

## Droga do 5G - wizje i wyzwania

### 1.1 Wizja dla 5G (w pigułce)

"Od najmniejszych rzeczy osobistych po największe kontynenty, wszystko, wszędzie będzie połączone cyfrowo i reaguje na nasze potrzeby i polubienia" jest typową wizją przyszłości sieci komunikacyjnych, jak określono w " Świecie w 2025 roku: 10 Prognozy innowacji "(Thomson Reuters). W szczególności oczekuje się, że komunikacja bezprzewodowa zdominuje wszystko, wszędzie i zmieni codzienne życie, głównie za sprawą rewolucyjnych technologii piątej generacji (5G), charakteryzujących się takimi pojęciami, jak: bezkomórkowe architektury, masowe przetwarzanie przestrzenne, czasy reakcji dotykowej, duże przetwarzanie danych i wirtualizacja, żeby wymienić tylko kilka. W tym celu rozpoczęto badania, ocenę wyników, eksperymenty i działania normalizacyjne: międzynarodowe konsorcja przemysłu i środowiska akademickiego (METIS2020, IWPC, Korea 5G), stowarzyszenia partnerstwa publiczno-prywatnego (5G-PPP), organy regulacyjne (ITU-R WP5D) , a fora operatorów (NGMN) skupiły ostatnio uwagę na ustalaniu wymagań, badaniu charakterystyk koncepcji systemu i badaniu technologii wspomagających dla 5G. Ewolucja bezprzewodowa przeszła długą drogę przez ostatnie 40 lat, poczynając od technologii zorientowanej na technologię aż po przejrzystą przyszłość technologii, przenosząc nacisk innowacji z komponentów i technologii nadawczo-odbiorczych (np. Kodowanie turbo, MIMO) na zasoby zarządzanie przydziałami i zakłóceniami (np. CDMA, OFDMA, skoordynowane wielopunktowe przesyłanie / odbiór-CoMP), a na koniec, komunikacja zorientowana na użytkownika i na urządzenia, w "Internecie wszystkiego". Przyspieszone przez dramatyczny wpływ aplikacji rodzajów sieci społecznościowych, środek ciężkości dla 5G zostanie przesunięty dalej na łączność sterowaną aplikacjami (wdrażaną w przejrzysty sposób przez technologie, infrastrukturę, użytkowników i urządzenia), co będzie głównie możliwe dzięki gęstemu dostępowi / obsłudze węzłów i wykorzystanie komunikacji proksymalnej, pojęcia, które są zwykle nazywane zagęszczaniem sieci. W przeciwieństwie do poprzednich etapów ewolucji komórkowej, które wynikały głównie ze zwiększenia przepustowości systemu i poprawy wydajności spektralnej dla wielu użytkowników, oczekuje się, że zagęszczenie sieci będzie kamieniem węgielnym przyszłych sieci bezprzewodowych. Mimo że zagęszczenie sieci było również brane pod uwagę w poprzednich generacjach komórkowych, było ono używane tylko jako "dodatkowy" środek do lokalnego radzenia sobie z mocno obciążonymi częściami systemu, a nie jako podstawowa, wielkoskalowa koncepcja projektu / systemu. Aby zrealizować ambitne cele ustalone dla 5G, gęstość węzłów dostępu / obsługi powinna wzrosnąć do punktu, w którym jest porównywalna lub nawet znacznie wyższa niż (również zwiększająca się) gęstość sprzętu użytkownika (UE), wprowadzając tym samym paradygmat ultra-gęstej sieci (UDN). Realizacja takiej destrukcyjnej topologii sieci, w odniesieniu do ścieżki ewolucji poprzednich generacji komórkowych, zostanie osiągnięta przez wykorzystanie, oprócz zagęszczonych tradycyjnych węzłów dostępowych (AN), takich jak małe komórki, wdrożone przez użytkownika AN (np. Wi-Fi, femtokomórki), jak również "UE" prosumentów infrastruktury, tj. (urządzenia mobilne) z możliwościami obliczeniowymi i pamięciowymi pozwalającymi im działać jako infrastruktura ANs. Oczywiście, udane wdrożenie UDN nie jest zadaniem łatwym ze względu na trudny obszar zakłóceń wprowadzony przez gęstą i, w dużym stopniu, nieregularne rozmieszczenie AN z transmisjami w ramach uniwersalnego schematu ponownego wykorzystania częstotliwości. W tym sensie UDN można postrzegać jako nową granicę bezprzewodową, którą można urzeczywistnić tylko poprzez ponowne przemyślenie obecnych praktyk komórkowych pod względem: (i) podstawowych mechanizmów umożliwiających rozwój technologii, a mianowicie interfejsu powietrznego, w tym zoptymalizowanych struktur ramowych i związana z tym sygnalizacja napowietrzna, schematy wielodostępu i strategię zarządzania zasobami, tryby łączności, topologia dostępu i projekt transmisji satelitarnej , oraz (ii) implikacje związane z wdrażaniem, takie

