Sprzężenie zwrotne kanału dla mmWave massive MIMO

WPROWADZENIE

Powszechnie wiadomo, że potencjalne korzyści z masywnych tablic antenowych można w pełni wykorzystać, jeśli w nadajniku dostępna jest dokładna informacja o stanie kanału (CSI). W szczególności stacja bazowa z wielkoskalowym układem antenowym może przeprowadzać formowanie wiązki, które zapewnia bardzo duże wzmocnienie układu przy odpowiednim wstępnym kodowaniu łącza w dół . Jest to realizowane przez dopasowanie prekodera z chwilową odpowiedzią kanału, więc oczywiście optymalny prekoder jest uzyskiwany ze stanu kanału z wieloma wejściami i wieloma wyjściami (MIMO). W wielu badaniach założono, że dla bezprzewodowych systemów nadawczo-odbiorczych wyposażonych w duże układy antenowe charakter wzajemności kanałów w trybie dupleksowania z podziałem czasu (TDD) . W przypadku wzajemności kanałów można postulować, że kanał łącza w dół jest równoważny sprzężeniu hermitowskiemu kanału łącza w górę. W związku z tym w badaniach dotyczących masywnych systemów MIMO powszechnie zakłada się TDD ponieważ idealnie stacja bazowa może pozyskiwać CSI łącza nadawczego po prostu przez oszacowanie sygnałów pilota łącza w górę. Niemniej jednak istnieją pewne wady przejęcia CSI opartego na TDD. Na przykład, CSI mierzony przez samą stację bazową za pomocą sygnałów pilota łącza zwrotnego nie jest w stanie wychwycić potencjalnie istniejących zakłóceń łącza nadawczego (np. Zakłóceń międzykomórkowych). Ponadto, każde niedopasowanie między odpowiedziami obwodu nadajnika i odbiornika sprzętu użytkownika (UE) z powodu implementacji sprzętu może skutkować niedokładnym CSI, oraz takich sytuacjach mogą być potrzebne dodatkowe procesy kalibracji, aby ułatwić akwizycję CSI [3]. Co więcej, kwestia zanieczyszczenia pilota przy akwizycji kanału TDD została zidentyfikowana jako główna przeszkoda ograniczająca wydajność masywnych systemów MIMO. Obecnie dupleks z podziałem częstotliwości (FDD) pozostaje dominującym sposobem komunikacji komórkowej na całym świecie. Wygląda na to, że FDD jest bardziej odpowiedni dla systemów z symetrycznym ruchem między łączem w dół a łączem w górę. Ponadto FDD jest mniej wrażliwy na opóźnienia, ponieważ zasoby łącza w dół i łącza w górę są zawsze dostępne; Jest to ważna cecha aplikacji o znaczeniu krytycznym, które zostały podkreślone jako perspektywiczny trend 5G [5]. W przeciwieństwie do TDD, w FDD różne pasma częstotliwości są wykorzystywane do transmisji w dół i w górę, więc wzajemność kanałów nie może być wykorzystana do akwizycji CSI. Innymi słowy, stacja bazowa nie jest w stanie uzyskać szczegółowego CSI łącza w dół po prostu przez pomiar symboli sygnału pilota łącza zwrotnego transmitowanych z UE. Znajomość CSI po stronie nadajnika jest niezbędna w przypadku ogromnych schematów MIMO, takich jak formowanie wiązki w zamkniętej pętli, dlatego projekt mechanizmu akwizycji CSI jest szczególnie ważny w przypadku masywnych systemów komunikacyjnych MIMO działający w trybie FDD. W nowoczesnych systemach komórkowych, takich jak LTE-A, zastosowano dedykowany mechanizm sprzężenia zwrotnego, aby ułatwić wstępne kodowanie MIMO w łączu w dół. Dzięki mechanizmowi sprzężenia zwrotnego UE zgłasza zmierzone informacje o kanale łącza w dół do stacji bazowej przy użyciu dedykowanych zasobów łącza w górę. W porównaniu z pozyskiwaniem CSI opartym na wzajemności w TDD, CSI wyprowadzone i zgłoszone przez UE może być dokładniejsze, ponieważ jest w stanie uchwycić wpływ zakłóceń łącza nadawczego. Aby uzyskać sprzężenie zwrotne CSI w bardziej efektywny sposób, informacje o amplitudzie i fazie odpowiedzi kanału MIMO są kwantowane przy użyciu predefiniowanego zestawu wektorów / macierzy zwanych „książką kodową” w praktycznych systemach. W ten sposób sprzężenie zwrotne kanału można uzyskać, używając jedynie ograniczonej liczby bitów reprezentujących wskazanie jednego wpisu książki kodowej, którym może być na przykład skwantowana wersja odpowiedzi kanału zgodnej z podstawą lub wektory własne macierzy kanału MIMO. W specyfikacjach 3GPP LTE taki składnik CSI jest nazywany wskaźnikiem macierzy wstępnego kodowania (PMI). Opierając się na podejściu opartym na książce kodów z ograniczoną informacją zwrotną, zasoby radiowe wymagane do raportowania CSI są znacznie zmniejszone, a zatem bardziej

namacalne. W miarę wzrostu liczby anten nadawczych może zaistnieć potrzeba znacznego zwiększenia rozmiaru książki kodowej w celu dostosowania do drobnoziarnistych przestrzennych struktur kanałów. Jednakże, jeśli założymy, że zasób radiowy przeznaczony do raportowania CSI pozostaje niezmieniony, użycie większej książki kodowej (więcej bitów sprzężenia zwrotnego) może wpłynąć na niezawodność sprzężenia zwrotnego z powodu niższej szybkości kodowania. Taki dylemat zmotywował liczne wysiłki badawcze nad mechanizmami sprzężenia zwrotnego CSI dla masywnych systemów MIMO. Na przykład autorzy [6] zaproponowali nowatorską technikę projektowania książki kodowej dla masywnego MIMO, która umożliwia proces kodowania o niskiej złożoności, który jest bardziej skalowalny niż konwencjonalne podejście z tablicą przeglądową. W literaturze dostępnych jest również wiele innych rozwiązań alternatywnych. Aby dokładniej zbadać potencjalne schematy, które umożliwiają operacje wstępnego kodowania łącza w dół dla ogromnych stacji bazowych MIMO, w tym rozdziale przeanalizowano dwie alternatywne metody, które pozwalają na bardziej wydajne mechanizmy sprzężenia zwrotnego CSI dla masywnych systemów MIMO w trybie FDD, a także metodę wstępnego kodowania łącza w dół, która nie wymaga sprzężenia zwrotnego CSI od użytkowników. W pozostałej części tego rozdziału omówiono następujące schematy:

• Kompresja CSI oparta na wykrywaniu kompresji: wykorzystując rzadką reprezentację macierzy kanałowej MIMO z przestrzennie skorelowanymi macierzami anten, opracowujemy wykrywanie kompresji oparte na metodzie kompresji CSI, a także jej ulepszoną wersję z adaptacyjną podstawą rozrzedzającą.

• Sprzężenie zwrotne CSI oparte na wielostopniowym kształtowaniu wiązki: Zamiast pozwalać użytkownikom na mierzenie odpowiedzi kanałów odpowiadających każdej z anten nadajnika, badamy alternatywny schemat oparty na symbolach pilotów formowanych wiązką, co umożliwia wydajniejsze pomiary CSI i raportowanie.

• Wstępne kodowanie w łączu w dół bez wiedzy o CSI: Badamy technikę wstępnego kodowania w łączu w dół, która umożliwia stacji bazowej wykorzystanie jedynie wiedzy o kątach nadejścia (AoA) ścieżek transmisji, w których można odciążyć sprzężenie zwrotne CSI.

KANAŁOWE INFORMACJE ZWROTNE Z KOMPRESYWNYM WYKRYWANIEM

W ciągu ostatniej dekady nowatorskie narzędzie matematyczne zwane wykrywaniem kompresyjnym (znane również jako próbkowanie kompresyjne) było intensywnie badane i stosowane w różnych obszarach badań naukowych i inżynieryjnych, takich jak przetwarzanie obrazu i rozpoznawanie wzorców. Zgodnie z teorią wykrywania kompresyjnego sygnały mogą być rekonstruowane poprzez optymalizację przy użyciu znacznie mniejszej liczby próbek niż wymagane przez tradycyjne twierdzenie o próbkowaniu Shannona-Nyquista, o ile sam sygnał jest rzadki w określonej dziedzinie. W związku z tym technika ta jest szczególnie przydatna w przypadku problemów inżynierskich dotyczących sygnałów z potencjalnymi rzadkimi reprezentacjami. W przypadku dużych tablic antenowych, tak jak w mmWave massive MIMO, odstępy między antenami są zwykle niewielkie ze względu na ograniczenia związane ze współczynnikami kształtu sprzętu. Na przykład rozważ węzły infrastruktury, takie jak stacje bazowe małych komórek i zdalne głowice radiowe o ograniczonych rozmiarach fizycznych. W tych węzłach radiowych należy zainstalować bardzo zwarty układ anten, jeśli ma być obsługiwana potężna funkcja MIMO z wieloma antenami. W ten sposób można oczekiwać silnej korelacji przestrzennej między odpowiednimi wpisami macierzy kanału przestrzennego MIMO [8], co może skutkować rzadką reprezentacją w pewnej dziedzinie innej niż przestrzeń. Dlatego do kompresji odpowiedzi kanału można zastosować techniki wykrywania kompresyjnego, co skutecznie zmniejsza wymagany narzut sprzężenia zwrotnego w raportowaniu CSI. Przed zagłębieniem się w szczegóły metod sprzężenia

zwrotnego CSI opartych na wykrywaniu kompresji, najpierw przyjrzymy się nakrętkom i śrubom wykrywania kompresji i jego ogólnym algorytmom dla rzadkiej kompresji i odzyskiwania sygnału.

