\|^2 \right) \quad (44)$$

gdzie h_k jest k-tym wierszem H.

Ustaw $\mathcal{A}_1 = \{k1\}$ i oznacz uzyskaną stawkę sumy R_{max} ($\mathcal{A}_1$).

2. Podczas gdy $n \leqslant K$

Znajdź użytkownika k_n spośród niewybranych użytkowników takiego, że

$$k_n = \arg \max_{k \in \{1,2,\ldots,K\} - \mathcal{A}_{n-1}} R_{\max}(\mathcal{A}_{n-1} \cup \{k\}) \quad (45)$$

Ustaw $\mathcal{A}_n = \mathcal{A}_{n-1} + \{k_n\}$ i oznacz uzyskaną szybkość jako R_{max} ($\mathcal{A}_n$). Jeśli R_{max} ($\mathcal{A}_n$) < R_{max} ($\mathcal{A}_{n-1}$) zatrzymaj się i ustaw n = n − 1;

3. Określ macierz prekodowania W dla zbioru wybranych użytkowników.

Diagonalizacja bloków

W przypadku odbiorników wieloantenowych, wstępne kodowanie ZF można stosować w sposób bezpośredni, o ile traktuje się wiele anten odbiorczych każdego użytkownika jako indywidualne odbiorniki jednoantenowe bez współpracy. Jednakże, pomimo prostej architektury odbiornika, nie pozwala ona na wykorzystanie zysku współpracy wielu anten odbiorczych w przetwarzaniu wykrywania

sygnału. Jedno podejście do przezwyciężenia tej wady jest reprezentowane przez schemat BD zaproponowany niezależnie przez Choi i Murcha oraz Spencera. Jego zasada polega na tym, że nadajnik całkowicie tłumi zakłócenia między użytkownikami, podczas gdy każdy odbiornik łagodzi zakłócenia między strumieniami między swoimi odpowiednimi strumieniami danych. Używając techniki rekodowania transmisji opartej na podejściu dekompozycji w stacji bazowej, wielоużytkownikowy kanał rozgłoszeniowy MIMO jest przekształcany w wiele równoległych kanałów SU-MIMO. Każdy równoważny kanał SU-MIMO ma takie same właściwości jak konwencjonalny kanał SU-MIMO. Dlatego też, każda technika SU-MIMO, taka jak Vertical Bell Laboratories Layer Space-Time (V-BLAST), detekcja maksymalnego prawdopodobieństwa, detekcja liniowa (np. ZF,MF i MMSE) oraz prekodowanie oparte na rozkładzie wartości osobliwych, może być stosowana dla każdego użytkownika wielodostępnego systemu MIMO. Tymczasem zwiększenie liczby anten transmisyjnych wielodostępnego systemu o jedną zwiększa liczbę kanałów przestrzennych dla każdego użytkownika o jeden. Zgodnie z równaniem (14) taki system można modelować w

$$\begin{aligned}\mathbf{r}_k &= \mathbf{H}^{(k)}\sum_{k=1}^{K}\mathbf{B}_k\mathbf{u}_k + \mathbf{n}_k \\ &= \underbrace{\mathbf{H}^{(k)}\mathbf{B}_k\mathbf{u}_k}_{\text{Desired signal}} + \underbrace{\mathbf{H}^{(k')}\sum_{k'=1,k'\neq k}^{K}\mathbf{B}_{k'}\mathbf{u}_{k'}}_{\text{Multi-user Interference}} + \mathbf{n}_k\end{aligned} \quad (46)$$

Drugi element przedstawia zakłócenia dla użytkownika k spowodowane przez innych użytkowników K − 1, a zatem głównym celem jest całkowite zniwelowanie tych zakłóceń. Cel ten można przedstawić matematycznie jako

$$\mathbf{H}^{(k')}\sum_{k'=1,k'\neq k}^{K}\mathbf{B}_{k'}\mathbf{u}_{k'} = \mathbf{0}, \quad (47)$$

gdy $B_{k'} \neq \mathbf{0}$. Jest to równoważne

$$\mathbf{H}^{(k')}\sum_{k'=1,k'\neq k}^{K}\mathbf{B}_{k'} = \mathbf{0}, \quad (48)$$

ponieważ $u_{k'} \neq \mathbf{0}$. Jak przedstawiono w Choi i Murch, problem ten można rozwiązać za pomocą rozkładu wartości osobliwych (SVD) podmacierzy H. To jest

$$\tilde{\mathbf{H}}_k = \begin{bmatrix}\mathbf{H}^{(1)}\\ \vdots \\ \mathbf{H}^{(k-1)}\\ \mathbf{H}^{(k+1)}\\ \vdots \\ \mathbf{H}^{(K)}\end{bmatrix} = \mathbf{U}_k\Sigma\mathbf{V}_k^H = \mathbf{U}_k\begin{bmatrix}\Sigma' & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{V}_k^{\varnothing} & \mathbf{V}_k^{0}\end{bmatrix}^H, \quad (49)$$

gdzie $\tilde{\mathbf{H}}_k \in \mathbb{C}^{\tilde{N}_k \times N_b}$ z

$$\tilde{N}_k = \sum_{k'=1,k'\neq k}^{K} N_k$$

jest podmacierzą H usuwania wierszy dla użytkownika k. Liczba kolumn zerowych $\mathbf{0}$ w co najmniej kolumnach macierzy diagonalnej wynosi $\tilde{N}_x$, spełniają $\tilde{N}_x \geq N_b - \tilde{N}_k$. Macierz $\mathbf{V}_k^0 \in \mathbb{C}^{N_b \times \tilde{N}_x}$ odpowiada tym kolumnom zerowym. Następnie macierz prekodowania dla użytkownika k wynosi

$$\mathbf{B}_k = \mathbf{V}_k^0 \mathbf{A}_k \quad (50)$$

gdzie A_k jest niezerową macierzą $\tilde{N}_k \times L_k$, która może być zaprojektowana samodzielnie według pewnych kryteriów lub może być zaprojektowana wspólnie ze strukturą odbiornika. Aby zapewnić wgląd w tę technikę, przedstawiamy konkretny przykład demonstrujący proces osiągania BD. Dla systemu MU-MIMO składającego się ze stacji bazowej z czterema antenami i dwóch użytkowników, gdzie użytkownik 1 ma 2 anteny, a użytkownik 1 ma jedną antenę. Realizacja kanału to

$$\mathbf{H} = \begin{bmatrix} 2.0563 - 1.1151i & -0.7482 + 0.0237i & 0.7767 + 0.2476i & -1.4509 - 0.1853i \\ 0.5835 + 0.3592i & -0.3314 - 0.9431i & -0.1965 - 0.2115i & -0.2502 - 1.2376i \\ 0.9751 + 0.1994i & -0.1927 + 0.7973i & 0.4961 + 0.0162i & -0.5824 - 0.2020i \end{bmatrix} \quad (51)$$

Zastosowanie SVD

$$\tilde{\mathbf{H}}_1 = \begin{bmatrix} 0.9751 + 0.1994i & -0.1927 + 0.7973i & 0.4961 + 0.0162i & -0.5824 - 0.2020i \end{bmatrix} \quad (52)$$

prowadzi

$$\mathbf{V}_1 = \begin{bmatrix} -0.6444 + 0.1318i & 0.0418 - 0.5404i & -0.3254 + 0.0416i & 0.4012 + 0.0705i \\ 0.1273 + 0.5269i & 0.8225 + 0.0142i & 0.0302 + 0.1034i & 0.0022 - 0.1338i \\ -0.3278 + 0.0107i & 0.0134 - 0.1069i & 0.9350 + 0.0052i & 0.0790 + 0.0178i \\ 0.3849 - 0.1335i & 0.0235 + 0.1318i & 0.0752 - 0.0301i & 0.8997 + 0.0080i \end{bmatrix} \quad (53)$$

a następnie wyprowadza

$$\mathbf{V}_1^0 = \begin{bmatrix} 0.0418 - 0.5404i & -0.3254 + 0.0416i & 0.4012 + 0.0705i \\ 0.8225 + 0.0142i & 0.0302 + 0.1034i & 0.0022 - 0.1338i \\ 0.0134 - 0.1069i & 0.9350 + 0.0052i & 0.0790 + 0.0178i \\ 0.0235 + 0.1318i & 0.0752 - 0.0301i & 0.8997 + 0.0080i \end{bmatrix} \quad (54)$$

Podobnie możemy ustalić

$$\mathbf{V}_2^0 = \begin{bmatrix} -0.3605 - 0.2350i & 0.2311 + 0.1576i \\ -0.4085 - 0.1008i & -0.5849 - 0.3009i \\ 0.7751 + 0.0742i & -0.1731 + 0.0567i \\ -0.0554 + 0.1684i & 0.6666 + 0.1070i \end{bmatrix} \quad (55)$$

Zakładając $V^0 = [V^0{}_1, V^0{}_2]$, mamy

$$\mathbf{HV^0} = \begin{bmatrix} -1.11 - 1.42i & -0.04 + 0.64i & -0.34 - 0.35i & 0 & 0 \\ 0.09 - 1.12i & -0.36 - 0.46i & -0.15 - 0.91i & 0 & 0 \\ 0 & 0 & 0 & 0.30 - 0.64i & 0.09 - 0.38i \end{bmatrix}, \quad (56)$$

pokazując, że macierz kanału jest pomyślnie diagonalizowana blokowo. Wraz z $A_1 \in \mathbb{C}^{3\times L1}$, gdzie $L_1 = 1$ lub 2, i $A_2 \in \mathbb{C}^{2\times 1}$, tworzone są dwa równoległe kanały SU-MIMO. Rozwiązanie dostarczone przez Choi i Murch nie jest jedyną metodą osiągnięcia BD, istnieją inne sposoby, takie jak rozwiązanie przedstawione w Chen i innych.

Massive MIMO

Z perspektywy sieci komórkowej stacja bazowa musi jednocześnie obsługiwać rozsądną liczbę aktywnych terminali. SU-MIMO to system MIMO typu punkt-punkt, w którym wiele anten nadawczych i odbiorczych jest dedykowanych pojedynczemu użytkownikowi do multipleksowania przestrzennego. Nie oznacza to jednak, że system ma tylko jednego użytkownika. Zamiast tego różni użytkownicy są obsługiwani w ortogonalnych jednostkach zasobów czasowo-częstotliwościowych, wykorzystując multipleksowanie z podziałem czasu i multipleksowanie z podziałem częstotliwości. Teoretycznie pojemność kanału rośnie liniowo poprzez jednoczesne zwiększanie liczby anten nadawczych i anten odbiorczych. Jednak SU-MIMO nie jest skalowalne do multipleksowania przestrzennego wyższego rzędu ze względu na trzy praktyczne czynniki. Po pierwsze, trudno jest osadzić zbyt wiele anten w terminalu i zastosować zaawansowane algorytmy przetwarzania sygnału w celu oddzielenia wielowymiarowych strumieni danych ze względu na ograniczenia rozmiaru sprzętu, zasilania i kosztów sprzętu. Po drugie, kompaktowa matryca antenowa ma trudności z obsługą dużej liczby niezależnych podkanałów w łączu punkt-punkt, nawet w środowisku o dużym rozproszeniu. W szczególności matryca kanałów ma minimalną rangę jednego w warunkach widoczności. Po trzecie, pojemność kanału skaluje się powoli w warunkach niskiego współczynnika SNR, np. na skraju komórki, gdzie zwykle znajduje się większość terminali, wykazując dużą utratę ścieżki i silne zakłócenia międzykomórkowe. Poprzez rozbicie strumieni multipleksowanych przestrzennie pomiędzy wieloma terminalami, MU-MIMO zyskuje przewagę nad SU-MIMO dzięki dwóm podstawowym zaletom. Po pierwsze, MU-MIMO wymaga tylko terminali z jedną anteną, co ułatwia korzystanie ze sprzętu o niskiej złożoności, niskich kosztach i oszczędzaniu energii. Po drugie, jest mniej podatny na środowisko propagacji ze względu na przestrzenne rozmieszczenie terminali. Może dobrze działać nawet w warunkach widoczności, jeśli typowy kątowy odstęp pomiędzy terminalami jest większy niż kątowa rozdzielczość matrycy stacji bazowej. niemniej jednak konwencjonalny MU-MIMO jest nadal trudny do skalowania w celu multipleksowania przestrzennego wyższego rzędu, ponieważ osiągające pojemność prekodowanie i dekodowanie narzucają wykładniczo rosnącą złożoność. Co najważniejsze, nadajnik wymaga znajomości kanału downlink, a zasoby wydane na pozyskiwanie CSI rosną wraz z liczbą anten usługowych i liczbą użytkowników. Massive MIMO zaproponowane przez Marzettę [2010] przełamuje tę barierę skalowalności, nie próbując osiągnąć pełnego limitu Shannona i paradoksalnie zwiększając rozmiar systemu. Odchodzi od praktyki teoretycznej Shannona na trzy sposoby:

● Tylko stacja bazowa uczy się wiedzy o kanale do prekodowania w downlinku i dekodowania w uplinku, podczas gdy terminale nie muszą jej znać. Korzystając z wzajemności kanałów systemu TDD, narzut wymagany do pozyskania CSI zależy od liczby terminali, niezależnie od liczby anten stacji bazowych.

● Po stronie stacji bazowej instaluje się dużą tablicę antenową, tak aby liczba anten usługowych była zazwyczaj kilkakrotnie większa od liczby aktywnych użytkowników. Skala użytkowników pozostaje niewielka, tak aby złożoność implementacji była niska.

● W łączu w dół stosuje się proste liniowe multipleksowanie prekodowania, połączone z liniowym dekodowaniem w łączu w górę. W miarę wzrostu liczby anten stacji bazowych wydajność liniowego prekodowania i dekodowania może zbliżać się do limitu Shannona.Akwizycja CSI

Akwizycja CSI

Możemy użyć bloku koherencji, aby zdefiniować płaszczyznę czasowo-częstotliwościową, w której kanał jest uważany za niezmienny w czasie i płaski częstotliwościowo. W domenie czasowej czas trwania bloku koherencji jest równy czasowi koherencji kanału T_c, a jego szerokość w domenie częstotliwości jest taka sama jak szerokość pasma koherencji kanału B_c. Liczba jednostek zasobów czasowo-częstotliwościowych wynosi $\tau_c = T_c B_c$, co można wykorzystać do przenoszenia symboli o wartościach zespolonych τ_c. Massive MIMO opiera się na pomiarze rzeczywistych odpowiedzi kanałów propagacyjnych. W tym celu każdemu terminalowi (na blok koherencji) przypisywany jest unikalny sygnał odniesienia, a te sygnały odniesienia muszą być wzajemnie ortogonalne. Bez utraty ogólności skupiamy się tylko na pojedynczym bloku koherencji, w którym transmisja sygnału jest podzielona na trzy fazy: transmisja danych w górę łącza, szkolenie w górę łącza i transmisja danych w dół łącza. Pierwsza część tej podsekcji będzie dotyczyć sposobu akwizycji CSI w systemie massive MIMO. Systemy Asilly-ellmassiveMIMO składają się zazwyczaj ze stacji bazowej z M antenami i K terminalami pojedynczej anteny, gdzie M ≫ K. Typowy terminal k k = 1, 2,..., K jest przypisany do sygnału odniesienia o długości τ_p, oznaczonego wektorem $\boldsymbol{\phi}k \in \mathbb{C}^{\tau p \times 1}$, gdzie $\tau_c \geq \tau_p \geq K$. Aby utworzyć K ortogonalnych sygnałów odniesienia, warunek

$$\boldsymbol{\Phi}^H \boldsymbol{\Phi} = \mathbf{I}_K \quad (57)$$

powinno być spełnione, gdzie $\boldsymbol{\Phi} \in \mathbb{C}^{\tau p \times}K$ jest podane przez

$$\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_K]. \quad (58)$$

Te terminale jednocześnie przesyłają swoje sygnały odniesienia przez τ_p jednostek zasobów czasowo-częstotliwościowych. Przesyłane sygnały można oznaczyć za pomocą

$$\mathbf{X}_p = \sqrt{p_u \tau_p} \boldsymbol{\Phi}^H \quad (59)$$

gdzie normalizacja jest stosowana tak, że każdy terminal zużywa całkowitą moc równą długości sygnałów odniesienia, a p_u oznacza ograniczenie mocy łącza w górę. Sygnały odniesienia są zawsze przesyłane z maksymalną możliwą mocą bez kontroli mocy. Następnie stacja bazowa obserwuje otrzymane symbole $M \times \tau_p$

$$\mathbf{Y}_p = \mathbf{G}_u \mathbf{X}_p + \mathbf{Z}_p, \quad (60)$$

gdzie Z_p odpowiada niezależnemu zespolonemu szumowi Gaussa $M \times \tau_p$ z każdym wpisem $z \sim \mathcal{CN}(0, \sigma_n^2), \mathbf{G}_u \in \mathbb{C}^{M \times K}$ modeluje macierz kanałów łącza w górę od K terminali do M anten stacji bazowej. Współczynnik kanału o wartościach zespolonych między terminalem k a anteną m jest oznaczony przez

$$g_{mk} = \sqrt{\beta_{mk}} h_{mk} \quad (61)$$

ze współczynnikiem zaniku na dużą skalę β_{mk} i wzmocnieniem zaniku na małą skalę, który jest ogólnie i.i.d. zanikiem Rayleigha, tj. $h_{mk} \sim \mathcal{CN}(0,1)$. Rozsądnym założeniem jest, że β_{mk} jest znane (np. z pomiaru) przez system. Możemy dalej założyć, że całe β_{mk} dla typowego terminala jest takie samo, ponieważ zanik na dużą skalę zależy od odległości propagacji i zacienienia, co daje $\beta_{mk} = \beta_k$, $\forall m = 1, 2,\ldots,M$. Następnie otrzymujemy rozkład a priori $g_{mk} \sim \mathcal{CN}(0, \beta_k)$. Równanie (61) jest dalej uproszczone do

$$g_{mk} = \sqrt{\beta_k} h_{mk}. \quad (62)$$

Stacja bazowa dekoreluje odebrane sygnały ze znanymi sygnałami odniesienia

$$\begin{aligned} \tilde{\mathbf{Y}}_p = \mathbf{Y}_p \mathbf{\Phi} &= \mathbf{G}_u \mathbf{X}_p \mathbf{\Phi} + \mathbf{Z}_p \mathbf{\Phi} \\ &= \sqrt{p_u \tau_p} \mathbf{G}_u \mathbf{\Phi}^H \mathbf{\Phi} + \mathbf{Z}_p \mathbf{\Phi} \\ &= \sqrt{p_u \tau_p} \mathbf{G}_u + \tilde{\mathbf{Z}}_p, \end{aligned} \quad (63)$$

gdzie każdy wpis $\tilde{\mathbf{Z}}_p \in \mathbb{C}^{M \times K}$ jest również niezależnym zespolonym szumem australijskim $\tilde{z} \sim \mathcal{CN}(0, \sigma_n^2)$ ze względu na mnożenie przez macierz unitarną. Ze względu na niezależność g i z, równanie (63) można rozłożyć na

$$\tilde{y}_{mk,p} = \sqrt{p_u \tau_p} g_{mk} + \tilde{z}_{mk} \quad (64)$$

Przeprowadzając estymację kanału za pomocą liniowej MMSE, estymację uzyskuje się przez

$$\hat{g}_{mk} = \mathbb{E}\left[g_{mk} | \tilde{y}_{mk,p}\right] = \frac{\mathbb{E}\left[\tilde{y}^*_{mk,p} g_{mk}\right] \tilde{y}_{mk,p}}{\mathbb{E}\left[\left|\tilde{y}_{mk,p}\right|^2\right]} = \left(\frac{\sqrt{p_u \tau_p} \beta_k}{p_u \tau_p \beta_k + \sigma_n^2}\right) \tilde{y}_{mk,p}.$$