jak modelowanie biznesowe, koszt / złożoność, zgodność wsteczna i problemy związane z wprowadzaniem na rynek. Aby skutecznie i skutecznie rozwiązywać te problemy, niezwykle ważne jest zrozumienie podstawowych wartości granicznych niezawodności UDN uzyskanych za pomocą odpowiednich modeli matematycznych. W tym celu można zidentyfikować dwa istotne wyzwania badawcze:

1. Jak powinna być skala zagęszczenia wraz ze wzrostem obciążenia ruchem, szybkością i niezawodnością

wymagania uwzględniające różne interferencje, wielokrotny dostęp, zarządzanie zasobami oraz kwestie związane z kosztami i złożonością?

2. Czy sam proces zagęszczania jest wystarczający, aby spełnić wymagania dotyczące sieci w przyszłości? Czy w związku z tym należy rozważyć współpracę, koordynację i funkcje poznawcze (między węzłami) jako integralną część projektu UDN, która mogłaby zapewnić znaczny wzrost wydajności?

Motywuując powyższymi pytaniami, niniejsza książka ma na celu odkrycie 10 ważnych pomysłów / tematów związanych z bezprzewodowymi badaniami, niektóre z najważniejszych podstawowych ograniczeń, obiecujące czynniki ułatwiające rozwój technologii, zmiany paradygmatów koncepcji systemu oraz krytyczne czynniki projektowania wdrożenia dla realizacji 5G i przedstawia jakościową i ilościową ocenę ich potencjału.

## **1.2 Najważniejsze tendencje w zakresie koncepcji systemu zakłócającego**

Wiele przełomowych osiągnięć technologicznych uważa się obecnie za obiecujące. Ułatwiają realizację oczekiwań 5G, takich jak ogromne MIMO, pełny duplex, komunikacja między urządzeniami, buforowanie i komunikacja mm-Wave. Często są one badane w kontekście tradycyjnych zasad łańcucha wartości, głównie dlatego, że nie określono nowych odpowiednich modeli systemów, odzwierciedlających krytyczne parametry scenariuszy użytkownika 5G, w odniesieniu do architektury sieci, funkcjonalności elementów infrastruktury i własności zasobów. Większość głównego nurtu rozumienia 5G postrzega świadczenie usług telekomunikacyjnych jako udoskonaloną wersję dzisiejszego modelu świadczenia usług oferowanych przez operatorów telefonii komórkowej. Ekstrapolacja na tej samej ścieżce ewolucji następnej generacji (od 2G do 3G, od 3G do 4G itd.) Może nie dawać optymalnej drogi do spełnienia oczekiwań 5G. Trendy użytkownika, wdrażania i technologii, które pojawiają się obecnie, dają pierwsze oznaki, że najważniejszym czynnikiem decydującym o sukcesie 5G jest systemowa koncepcja integracji różnych zaawansowanych technologii. Niektóre z kluczowych trendów dotyczących koncepcji systemu obejmują:

1. Krawędź sieci jest coraz bardziej zagęszczona. Rozbudowane rozmieszczenie małych komórek, nie tylko za pomocą operatora, ale także infrastruktury wdrożonej przez użytkownika, w nieplanowany sposób zmienia właściciela i kontrolę nad krawędzią sieci.

2. Sieć bazowa jest przekształcana poprzez dystrybucję funkcjonalności sieci i ich migrację do krawędzi sieci oraz pojawienie się systemów sieci rdzeniowej typu "plug and play".

3. Pojęcie "zasobów sieciowych" jest ponownie definiowane. Komodyfikacja zasobów sieciowych oznacza nową erę w optymalizacji, architekturze i wdrażaniu sieci: moc obliczeniowa, pamięć masowa, łącze dosyłowe, widmo, przepustowość itd. Mogą być, ze względu na postęp technologiczny, produkowane / oferowane / wdrażane / wykorzystywane w sposób, który pozwala na wypożyczanie, sprzedaż i sprzedaż jednostek tych zasobów. W ten sposób technologia i łańcuchy wartości są przekształcane wraz z podstawowymi modelami biznesowymi.