ALGORYTMICZNE RAMY WYKRYWANIA KOMPRESYJNEGO

Wykrywanie kompresyjne jest przydatne do odzyskiwania rzadkich sygnałów z niewielkiej liczby przypadkowych pomiarów oryginalnego sygnału, a także do wydajnej kompresji rzadkich sygnałów. Aby opisać algorytmiczne procedury wykrywania kompresyjnego, najpierw rozważymy wektor sygnału rzadkiego N x 1 oznaczony jako x, w którym tylko elementy K≪N x mają wartości niezerowe, podczas gdy wszystkie pozostałe wpisy N- K są zerowe. Kodowanie kompresyjne jest w stanie wychwycić te niezerowe składowe K poprzez kompresję sygnału do pomiarów M≪N poprzez losowe projekcje. Mówiąc konkretnie, rzadki sygnał N x 1 można skompresować do wektora Mx1 oznaczonego jako y przez

$$\boldsymbol{y} = \boldsymbol{\Phi x} \quad (1)$$

gdzie Φ jest macierzą M x N z niezależnymi i identycznie rozmieszczonymi (i.i.d.) losowymi wpisami, których wpisy można po prostu wygenerować zgodnie z rozkładami, takimi jak Gaussian lub Bernoulli. Ponieważ M≪N, (1) jest oczywiście niedookreślonym układem liniowym (jest mniej równań niż niewiadomych), który może mieć nieskończenie wiele rozwiązań, jeśli chce się rozwiązać x od y. Na szczęście koncepcja wykrywania kompresyjnego stwierdza, że tak długo, jak wektor sygnału x jest rzadki i zastosowano dostatecznie niespójne próbkowanie (do Φ), x można odzyskać z y z bardzo dużym prawdopodobieństwem za pomocą procedury optymalizacji. W wielu praktycznych przypadkach, docelowy sygnał N x 1 nie wykazuje żadnej rzadkości, co sprawia, że wykrywanie kompresyjne staje się niewykonalne. Niemniej jednak, jeśli korelacje między próbkami sygnału (elementami x) są wystarczająco wysokie, możemy być w stanie uzyskać rzadką reprezentację x, jeśli przejdzie pewną rzadką transformację:

$S = \Psi x$ (2)

gdzie S jest wektorem N x 1, który reprezentuje x w innej domenie, a Ψ jest podstawą transformacji N x N, która rozrzedza x. Kwintesencyjne przykłady Ψ obejmują dyskretną transformatę Fouriera (DFT) i dyskretną transformatę kosinusową (DCT), a te transformacje są zwykle ortogonalne, co oznacza, że odwrotna transformacja jest po prostu

$x=\Psi^T S$ (3)

W rzeczywistości rzadka reprezentacja była szeroko stosowana w konwencjonalnych schematach kompresji danych, w których informacje o położeniu i wartościach niezerowych elementów w S są kodowane w celu przywrócenia danych. Wykrywanie kompresyjne eliminuje konieczność takich narzutów, ponieważ S jest ślepo zakodowane w celu utworzenia y poprzez losową projekcję, jak w równaniu. (1):

$y=\Phi x=\Phi\Psi^T S$ (4)

Jak wspomniano wcześniej, rekonstrukcja x z y może być sformułowana jako proces optymalizacji. Aby być bardziej szczegółowym, dekoder powinien najpierw oszacować S z minimalizacją normy l_1:

$$\min \|\boldsymbol{S}\|_1, \text{ subject to } \boldsymbol{y} = \boldsymbol{\Phi\Psi}^T\boldsymbol{S} \quad (5)$$

Następnie x można odzyskać w oparciu o równanie. (3). Zwykle problem optymalizacji pokazany w równaniu. (5) można rozwiązać za pomocą dobrze znanych algorytmów, w tym programowania

liniowego, podstawowego śledzenia i ortogonalnego dopasowania. Szczegóły tych algorytmów optymalizacji zostały tutaj pominięte, ponieważ są dostępne w obszernej literaturze (np. Ref. [9]). Należy zauważyć, że S można dokładnie rozwiązać z dużym prawdopodobieństwem, o ile M jest wystarczająco duże, a $\Phi\Psi^T$ spełnia ograniczoną właściwość izometrii. Stąd wartość M powinna być starannie dobrana. Teoretycznie wielkość y powinna spełniać $M \geq cK \log (N/K)$, gdzie c jest małą stałą. Praktyczną zasadą jest ustawienie M jako dwa do czterech razy K.

ZASTOSOWANIE KOMPRESYJNEGO WYKRYWANIA W SYSTEMACH INFORMACJI ZWROTNYCH CSI

Rozważamy ogromny system komunikacji bezprzewodowej MIMO punkt-punkt z antenami nadajnika $N_t >>1$ w stacji bazowej i antenami odbiorczymi $N_r >> 1$ w UE. Model sygnału łącza nadawczego można zapisać jako

$r=Hd + n$ (6),

gdzie d i r są odpowiednio wektorem sygnału nadajnika N_t x 1 i wektorem sygnału odbiornika N_r x 1, a H jest macierzą kanału MIMO N_r x N_t z płaskimi wejściami. Wreszcie n jest wektorem N_r x 1 reprezentującym addytywny biały szum gaussowski. Zakładając, że H jest doskonale oszacowane w UE za pośrednictwem sygnałów pilota łącza w dół, stacja bazowa musi uzyskać co najmniej częściowe, jeśli nie pełne, informacje dotyczące szacowanego kanału status w celu adaptacyjnej konfiguracji parametrów warstwy fizycznej, takich jak wagi prekoderów MIMO. Na przykład koniugat hermitowski lub pseudoinwersja macierzy MIMO H może być użyty do konstrukcji prekodera masywnych systemów MIMO (a mianowicie formowania wiązki sprzężonej i formowania wiązki z wymuszeniem zerowym) . Alternatywnie, rozkład wartości osobliwych (SVD) można zastosować na H, aby wyodrębnić wektor formujący wiązkę, który odpowiada głównemu wektorowi osobliwemu H. W masywnych systemach MIMO sprzężenie zwrotne H może być żmudnym zadaniem w trybie FDD. Jak wspomniano powyżej, w konwencjonalnych systemach komunikacji komórkowej przyjmuje się podejście kwantyzacji wektor / macierz, tak że odpowiedź kanału H może być kwantyzowana przez predefiniowaną książkę kodową ze skończonymi wpisami, a sprzężenie zwrotne z UE może być realizowane po prostu przez zgłoszenie indeksu wpisów spisu kodów. Ze względu na dużą liczbę anten nadawczych w masywnym MIMO, rozmiar książki kodowej musi zostać rozszerzony, aby prawidłowo przechwycić H. Innymi słowy, należy zwiększyć ładunek sprzężenia zwrotnego. Jest to oczywiście niepożądane z punktu widzenia wykorzystania zasobów radiowych łącza w górę. Przedstawione tutaj techniki wykrywania kompresji można by zastosować do kompresji wymaganego ładunku sprzężenia zwrotnego poprzez zmniejszenie wymiaru H poprzez losową projekcję. Ze względu na wygodę implementacji, rzeczywiste i urojone części macierzy kanału H MIMO są przetwarzane oddzielnie we wszystkich proponowanych schematach ze wspólną operacją, a notacja $\widehat{\mathbf{H}}$ jest używana do reprezentowania części rzeczywistej lub urojonej H. ustawienia początkowe, sygnał docelowy $\widehat{\mathbf{H}}$ powinien być najpierw wektoryzowany do wektora (N_r x N_t) x 1:

$$\boldsymbol{h} = \mathrm{vec}\left(\widehat{\mathbf{H}}\right) \quad (7)$$

Należy zauważyć, że ten krok można pominąć w przypadku kanału z wieloma wejściami i jednym wyjściem, ponieważ kanał ma przede wszystkim postać wektorową. Aby skompresować h, losowa projekcja w równaniu (1) można zastosować, więc mamy

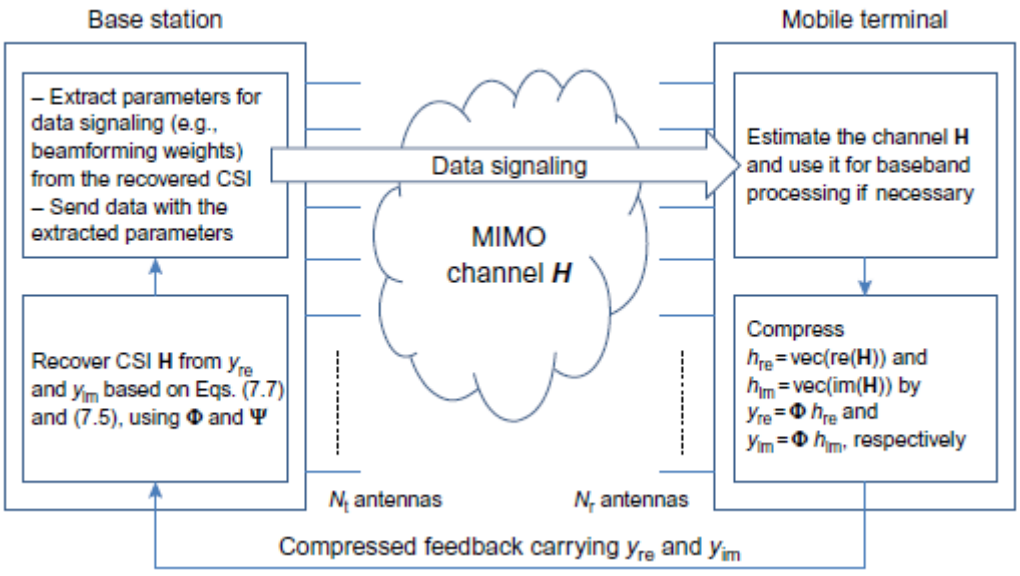
$y=\Phi h$ (8)

Ponieważ Φ jest macierzą M x (N_r x N_t) składającą się z niezależnych losowych wpisów, a wartość M jest mniejsza niż (N_r x N_t), kanał h jest kompresowany i kodowany jako wektor M x 1 y w równaniu. (8),

a obciążenie sprzężeniem zwrotnym jest w ten sposób zmniejszane o stopień sprężania $\eta = M / (N_r \times N_t)$. Ponieważ założyliśmy, że w mmWave massive MIMO, tablice antenowe zarówno w nadajniku, jak i odbiorniku są ściśle upakowane, wynikające z tego wysokie korelacje przestrzenne między elementami h mogą prowadzić do rzadkiej reprezentacji w określonej dziedzinie, jak w równaniu. (2). W związku z tym, w oparciu o teorię wykrywania kompresyjnego, informacje o kanale można dokładnie odzyskać z y, nawet jeśli wartość M jest ustawiona na znacznie mniejszą niż (N_r x N_t), więc można osiągnąć niski współczynnik kompresji, aby umożliwić skuteczny schemat sprzężenia zwrotnego CSI. Warto zauważyć, że macierz transformacji rozrzedzającej i elementy Φ można określić offline (na przykład ustaloną przez specyfikację systemu); dlatego zarówno stacja bazowa, jak i UE są świadome Ψ, jak również zawartość w oparciu o wstępne konfiguracje. W ten sposób stacja bazowa jest w stanie podjąć procedury optymalizacji w równaniu. (7.5) aby oszacować S (rzadka forma h) i ostatecznie odzyskać h przez zastosowanie odwrotnej transformacji rozrzedzającej:

$h=\Psi^T S$ (9)

Ogólny schemat sprzężenia zwrotnego kanału przedstawiono na rysunku



PODSTAWOWA SPARSYFIKACJA

Wybór ograniczającej podstawy Ψ odgrywa kluczową rolę w przeprowadzaniu odzyskiwania. Przy danym współczynniku kompresji pożądane jest wybranie rzadkiej podstawy, która zapewnia bardziej rzadką reprezentację (mniej elementów niezerowych w S lub mniejsze K) h, ponieważ prowadzi to do odzyskiwania sygnału z większą dokładnością. W tym rozdziale rozważymy dwie takie podstawy: dwuwymiarową dyskretną transformację kosinusową (2D-DCT) i transformację Karhunena-Loeve'a (KLT). W kontekście cyfrowego przetwarzania obrazu, 2D-DCT jest szeroko stosowany do kompresji obrazu, ponieważ jest przydatny w wykorzystywaniu korelacji przestrzennej między pikselami w celu zmniejszenia wymiarów obrazu. W przypadku mmWave massive MIMO macierz składająca się z dużej liczby anten może być bardzo kompaktowa ze względu na ograniczony współczynnik kształtu sprzętu, co prowadzi do małych odstępów między antenami. Opierając się na takim przypuszczeniu, oczekuje się, że elementy rzeczywistej lub urojonej macierzy kanałowej MIMO $\widehat{\mathbf{H}}$ będą silnie skorelowane w co najmniej jednym wymiarze z powodu małych odstępów między antenami po stronie nadajnika. Dlatego analogicznie do wykorzystywania korelacji przestrzennej między pikselami obrazu cyfrowego,

