WPROWADZENIE

Podczas gdy podstawy kodowania wstępnego (formowania wiązki) są takie same niezależnie od częstotliwości nośnej, przetwarzanie sygnału w masywnych systemach MIMO z wieloma wejściami i wieloma wyjściami (MIMO) na falach milimetrowych (mmWave) podlega zestawowi nietrywialnych ograniczeń praktycznych . Tradycyjne analogowe formowanie wiązki w komunikacji mmWave jest zwykle realizowane przez obwody analogowe, takie jak sieć przesuwnika fazowego (PS) z niewielką liczbą łańcuchów RF, co nakłada ograniczenia na stałą amplitudę konstrukcji analogowego formatora wiązki . W rezultacie analogowe formowanie wiązki ucierpi z powodu poważnej utraty wydajności, chociaż jest łatwe do wdrożenia w sprzęcie. Z kolei cyfrowe kodowanie wstępne przy niskich częstotliwościach, takie jak kodowanie brudnego papieru (DPC), może sterować zarówno fazą, jak i amplitudą sygnału, aby wyeliminować zakłócenia i osiągnąć optymalną wydajność. Jednak cyfrowe kodowanie wstępne wymaga dedykowanego pasma podstawowego i łańcucha RF dla każdego elementu anteny, co jest obecnie kosztowne i energochłonne. Jeśli cyfrowe prekodowanie jest bezpośrednio stosowane w masywnych systemach MIMO mmWave z dużą liczbą anten (np. setki lub nawet tysiące), związana z tym ogromna liczba łańcuchów RF spowoduje wysokie koszty i zużycie energii. Na przykład zużycie energii przez jeden łańcuch RF (w tym konwerter cyfrowo-analogowy, konwerter w górę itp.) Przy mmWave wynosi około 250 mW [1], a dla masywnego systemu MIMO mmWave z 64 antenami, 16 zostanie zużytych tylko przez łańcuchy RF, nie wspominając o przesyłanej energii. Aby rozwiązać ten problem, proponuje się hybrydowe, analogowe i cyfrowe kodowanie wstępne. W szczególności dzieli optymalny cyfrowy prekoder na dwa etapy. W pierwszym kroku niewielkich rozmiarów cyfrowy prekoder (realizowany przez niewielką liczbę łańcuchów RF) jest używany do eliminacji zakłóceń, podczas gdy w drugim etapie duży analogowy formator wiązki (realizowany przez dużą liczbę analogowych przesuwników fazowych bez łańcuchów RF).) służy do zwiększania wzmocnienia tablicy antenowej. Dlatego wstępne kodowanie hybrydowe może znacznie zmniejszyć liczbę wymaganych łańcuchów RF bez oczywistych strat wydajności dzięki starannej konstrukcji, co czyni ją obiecującą technologią wstępnego kodowania dla systemów MIMO typu mmWave. Pozostała część tego rozdziału jest zorganizowana w następujący sposób. Rozdział 5.2 w skrócie przedstawia model kanałowy mmWave Masywny MIMO. Następnie wprowadzane są tradycyjne schematy, w tym cyfrowe kodowanie wstępne i analogowe formowanie wiązki. Wreszcie hybryda cyfrowa i omówione jest wstępne kodowanie analogowe, które jest bardziej odpowiednie dla masywnych systemów MIMO mmWave.

KANAŁOWY MODEL DLA mmWave MASSIVE MIMO

Wysoka tłumienność ścieżki w wolnej przestrzeni jest charakterystyczna dla propagacji fal milimetrowych i prowadzi do ograniczonej przestrzennej selektywności lub rozpraszania. Z drugiej strony, duże, ciasno upakowane tablice antenowe są charakterystyką transceiverów mmWave, co prowadzi do wysokich poziomów korelacji anten. Ta cecha ciasno upakowanych macierzy w środowiskach o rzadkim rozproszeniu sprawia, że wiele statystycznych rozkładów zanikania używanych w tradycyjnej analizie MIMO jest niedokładnych w modelowaniu kanałów mmWave. Z tego powodu przyjmujemy wąskopasmową, klastrowaną reprezentację kanałów, opartą na rozszerzonym modelu Saleha-Valenzueli, która pozwala na dokładne uchwycenie charakterystyk w kanałach mmWave [5,6]. Używając klastrowego modelu kanału, zakłada się, że macierz kanału H jest sumą wkładów ścieżek propagacji L. Dlatego dyskretny kanał wąskopasmowy H można zapisać jako

$$\mathbf{H}=\sqrt{\frac{N_t N_r}{L}}\sum_{l=1}^{L}\alpha_l \Lambda_r\left(\phi_l^r,\theta_l^r\right)\Lambda_t\left(\phi_l^t,\theta_l^t\right)\mathbf{a}_r\left(\phi_l^r,\theta_l^r\right)\mathbf{a}_t^H\left(\phi_l^t,\theta_l^t\right), \tag{1}$$

gdzie αl jest złożonym wzmocnieniem ścieżki l, podczas gdy ϕ^r_l (θ^r_l) i ϕ^t_l (θ^t_l) są odpowiednio jego azymutowymi (elewacyjnymi) kątami nadejścia i odlotu (AoAs / AoDs) odpowiednio. Funkcje $\Lambda_r(\phi^r_l, \theta^r_l)$ i $\Lambda_t(\phi^t_l \theta^t_l)$ reprezentują zysk elementu anteny odbiornika i nadajnika przy odpowiednich AoAs / AoDs. Dla uproszczenia, ale bez utraty ogólności $\Lambda_r(\phi^r_l, \theta^r_l)$ i $\Lambda_r(\phi^t_l, \theta^t_l)$ można ustawić jako jedne z zakresu AoDs / AoAs. Wreszcie wektory $a_r(\phi^r_l, \theta^r_l)$ i a_t (ϕ^t_l, θ^t_l) reprezentują znormalizowane wektory odpowiedzi układu odbiornika i nadajnika przy kącie azymutu (elewacji) (ϕ^r_l, θ^r_l) i (ϕ^t_l, θ^t_l), odpowiednio, które zależą od struktury układu anten nadajnika i odbiornika. Dla jednorodnej macierzy liniowej (ULA) z N elementami wektor odpowiedzi macierzy można przedstawić jako

$$\mathbf{a}_{\mathrm{ULA}}(\phi) = \frac{1}{\sqrt{N}}\left[1, e^{j\frac{2\pi}{\lambda}d\sin(\phi)}, \ldots, e^{j(N-1)\frac{2\pi}{\lambda}d\sin(\phi)}\right]^T, \quad (2)$$

gdzie λ oznacza długość fali sygnału, a d jest odstępem międzyprzedmiotowym. Zauważ, że nie uwzględniamy θ w argumentach a_{ULA}, ponieważ odpowiedź tablicy jest niezmienna w dziedzinie podniesienia. W przypadku jednorodnej tablicy planarnej (UPA) z elementami W_1 i W_2 ($W_1W_2 = N$) odpowiednio na poziomie i pionie, wektor odpowiedzi tablicy można podać jako

$$\mathbf{a}_{\mathrm{UPA}}(\phi, \theta) = \frac{1}{\sqrt{N}}\left[1, \ldots, e^{j\frac{2\pi}{\lambda}d(x\sin(\phi)\sin(\theta)+y\cos(\theta))}, \ldots, e^{j\frac{2\pi}{\lambda}d((W_1-1)\sin(\phi)\sin(\theta)+(W_2-1)\cos(\theta))}\right]^T, \quad (3)$$

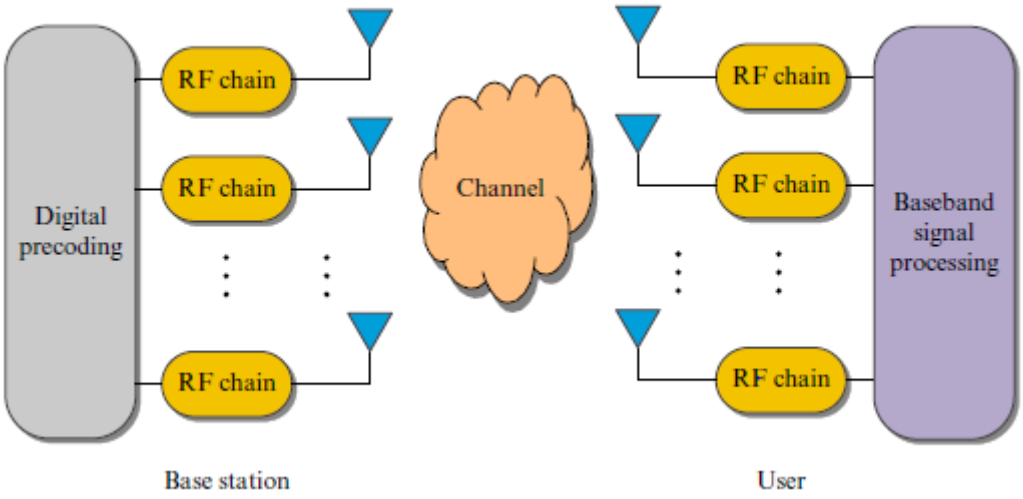
gdzie $0 \le x \le W_1 - 1$ i $0 \le y \le W_2 - 1$. Biorąc pod uwagę UPA są interesujące mmWave beamforming, ponieważ (1) dają mniejsze wymiary tablicy antenowej; (2) ułatwiają pakowanie większej liczby elementów a tenna w szyku o rozsądnych rozmiarach; i (3) umożliwiają formowanie wiązki w domenie elewacji (znane również jako formowanie wiązki 3D).

CYFROWE KODOWANIE WSTĘPNE

Cyfrowe kodowanie wstępne jest tradycyjną technologią szeroko stosowaną w systemach MIMO niskiej częstotliwości .Podstawową ideą cyfrowego kodowania wstępnego jest kontrolowanie zarówno faz, jak i amplitud oryginalnych sygnałów, aby z wyprzedzeniem wyeliminować zakłócenia. Ogólnie mówiąc, wstępne kodowanie cyfrowe można podzielić na dwie kategorie: wstępne kodowanie liniowe, w którym przesyłane sygnały składają się z liniowej kombinacji oryginalnych sygnałów; i nieliniowe kodowanie wstępne, gdzie przesyłane sygnały są otrzymywane w sposób nieliniowy. Poza tym cyfrowe kodowanie wstępne można również podzielić na wstępne kodowanie dla pojedynczego użytkownika i wstępne kodowanie dla wielu użytkowników. W tej sekcji skupimy się najpierw na systemach jednego użytkownika i omówimy kilka prostych liniowych schematów kodowania cyfrowego, takich jak wstępne kodowanie z filtrem dopasowanym (MF) i wymuszeniem zerowym (ZF). Następnie rozważane są systemy wielożytkownikowe i zostanie wprowadzone klasyczne nieliniowe kodowanie diagonalizacji bloków (BD).

CYFROWE KODOWANIE WSTĘPNE DLA JEDNEGO UŻYTKOWNIKA

W tej podsekcji skupiamy się głównie na prostym liniowym kodowaniu cyfrowym. Rozważmy mmWave massive MIMO dla pojedynczego użytkownika z cyfrowym kodowaniem wstępnym, jak pokazano na rysunku, w którym stacja bazowa (BS) wykorzystuje anteny N_T do jednoczesnego przesyłania strumieni danych do użytkownika z liczbą anten ($N_r < N_T$).



Stacja BS stosuje cyfrowy prekoder D , N_t x N_r wykorzystujący swoje łańcuchy N_t RF, a transmitowany sygnał może być prezentowany przez

x= Ds. (4)

gdzie s jest pierwotnym wektorem sygnału N_r x 1 przed wstępnym kodowaniem ze znormalizowaną mocą

jako $\mathbb{E}(\mathbf{ss}^H) = (1/N_t)\mathbf{I}_{N_t}$. Zauważ, że aby spełnić ograniczenie całkowitej mocy nadawania, D.

powinien również spełniać $||D||_F^2 = tr(DD^H) = Nr$

W systemie wąskopasmowym odebrany wektor sygnału y o rozmiarze N_r x 1 może

być odpowiednio przedstawione jako

$$\mathbf{y} = \sqrt{\rho}\mathbf{HDs} + \mathbf{n}, \quad (5)$$

gdzie H jest macierzą kanału N_r x N_t ze znormalizowaną mocą $\mathbb{E}\left(\|\mathbf{H}\|_F^2\right) = N_t N_r$, ρ oznacza średnią odbieraną moc, a n jest addytywnym wektorem białego szumu gaussowskiego (AWGN), którego wpisy są zgodne z niezależnym i identycznym rozkładem (i.i.d.) CN(0, σ^2_n) . Ponadto zakładamy, że macierz kanału H jest doskonale znana w BS, aby umożliwić wstępne kodowanie. Najprostszym liniowym cyfrowym kodowaniem wstępnym jest wstępne kodowanie MF [10], które można przedstawić jako

$$\mathbf{D} = \sqrt{\frac{N_t}{\mathrm{tr}(\mathbf{FF}^H)}}\mathbf{F}, \quad \mathbf{F} = \mathbf{H}^H. \quad (6)$$

Wstępne kodowanie MF może maksymalizować odbierany stosunek sygnału do szumu (SNR) po stronie użytkownika. Jednak zwykle wiąże się to z poważnymi zakłóceniami między różnymi strumieniami danych. W tym celu proponuje się dobrze znane wstępne kodowanie ZF . Odpowiedni cyfrowy prekoder D może być przedstawiony jako

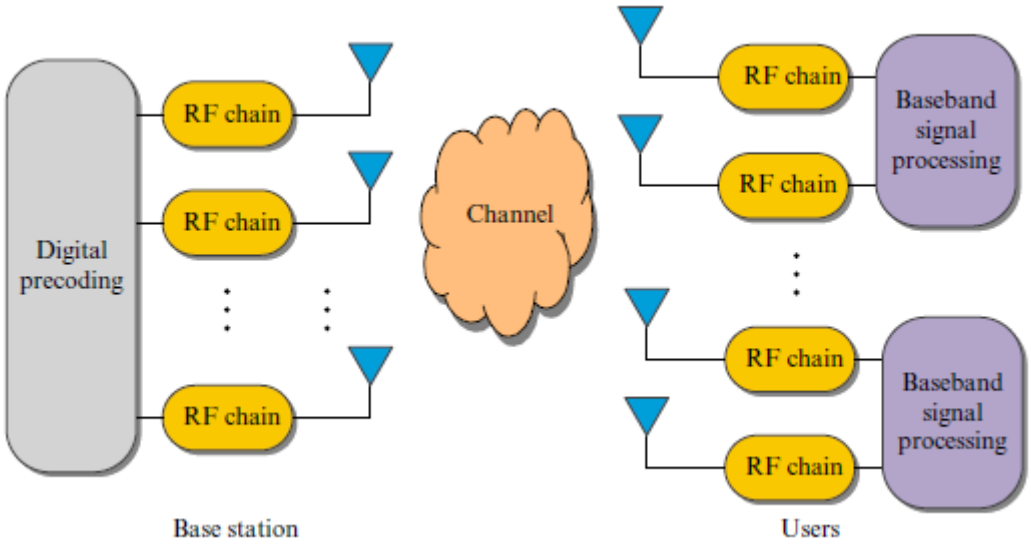
$$\mathbf{D} = \sqrt{\frac{N_t}{\mathrm{tr}(\mathbf{F}\mathbf{F}^H)}}\mathbf{F},$$
$$\mathbf{F} = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}, \quad (7)$$

co może całkowicie wyeliminować zakłócenia między różnymi strumieniami danych. Jednakże, ponieważ cyfrowy prekoder D jest wymagany do spełnienia ograniczenia całkowitej mocy nadajnika, wstępne kodowanie ZF może zwiększyć moc szumów, prowadząc do pewnej utraty wydajności w porównaniu z idealną pojemnością kanału. Aby uzyskać lepszy kompromis między odebranym SNR a zakłóceniami, proponuje się wstępne kodowanie filtru Wienera (WF) , które można przedstawić jako

$$\mathbf{D} = \sqrt{\frac{N_t}{\mathrm{tr}(\mathbf{F}\mathbf{F}^H)}}\mathbf{F},$$
$$\mathbf{F} = \mathbf{H}^H\left(\mathbf{H}\mathbf{H}^H + \frac{\sigma_n^2 N_t}{\rho}\mathbf{I}\right)^{-1}. \quad (8)$$

CYFROWE KODOWANIE WSTĘPNE DLA WIELU UŻYTKOWNIKÓW

Rozważmy wieloużytkownikowy system mmWave, jak pokazano poniżej, gdzie stacja BS wyposaża anteny N_{BS} i łańcuchy RF N_{BS} do jednoczesnej komunikacji ze stacjami ruchomymi U (MSs).