Niech $\hat{g}_{mk}$ będzie oszacowaniem g_{mk}, a $\hat{g}_{mk}$ będzie błędem oszacowania wywołanym przez szum addytywny. Mamy

$$\hat{g}_{mk} = g_{mk} - \tilde{g}_{mk}. \quad (65)$$

Wariancję $\hat{g}_{mk}$ oblicza się za pomocą

$$\mathbb{E}\left[\left|\hat{g}_{mk}\right|^2\right] = \frac{p_u \tau_p \beta_k^2}{p_u \tau_p \beta_k + \sigma_n^2}. \quad (66)$$

Następnie możemy zapisać $\hat{g}_{mk} \sim \mathcal{CN}(0, \alpha_k)$ z $\alpha_k = \frac{p_u \tau_p \beta_k^2}{p_u \tau_p \beta_k + \sigma_n^2}$, a MSE jest zatem

$$\mathbb{E}\left[\left|\tilde{g}_{mk}\right|^2\right] = \beta_k - \alpha_k = \frac{\sigma_n^2 \beta_k}{p_u \tau_p \beta_k + \sigma_n^2}. \quad (67)$$

Liniowa detekcja w łączu w górę

W łączu w górę, K terminali jednocześnie przesyła swoje symbole w kierunku stacji bazowej. Nie ma wyraźnej współpracy między terminalami w celu wykonania wspólnego wstępnego kodowania. Jedyne, co te terminale mogą zrobić, to niezależnie ważyć swoje symbole. Piszemy η_k, aby oznaczyć współczynnik kontroli mocy dla typowego terminala k, spełniającego $0 \leqslant \eta_k \leqslant 1$. Przesyłane symbole u_k, k = 1, 2,…, K są nieskorelowane z ograniczeniem mocy p_u. W konsekwencji macierz kowariancji dla wektora przesyłanych symboli $u = [u_1, u_2,\ldots, u_K]^T$ wynosi

$$\mathbb{E}\left[\mathbf{u}\mathbf{u}^H\right] = p_u \mathbf{I}_K. \quad (68)$$

Przekazywany symbol dla typowego terminala k to $\sqrt{\eta_k}u_k$. Zatem stacja bazowa obserwuje wektor M × 1 odebranych symboli

$$\mathbf{r} = \mathbf{G}_u \mathbf{D}_\eta \mathbf{u} + \mathbf{n} \quad (69)$$

stosując macierz diagonalną $\mathbf{D}_\eta \in \mathbb{C}^{K\times K}$ utworzoną ze współczynników sterujących potęgą η_k, k = 1, 2,…, K, mianowicie

$$\mathbf{D}_\eta = \begin{bmatrix} \sqrt{\eta_1} & 0 & \ldots & 0 \\ 0 & \sqrt{\eta_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sqrt{\eta_K} \end{bmatrix}. \quad (70)$$

Stacja bazowa wykonuje przetwarzanie detekcji w kategoriach obserwacji i wiedzy o kanale, aby odzyskać przesyłane symbole. Proces ten można oznaczyć matematycznie jako

$$\hat{\mathbf{u}} = f(\mathbf{r}, \mathbf{G}_u), \quad (71)$$

gdzie zakładamy, że wiedza o kanale w stacji bazowej jest idealna (wpływ błędu oszacowania kanału na wydajność można znaleźć w literaturze. Detekcja liniowa jest atrakcyjna z praktycznego punktu widzenia ze względu na niską złożoność przy jednoczesnym osiągnięciu dobrej wydajności. Trzy algorytmy liniowe są zwykle stosowane do wykrywania łącza w górę masywnego systemu MIMO.

Filtrowanie dopasowane

Filozofia stojąca za filtrowaniem dopasowanym, znanym również jako łączenie o maksymalnym współczynniku, polega na wzmocnieniu pożądanego sygnału w jak największym stopniu, przy jednoczesnym pominięciu zakłóceń między użytkownikami. W przypadku transmisji pojedynczego użytkownika byłoby to optymalne. Macierz dekodowania może być podana przez GHu, co skutkuje wyjściem post-processingu

$$\tilde{\mathbf{u}} = \mathbf{G}_u^H \mathbf{r} = \mathbf{G}_u^H \mathbf{G}_u \mathbf{D}_\eta \mathbf{u} + \mathbf{G}_u^H \mathbf{n}$$

Rozłożenie tego równania daje k-tą miękką ocenę

$$\tilde{u}_k = \underbrace{\|\mathbf{g}_k\|^2 \sqrt{\eta_k} u_k}_{\text{Desired signal}} + \underbrace{\sum_{i=1, i\neq k}^{K} \mathbf{g}_k^H \mathbf{g}_i \sqrt{\eta_i} u_i}_{} + \underbrace{\mathbf{g}_k^H \mathbf{n}}_{\text{Noise}}, \quad (72)$$

gdzie $g_k \in \mathbb{C}^{M\times 1}$ jest k-tą kolumną G_u, lub $G_u = [g_1, g_2,\ldots, g_K]$. Traktując zakłócenia między użytkownikami jako kolorowy szum, stacja bazowa może uzyskać twardą ocen $\tilde{u}_k$ dla każdego przesyłanego symbolu u_k.

Wykrywanie ZF

Zamiast maksymalizować siłę pożądanego sygnału, lepszą wydajność można osiągnąć, całkowicie anulując zakłócenia między użytkownikami. Macierz dekodowania jest pseudoodwrotnością macierzy kanału, tj. $(G^H_u G_u)^{-1}G^H_u$. Zatem wyjściem przetwarzania końcowego jest

$$\begin{aligned}\tilde{\mathbf{u}} &= (\mathbf{G}_u^H\mathbf{G}_u)^{-1}\mathbf{G}_u^H\mathbf{r}\\ &= (\mathbf{G}_u^H\mathbf{G}_u)^{-1}\mathbf{G}_u^H\mathbf{G}_u\mathbf{D}_\eta\mathbf{u} + (\mathbf{G}_u^H\mathbf{G}_u)^{-1}\mathbf{G}_u^H\mathbf{n}\\ &= \mathbf{D}_\eta\mathbf{u} + (\mathbf{G}_u^H\mathbf{G}_u)^{-1}\mathbf{G}_u^H\mathbf{n}.\end{aligned} \quad (73)$$

Podobnie, rozkład powyższego równania daje k-tą miękką ocenę

$$\tilde{u}_k = \underbrace{\sqrt{\eta_k}u_k}_{\text{Desired signal}} + \underbrace{\mathbf{g}_k\mathbf{n}}_{\text{Noise}}, \quad (74)$$

gdzie $g_k \in \mathbb{C}^{1\times M}$ oznacza k-ty wiersz $(G^H_u G_u)^{-1}G^H_u$. Interferencja między użytkownikami jest teraz całkowicie wyeliminowana, ale możliwe jest, że szum jest wzmacniany, jeśli macierz dekodowania jest źle uwarunkowana

Wykrywanie MMSE

Aby złagodzić efekt wzmocnienia szumu przy wymuszaniu zera, istnieje zregularizowana wersja z macierzą dekodowania $(G^H_u G_u + \sigma^2_n I)^{-1}G^H_u$. Może ona zminimalizować MSE szacowanych symboli, dlatego nazywa się ją wykrywaniem MMSE lub zregularizowanym wykrywaniem ZF. Wyjście przetwarzania końcowego staje się

$$\begin{aligned}\tilde{\mathbf{u}} &= (\mathbf{G}_u^H\mathbf{G}_u + \sigma_n^2\mathbf{I})^{-1}\mathbf{G}_u^H\mathbf{r}\\ &= (\mathbf{G}_u^H\mathbf{G}_u + \sigma_n^2\mathbf{I})^{-1}\mathbf{G}_u^H\mathbf{G}_u\mathbf{u} + (\mathbf{G}_u^H\mathbf{G}_u + \sigma_n^2\mathbf{I})^{-1}\mathbf{G}_u^H\mathbf{n}.\end{aligned} \quad (75)$$

W niskim reżimie SNR $\sigma^2_n \rightarrow 0$, regularizowana detekcja ZF osiąga porównywalną wydajność z detekcją ZF. W wysokim reżimie SNR $\sigma 2n \gg 0$ zachowuje się podobnie do dopasowanego filtrowania. Dlatego regularizowana detekcja ZF działa dobrze w całym zakresie SNR.

Liniowe prekodowanie w łączu w dół

W łączu w dół masywnego systemu MIMO stacja bazowa multipleksuje przestrzennie symbole niosące informacje przeznaczone dla K terminali, oznaczone jako $u = [u_1, u_2,\ldots, u_K]^T$, poprzez prekodowanie lub formowanie wiązki nadawczej. Następnie wysyła multipleksowane przestrzennie sygnały w tej samej jednostce zasobów czasowo-częstotliwościowych. Istotną różnicą w przypadku transmisji w górę jest to, że wspólne przetwarzanie może być wykonywane między antenami nadawczymi M. Piszemy η_k, aby oznaczyć współczynnik kontroli mocy dla k-tego symbolu informacji, a η_k, k = 1, 2,…, K są również wspólnie określane, z zastrzeżeniem

$$\sum_{k=1}^{K} \eta_k \leqslant 1 \quad (76)$$

Następnie wektor przekazywanych symboli $s = [s_1, s_2, \ldots, s_M]^T$ można utworzyć za pomocą

$$\mathbf{s} = \mathbf{P}\mathbf{D}_\eta \mathbf{u}, \quad (77)$$

gdzie P oznacza macierz prekodowania M × K, a D_η jest określone w równaniu (70). Wybór nieujemnych współczynników kontroli mocy i skalowanie macierzy prekodowania zapewniają, że całkowita moc transmisji spełnia

$$\mathbb{E}\left[\mathbf{s}^H \mathbf{s}\right] = tr\left(\mathbb{E}\left[\mathbf{s}\mathbf{s}^H\right]\right) \leqslant P. \quad (78)$$

Łącznie wektor K × 1 otrzymanych symboli dla wszystkich terminali jest podany przez

$$\mathbf{r} = \mathbf{G}_d \mathbf{s} + \mathbf{n} = \mathbf{G}_u^T \mathbf{s} + \mathbf{n}, \quad (79)$$

gdzie $G_d \in \mathbb{C}^{K \times M}$ oznacza macierz kanału od stacji bazowej do terminali, a zakładamy $G_d = G^T_u$ ze względu na wzajemność kanału w systemie TDD. Podobnie jak w przypadku wykrywania liniowego, istnieją trzy typowe liniowe metody prekodowania, tj. prekodowanie o maksymalnym współczynniku lub zwane również sprzężonym formowaniem wiązki, prekodowanie ZF i zregularyzowane prekodowanie ZF.

Sprzężone formowanie wiązki

Aby zmaksymalizować wzmocnienie tablicy transmisji, macierz prekodowania jest podana przez $P_{cb} = \alpha_{cb} G^H_d$ ze skalarem normalizującym α_{cb}. W rezultacie otrzymujemy wektor symboli

$$\mathbf{r} = \mathbf{G}_d \mathbf{s} + \mathbf{n} = \mathbf{G}_d \mathbf{P}_{cb} \mathbf{D}_\eta \mathbf{u} + \mathbf{n} = \alpha_{cb} \mathbf{G}_d \mathbf{G}_d^H \mathbf{D}_\eta \mathbf{u} + \mathbf{n}. \quad (80)$$

Równoważnie, k-ty terminal ma obserwację

$$r_k = \underbrace{\alpha_{cb} \|\mathbf{g}_k\|^2 \sqrt{\eta_k} u_k}_{\text{Desired signal}} + \underbrace{\alpha_{cb} \sum_{i=1, i \neq k}^{K} \mathbf{g}_k \mathbf{g}_i^H \sqrt{\eta_i} u_i}_{} + \underbrace{\mathbf{n}}_{\text{Noise}}. \quad (81)$$

gdzie $g_k \in \mathbb{C}^{1 \times M}$ jest k-tym wierszem G_d. W porównaniu z równaniem (72) wiadomo, że sprzężone formowanie wiązki może utworzyć odebrany sygnał równy miękkiemu oszacowaniu po post-processingu . Stąd detekcja sygnału po stronie terminala jest uproszczona.

Wstępne kodowanie ZF

Dzięki wstępnemu kodowaniu wymuszającemu zero w nadajniku, interferencja między użytkownikami odebranych sygnałów może zostać całkowicie stłumiona. Macierz wstępnego kodowania jest pseudoodwrotnością macierzy kanału, tj. $P_{ZF} = \alpha_{ZF} G^H_d (G_d G^H_d)^{-1}$. Otrzymany wektor symboli staje się wtedy

$$\begin{aligned} \mathbf{r} = \mathbf{G}_d \mathbf{s} + \mathbf{n} &= \mathbf{G}_d \mathbf{P}_{ZF} \mathbf{D}_\eta \mathbf{u} + \mathbf{n} \\ &= \alpha_{ZF} \mathbf{G}_d \mathbf{G}_d^H (\mathbf{G}_d \mathbf{G}_d^H)^{-1} \mathbf{D}_\eta \mathbf{u} + \mathbf{n} \\ &= \alpha_{ZF} \mathbf{D}_\eta \mathbf{u} + \mathbf{n}. \end{aligned} \quad (82)$$

Zatem obserwacja k-tego terminala jest wyrażona wzorem

$$r_k = \underbrace{\alpha_{\mathrm{ZF}}\sqrt{\eta_k}u_k}_{\text{Desired signal}} + \underbrace{n_k}_{\text{Noise}}, \quad (83)$$

co jest równoważne kanałowi AWGN

$$\tilde{r}_k = u_k + \tilde{n}_k \quad (84)$$

z tego powodu, że czynnik $\alpha_{\mathrm{ZF}}\sqrt{\eta_k}$ jest deterministyczny i łatwy do poznania

Regularized ZF Precoding

Możliwe jest utworzenie liniowej kombinacji prekodowania ZF i sprzężonego formowania wiązki za pomocą regularizacji. Przed inwersją macierzy $G_dG^H_d$ dodawany jest diagonalny współczynnik obciążenia. W konsekwencji macierz prekodowania staje się

$$\mathbf{P}_{\mathrm{rZF}} = \alpha_{\mathrm{rZF}}\mathbf{G}_d^H(\mathbf{G}_d\mathbf{G}_d^H + \delta\mathbf{I})^{-1} \quad (85)$$

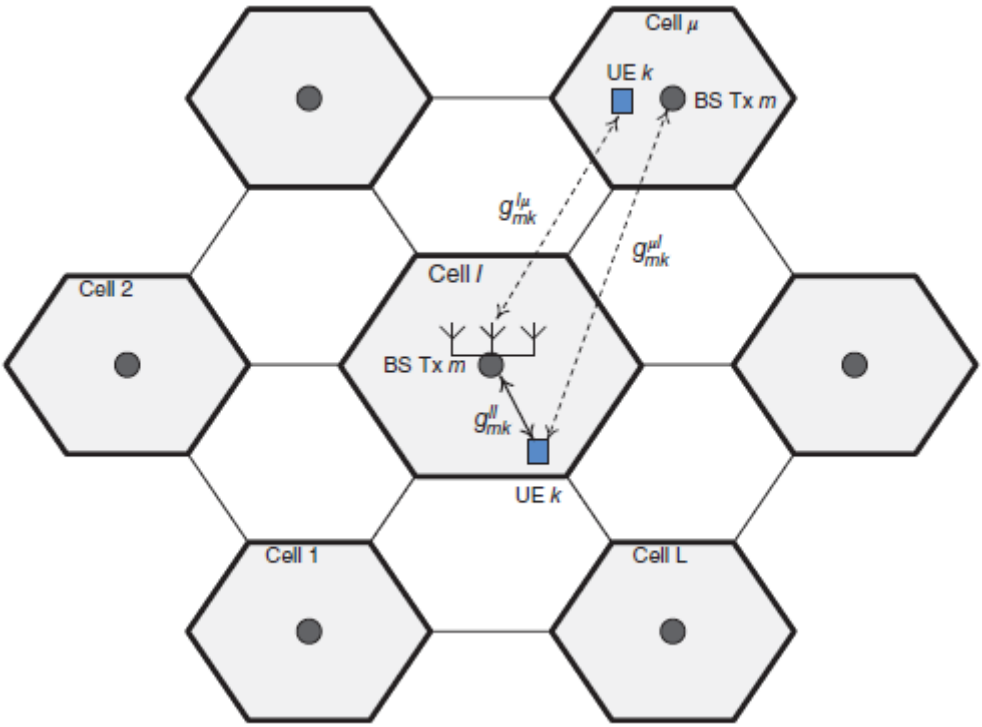
Następnie przesyłany wektor symboli jest

$$\mathbf{s} = \alpha_{\mathrm{rZF}}\mathbf{G}_d^H(\mathbf{G}_d\mathbf{G}_d^H + \delta\mathbf{I})^{-1}\mathbf{D}_\eta\mathbf{u}, \quad (86)$$

gdzie $\delta > 0$ jest współczynnikiem regularizacji i może być optymalizowany na podstawie wymagań projektowych. Regularyzowane prekodowanie ZF staje się prekodowaniem ZF, gdy $\delta \to 0$, i staje się sprzężonym formowaniem wiązki, gdy$\delta \gg 0$.

Multi-Cell Massive MIMO

Z perspektywy sieci komórkowej transmisja downlink i uplink, zwłaszcza trening uplink, komórki jest dotknięta przez sąsiednie komórki. Rozważmy system komórkowy z siecią nienakładających się komórek. Dwie sąsiednie komórki są zazwyczaj przypisane do ortogonalnych pasm częstotliwości w celu wyeliminowania interferencji między komórkami. Rysunek przedstawia sieć składającą się z heksagonalnych komórek o współczynniku ponownego wykorzystania częstotliwości wynoszącym siedem.