2D-DCT można wykorzystać jako podstawę do realizacji rzadkiej reprezentacji bH w dziedzinie częstotliwości przestrzennej poprzez pełne wykorzystanie takiej struktury korelacji. Operację macierzową stosowania 2D-DCT na $\widehat{\mathbf{H}}$ można zapisać jako $\boldsymbol{C}_{N_t}^{\mathrm{T}}\widehat{\mathbf{H}}\boldsymbol{C}_{N_r}$, gdzie C_L jest macierzą L x L DCT. Ponieważ macierz DCT jest podstawą do rozdzielenia 2D], rzadką reprezentację h można wyrazić jako

$$S=(\boldsymbol{C}_{N_t}\otimes\boldsymbol{C}_{N_r})^{\mathrm{T}}\mathrm{vec}\left(\widehat{\mathbf{H}}\right)=(\boldsymbol{C}_{N_t}\otimes\boldsymbol{C}_{N_r})^{\mathrm{T}}\boldsymbol{h} \quad (10)$$

gdzie ⊗ oznacza iloczyn Kronecker. Stąd jest oczywiste, że ograniczoną podstawą Ψ związaną z 2D-DCT jest

$$\boldsymbol{\Psi}_{\mathrm{2D-DCT}}=(\boldsymbol{C}_{N_t}\otimes\boldsymbol{C}_{N_r})^{\mathrm{T}} \quad (11)$$

Sparsyfikaja, jaką może osiągnąć 2D-DCT, jest silnie zależna od struktury korelacji h. W przeciwieństwie do tego KLT jest w stanie zapewnić optymalną rzadką reprezentację z tylko jednym niezerowym elementem (K=1) niezależnie od struktury korelacji , co obiecuje dokładne odtworzenie kanału, nawet jeśli dostępna jest tylko niewielka liczba pomiarów. Oznacza to, że można zagwarantować bardzo niski współczynnik kompresji. Niemniej jednak zastosowanie KLT jest raczej trudne w praktyce, ponieważ jego podstawa rozpraszająca jest zależna od sygnału i musi pochodzić z samego sygnału docelowego. Konkretnie, biorąc pod uwagę, że mamy macierz korelacji chwilowej (Nr x Nt) (Nr x Nt) h:

$W = hh^T$ (12)

podstawy sparsyfikacji KLT, Ψ_{KLT}, można obliczyć jako wektory własne W:

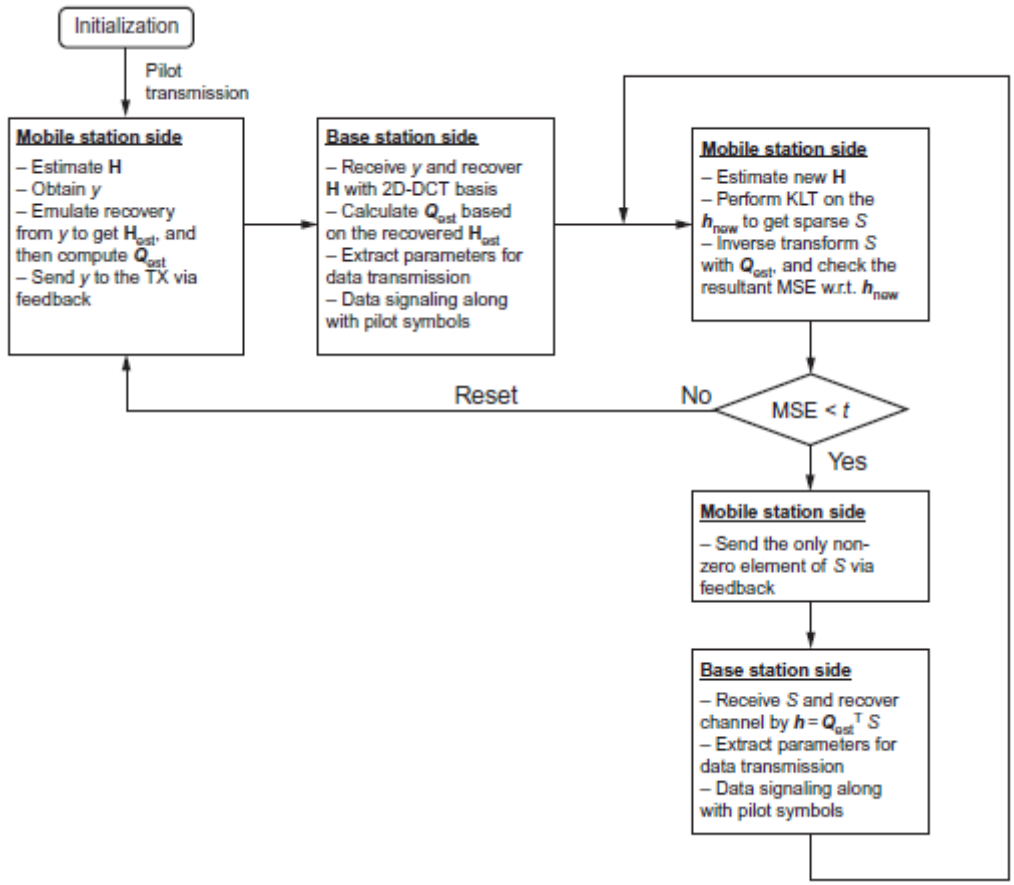
$W= \Psi_{KLT}\Lambda\Psi_{KLT}^T$ (13)

gdzie Λ jest macierzą diagonalną, której niezerowe wpisy są wartościami własnymi W. Tak więc, pomimo optymalnej rzadkości, jaką może osiągnąć KLT, zastosowanie KLT jest niepraktyczne, ponieważ jest to zasadniczo paradoks kurczaka i jaja, w którym podstawa pochodzi z sygnału docelowego, który chcemy odzyskać. Od tego momentu , podstawa sparsyfikacji o stałej wartości, taka jak 2D-DCT, jest bardziej wykonalna, jeśli mechanizm sprzężenia zwrotnego CSI oparty na wykrywaniu kompresyjnym zostanie zastosowany dla masywnych systemów MIMO. Jeśli jednak kanał jest względnie statyczny (na przykład w scenariuszach o niskiej mobilności), ważność Ψ_{KLT} uzyskanego wcześniej może zostać utrzymana przez kilka przedziałów czasu transmisji. W następnej sekcji opracujemy protokół sprzężenia zwrotnego CSI, który wykorzystuje 2D-DCT i KLT w sposób hybrydowy.

PROTOKÓŁ ZWROTNY CSI Z ADAPTACYJNĄ PODSTAWĄ SPARSYFIKACJI

Wiadomo, że optymalna rzadka reprezentacja macierzy kanału MIMO dostarczona przez KLT ma tylko jedną wartość niezerową (oznaczoną jako s), która jest przypuszczalnie pierwszym wpisem rzadkiego wektora (S) ze względu na właściwości struktury własnej. W ten sposób obciążenie sprzężeniem zwrotnym można znacznie zmniejszyć, ponieważ sama wartość s jest wystarczająca do przywrócenia CSI w stacji bazowej. Chociaż KLT jest zależne od sygnału, jak wspomniano wcześniej, może być przydatne, jeśli kanał bezprzewodowy nie zmienia się szybko. W szczególności podstawa KLT uzyskana w jednym przypadku może nadal obowiązywać w kilku kolejnych przypadkach. Tutaj proponujemy protokół sprzężenia zwrotnego CSI oparty na wykrywaniu kompresyjnym, który wykorzystuje zarówno 2D-DCT, jak i KLT, który jest przeznaczony dla scenariuszy o niskiej mobilności z wolno zmieniającymi się kanałami. Schemat umożliwia przełączanie zawartości sprzężenia zwrotnego między pomiarami CS y (dla początkowej konstrukcji CSI opartej na 2D-DCT) i wartością s (dla późniejszego przywrócenia CSI

na podstawie KLT) w celu poprawy wydajności. Schemat blokowy tego proponowanego schematu przedstawiono na rysunku poniższym .



Procedurę operacyjną proponowanego schematu wyjaśniono w następujący sposób:

1. UE szacuje chwilową macierz kanału H N_r x N_T MIMO na podstawie symboli pilota transmitowanych przez stację bazową, a następnie kompresuje zmierzony kanał do wektora M x1 y poprzez rzutowanie losowe opisane przez równanie (8). Skompresowany CSI, y, jest następnie wysyłany do stacji bazowej przy użyciu zasobu radiowego łącza w górę przeznaczonego do sprzężenia zwrotnego CSI. Dodatkowo, UE emuluje procedurę odtwarzania kanału przez wykonanie minimalizacji normy l_1, jak opisano w równaniu (5) na podstawie 2D-DCT, w celu przewidzenia odzyskanego kanału, który ma być postrzegany przez stację bazową. Warto zauważyć, że z powodu nieuniknionych błędów odtwarzania, odzyskana macierz kanału (oznaczona jako H_{est}) mogłaby nieznacznie odchylić się od kanału naziemnej prawdy H. UE powinno następnie wykonać eigendecomposition H_{est}, aby obliczyć jego podstawę KLT, oznaczoną jako Q_{est} . Podsumowując, oprócz y, UE również wyprowadza i przechowuje $Q_{e}st$ do wykorzystania w przyszłości.

2. Po odebraniu skompresowanego CSI y przez stację bazową poprzez sprzężenie zwrotne, kanał mmWave massive MIMO może zostać odtworzony przez minimalizację l_1-normy przez stację bazową z wykorzystaniem podstawy 2D-DCT. Jest to identyczne z procesem emulacji, który UE przeprowadził w kroku 1. Tutaj zakładamy, że ta sama wersja Hest jest uzyskiwana przez stację bazową poprzez rekonstrukcję kanału dzięki identycznemu algorytmowi odtwarzania. Bazując na odzyskanej macierzy kanału Hest, stacja bazowa dalej oblicza i przechowuje odpowiednią bazę KLT, Q_{est}. Ponieważ CSI jest odzyskiwany, do transmisji danych w łączu w dół można zastosować odpowiednie wagi kształtowania

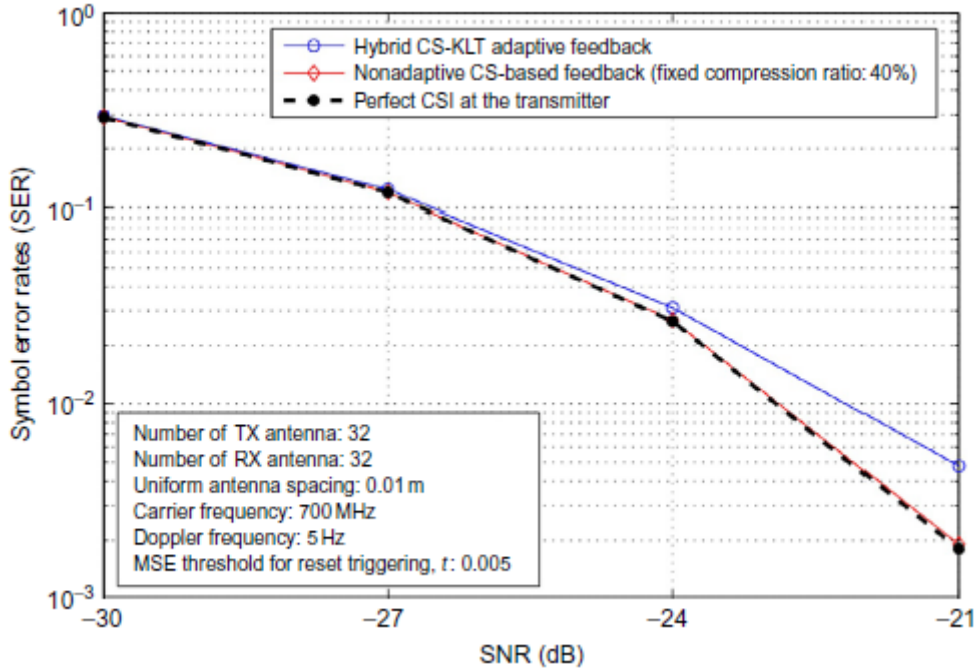
wiązki. W międzyczasie nowe symbole pilota są również przesyłane w łączu w dół w celu śledzenia zmian kanału