Każde MS wyposaża anteny N_{MS}. Całkowita liczba strumieni danych do komunikacji to $N_{MS}U$ ($N_{MS}U \leq N_{BS}$). W łączu w dół stacja BS wykorzystuje cyfrowy prekoder N_{BS} x $N_{MS}U$ w paśmie podstawowym $D=[D_1, D_2,\ldots, D_U]$ gdzie D_u o rozmiarze N_{BS} x N_{MS} jest cyfrowym prekoderem dla u-tego użytkownika. Ponadto D_u powinien spełniać ograniczenie całkowitej mocy nadawania $||D_u||F = N_{MS}$. Rozważmy wąskopasmowy model kanału zanikającego blokami .Odebrany wektor sygnału r_u obserwowany przez u-ty MS można przedstawić za pomocą

$$\mathbf{r}_u = \mathbf{H}_u \sum_{n=1}^{U} \mathbf{D}_n \mathbf{s}_n + \mathbf{n}_u, \quad (9)$$

gdzie s_n o rozmiarze N_{Mx} x 1 jest oryginalnym wektorem sygnału przed wstępnym kodowaniem ze znormalizowaną mocą, Hu o rozmiarze N_{MS} x N_{BS} oznacza kanał mmWave massive MIMO między BS a u-tym MS, a n_u jest wektorem AWGN, którego wpisy następują po niezależnym i identycznym rozkładzie (iid) CN(0, σ^2_n)

Można to zaobserwować z równania. (9), że terminy $H_uD_ns_n$ dla n ≠ u są interferencjami dla u-tego MS. Dlatego prostym pomysłem jest zaprojektowanie całego D_n tak, aby spełniał $H_uD_n = 0$, co nazywa się kodowaniem wstępnym BD . W szczególności możemy najpierw zdefiniować macierz $\overline{\mathbf{H}}_u$ jako

$$\overline{\mathbf{H}}_u = \left[\mathbf{H}_1^H, \ldots, \mathbf{H}_{u-1}^H, \mathbf{H}_{u+1}^H, \ldots, \mathbf{H}_U^H\right]^H \quad (10)$$

a D_n powinien leżeć w pustej przestrzeni $\overline{\mathbf{H}}_u$. Aby to zrobić, musimy obliczyć rozkład według wartości osobliwej (SVD) $\overline{\mathbf{H}}_u$ jako

$$\overline{\mathbf{H}}_u = \overline{\mathbf{U}}_u \overline{\mathbf{\Lambda}}_u \overline{\mathbf{V}}_u^H = \overline{\mathbf{U}}_u \overline{\mathbf{\Lambda}}_u \left[\overline{\mathbf{V}}_u^{\text{nonzero}}, \overline{\mathbf{V}}_u^{\text{zero}}\right]^H, \quad (11)$$

gdzie $\overline{\mathbf{V}}_u^{\text{nonzero}}$ i $\overline{\mathbf{V}}_u^{\text{zero}}$ składają się z prawych wektorów osobliwych, które odpowiadają niezerowym wartościom osobliwym i zerowym wartościom osobliwym $\overline{\mathbf{H}}_u$, odpowiednio. Następnie cyfrowy prekoder Du dla u-tego MS może być zaprojektowany tak, aby zawierał pierwsze kolumny N_{MS} $\overline{\mathbf{V}}_u^{\text{zero}}$, czyli $\bar{\mathbf{D}}_u = \overline{\mathbf{V}}_u^{\text{zero}}(:, 1:N_{\text{MS}})$.

PODSUMOWANIE WSTĘPNEGO KODOWANIA CYFROWEGO

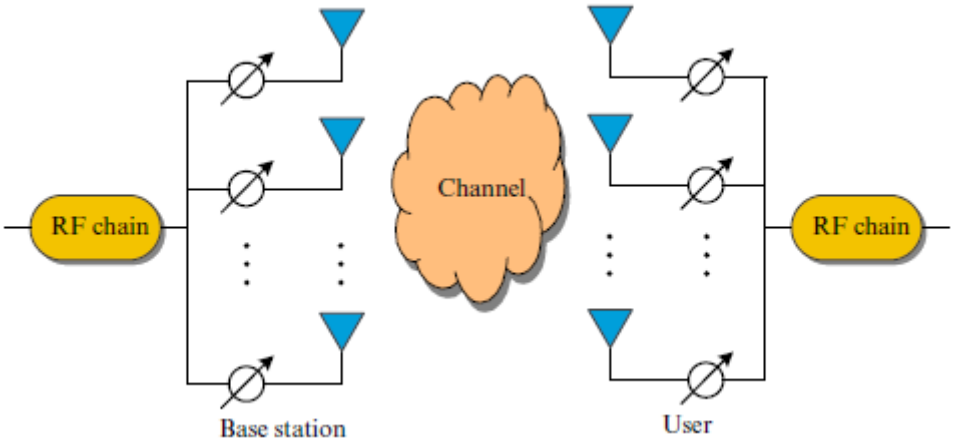
Oprócz schematów wprowadzonych wcześniej, istnieją również doskonałe schematy cyfrowego kodowania wstępnego, takie jak optymalne DPC i prawie optymalne wstępne kodowanie Tomlinsona-Harashimy (TH) .Jednak schematy te zwykle wiążą się z dużą złożonością obliczeniową. Dlatego wspominamy o nich tylko tutaj, bez dalszych wyjaśnień dyskusja. Ogólnie rzecz biorąc, cyfrowe kodowanie wstępne zwykle zapewnia satysfakcjonującą wydajność, ponieważ umożliwia sterowanie zarówno fazami, jak i amplitudami sygnałów. Niestety, wstępne kodowanie cyfrowe wiąże się z wysokim zużyciem energii i kosztami sprzętu ze względu na jeden dedykowany Dla każdej anteny wymagany jest łańcuch RF, co utrudnia bezpośrednie rozszerzenie do masywnych systemów MIMO mmWave z dużą liczbą anten.

ANALOGOWE KSZTAŁTOWANIE ŚWIATŁA

Analogowe formowanie wiązki jest rozwijane w systemach typu punkt-punkt mmWave z dużymi antenami. W takim systemie tylko jeden łańcuch RF jest używany do transmisji pojedynczego strumienia danych, a analogowe formowanie wiązki jest wykorzystywane do sterowania fazami oryginalnych sygnałów w celu uzyskania maksymalnego wzmocnienia tablicy antenowej i efektywnego SNR. W tej sekcji najpierw przedstawimy szeroko stosowany schemat analogowego kształtowania wiązki zwany sterowaniem wiązką, w którym analogowe wektory kształtowania wiązki są ograniczone do wektorów odpowiedzi macierzy, takich jak równania. (2) i (3). Następnie skupimy się na bardziej praktycznym systemie, w którym doskonała informacja o stanie kanału (CSI) nie jest dostępna, i przedstawimy kilka schematów uczenia wiązki w celu uzyskania najlepszych analogowych wektorów kształtujących wiązkę bez znajomości macierzy kanału.

STEROWANIE WIĄZKĄ

Rozważmy mmWave massive MIMO dla pojedynczego użytkownika z analogowym kształtowaniem wiązki, jak pokazano na rysunku w którym stacja BS wykorzystuje anteny N_t, ale tylko jeden łańcuch RF do przesyłania jednego strumienia danych do użytkownika z liczbą anten i jednym łańcuchem RF.



Zdefiniuj f o rozmiarze N_t x 1 jako analogowy wektor kształtujący wiązkę w BS i w o rozmiarze N_r x 1 jako analogowy wektor łączący użytkownika. Naszym celem jest zaprojektowanie f i w, aby zmaksymalizować efektywny SNR, który może być przedstawiony przez

$$(\mathbf{w}^{\text{opt}}, \mathbf{f}^{\text{opt}}) = \arg\max \left|\mathbf{w}^H \mathbf{H} \mathbf{f}\right|^2$$
$$\text{s.t.} \quad w_i = \sqrt{N_r^{-1}} e^{j\varphi_i}, \quad \forall i,$$
$$f_l = \sqrt{N_t^{-1}} e^{j\phi_l}, \quad \forall l. \qquad (12)$$

Zdefiniuj SVD kanału jako $H = U\Sigma V^H$. Wtedy wiemy, że optymalne rozwiązania nieograniczające równania. (12) powinno być $w^{opt} = U(:, 1)$ i $f^{opt} = V(:, 1)$. Jednak nie przestrzegają ograniczenia amplitudy w równaniu. (12). Zgodnie z naszą najlepszą wiedzą nie są dostępne żadne wyniki analityczne dla równania. (12) ogólnie, a algorytmy formowania wiązki są ograniczone do wyszukiwania w ustalonej liczbie predefiniowanych kierunków. Możliwą metodą rozwiązania tego problemu jest zaprojektowanie praktycznych rozwiązań f i w spełniających ograniczenie amplitudy tak, aby było jak najbliższe optymalnym rozwiązaniom nieograniczonym f^{opt} i wopt. Twierdzenie 1 pokazuje, że w przypadku systemów mmWave z dużymi antenami taki projekt można znacznie uprościć.

Twierdzenie 1. Każdy prawy wektor osobliwy kanału macierzy H danego przez model równania (1) z L =o(N_t) i L=o(N_r) zbiega się w odległości cięciwy do wektora odpowiedzi tablicy $a_t(\phi^t_l, \theta^t_l)$ Każdy lewy wektor osobliwy podobnie zbiega się do $a_r(\phi^r_l, \theta^r_l)$. Osobliwe wartości z kolei zbiegają się do $N_t N_r |\alpha_l|^2 /L$.

Twierdzenie 1 stwierdza, że reprezentacja kanału „zbiega się” do jego SVD, w wyniku czego optymalny wektor formowania wiązki SVD jest w rzeczywistości podany w postaci zamkniętej i równy wektorowi odpowiedzi macierzy w najsilniejszym kierunku. Dlatego w przypadku układów mmWave z dużymi N_t i N_r wektor kształtujący wiązkę f i wektor łączący w można po prostu ograniczyć do wektorów odpowiedzi macierzy, takich jak równania (2),(3), które są wystarczająco bliskie optymalnym rozwiązaniom nieograniczającym. Innymi słowy, możemy wybrać $\mathbf{f} = \mathbf{a}_t(\phi^t_{k^*}, \theta^t_{k^*})$ i $\mathbf{w} = \mathbf{a}_r(\phi^r_{k^*}, \theta^r_{k^*})$, gdzie $k^* = \arg\ \max_l |a_l|^2$, aby skierować wiązkę w najsilniejszym kierunku, w którym oczekuje się osiągnięcia prawie optymalnej wydajności, szczególnie dla dużych N_t i N_r.

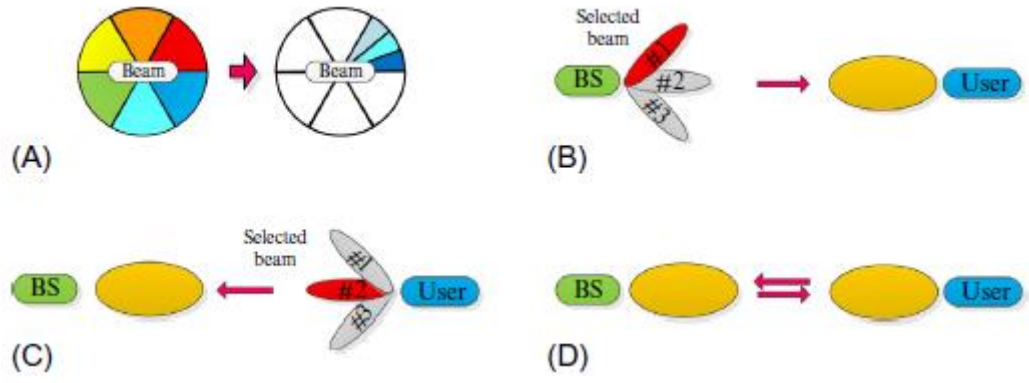
SZKOLENIE Z WIĄZKI

Kiedy mówimy o sterowaniu wiązką, zakładamy, że idealny CSI może zostać osiągnięty zarówno przez BS, jak i użytkownika, co jest niepraktyczne w realistycznych systemach. W rzeczywistości, dla systemów mmWave z tylko jednym łańcuchem RF, BS lub użytkownik nie mogą bezpośrednio obserwować matrycy kanału; raczej obserwuje hałaśliwą wersję efektywnego kanału mniejszego rozmiaru. Innymi słowy, będzie cierpieć z powodu ograniczenia próbkowania podprzestrzeni. Takie ograniczenie oznacza, że tradycyjne algorytmy estymacji kanału, które zostaną omówione później, nie mogą być bezpośrednio zastosowane, a także praktycznie nie jest możliwe oszacowanie wszystkich elementów macierzy kanału, nie wspominając o doskonałym CSI. Bez pełnego CSI problem przekształca się teraz w ogólny problem próbkowania podprzestrzennego w celu uczenia wiązki. Stacja BS i użytkownik muszą współpracować w celu wyszukania najlepszej pary formatora wiązki i łącznika z predefiniowanych książek kodowych podczas uczenia wiązki. Zgodnie z dyskusją wiemy, że dziennik można zaprojektować w oparciu o schemat sterowania wiązką, który można przedstawić jako

$$\mathbf{f} \in \mathcal{F} = \left\{ \mathbf{a}_t\left(\overline{\phi}_1^t, \overline{\theta}_1^t\right), \mathbf{a}_t\left(\overline{\phi}_2^t, \overline{\theta}_2^t\right), \ldots, \mathbf{a}_t\left(\overline{\phi}_{|\mathcal{F}|}^t, \overline{\theta}_{|\mathcal{F}|}^t\right) \right\},$$

$$\mathbf{w} \in \mathcal{W} = \left\{ \mathbf{a}_r(\overline{\phi}_1^r, \overline{\theta}_1^r), \mathbf{a}_t(\overline{\phi}_2^r, \overline{\theta}_2^r), \ldots, \mathbf{a}_t\left(\overline{\phi}_{|\mathcal{W}|}^r, \overline{\theta}_{|\mathcal{W}|}^r\right) \right\}, \quad (13, 14)$$

gdzie $\overline{\phi}_i^t\,(\overline{\theta}_i^t)$ i $\overline{\phi}_l^r\,(\overline{\theta}_l^r)$ są odpowiednio skwantyfikowanym azymutem (wzniesieniem) kątem odlotu i przylotu, przy czym zakłada się, że równomiernie pokrywa cały zakres AoD / AoAs. Najbardziej intuicyjnym, ale optymalnym schematem uczenia wiązki jest wyczerpujące przeszukiwanie wszystkich możliwych par |Fj||W| formowania wiązki i łączenia wektorów w oparciu o kryterium maksymalizacji efektywnego SNR (12). Jednakże, w systemach mmWave, potencjalnie duża liczba anten i wymóg znacznego wzmocnienia kształtowania wiązki będą wymagały, aby rozmiary książki kodowej |F| i|W| były bardzo duże, co oznacza, że wyczerpujące wyszukiwanie może wiązać się z nadmiernym narzutem. Aby rozwiązać ten problem, proponuje się hierarchiczny schemat uczenia wiązki. Najpierw konstruujemy serię książek kodowych F_1, F_2,…, F_K (W_1, W_2,…, W_K) o rosnącej rozdzielczości, jak pokazano na rysunku A



Następnie BS i użytkownik wspólnie szkolą wiązki na pierwszym poziomie (książka kodowa F1 o najniższej rozdzielczości) przez wysyłanie danych szkoleniowych, które składają się z trzech etapów: (1) stacja BS wysyła dane uczące za pomocą jednego możliwego wektora kształtowania wiązki z książki kodowej F_1 do użytkownik, a użytkownik może określić najlepszy wektor łączący zgodnie z równaniem. (12) (Rysunek B); (2) użytkownik i BS zamieniają się rolami i określają najlepszy wektor formowania wiązki w podobny sposób (Rysunek C); i (3) zwracają do siebie indeks wybranego wektora formującego

wiązkę (Rysunek D). Taka procedura opisana wcześniej będzie powtarzana z książką kodową o wyższej rozdzielczości w ramach wybranej wiązki, aż do rozpatrzenia ostatniego poziomu (książki kodowej F_K o najwyższej rozdzielczości). W ten sposób hierarchiczne uczenie wiązki może skutecznie zmniejszyć narzut w porównaniu z wyszukiwaniem wyczerpującym. Warto też zaznaczyć, że za każdym razem, gdy przechodzimy na kolejny poziom, możemy korzystać z krótszych sekwencji treningowych ze względu na dodatkowe wzmocnienie macierzy.