Załóżmy, że łącznie L komórek, indeksowanych przez l = 1, 2,…, L, dzieli to samo pasmo częstotliwości, odnosząc się do komórek współkanałowych, ignorując inne komórki współkanałowe o pomijalnej wzajemnej interferencji ze względu na znaczną odległość separacji. Każda komórka składa się z jednej stacji bazowej z antenami M i użytkowników pojedynczej anteny K. Piszemy $g^{l\mu}{}_{mk}$, aby modelować kanał między m-tą anteną serwisową na stacji bazowej komórki l a k-tym użytkownikiem komórki μ, gdzie skupiamy się na komórce l, a μ oznacza jedną z pobliskich komórek współkanałowych. Ogólny zysk kanału to

$$g_{mk}^{l\mu} = \sqrt{\beta_{mk}^{l\mu}} h_{mk}^{l\mu} \quad (87)$$

gdzie współczynnik zanikania na dużą skalę $\beta^{l\mu}{}_{mk}$ jest stałą nieujemną i zakłada się, że jest znany każdemu, a wzmocnienie zanikania na małą skalę $h^{l\mu}{}_{mk}$ jest niezależnymi, zerowymi, kołowo symetrycznymi zespolonymi zmiennymi losowymi Gaussa, tj. $h_{mk}^{l\mu} \in \mathcal{CN}(0,1)$. Wartości $\beta^{l\mu}{}_{mk}$ modelują stratę ścieżki i zacienienie, które zmieniają się powoli i można ich nauczyć się w długim okresie czasu, podczas gdy wartości $h^{l\mu}{}_{mk}$ modelują zanikanie, które zmienia się stosunkowo szybko i muszą być nauczone i wykorzystane bardzo szybko. Ponieważ układ komórek i zacienienie są przechwytywane przy użyciu stałych wartości $\beta^{l\mu}{}_{mk}$, szczegółowe szczegóły układu komórek i modelu zacienienia są nieistotne. W systemie TDD zakładamy wzajemność kanałów dla łączy do przodu i do tyłu, tj. $g^{l\mu}{}_{mk} = g^{l\mu}{}_{km}$, oraz zanikanie bloków, mianowicie $h^{l\mu}{}_{mk}$ pozostaje stałe przez szereg okresów symboli. Współczynniki zanikania na małą skalę są zwykle różne dla różnych par anten użytkownika ze względu na konstruktywne i destruktywne dodawanie sygnałów wielodrogowych. Natomiast współczynniki zanikania na dużą skalę są identyczne dla wszystkich anten w tej samej stacji bazowej, ale zależą tylko od użytkownika, ponieważ są związane z odległością propagacji i zacienianiem. Tak więc $\beta^{l\mu}{}_{mk}$ w równaniu (87) można zastąpić $\beta^{l\mu}{}_{k}$, mianowicie

$$g_{mk}^{l\mu} = \sqrt{\beta_k^{l\mu}} h_{mk}^{l\mu} \quad (88)$$

Macierz kanału łącza w górę od wszystkich K użytkowników w μ-tej komórce do stacji bazowej l-tej komórki można wyrazić jako

$$\mathbf{G}^{l\mu} = \begin{pmatrix} g_{11}^{l\mu} & g_{12}^{l\mu} & \cdots & g_{1K}^{l\mu} \\ g_{21}^{l\mu} & g_{22}^{l\mu} & \cdots & g_{2K}^{l\mu} \\ \vdots & \vdots & \ddots & \vdots \\ g_{M1}^{l\mu} & g_{M2}^{l\mu} & \cdots & g_{MK}^{l\mu} \end{pmatrix} = \mathbf{H}^{l\mu} \left(\mathbf{B}^{l\mu}\right)^{1/2}, \quad (89)$$

gdzie

$$\mathbf{H}^{l\mu} = \begin{pmatrix} h_{11}^{l\mu} & h_{12}^{l\mu} & \cdots & h_{1K}^{l\mu} \\ h_{21}^{l\mu} & h_{22}^{l\mu} & \cdots & h_{2K}^{l\mu} \\ \vdots & \vdots & \ddots & \vdots \\ h_{M1}^{l\mu} & h_{M2}^{l\mu} & \cdots & h_{MK}^{l\mu} \end{pmatrix}, \quad (90)$$

i

$$\mathbf{B}^{l\mu} = \begin{pmatrix} \beta_1^{l\mu} & & & \\ & \beta_2^{l\mu} & & \\ & & \ddots & \\ & & & \beta_K^{l\mu} \end{pmatrix}. \quad (91)$$

Macierz kanałów łącza wstecznego od stacji bazowej l-tej komórki do wszystkich K użytkowników w μ-tej komórce jest transpozycją $G^{l\mu}$, tj.

$$\left(\mathbf{G}^{l\mu}\right)^T = \begin{pmatrix} g_{11}^{l\mu} & g_{21}^{l\mu} & \cdots & g_{M1}^{l\mu} \\ g_{12}^{l\mu} & g_{22}^{l\mu} & \cdots & g_{M2}^{l\mu} \\ \vdots & \vdots & \ddots & \vdots \\ g_{1K}^{l\mu} & g_{2K}^{l\mu} & \cdots & g_{MK}^{l\mu} \end{pmatrix}. \quad (92)$$

Należy zauważyć, że macierz kanału łącza w górę od wszystkich K użytkowników w l-tej komórce do stacji bazowej μ-tej komórki jest wyrażona jako

$$\mathbf{G}^{\mu l} = \begin{pmatrix} g_{11}^{\mu l} & g_{12}^{\mu l} & \cdots & g_{1K}^{\mu l} \\ g_{21}^{\mu l} & g_{22}^{\mu l} & \cdots & g_{2K}^{\mu l} \\ \vdots & \vdots & \ddots & \vdots \\ g_{M1}^{\mu l} & g_{M2}^{\mu l} & \cdots & g_{MK}^{\mu l} \end{pmatrix} \quad (93)$$

gdzie pierwsza litera indeksu górnego jest używana do indeksu komórki, w której znajduje się stacja bazowa, a druga litera indeksu górnego jest używana do indeksu komórki, w której znajdują się użytkownicy.

Zanieczyszczenie pilota

W wielokomórkowym systemie Massive MIMO typowy terminal k, k = 1, 2,…, K w komórce l, l = 1, 2,…, L jest przypisany do sygnału odniesienia $\boldsymbol{\phi}_{k,l} \in \mathbb{C}^{\tau p \times 1}$. W idealnym przypadku sygnały odniesienia wykorzystywane przez użytkowników w tej samej komórce i w sąsiednich komórkach współkanałowych są ortogonalne, tj.

$$\boldsymbol{\phi}_{k,l}^{H}\boldsymbol{\phi}_{i,\mu} = \delta[k-i]\delta[l-\mu], \quad (94)$$

gdzie $\delta[\cdot]$ jest funkcją delta

$$\delta[n] = \begin{cases} 1 & n = 0 \\ 0 & \text{otherwise} \end{cases} \quad (95)$$

W formie wektorowej spełnia

$$\boldsymbol{\Phi}_{l}^{H}\boldsymbol{\Phi}_{\mu} = \delta[l-\mu]\mathbf{I}_{K}, \quad (96)$$

gdzie $\tau_p \geq K$, a $\boldsymbol{\Phi}l \in \mathbb{C}^{\tau p \times K}$ jest podane wzorem

$$\boldsymbol{\Phi}_{l} = [\boldsymbol{\phi}_{1,l}, \boldsymbol{\phi}_{2,l}, \ldots, \boldsymbol{\phi}_{K,l}]. \quad (97)$$

Jednak liczba ortogonalnych sekwencji odniesienia o danym okresie i szerokości pasma jest ograniczona, co z kolei ogranicza liczbę użytkowników, którym można obsłużyć. Aby obsłużyć więcej użytkowników, w sąsiednich komórkach używane są nieortogonalne sekwencje odniesienia. W rezultacie oszacowanie wektora kanału dla użytkownika staje się skorelowane z wektorami kanału użytkowników z nieortogonalnymi sekwencjami odniesienia. Istnieje wiele schematów przypisywania sekwencji odniesienia użytkownikom w różnych komórkach. Jednym z prostych schematów jest ponowne użycie tego samego zestawu ortogonalnych sekwencji odniesienia we wszystkich komórkach współkanałowych. Oznacza to, że k-temu użytkownikowi w komórce zostanie przypisana sekwencja odniesienia $\boldsymbol{\phi}_k$. Identyczne sekwencje odniesienia przypisane użytkownikom w sąsiednich komórkach współkanałowych będą się wzajemnie zakłócać, co doprowadzi do skażenia pilota. Rozważmy system z L komórkami współkanałowymi, a pozostałe komórki obsługiwane w innych pasmach częstotliwości są uważane za idealnie odizolowane. Wszystkie komórki L używają tego samego zestawu sekwencji odniesienia K $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_K]$, a k-ty użytkownik w każdej komórce jest przypisany do identycznej sekwencji odniesienia $\boldsymbol{\phi}_k$. Terminale we wszystkich komórkach L jednocześnie przesyłają swoje sygnały odniesienia i zakładają ponadto, że transmisja z różnych komórek jest zsynchronizowana (z punktu widzenia zanieczyszczenia pilota stanowi to najgorszy możliwy przypadek). Przesyłane sygnały można oznaczyć za pomocą

$$\mathbf{X}_{p} = \sqrt{P_r \tau_p}\boldsymbol{\Phi}^{H} \quad (98)$$

gdzie normalizacja jest stosowana tak, że każdy terminal zużywa całkowitą moc równą długości sygnałów odniesienia, a P_r oznacza ograniczenie mocy w łączu odwrotnym. Następnie stacja bazowa w komórce l obserwuje M × τ_p odebranych symboli

$$\mathbf{Y}_p^l = \sum_{\mu=1}^{L} \mathbf{G}^{l\mu}\mathbf{X}_p + \mathbf{Z}_p \quad (99)$$

gdzie Z_p odpowiada niezależnemu zespolonemu szumowi Gaussa M × τ_p z każdym wpisem $z \sim \mathcal{CN}(0, \sigma_n^2), \mathbf{G}^{l\mu} \in \mathbb{C}^{M\times K}$ modeluje macierz kanałów łącza w górę od K terminali komórki μ do M anten stacji bazowej komórki l, jak zdefiniowano w równaniu (89). Stacja bazowa dekoreluje odebrane sygnały ze znanymi sygnałami odniesienia

$$\begin{aligned}\tilde{\mathbf{Y}}_p^l &= \frac{1}{\sqrt{P_r\tau_p}}\mathbf{Y}_p^l\mathbf{\Phi} = \frac{1}{\sqrt{P_f\tau_p}}\sum_{\mu=1}^{L}\mathbf{G}^{l\mu}\mathbf{X}_p\mathbf{\Phi} + \frac{1}{\sqrt{P_r\tau_p}}\mathbf{Z}_p\mathbf{\Phi} \\ &= \sum_{\mu=1}^{L}\mathbf{G}^{l\mu}\mathbf{\Phi}^H\mathbf{\Phi} + \frac{1}{\sqrt{P_r\tau_p}}\mathbf{Z}_p\mathbf{\Phi} \\ &= \underbrace{\mathbf{G}^{ll}}_{\text{Desired CSI}} + \underbrace{\sum_{\mu=1,\mu\neq l}^{L}\mathbf{G}^{l\mu}}_{\text{Pilot contamination}} + \frac{1}{\sqrt{P_r\tau_p}}\tilde{\mathbf{Z}}_p\end{aligned} \quad (100)$$

gdzie każdy wpis $\tilde{\mathbf{Z}}_p = \mathbf{Z}_p\mathbf{\Phi}$ jest również niezależnym zespolonym szumem Gauss $\tilde{z} \sim \mathcal{CN}(0, \sigma_n^2)$ wynikającym z mnożenia przez macierz unitarną. Powyższe równanie można rozłożyć na

$$\tilde{y}_{mk,p}^l = g_{mk}^{ll} + \sum_{\mu=1,\mu\neq l}^{L} g_{mk}^{l\mu} + \frac{1}{\sqrt{P_r\tau_p}}\tilde{z}_{mk} \quad (101)$$

Przeprowadzając estymację kanału za pomocą liniowej MMSE, estymację uzyskuje się przez

$$\begin{aligned}\hat{g}_{mk}^{ll} = \mathbb{E}\left[g_{mk}^{ll}|\tilde{y}_{mk,p}^l\right] &= \frac{\mathbb{E}\left[\left(\tilde{y}_{mk,p}^l\right)^* g_{mk}^{ll}\right]\tilde{y}_{mk,p}^l}{\mathbb{E}\left[\left|\tilde{y}_{mk,p}^l\right|^2\right]} \\ &= \left(\frac{\beta_k^{ll}}{\sum_{\mu=1}^{L}\beta_k^{l\mu} + \frac{\sigma_n^2}{P_r\tau_p}}\right)\tilde{y}_{mk,p}^l\end{aligned} \quad (102)$$

Wariancję $\hat{g}_{mk}^{ll}$ oblicza się za pomocą

$$\mathbb{E}\left[\left|\hat{g}_{mk}^{ll}\right|^2\right] = \frac{P_r\tau_p\left(\beta_k^{ll}\right)^2}{P_r\tau_p\sum_{\mu=1}^{L}\beta_k^{l\mu} + \sigma_n^2} = \alpha_k^{ll} \quad (103)$$

Następnie możemy zapisać $\hat{g}_{mk}^{ll} \sim \mathcal{CN}(0, \alpha_k^{ll})$. Niech $\tilde{g}_{mk}^{ll}$ będzie błędem oszacowania wywołanym przez szum addytywny i zanieczyszczenie pilota, mamy

$$\tilde{g}_{mk}^{ll} = g_{mk}^{ll} - \hat{g}_{mk}^{ll} \quad (104)$$

W ten sposób oblicza się MSE oszacowania tego kanału:

$$\mathbb{E}\left[\left|\tilde{g}_{mk}\right|^2\right] = \beta_k^{ll} - \alpha_k^{ll}. \quad (105)$$

Równoważnie równanie (9.100) można również zapisać jako

$$\begin{aligned}\tilde{\mathbf{Y}}_p^l &= \mathbf{G}^{ll} + \sum_{\mu=1,\mu\neq l}^{L} \mathbf{G}^{l\mu} + \frac{1}{\sqrt{P_r\tau_p}}\tilde{\mathbf{Z}}_p \\ &= \mathbf{H}^{ll}\left(\mathbf{B}^{ll}\right)^{1/2} + \sum_{\mu=1,\mu\neq l}^{L} \mathbf{H}^{l\mu}\left(\mathbf{B}^{l\mu}\right)^{1/2} + \frac{1}{\sqrt{P_r\tau_p}}\tilde{\mathbf{Z}}_p.\end{aligned} \quad (106)$$

Zakładając, że $B^{l\mu}$ jest znane, oszacowanie MMSE Hll wynosi :

$$\hat{\mathbf{H}}^{ll} = \sqrt{P_r\tau_p}\left(\mathbf{B}^{ll}\right)^{1/2}\left(\sigma_n^2\mathbf{I} + P_r\tau_p\sum_{\mu=1}^{L}\mathbf{B}^{l\mu}\right)^{-1}\tilde{\mathbf{Y}}_p^l. \quad (107)$$

Transmisja danych w łączu w górę

W łączu w górę wielokomórkowego systemu MIMO, K terminali w każdej komórce niezależnie przesyła sygnały w kierunku odpowiedniej stacji bazowej. Niech u^{μ}_{k} , ∀k = 1, 2,…, K i μ = 1, 2,…, L oznaczają symbol średniej zerowej, wariancji jednostkowej od użytkownika k w μ-tej komórce. Wektor symboli niosących informacje od wszystkich K użytkowników w komórce μ jest wyrażony jako $u_{\mu} = [u^{\mu}_{1}, u^{\mu}_{2}, \ldots, u^{\mu}_{K}]^{T}$. Ignorując kontrolę mocy, stacja bazowa w l-tej komórce odbiera wektor M × 1 obejmujący otrzymane sygnały ze wszystkich terminali w L komórkach, tj.

$$\begin{aligned}\mathbf{r}_l &= \sqrt{P_r}\sum_{\mu=1}^{L}\mathbf{G}^{l\mu}\mathbf{u}_{\mu} + \mathbf{n}_l \\ &= \underbrace{\sqrt{P_r}\mathbf{G}^{ll}\mathbf{u}_l}_{\text{Desired signal}} + \underbrace{\sqrt{P_r}\sum_{\mu=1,\mu\neq l}^{L}\mathbf{G}^{l\mu}\mathbf{u}_{\mu}}_{\text{Inter-cell interference}} + \mathbf{n}_l\end{aligned} \quad (108)$$

z ograniczeniem mocy łącza zwrotnego P_r i macierzą kanału łącza w górę $G^{l\mu} \in \mathbb{C}^{M\times K}$ od K użytkowników w komórce μ do stacji bazowej w komórce l. Równoważnie, odebrany sygnał na m-tej antenie stacji bazowej komórki l jest wyrażony jako

$$\begin{aligned}r_m^l &= \sqrt{P_r}\sum_{\mu=1}^{L}\sum_{k=1}^{K} g_{mk}^{l\mu}u_k^{\mu} + n_m^l \\ &= \underbrace{\sqrt{P_r}\sum_{k=1}^{K} g_{mk}^{ll}u_k^{l}}_{\text{Desired signal}} + \underbrace{\sqrt{P_r}\sum_{\mu=1,\mu\neq l}^{L}\sum_{k=1}^{K} g_{mk}^{l\mu}u_k^{\mu}}_{\text{Inter-cell interference}} + n_m^l.\end{aligned}$$

(109)

Załóżmy, że używamy dopasowanego filtrowania do wykrywania ul, l-ta stacja bazowa przetwarza odebrany sygnał, mnożąc go przez sprzężenie jego szacowanego CSI $\hat{\mathbf{G}}_{ll}$, patrz równanie (100), które można zapisać jako

$$\begin{aligned}\hat{\mathbf{G}}^{ll} &= \sum_{\mu=1}^{L} \mathbf{G}^{l\mu} + \mathbf{W}_l \\ &= \sum_{\mu=1}^{L} \mathbf{H}^{l\mu} \left(\mathbf{B}^{l\mu}\right)^{1/2} + \mathbf{W}_l.\end{aligned} \tag{110}$$

To powoduje

$$\begin{aligned}\mathbf{y}_l &= \left(\hat{\mathbf{G}}^{ll}\right)^H \mathbf{r}_l \\ &= \left[\sum_{\mu=1}^{L} \mathbf{G}^{l\mu} + \mathbf{W}_l\right]^H \left[\sqrt{P_r} \sum_{\mu'=1}^{L} \mathbf{G}^{l\mu'} \mathbf{u}_{\mu'} + \mathbf{n}_l\right]\end{aligned} \tag{111}$$

Według Marzetty

$$\frac{1}{M}\left[\mathbf{G}^{l\mu}\right]^H \mathbf{G}^{l\mu'} = (\mathbf{B}^{l\mu})^{1/2} \left(\frac{\left[\mathbf{H}^{l\mu}\right]^H \mathbf{H}^{l\mu'}}{M}\right) \left(\mathbf{B}^{l\mu'}\right)^{1/2} \tag{112}$$

W miarę jak liczba anten stacji bazowych rośnie bez ograniczeń M → +∞, mamy

$$\frac{\left[\mathbf{H}^{l\mu}\right]^H \mathbf{H}^{l\mu'}}{M} \longrightarrow \mathbf{I}_K \delta[\mu - \mu'].$$

(113)

Podstawienie równań (112) i (113) do równania (111) daje

$$\frac{1}{\sqrt{P_r} M} \mathbf{y}_l \longrightarrow \sum_{\mu=1}^{L} \mathbf{B}^{l\mu} \mathbf{u}_{\mu}$$

(114)

$$\frac{1}{\sqrt{P_r} M} y_k^l \longrightarrow \beta_k^{ll} u_k^l + \sum_{\mu=1, \mu \neq l}^{L} \beta_k^{l\mu} u_k^{\mu}$$

(115)

Zbawiennym efektem korzystania z nieograniczonej liczby anten stacji bazowych jest to, że efekty nieskorelowanego szumu odbiornika i szybkiego zanikania są całkowicie eliminowane, a transmisje z terminali w obrębie własnej komórki nie zakłócają się. Jednak transmisje z terminali w innych komórkach, które używają tej samej sekwencji pilota, stanowią resztkową interferencję. Efektywny stosunek sygnału do zakłóceń (SIR) wynosi

$$\gamma_k^l = \frac{\left(\beta_k^{ll}\right)^2}{\sum_{\mu=1, \mu \neq l}^{L} \left(\beta_k^{l\mu}\right)^2},$$

(116)

która jest wielkością losową zależną od losowego położenia zacisków i zanikania cienia.