3. UE wykrywa dane łącza w dół, jak również szacuje najnowszy stan kanału oznaczony jako H_{new} (wersja zwektoryzowana jest zapisywana jako H_{new}). Zamiast kompresji oszacowanego kanału H_{new} przez losową projekcję, jak poprzednio, KLT jest stosowany w celu uzyskania optymalnej rzadkiej reprezentacji H_{new}, która jest rzadkim wektorem z tylko jednym niezerowym elementem (oznaczonym jako S) umieszczonym w pierwszym wpisie. Idealnie, obciążenie sprzężenia zwrotnego można znacznie zmniejszyć, ponieważ UE musi jedynie przekazać zwrotne S, co przypuszczalnie jest wystarczające do przywrócenia kanału w stacji bazowej, biorąc pod uwagę, że Q_{est} (wyprowadzone w kroku 1) jest nadal ważne ze względu na powolną zmianę kanału. Jednak przed sprzężeniem zwrotnym S, UE powinno sprawdzić ważność Q_{est} używając go do emulacji CSI odzyskiwane w oparciu o KLT:

$h_{try} = Q_{es}tS$ (14)

a następnie sprawdź średni kwadrat błędu (MSE) między $h_{tr}y$ a ostatnim kanałem h_{new}. Jeżeli MSE jest mniejsze niż wstępnie skonfigurowany poziom progowy t, wówczas UE może po prostu zgłosić S w celu uzyskania wydajnej informacji zwrotnej. W przeciwnym razie UE powinno powtórzyć krok 1 i zamiast tego zgłosić skompresowany CSI y.

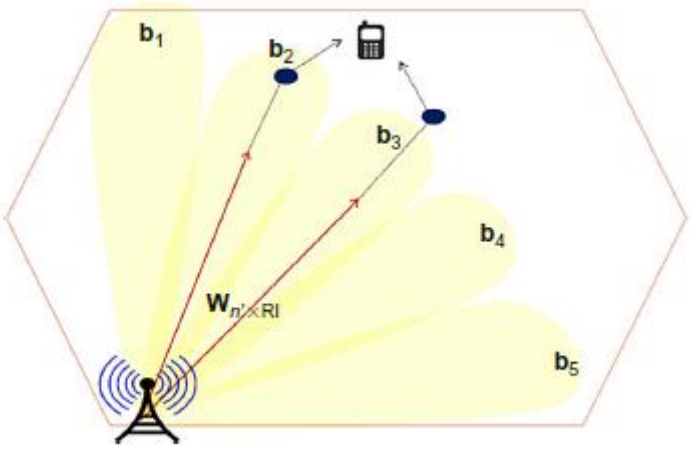
4. Po odebraniu S stacja bazowa mogłaby po prostu odzyskać kanał wymagany przez masowe operacje MIMO przy użyciu równania (9), ponieważ Q_{est} jest już dostępny (uzyskany w kroku 2), a jego ważność została przetestowana przez UE w kroku 3. Z protokołem opisanym wcześniej, raportowanie CSI dla masywnego MIMO staje się bardzo wydajne, ponieważ ładunek zawiera tylko jedną wartość skalarną (S) w większości sesji ze sprzężeniem zwrotnym, ze sporadycznym sprzężeniem zwrotnym skompresowanego wektora kanału (y), o ile kanał nie zmienia się zbyt szybko. Podczas całej operacji stosowane są zarówno podstawy 2D-DCT, jak i KLT. Wyniki symulacji na rysunku poniżej pokazują porównanie wydajności między dwoma metodami sprzężenia zwrotnego opartymi na wykrywaniu kompresyjnym. Najwyraźniej, gdy stosowana jest tylko podstawa 2D-DCT, a współczynnik kompresji jest ustalony na 40%, wynikowa wydajność jest prawie zgodna z przypadkami, w których doskonały CSI jest dostępny w nadajniku. Jednak skompresowany wektor kanałowy y musi być zgłaszany przez UE w każdej sesji sprzężenia zwrotnego. Z drugiej strony, gdy stosuje się adaptacyjną podstawę rozrzedzania schematu, obciążenie sprzężeniem zwrotnym CSI jest zmniejszane pomimo marginalnej degradacji wydajności, jak pokazano



Podejście oparte na wykrywaniu kompresyjnym w tej sekcji zapewnia potencjalny sposób na zmniejszenie wymaganego ładunku sprzężenia zwrotnego CSI dla ogromnych systemów MIMO. Jednak aby uzyskać rzadką reprezentację kanału, konieczne jest założenie wysokiej korelacji przestrzennej między antenami. W praktyce takie założenie może stać się nieaktualne, zwłaszcza gdy częstotliwość nośna zostanie zwiększona do zakresów fal milimetrowych (mmWave). Aby dostosować się do systemów pracujących na wyższych częstotliwościach, w następnej sekcji przeanalizujemy techniki akwizycji CSI oparte na formowaniu wiązki.

AKWIYJA CSI Z FORMAOWANIEM WIĄZKI W DOMENACH KĄTOWYCH

W tej sekcji omówimy zastosowania formowania wiązki w domenie kątowej w mechanizmach akwizycji CSI dla masywnych systemów MIMO. Na rysuku przedstawiono schemat transmisji sygnałów pilota łącza w dół z formowaniem wiązki w dziedzinie kątowej, w którym UE może odbierać energię sygnału i przeprowadzać estymację kanału tylko w kilku kierunkach formowania wiązki.



W oparciu o takie ustawienie w tej sekcji przedstawiono schemat sprzężenia zwrotnego z wielostopniowym kształtowaniem wiązki (np. Dwu- i trzystopniowy).

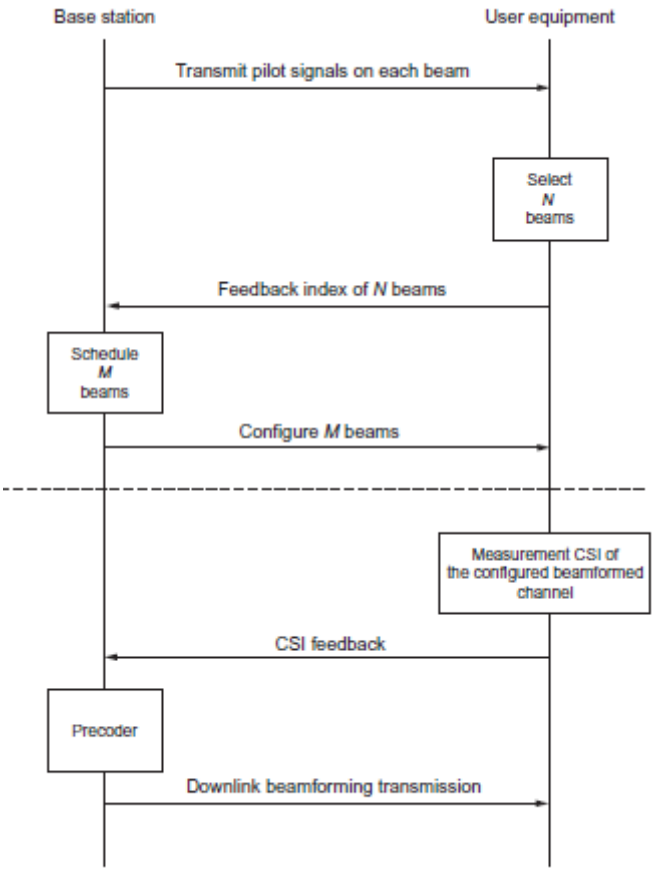
WIELOSTOPNIOWE WYŚWIETLANIE ŚWIATŁA I INFORMACJE ZWROTNE

Zamiast mierzyć i raportować CSI odpowiadające dużej liczbie anten dla masywnych systemów MIMO, jak w konwencjonalnym podejściu, możliwe jest zmniejszenie wymiarów żądanego kanału poprzez formowanie wiązki. W szczególności liczba propagacji ścieżek sygnału formowanego wiązką będzie znacznie niższa; użytkownik może skupić się na tych „wiązkach" zamiast mierzyć odpowiedzi kanałów odpowiadające wszystkim antenom nadajnika stacji bazowej. Z tej perspektywy obciążenie sprzężeniem zwrotnym może zostać zmniejszone, ponieważ użytkownik musi jedynie wziąć pod uwagę stan kanału w obszarze wiązki. Taki schemat akwizycji CSI można zrealizować za pomocą procedury sekwencyjnej składającej się z wielu etapów. Tutaj szczególnie przyjrzymy się schematowi dwustopniowemu i schematowi trójstopniowemu

Schemat akwizycji CSI z dwustopniowym kształtowaniem wiązki

W dwustopniowym schemacie kształtowania wiązki akwizycja CSI obejmuje dwie różne fazy, a mianowicie wybór wiązki i raportowanie CSI na podstawie wiązki, jak opisano poniżej:

• W pierwszym etapie stacja bazowa transmituje wiązki w wielu ustalonych kierunkach, a następnie użytkownik wybiera preferowane wiązki i informacje zwrotne w pierwszym etapie. Schemat blokowy schematu przedstawiono na rysunku



Kryterium wyboru może być oparte na odebranym poziomie energii lub stosunku sygnału do zakłóceń plus szum (SINR) mierzonych sygnałów pilota na każdej z wiązek. Po wykonaniu pomiaru użytkownik

zgłasza indeks (indeksy) wybranych belek N ≥ 1. W oparciu o raport wyboru wiązki (wiązek) otrzymany od użytkownika, stacja bazowa konfiguruje użytkownika z zestawem pewnych kierunków wiązki, które planuje zastosować do transmisji w łączu w dół. W przypadku schematów komunikacyjnych opartych na formowaniu wiązki dokładne śledzenie wiązki jest kluczem do wysokiej wydajności systemu. Jeśli śledzenie wiązki nie jest precyzyjnie wykonane, nie tylko znacznie zmniejszy odbierany SINR, ale także spowoduje zakłócenia. Zatem, aby zdecydować o konfiguracjach wiązek dla użytkownika, stacja bazowa powinna również wziąć pod uwagę inne dostępne informacje, takie jak mobilność użytkownika, kierunki ruchu i potencjalne zakłócenia ze strony wiązek skonfigurowanych dla innych użytkowników, w celu ulepszenia wydajności systemu.