PODSUMOWANIE SYGNAŁU ANALOGOWEGO

Oprócz wprowadzonych analogowych schematów kształtowania i uczenia wiązki, istnieje również kilka powiązanych prac. Na przykład projektuje nową książkę kodów dla analogowego kształtowania wiązki, która może osiągnąć lepsze wyniki niż książka kodów sterująca wiązką. Zaproponowano jednostronny schemat uczenia wiązki w celu znalezienia optymalnej wiązki w dwóch etapach: sumator jest ustalony tak, aby najpierw wyczerpująco wyszukiwał najlepszy prekoder, a następnie ten najlepszy prekoder jest ustalany w celu wyczerpującego wyszukiwania najlepszego sumatora. Schemat ten został przyjęty w standardzie IEEE 802.11ad. Podsumowując, analogowe formowanie wiązki wymaga tylko jednego łańcucha RF, co ułatwia jego implementację. Jednak analogowe formowanie wiązki zwykle cierpi z powodu poważnych strat wydajności, ponieważ można sterować tylko fazami nadawanych sygnałów. Co ważniejsze, analogowe formowanie wiązki jest zwykle stosowane w systemach z jednym strumieniem dla jednego użytkownika, a rozszerzenie na systemy wielostrumieniowe z wieloma użytkownikami nie wydaje się proste.

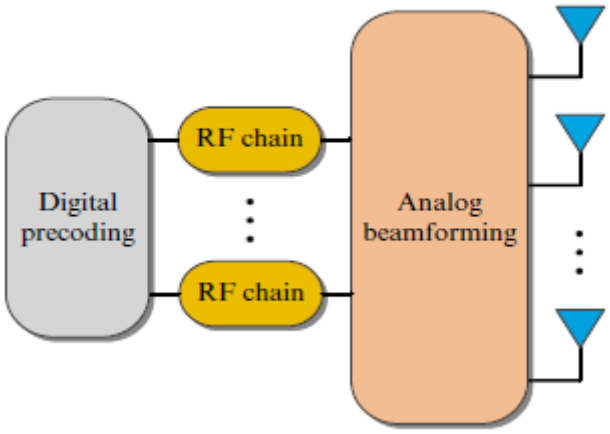
WSTĘPNE KODOWANIE HYBRYDOWE

Jak wspomnieliśmy wcześniej, zarówno cyfrowe prekodowanie, jak i analogowe formowanie wiązki napotykają pewne wyzwania, gdy rozszerzymy je na systemy MIMO mmWave. Aby rozwiązać ten problem, zaproponowano hybrydowe analogowe i cyfrowe kodowanie wstępne. W szczególności dzieli optymalne cyfrowe kodowanie wstępne na dwa etapy. W pierwszym etapie mały cyfrowy prekoder jest używany do eliminacji zakłóceń, podczas gdy w drugim etapie duży analogowy formator wiązki jest używany do zwiększenia wzmocnienia szyku antenowego. Pod względem strukturalnym, hybrydowe formowanie wiązki można podzielić na dwie kategorie: w pełni połączoną architekturę, w której każdy łańcuch RF jest połączony ze wszystkimi antenami BS za pośrednictwem PS; oraz architekturę podpołączoną, w której każdy łańcuch RF jest połączony tylko z podzbiorem anten BS. Poza tym wstępne kodowanie hybrydowe można również podzielić według jednego użytkownika lub wielu użytkowników. W tej sekcji skupimy się najpierw na systemie jednego użytkownika, w którym dwa klasyczne hybrydowe schematy kształtowania wiązki, to znaczy przestrzennie rzadkie hybrydowe kodowanie wstępne (architektura w pełni połączona) [4] i hybrydowe wstępne kodowanie oparte na sukcesywnej eliminacji zakłóceń (SIC). (architektura podporządkowana) [16]. Następnie rozważany jest system wielu użytkowników i badany jest dwuetapowy hybrydowy schemat wstępnego kodowania. Na koniec podsumowujemy wstępne kodowanie hybrydowe i przedstawiamy szczegółowe informacje.

WSTĘPNE KODOWANIE HYBRYDOWE DLA JEDNEGO UŻYTKOWNIKA

Model systemu

Rozważmy masywny system MIMO mmWave dla pojedynczego użytkownika z hybrydowym kodowaniem wstępnym, jak pokazano na rysunku, gdzie stacja BS wykorzystuje anteny N_t do jednoczesnego przesyłania N_s strumieni danych do użytkownika z liczbą anten.



Aby umożliwić transmisję wielostrumieniową, stacja BS jest wyposażona w łańcuchy RF N_t^{RF}, takie jak $N_s \leq N_t^{RF} \leq N_t$. BS stosuje N_t^{RF} x N_s cyfrowy prekoder D wykorzystujący jego łańcuchy RF N_t^{RF}, a następnie N_t x N_t^{RF} analogowy formator wiązki A wykorzystujący obwody analogowe, takie jak przesuwniki fazy (PS). Następnie transmitowany sygnał można przedstawić za pomocą

x=ADs, (15)

gdzie s jest pierwotnym wektorem sygnału N_s x 1 przed wstępnym kodowaniem ze znormalizowaną mocą jako $E(ss^H) = (1/N_s)I_{Ns}$.

Dla uproszczenia rozważamy system wąskopasmowy, który można uznać za rozsądny pierwszy krok , ponieważ szerokość pasma koherencji jest zwykle bardzo duża przy fali mm (rzędu 100 MHz). Odebrany wektor sygnału y o rozmiarze Nr 1 można odpowiednio przedstawić jako

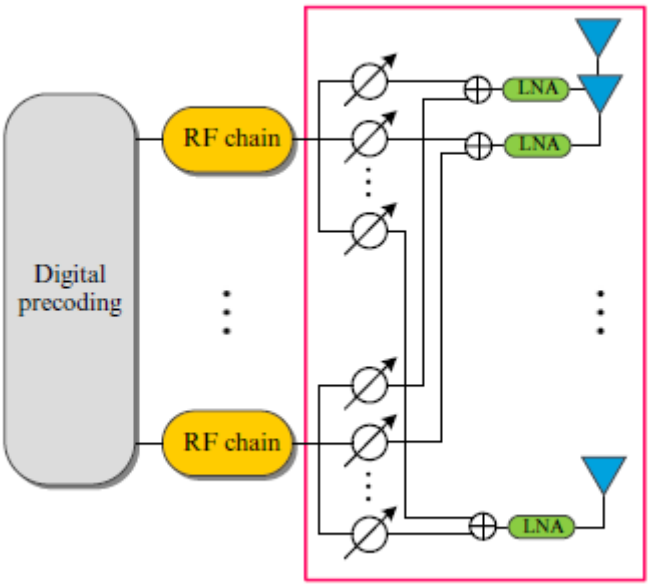
$$\mathbf{y} = \sqrt{\rho}\mathbf{HADs} + \mathbf{n}. \quad (16)$$

Zakładamy, że macierz kanałów jest doskonale znana zarówno w stacji bazowej, jak i użytkownika, aby umożliwić wstępne kodowanie. W praktycznych systemach CSI w odbiorniku można uzyskać poprzez szkolenie, a następnie udostępnić nadajnikowi za pomocą ograniczonej informacji zwrotnej.

Przestrzennie rzadkie wstępne kodowanie hybrydowe (w pełni połączone)

Architektura w pełni połączona

W przypadku architektury w pełni połączonej każdy łańcuch RF jest połączony ze wszystkimi antenami BS za pośrednictwem PS, jak pokazano na rysunku



Ponieważ analogowy kształtownik wiązki A jest realizowany przez analogowe przesuwniki fazy, wszystkie elementy A mają tę samą amplitudę

$$1/\sqrt{N_t}$$

ale różne fazy. Poza tym ograniczenie całkowitej mocy nadawania jest wymuszane przez normalizację D w celu spełnienia wymagań $||AD||^2_F = N_s$. Nie są nakładane żadne dalsze ograniczenia sprzętowe na analogowy beamformer i cyfrowy prekoder dla w pełni połączonej architektury.

Podstawowy pomysł

Staramy się zaprojektować (A, D), aby zmaksymalizować sumaryczną szybkość R (A, D) osiągniętą przez sygnalizację Gaussa w kanale mmWave

$$R(\mathbf{A},\mathbf{D}) = \log_2\left(\left|\mathbf{I} + \frac{\rho}{N_s\sigma_n^2}\mathbf{HADD}^H\mathbf{A}^H\mathbf{H}^H\right|\right). \quad (17)$$

Przy projektowaniu (A, D) odpowiedni problem optymalizacji szybkości sumy można przedstawić jako

$$\begin{aligned}(\mathbf{A}^{\text{opt}},\mathbf{D}^{\text{opt}}) = &\arg\max_{\mathbf{A},\mathbf{D}} R(\mathbf{A},\mathbf{D}),\\ \text{s.t}\ \ &\mathbf{A}\in\mathcal{F},\\ &\|\mathbf{AD}\|_F^2 = N_s,\end{aligned} \quad (18)$$

gdzie F jest zbiorem zawierającym wszystkie możliwe analogowe formatory wiązki, to znaczy zbiorem macierzy N_t x N^{RF}_t z wpisami o stałej wielkości. Wiadomo, że nie ma ogólnych rozwiązań równania (18) w obecności niewypukłego ograniczenia amplitudy A ∈ F. Aby znaleźć praktyczne rozwiązania, które można zaimplementować w systemie z rysunku, proponuje przybliżenie równania (18), który jest łatwiejszy do rozwiązania. Podstawową ideą takiego przybliżenia jest przekształcenie osiągalnej szybkości sumarycznej w „odległość" między praktycznym hybrydowym prekoderem AD a optymalnym nieograniczonym prekoderem P_{opt}. Zacznijmy od zdefiniowania uporządkowanej SVD macierzy kanału

H jako $H=U\Sigma V^H$, gdzie U jest macierzą unitarną N_r x rank(H), Σ jest macierzą diagonalną rank(H) x rank(H) zawierającą wartości osobliwe H w kolejności malejącej, a V jest macierzą unitarną N_t x rank(H). Wykorzystując SVD dla H, możemy przepisać równanie. (18) jak

$$R(\mathbf{A},\mathbf{D}) = \log_2\left(\left|\mathbf{I} + \frac{\rho}{N_s\sigma_n^2}\boldsymbol{\Sigma}^2\mathbf{V}^H\mathbf{A}\mathbf{D}\mathbf{D}^H\mathbf{A}^H\mathbf{V}\right|\right). \quad (19)$$

Ponadto możemy rozłożyć macierze Σ i V jako

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix}, \quad \mathbf{V} = [\mathbf{V}_1 \ \mathbf{V}_2], \quad (20)$$

gdzie Σ_1 ma rozmiar N_s x N_s, a V_1 ma rozmiar N_t x Ns. Możemy zauważyć, że optymalny nieograniczony prekoder P_{opt} jest po prostu dany przez $P_{opt} = V_1$. Niestety, prekodera V_1 nie można wyrazić jako AD z A ∈ F, co oznacza, że nie może być zrealizowany przez architekturę będącą przedmiotem zainteresowania, jak pokazano na rysunku. Jeśli jednak potrafimy zaprojektować praktyczny hybrydowy prekoder AD wystarczająco „blisko" optymalnego nieograniczonego prekodera V_1, można oczekiwać, że osiągnie on prawie optymalną wydajność. Aby zweryfikować ten pomysł, przyjmujemy następujące założenie systemu

Aproksymacja 5.1 Zakładamy, że parametry systemu mmWave (N_t, Nr, N_t^{RF}, N_r^{RF}), a także parametry kanału propagacji mmWave (L, α_l...) są takie, że hybrydowe prekodery AD mogą być dostatecznie „bliskie" optymalnemu unitarnemu prekoderowi $P_{opt} = V_1$. Matematycznie tę „bliskość" definiują następujące dwa równoważne przybliżenia:

(1) Wartości własne macierzy $I - V_l^H ADD^HA^HV_1$ są małe. W przypadku wstępnego kodowania mmWave można to równoważnie określić jako $V_l^H AD \approx I$.

(2) Osobliwe wartości macierzy V_2^HAD są małe; alternatywnie $V_2^H AD \approx 0$.