Transmisja danych w łączu w dół

W łączu w dół wielokomórkowego systemu MIMO, μ-ta stacja bazowa transmituje wektor symboli niosących wiadomości $u_\mu = [u^{\mu}_1 , u^{\mu}_2 ,\ldots, u^{\mu K}]T$ przez macierz prekodowania w kierunku K terminali w swojej odpowiedniej komórce niezależnie, gdzie u^{μ}_k , $\forall k = 1, 2,\ldots, K$ i $\mu = 1, 2,\ldots, L$ oznacza symbol średniej zerowej, wariancji jednostkowej przeznaczony dla użytkownika k w μ-tej komórce. Stosując sprzężone formowanie wiązki, $u\mu$ jest mnożone przez sprzężenie jego oszacowania dla macierzy kanału. Zatem wektor transmitowanych symboli w μ-tej stacji bazowej jest obliczany przez

$$\mathbf{s}_\mu = \left(\hat{\mathbf{G}}^{\mu\mu}\right)^* \mathbf{u}_\mu, \tag{117}$$

gdzie $\hat{\mathbf{G}}^{\mu\mu} \in \mathbb{C}^{M\times K}$ oznacza oszacowanie macierzy kanału uplink pomiędzy K użytkownikami w komórce μ i stacją bazową w komórce μ. K użytkowników w l-tej komórce otrzymuje wektor K × 1 obejmujący otrzymane sygnały ze wszystkich L stacji bazowych, tj.

$$\begin{aligned}\mathbf{r}_l &= \sqrt{P_f}\sum_{\mu=1}^{L}\left(\mathbf{G}^{l\mu}\right)^T \mathbf{s}_\mu + \mathbf{n}_l \\ &= \sqrt{P_f}\sum_{\mu=1}^{L}\left(\mathbf{G}^{l\mu}\right)^T \left(\hat{\mathbf{G}}^{\mu\mu}\right)^* \mathbf{u}_\mu + \mathbf{n}_l \\ &= \sqrt{P_f}\sum_{\mu=1}^{L}\left(\mathbf{G}^{l\mu}\right)^T \left(\sum_{\mu'=1}^{L}\mathbf{G}^{\mu\mu'} + \mathbf{W}_\mu\right)^* \mathbf{u}_\mu + \mathbf{n}_l\end{aligned} \tag{118}$$

gdzie P_f jest ograniczeniem mocy łącza do przodu, $Gl\mu \in \mathbb{C}^{M\times K}$ oznacza macierz kanałów łącza w górę od K użytkowników w komórce μ do stacji bazowej w komórce l, a macierz kanałów łącza w dół jest równa $(G^{l\mu})^T$ ze względu na wzajemność kanałów. Ponieważ liczba anten stacji bazowych rośnie bez ograniczeń $M \to +\infty$, podobnie jak w równaniu (114), mamy

$$\frac{1}{\sqrt{P_f}M}\mathbf{r}_l \longrightarrow \sum_{\mu=1}^{L}\mathbf{B}^{\mu l}\mathbf{u}_\mu. \tag{119}$$

k-ty wpis przetworzonego sygnału staje się

$$\frac{1}{\sqrt{P_f}M}r_k^l \longrightarrow \sum_{\mu=1}^{L}\beta_k^{\mu l}u_k^\mu = \beta_k^{ll}u_k^l + \sum_{\mu=1,\mu\neq l}^{L}\beta_k^{\mu l}u_k^\mu. \tag{120}$$

Efektywny SIR wynosi zatem

$$\gamma_k^l = \frac{\left(\beta_k^{ll}\right)^2}{\sum_{\mu=1,\mu\neq l}^{L}\left(\beta_k^{\mu l}\right)^2} \tag{121}$$

Bezkomórkowe Massive MIMO

Stacja bazowa z dużym układem anten jednocześnie obsługuje wielu użytkowników w komórce sieci komórek w tym samym zasobie czasowo-częstotliwościowym, co jest obiecującą technologią dostępu bezprzewodowego. Dzięki prostemu przetwarzaniu sygnału może zapewnić wysoką przepustowość, niezawodność i energooszczędność. Ogromna liczba anten usługowych w komórce może być rozmieszczona w konfiguracjach kolokowanych lub rozproszonych. Kolokowane architektury Massive MIMO, w których wszystkie anteny usługowe znajdują się na zwartym obszarze, mają niskie wymagania dotyczące transmisji wstecznej i wspólnego przetwarzania. Niemniej jednak wysoką wydajność osiągają przede wszystkim użytkownicy, którzy pozostają w pobliżu centrów komórkowych. Tymczasem większość użytkowników na obrzeżach komórek ogranicza się do znacznie gorszej jakości usług z powodu zakłóceń międzykomórkowych i problemów z przekazywaniem, które są nieodłączną cechą architektury komórkowej. Zagęszczenie sieci w celu uzyskania dużej przepustowości systemu prowadzi również do poważnych zakłóceń międzykomórkowych i częstszych przekazywań. W związku z tym większość zatorów ruchu w sieciach komórkowych ma obecnie miejsce na obrzeżach komórek. Tak zwane 95%-prawdopodobne szybkości transmisji danych użytkownika, które mogą być zagwarantowane dla 95% użytkowników i w ten sposób określają wydajność odczuwaną przez użytkownika, pozostają przeciętne w sieciach 5G. Rozwiązaniem tych problemów może być połączenie każdego użytkownika z wieloma rozproszonymi antenami. Jeśli w sieci jest tylko jedna ogromna komórka, nie pojawiają się żadne zakłócenia międzykomórkowe i nie jest potrzebne przekazywanie. To rozwiązanie było badane w przeszłości, przy użyciu takich nazw, jak sieciowy MIMO, rozproszony MIMO, rozproszony układ antenowy i skoordynowana transmisja i odbiór wielopunktowy (CoMP). Dzięki ich zdolności do wykorzystywania różnorodności przestrzennej przeciwko zanikaniu cienia bardziej wydajnie, rozproszony system może oferować znacznie większe prawdopodobieństwo zasięgu niż system kolokowany kosztem zwiększonych wymagań dotyczących sieci szkieletowej. W Ngo i in. [2017] zaproponowano rozproszony system MIMO, w którym duża liczba anten usługowych obsługuje znacznie mniejszą liczbę autonomicznych użytkowników rozproszonych na dużym obszarze. Wszystkie anteny współpracują fazowo spójnie za pośrednictwem sieci fronthaul i obsługują wszystkich użytkowników w tym samym zasobie czasowo-częstotliwościowym. Aby uniknąć ogromnego narzutu związanego z pozyskiwaniem CSI, system działa w trybie TDD i wykorzystuje wzajemność kanałów. Nie ma komórek ani granic komórek. Dlatego ten system jest określany jako bezkomórkowy massive MIMO. Ponieważ ta konfiguracja łączy koncepcje distributedMIMO i massive MIMO, oczekuje się, że będzie czerpać wszystkie korzyści z tych dwóch systemów.

Układ sieci bezkomórkowej

Początkowa konfiguracja bezkomórkowego systemu massive MIMO składa się z M punktów dostępowych (AP) i K użytkowników, gdzie $M \gg K$. Wszystkie AP i terminale użytkowników są wyposażone w jedną antenę i losowo rozproszone na obszarze geograficznym. Ponadto wszystkie AP łączą się z jednostką centralną (CPU) za pośrednictwem sieci backhaul, jak pokazano na rysunku 9.4, a przepustowość backhaulingu jest nieograniczona, a transmisja jest wolna od błędów, aby skupić się na kodowaniu wstępnym i wykrywaniu. Wszystkie M AP jednocześnie obsługują wszystkich K użytkowników w tym samym zasobie czasowo-częstotliwościowym. Transmisja downlink od AP do użytkowników i transmisja uplink od użytkowników do AP są rozdzielone przez operację TDD. Każdy interwał koherencji jest podzielony na trzy fazy: szkolenie uplink, transmisja danych ładunku downlink i transmisja danych ładunku uplink. W fazie szkolenia uplink użytkownicy wysyłają sygnały referencyjne do AP, a każdy AP szacuje kanał dla wszystkich użytkowników niezależnie. Wykorzystując wzajemność kanałów systemu TDD, stacja bazowa zna wiedzę o kanale downlink z szacowanego CSI uplink. Uzyskane oszacowania kanału są wykorzystywane do wstępnego kodowania przesyłanych sygnałów w downlink i do wykrywania sygnałów przesyłanych od użytkowników w uplink. Możemy napisać

$$g_{mk} = \sqrt{\beta_{mk}} h_{mk} \quad (122)$$

modelowanie kanału zanikania między ogólnym AP m = 1,…,M a typowym sprzętem użytkownika (UE) k = 1,…, K, gdzie β_{mk} i h_{mk} reprezentują odpowiednio zanikanie na dużą i małą skalę. Zakłada się, że zanikanie na małą skalę jest płaskie pod względem częstotliwości i jest modelowane przez kołowo symetryczną zespoloną zmienną losową Gaussa o zerowej średniej i wariancji jednostkowej, tj. $h_{mk} \sim \mathcal{CN}(0, 1)$. Zanikanie na dużą skalę jest niezależne od częstotliwości i pozostaje stałe przez stosunkowo długi okres. Jest obliczane przez

$$\beta_{mk} = 10^{\frac{PL_{mk} + X_{mk}}{10}} \quad (123)$$

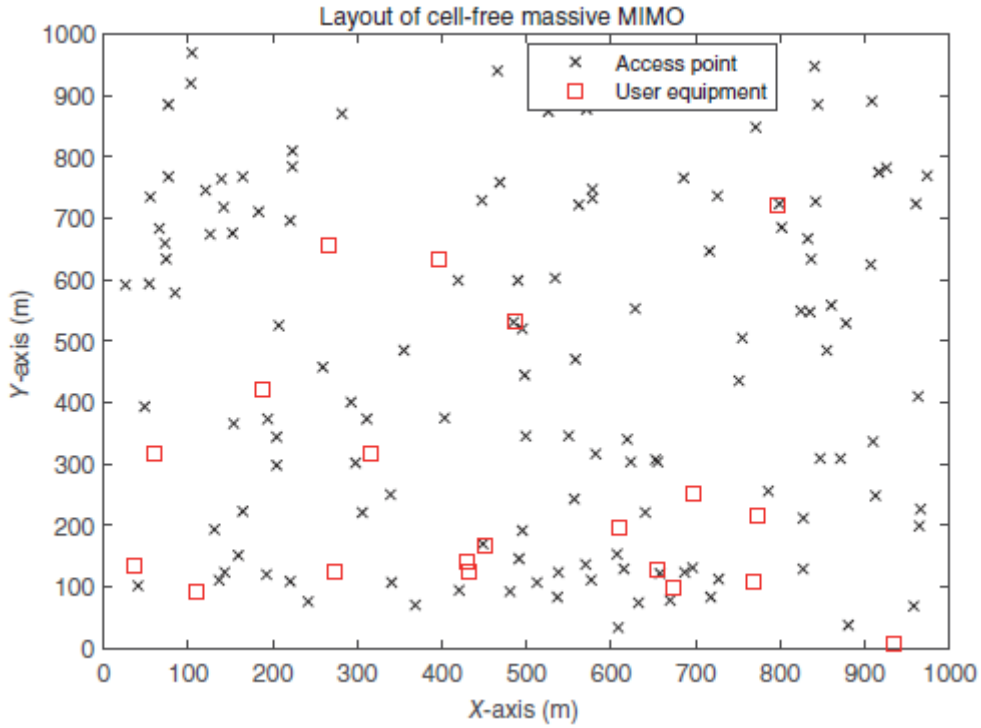
z zanikaniem cienia $X_{mk} \sim \mathcal{N}(0, \sigma_{sd}^2)$ i utratą sygnału PL_{mk}. Ngo i inni stosują model COST-Hata, tj.

$$PL_{mk} = \begin{cases} -L - 35 \log_{10}(d_{mk}), & d_{mk} > d_1 \\ -L - 15 \log_{10}(d_1) - 20 \log_{10}(d_{mk}), & d_0 < d_{mk} \leq d_1 \\ -L - 15 \log_{10}(d_1) - 20 \log_{10}(d_0), & d_{mk} \leq d_0 \end{cases} \quad (124)$$

gdzie d_{mk} oznacza odległość między AP_m i UE k, d0 i d_1 są punktami przerwania modelu trójzboczowego, a

$$L = 46.3 + 33.9 \log_{10}(f_c) - 13.82 \log_{10}(h_{AP}) - [1.1 \log_{10}(f_c) - 0.7] h_{UE} + 1.56 \log_{10}(f_c) - 0.8 \quad (125)$$

z częstotliwością nośną f_c, wysokością anteny AP h_{AP} i wysokością anteny UE h_{UE}. W przeciwieństwie do kolokowanego masywnego MIMO, gdzie zanikanie na dużą skalę od typowego terminala k do wszystkich anten stacji bazowych jest identyczne, oznaczone jako $\beta_{mk} = \beta_k$, ∀m = 1, 2,…,M, każda para anten między terminalem k i AP m ma unikalny β_{mk}. Przykładowy bezkomórkowy system masywnego MIMO składający się z M = 128 AP i K = 20 UE jest pokazany na rysunku 9.5.



Szkolenie łącza w górę

Jako system masywnego MIMO, typowy terminal k jest przypisany do ortogonalnej sekwencji odniesienia $\boldsymbol{\phi}k \in \mathbb{C}^{\tau p \times 1}$, gdzie $\tau_p \geq K$. Te terminale jednocześnie przesyłają swoje sygnały odniesienia przez τ_p jednostek zasobów czasowo-częstotliwościowych, co powoduje, że przesyłane sygnały

$$\mathbf{X}_p = \sqrt{p_u \tau_p}\boldsymbol{\Phi}^H \quad (126)$$

gdzie

$$\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_K]. \quad (127)$$

Następnie m-ty AP obserwuje otrzymany wektor symboli $1 \times \tau_p$

$$\mathbf{y}_m^p = \mathbf{g}_m \mathbf{X}_p + \mathbf{z}_m^p \quad (128)$$

gdzie $z^p{}_m$ odpowiada niezależnemu zespolonemu szumowi Gaussa $1 \times \tau_p$ przy każdym wpisie $z \sim \mathcal{CN}(0, \sigma_n^2), \mathbf{g}_m \in \mathbb{C}^{1\times k}$ modeluje sygnaturę przestrzenną od K terminali do AP m, którą można wyrazić za pomocą

$$\mathbf{g}_m = [g_{m1}, g_{m2}, \ldots, g_{mK}], \quad (129)$$

odpowiadający m-temu wierszowi macierzy kanału uplink $G_u \in \mathbb{C}^{M\times K}$. W przeciwieństwie do masywnego systemu MIMO, w którym stacja bazowa wykonuje wspólną estymację kanału, każdy AP w systemie bezkomórkowym szacuje swój własny sygnaturę przestrzenną g_m niezależnie

$$\begin{aligned}\tilde{\mathbf{y}}_m^p &= \mathbf{y}_m^p \mathbf{\Phi} = \mathbf{g}_m \mathbf{X}_p \mathbf{\Phi} + \mathbf{z}_m^p \mathbf{\Phi} \\ &= \sqrt{p_u \tau_p} \mathbf{g}_m \mathbf{\Phi}^H \mathbf{\Phi} + \mathbf{z}_m^p \mathbf{\Phi} \\ &= \sqrt{p_u \tau_p} \mathbf{g}_m + \tilde{\mathbf{z}}_m^p,\end{aligned} \quad (130)$$

gdzie każdy wpis $\tilde{\mathbf{z}}_m^p \in \mathbb{C}^{1\times K}$ jest również niezależnym zespolonym szumem Gaussa $\tilde{z} \sim \mathcal{CN}(0, \sigma_n^2)$ wynikającym z mnożenia przez macierz unitarną. Równoważnie mamy

$$\tilde{y}_{mk,p} = \sqrt{p_u \tau_p} g_{mk} + \tilde{z}_{mk}, \quad \forall k = 1, 2, \ldots, K. \quad (131)$$

Stosując liniową estymację MMSE, AP m otrzymuje oszacowania

$$\hat{g}_{mk} = \mathbb{E}\left[g_{mk} | \tilde{y}_{mk,p}\right] = \left(\frac{\sqrt{p_u \tau_p}\beta_{mk}}{p_u \tau_p \beta_{mk} + \sigma_n^2}\right) \tilde{y}_{mk,p}, \quad \forall k = 1, 2, \ldots, K.$$

Wariancję $\hat{g}_{mk}$ oblicza się za pomocą

$$\mathbb{E}\left[\left|\hat{g}_{mk}\right|^2\right] = \frac{p_u \tau_p \beta_{mk}^2}{p_u \tau_p \beta_{mk} + \sigma_n^2} = \alpha_{mk} \quad (132)$$

co daje $\hat{g}_{mk} \sim \mathcal{CN}(0, \alpha_{mk})$. Średnia skumulowana wartość szacunkowa kanału jest równa

$$\mathbb{E}\left[\left|\tilde{g}_{mk}\right|^2\right] = \epsilon_{mk} = \beta_{mk} - \alpha_{mk} = \frac{\sigma_n^2 \beta_{mk}}{p_u \tau_p \beta_{mk} + \sigma_n^2} \quad (133)$$

odpowiadające $\tilde{g}_{mk} \sim \mathcal{CN}(0, \epsilon_{mk})$. W następującym kodowaniu wstępnym i dekodowaniu zakłada się, że AP m zna oszacowanie własnego sygnatury przestrzennej

$$\hat{\mathbf{g}}_m = [\hat{g}_{m1}, \hat{g}_{m2}, \ldots, \hat{g}_{mK}]. \quad (134)$$

Wykrywanie sygnału łącza w górę

Podobnie jak w łączu w górę systemu MIMO komórkowego, terminale K w sieci bezkomórkowej jednocześnie przesyłają swoje symbole w kierunku M punktów dostępowych. Nie ma bezpośredniej współpracy między terminalami w celu wykonania wspólnego wstępnego kodowania. Jedyne, co te terminale mogą zrobić, to niezależnie ważyć swoje symbole. Przesyłane symbole u_k, k = 1, 2,…, K mają średnią zerową, wariancję jednostkową i są wzajemnie nieskorelowane, a zatem macierz kowariancji dla wektora $u = [u_1, u_2, \ldots, u_K]^T$ wynosi

$$\mathbb{E}\left[\mathbf{u}\mathbf{u}^H\right] = \mathbf{I}_K \quad (135)$$

Przekazywanym symbolem dla typowego terminala k jest $\sqrt{\eta_k} u_k$, gdzie η_k oznacza współczynnik sterowania mocą, spełniający warunek $0 \leqslant \eta_k \leqslant 1$. Zatem wektor odebranych symboli oblicza się za pomocą

$$\mathbf{r} = \sqrt{P_r}\mathbf{G}\mathbf{D}_\eta\mathbf{u} + \mathbf{n} \quad (136)$$

z ograniczeniem mocy P_r, macierzą kanału łącza wstecznego $G \in \mathbb{C}^{M\times K}$ i macierzą diagonalną D_η = diag{ $\sqrt{\eta_1},\ldots,\sqrt{\eta_K}$}. Równoważnie, obserwacja w m-tym punkcie dostępowym jest wyrażona wzorem

$$\begin{aligned} r_m &= \sqrt{P_r}\mathbf{g}_m\mathbf{D}_\eta\mathbf{u} + n_m \\ &= \sqrt{P_r}\sum_{k=1}^{K} g_{mk}\sqrt{\eta_k}u_k + n_m, \end{aligned} \quad (137)$$

gdzie $g_m = [g_{m1}, g_{m2},\ldots, g_{mK}]$ jest przestrzenną sygnaturą AP m, równoważną m-temu wierszowi G. Podobnie, trzy typowe algorytmy mogą być stosowane do odzyskiwania symboli informacyjnych, tj. dopasowane filtrowanie, wykrywanie wymuszania zera i wykrywanie minimalnego błędu średniokwadratowego.