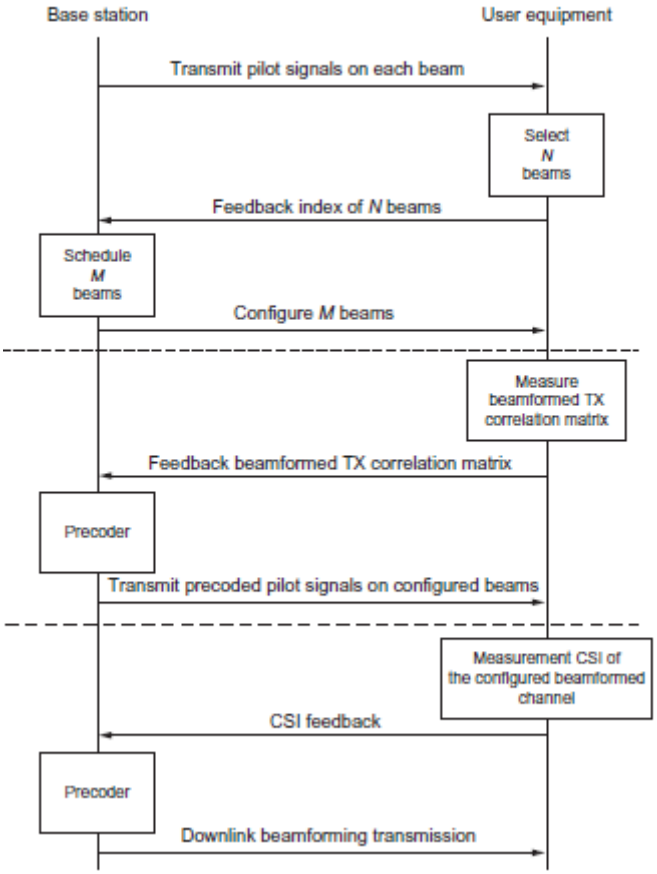
Na początku drugiego etapu stacja bazowa wykorzystuje skonfigurowane wektory wagi formowania wiązki $\{b_{Ii}\}$, aby użytkownik transmitował symbole pilota łącza nadawczego; czyli przesyłany wektor sygnału można zapisać jako

$x=Bs$ (15)

gdzie B składa się z ułożonych w stos wektorów kolumnowych $\{b_{Ii}\}$, to jest $B =[b_{I1,} b_{I2} \cdots b_{IM}]$ w którym $\{I; i = 1, 2,\ldots, M, I_i \in \{1, 2,\ldots N\}$ to zestaw indeksu wiązki wybrany przez użytkownika. W ten sposób użytkownik mierzy kanał $H_{bf} =HB$ formowany wiązką i ocenia odpowiadający mu współczynnik CSI. Dlatego rozmiar kanału formowanego wiązką jest zmniejszony z N_r x N_t do N_r x M: Ponieważ liczba wiązek (M) skierowanych do określonego UE jest zwykle znacznie mniejsza niż liczba anten nadawczych (N_t), wymiar MIMO kanał jest znacznie zmniejszony. Do transmisji danych stacja bazowa stosuje prekoder W z CSI zgłoszonego przez użytkownika do transmisji danych w łączu w dół; to znaczy, odebrany sygnał w UE może być zapisany jako x= BWs. Najwyraźniej sygnał przechodzi dwa procesy kodowania wstępnego, a zatem schemat ten jest określany jako dwuetapowe formowanie wiązki.

Schemat akwizycji CSI z trzystopniowym formowaniem wiązki

Wykazano, że wymiar efektywnego kanału po formowaniu wiązki jest drastycznie zmniejszony w dwustopniowym schemacie kształtowania wiązki. Co ciekawe, wymiar można jeszcze bardziej zmniejszyć, jeśli znana jest struktura własna macierzy kanału. Wynika to z faktu, że anteny nadajnika są ze sobą skorelowane i / lub istnieje tylko niewielka liczba dominujących ścieżek, które istnieją między stacją bazową a użytkownikiem. Obydwa czynniki spowodują, że macierz kanału H_{bf} = HB ukształtowana wiązką będzie miała niedobór rang. Wykorzystując tę właściwość, stacja bazowa może skoncentrować swoją moc transmisji tylko na dominujących wektorach własnych, poprawiając w ten sposób efektywność energetyczną stacji bazowej i dodatkowo ograniczając narzut sprzężenia zwrotnego dla UE. Aby to osiągnąć, użytkownik musi dostarczyć więcej informacji do stacji bazowej. W szczególności stacja bazowa potrzebuje informacji, aby ułatwić alokację mocy między wektorami własnymi. Wiadomo, że macierz kanału można rozłożyć na czynniki jako $H_{bf} = R^{1/2}{}_{R_bf} H_{iid} R^{1/2}{}_{T_bf}$, gdzie $R_{T_bf} =E\{H_{bf}{}^H H_{bf}\}$ jest macierzą korelacji nadajnika, $R_{Rbf} = E\{H_{bf} H_{bf}{}^H\}$ to macierz korelacji odbiornika, a H_{iid} to macierz złożona z i.i.d. zmiennych losowyh. SVD dla R_{R_bf} można zapisać jako $R_{R_bf} = UDU^H$, w którym kolumny U są jego wektorami własnymi, a elementy diagonalne macierzy diagonalnej D są powiązanymi wartościami własnymi. Podobnie mamy $R^{1/2}{}_{T_bf}= UD^{1/2}UH$: Zwróć uwagę, że ranga $R_{T_bf}{}^{1/2}$, to znaczy R jest mniejsza lub równa liczbie skonfigurowanych belek M, stąd możliwa jest dalsza redukcja wymiarów. Dokładniej, jeśli stacja bazowa nadaje tylko na wektorach własnych, to znaczy przez prekoder U o wymiarze M x R, wówczas efektywny kanał staje się $H_{bf}U=HBU$, którego wymiar wynosi N_r x R: Przypomnij sobie, że R ≤ M; w ten sposób wymiar kanału formowanego wiązką jest zmniejszony. Wreszcie, po przesłaniu CSI kanału $H_{bf}U$ do stacji bazowej, sygnał danych łącza w dół może być wstępnie zakodowany jako x=BUWs. Sieć działań tego trzystopniowego schematu kształtowania wiązki jest zilustrowana poniżej



Zwróć uwagę, że pierwszy i trzeci stopień są identyczne z dwoma stopniami pokazanymi na rysunku wzweśniejszym. W drugim etapie użytkownik mierzy macierz korelacji nadajnika i zwraca jej wektory własne do stacji bazowej. W porównaniu z dwustopniowym formowaniem wiązki, dodatkowo potrzebna jest informacja zwrotna. Na szczęście dzięki temu, że macierz korelacji nadajnika jest statystyką długoterminową, okresowość raportowania można ustawić na dłuższą niż w raporcie CSI trzeciego etapu. Tak więc, jeśli $R \ll M$, możliwe jest osiągnięcie jeszcze niższego narzutu sprzężenia zwrotnego niż w dwustopniowym formowaniu wiązki.

WSTĘPNE KODOWANIE WSTĘPNE W FDD NA PODSTAWIE KĄTA PRZYBYCIA

W przeciwieństwie do poprzednich sekcji, które koncentrują się na bardziej wydajnych sprzężeniach zwrotnych CSI, ta sekcja bada alternatywne mechanizmy akwizycji CSI dla masywnych systemów MIMO FDD wykorzystujących techniki formowania wiązki. Stacje bazowe wyposażone w dużą liczbę anten mają możliwości formowania wiązki nadawczej i odbiorczej, biorąc pod uwagę, że anteny są rozmieszczone jako jednolity liniowy układ (ULA). Wiadomo, że użycie kształtowania wiązki w ogromnym systemie MIMO ma duży potencjał w zwiększaniu wydajności sumarycznej szybkości w scenariuszach z wieloma użytkownikami, zakładając, że informacje CSI są doskonale znane. Donoszono, że formowanie wiązki jest również pomocne w pozyskiwaniu częściowych informacji CSI, które umożliwiają wstępne kodowanie łącza w dół w masowych systemach MIMO FDD. W tej sekcji przedstawiamy niektóre z tych technik . Główną cechą wyróżniającą podejście przedstawione w tej sekcji jest to, że rezygnuje z wszelkich wysiłków na rzecz oszacowania CSI łącza w dół w celu wstępnego kodowania łącza w dół. Następnie żadne wyraźne sprzężenie zwrotne CSI nie będzie również

dostarczane do stacji bazowej w kanale zwrotnym, ponieważ UE i tak nie ma żadnych informacji CSI przed transmisją w łączu w dół. Zamiast tego stacja bazowa opiera się na swojej własnej zdolności do odbierania kształtowania wiązki w celu uzyskania częściowych informacji CSI dla UE za pośrednictwem wcześniejszej sygnalizacji łącza w górę z UE. Przykłady takich sygnałów łącza w górę zainicjowanych przez aUE obejmują żądanie dostępu swobodnego, potwierdzenie poprzedniej transmisji łącza w dół, transmisja danych w łączu w górę i tak dalej. Wszystkie te rodzaje sygnalizacji nie zawierają informacji o CSI łącza w dół. Jednak stacja bazowa wyposażona w wielkoskalowy układ antenowy jest w stanie uzyskać oszacowania kątów docierających ścieżek propagacji, przez które przesyłany był sygnał łącza w górę. Główną zaletą tego podejścia jest to, że nie wymaga przydziału zasobów transmisyjnych do celów wstępnego kodowania łącza w dół. Chociaż w FDD właściwość wzajemności kanałów nie może być zastosowana jak w TDD, należy zauważyć, że kąty przychodzące ścieżek propagacji łącza w górę z UE mają silne korelacje z odchylającymi się kątami ścieżek propagacji w transmisji w łączu w dół do tego samego UE. Wykorzystując tę właściwość, można zaprojektować współczynniki kodowania wstępnego dla transmisji w łączu w dół bez dodatkowych kosztów związanych z przepustowością. Zatem wewnętrzne korzyści z łatwiejszego pozyskiwania CSI w TDD mogą być również wykorzystane przez FDD. Podstawową ideą przedstawionych tutaj konstrukcji prekodera łącza w dół jest zastosowanie technik formowania wiązki i sterowanie mocą nadawania w kierunku, w którym określamy, że docelowy UE otrzymuje największą moc, a jednocześnie minimalizowanie mocy kierowanej do innych UE uzyskujących dostęp do tego samego zasób transmisji. Chociaż UE może odbierać moc ze stacji bazowej pod wieloma kątami, wybralibyśmy tylko jeden z kierunków, aby uniknąć potencjalnych destrukcyjnych zakłóceń. Prostym sposobem realizacji tego pomysłu jest zastosowanie metody podziału przestrzeni wiązek, w której przestrzeń w dziedzinie kątów jest równo podzielona na M wzajemnie prostopadłych wiązek. Jednak metoda podziału przestrzeni wiązki może mieć następujące wady. Po pierwsze, gdy liczba anten w stacji bazowej jest mała, sposób ten powodowałby zakłócenia dla innych UE z powodu dużej mocy listwy bocznej formującej wiązkę. Po drugie, gdy liczba anten w stacji bazowej jest duża, wynikowa szerokość wiązki jest wąska, a zatem wydajność transmisji w łączu w dół byłaby bardzo wrażliwa na błędy w oszacowaniu kątów docierających sygnału łącza w górę. Aby złagodzić te dwa problemy, zaproponowano pewne konstrukcje prekoderów w celu zwiększenia energii transmisji w kierunku docelowego UE i stłumienia zakłóceń powodowanych przez inne UE usytuowane w różnych kierunkach. Opierając się na poprzednim pomyśle, przedstawiamy tutaj dwie metody projektowania prekoderów dla wielu użytkowników bez sprzężenia zwrotnego CSI. Pierwsza wykorzystuje szeroko stosowane w literaturze radarowej techniki obciążenia diagonalnego (DL) [17]; druga traktuje problem formowania wiązki jako problem projektowania filtra w dziedzinie przestrzennej i wykorzystuje klasyczny algorytm Parks-McClellan (PM) [18]. Zaobserwowano, że proponowane metody mają tylko niewielki spadek wydajności w porównaniu z metodą idealnej diagonalizacji bloków (BD), która wymaga pełnego, dokładnego sprzężenia zwrotnego CSI.