. Opierając się na przybliżeniu 5.1, sumaryczną stawkę R(A,D) (19) można uprościć jako

$$R(\mathbf{A},\mathbf{D}) = \log_2\left(\left|\mathbf{I} + \frac{\rho}{N_s\sigma_n^2}\boldsymbol{\Sigma}_1^2\right|\right) - \left(N_s - \left\|\mathbf{V}_1^H\mathbf{A}\mathbf{D}\right\|_F^2\right). \quad (21)$$

Obserwujemy z równania. (21), że zależność między sumaryczną szybkością R(A,D) a zaprojektowanym hybrydowym prekoderem AD jest określona tylko przez termin $||V_1^HAD||_F^2$. Zgodnie z przybliżeniem 5.1, termin $||V_1^HAD||_F^2$ można w przybliżeniu zmaksymalizować poprzez maksymalizację $tr(V_1^HAD)$. Poza tym maksymalizacja $tr(V_1^HAD)$ jest ponownie równoważna minimalizacji $||P_{opt} - AD||_F$. Dlatego problem optymalizacji kursu sumarycznego (18) można dobrze przybliżyć za pomocą

$$\begin{aligned} (\mathbf{A}^{\text{opt}}, \mathbf{D}^{\text{opt}}) = \underset{\mathbf{A},\mathbf{D}}{\arg\min} \ & \left\|\mathbf{P}_{\text{opt}} - \mathbf{A}\mathbf{D}\right\|_F, \\ \text{s.t} \ \ & \mathbf{A} \in \mathcal{F}, \\ & \|\mathbf{A}\mathbf{D}\|_F^2 = N_s, \end{aligned} \quad (22)$$

co jest równoważne znalezieniu rzutowania P_{opta} na zbiór hybrydowych prekoderów postaci AD z A ∈ F. Niestety, niepowypukłe ograniczenie zbioru wykonalnego F sprawia, że znalezienie takiego odwzorowania zarówno analitycznie (w postaci zamkniętej), jak i algorytmicznie jest niemożliwe. Aby

uzyskać prawie optymalne i zamknięte rozwiązanie równania (22), musimy nałożyć więcej ograniczeń na F. W szczególności, wykorzystując strukturę macierzy H kanału mmWave massive MIMO, zauważamy, że prawie optymalne hybrydowe prekodery można znaleźć poprzez dalsze ograniczenie F do zbiór wektorów postaci w $a_t(\phi^t_l\ (\theta^t_l)$ i rozwiązywanie

$$\begin{aligned}(\mathbf{A}^{\text{opt}}, \mathbf{D}^{\text{opt}}) &= \arg\min_{\mathbf{A},\mathbf{D}} \left\|\mathbf{P}_{\text{opt}} - \mathbf{AD}\right\|_F, \\ \text{s.t}\ \ &\mathbf{A}(:,i) \in \left\{\mathbf{a}_t\left(\phi_l^t, \theta_l^t\right),\ \ \forall l\right\}, \\ &\|\mathbf{AD}\|_F^2 = N_s,\end{aligned} \tag{23}$$

gdzie A (:, i) przedstawia i-tą kolumnę A. Równanie (23) ma na celu znalezienie najlepszej niskowymiarowej reprezentacji P_{opt} za pomocą wektorów bazowych w $a_t(\phi^t_l\ (\theta^t_l)$. Innymi słowy, równanie. (23) polega na wybraniu „najlepszych" wektorów odpowiedzi macierzy N_t^{RF} i znalezieniu ich optymalnej kombinacji pasma podstawowego. Należy zauważyć, że ograniczenie A (:, i) można bezpośrednio osadzić w funkcji celu optymalizacji, aby uzyskać następujący równoważny problem:

$$\begin{aligned}\widetilde{\mathbf{D}}^{\text{opt}} &= \arg\min_{\widetilde{\mathbf{D}}} \left\|\mathbf{P}_{\text{opt}} - \mathbf{A}_t\widetilde{\mathbf{D}}\right\|_F, \\ \text{s.t}\ \ &\left\|\text{diag}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{D}}^H\right)\right\|_0 = N_t^{\text{RF}}, \\ &\left\|\mathbf{A}_t\widetilde{\mathbf{D}}\right\|_F^2 = N_s,\end{aligned} \tag{24}$$

gdzie $\mathbf{A}_t = \left[\mathbf{a}_t\left(\phi_1^t, \theta_1^t\right), \ldots, \mathbf{a}_t\left(\phi_L^t, \theta_L^t\right)\right]$ jest macierzą N_t x L zawierającą wszystkie wektory odpowiedzi macierzy i $\widetilde{\mathbf{D}}$ jest macierzą L x N_s A_t i $\widetilde{\mathbf{D}}$ są zdefiniowane jako macierze pomocnicze, z którego otrzymujemy odpowiednio A^{opt} i D^{opt}. W szczególności, zgodnie z diagnozy ograniczenia $\left\|\text{diag}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{D}}^H\right)\right\|_0 = N_t^{\text{RF}}$, zakłada się, że $\widetilde{\mathbf{D}}$ nie może mieć więcej niż N_t^{RF} niezerowych wierszy. Dlatego, gdy tylko N_t^{RF} wiersze $\widetilde{\mathbf{D}}$ są niezerowe, tylko kolumny N_t^{RF} macierzy A_t są efektywnie „wybierane". W konsekwencji, D^{opt} prekodera pasma podstawowego można uzyskać przez niezerowe wiersze N_t^{RF} $\widetilde{\mathbf{D}}^{\text{opt}}$ i prekodera A^{opt} mogą być wybrane za pomocą odpowiednich kolumn N_t^{RF} odpowiednio A_t.

Do tej pory problem wspólnego projektowania A i D został przekształcony w problem rekonstrukcji macierzy ograniczonej rzadkością z jedną zmienną. Oznacza to, że teraz do uzyskania rozwiązania równania można zastosować obszerne klasyczne algorytmy rzadkiej rekonstrukcji (24) i zaprojektuj hybrydowy prekoder o niemal optymalnej wydajności. Aby to wyjaśnić, możemy najpierw rozważyć najprostszy przypadek transmisji jednostrumieniowej, to znaczy N_s =1. Następnie równanie (5.24) jest uproszczona do

$$\begin{aligned}\widetilde{\mathbf{d}}^{\text{opt}} &= \arg\min_{\widetilde{\mathbf{d}}} \left\|\mathbf{p}_{\text{opt}} - \mathbf{A}_t\widetilde{\mathbf{d}}\right\|_F, \\ \text{s.t}\ \ &\left\|\widetilde{\mathbf{d}}\right\|_0 = N_t^{\text{RF}},\ \ \left\|\mathbf{A}_t\widetilde{\mathbf{d}}\right\|_F^2 = N_s.\end{aligned}$$

(25)

Równanie (25) odpowiada typowemu problemowi odzyskiwania rzadkiego sygnału, który można rozwiązać, na przykład, rozluźniając ograniczenie rzadkości i stosując optymalizację wypukłą do rozwiązania jego relaksacji l_2 - l_1. W bardziej ogólnym przypadku transmisji wielostrumieniowej, to znaczy $N_s \geq 2$, problem w równaniu (24) odpowiada typowemu problemowi rzadkiego odtwarzania sygnału z wieloma wektorami pomiarowymi, znanemu również jako problem jednoczesnej rzadkiej aproksymacji . Taki problem może rozwiązać dobrze znana koncepcja poszukiwania dopasowania ortogonalnego (OMP) .Algorytm 1 dostarcza pseudokod dla rozwiązania prekodera otrzymanego przez OMP.

ALGORITHM 1 SPATIALLY SPARSE PRECODING

Input: $\mathbf{P}_{\text{opt}}$

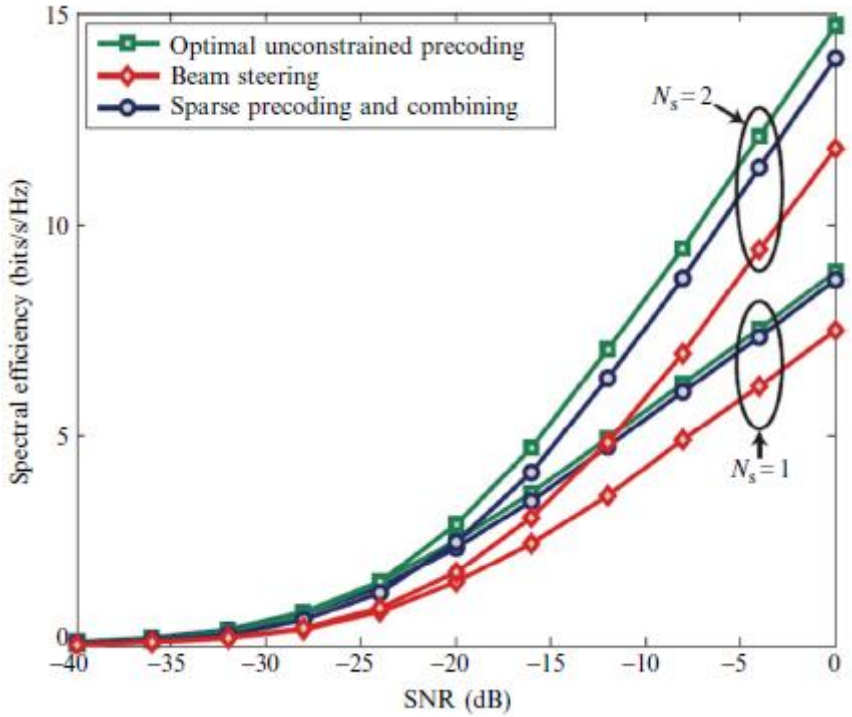
1: $\mathbf{A} = \text{Empty Matrix}$
2: $\mathbf{P}_{\text{res}} = \mathbf{P}_{\text{opt}}$
3: **for** $i \leq N_t^{\text{RF}}$ **do**
4: $\boldsymbol{\Psi} = \mathbf{A}_t^H \mathbf{P}_{\text{res}}$
5: $k = \arg\max_{l=1,\ldots,L} \left(\boldsymbol{\Psi}\boldsymbol{\Psi}^H\right)_{l,l}$
6: $\mathbf{A} = \left[\mathbf{A} \middle| \mathbf{A}_t^{(k)}\right]$
7: $\mathbf{D} = \left(\mathbf{A}^H\mathbf{A}\right)^{-1}\mathbf{A}^H\mathbf{P}_{\text{opt}}$
8: $\mathbf{P}_{\text{res}} = \dfrac{\mathbf{P}_{\text{opt}} - \mathbf{AD}}{\left\|\mathbf{P}_{\text{opt}} - \mathbf{AD}\right\|_F}$
9: **end for**
10: $\mathbf{D} = \sqrt{N_s}\dfrac{\mathbf{D}}{\|\mathbf{AD}\|_F}$
11: **return** $\mathbf{A}, \mathbf{D}$

Podsumowując, po inicjalizacji (kroki 1 i 2), algorytm 1 rozpoczyna od znalezienia wektora w $a_t((\phi^t{}_l, \theta^t{}_l))$, wzdłuż którego optymalny prekoder ma maksymalną projekcję w kroku 5. Następnie w kroku 6 dołącza wybrany wektor kolumnowy w $a_t(\phi^t{}_l, \theta^t{}_l)$ do analogowego formatora wiązki A. Po znalezieniu dominującego wektora i obliczeniu rozwiązania metodą najmniejszych kwadratów dla D w kroku 7, wkład wybranego wektora jest usuwany w kroku 8 i algorytm przechodzi do znalezienia kolumny, wzdłuż której „resztkowa macierz kodowania wstępnego" P_{res} ma największą projekcję. Proces trwa, aż wszystkie wektory prekodowania $N_t{}^{RF}$ zostaną wybrane i wyprowadzają zaprojektowane przez N_t x $N^{RF}{}_t$ analogowe formatory wiązki A i prekoder pasma podstawowego $N^{RF}{}_t$ x N_s, co minimalizuje $||P_{opt} - AD||_F$. Krok 10 zapewnia, że ograniczenie mocy nadawania $||AD||^2{}_F = N_s$ jest dokładnie spełnione.

Ocena wydajności

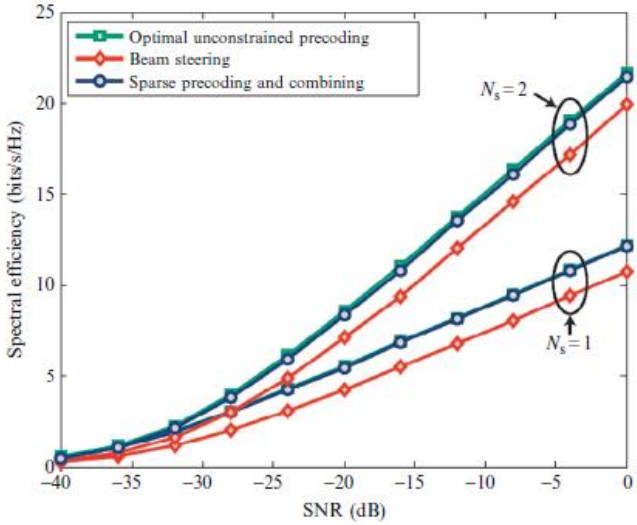
W tej sekcji przedstawiono wyniki symulacji w celu oceny wydajności przestrzennie rzadkiego wstępnego kodowania hybrydowego, w porównaniu z optymalnym nieograniczonym kodowaniem wstępnym $P_{opt} = V_1$ (znanym również jako w pełni cyfrowe kodowanie wstępne) i wstępnym kodowaniem sterującym wiązką, którego kolumny są ściśle związane z wektorami sterującymi macierzy (znane również jako w pełni analogowe formowanie wiązki). Dla kanału przedstawionego w rozdziale 5.2 parametry są ustawione w następujący sposób: Zakładamy, że środowisko propagacji ma L=3 ścieżki z równomiernym rozproszonym azymutem i elewacją AoA / AoDs . Zakłada się, że kąt sektora BS wynosi 60 stopni w dziedzinie azymutu i 20 stopni w elewacji w przypadku zastosowań zewnętrznych. W przeciwieństwie do tego, użytkownik ma stosunkowo mniejsze tablice antenowe z elementów dookólnych, ponieważ użytkownik musi mieć możliwość kierowania wiązkami w dowolnym kierunku, ponieważ ich lokalizacja i orientacja w rzeczywistych systemach jest przypadkowa. Zakłada się, że odstęp między elementami d jest równy połowie długości fali. SNR definiuje się jako $\rho/\sigma_n{}^2$. Rysunek przedstawia osiągalne porównanie szybkości sumarycznej w systemie MIMO o przepustowości 64 x 16 mm z kwadratowymi planarnymi macierzami zarówno w stacji bazowej, jak i u użytkownika. Stacja BS wykorzystuje łańcuchy RF $N^{RF}t$ = 4 do jednoczesnej transmisji N_s = 1 lub 2 strumieni. Na rysunku

obserwujemy, że przestrzennie rzadkie hybrydowe kodowanie wstępne może osiągnąć sumaryczną szybkość całkiem zbliżoną do optymalnego rozwiązania nieograniczonego w przypadku $N_s = 1$, podczas gdy w przypadku $N_s = 2$ jego wydajność mieści się w niewielkiej luce od optymalnego nieograniczonego rozwiązania rozwiązanie.



Oznacza to, że przestrzennie rzadkie wstępne kodowanie hybrydowe może bardzo dokładnie przybliżyć dominujące wektory osobliwe kanału i osiągnąć prawie optymalną wydajność oceniać.

Rysunek poniższy przedstawia osiągalne porównanie współczynnika sumarycznego w masywnym systemie MIMO 256 x 64 mmWave massive MIMO, w którym BS wykorzystuje $N^{RF}_t = 6$ łańcuchów RF, a zarówno BS, jak i użytkownik są wyposażeni w kwadratowe planarne tablice. Na rysunku możemy zaobserwować, że przestrzennie rzadkie hybrydowe kodowanie wstępne może osiągnąć prawie doskonałą wydajność zarówno w przypadkach $N_s = 1$, jak i $N_s = 2$.



Ponadto zauważamy, że chociaż oczekuje się, że sterowanie wiązką zapewni satysfakcjonującą wydajność dla mmWave Massive MIMO z dużymi macierzami, przestrzennie rzadka hybryda kodowania wstępnie nadal przewyższa sterowanie wiązką o ~5 dB.