Dopasowane filtrowanie

Aby wykryć u_k, m-ty AP wstępnie przetwarza odebrany sygnał, mnożąc go przez sprzężenie swojego lokalnego CSI $\hat{g}_{mk}$, i wysyła wynik $\hat{g}^*_{mk}r_m$ do CPU, co skutkuje

$$\begin{aligned} r_i &= \sum_{m=1}^{M}\hat{g}^*_{mi}r_m \\ &= \sum_{m=1}^{M}\hat{g}^*_{mi}\left(\sqrt{P_r}\sum_{k=1}^{K} g_{mk}\sqrt{\eta_k}u_k + n_m\right) \\ &= \underbrace{\sqrt{P_r}\sum_{m=1}^{M}\hat{g}^*_{mi}g_{mi}\sqrt{\eta_i}u_i}_{\text{Desired signal}} + \underbrace{\sqrt{P_r}\sum_{m=1}^{M}\hat{g}^*_{mi}\sum_{k=1,k\neq i}^{K} g_{mk}\sqrt{\eta_k}u_k}_{\text{Inter-user interference}} + \underbrace{\sum_{m=1}^{M}\hat{g}^*_{mi}n_m}_{\text{Noise}}. \end{aligned} \quad (138)$$

Traktując zakłócenia między użytkownikami jako kolorowy szum, procesor może uzyskać twardą ocenę $\hat{u}_i$ i dla ui, ∀i = 1, 2,…, K.

Wykrywanie ZF

W tym przypadku każdy AP musi wysłać swoją obserwację r_m i lokalny CSI $\hat{\mathbf{g}}_m$ do procesora za pośrednictwem sieci fronthaul. W konsekwencji procesor zna r i buduje macierz dekodowania jako $(\hat{\mathbf{G}}^H\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}^H$. W ten sposób wyjście przetwarzania końcowego to

$$\begin{aligned} \tilde{\mathbf{u}} &= (\hat{\mathbf{G}}^H\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}^H\mathbf{r} \\ &= \sqrt{P_r}(\hat{\mathbf{G}}^H\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}^H\mathbf{G}\mathbf{D}_\eta\mathbf{u} + (\hat{\mathbf{G}}^H\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}^H\mathbf{n}. \end{aligned} \quad (139)$$

Rozłożenie tego równania daje k-tą miękką ocenę

$$\tilde{u}_k = \underbrace{\sqrt{P_r\eta_k}u_k}_{\text{Desired signal}} + \underbrace{\mathbf{g}_k\mathbf{n}}_{\text{Noise}}, \quad (140)$$

gdzie $g_k \in \mathbb{C}^{1\times M}$ oznacza k-ty wiersz $(\hat{\mathbf{G}}^H\hat{\mathbf{G}})^{-1}\hat{\mathbf{G}}^H$. Zakłócenia między użytkownikami są całkowicie eliminowane, a twardą ocenę $\hat{u}_k$ można uzyskać z $\tilde{z}_k$.

Wykrywanie MMSE

Aby złagodzić efekt wzmocnienia szumu w wykrywaniu wymuszającym zero, procesor może zastosować metodę MMSE w celu zminimalizowania MSE szacowanych symboli. Każdy AP musi wysłać swoją obserwację r_m i lokalny CSI $\hat{\mathbf{g}}_m$ do procesora za pośrednictwem sieci fronthaul. W konsekwencji procesor zna r i buduje macierz dekodowania jako $(\hat{\mathbf{G}}^H\hat{\mathbf{G}}+\sigma_n^2\mathbf{I})^{-1}\hat{\mathbf{G}}^H$. Wyjście przetwarzania końcowego staje się

$$\begin{aligned}\tilde{\mathbf{u}} &= (\hat{\mathbf{G}}^H\hat{\mathbf{G}}+\sigma_n^2\mathbf{I})^{-1}\hat{\mathbf{G}}^H\mathbf{r} \\ &= \sqrt{P_r}(\hat{\mathbf{G}}^H\hat{\mathbf{G}}+\sigma_n^2\mathbf{I})^{-1}\hat{\mathbf{G}}^H\mathbf{G}\mathbf{D}_\eta\mathbf{u}+(\hat{\mathbf{G}}^H\hat{\mathbf{G}}+\sigma_n^2\mathbf{I})^{-1}\hat{\mathbf{G}}^H\mathbf{n}\end{aligned} \quad (141)$$

Formowanie wiązki sprzężonej

W łączu wstecznym bezkomórkowego systemu MIMO o dużej masie stosuje się dwie typowe metody transmisji, tj. formowanie wiązki sprzężonej i wstępne kodowanie z wymuszeniem zera, w celu multipleksowania przestrzennego symboli niosących informacje przeznaczonych dla terminali K, oznaczonych jako $u = [u_1, u_2,\ldots, u_K]^T$, gdzie symbole są znormalizowane

$$\mathbb{E}[|u_k|^2] = 1, \quad k = 1, 2, \ldots, K \quad (142)$$

Bezkomórkowy system MIMO masywny stosujący sprzężone formowanie wiązki działa w następujący sposób:

● Typowy AP m mierzy β_{mk}, k = 1, 2,…, K i raportuje je do CPU. Zwykle zanikanie na dużą skalę pozostaje stałe przez stosunkowo długi okres w odniesieniu do czasu koherencji kanału. W związku z tym wiedzę o β_{mk} można uznać za doskonałą, a narzut pomiaru i dystrybucji jest niewielki.

● CPU oblicza współczynniki sterowania mocą η_{mk}, ∀m, ∀k jako funkcję β_{mk} i wysyła je do odpowiednich AP. W międzyczasie CPU dystrybuuje symbole niosące informacje u do wszystkich AP. Należy zauważyć, że β_{mk} może być takie samo dla dziesiątek okresów symboli, dlatego też do przesłania jest tylko u.

● Użytkownicy synchronicznie przesyłają swoje sekwencje pilotażowe $\boldsymbol{\phi}$k, k = 1,…, K o czasie trwania τ_p.

● m-ty AP, gdzie m = 1,…,M, uzyskuje oszacowanie własnego sygnatury przestrzennej $\hat{\mathbf{g}}_m = [\hat{g}_{m1}, \hat{g}_{m2}, \ldots, \hat{g}_{mK}]^T$

● AP traktują oszacowania kanału jako prawdziwe kanały i używają sprzężonego kształtowania wiązki do generowania przesyłanych sygnałów. m-ty AP wysyła sygnał

$$s_m = \sqrt{\eta_{mk}P_m}\hat{\mathbf{g}}_m^H\mathbf{u} = \sqrt{P_m}\sum_{k=1}^{K}\sqrt{\eta_{mk}}\hat{g}_{mk}^* u_k \quad (143)$$

gdzie P_m jest limitem mocy nadawczej APm. Wybór współczynników sterowania mocą podlega

$$\mathbb{E}[|s_m|^2] \leqslant P_m, \quad \forall m = 1, 2, \ldots, M \tag{144}$$

co można zinterpretować jako

$$\sum_{k=1}^{K} \eta_{mk} \mathbb{E}[|\hat{g}_{mk}|^2] \leqslant 1 \tag{145}$$

lub

$$\sum_{k=1}^{K} \eta_{mk} \leqslant \frac{1}{\mathbb{E}[|\hat{g}_{mk}|^2]} = \frac{p_u \tau_p \beta_{mk} + \sigma_n^2}{p_u \tau_p \beta_{mk}^2} \tag{146}$$

stosując równania (132) i (142). Następnie obserwacja i-tego użytkownika, ∀i = 1, 2,…, K jest

$$\begin{aligned} r_i &= \sum_{m=1}^{M} g_{mi} s_m + n_i \\ &= \sum_{m=1}^{M} g_{mi} \left(\sqrt{P_m} \sum_{k=1}^{K} \sqrt{\eta_{mk}} \hat{g}_{mk}^* u_k \right) + n_i \\ &= \underbrace{\sum_{m=1}^{M} \sqrt{P_m \eta_{mi}} g_{mi} \hat{g}_{mi}^* u_i}_{} + \underbrace{\sum_{m=1}^{M} \sqrt{P_m} \sum_{k=1, k \neq i}^{K} \sqrt{\eta_{mk}} g_{mi} \hat{g}_{mk}^* u_k}_{} + n_i. \end{aligned}$$

Zbawiennym efektem używania nieograniczonej liczby anten stacji bazowych jest to, że efekty nieskorelowanego szumu i szybkiego zaniku kanału znikają. Jest to wynik utwardzania kanału w systemach Massive MIMO. W konsekwencji, wykrywanie odebranych sygnałów odbywa się pod warunkiem, że ogólny użytkownik k jest świadomy jedynie statystyk szacowanych współczynników kanału, tj. $\mathbb{E}\left[|\hat{g}_{mk}|^2\right] = \alpha_{mk}, \forall m = 1, 2, \ldots, M$, ponieważ w łączu w dół nie ma sygnałów odniesienia ani szacowania kanału. Tak więc obserwację i-tego użytkownika można zapisać jako (zakładając, że ograniczenie mocy każdego AP jest identyczne, tj. $P_m = P_f$, ∀m = 1, 2,…,M dla zwięzłości)

$$
\begin{aligned}
r_i &= \sum_{m=1}^{M} \sqrt{P_f \eta_{mi}} [\hat{g}_{mi} + \bar{g}_{mi}] \hat{g}_{mi}^* u_i + \sum_{m=1}^{M} \sum_{k=1, k\neq i}^{K} \sqrt{P_f \eta_{mk}} [\hat{g}_{mi} + \bar{g}_{mi}] \hat{g}_{mk}^* u_k + n_i \\
&= \sum_{m=1}^{M} \sqrt{P_f \eta_{mi}} |\hat{g}_{mi}|^2 u_i + \sum_{m=1}^{M} \sum_{k=1, k\neq i}^{K} \sqrt{P_f \eta_{mk}} \hat{g}_{mi} \hat{g}_{mk}^* u_k \\
&\quad + \sum_{m=1}^{M} \sum_{k=1}^{K} \sqrt{P_f \eta_{mk}} \bar{g}_{mi} \hat{g}_{mk}^* u_k + n_i \\
&= \underbrace{\sum_{m=1}^{M} \sqrt{P_f \eta_{mi}} \alpha_{mi} u_i}_{S_0:\ \text{Useful signal}} + \underbrace{\sum_{m=1}^{M} \sqrt{P_f \eta_{mi}} (|\hat{g}_{mi}|^2 - \alpha_{mi}) u_i}_{\mathcal{I}_1:\ \text{No CSI at user}} \\
&\quad + \underbrace{\sum_{m=1}^{M} \sum_{k=1, k\neq i}^{K} \sqrt{P_f \eta_{mk}} \hat{g}_{mi} \hat{g}_{mk}^* u_k}_{\mathcal{I}_2:\ \text{Multi-user interference}} + \underbrace{\sum_{m=1}^{M} \sum_{k=1}^{K} \sqrt{P_f \eta_{mk}} \bar{g}_{mi} \hat{g}_{mk}^* u_k}_{\mathcal{I}_3:\ \text{CSI estimate error}} + \underbrace{n_i}_{\mathcal{N}_4}
\end{aligned}
\tag{147}
$$

Ponieważ symbole informacyjne przeznaczone dla różnych użytkowników są niezależne, a szum gaussowski addytywny nie jest skorelowany z symbolami informacyjnymi i realizacjami kanałów, terminy S_0, I_1,I_2,I_3 i N_4 są wzajemnie nieskorelowane. Według Hassibi i Hochwaldnajgorszym przypadkiem szumu dla wzajemnej informacji jest szum gaussowski addytywny z wariancją kwalifikującą się do wariancji ▪ I1 + I2 + I3 + NI4. Zatem osiągalna szybkość łącza wstecznego dla użytkownika k jest ograniczona dolną granicą

$$
R_i = \log(1 + \gamma_i), \tag{148}
$$

gdzie

$$
\begin{aligned}
\gamma_i &= \frac{\mathbb{E}\left[|S_0|^2\right]}{\mathbb{E}\left[|\mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{N}_4|^2\right]} \\
&= \frac{\mathbb{E}\left[|S_0|^2\right]}{\mathbb{E}\left[|\mathcal{I}_1|^2\right] + \mathbb{E}\left[|\mathcal{I}_2|^2\right] + \mathbb{E}\left[|\mathcal{I}_3|^2\right] + \mathbb{E}\left[|\mathcal{N}_4|^2\right]}
\end{aligned}
\tag{149}
$$

z

$$
\mathbb{E}\left[|S_0|^2\right] = P_f \left(\sum_{m=1}^{M} \sqrt{\eta_{mi}} \alpha_{mi} \right)^2
$$

$$
\mathbb{E}\left[|\mathcal{I}_1|^2\right] = P_f \sum_{m=1}^{M} \eta_{mi} \alpha_{mi}^2
$$

$$
\mathbb{E}\left[|\mathcal{I}_2|^2\right] = P_f \sum_{m=1}^{M} \sum_{k=1, k\neq i}^{K} \eta_{mk} \alpha_{mi} \alpha_{mk}
$$

$$
\mathbb{E}\left[|\mathcal{I}_3|^2\right] = P_f \sum_{m=1}^{M} \sum_{k=1}^{K} \eta_{mk} \epsilon_{mi} \alpha_{mk}
$$

(15), (151), (152), (153)

Podstawiając równania od (149) do (153) do równania (148), otrzymujemy

$$R_i = \log\left(1 + \frac{P_f\left(\sum_{m=1}^{M}\sqrt{\eta_{mi}}\alpha_{mi}\right)^2}{\sigma_n^2 + P_f\sum_{m=1}^{M}\sum_{k=1}^{K}\eta_{mk}\beta_{mi}\alpha_{mk}}\right). \quad (154)$$

Zero-Forcing Precoding

Filozofia stojąca za prekodowaniem ZF polega na całkowitym wyeliminowaniu zakłóceń między różnymi użytkownikami, biorąc pod uwagę znajomość kanałów łącza wstecznego. System Massive MIMO bez komórek stosujący prekodowanie ZF działa w następujący sposób:

● AP m mierzy β_{mk}, k = 1, 2,…, K i raportuje je do CPU.

● Jako sprzężone formowanie wiązki, CPU oblicza współczynniki sterowania mocą w kategoriach β_{mk}. Konieczne jest, aby $\eta_{1k} = \cdots = \eta_{Mk}$, ∀k, a zatem współczynniki mocy powinny być tylko funkcjami k, tj. $\eta_{mk} = \eta k$.

● Użytkownicy synchronicznie przesyłają swoje sekwencje pilotażowe $\boldsymbol{\phi}_k$, k = 1,…, K.

● M-ty AP, gdzie m = 1,…,M, uzyskuje oszacowanie własnego sygnatury przestrzenne $\hat{\mathbf{g}}_m = [\hat{g}_{m1}, \hat{g}_{m2}, \cdots, \hat{g}_{mK}]^T$.

● Każdy AP wysyła swój lokalny CSI do CPU, a zatem CPU otrzymuje globalny CSI $\hat{\mathbf{G}} = [\hat{\mathbf{g}}_1, \hat{\mathbf{g}}_2, \cdots, \hat{\mathbf{g}}_M] \in \mathbb{C}^{K\times M}$.

● CPU wspólnie koduje symbole niosące informacje w kategoriach

$$\mathbf{s} = \hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\mathbf{D}_\eta\mathbf{u}, \quad (155)$$

gdzie $\mathbf{D}_\eta \in \mathbb{C}^{K\times K}$ jest macierzą diagonalną składającą się ze współczynników sterowania mocą, tj. $\mathbf{D}_\eta$ = diag{ $\sqrt{\eta_1}$,…, $\sqrt{\eta_K}$}.

● CPU dystrybuuje wstępnie zakodowany symbol s_m do AP m, a te AP synchronicznie wysyłają swoje odpowiednie transmitowane symbole w kierunku użytkowników. Następnie wektor odebranych symboli można zapisać jako

$$\begin{aligned}\mathbf{r} &= \sqrt{P_f}\mathbf{G}\mathbf{s} + \mathbf{n} \\ &= \sqrt{P_f}\mathbf{G}\hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\mathbf{D}_\eta\mathbf{u} + \mathbf{n}.\end{aligned} \quad (156)$$

Równoważnie, i-ty użytkownik obserwuje

$$
\begin{aligned}
r_i &= \sqrt{P_f}\mathbf{g}_i\mathbf{s} + n_i \\
&= \sqrt{P_f}\mathbf{g}_i\hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\mathbf{D}_\eta\mathbf{u} + n_i \\
&= \sqrt{P_f}\left(\hat{\mathbf{g}}_i + \tilde{\mathbf{g}}_i\right)\hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\mathbf{D}_\eta\mathbf{u} + n_i \\
&= \underbrace{\sqrt{P_f\eta_i}u_i}_{S_0:\ \text{Useful signal}} + \underbrace{\sqrt{P_f}\tilde{\mathbf{g}}_i\hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\mathbf{D}_\eta\mathbf{u}}_{I_1:\ \text{CSI estimate error}} + \underbrace{n_i}_{I_2:\ \text{Noise}},
\end{aligned}
\tag{157}
$$

gdzie gi ∈ $\mathbb{C}^{1\times M}$ = [g_{1i}, g_{2i},…, g_{Mi}] oznacza rzeczywistą sygnaturę kanału dla użytkownika i, która jest i-tym wierszem macierzy kanału G $\hat{\mathbf{g}}_i \in \mathbb{C}^{1\times M} = [\hat{g}_{1i}, \hat{g}_{2i}, \ldots, \hat{g}_{Mi}]$ wyraża oszacowanie gi i odpowiadający mu błąd oszacowania $\tilde{\mathbf{g}}_i = \mathbf{g}_i - \hat{\mathbf{g}}_i$.Ze względu na niezależność przekazywanych symboli, szumu addytywnego i realizacji kanału, terminy ▪ S_0, I_1 i I_2 są wzajemnie nieskorelowane. Na podstawie najgorszego przypadku nieskorelowanego szumu addytywnego , osiągalna szybkość użytkownika i z prekodowaniem ZF jest ograniczona dolną granicą

$$
R_i^{\mathrm{ZF}} = \log\left(1 + \gamma_i^{\mathrm{ZF}}\right) \tag{158}
$$

gdzie

$$
\gamma_i^{\mathrm{ZF}} = \frac{\mathbb{E}\left[|S_0|^2\right]}{\mathbb{E}\left[|I_1|^2\right] + \mathbb{E}\left[|I_2|^2\right]}. \tag{159}
$$

Według Nayebi i innych wariancję I_1 można obliczyć jako

$$
\begin{aligned}
\mathbb{E}\left[|I_1|^2\right] &= P_f\mathbb{E}\left[\left|\tilde{\mathbf{g}}_i\hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\mathbf{D}_\eta\mathbf{u}\right|^2\right] \\
&= P_f tr\left(\mathbf{D}_\eta^2\mathbb{E}[\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\hat{\mathbf{G}}\mathbb{E}\left[\tilde{\mathbf{g}}_i^H\tilde{\mathbf{g}}_i\right]\hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}]\right).
\end{aligned}
\tag{160}
$$

Piszemy χ^i_k, k = 1, 2,…, K aby oznaczyć k-ty element przekątnej macierzy K × K dedykowany użytkownikowi i:

$$
\mathbb{E}\left[\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\hat{\mathbf{G}}\mathbb{E}\left[\tilde{\mathbf{g}}_i^H\tilde{\mathbf{g}}_i\right]\hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\right] \tag{161}
$$

gdzie $\mathbb{E}\left[\tilde{\mathbf{g}}_i^H\tilde{\mathbf{g}}_i\right]$ jest macierzą diagonalną z ϵmi na m-tym elemencie diagonalnym, tj.

$$
\mathbb{E}\left[\tilde{\mathbf{g}}_i^H\tilde{\mathbf{g}}_i\right] = \begin{bmatrix} \epsilon_{1i} & 0 & \ldots & 0 \\ 0 & \epsilon_{2i} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \epsilon_{Mi} \end{bmatrix}. \tag{162}
$$

Następnie równanie (159) można wyrazić dalej za pomocą

$$\gamma_i^{ZF} = \frac{P_f \eta_i}{\sigma_n^2 + P_f \sum_{k=1}^{K} \eta_k \chi_k^i}. \quad (163)$$

Wpływ starzenia się kanału

W bezkomórkowych systemach Massive MIMO liniowe prekodowanie jest implementowane głównie poprzez sprzężone formowanie wiązki i wstępne kodowanie z zerowym wymuszaniem. Pierwsze z nich wykorzystuje lokalny CSI do niezależnego wytwarzania sygnałów przesyłanych w każdym punkcie dostępowym. Jest proste i ma niskie wymagania dotyczące transmisji wstecznej, ale cierpi na zakłócenia między użytkownikami. Stąd sprzężone formowanie wiązki jest gorsze od wstępnego kodowania z zerowym wymuszaniem pod względem wydajności widmowej i energetycznej. Jednak w architekturze bezkomórkowej wstępne kodowanie z zerowym wymuszaniem wymaga wymiany natychmiastowego CSI i wstępnie zakodowanych danych między procesorem a punktami dostępowymi za pośrednictwem sieci fronthaul. Oprócz wysokiej złożoności implementacji i znacznego obciążenia siecią wsteczną powoduje znaczne opóźnienia propagacji i przetwarzania. W praktyce wydajność systemu jest podatna na takie opóźnienie, ponieważ wiedza o CSI szybko się dezaktualizuje, co określa się jako starzenie się kanału, podlegając zanikowi kanału i niedoskonałemu sprzętowi.