TŁO I ZAŁOŻENIA

Przed zagłębieniem się w szczegóły proponowanego schematu, najpierw przedstawiamy model kanału i tło wymagane do analizy.

Model kanału

Rozważana jest stacja bazowa wyposażona w ULA z antenami M. Załóżmy, że stacja bazowa obsługuje K UE z pojedynczą anteną przy użyciu tych samych zasobów czasowo-częstotliwościowych ($K \ll M$). Stacja bazowa wykorzystuje FDD z częstotliwością nośną łącza w górę f_u i częstotliwością nośną łącza w dół f_d. Zmienne λ_u i λ_d oznaczają odpowiednie łącze w górę i odpowiednio długości fal łącza nadawczego; to znaczy $\lambda_u=c/f_u$ i $\lambda_d=c/f_d$, gdzie c jest prędkością światła. Oznaczamy D jako odstęp

anteny ULA i zakładamy, że odstęp między antenami jest połową długości fali częstotliwości nośnej łącza nadawczego, to znaczy $\lambda_u \approx \lambda_d = 2D$. Kanał propagacji radiowej z k-tego UE do stacji bazowej składa się ze ścieżek P i jest scharakteryzowany jako

$$c_k(t;\theta) = \sum_{p=1}^{P} \alpha_{kp}\delta(t-\tau_{kp})\delta(\theta-\theta_{kp}) \tag{16}$$

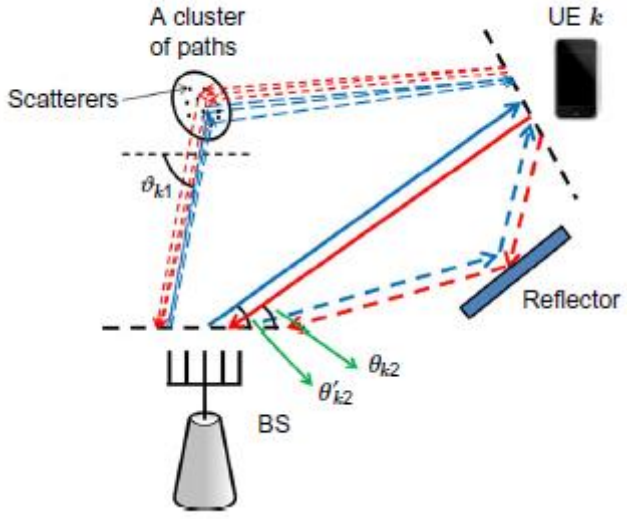
gdzie ścieżka p-ta ma złożone wzmocnienie α_{kp}, opóźnienie czasowe τ_{kp} z AoA θ_{kp}. AoA toru jest definiowane jako kąt między kierunkiem toru przybywającego a linią równoległą do szyku antenowego. Bez utraty ogólności mamy $-\pi/2 \le \theta_{kp} < \pi/2$. Załóżmy, że te ścieżki propagacji można pogrupować jako oddzielne klastry, z których każda składa się ze ścieżek, które mają bliskie AoA. Zgodnie z grupowaniem kanał można przepisać jako

$$c_k(t;\theta) = \sum_{l=1}^{L}\sum_{p=1}^{P} \alpha_{klp}\delta(t-\tau_{klp})\delta(\theta-\theta_{klp}) \tag{17}$$

gdzie L oznacza liczbę klastrów, P_l oznacza liczbę ścieżek l-tego klastra, a α_{klp}, τ_{klp} i θ_{klp} są odpowiednimi parametrami dla p-tej ścieżki w l-tym klastrze. Dla każdego klastra można zdefiniować reprezentatywny kąt ϑ_{kl}, aby umożliwić dowolnej ścieżce w tym klastrze spełnienie następującej nierówności z góry określonym progiem θ_{di}:

$$\left|\theta_{klp} - \vartheta_{kl}\right| < \theta_{di},\ \forall p \in \{1,\ldots,P_l\} \text{ and } \forall l \in \{1,\ldots,L\} \tag{18}$$

Przykład takiej grupy jest pokazany jako przerywane strzałki w okręgu oznaczonym „Cluster of Paths" na rysunku



Podobnie kanał łącza w dół można modelować jako

$$h_k(t,\theta) = \sum_{p=1}^{P} \beta_{kp}\delta(t-\tau_{kp}')\delta(\theta-\theta_{kp}') = \sum_{l=1}^{L}\sum_{p=1}^{P_l'} \beta_{klp}\delta(t-\tau_{klp}')\delta(\theta-\theta_{klp}') \quad (19)$$

gdzie P_l' oznacza liczbę ścieżek w l-tej ścieżce, β_{klp} jest wzmocnieniem zespolonym, τ_{klp}' oznacza opóźnienie czasowe, a θ_{klp}' jest kątem wyjścia (AoD) w łączu w dół. AoD łącza w dół pojedynczej ścieżki θ_{kp}' może być równe AoA jej ścieżek łącza w górę, gdy ścieżki propagacji są wyłącznie linią widzenia lub odbiciami. Taka wzajemność nie zachodzi już, gdy ścieżka przebiega przez załamanie, ponieważ rozproszenie fal elektromagnetycznych powoduje rozbieżności w ścieżkach przemieszczania się różnych częstotliwości radiowych. Jednak utrzymując $\lambda_u \approx \lambda_d$, taką rozbieżność w podróżowaniu kierunek może być ograniczony i możemy to założyć

$$\left|\theta_{klp}' - \vartheta_{kl}\right| < \theta_{\mathrm{di}},\ \forall p\left\{1,\ldots,P_l'\right\} \text{ and } \forall l \in \{1,\ldots,L\} \quad (20)$$

Zakładamy ponadto, że modulacja multipleksowania z ortogonalnym podziałem częstotliwości z prefiksami cyklicznymi jest stosowana w nadajnikach-odbiornikach pasma podstawowego tak, że dla każdej podnośnej można założyć, że kanał łącza zwrotnego ma płaskie częstotliwości. Wektor sterujący łącze w dół można zdefiniować jako

$$\boldsymbol{a}_{\mathrm{d}}(\theta) = \frac{1}{\sqrt{M}}\left[1, \mathrm{e}^{-j2\pi(D/\lambda_{\mathrm{d}})\sin\theta}, \ldots, \mathrm{e}^{-j2\pi(M-1)(D/\lambda_{\mathrm{d}})\sin\theta}\right]^T \quad (21)$$

Niech x[n] oznacza przesyłany sygnał wektorowy; następnie odebrany sygnał w k-tym UE jest

$$y_k[n] = \mathbf{h}_k^{\mathrm{T}} x[n] + v_k[n] \quad (22)$$

gdzie $v_k[n]$ jest średnią złożoną zmienną losową Gaussa o zerowej średniej z wariancją σ_v^2, i

$$\mathbf{h}_k = \sum_{l=1}^{L}\sum_{p=1}^{P_l} \beta_{klp}\mathrm{e}^{-j2\pi f_{\mathrm{d}}\tau_{klp}'} a_{\mathrm{d}}\left(\theta_{klp}'\right) \quad (23)$$

W transmisji w dół używamy terminu $s_k[n]$ do oznaczenia sygnału przenoszącego informacje dedykowanego dla k-tego UE. Sygnał $s_k[n]$ jest wstępnie kodowany przez wektor $_{wk}$, w wyniku czego przesyłany jest sygnał jako

$$x[n] = \sum_{k=1}^{K} \boldsymbol{w}_k^* s_k[n] \quad (24)$$

W konsekwencji odebrany sygnał w k-tym UE w równaniu. (22) jest

$$y_k[n] = \mathbf{h}_k^{\mathrm{T}}\left(\boldsymbol{w}_k^* s_k[n] + \sum_{j\neq k} \boldsymbol{w}_j^* s_j[n]\right) + v_k[n] \quad (25)$$

Wstępne kodowanie łącza w dół dla wielu użytkowników

Celem tej sekcji jest opracowanie metod wstępnego kodowania, aby umożliwić transmisję w łączu w dół, która może jednocześnie obsługiwać wiele UE poprzez zaprojektowanie wektora (tj. w_k, k1, 2,…, K) w równaniu. (24). Projekt opiera się na następujących założeniach:

• Stacja bazowa może uzyskać oszacowanie ϑ_{kl}, oznaczone jako $\hat{\vartheta}_{kl}$, dla wszystkich k ∈ {1, 2,… f, K} oraz l ∈ {1,2,…L} poprzez poprzednią sygnalizację łącza w górę. Błąd oszacowania reprezentatywnego kąta jest ograniczony z góry określoną wartością θ_{er}; innymi słowy,

$$\left|\vartheta_{kl}-\hat{\vartheta}_{kl}\right|<\theta_{er} \quad (26)$$

• Zdefiniuj $\nu_{kl}=\left|\mathbf{a}_{\mathrm{d}}^{\dagger}\left(\vartheta_{kl}^{\prime}\right)\mathbf{h}_{k}\right|$, co z grubsza reprezentuje energię przesyłaną przez kierunek ϑ_{kl} w łączu w dół. Podobnie zdefiniuj μ_{kl} jako odpowiednik łącza w górę. Zakładamy, że stacja bazowa zna μ_{kl}, ale nie jest świadoma ν_{kl}. Niemniej jednak przypuszcza się, że ν_{kl} jest silnie skorelowane z μ_{kl}, jeśli częstotliwości nośne łącza w górę i łącza w dół są wystarczająco blisko siebie. A konkretnie zakłada się, że ν_{kl} ma dużą wartość z dużym prawdopodobieństwem, ilekroć μ_{kl} jest duże, jak sugerowano w poprzednim badaniu.

• Różnica kątów między skupieniami spełnia

$$\left|\vartheta_{kl}-\vartheta_{km}\right|>\theta_{\mathrm{th}}, l\neq m, \forall l,m\in\{1,2,\ldots,L\} \quad (27)$$

gdzie θ_{th} oznacza próg rozstawu kątów.

• Dowolna para UE ze ścieżkami blisko siebie może być przypisana do różnych zasobów częstotliwości czasowych. Dlatego w przypadku UE uzyskujących dostęp do tych samych zasobów czasowo-częstotliwościowych możemy to założyć

$$\left|\vartheta_{kl}-\vartheta_{jm}\right|>\theta_{\mathrm{th}}, j\neq k, \forall j,k\in\{1,2,\ldots,K\} \text{ and } l,m\in\{1,2,\ldots,L\} \quad (28)$$

• Wreszcie

$$\left|\theta_{kp}\right|\leq\theta_{\mathrm{ext}}<\pi,\ \forall k\in\{1,2,\ldots,K\} \text{ and } \forall p\in\{1,2,\ldots,P\} \quad (29)$$

gdzie θ_{ext} oznacza skrajną wartość θ, co sugeruje, że kąty UE obsługiwane przez ten zestaw anten nie mogą znajdować się w obszarach [-π, θ_{ext}) i (θ_{ext}, π). Założenie to jest uzasadnione następującym argumentem: jeśli w tych regionach znajdują się UE, to mogą być obsługiwane przez inny prostopadły układ anten stacji bazowej.