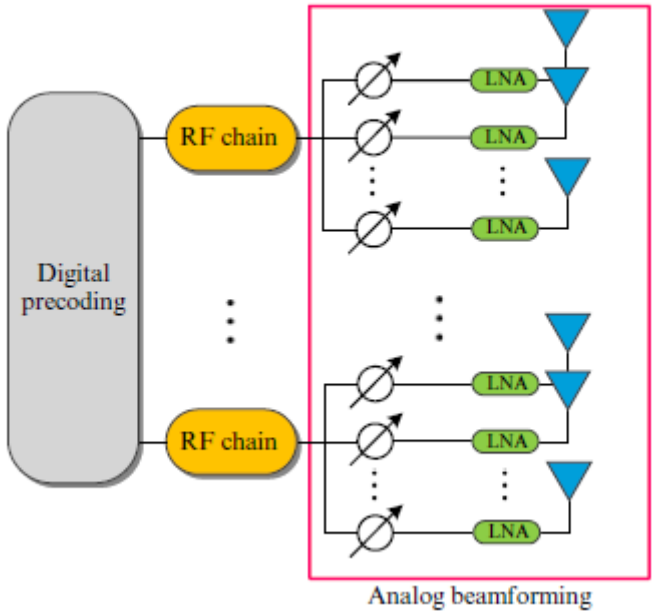
Podsumowanie przestrzennie rzadkiego hybrydowego wstępnego kodowania

Podstawową ideą przestrzennie rzadkiego wstępnego kodowania jest sformułowanie problemu optymalizacji sumarycznej szybkości jako problemu rzadkiego przybliżenia. Następnie można wykorzystać klasyczne algorytmy odtwarzania rzadkich sygnałów, takie jak algorytm OMP. Algorytm ten przyjmuje jako dane wejściowe optymalny nieograniczony prekoder i aproksymuje go jako liniową kombinację wektorów sterujących wiązką, które mogą być zastosowane przez układ analogowy wraz z cyfrowym prekoderem w paśmie podstawowym w celu uzyskania prawie optymalnego praktycznego prekodera hybrydowego. Wyniki symulacji pokazują, że dla ypicalmmWavemassiveMIMOsystems, przestrzennie rzadkie hybrydowe kodowanie wstępne może osiągnąć wydajność bardzo zbliżoną do optymalnego nieograniczonego prekodera, podczas gdy liczba wymaganych łańcuchów RF jest znacznie zmniejszona. W oparciu o przestrzennie rzadkie hybrydowe kodowanie wstępne, zaproponowano kilka rozwiniętych hybrydowych schematów wstępnego kodowania. W ref. [20], zaproponowano wersję przestrzennie rzadkiego hybrydowego wstępnego kodowania o niskiej złożoności. Główne wkłady tej pracy obejmują: (1) wyprowadzenie i integrację algorytmu macierzy-inwersji-obejścia OMP w celu wyeliminowania operacji odwracania macierzy; (2) opracowanie konkretnego algorytmu rekonstrukcji wstępnego kodowania dla implementacji sprzętowej poprzez uwzględnienie właściwości kanału mWave; oraz (3) zaprojektowanie i wdrożenie procesora rekonstrukcji wstępnego kodowania w chipie układu scalonego specyficznego dla aplikacji (ASIC). W ref. [21], rozważany jest masywny system MIMO mmWave z częściową wiedzą o kanale, w którym BS i użytkownik znają tylko własne lokalne kąty nadejścia (AoAs). W takim scenariuszu ref. [21] proponuje odpowiednio zmodyfikowane przestrzennie rzadkie wstępne kodowanie hybrydowe w celu osiągnięcia prawie optymalnej wydajności.

Hybrydowe kodowanie wstępne oparte na SIC (podpołączone)

Architektura podporządkowana

W przypadku architektury podpołączonej każdy łańcuch RF jest połączony tylko z podzbiorem anten BS, jak pokazano na rysunku



W porównaniu z architekturą w pełni połączoną, architektura podporządkowana może znacznie zmniejszyć liczbę wymaganych PS z N_t x N^{RF}_t do N_t, co przyniesie trzy korzyści: (1) może zaoszczędzić energię do wzbudzenia PS; (2) może zaoszczędzić energię, aby zrekompensować tłumienie PS; i (3) wiąże się z mniejszą złożonością obliczeniową ze względu na uproszczoną architekturę sprzętową. Rozważmy system, w którym stacja bazowa ma anteny N_t, ale tylko łańcuchy RF N_t^{RF}. Każdy łańcuch RF jest podłączony do jednej macierzy podantenowej ze stałą i niewielką liczbą anten M, to znaczy $N_t = N^{RF}_t M$. Aby w pełni osiągnąć zysk multipleksowania przestrzennego, stacja BS zwykle przesyła $N_s = N^{RF}_t$ niezależnych strumieni danych do użytkowników korzystających z N_r odbiornika anteny . W przypadku architektury podpołączonej cyfrowy prekoder D o rozmiarze N^{RF}_t x $N_s = N_s$ x N_s można wyspecjalizować jako macierz diagonalną jako $D = diag[d_1, d_2,\ldots, d_N]$, gdzie $d_n \in \mathbb{R}$ dla $n=1,2,\ldots, N_s$. Następnie rola D zasadniczo wykonuje pewną alokację mocy. Ponadto, w przeciwieństwie do architektury w pełni połączonej, analogowy kształtownik wiązki A o rozmiarze N_t x $N^{RF}_t = N_t$ x N_s w architekturze z połączeniem podrzędnym będzie miał specjalną blokową strukturę diagonalną,

$$\mathbf{A} = \begin{bmatrix} \overline{\mathbf{a}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \overline{\mathbf{a}}_2 & & \mathbf{0} \\ \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \overline{\mathbf{a}}_{N_s} \end{bmatrix}_{N_t \times N_s}, \quad (26)$$

gdzie $\overline{\mathbf{a}}_n \in \mathbb{C}^{M\times 1}$ jest analogowym wektorem wagowym dla n-tej podanteny z M antenami, których elementy mają tę samą amplitudę $1/\sqrt{M}$, ale różne fazy.

Podstawowa idea

Naszym celem jest maksymalizacja całkowitej osiągalnej szybkości R(P) systemów mmWave MIMO poprzez zaprojektowanie hybrydowego prekodera P = AD, który można wyrazić jako

$$R(\mathbf{P}) = \log_2\left(\left|\mathbf{I} + \frac{\rho}{N_s\sigma_n^2}\mathbf{H}\mathbf{P}\mathbf{P}^H\mathbf{H}^H\right|\right). \quad (27)$$

Na podstawie poprzedniego opisu hybrydową macierz wstępnego kodowania P można przedstawić jako $\mathbf{P} = \mathbf{AD} = \mathrm{diag}\{\bar{\mathbf{a}}_1, \ldots, \bar{\mathbf{a}}_{N_s}\} \cdot \mathrm{diag}\{d_1, \ldots, d_{N_s}\}$. Dlatego będą istnieć trzy ograniczenia dotyczące projektowania P:

Wiązanie 1: P powinno być blokową macierzą diagonalną podobną do formy A, jak pokazano w równaniu. (26); to znaczy $\mathbf{P} = \mathrm{diag}\{\bar{\mathbf{p}}_1, \ldots, \bar{\mathbf{p}}_{N_s}\}$, gdzie $\bar{\mathbf{p}}_n = d_n \bar{\mathbf{a}}_n$ jest niezerowym wektorem M x 1 z n-tej kolumny $\mathbf{p}_n = \left[\mathbf{0}_{1\times M(n-1)}, \bar{\mathbf{p}}_n^T, \mathbf{0}_{1\times M(N_s-n)}\right]^T$.

Ograniczenie 2: Niezerowe elementy każdej kolumny P powinny mieć taką samą amplitudę, ponieważ cyfrowa macierz kodowania wstępnego D jest macierzą diagonalną, a amplituda niezerowych elementów analogowej macierzy kształtowania wiązki A jest ustalona na 1 /√M.

Ograniczenie 3: Norma Frobeniusa P powinna spełniać $||Pk||_F \le Ns$, aby spełnić ograniczenie całkowitej mocy nadawania, gdzie N_s jest liczbą łańcuchów RF równą liczbie transmitowanych strumieni danych.

Niestety, niepowypukłe ograniczenia 1 i 2 na P sprawiają, że maksymalizacja całkowitej osiągalnej szybkości (27) jest bardzo trudna do rozwiązania. Jednak w oparciu o specjalną blokową strukturę diagonalną hybrydowej macierzy wstępnego kodowania P, możemy zauważyć, że wstępne kodowanie na różnych macierzach podanten jest niezależne. To inspiruje nas do rozłożenia całkowitej osiągalnej szybkości (27) na serię problemów optymalizacji podrzędnej szybkości, z których każdy bierze pod uwagę tylko jedną tablicę podanteny. W szczególności możemy podzielić hybrydową macierz kodowania wstępnego P jako P = $[P_{N_{s-1}}\ P_{N_s}]$, gdzie p_{N_s} to Ns-ta kolumna P, a $P_{N_{s-1}}$ to macierz N_sM x (N_s-1) zawierająca pierwsze (N_s-1) kolumny P. Następnie całkowitą osiągalnąa szybkość R(P) w równaniu (27) można przepisać jako

$$\begin{aligned} R(\mathbf{P}) &\overset{(a)}{=} \log_2(|\mathbf{T}_{N_s-1}|) + \log_2\left(\left|\mathbf{I} + \frac{\rho}{N_s\sigma_n^2}\mathbf{T}_{N_s-1}^{-1}\mathbf{H}\mathbf{p}_{N_s}\mathbf{p}_{N_s}^H\mathbf{H}^H\right|\right) \\ &\overset{(b)}{=} \log_2(|\mathbf{T}_{N_s-1}|) + \log_2\left(1 + \frac{\rho}{N_s\sigma_n^2}\mathbf{p}_{N_s}^H\mathbf{H}^H\mathbf{T}_{N_s-1}^{-1}\mathbf{H}\mathbf{p}_{N_s}\right), \end{aligned} \quad (28)$$

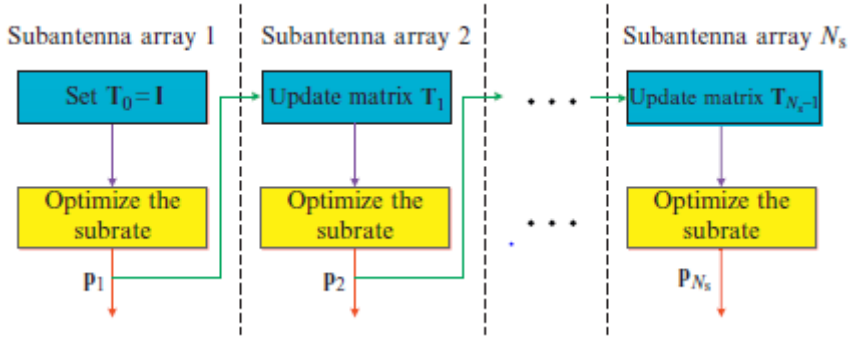
gdzie (a) uzyskuje się przez zdefiniowanie macierzy pomocniczej $\mathbf{T}_{N_s-1} = \mathbf{I} + \frac{\rho}{N_s\sigma_n^2}\mathbf{H}\mathbf{P}_{N_s-1}\mathbf{P}_{N_s-1}^H\mathbf{H}^H$, a (b) jest prawdziwe ze względu na fakt, że $|\mathbf{I}+\mathbf{XY}| = |\mathbf{I}+\mathbf{YX}|$ przez zdefiniowanie $\mathbf{X} = \mathbf{T}_{N_s-1}^{-1}\mathbf{H}\mathbf{p}_{N_s}$ i $\mathbf{Y} = \mathbf{p}_{N_s}^H\mathbf{H}^H$. Zwróć uwagę, że drugi termin $\log_2\left(1 + \frac{\rho}{N_s\sigma_n^2}\mathbf{p}_{N_s}^H\mathbf{H}^H\mathbf{T}_{N_s-1}^{-1}\mathbf{H}\mathbf{p}_{N_s}\right)$ po prawej stronie równania. (28) to osiągalny współczynnik podrzędny tablicy podanteny N_s-tej, podczas gdy pierwszy wyraz $\log_2(TN_{s-1})$ ma taką samą postać jak równanie (27). Ta obserwacja sugeruje, że możemy dalej rozłożyć $\log_2(TN_{s-1})$ przy użyciu podobnej metody w równaniu. (28) jak

$$\log_2(|\mathbf{T}_{N_s-2}|) + \log_2\left(1 + \frac{\rho}{N_s\sigma_n^2}\mathbf{p}_{N_s-1}^H\mathbf{H}^H\mathbf{T}_{N_s-2}^{-1}\mathbf{H}\mathbf{p}_{N_s-1}\right)$$

Następnie, po N takich rozkładów, całkowity osiągalny współczynnik R w równaniu (27) można przedstawić jako

$$R=\sum_{n=1}^{N_s}\log_2\left(1+\frac{\rho}{N_s\sigma_n^2}\mathbf{p}_n^H\mathbf{H}^H\mathbf{T}_{n-1}^{-1}\mathbf{H}\mathbf{p}_n\right), \quad (29)$$

gdzie mamy $\mathbf{T}_n=\mathbf{I}+\frac{\rho}{N_s\sigma_n^2}\mathbf{H}\mathbf{P}_n\mathbf{P}_n^H\mathbf{H}^H$ and $\mathbf{T}_0=\mathbf{I}_{N_s}$. Z równania (29), obserwujemy, że całkowity osiągalny problem optymalizacji szybkości można przekształcić w serię problemów optymalizacji podrzędnej szybkości macierzy podantenowych, które można optymalizować jeden po drugim. Następnie, zainspirowani pomysłem SIC do wykrywania sygnału wielu użytkowników , możemy zoptymalizować osiągalny subrate pierwszej macierzy subantenna i zaktualizować macierz T_1. Następnie podobną metodę można wykorzystać do optymalizacji osiągalnej szybkości podrzędnej drugiej tablicy podantenowej. Taka procedura będzie wykonywana do czasu rozpatrzenia ostatniej tablicy podantennej. Rysunek przedstawia schemat proponowanego wstępnego kodowania hybrydowego opartego na SIC.



Następnie omówimy, jak zoptymalizować osiągalny współczynnik podrzędny każdej tablicy podantenny. Następnie skupimy się na problemie optymalizacji współczynnika podrzędnego n-tej tablicy podantenowej, która może być bezpośrednio zastosowana do innych tablic podantenowych. Zgodnie z równaniem (29), problem optymalizacji szybkości podrzędnej n-tej tablicy podantenowej poprzez zaprojektowanie n-tego wektora prekodującego pn można określić jako

$$\mathbf{p}_n^{\text{opt}}=\arg\max_{\mathbf{p}_n\in\mathcal{F}}\log_2\left(1+\frac{\rho}{N_s\sigma_n^2}\mathbf{p}_n^H\mathbf{G}_{n-1}\mathbf{p}_n\right), \quad (30)$$

gdzie G_{n-1} jest zdefiniowane jako G_{n-1} $=H^HT^{-1}{}_{n-1}H$, F jest zbiorem wszystkich możliwych wektorów spełniających trzy opisane wcześniej ograniczenia. Zauważ, że n-ty wektor kodowania wstępnego p_n ma tylko M niezerowych elementów od (M(n-1) + 1)-tego do (Mn)-tego. Dlatego problem optymalizacji szybkości podrzędnej (30) można równoważnie zapisać jako

$$\overline{\mathbf{p}}_n^{\text{opt}}=\arg\max_{\overline{\mathbf{p}}_n\in\overline{\mathcal{F}}}\log_2\left(1+\frac{\rho}{N_s\sigma_n^2}\overline{\mathbf{p}}_n^H\overline{\mathbf{G}}_{n-1}\overline{\mathbf{p}}_n\right) \quad (31)$$

gdzie $\overline{\mathcal{F}}$ obejmuje wszystkie możliwe wektory M x1 spełniające Wiązanie 2 i Ograniczenie 3, $\overline{\mathbf{G}}_{n-1}$ o rozmiarze M x M jest odpowiednią podmacierzem Gn-1, zachowując jedynie wiersze i kolumny Gn-1 od (M(n-1) + 1)-tej do (Mn)tego, który można przedstawić jako

$$\overline{\mathbf{G}}_{n-1} = \mathbf{R}\mathbf{G}_{n-1}\mathbf{R}^H = \mathbf{R}\mathbf{H}^H\mathbf{T}_{n-1}^{-1}\mathbf{H}\mathbf{R}^H, \quad (32)$$

gdzie R= $0_{MxM(n-1)}$) I_M $0_{MxM(n-1)}$ jest odpowiednią macierzą wyboru. Zdefiniuj SVD macierzy hermitowskiej $\overline{\mathbf{G}}_{n-1}$ jako $\overline{\mathbf{G}}_{n-1} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^H$, gdzie Σ jest macierzą diagonalną M x M zawierającą wartości osobliwe $\overline{\mathbf{G}}_{n-1}$ w kolejności malejącej, a V jest macierzą jednostkową M x M. Wiadomo, że optymalny nieograniczony wektor wstępnego kodowania równania. (31) to pierwsza kolumna v_1 V, to znaczy pierwszy prawy wektor osobliwy z $\overline{\mathbf{G}}_{n-1}$. Jednak zgodnie z ograniczeniami wymienionymi powyżej, nie możemy bezpośrednio wybrać $\overline{\mathbf{p}}_n^{\text{opt}}$ jako v_1, ponieważ elementy v_1 nie podlegają ograniczeniu o tej samej amplitudzie (tj. Ograniczenie 2). Aby znaleźć wykonalne rozwiązanie problemu optymalizacji współczynnika podrzędnego (31), musimy dalej przekształcić równanie. (31) na inną formę, którą daje Twierdzenie 1.