Starzenie się kanału

Ze względu na opóźnienia przetwarzania i propagacji istnieje luka czasowa między momentem, w którym sygnały odniesienia brzmią w kanałach łącza w górę, a momentem, w którym następuje transmisja danych łącza w dół na podstawie zmierzonego CSI. Uzyskany CSI może być nieaktualny w wyniku wahań kanałów wywołanych przez mobilność użytkownika i szum fazowy.

Mobilność użytkownika. Względny ruch między AP i UE, a także otaczającymi je reflektorami, prowadzi do kanału zmieniającego się w czasie. Biorąc pod uwagę prędkość ruchu v_k typowego UE k, jego maksymalne przesunięcie Dopplera uzyskuje się za pomocą $f^k_d = v_k/\lambda$, gdzie λ reprezentuje długość fali częstotliwości nośnej. Im wyższa mobilność, tym szybciej zmienia się kanał. Aby określić ilościowo starzenie się CSI wywołane efektem Dopplera, stosuje się metrykę znaną jako współczynnik korelacji, zgodnie z definicją Jiang i Schotten:

$$\rho_k = \frac{\mathbb{E}\left[h_{mk,d} h^*_{mk,p}\right]}{\sqrt{\mathbb{E}[|h_{mk,p}|^2]\mathbb{E}[|h_{mk,d}|^2]}}, \quad (164)$$

gdzie $h_{mk,p}$ i $h_{mk,d}$ oznaczają zanikanie kanału na małą skalę między AP m i UE k w momentach treningu łącza w górę (oznaczonego jako p) i rzeczywistej transmisji danych łącza w dół (oznaczonej jako d), odpowiednio. Zgodnie z klasycznym widmem Dopplera modelu Jakesa przyjmuje wartość

$$\rho_k = J_0(2\pi f_d^k \Delta\tau) \quad (165)$$

gdzie $\Delta\tau$ oznacza całkowite opóźnienie, a $J_0(\cdot)$ oznacza funkcję Bessela zerowego rzędu pierwszego rodzaju. Według Jiang mamy

$$h_{mk,d} = \left(\rho_k h_{mk,p} + \kappa_{mk}\sqrt{1-\rho_k^2}\right) \quad (166)$$

ze składnikiem innowacyjnym κ_{mk}, który jest zmienną losową o standardowym rozkładzie normalnym $\kappa_{mk} \sim \mathcal{CN}(0,1)$.

Szum fazowy. Jest atrakcyjny dla ekonomicznej implementacji systemów massive MIMO z tanimi transceiverami, jednocześnie podnosząc problem uszkodzeń sprzętowych. Tymczasem każdy rozproszony AP w bezkomórkowym systemie massive MIMO musi obsługiwać lokalny oscylator, w przeciwieństwie do wspólnego oscylatora w kolokowanej konfiguracji massive MIMO. Ze względu na niedoskonałe oscylatory w nadajniku, przesyłane sygnały cierpią z powodu szumu fazowego podczas przetwarzania konwersji w górę z sygnałów pasma podstawowego do sygnałów pasma przepustowego i odwrotnie w odbiorniku. Taki szum fazowy jest nie tylko losowy, ale także zmienny w czasie, co prowadzi do przestarzałego CSI, który jest równoważny z mobilnością użytkownika. Wykorzystując dobrze ugruntowany proces Wienera , szum fazowy m-tego punktu dostępowego i k-tego użytkownika w dyskretnej chwili czasu t można modelować jako

$$\begin{cases} \phi_{m,t} = \phi_{m,t-1} + \Delta\phi_t, & \Delta\phi_t \sim \mathcal{CN}(0,\sigma_\phi^2) \\ \varphi_{k,t} = \varphi_{k,t-1} + \Delta\varphi_t, & \Delta\varphi_t \sim \mathcal{CN}(0,\sigma_\varphi^2), \end{cases} \quad (167)$$

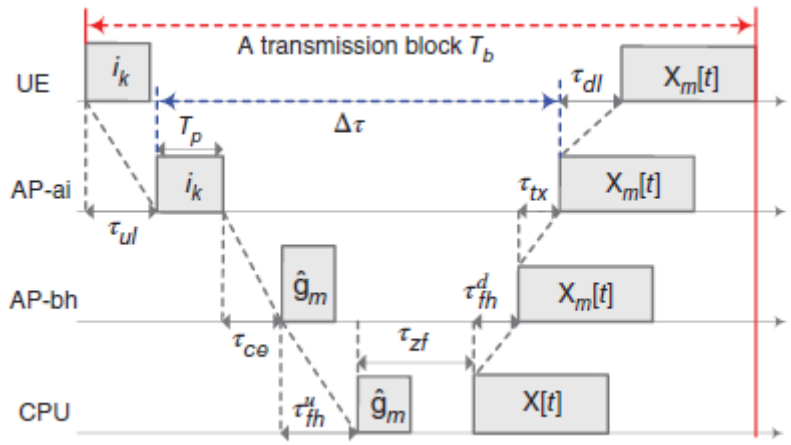
gdzie wariancje przyrostu są podane przez $\sigma^2_i = 4\pi^2 f_c c_i T_s$, $\forall i = \phi, \varphi$ z okresem symbolu Ts i stałą zależną od oscylatora ci. Do tej pory możemy zapisać

$$g_{mk,t} = \sqrt{\beta_{mk}} h_{mk,t} e^{j(\phi_{m,t}+\varphi_{k,t})} \quad (168)$$

aby oznaczyć całkowite wzmocnienie kanału między AP m i UE k w chwili t, łącząc efekty utraty ścieżki, zacieniania, zanikania na małą skalę i szumu fazowego. W szczególności uzyskany CSI $g_{mk,p} = \sqrt{\beta_{mk}} h_{mk,p} e^{j(\phi_{m,p}+\varphi_{k,p})}$ jest przestarzałą wersją jego rzeczywistej wartości $g_{mk,d} = \sqrt{\beta_{mk}} h_{mk,d} e^{j(\phi_{m,d}+\varphi_{k,d})}$. W dobrych warunkach, gdy kanały wykazują powolne zanikanie przy niskiej mobilności, a jakość oscylatorów jest wysoka, efekt starzenia się kanału nie jest wyraźny, a utrata wydajności może być niewielka. W przeciwnym razie wpływ powinien być poważny albo w środowiskach o szybkim zanikaniu, albo przy wykorzystaniu taniego sprzętu.

Opóźnienia propagacji i przetwarzania

Załóżmy, że AP i UE są dobrze zsynchronizowane, wiedza o β_{mk} jest w pełni dostępna, a sieć fronthaul zapewnia bezbłędną i nieskończoną przepustowość. Jak pokazano na rysunku ,



opóźnienia propagacji i przetwarzania można modelować w następujący sposób:

● Użytkownicy jednocześnie przesyłają swoje sygnały referencyjne i_k, k = 1,…, K w kierunku AP z czasem trwania T_p. Opóźnienie propagacji wynosi τ_{ul}.

● AP szacuje swój własny podpis kanału $\hat{g}_{mk,p}$, ∀k z czasem przetwarzania τ_{ce}.

● AP m wysyła swój lokalny CSI $\hat{\mathbf{g}}_m = \left[\hat{g}_{m1,p}, \cdots, \hat{g}_{mK,p}\right]^T \in \mathbb{C}^{K\times 1}$ do CPU, co powoduje opóźnienie propagacji τ^u_{fh}

.● Używając $\hat{\mathbf{G}} = [\hat{\mathbf{g}}_1, \cdots, \hat{\mathbf{g}}_M] \in \mathbb{C}^{K\times M}$, CPU prekoduje blok informacji zawierający symbole U ∈ $\mathbb{C}^{K\times NT}$, gdzie N_T oznacza liczbę symboli na użytkownika. Przekazywany blok symboli jest podany przez $\mathbf{X} = \hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\mathbf{D}_\eta\mathbf{U}$. Prekodowanie kosztuje czas τZF.

● Procesor CPU dystrybuuje wstępnie zakodowany wektor symboli x_m ∈ $\mathbb{C}^{1\times NT}$ do AP m, używając czasu τ^d_{fh}

.● Nadajnik AP m potrzebuje czasu przygotowania τ_{tx}, aby rozpocząć transmisję po odebraniu x_m, a propagacja sygnału trwa τ_{dl}.

W szczególności niech Δτ oznacza przerwę między czasem, gdy sygnały odniesienia badają kanały, a momentem, w którym wszystkie AP synchronicznie przesyłają wstępnie zakodowane symbole. Jak pokazano na rysunku powyżej, otrzymujemy

$$\Delta\tau = T_p + \tau_{ce} + \tau^u_{fh} + \tau_{\mathrm{ZF}} + \tau^d_{fh} + \tau_{tx}, \quad (169)$$

który jest znormalizowany przez okres próbkowania do $n_{\Delta\tau} = \left\lceil \frac{\Delta\tau}{T_s} \right\rceil$

Degradacja wydajności

Zgodnie z równaniem (168) całkowity CSI podczas transmisji danych w łączu wstecznym jest podany przez

$$g_{mk,d} = \sqrt{\beta_{mk}} h_{mk,d} e^{j(\phi_{m,d}+\varphi_{k,d})}, \quad (170)$$

Piszemy $\hat{g}_{mk,p}$, aby oznaczyć oszacowanie $g_{mk,p}$, a następnie błąd oszacowania wywołany przez szum addytywny obliczany jest przez

$$\tilde{g}_{mk,p} = g_{mk,p} - \hat{g}_{mk,p} \quad (171)$$

Składnik innowacji w równaniu (166) odpowiada elementowi złożonemu w ogólnym CSI, który jest zapisany jako $e_{mk} = \sqrt{\beta_{mk}}\kappa_{mk}e^{j(\phi^{m,p}+\varphi^{k,p})}$. Podstawienie równania (166) do równania (170) i zastosowanie równania (171) daje

$$g_{mk,d} = \sqrt{\beta_{mk}}\left(\rho_k h_{mk,p} + \kappa_{mk}\sqrt{1-\rho_k^2}\right) e^{j(\phi_{m,d}+\varphi_{k,d}+\phi_{m,p}+\varphi_{k,p}-\phi_{m,p}-\varphi_{k,p})}$$
$$= \left(\rho_k g_{mk,p} + e_{mk}\sqrt{1-\rho_k^2}\right) e^{j(\phi_{m,d}+\varphi_{k,d}-\phi_{m,p}-\varphi_{k,p})}$$
$$= \left(\rho_k \hat{g}_{mk,p} + \rho_k \tilde{g}_{mk,p} + e_{mk}\sqrt{1-\rho_k^2}\right) e^{j(\phi_{m,d}-\phi_{m,p})} e^{j(\varphi_{k,d}-\varphi_{k,p})}. \quad (172)$$

Oznaczmy k-ty wiersz $\hat{\mathbf{G}}$ jako $\hat{\mathbf{g}}_k = [\hat{g}_{1k,p}, \ldots, \hat{g}_{Mk,p}]$, $\tilde{\mathbf{g}}_k = [\tilde{g}_{1k,p}, \tilde{g}_{2k,p}, \ldots, \tilde{g}_{Mk,p}]$, $\mathbf{e}_k = [e_{1k}, \ldots, e_{Mk}]$, i macierz diagonalną

$$\Delta\,\boldsymbol{\Phi} = \mathrm{diag}\{e^{j(\phi_{1,d}-\phi_{1,p})}, \ldots, e^{j(\phi_{M,d}-\phi_{M,p})}\} \quad (173)$$

Zbudowanie wektora kanałowego $g_{k,d} = [g_{1k,d}, \ldots, g_{Mk,d}] \in \mathbb{C}^{1\times M}$ i podstawienie do niego równania (172) daje

$$\mathbf{g}_{k,d} = e^{j(\varphi_{k,d}-\varphi_{k,p})}\left(\rho_k \hat{\mathbf{g}}_k + \rho_k \tilde{\mathbf{g}}_k + \sqrt{1-\rho_k^2}\mathbf{e}_k\right) \Delta\,\boldsymbol{\Phi}. \quad (174)$$

W przypadku nieaktualnego CSI odebrany sygnał użytkownika k podany w równaniu (157) można zapisać jako

$$r_k = \sqrt{P_f}\mathbf{g}_{k,d}\mathbf{s} + n_k$$
$$= \sqrt{P_f}\mathbf{g}_{k,d}\hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\mathbf{D}_\eta\mathbf{u} + n_k$$
$$= \sqrt{P_f}e^{j(\varphi_{k,d}-\varphi_{k,p})}\left(\rho_k\hat{\mathbf{g}}_k + \rho_k\tilde{\mathbf{g}}_k + \sqrt{1-\rho_k^2}\mathbf{e}_k\right)$$
$$\times\,\Delta\,\boldsymbol{\Phi}\hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\mathbf{D}_\eta\mathbf{u} + n_k$$
$$= \underbrace{\sqrt{P_f\eta_k}e^{j(\varphi_{k,d}-\varphi_{k,p})}e^{-\frac{n_{\Delta\tau}\sigma_\phi^2}{2}}\rho_k u_k}_{D_0:\ \text{desired signal}}$$
$$+ \underbrace{\sqrt{P_f}e^{j(\varphi_{k,d}-\varphi_{k,p})}\rho_k\tilde{\mathbf{g}}_k\,\Delta\,\boldsymbol{\Phi}\hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\mathbf{D}_\eta\mathbf{u}}_{I_1:\ \text{effective noise}}$$
$$+ \underbrace{\sqrt{P_f(1-\rho_k^2)}e^{j(\varphi_{k,d}-\varphi_{k,p})}\mathbf{e}_k\,\Delta\,\boldsymbol{\Phi}\hat{\mathbf{G}}^H\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\mathbf{D}_\eta\mathbf{u}}_{I_2:\ \text{effective noise}} + \underbrace{n_k}_{I_3}. \quad (175)$$

Podczas wyprowadzania $T_{PN} = \lim_{M\to\infty} 1/M\ \mathrm{tr}\{\Delta\boldsymbol{\Phi}\} = e^{-n_\Delta \tau\sigma^2\phi/2}$ jest stosowane zgodnie z Krishnan , co oznacza, że szum fazowy twardnieje do wartości deterministycznej, gdy $M \to \infty$. Symbole informacyjne, błędy oszacowania, składniki innowacji i szum addytywny są niezależne, tak że człony D_0, I_1, I_2 i I_3 w równaniu (175) są wzajemnie nieskorelowane. Stosując fakt, że niekorygowany szum gaussowski reprezentuje najgorszy przypadek , osiągalna szybkość dla użytkownika k jest ograniczona dolną granicą $\log2\,(1+\gamma_k)$ przy efektywnym stosunku sygnału do zakłóceń i szumu (SINR)

$$\gamma_k = \frac{\mathbb{E}\left[|\mathcal{D}_0|^2\right]}{\mathbb{E}\left[|\mathcal{I}_1|^2\right] + \mathbb{E}\left[|\mathcal{I}_2|^2\right] + \mathbb{E}\left[|\mathcal{I}_3|^2\right]} \quad (176)$$

Łatwo to rozgryźć

$$\mathbb{E}\left[|\mathcal{D}_0|^2\right] = P_f \eta_k \rho_k^2 e^{-n_{\Delta\tau}\sigma_\phi^2} \quad (177)$$

i $\mathbb{E}[|I_3|^2] = \sigma^2_n$. Podobnie jak w równaniu (160), wariancję I_1 oblicza się za pomocą

$$\begin{aligned}
\mathbb{E}\left[|\mathcal{I}_1|^2\right] &= \mathbb{E}\left[\left|\sqrt{P_f} e^{j(\varphi_{k,d}-\varphi_{k,p})} \rho_k \tilde{\mathbf{g}}_k \vartriangle \mathbf{\Phi}\hat{\mathbf{G}}^H \left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1} \mathbf{D}_\eta \mathbf{u}\right|^2\right] \\
&= P_f \rho_k^2 \mathbb{E}\left[\left|\tilde{\mathbf{g}}_k \vartriangle \mathbf{\Phi}\hat{\mathbf{G}}^H \left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1} \mathbf{D}_\eta \mathbf{u}\right|^2\right] \\
&= P_f \rho_k^2 e^{-n_{\Delta\tau}\sigma_\phi^2} tr\left\{\mathbf{D}_\eta^2 \mathbb{E}\left[\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1} \hat{\mathbf{G}} \mathbb{E}\left[\tilde{\mathbf{g}}_k^H \tilde{\mathbf{g}}_k\right] \hat{\mathbf{G}}^H \left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\right]\right\} \\
&= P_f \rho_k^2 e^{-n_{\Delta\tau}\sigma_\phi^2} \sum_{i=1}^{K} \eta_i \chi_{ki},
\end{aligned} \quad (178)$$

gdzie χki oznacza i-ty element przekątnej

$$\mathbb{E}\left[\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1} \hat{\mathbf{G}} \mathbb{E}\left[\tilde{\mathbf{g}}_k^H \tilde{\mathbf{g}}_k\right] \hat{\mathbf{G}}^H \left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\right] \quad (179)$$