ŁĄCZE WSTĘPNE PROJEKTOWANIE WSTĘPNE PRZY UŻYCIU CZĘŚCIOWEGO CSI OPARTEGO NA BEAMFORMING

Metoda MVDR

Formowanie wiązki z minimalną wariancją zniekształcenia bez zniekształceń (MVDR) jest dobrze znane z łagodzenia zakłóceń z AoA, które różnią się od żądanego sygnału. Jeżeli AoA żądanego sygnału nie jest dokładnie znana, metody DL można użyć, aby uniknąć nieoczekiwanego tłumienia żądanego sygnału. Metoda MVDR-DL wykorzystuje macierz korelacji w dziedzinie przestrzennej. W zastosowaniu będącym przedmiotem zainteresowania macierz korelacji przestrzennej można wyrazić jako

$$\mathbf{R} = E_s \sum_{k=1}^{K} \sum_{l=1}^{L} |\mu_{kl}|^2 \boldsymbol{a}_d(\hat{\vartheta}_{kl}) \boldsymbol{a}_d^{\dagger}(\hat{\vartheta}_{kl}) + \frac{\sigma_v^2}{M} \mathbf{I}_M \quad (30)$$

gdzie E_s oznacza średnią energię $s_k[n]$. Formator wiązki DL rozwiązuje następujący problem optymalizacji:

$$\min_{\mathbf{w}_k} \mathbf{w}_k^{\dagger} (\mathbf{R} + \gamma_k \mathbf{I}_M) \mathbf{w}_k \text{ subject to } \boldsymbol{a}_d^{\dagger}(\theta_k) \mathbf{w}_k = 1 \quad (31)$$

gdzie $\theta_k = \hat{\vartheta}_{k\hat{l}_k}$ oznaczają współczynnik DL UE k. Rozwiązaniem tego problemu jest $w_k = u_k||u_k||$gdzie

$$\mathbf{u}_k = \frac{(\mathbf{R} + \gamma_k \mathbf{I}_M)^{-1} \boldsymbol{a}_d(\theta_k)}{\boldsymbol{a}_d^{\dagger}(\theta_k)(\mathbf{R} + \gamma_k \mathbf{I}_M)^{-1} \boldsymbol{a}_d(\theta_k)} \quad (32)$$

Współczynnik DL γ_k dla UE k jest wybierany jako

$$\gamma_k = -\left(\frac{\sigma_v^2}{M} + \left| \mu_{k\hat{l}_k} \right|^2 E_s \right) \quad (33)$$

maksymalizować SINR. Prekoder DL jest skuteczny, gdy oszacowany kierunek żądanego sygnału (tj. θ_k) jest dokładny. Niemniej jednak kąt θk może być niedopasowany z powodu niedokładnych szacunków AoA. W takich przypadkach niedopasowane kąty zmniejszyłyby wydajność transmisji w łączu w dół. Dlatego proponujemy inny sposób wstępnego kodowania mający na celu poprawę odporności na niedopasowanie kątów.

Podejście do projektowania filtrów w dziedzinie przestrzennej

Tutaj przedstawiamy inną metodę projektowania prekodera na podstawie spojrzenia na problem projektowy kształtowania wiązki jako problem z projektowaniem filtru o skończonej odpowiedzi impulsowej (FIR) w dziedzinie przestrzennej. Stosujemy uogólniony algorytm PM , który rozszerza znany algorytm PM FIR na przypadek dopuszczenia parametrów o wartościach zespolonych. Załóżmy, że projektujemy filtr FIR, oznaczony jako $w_{[n]}$, n=0, 1,…, M-1, z jego transformatą Fouriera wyrażoną jako

$$W(e^{j\omega}) = \sum_{n=0}^{M-1} w[n] e^{-j\omega n} \quad (34)$$

Definiujemy $w \triangleq w[0]w[1] \cdots w[M-1]^T$. Problem projektu optymalnego filtra polega na minimalizacji maksymalnej wartości ważonego błędu aproksymacji odpowiedzi częstotliwościowej; innymi słowy

$$\min_{w} \left\{ \max_{w \in B} \left[G(e^{j\omega}) \left| F(e^{j\omega}) - W(e^{j\omega}) \right| \right] \right\} \quad (35)$$

gdzie $F(e^{j\omega})$ oznacza pożądaną odpowiedź częstotliwościową, $G(e^{j\omega})$ jest nieujemną funkcją ważenia, a B reprezentuje zwarty podzbiór w [-π, π]. W przypadku problemu projektowania wektora wstępnego kodowania domena kątowa odpowiada domenie częstotliwości. W szczególności rozważaliśmy $\omega \triangleq \pi \sin\theta$. W przypadku konkretnego UE chcieliśmy zwiększyć moc w pożądanym kierunku i wyeliminować zakłócenia w klastrach AoA innych UE. Dlatego idealny wektor wstępnego kodowania w_k dla UE k

powinien mieć tę właściwość, że jego odpowiedź w dziedzinie kątowej $W_k(e^{j\omega})$, $\omega \in B$ jest jednością w sąsiedztwie $\pi \sin \theta$ i 0 w sąsiedztwie $\pi \sin_{\theta j}$ dla $j \neq k$. Niech δ_θ oznacza maksymalny dopuszczalny zakres niedopasowania kątów. Najpierw definiujemy

$$C_r = [\pi \sin(\theta_r - \delta_\theta), \pi \sin(\theta_r + \delta_\theta)], r \in \{0, 1, \ldots, KL\} \quad (36)$$

Należy zauważyć, że $C_r \cap C_q = \emptyset$, $q \neq r$, ponieważ klastry AoA różnych UE są rozdzielone, jak wyrażono w (28). Niech q(k) oznacza indeks pożądanego klastra dla UE k (tj. $q(k) = \hat{l}_k$); następnie można zdefiniować zwarty podzbiór jako

$D_{kr} = [a, b]$ (37)

gdzie

$$a = \begin{cases} \pi \sin(\theta_r + \beta\delta_\theta), & \text{if } r = q(k) \\ -\pi/2, & \text{if } r = 0 \\ \pi \sin(\theta_r + \delta_\theta), & \text{otherwise.} \end{cases} \quad (38)$$

i

$$b = \begin{cases} \pi \sin(\theta_r - \beta\delta_\theta), & \text{if } r + 1 = q(k) \\ -\pi/2, & \text{if } r = KL \\ \pi \sin(\theta_r - \delta_\theta), & \text{otherwise.} \end{cases} \quad (39)$$

Ten podzbiór zwarty nazywamy dzikim zestawem kompaktowym, który nie zawiera klastrów AoA UE. W związku z tym kompaktowy podzbiór można zapisać jako

$$\boldsymbol{B}_k = \left(\cup_{r=1}^{KL} C_r\right) \cup \left(\cup_{r=0}^{KL} D_{kr}\right) \quad (40)$$

W oparciu o te definicje, pożądana odpowiedź domeny kątowej dla UE k wynosi

$$F_k\left(e^{j\omega}\right) = \begin{cases} 1, & \omega \in C_{q(k)} \\ 0, & \omega \in \boldsymbol{B}_k\, C_{q(k)} \end{cases} \quad (41)$$

Funkcja ważenia $G_k(e^{j\omega})$ została zaprojektowana jako

$$G_k\left(e^{j\omega}\right) = \begin{cases} \rho_d, & \omega \in C_{q(k)} \\ \rho_i, & \omega \in C_r, r \neq q(k) \\ \rho_v, & \omega \in \boldsymbol{D}_{kr} \end{cases} \quad (42)$$

gdzie ρ_d, ρ_i i ρ_v oznaczają odpowiednio wagi dla pożądanych, interferencyjnych i dzikich kierunków. Aby uzyskać wysoką SINR i energooszczędną transmisję w dół, wybraliśmy $\rho_d \approx \rho_i \gg \rho_v$. Podstawiając równania (40) - (42) do równania (35), otrzymujemy optymalny wektor wstępnego kodowania dla UE k, oznaczony jako w_k.

WYNIKI SYMULACJI

W naszej symulacji średnią sumaryczną stawkę, która została użyta jako miernik wydajności, można obliczyć według

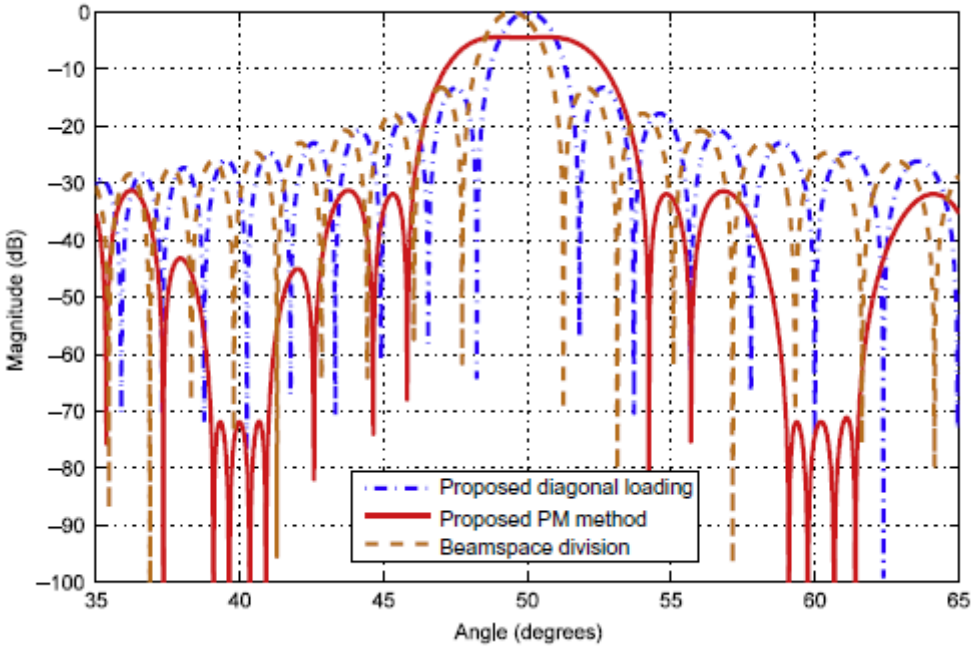
$$C_{\text{avg}} = \frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{K}\log_2(1+\text{SINR}_k(t)) \tag{43}$$

gdzie T jest liczbą prób, a SINR UE k podczas próby t jest obliczane przez

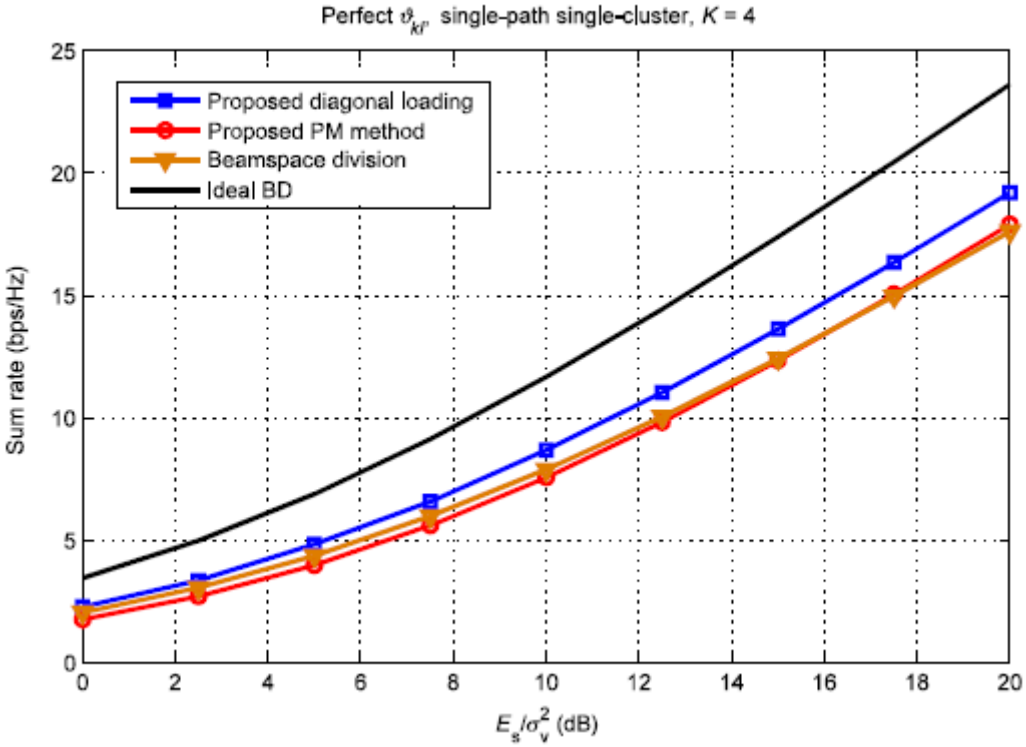
$$\text{SINR}_k(t) = \frac{\left|\mathbf{w}_k^{\dagger}(t)\boldsymbol{h}_k(t)\right|^2 E_s}{\sum_{j\neq k}\left|\mathbf{w}_j^{\dagger}(t)\boldsymbol{h}_k(t)\right|^2 E_s + \sigma_v^2} \tag{44}$$

Liczbę prób na wszystkich wykresach symulacyjnych ustawiono na T=1000.