Twierdzenie 1 Problem optymalizacji

$$\overline{\mathbf{p}}_n^{\text{opt}} = \underset{\overline{\mathbf{p}}_n \in \overline{\mathcal{F}}}{\arg\max}\ \log_2\left(1 + \frac{\rho}{N_s\sigma_n^2}\overline{\mathbf{p}}_n^H\overline{\mathbf{G}}_{n-1}\overline{\mathbf{p}}_n\right)$$

jest równoważne z następującym problemem:

$$\overline{\mathbf{p}}_n^{\text{opt}} = \underset{\overline{\mathbf{p}}_n \in \overline{\mathcal{F}}}{\arg\min}\ \|\mathbf{v}_1 - \overline{\mathbf{p}}_n\|_2^2, \quad (33)$$

gdzie v_1 jest pierwszym prawym wektorem osobliwym $\overline{\mathbf{G}}_{n-1}$.

Twierdzenie 1 wskazuje, że możemy znaleźć wykonalny wektor wstępnego kodowania $\overline{\mathbf{p}}_n$, który jest wystarczająco bliski (pod względem odległości euklidesowej) optymalnemu, ale niepraktycznemu wektorem prekodującym v_1, aby zmaksymalizować osiągalny współczynnik podrzędny n-tej tablicy subantenowej. Ponieważ $\overline{\mathbf{p}}_n = d_n\overline{\mathbf{a}}_n$, docelowa wartość $\|\mathbf{v}_1 - \overline{\mathbf{p}}_n\|_2^2$ w równaniu (33) można przepisać jako

$$\|\mathbf{v}_1 - \overline{\mathbf{p}}_n\|_2^2 \overset{(a)}{=} 1 + d_n^2 - 2d_n\mathrm{Re}\left(\mathbf{v}_1^H\overline{\mathbf{a}}_n\right) = \left(d_n - \mathrm{Re}\left(\mathbf{v}_1^H\overline{\mathbf{a}}_n\right)\right)^2 + \left(1 - \left[\mathrm{Re}\left(\mathbf{v}_1^H\overline{\mathbf{a}}_n\right)\right]^2\right), \quad (34)$$

gdzie (a) uzyskuje się na podstawie faktu, że $v^H{}_1v_1 = 1$ i $\overline{\mathbf{a}}_n^H\overline{\mathbf{a}}_n = 1$, ponieważ v_1 jest pierwszą kolumną unitarnej macierzy V, a każdy element $\overline{\mathbf{a}}_n$ ma taką samą amplitudę 1 /√M

Z równania (34), obserwujemy, że odległość między $\overline{\mathbf{p}}_n$ i v_1 składa się z dwóch części. Pierwsza to

$$\left(d_n - \mathrm{Re}\left(\mathbf{v}_1^H\overline{\mathbf{a}}_n\right)\right)^2$$

, które można zminimalizować do zera, wybierając $d_n = \mathrm{Re}\left(\mathbf{v}_1^H\overline{\mathbf{a}}_n\right)$. Druga to $\left(1 - \left[\mathrm{Re}\left(\mathbf{v}_1^H\overline{\mathbf{a}}_n\right)\right]^2\right)$

, co można zminimalizować, maksymalizując $|\mathrm{Re}(\mathbf{v}_1^H \bar{\mathbf{a}}_n)|$. Zauważ, że zarówno $\bar{\mathbf{a}}_n$, jak i v_1 mają stałą moc równą jeden, to znaczy $\mathbf{v}_1^H \mathbf{v}_1 = 1$ and $\bar{\mathbf{a}}_n^H \bar{\mathbf{a}}_n = 1$. Dlatego optymalna $\bar{\mathbf{a}}_n^{\mathrm{opt}}$ maksymalizacji $|\mathrm{Re}(\mathbf{v}_1^H \bar{\mathbf{a}}_n)|$ to

$$\bar{\mathbf{a}}_n^{\mathrm{opt}} = \frac{1}{\sqrt{M}} e^{j\mathrm{angle}(\mathbf{v}_1)}, \tag{35}$$

gdzie kąt (v_1) oznacza wektor fazowy v_1, to znaczy każdy element $\bar{\mathbf{a}}_n^{\mathrm{opt}}$ ma tę samą fazę, co odpowiadający mu element v_1. W związku z tym optymalny wybór opcji d_n^{opt} to

$$d_n^{\mathrm{opt}} = \mathrm{Re}\left(\mathbf{v}_1^H \bar{\mathbf{a}}_n\right) = \frac{1}{\sqrt{M}} \mathrm{Re}\left(\mathbf{v}_1^H e^{j\mathrm{angle}(\mathbf{v}_1)}\right) = \frac{\|\mathbf{v}_1\|_1}{\sqrt{M}}. \tag{36}$$

W oparciu o równania. (35), (36), optymalne rozwiązanie $\bar{\mathbf{p}}_n^{\mathrm{opt}}$ problemu optymalizacji (33) (lub równoważnie (31)) można otrzymać przez

$$\bar{\mathbf{p}}_n^{\mathrm{opt}} = d_n^{\mathrm{opt}} \bar{\mathbf{a}}_n^{\mathrm{opt}} = \frac{1}{M} \|\mathbf{v}_1\|_1 e^{j\mathrm{angle}(\mathbf{v}_1)}. \tag{37}$$

Warto zwrócić uwagę, że v_1 to pierwsza kolumna unitarnej macierzy V, każdy element v_i z v_1 (dla i = 1,…, M) ma amplitudę mniejszą niż jeden. Dlatego mamy $\left\|\bar{\mathbf{p}}_n^{\mathrm{opt}}\right\|_2^2 \le 1$. Zauważ, że dla wszystkich tablic podantenowych optymalne rozwiązanie $\bar{\mathbf{p}}_n^{\mathrm{opt}}$ dla n = 1, 2,…, N_s ma podobną postać. Zatem możemy to wywnioskować

$$\|\mathbf{P}^{\mathrm{opt}}\|_F^2 = \left\|\mathrm{diag}\left\{\bar{\mathbf{p}}_1^{\mathrm{opt}}, \ldots, \bar{\mathbf{p}}_N^{\mathrm{opt}}\right\}\right\|_F^2 \le N_s, \tag{38}$$

co pokazuje, że ograniczenie całkowitej mocy nadawania (Ograniczenie 3) jest spełnione.

Po uzyskaniu $\bar{\mathbf{p}}_n^{\mathrm{opt}}$ n dla n-tej tablicy podanteny, można zaktualizować macierze $\mathbf{T}_n = \mathbf{I} + \frac{\rho}{N_s \sigma_n^2} \mathbf{H} \mathbf{P}_n \mathbf{P}_n^H \mathbf{H}^H$ (29) i $\bar{\mathbf{G}}_n = \mathbf{R}\mathbf{H}^H \mathbf{T}_n^{-1} \mathbf{H}\mathbf{R}^H$ (32). Następnie można ponownie wykorzystać metodę opisaną wcześniej dla n-tej tablicy podanteny ponownie, aby zoptymalizować osiągalny współczynnik podrzędny tablicy (n + 1)- tej subantenny. Podsumujmy, rozwiązując problem optymalizacji szybkości podrzędnej w tablicy n-tej podanteny składa się z następujących trzech kroków:

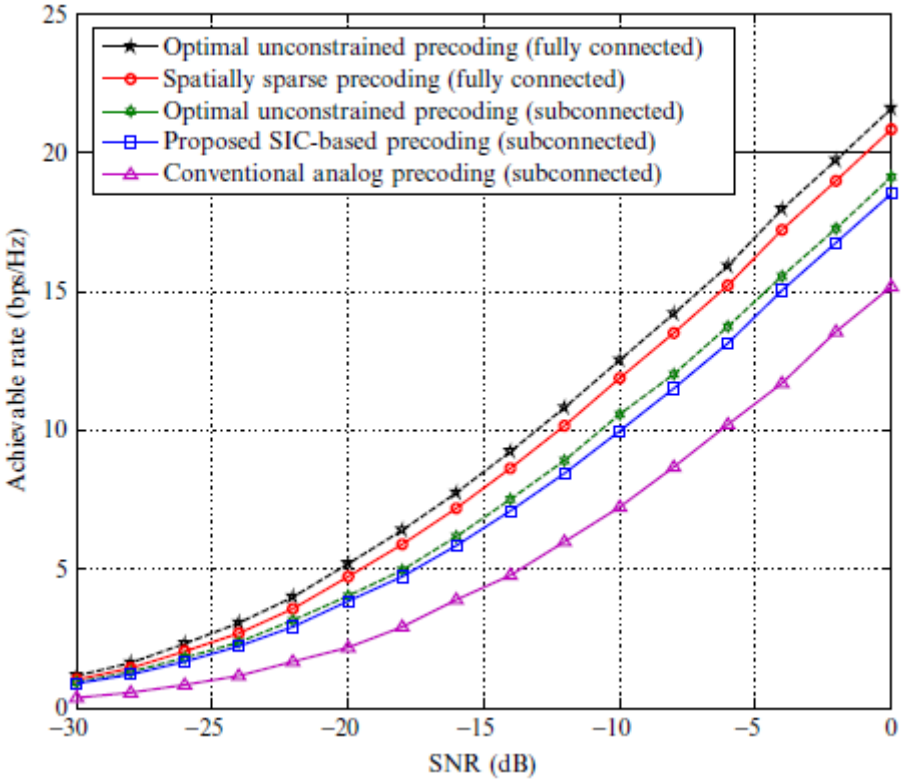
Krok 1: Wykonaj SVD z $\bar{\mathbf{G}}_{n-1}$, aby uzyskać v_1.

Krok 2: Niech $\bar{\mathbf{p}}_n^{\mathrm{opt}} = \frac{1}{M} \|\mathbf{v}_1\|_1 e^{j\mathrm{angle}(\mathbf{v}_1)}$ będzie optymalnym rozwiązaniem dla bieżącej n-tej tablicy podanteny.

Krok 3: Zaktualizuj macierze $\mathbf{T}_n = \mathbf{I} + \frac{\rho}{N_s \sigma_n^2} \mathbf{H} \mathbf{P}_n \mathbf{P}_n^H \mathbf{H}^H$ i $\overline{\mathbf{G}}_n = \mathbf{R} \mathbf{H}^H \mathbf{T}_n^{-1} \mathbf{H} \mathbf{R}^H$ dla następnej tablicy podanteny (n+1)-tej

Ocena wydajności

W tej sekcji przedstawiamy wyniki symulacji osiągalnej szybkości w celu oceny wydajności proponowanego hybrydowego wstępnego kodowania opartego na SIC. Porównujemy wydajność wstępnego kodowania hybrydowego opartego na SIC z przestrzennie rzadkim kodowaniem wstępnym [4] i optymalnym nieograniczonym kodowaniem wstępnym opartym na SVD macierzy kanału, które mają w pełni połączoną architekturę. Dodatkowo uwzględniamy również konwencjonalne analogowe wstępne kodowanie [24] i optymalne nieograniczone wstępne kodowanie (tj. $\overline{\mathbf{p}}_n^{\text{opt}} = \mathbf{v}_1$), które mają podporządkowaną architekturę jako punkty odniesienia dla porównania. Parametry symulacji opisano poniżej. Generujemy macierz kanałów zgodnie z modelem kanału opisanym w rozdziale 5.2. Liczba efektywnych ścieżek kanału wynosi L=3 . Częstotliwość nośną jest ustawiona na 28 GHz. Zarówno układy anten nadajnika, jak i odbiornika są ULA z odstępami między antenami d = λ / 2. Ponieważ stacja bazowa zwykle wykorzystuje anteny kierunkowe w celu wyeliminowania zakłóceń i zwiększenia wzmocnienia anteny, zakłada się, że AoD podążają za równomiernym rozkładem w zakresie [-π/6, π/6]. Tymczasem ze względu na losową pozycję użytkowników zakładamy, że AoA mają rozkład równomierny w granicach [- π, π], co oznacza, że anteny dookólne są przyjmowane przez użytkowników. Wreszcie SNR definiuje się jako ρ/σ_n^2. Rysunek przedstawia osiągalne porównanie szybkości w masywnym systemie MIMO mmWave, w którym N_t xN_r = 64 x 16, a liczba łańcuchów RF wynosi $N^{RF}_t = N_s = 8$.