Podobnie wariancja I_2 jest podana przez

$$\begin{aligned}
\mathbb{E}\left[|\mathcal{I}_2|^2\right] &= \mathbb{E}\left[\left|\sqrt{P_f(1-\rho_k^2)} e^{j(\varphi_{k,d}-\varphi_{k,p})} \mathbf{e}_k \vartriangle \mathbf{\Phi}\hat{\mathbf{G}}^H \left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1} \mathbf{D}_\eta \mathbf{u}\right|^2\right] \\
&= P_f \left(1-\rho_k^2\right) \mathbb{E}\left[\left|\mathbf{e}_k \vartriangle \mathbf{\Phi}\hat{\mathbf{G}}^H \left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1} \mathbf{D}_\eta \mathbf{u}\right|^2\right] \\
&= P_f \left(1-\rho_k^2\right) e^{-n_{\Delta\tau}\sigma_\phi^2} tr\left\{\mathbf{D}_\eta^2 \mathbb{E}\left[\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1} \hat{\mathbf{G}} \mathbf{E}_k \hat{\mathbf{G}}^H \left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\right]\right\} \\
&= P_f \left(1-\rho_k^2\right) e^{-n_{\Delta\tau}\sigma_\phi^2} \sum_{i=1}^{K} \eta_i \xi_{ki}
\end{aligned} \quad (180)$$

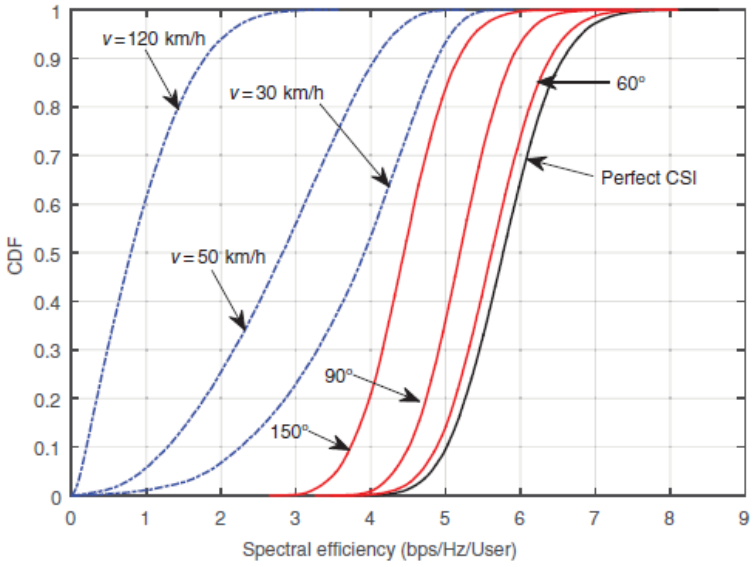
gdzie $E_k = \mathbb{E}[e^H_k e_k] = \text{diag}\{\beta_{1k}, \beta 2k, \ldots, \beta_{Mk}\} \in$ CM×M, a ξ_{ki} reprezentuje i-ty element przekątnej $\mathbb{E}\left[\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1} \hat{\mathbf{G}} \mathbf{E}_k \hat{\mathbf{G}}^H \left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\right]$. Podstawiając równania (178) i (180) do równania (176), otrzymujemy

$$\gamma_k = \frac{\rho_k^2 \eta_k}{\rho_k^2 \sum_{i=1}^{K} \eta_i \chi_{ki} + \left(1-\rho_k^2\right) \sum_{i=1}^{K} \eta_i \xi_{ki} + \frac{\sigma_n^2}{P_f e^{-n_{\Delta\tau}\sigma_\phi^2}}} \quad (181)$$

Biorąc pod uwagę opóźnienie propagacji przez interfejs radiowy $n_{ai} = \left\lceil \frac{\tau_{ul}+\tau_{dl}}{T_s} \right\rceil$ i opóźnienie $n_{\Delta\tau}$, osiągalna wydajność widmowa k-tego użytkownika jest podana wzorem

$$R_k = \left(1 - \frac{n_{ai} + n_{\Delta\tau}}{T_b}\right) \log_2 (1 + \gamma_k) \quad (182)$$

Rysunek przedstawia porównanie funkcji rozkładu kumulacyjnego (CDF) wydajności widmowej (SE) na użytkownika poprzez zmianę prędkości v lub akumulacyjnego szumu fazowego TPN.



Krzywa wydajności kodowania wstępnego z zerowym wymuszaniem (ZFP) przy użyciu idealnego CSI jest stosowana jako punkt odniesienia, gdzie UE są nieruchome (v = 0 km⁄h), a transceivery mają idealne oscylatory lokalne (TPN = 0◦). Aby zaobserwować wpływ mobilności użytkownika, najpierw ustawiamy TPN = 0◦ i wybieramy trzy typowe wartości: v = 30, 50 i 120 km⁄h. Bez utraty ogólności całkowite opóźnienie jest po prostu ustawione na $\Delta\tau$ = 1 ms, ponieważ efekt starzenia się mobilności użytkownika jest ustalany przez kombinację prędkości i opóźnienia. Nawet przy niskiej mobilności v = 30 km⁄h, co odpowiada bardzo wysokiej korelacji ρ = 0,97, pogorszenie wydajności jest już znaczące. Mówiąc konkretnie, 5%-owy prawdopodobny SE na użytkownika zmniejsza się do 1,8 bps⁄Hz, w porównaniu z 4,8 bps⁄Hz punktu odniesienia, co stanowi stratę 62,5%. 50%-owy prawdopodobny (mediana) SE na użytkownika pogarsza się o 32%, spadając z 5,7 do 3,9 bps⁄Hz. Wraz ze wzrostem v, utrata wydajności staje się bardziej znacząca. Przy wysokiej mobilności v = 120 km⁄h, 5%-owy prawdopodobny i mediana SE dalej zmniejszają się do 0,13 i 0,79 bps⁄Hz, co stanowi bardzo wysoką stratę odpowiednio 97% i 86%. Ponadto badano wpływ szumu fazowego, używając wybranego szumu fazowego T2 PN = 60◦, 90◦ i 150◦, gdzie UE są ustawione tak, aby były nieruchome v = 0 km⁄h. Przy małym szumie fazowym 60◦, jak pokazano na rysunku, utrata wydajności jest marginalna. Zwiększona do 150◦, 5% prawdopodobieństwa i mediana SE degradują się do 3,5 i 4,5 bps⁄Hz, co odpowiada utracie odpowiednio 27% i 23%.

Komunikacja oportunistyczna bezkomórkowa

Wykorzystując stopień swobody w domenie częstotliwości umożliwiony przez transmisję OFDM w systemach szerokopasmowych i efekt blisko-daleko między różnymi AP, system Massive MIMO bezkomórkowy może wdrożyć komunikację oportunistyczną w celu poprawy swojej efektywności energetycznej i wydajności widmowej. Kluczowym pomysłem jest przypisanie ortogonalnych zasobów domeny częstotliwości różnym użytkownikom, tak aby każda podnośna lub blok zasobów (RB) przenosił tylko jednego użytkownika. Taka konfiguracja nie tylko zapobiega interferencji wielu użytkowników, ale także upraszcza projekt systemu. Następnie szereg punktów dostępowych z silnym

zanikaniem na dużą skalę (zdefiniowanych jako bliskie AP) jest oportunistycznie wybieranych do obsługi tego użytkownika. Jednocześnie dalekie AP ze słabym zanikaniem na dużą skalę są dezaktywowane na przypisanych podnośnych lub RB dla tego użytkownika. Jako efekt uboczny liczba aktywnych AP na podnośną staje się mała, co umożliwia wykorzystanie pilotów łącza wstecznego, dzięki czemu użytkownik może wykonywać spójne wykrywanie. Główne korzyści techniczne schematu Opportunistic AP Selection (OAS) są dwojakie:

● Zysk oportunistyczny: Z punktu widzenia typowego użytkownika bliski AP ma korzystny kanał z małą utratą ścieżki. Natomiast energia emitowana z dalekiego AP jest marnowana na długim dystansie propagacji. Innymi słowy, ta sama ilość mocy przesyłana z bliskiego AP generuje o wiele silniejszą moc odbieraną niż daleki AP, co skutkuje wysoką mocą i wydajnością widmową.

● Spójny zysk: Z perspektywy każdej podnośnej lub bloku zasobów, tylko kilka AP obsługuje pojedynczego użytkownika. Następnie wielowymiarowy system Massive MIMO jest przekształcany w niskowymiarowy system Multiple-Input Single-Output (MISO). W rezultacie można złagodzić zaporowy narzut związany z wstawianiem pilotów łącza wstecznego, który jest proporcjonalny do ogromnej liczby anten stacji bazowych. Użytkownik może uzyskać natychmiastowy CSI, szacując piloty łącza wstecznego, zamiast znać tylko statystyczny CSI, a następnie wykonać spójne wykrywanie. Stąd podstawowy problem ograniczający wydajność łącza wstecznego Massive MIMO można rozwiązać dzięki oportunistycznemu wyborowi AP.

Bezkomórkowe masywne systemy szerokopasmowe

Rozważ bezkomórkowy masywny system MIMO, w którym M losowo rozproszonych punktów dostępowych podłączonych do procesora obsługuje K użytkowników na danym obszarze geograficznym. Nie tracąc ogólności, załóż, że każdy punkt dostępowy i UE jest wyposażony w pojedynczą antenę w celu prostej analizy. Rozważ selektywne zanikanie częstotliwości w systemach szerokopasmowych, w których kanał między punktem dostępowym m a użytkownikiem k można modelować jako liniowy filtr zmienny w czasie w równoważnej bazie pasma podstawowego, tj.

$$\mathbf{h}_{mk}[t] = [h_{mk,0}[t], h_{mk,1}[t], \ldots, h_{mk,L_{mk}-1}[t]]^T, \quad (183)$$

gdzie długość filtra L_{mk} zależy od rozproszenia opóźnienia i interwału próbkowania. Biorąc pod uwagę zanikanie na dużą skalę β_{mk}, filtr kanałowy między AP m i użytkownikiem k można modelować za pomocą

$$\begin{aligned}\mathbf{g}_{mk}[t] &= [g_{mk,0}[t], g_{mk,1}[t], \ldots, g_{mk,L_{mk}-1}[t]]^T \\ &= \sqrt{\beta_{mk}}\mathbf{h}_{mk}[t],\end{aligned} \quad (184)$$

z $g_{mk,l}[t] = \sqrt{\beta_{mk}} h_{mk,l}[t]$, $\forall l = 0, 1, \ldots, L_{mk} - 1$. Obserwuje się, że struktura bezkomórkowa powoduje efekt bliski-daleki pomiędzy różnymi AP z perspektywy typowego użytkownika. AP można zatem podzielić na dwie kategorie: bliskie AP i dalekie AP, podobnie jak bliscy i dalecy użytkownicy z perspektywy stacji bazowej w konwencjonalnych systemach komórkowych. Transmisja sygnału w systemie OFDM jest zorganizowana blokowo. Oznacz blok symboli domeny częstotliwości AP m na t-tym symbolu OFDM przez

$$\tilde{\mathbf{x}}_m[t] = \left[\tilde{x}_{m,0}[t], \tilde{x}_{m,1}[t], \ldots, \tilde{x}_{m,N-1}[t]\right]^T \quad (185)$$

Wykonując N-punktową odwrotną dyskretną transformację Fouriera (IDFT), $\tilde{\mathbf{x}}_m[t]$ jest konwertowany na sekwencję w dziedzinie czasu

$$\mathbf{x}_m[t] = \left[x_{m,0}[t], x_{m,1}[t], \ldots, x_{m,N-1}[t]\right]^{\mathrm{T}} \quad (186)$$

pod względem

$$x_{m,n'}[t] = \frac{1}{N}\sum_{n=0}^{N-1} \tilde{x}_{m,n}[t] e^{\frac{2\pi j n' n}{N}}, \quad \forall n' \quad (187)$$

Definiowanie macierzy dyskretnej transformacji Fouriera (DFT)

$$\mathbf{D} = \begin{bmatrix} \Omega_N^{00} & \cdots & \Omega_N^{0(N-1)} \\ \vdots & \ddots & \vdots \\ \Omega_N^{(N-1)0} & \cdots & \Omega_N^{(N-1)(N-1)} \end{bmatrix} \quad (188)$$

przy pierwotnym pierwiastku N-tego stopnia z jedności $\Omega^{nn'}{}_N = e^{-2\pi j n'n/N}$ modulacja OFDM jest wyrażona w postaci macierzowej jako

$$\mathbf{x}_m[t] = \mathbf{D}^{-1}\tilde{\mathbf{x}}_m[t] = \frac{1}{N}\mathbf{D}^{*}\tilde{\mathbf{x}}_m[t]. \quad (189)$$

Prefiks cykliczny (CP) jest wstawiany pomiędzy dwa bloki transmisyjne w celu zachowania ortogonalności podnośnych i absorbowania interferencji międzysymbolowej. Przekazywany sygnał pasma podstawowego z CP jest oznaczany jako $x^{cp}{}_m[t]$. Przechodząc przez kanał bezprzewodowy, skutkuje on odebranym składnikiem sygnału $x^{cp}{}_m[t] * g_{mk}[t]$ u typowego użytkownika k, gdzie $*$ oznacza splot liniowy. W konsekwencji odebrany sygnał u użytkownika k jest podany przez

$$\mathbf{y}_k^{cp}[t] = \sum_{m=1}^{M} \mathbf{g}_{mk}[t] * \mathbf{x}_m^{cp}[t] + \mathbf{z}_k[t] \quad (190)$$

gdzie $z_k[t]$ oznacza wektor addytywnego białego szumu gaussowskiego o zerowej średniej i wariancji $\sigma^2{}_z$, tj. $\mathbf{z}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\sigma}_z^2\mathbf{I})$. Usuwając CP, otrzymujemy

$$\mathbf{y}_k[t] = \sum_{m=1}^{M} \mathbf{g}_{mk}^{N}[t] \otimes \mathbf{x}_m[t] + \mathbf{z}_k[t], \quad (191)$$

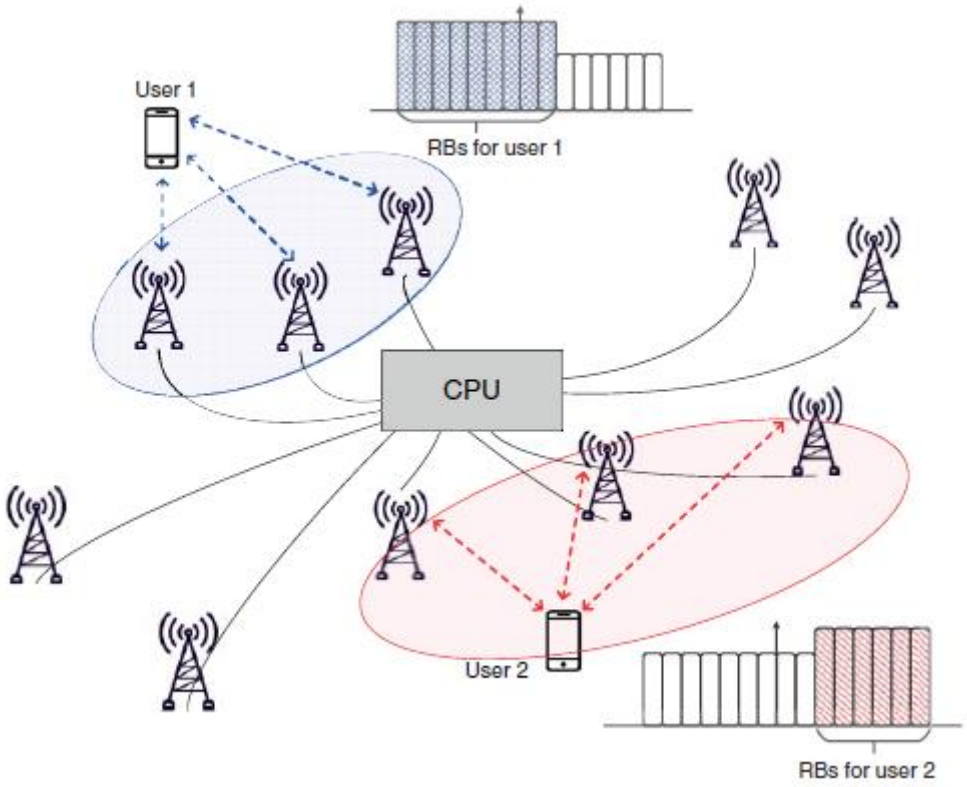
gdzie $\otimes$ oznacza splot cykliczny, a $g^N{}_{mk}[t]$ jest wektorem $g_{mk}[t]$ o długości N wypełnionym zerami. Następnie odebrany sygnał w dziedzinie częstotliwości jest obliczany przez

$$\tilde{\mathbf{y}}_k[t] = \mathbf{D}\mathbf{y}_k[t]. \quad (192)$$

Podstawiając równanie (191) do równania (192) i stosując twierdzenie splotu, otrzymujemy

$$\tilde{\mathbf{y}}_k[t] = \sum_{m=1}^{M} \mathbf{D}\left(\mathbf{g}_{mk}^{N}[t] \otimes \mathbf{x}_m[t]\right) + \mathbf{D}\mathbf{z}_k[t]$$
$$= \sum_{m=1}^{M} \tilde{\mathbf{g}}_{mk}[t] \odot \tilde{\mathbf{x}}_m[t] + \tilde{\mathbf{z}}_k[t], \quad (193)$$

gdzie ⊙ oznacza iloczyn Hadamarda .



Dla typowej podnośnej n model sygnału łącza wstecznego jest wyrażony przez

$$\tilde{y}_{k,n}[t] = \sum_{m=1}^{M} \tilde{g}_{mk,n}[t]\tilde{x}_{m,n}[t] + \tilde{z}_{k,n}[t], \quad k \in \{1, \ldots, K\} \quad (194)$$

Wybór AP oportunistyczny

Transmisja downlink od AP do użytkowników i transmisja uplink od użytkowników do AP są rozdzielone multipleksowaniem z podziałem czasu (TDD) przy założeniu doskonałej wzajemności kanału. Długość ramki radiowej jest na ogół mniejsza niż czas koherencji kanału, dlatego stan kanału jest uważany za stały w ramce. Nie tracąc ogólności, indeksowanie czasowe sygnałów jest ignorowane w przypadku prostej analizy. Proces komunikacji schematu OAS w bezkomórkowym systemie masywnego

MIMO-OFDM jest przedstawiony następująco:

● AP m, ∀m mierzy zanikanie na dużą skalę β_{mk}, k = 1, 2,…, K, w długoterminowej perspektywie i okresowo raportuje te informacje do procesora. W ten sposób procesor ma globalną wiedzę o CSI na dużą skalę B ∈ $\mathbb{C}^{M\times K}$, gdzie $[B]_{m,k} = \beta_{mk}$, gdzie $[\cdot]_{m,k}$ oznacza (m, k)-ty wpis macierzy. Ponieważ β_{mk} jest niezależne od częstotliwości i zmienia się powoli, ten pomiar jest praktycznie łatwy do wdrożenia.

● UE k, ∀k okresowo raportuje swoje żądanie szybkości transmisji danych za pomocą skalara $r_{q,k}$ poprzez sygnalizację łącza w górę. Następnie procesor wie, że $r_q = \{r_{q,1}, r_{q,2},\ldots, r_{q,K}\}$.