Rysunek poniższy ilustruje wzór prekodera proponowanych metod i metodę podziału przestrzeni wiązki.



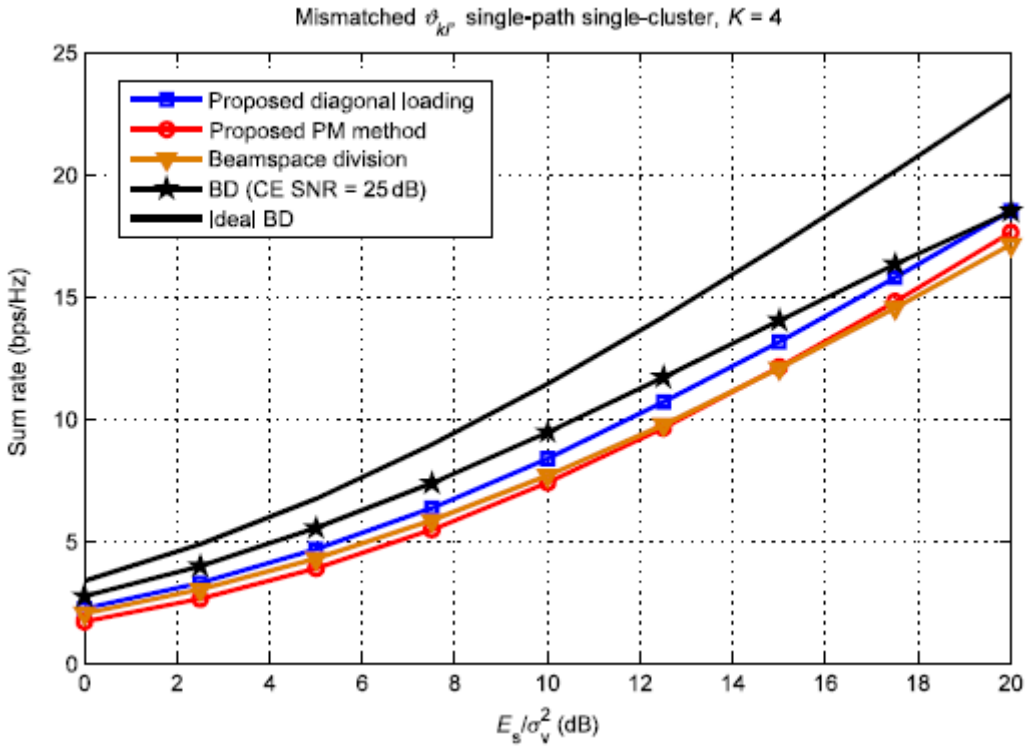
Na tym rysunku docelowe UE znajduje się pod kątem 50 stopni, a dwa inne UE są usytuowane pod kątem 40 i 60 stopni. Wzorzec proponowanego prekodera, który wykorzystuje metodę DL, koncentruje moc transmisji w pożądanym kierunku z wąską dopuszczalną szerokością niedopasowania kątowego. Natomiast wzór proponowanego prekodera, który wykorzystuje metodę PM, ma około 4 dB straty wielkości, ponieważ moc transmisji jest rozłożona w rozszerzonej szerokości wiązki, aby złagodzić problem niedopasowania kątów. Pod kątami innych UE, prekoder ma korzystne tłumienie tylko -70 dB. Ta właściwość sprzyja uzyskaniu korzystnego SINR. Poniżej przedstawiono sumaryczną wydajność systemu zawierającego cztery UE.

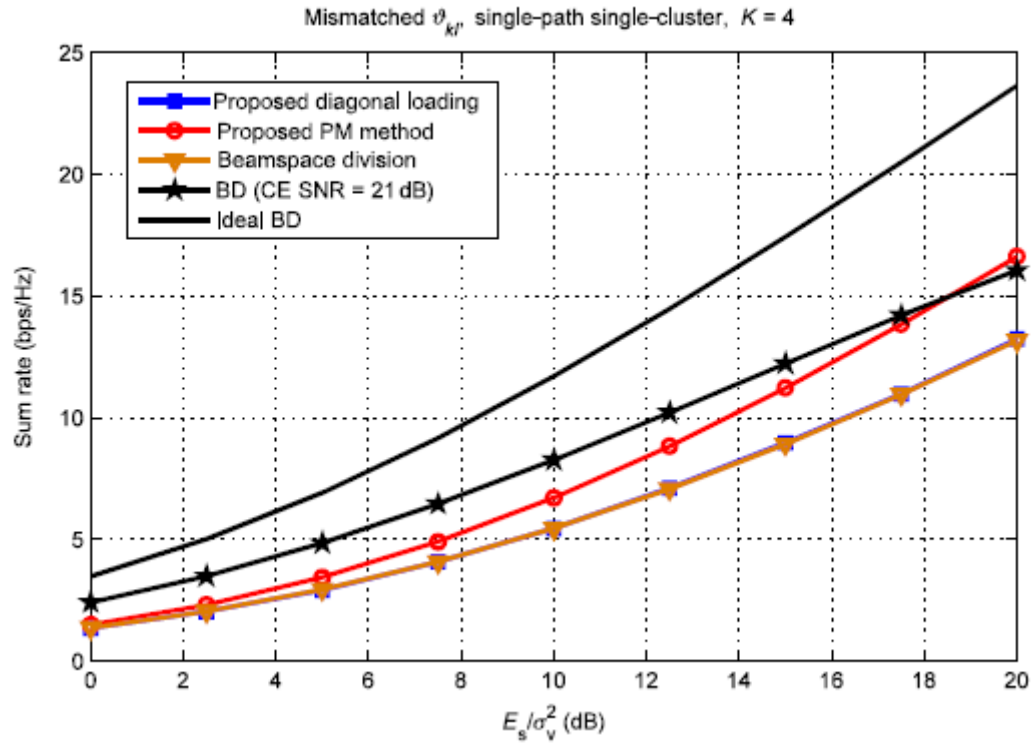


Rozważana jest tutaj doskonała znajomość reprezentatywnych kątów (tj. $\hat{\vartheta}_{kl} = \vartheta_k$, ∀k=1, 2,…, K a l = 1, 2,…, L). Ten rysunek pokazuje, że proponowane metody i metoda podziału przestrzeni wiązki różnią się nieznacznie pod względem wydajności. Wydajność proponowanej metody PM jest nieco gorsza niż proponowanej metody DL z powodu strat wielkości w pożądanych kątach. Na poniższych wykresach przedstawiamy wyniki symulacji niedopasowanych przypadków (tj. Oszacowane reprezentatywne kąty są niedopasowane). W szczególności rozkład oszacowanego reprezentatywnego kąta jest wyrażony jako

$$\hat{\vartheta}_{kl} \backsim U(\vartheta_{kl} - \theta_{er}, \vartheta_{kl} + \theta_{er}) \quad (45)$$

gdzie θ_{er} oznacza granicę niedopasowania kąta zdefiniowanego w równaniu (26). Ograniczone θ_{er} są ustawione na 0,5 stopnia i 2 stopnie na poniższych rysunkach odpowiednio.





Wykresy symulacyjne opierają się na uśrednieniu wyników dla różnych niedopasowanych wartości wygenerowanych z danych rozkładów. Rysunki przedstawiają symulowaną wydajność każdej metody z niedopasowanym ϑ_{kl}, pokazując, że wydajność proponowanej metody DL jest nieco bardziej korzystna niż proponowana metoda PM. Z drugiej strony, jak pokazano na ostatnim rysunku, gdy wartość θ_{er} zostanie zwiększona do 2 stopni, proponowana metoda PM przewyższyła inne metody i

uzyskała podobną wydajność z metodą BD, ponieważ ma szerszą szerokość wiązki, aby złagodzić spadek wydajności spowodowane niedopasowaniem kątów. Obserwacje te wykazały, że proponowana metoda DL przewyższa proponowaną metodę PM, gdy niedopasowany kąt jest mniejszy niż 0,5 stopnia, ale jej przewaga znika, gdy niedopasowany kąt jest większy.

PODSUMOWANIE

Przejęcie CSI jest ważnym, a jednocześnie trudnym problemem dla ogromnych systemów MIMO. Odpowiednio, ten rozdział zawiera przegląd kilku różnych kierunków w tym kontekście. Zasugerowano, że wykrywanie kompresyjne zmniejsza obciążenie sprzężeniem zwrotnym CSI, które polega na potencjalnej rzadkiej reprezentacji kanałów MIMO w korelacji przestrzennej. W przypadku wolno zmieniającego się środowiska propagacji hybrydowa metoda kompresji ze sprzężeniem zwrotnym wykorzystująca dwie różne podstawy rozpraszania (DCT i KLT) może dodatkowo zapewnić równowagę między obciążeniem sprzężeniem zwrotnym a wydajnością odzyskiwania CSI. Niemniej jednak, ponieważ masywne MIMO prawdopodobnie zostaną rozmieszczone w zakresach częstotliwości mmWave w sieciach komórkowych nowej generacji, podejście oparte na wykrywaniu kompresyjnym może nie być wykonalne, ponieważ wystąpienie silnej korelacji przestrzennej jest mało prawdopodobne w przypadku wysokich częstotliwości nośnych, takich jak mmWave. Zamiast tego, ze względu na ograniczoną liczbę ścieżek propagacji, podejście oparte na formowaniu wiązki jest prawdopodobnie bardziej praktyczne, w którym użytkownik może skupić się tylko na kilku wiązkach kątowych do pomiarów CSI i raportowania. Ponadto, wykorzystując fakty, że AoA sygnałów łącza w górę mogą być wyrównane z odchylającymi się kątami łącza w dół, w tym rozdziale omówiono również technikę wstępnego kodowania łącza w dół, która nie wymaga sprzężenia zwrotnego CSI od użytkowników, co jest bardzo wydajne pod względem oszczędności zasobów i energii przy po stronie użytkownika. Schematy przeanalizowane w tym rozdziale stworzyły podstawę do przyszłego rozwoju technologii mmWave MIMO w systemach 5G.