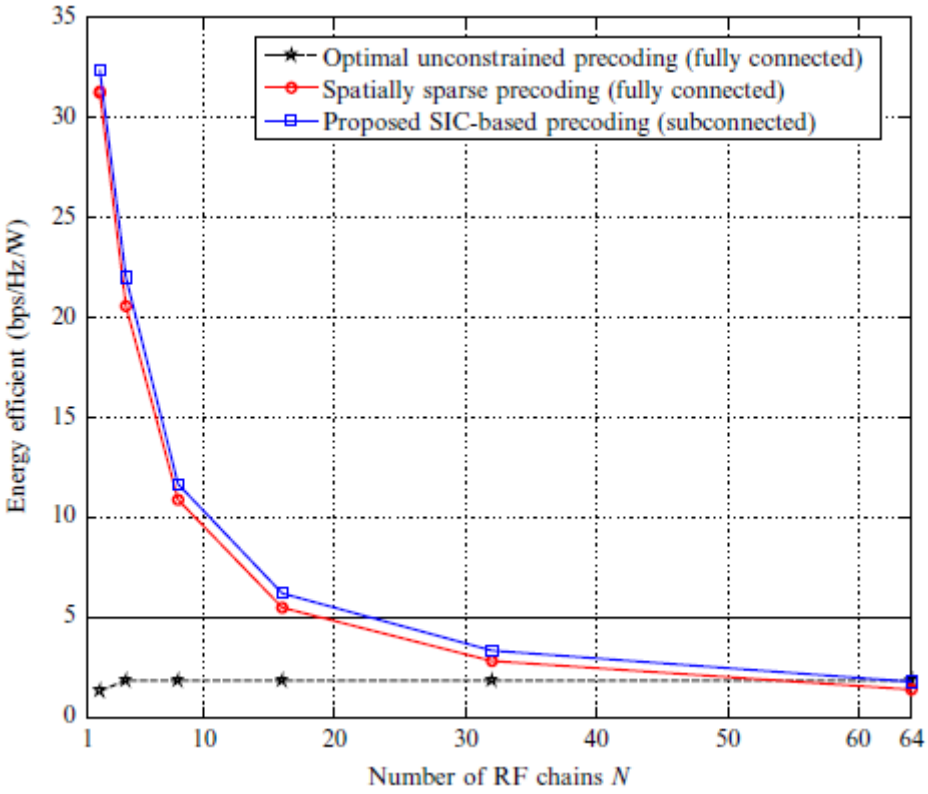


Na rysunku możemy zauważyć, że hybrydowe kodowanie wstępne oparte na SIC przewyższa konwencjonalne wstępne kodowanie analogowe z architekturą podpołączoną w całym zakresie

symulowanego SNR. W międzyczasie weryfikuje również prawie optymalną wydajność wstępnego kodowania hybrydowego opartego na SIC, ponieważ może osiągnąć około 99% szybkości uzyskiwanej dzięki optymalnemu nieograniczonemu wstępnemu kodowaniu z podpołączoną architekturą. Co ważniejsze, rysunek pokazuje, że wydajność wstępnego kodowania hybrydowego opartego na SIC jest zbliżona do przestrzennie rzadkiego wstępnego kodowania i optymalnego nieograniczonego wstępnego kodowania przy w pełni połączonej architekturze. Na przykład, gdy SNR = 0 dB, wstępne kodowanie hybrydowe oparte na SIC może osiągnąć ponad 90% szybkości osiąganej przez prawie optymalne przestrzennie rzadkie wstępne kodowanie w obu symulowanych konfiguracjach MIMO mmWave. Następnie oceniamy efektywność energetyczną hybrydowego kodowania wstępnego opartego na SIC. Na podstawie modelu zużycia energii, efektywność energetyczną η można zdefiniować jako

$$\eta = \frac{R}{P_{\text{total}}} = \frac{R}{P_t + N_{\text{RF}} P_{\text{RF}} + N_{\text{PS}} P_{\text{PS}}} (\text{bps/Hz/W}) \quad (39)$$

gdzie $P_{\text{total}} \triangleq P_t + N_{\text{RF}} P_{\text{RF}} + N_{\text{PS}} P_{\text{PS}}$ i to całkowite zużycie energii, P_t to energia przesyłana, P_{RF} to energia zużyta przez łańcuch RF, P_{PS} to energia zużyta przez PS (w tym energia do wzbudzenia i energia do kompensacji tłumienie wtrąceniowe), a N_{RF} i N_{PS} to odpowiednio numery wymaganych łańcuchów RF i PS. Tutaj używamy praktycznych wartości P_{RF} =250mW , P_{PS} =1mW i P_t = 1W (około 30 dBm) w scenariuszu transmisji w małych komórkach , ponieważ istnieje większe prawdopodobieństwo, że mmWave zostanie zastosowane w małych komórkach. Rysunek poniżej przedstawia porównanie efektywności energetycznej z liczbą łańcuchów RF Nt RF, gdzie SNR = 0 dB, N_t x N_r = 64 x 64 (N^{RF}_t = 1, 2, 4,…, 64, aby upewnić się, że M jest liczbą całkowitą).



Możemy zauważyć, że zarówno przestrzennie rzadkie wstępne kodowanie, jak i wstępne kodowanie oparte na SIC mogą osiągnąć wyższą efektywność energetyczną niż optymalne nieograniczone wstępne

kodowanie (znane również jako w pełni cyfrowe wstępne kodowanie), zwłaszcza gdy liczba łańcuchów RF N_t^{RF} jest ograniczona (np. $N^{RF}_t \leq 30$). Poza tym możemy również zauważyć, że wstępne kodowanie oparte na SIC jest bardziej wydajne energetycznie niż przestrzennie rzadkie kodowanie wstępne.

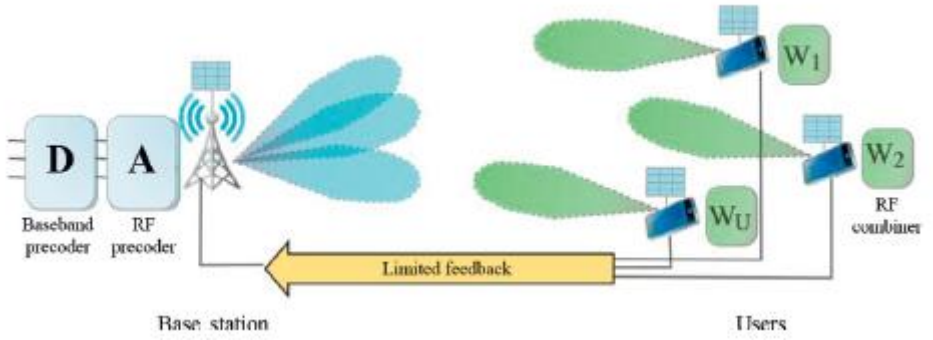
Podsumowanie hybrydowego wstępnego kodowania opartego na SIC

Podstawową ideę hybrydowego wstępnego kodowania opartego na SIC można podsumować w następujący sposób: (1) rozkłada on całkowity osiągalny problem optymalizacji szybkości z nieprzekraczającymi ograniczeniami na serię prostych problemów optymalizacji podrzędnej szybkości, z których każdy rozważa jedną tablicę subantenna; (2) dowodzi, że maksymalizacja osiągalnego współczynnika podrzędnego każdej tablicy subanten jest równoznaczna z prostym poszukiwaniem wektora prekodującego wystarczająco blisko (pod względem odległości euklidesowej) nieograniczonego optymalnego rozwiązania; i (3) maksymalizuje osiągalny współczynnik podrzędny każdej tablicy podanten jeden po drugim, aż do rozważenia ostatniej tablicy podanten. Wyniki symulacji potwierdzają, że hybrydowe kodowanie wstępne oparte na SIC jest bliskie optymalnego i zapewnia wyższą efektywność energetyczną niż wstępne kodowanie przestrzennie rzadkie i w pełni cyfrowe kodowanie wstępne. Rozwiniętą wersję hybrydowego kodowania wstępnego opartego na SIC można również znaleźć w Ref. [16], gdzie zaproponowano algorytm o niskiej złożoności do realizacji hybrydowego kodowania wstępnego opartego na SIC, co pozwala uniknąć potrzeby SVD i inwersji macierzy.

WSTĘPNE KODOWANIE HYBRYDOWE WIELU UŻYTKOWNIKÓW

Model systemu

Rozważmy wieloużytkownikowy system mmWave, jak pokazano na rysunku, gdzie stacja bazowa wyposaża anteny N_{BS} i łańcuchy RF N_{RF} ($N_{RF} \leq N_{BS}$) do jednoczesnej komunikacji z U MSs.



Każde MS wyposaża anteny N_{MS}, ale tylko jeden łańcuch RF. Oznacza to, że skupiamy się na przypadku, w którym BS komunikuje się z każdym MS tylko za pośrednictwem jednego strumienia. Dlatego całkowita liczba strumieni danych do komunikacji wynosi $N_S = U \leq N_{RF}$. Aby w pełni osiągnąć zysk multipleksowania przestrzennego, zakładamy, że stacja BS będzie wykorzystywać U z dostępnych łańcuchów RF N_{RF} do obsługi takich jednostek U MSs. Skupiamy się na architekturze w pełni połączonej. W łączu w dół stacja BS wykorzystuje cyfrowy prekoder U x U w paśmie podstawowym $D= [d_1, d_2,\ldots, d_U]$, po którym następuje analogowy prekoder N_{BS} x U $A = [a_1, a_2,\ldots, a_U]$. Następnie transmitowany sygnał może być przedstawiony jako

$$x = ADs \quad (40)$$

gdzie $s = [s_1, s_2,\ldots,s_U]^T$ o rozmiarze Ux1 jest pierwotnym wektorem sygnału przed wstępnym kodowaniem ze znormalizowaną mocą $\mathbb{E}(\mathbf{ss}^H) = (\rho/U)\mathbf{I}_U$, a ρ jest średnią przesyłaną mocą.

Zakładamy równy przydział mocy między strumieniami różnych MS. Ponieważ analogowy prekoder A jest zaimplementowany przy użyciu analogowych PS, wszystkie elementy A mają tę samą stałą amplitudę N^{-1}_{BS}. Poza tym ograniczenie całkowitej transmitowanej mocy jest wymuszane przez normalizację D w celu spełnienia $||AD||^2_F = U$. Weź również pod uwagę model kanału z zanikaniem bloków wąskopasmowych. Odebrany wektor sygnału ru obserwowany przez Uth MS można przedstawić jako

$$\mathbf{r}_u = \mathbf{H}_u \sum_{n=1}^{U} \mathbf{A}\mathbf{d}_n s_n + \mathbf{n}_u, \quad (41)$$

gdzie H_u o wielkości N_{MS} x N_{BS} oznacza masową macierz kanałów mmWave massive MIMO między BS a u-tym MS, a n_u jest wektorem AWGN, którego wpisy są zgodne z niezależnym i identycznym rozkładem (i.i.d.) $CN(0,\sigma^2_n)$

W u-tym MS analogowy sumator w_u jest używany do łączenia odebranego sygnału r_u jako

$$y_u = \mathbf{w}_u^H \mathbf{r}_u = \mathbf{w}_u^H \mathbf{H}_u \sum_{n=1}^{U} \mathbf{A}\mathbf{d}_n s_n + \mathbf{w}_u^H \mathbf{n}_u, \quad (42)$$

gdzie wu ma podobne ograniczenia jak analogowy prekoder A, to znaczy wszystkie elementy w_u mają tę samą amplitudę N^{-1}_{MS}, ale różne fazy. Tutaj zakładamy, że po stronie MS używane jest tylko łączenie analogowe, ponieważ MS jest bardziej prawdopodobne, że wdroży tańszy i prostszy sprzęt o niższym zużyciu energii.

Dwuetapowe wstępne kodowanie hybrydowe

Podstawowy pomysł

Naszym celem jest zaprojektowanie analogowego prekodera A, cyfrowego prekodera D w BS i analogowych łączników $\{w_u\}^U_{u=1}$ w MSs, aby zmaksymalizować sumaryczną szybkość $R = \sum_{u=1}^{U} R_u,$, gdzie R_u jest szybkością osiągniętą przez u-ty MS wyrażoną przez

$$R_u = \log_2 \left(1 + \frac{\frac{P}{U} \left| \mathbf{w}_u^H \mathbf{H}_u \mathbf{A}\mathbf{d}_u \right|^2}{\frac{P}{U} \sum_{n \neq u} \left| \mathbf{w}_u^H \mathbf{H}_u \mathbf{A}\mathbf{d}_n \right|^2 + \sigma_n^2} \right). \quad (43)$$

Następnie można sformułować problem projektu wstępnego kodowania, aby znaleźć A^{opt}, D^{opt} i $\{w^{opt}{}_u\}^U_{u=1}$, które rozwiązują

$$\left\{ \mathbf{A}^{\text{opt}}, \mathbf{D}^{\text{opt}}, \left\{ \mathbf{w}_u^{\text{opt}} \right\}_{u=1}^{U} \right\} = \arg\max \sum_{u=1}^{U} \log_2 \left(1 + \frac{\frac{P}{U} \left| \mathbf{w}_u^H \mathbf{H}_u \mathbf{A}\mathbf{d}_u \right|^2}{\frac{P}{U} \sum_{n \neq u} \left| \mathbf{w}_u^H \mathbf{H}_u \mathbf{A}\mathbf{d}_n \right|^2 + \sigma_n^2} \right)$$

$$\text{s.t.} \quad \mathbf{a}_u \in \mathcal{F}, \quad u = 1, 2, \ldots, U,$$
$$\mathbf{w}_u \in \mathcal{W}, \quad u = 1, 2, \ldots, U,$$
$$\|\mathbf{A}\mathbf{D}\|_F^2 = U. \quad (44)$$

Problem w równaniu. (44) to problem z programowaniem mieszanych liczb całkowitych. Jego rozwiązanie wymaga przeszukania całej przestrzeni F^U x W^U wszystkich możliwych kombinacji $\{a_u\}^U_{u=1}$ i $\{w_u\}^U_{u=1}$. Ponadto cyfrowy prekoder D musi być zaprojektowany wspólnie z analogowymi wektorami wstępnego kodowania / łączenia. Niestety optymalne rozwiązanie równania (44) nie jest ogólnie znany nawet bez ograniczeń RF i istnieją tylko rozwiązania iteracyjne. W tym celu sygn. [17] proponuje dwuetapowe hybrydowe kodowanie wstępne dla wielu użytkowników. Główną ideą proponowanego algorytmu jest podzielenie projektu hybrydowego prekodera na dwa etapy, jak pokazano w algorytmie 2. W pierwszym etapie, analogowy prekoder BS i analogowe łączniki MS są wspólnie projektowane, aby zmaksymalizować żądaną moc sygnału każdego z nich. użytkownika, zaniedbując wynikające z tego zakłócenia między użytkownikami. W drugim etapie, cyfrowy prekoder BS jest przeznaczony do zarządzania zakłóceniami wielu użytkowników.

ALGORITHM 2 TWO-STAGE HYBRID PRECODING

Inputs: $\mathcal{F}$BS analog precoding codebook
$\mathcal{W}$MS analog combining codebook
First stage: Single-user analog precoding/combining design
For each MS u, $u = 1,2,\ldots,U$
The BS and MS u select $\mathbf{v}_u^{\text{opt}}$ and $\mathbf{g}_u^{\text{opt}}$ respectively that solve

$$\left\{\mathbf{g}_u^{\text{opt}}, \mathbf{v}_u^{\text{opt}}\right\} = \arg\max_{\substack{\forall \mathbf{g}_u \in \mathcal{W} \\ \forall \mathbf{v}_u \in \mathcal{F}}} \left\|\mathbf{g}_u^H \mathbf{H}_u \mathbf{v}_u\right\|$$

MS u sets $\mathbf{w}_u = \mathbf{g}_u^{\text{opt}}$
BS sets $\mathbf{A} = \left[\mathbf{v}_1^{\text{opt}}, \mathbf{v}_2^{\text{opt}}, \ldots, \mathbf{v}_U^{\text{opt}}\right]$
Second stage: Multiuser digital precoding design
For each MS u, $u = 1,2,\ldots,U$
MS u estimates its effective channel $\overline{\mathbf{h}}_u = \mathbf{w}_u^H \mathbf{H}_u \mathbf{A}$
MS u feeds back $\overline{\mathbf{h}}_u$ to the BS
BS employs ZF digital precoder $\mathbf{D} = \overline{\mathbf{H}}^H \left(\overline{\mathbf{H}}\overline{\mathbf{H}}^H\right)^{-1}$, where $\overline{\mathbf{H}} = \left[\overline{\mathbf{h}}_1^T, \overline{\mathbf{h}}_2^T, \ldots, \overline{\mathbf{h}}_U^T\right]^T$

$$\mathbf{d}_u = \frac{\mathbf{d}_u}{\left\|\mathbf{A}\mathbf{d}_u\right\|_F}, \quad u = 1,2,\ldots,U$$

W szczególności algorytm 2 można podsumować w następujący sposób. W pierwszym etapie BS i każdy MS u projektują analogowe wektory prekodowania / łączenia au i wu, aby zmaksymalizować żądaną moc sygnału dla szóstego MS bez uwzględniania zakłóceń innych użytkowników. Należy zauważyć, że jest to typowy problem projektowy dotyczący wstępnego kodowania / łączenia analogowego dla jednego użytkownika. W związku z tym klasyczne algorytmy uczenia wiązki opracowane dla systemów jednego użytkownika, można wykorzystać do projektowania wektorów wstępnego kodowania / łączenia analogowego bez jawnej estymacji kanału. W drugim etapie każdy MS u najpierw szacuje swój efektywny kanał $\overline{\mathbf{h}}_u = \mathbf{w}_u^H \mathbf{H}_u \mathbf{A}$ dla u = 1, 2,…, U. Następnie efektywny kanał $\overline{\mathbf{h}}_u$ jest przekazywany z powrotem do stacji bazowej. Należy zauważyć, że wymiar każdego efektywnego kanału $\overline{\mathbf{h}}_u$ wynosi tylko Ux1, czyli znacznie mniej niż pierwotna macierz kanału. Dlatego można się spodziewać, że takie sprzężenie zwrotne będzie miało niski narzut sprzężenia zwrotnego dzięki zastosowaniu klasycznych schematów . Wreszcie BS wykorzystuje klasyczny cyfrowy prekoder ZF oparty na efektywnej macierzy kanału $\overline{\mathbf{H}} = \left[\overline{\mathbf{h}}_1, \overline{\mathbf{h}}_2, \ldots, \overline{\mathbf{h}}_U\right]$. Dzięki rzadkim, kanałom mmWave massive MIMO mmWave, oczekuje się, że efektywny kanał MIMO będzie dobrze uwarunkowany , co sprawia, że przyjęcie prostej strategii wstępnego kodowania cyfrowego dla wielu użytkowników, takiej jak ZF, jest w stanie osiągnąć niemal optymalną wydajność, o czym świadczą poniższe:

Twierdzenie 2 Niech algorytm 2 zostanie użyty przy następujących założeniach:

(1) Wszystkie kanały mają jedną ścieżkę, to znaczy $L_u = 1$, u = 1, 2,…, U.