● Przydział zasobów w dziedzinie częstotliwości: procesor podejmuje decyzję o przydziale zasobów jako funkcję żądań użytkowników, tj. $\{\mathbb{B}_1,\ldots,\mathbb{B}_K\} = f(r_q)$, gdzie konkretna implementacja f(·) opiera się na pewnych szczególnych kryteriach, np. uczciwości, priorytecie i wydajności. Pula zasobów składa się z N podnośnych OFDM, oznaczonych przez zestaw indeksów podnośnych $\mathbb{B} = \{0, 1, 2\ldots,N - 1\}$. Używając $\mathbb{B}_k$ do oznaczenia indeksów podnośnych przypisanych do użytkownika k, mamy $\bigcup^{K}_{k=1} \mathbb{B}_k \in \mathbb{B}$ (gdy wszystkie podnośne są przydzielone, $\bigcup_{Kk} =1 \mathbb{B}_k = \mathbb{B}$). Podnośne są przydzielane ortogonalnie, spełniając $\mathbb{B}_k \cap \mathbb{B}_{k'} = \emptyset$, ∀k′ ≠ k. Przedział czasowy przydziału zasobów zależy od projektu systemu.

● Wybór oportunistyczny: CPU wybiera oportunistyczne AP dla każdego użytkownika pod kątem zaniku na dużą skalę. Załóżmy, że liczba wybranych AP wynosi Ms, gdzie $1 \leqslant M_s \leqslant M$. Uporządkuj indeksy AP w kategoriach ich zaniku na dużą skalę w kolejności malejącej, a następnie wybierz pierwsze Ms AP. Zbiór oportunistycznych AP dla użytkownika k oznaczamy przez $\mathbb{M}k = \{\dot{m}^k_1,\ldots,\ddot{m}^k_{Ms}\}$. Jeśli $M_s = M$, wszystkie AP uczestniczą w transmisji, bez żadnego wyboru. Jeśli $M_s = 1$, określany jest tylko jeden AP z największym zanikiem na dużą skalę, tj.

$$\ddot{m}^k_1 = \arg \max_{m=1,\ldots,M} (\mathbf{b}_k), \quad (195)$$

gdzie b_k oznacza k-ty wiersz B.

● Transmisja łącza w górę: Użytkownik k, ∀k przesyła swoje dane i określoną sekwencję pilota przez przypisane mu podnośne $\mathbb{B}_k$. Oportunistyczne AP w $\mathbb{M}_k$ szacują CSI łącza w górę $\hat{g}_{mk,n}$, gdzie m ∈ $\mathbb{M}_k$ i n ∈ $\mathbb{B}_k$. AP wykrywają dane łącza w górę spójnie ze znajomością CSI łącza w górę.

● Sprzężone formowanie wiązki: Następnie AP m, ∀m ∈ $\mathbb{M}_k$ zna CSI łącza w dół $g_{mk,n}$ zgodnie z wzajemnością kanału. Następnie przesyła zmodulowany symbol $s_{k,n}$, n ∈ $\mathbb{B}_k$ z $\mathbb{E}[|s_{k,n}|^2] = 1$ i sekwencje pilota łącza w dół przez przypisane podnośne $\mathbb{B}_k$. Stosując sprzężone formowanie wiązki w dziedzinie częstotliwości, transmitowany symbol w m-tym punkcie dostępowym jest

$$\tilde{x}_{m,n} = \sqrt{\eta_{mk} P_d} \hat{g}^*_{mk,n} s_{k,n}, \quad (196)$$

gdzie $\sqrt{\eta_{mk}}$, $0 \leq \eta_{mk} \leq 1$ oznacza współczynnik kontroli mocy, a Pd jest ujednoliconym ograniczeniem mocy każdego AP.

● Koherentne wykrywanie: Użytkownik k szacuje CSI łącza w dół $\hat{g}_{mk,n}$, gdzie m ∈ $\mathbb{M}_k$ i n ∈ $\mathbb{B}_k$, i wykrywa dane łącza w dół koherentnie za pomocą $\hat{g}_{mk,n}$.

Analiza wydajności widmowej

Badanie wydajności trzech różnych schematów ma na celu rzucenie światła na korzyści z oportunistycznego wyboru i oszacowania kanału downlink. Najpierw analizuje się wydajność w kategoriach SE dla konwencjonalnego systemu CFmMIMO-OFDM bez oportunistycznego wyboru AP, oznaczonego jako Full AP, jako punkt odniesienia do porównania. Po drugie, wyprowadza się również wydajność systemu, który wybiera Ms oportunistycznych AP spośród M AP, ale nie wstawia pilotów downlink. Na koniec analizuje się SE oportunistycznego wyboru AP z wstawieniem pilotów downlink, oznaczonego jako OAS-DP. Jako punkt odniesienia konwencjonalne sprzężone formowanie wiązki w CFmMIMO jest skierowane do każdej podnośnej w CFmMIMO-OFDM. Aby to zrobić, każdy AP

multipleksuje łącznie K symboli, tj. $s_{k,n}$ przeznaczonych dla użytkownika k, k = 1,…, K, przed transmisją. Przy współczynniku sterowania mocą $\sqrt{\eta_{mk}}$, $0 \le \eta_{mk} \le 1$, przesyłany sygnał m-tego AP na podnośnej n wynosi

$$\bar{x}_{m,n} = \sqrt{P_d} \sum_{k=1}^{K} \sqrt{\eta_{mk}} \hat{g}^*_{mk,n} s_{k,n} \tag{197}$$

gdzie $\hat{g}_{mk,n}$ oznacza oszacowani$\hat{g}_{mk,n}$, a $\hat{g}_{mk,n} = \bar{g}_{mk,n} - \xi_{mk,n}$ z błędem oszacowania $\xi_{mk,n}$ podniesionym przez szum addytywny. Stosując oszacowanie MMSE otrzymujemy $\hat{g}_{mk,n} \in \mathcal{CN}(0, \alpha_{mk})$ z $\alpha_{mk} = \frac{P_u \beta^2_{mk}}{P_u \beta_{mk} + \sigma^2_z}$, gdzie P_u jest ograniczeniem mocy łącza w górę, w porównaniu z $\bar{g}_{mk,n} \in \mathcal{CN}(0, \beta_{mk})$. W konwencjonalnym CFmMIMO nie ma pilota łącza wstecznego i oszacowania kanału ze względu na zaporowy narzut wstawiania pilotów na ogromnej liczbie anten. W konsekwencji zakłada się, że każdy użytkownik ma tylko wiedzę o statystykach kanału $\mathbb{E}\left[\left|\hat{g}_{mk,n}\right|^2\right] = \alpha_{mk}$, czyli hartowaniu kanału, a nie o realizacji kanału $\hat{g}_{mk,n}$ Podstawienie równania (197) do równania (194) daje odebrany sygnał dla użytkownika k:

$$\begin{aligned}
\bar{y}_{k,n} &= \sqrt{P_d} \sum_{m=1}^{M} \bar{g}_{mk,n} \sum_{k'=1}^{K} \sqrt{\eta_{mk'}} \hat{g}^*_{mk',n} s_{k',n} + \bar{z}_{k,n} \\
&= \underbrace{\sqrt{P_d} \sum_{m=1}^{M} \sqrt{\eta_{mk}} \mathbb{E}\left[\left|\hat{g}_{mk,n}\right|^2\right] s_{k,n}}_{\text{Desired signal}} + \underbrace{\sqrt{P_d} \sum_{m=1}^{M} \hat{g}_{mk,n} \sum_{k' \ne k}^{K} \sqrt{\eta_{mk'}} \hat{g}^*_{mk',n} s_{k',n}}_{\text{Inter-user interference}} \\
&\quad + \underbrace{\sqrt{P_d} \sum_{m=1}^{M} \sqrt{\eta_{mk}} \left(\left|\hat{g}_{mk,n}\right|^2 - \mathbb{E}\left[\left|\hat{g}_{mk,n}\right|^2\right]\right) s_{k,n}}_{\text{Error due to CSI statistics}} \\
&\quad + \underbrace{\sqrt{P_d} \sum_{m=1}^{M} \xi_{mk,n} \sum_{k'=1}^{K} \sqrt{\eta_{mk'}} \hat{g}^*_{mk',n} s_{k',n}}_{\text{Channel-estimation error}} + \underbrace{\bar{z}_{k,n}}_{\text{Noise}}.
\end{aligned}$$

Wydajność widmowa użytkownika k na podnośnej n, ∀n = 0, 1,…,N − 1 jest ograniczona dolną granicą $\log_2(1 + \gamma^{(n)}_k)$ z

$$\gamma_k^{(n)} = \frac{\left(\sum_{m=1}^{M} \sqrt{\eta_{mk}} \alpha_{mk}\right)^2}{\sum_{m=1}^{M} \beta_{mk} \sum_{k'=1}^{K} \eta_{mk'} \alpha_{mk'} + \frac{1}{\gamma_t}}, \tag{198}$$

ze stosunkiem sygnału do szumu transmisji $\gamma_t = P_d/\sigma^2_z$. Schemat OAS wykorzystuje stopień swobody włączony przez dziedzinę częstotliwości, aby przypisać różnych użytkowników do zasobów ortogonalnych. W rezultacie interferencja między użytkownikami znika, ponieważ każda podnośna OFDM obsługuje pojedynczego użytkownika. Dlatego podstawienie K = 1 do równania (198) daje wydajność pierwszego schematu z transmisją Full AP, tj.

$$\gamma_k^{\langle n \rangle} = \frac{\left(\sum_{m=1}^{M} \sqrt{\eta_{mk}}\alpha_{mk}\right)^2}{\sum_{m=1}^{M} \beta_{mk}\eta_{mk}\alpha_{mk} + \frac{1}{\gamma_t}}. \quad (199)$$

Aby rzucić światło na wpływ selekcji oportunistycznej, zbadano wydajność wybierania punktów dostępowych M_s bez dodawania pilotów downlink. Oznacza to, że każdy użytkownik ma wiedzę jedynie o statystykach kanału, a nie o realizacji kanału. Podstawiając równanie (196) do równania (194) w celu uzyskania sygnału odebranego u użytkownika k na podnośnej $n \in \mathbb{B}_k$ jako

$$\begin{aligned}
\tilde{y}_{k,n} &= \sqrt{P_d} \sum_{m \in \mathbb{M}_k} \bar{g}_{mk,n} \sqrt{\eta_{mk}} \hat{g}^*_{mk,n} s_{k,n} + \bar{z}_{k,n} \\
&= \underbrace{\sqrt{P_d} \sum_{m \in \mathbb{M}_k} \sqrt{\eta_{mk}} \mathbb{E}\left[\left|\hat{g}_{mk,n}\right|^2\right] s_{k,n}}_{\text{Desired signal}} \\
&+ \underbrace{\sqrt{P_d} \sum_{m \in \mathbb{M}_k} \sqrt{\eta_{mk}} \left(\left|\hat{g}_{mk,n}\right|^2 - \mathbb{E}\left[\left|\hat{g}_{mk,n}\right|^2\right]\right) s_{k,n}}_{\text{Error due to CSI statistics}} \\
&+ \underbrace{\sqrt{P_d} \sum_{m \in \mathbb{M}_k} \xi_{mk,n} \sqrt{\eta_{mk}} \hat{g}^*_{mk,n} s_{k,n}}_{\text{Channel-estimation error}} + \underbrace{\bar{z}_{k,n}}_{\text{Noise}}.
\end{aligned} \quad (200)$$

Należy zauważyć, że odebrany sygnał użytkownika k na podnośnej $n \in \{\mathbb{B} - \mathbb{B}_k\}$ wynosi $\tilde{y}_{k,n} = 0$. Podobnie, wydajność widmowa użytkownika k na podnośnej $n \in \mathbb{B}_k$ jest ograniczona dolną granicą $\log_2(1 + \gamma^{\langle n \rangle}_k)$ przez

$$\gamma_k^{\langle n \rangle} = \frac{\left(\sum_{m \in \mathbb{M}_k} \sqrt{\eta_{mk}}\alpha_{mk}\right)^2}{\sum_{m \in \mathbb{M}_k} \eta_{mk}\beta_{mk}\alpha_{mk} + \frac{1}{\gamma_t}}. \quad (201)$$

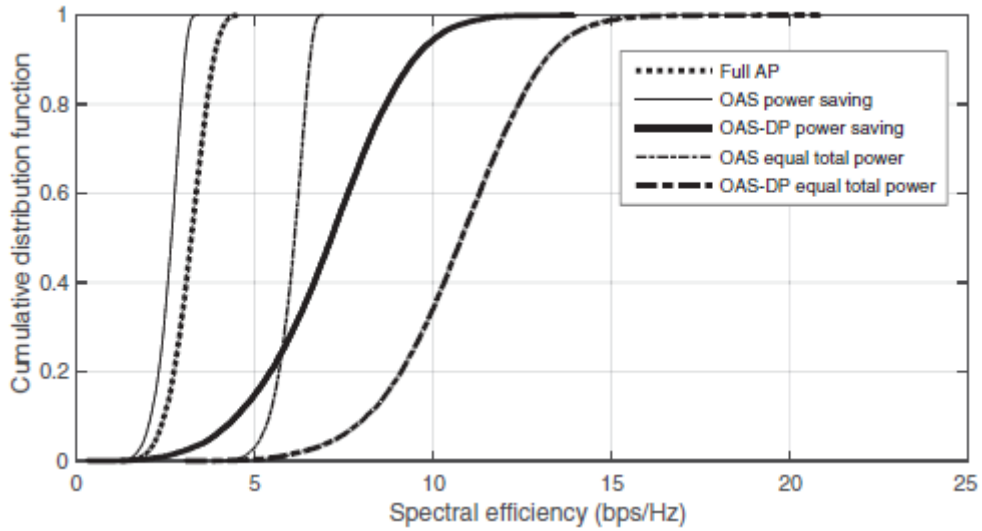
Dzięki oportunistycznemu wyborowi AP, w proponowanym schemacie istnieje tylko kilka aktywnych AP na każdej podnośnej, podczas gdy inne dalekie AP są wyłączone. Z perspektywy typowej podnośnej jest to niskowymiarowy system MISO, w którym narzut wstawiania pilotów łącza wstecznego jest akceptowalny. W rezultacie użytkownik k otrzymuje szacowany CSI , a nie statystyki kanału α_{mk}. Tak więc odebrany sygnał u użytkownika k w równaniu (200) można przekształcić jako

$$\begin{aligned}
\tilde{y}_{k,n} &= \sqrt{P_d} \sum_{m \in \mathbb{M}_k} \bar{g}_{mk,n} \sqrt{\eta_{mk}} \hat{g}^*_{mk,n} s_{k,n} + \bar{z}_{k,n} \\
&= \underbrace{\sqrt{P_d} \sum_{m \in \mathbb{M}_k} \sqrt{\eta_{mk}} \left|\hat{g}_{mk,n}\right|^2 s_{k,n}}_{\text{Desired signal}} + \underbrace{\sqrt{P_d} \sum_{m \in \mathbb{M}_k} \xi_{mk,n} \sqrt{\eta_{mk}} \hat{g}^*_{mk,n} s_{k,n}}_{\text{Channel-Estimation error}} + \underbrace{\bar{z}_{k,n}}_{\text{Noise}}.
\end{aligned} \quad (202)$$

Stosując koherentną detekcję, wydajność widmowa użytkownika k na podnośnej $n \in \mathbb{B}_k$ jest wyrażona przez $\log_2(1 + \gamma^{\langle n \rangle}_k)$ z

$$\gamma_k^{\langle n \rangle} = \frac{\left(\sum_{m \in \mathbb{M}_k} \sqrt{\eta_{mk}} \left| \hat{g}_{mk,n} \right|^2 \right)^2}{\sum_{m \in \mathbb{M}_k} \eta_{mk} (\beta_{mk} - \alpha_{mk}) \alpha_{mk} + \frac{1}{\gamma_t}}. \quad (203)$$

Rysunek przedstawia CDF różnych schematów.



Po pierwsze, krzywa pełnego AP oznacza konwencjonalny system CFmMIMO-OFDM, w którym wszystkie M = 128 AP obsługuje przypisanego użytkownika na typowej podnośnej bez oportunistycznego wyboru AP. Osiągnięta 95% prawdopodobna wydajność widmowa wynosi około 2,2 bps/Hz, a 50% prawdopodobna lub mediana wydajności widmowej wynosi około 3,2 bps/Hz. Jeśli Ms = 10 aktywnych AP zostanie wybranych pod kątem zaniku na dużą skalę, a ograniczenie mocy każdego AP jest takie samo jak w pełnym AP, osiągnięty SE Oszczędzania energii OAS jest nieznacznie gorszy od pełnego AP. Ma 95% prawdopodobny SE około 1,9 bps/Hz i medianę SE około 2,7 bps/Hz. Jednakże, znacznie przewyższa pod względem efektywności energetycznej, ponieważ tylko Ms = 10 AP jest aktywnych, w porównaniu z M = 128 AP w pełnym AP, co daje oszczędność energii na poziomie 92,19%. Dzieje się tak, ponieważ moc dalekich AP nie może skutecznie przenieść się na moc odebraną z powodu poważnych strat propagacyjnych. Wyłączenie dalekich AP nie wpływa na całkowitą moc odebraną u użytkownika. Jako uczciwe porównanie załóżmy, że wybrane AP mają taką samą całkowitą moc jak pełny AP, tj. każdy oportunistyczny AP używa mocy M⁄Ms = 12,8 razy wyższej. Jak pokazuje CDF OAS Equal Total Power, 95% prawdopodobny SE znacznie wzrasta do 5,2 bps/Hz, a mediana SE osiąga 6,1 bps/Hz. Następnie możemy zaobserwować znaczący wzrost wydajności pilotów łącza wstecznego, które umożliwiają spójne wykrywanie u użytkownika. Nawet jeśli całkowita moc transmisji jest mniejsza niż 10% pełnego AP, OAS-DP Power Saving osiąga 95% prawdopodobny SE około 3,8 bps/Hz i medianę SE 7,2 bps/Hz. W porównaniu z pełnym AP, osiąga wzrost wydajności około 70% i 125% odpowiednio w 95% prawdopodobnym i medianie SE, jednocześnie osiągając 10-krotną wydajność energetyczną. Przy tej samej całkowitej mocy, wyższość oportunistycznego wyboru AP za pomocą pilota łącza wstecznego jest bardziej znacząca. W tym przypadku 95% prawdopodobny SE znacznie wzrasta do 7,4 bps/Hz, a mediana SE osiąga 10,8 bps/Hz. Krótko mówiąc, wyniki numeryczne potwierdzają wielką zaletę oportunistycznego wyboru AP, a także wzmocnionego CSI łącza wstecznego, w celu zwiększenia zarówno mocy, jak i wydajności widmowej w bezkomórkowym systemie massive MIMO.

Podsumowanie

W tej części najpierw przedstawiono kluczowe kwestie technik MIMO dla wielu użytkowników, w tym zasadę dobrze znanej metody osiągania przepustowości zwanej kodowaniem brudnego papieru. MU-MIMO ułatwia korzystanie z terminali o niskiej złożoności i niskich kosztach oraz jest mniej podatne na środowiska propagacyjne. Co najważniejsze, osiąga przepustowość sumaryczną wyższą niż przepustowość kanału SU-MIMO. Niemniej jednak konwencjonalne MU-MIMO nadal trudno skalować w celu multipleksowania przestrzennego wyższego rzędu. W tym rozdziale zbadano rewolucyjną technikę zwaną massive MIMO, która przełamuje tę barierę skalowalności, nie próbując osiągnąć pełnego limitu Shannona i paradoksalnie zwiększając rozmiar systemu. Na koniec przedstawiono rozproszony system massive MIMO zwany cell-free massive MIMO, w którym duża liczba anten usługowych jest losowo rozproszona na dużym obszarze. Konfiguracja bezkomórkowa jest szczególnie atrakcyjna w przypadku niektórych scenariuszy wdrażania 5G i nadchodzących 6G, takich jak kampus lub sieć prywatna dedykowana obiektowi przemysłowemu.