(2) Analogowe wektory wstępnego kodowania a_u, u = 1, 2,…, U i analogowe wektory łączące w_u, u=1, 2,…, U są wektorami sterującymi wiązką z ciągłymi kątami.

(3) Każdy MS u doskonale zna swój kanał H_u, u = 1, 2,…, U.

(4) BS doskonale zna efektywne kanały $\overline{\mathbf{h}}_u$, u=1, 2,…, U.

i zdefiniuj N_{BS} x U macierzy A_{BS} w celu zebrania wektorów odpowiedzi macierzy BS powiązanych z U AoDs, to znaczy $A_{BS} = [a_{BS}(\phi_1), a_{BS}(\phi_2),\ldots, a_{BS}(\phi_U)$, z maksymalnymi i minimalnymi wartościami osobliwymi $\sigma_{max}(A_{BS})$ i $\sigma_{min}(A_{BS})$, odpowiednio. Wtedy osiągalny współczynnik użytkownika u jest ograniczony przez

$$R_u \geq \log_2\left(1+\frac{P}{\sigma_n^2 U}N_{\mathrm{BS}}N_{\mathrm{MS}}|\alpha_u|^2 G\left(\{\phi_u\}_{u=1}^U\right)\right),$$
$$where\ G\left(\{\phi_u\}_{u=1}^U\right)=4\left(\frac{\sigma_{\max}^2(\mathbf{A}_{\mathrm{BS}})}{\sigma_{\min}^2(\mathbf{A}_{\mathrm{BS}})}+\frac{\sigma_{\min}^2(\mathbf{A}_{\mathrm{BS}})}{\sigma_{\max}^2(\mathbf{A}_{\mathrm{BS}})}+2\right)^{-1}.$$

(45)

Twierdzenie 3 Oznaczmy wskaźnik dla pojedynczego użytkownika jako

$$\overline{R}=\log_2\left(1+\frac{P}{\sigma_n^2 U}N_{\mathrm{BS}}N_{\mathrm{MS}}|\alpha_u|^2\right).$$

Gdy algorytm 2 jest używany z założeniami przedstawionymi w Twierdzeniu 2, zależność między szybkością osiągalną przez MS u a szybkością $\overline{R}$ jednego użytkownika spełnia

$$(1)\quad \mathbb{E}\left(\overline{R}-R_u\right)\leq K(N_{\mathrm{BS}},U)$$
$$(2)\quad \lim_{N_{\mathrm{BS}}\to\infty}R_u=\overline{R}\ almost\ surely$$

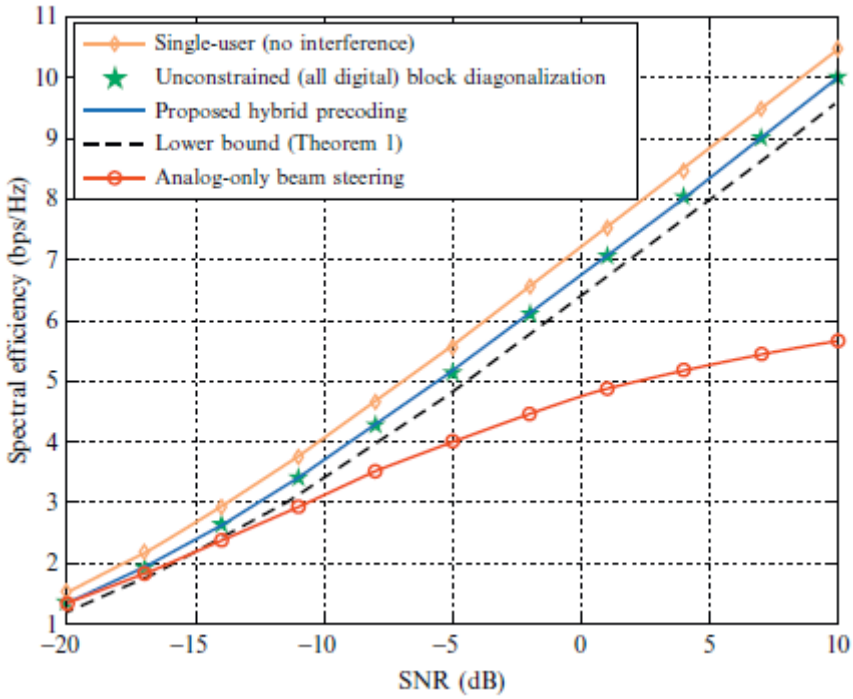
gdzie K (N_{BS},U) jest stałą, której wartość zależy tylko od N_{BS} i U.

Twierdzenie 3 wskazuje, że średnia osiągalna szybkość MS u przy użyciu dwustopniowego hybrydowego wstępnego kodowania dla wielu użytkowników rośnie z tym samym nachyleniem szybkości jednego użytkownika przy wysokim SNR i pozostaje w ramach stałej przerwy K(N_{BS}, U), która zależy tylko od tego na liczbę państw członkowskich U i liczbę anten BS N_{BS}. Co ważniejsze, gdy N_{BS} wzrośnie, A_{BS} macierzy będzie lepiej kondycjonowany, a stosunek między jego maksymalnymi i minimalnymi wartościami osobliwymi zbliży się do jedności. W rezultacie wyraz $G\left(\{\phi_u\}_{u=1}^U\right)$ w równaniu. (45) będzie bliżej jedności, a różnica między tym, co osiągalne przy użyciu algorytmu 2, a wskaźnik dla jednego użytkownika spadnie.

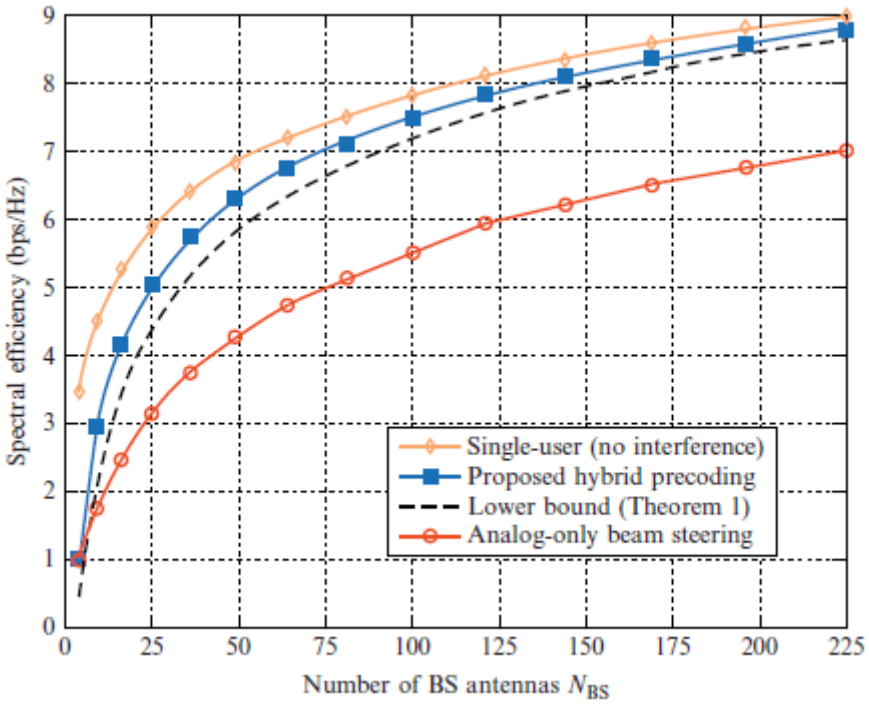
Ocena wydajności

W tej sekcji wydajność dwuetapowego wstępnego kodowania hybrydowego jest oceniana pod względem osiągalnej szybkości.

Na rysunku rozważamy model systemu z BS wykorzystującym 8 8 UPA z czterema MSs, z których każdy ma 4 x 4 UPA.



Zakłada się, że kanał dla każdego użytkownika ma jedną ścieżkę. Zakłada się, że azymuty AoAs / AoD są równomiernie rozmieszczone w [0,2π], a elewacje AoAs / AoD są równomiernie rozłożone w in [-π /2, π/2]. Z rysunku można zauważyć, że wydajność dwustopniowego wstępnego kodowania hybrydowego jest bardzo zbliżona do szybkości jednego użytkownika. Poza tym dwuetapowe hybrydowe kodowanie wstępne może również osiągnąć prawie taką samą wydajność jak w pełni cyfrowe kodowanie wstępne BD i cieszyć się oczywistym wzmocnieniem w porównaniu z w pełni analogowym rozwiązaniem sterowania wiązką, które wzrasta wraz z SNR, gdy rozwiązanie sterowania wiązką zaczyna być ograniczone zakłóceniami. Szczelność wyprowadzonej dolnej granicy w Twierdzeniu 1 jest również zweryfikowana na rysunku. Na kolejnym rysunku to samo ustawienie na rysunku jest rozważane przy SNR = 0 dB, ale z inną liczbą anten BS NBS. Rys. 5.15 pokazuje, że nawet przy bardzo dużej liczbie anten, nadal istnieje znaczny zysk wstępnego kodowania hybrydowego w porównaniu ze sterowaniem wiązką.



Na rysunku możemy również zaobserwować, że różnica między szybkością osiąganą przez dwustopniowe hybrydowe wstępne kodowanie a szybkością dla pojedynczego użytkownika maleje wraz ze wzrostem liczby anten BS, co weryfikuje wnioski w Twierdzeniu 3.

Podsumowanie dwustopniowego wstępnego kodowania hybrydowego dla wielu użytkowników

Podstawową ideą dwustopniowego wstępnego kodowania hybrydowego dla wielu użytkowników jest podzielenie projektu hybrydowego prekodera na dwa etapy. W pierwszym etapie prekoder BS RF i kombinatory MS RF są wspólnie projektowane, aby zmaksymalizować pożądaną moc sygnału każdego użytkownika.

W drugim etapie cyfrowy prekoder BS jest przeznaczony do zarządzania zakłóceniami wielu użytkowników. Wyniki symulacji pokazują, że dwuetapowe hybrydowe kodowanie wstępne przewyższa w pełni analogowe rozwiązanie do sterowania wiązką i zbliża się do wydajności w pełni cyfrowego kodowania wstępnego BD w typowym, wielodostępnym systemie MIMO mmWave. Oprócz dwuetapowego wstępnego kodowania hybrydowego, zaproponowano również kilka schematów wstępnego kodowania hybrydowego dla wielu użytkowników. Na przykład Ref. [28] proponuje wielożytkownikowy hybrydowy schemat wstępnego kodowania w celu zminimalizowania odległości euklidesowej między odebranymi sygnałami systemu z wstępnym kodowaniem hybrydowym a sygnałami systemu z w pełni cyfrowym kodowaniem wstępnym. W ref. [29], zaproponowano podobną koncepcję zwaną hybrydowym kodowaniem wstępnym opartym na średniej kwadracie (MSE), która ma na celu zminimalizowanie MSE między pierwotnie przesłanymi sygnałami przed wstępnym kodowaniem i odebranymi sygnałami po połączeniu.

WNIOSKI

Projektowanie prekodowania i kształtowania wiązki będzie ważnym elementem przyszłych systemów MIMO mmWave masywnych i odpowiadających im standardów 5G. W tym rozdziale najpierw przedstawiamy tradycyjne technologie prekodowania (formowania wiązki), w tym cyfrowe kodowanie wstępne dla systemów MIMO o niskiej częstotliwości oraz analogowe formowanie wiązki do

komunikacji mmWave z dużymi antenami. Cyfrowe kodowanie wstępne ma na celu wyeliminowanie zakłóceń między różnymi strumieniami danych, co może zapewnić satysfakcjonującą wydajność. Ale zwykle wymaga dużej liczby łańcuchów RF, gdy jest rozszerzony na masywne systemy MIMO mmWave, co prowadzi do nieosiągalnych kosztów sprzętu i zużycia energii. W przeciwieństwie do tego, analogowe formowanie wiązki próbuje poprawić wzmocnienie tablicy antenowej, wykorzystując analogową sieć przesuwnika fazy, która wymaga tylko niewielkiej liczby łańcuchów RF. Jednak zwykle cierpi z powodu utraty wydajności w porównaniu z cyfrowym kodowaniem wstępnym. Co więcej, trudno jest go również rozszerzyć na systemy wielostrumieniowe z wieloma użytkownikami. Po wprowadzeniu tradycyjnych technologii wstępnego kodowania (formowania wiązki) badamy wstępne kodowanie hybrydowe, które łączy w sobie zalety zarówno cyfrowego kodowania wstępnego, jak i analogowego kształtowania wiązki. Wstępne kodowanie hybrydowe może osiągnąć wydajność zbliżoną do cyfrowego kodowania wstępnego przy niewielkiej liczbie łańcuchów RF, co oznacza liczbę strumieni danych zamiast liczby anten. W rezultacie hybrydowe kodowanie wstępne wydaje się bardziej odpowiednie dla przyszłych masywnych systemów MIMO mmWave. Warto zauważyć, że oprócz wstępnego kodowania hybrydowego, ostatnio zaproponowano również pewne zaawansowane technologie, które mają na celu zmniejszenie potencjalnego wysokiego zużycia energii i kosztów sprzętu systemów mmWave MassMIMO. W ref. [1], zaproponowano masywną architekturę MIMO anmmWave z niskorozdzielczymi (np. 1-bitowymi) przetwornikami analogowo-cyfrowymi (ADC). Przetworniki ADC o niskiej rozdzielczości mogą być implementowane z prostszymi obwodami i zużywają mniej energii, co może być postrzegane jako alternatywa dla wstępnego kodowania hybrydowego. Co ważniejsze, udowodniono, że przy niskim SNR powoduje to tylko niewielką utratę szybkości w porównaniu z kwantyzacją o wysokiej rozdzielczości .Zaproponowano obszar wiązki MIMO. Poprzez pełne wykorzystanie kątowej rzadkości kanału mmWave, przestrzeń wiązki MIMO wykorzystuje soczewkę do przekształcenia konwencjonalnego kanału MIMO w rzadki kanał przestrzeni wiązki. Następnie można wybrać kilka wiązek uzyskanego kanału przestrzeni wiązek zgodnie z określonym kryterium, aby zmniejszyć efektywny wymiar kanału MIMO bez oczywistej utraty wydajności. Następnie potrzebny będzie tylko mały cyfrowy prekoder zrealizowany przez niewielką liczbę łańcuchów RF.