4. Wirtualizacja funkcjonalności sieci i współużytkowanie zasobów zaangażowane, zmienić znaczenie / znaczenie bliskości, podziału OPEX / CAPEX i modeli własności. Istnieją możliwości projektowania infrastruktury w celu zminimalizowania kosztów / złożoności, maksymalizacji otwartości itp., które wprowadzają kilka stopni swobody w optymalizacji i wdrażaniu sieci.

### 1.3 Ograniczenia wydajności, nowe zasady projektowania i trzy zmiany paradygmatu

Wszyscy zainteresowani są zgodni co do tego, że przyszła sieć bezprzewodowa powinna oferować większe zamówienia pod względem wydajności, pod względem pojemności, gęstości podłączonych urządzeń, wartości szczytowej i liczby użytkowników, opóźnień i niezawodności. Niemniej jednak najbardziej krytyczną cechą jest to, że przyszła sieć będzie musiała sprostać niezwykle zróżnicowanemu środowisku użytkowników, urządzeń i usług. Przykłady obejmują niskoprądową i masową komunikację Internetu rzeczy (IoT), aplikacje sterujące o znaczeniu krytycznym z wysoką jakością wymagań usługowych, wszechobecny mobilny dostęp do Internetu oraz aplikacje do realistycznych timestrycznych aplikacji wymagających dużej przepustowości (takie jak ultra-wysokie definicja wideo, zdalny wyświetlacz 3D, tj. "obecność wirtualna"). Aby osiągnąć te ambitne cele, badania bezprzewodowe, jako pierwszy krok, próbują "liniowo ekstrapolować" na podejście poprzednich generacji sieci komórkowych (2G, 3G, 4G) i powszechnie uważa następujące trzy (oczywiste) kierunki techniczne za oznaczające spełnienie nowych wymagań 5G:

- Uwzględniając więcej częstotliwości, najprawdopodobniej uwzględnimy licencjonowane pasma cm-wave (takie jak pasmo 3,4-3,8 GHz, które pojawia się jako nowy pasmo małej komórki), pasma nielicencjonowane (takie jak pasmo Wi-Fi 5 GHz), co jest również brane pod uwagę dla nielicencjonowanego pasma LTE, znanego również jako LTE-U / LAA), a nawet nieprzydzielone, wyższe pasma częstotliwości w reżimie mm-Wave (28, 38, 70 GHz).
- Infrastruktura zagęszczająca skorzystałaby z większego ponownego wykorzystania przestrzeni, poprzez wdrażanie hiper-gęstych klastrów małych komórek i / lub wykorzystywania technologii Wi-Fi do rozładowywania znacznych ilości ruchu w scenariuszach hotspotów i wnętrzach.
- Wymyślanie nowych technologii radiowych będzie próbowało zmaksymalizować wydajność widmową za pomocą relaksującej ortogonalności w interfejsie radiowym (np. kształty fal GFDM), wzbogacając systemy prądowe o asynchroniczne i nieortogonalne wielokrotny dostęp (schematy SCMA, NOMA), postępująca współpraca między komórkami i koordynacja dla zarządzania zakłóceniami (np. CoMP), wykorzystanie masywnych elementów antenowych do tworzenia przestrzennie izolowanych strumieni w tej samej spektralnej / czasowej jednostce zasobów (Massive / FD-MIMO) i włączenie MIMO wielu użytkowników do nowych standardów Wi-Fi (802.11ad, ax). O ile można udowodnić, że rozwiązania te zapewniają znaczny wzrost wydajności, gdy są wdrażane osobno iw wyidealizowanych warunkach sieci (pod względem sygnalizacji, wymagań dotyczących dosyłania i przetwarzania / złożoności kosztów), to jest wysoce wątpliwe, czy rzeczywiście można je włączyć jako integralną część sieci przed niektórymi nieodłącznymi ograniczeniami jest uwzględniana jako pierwsza:
- Ograniczenia koncepcji systemu komórkowego: Paradygmat komórkowy ma na celu zapewnienie pewnych poziomów jakości usług dla różnych aplikacji w dużych regionach geograficznych poprzez dobrze zaplanowane rozmieszczenie węzłów stacji bazowej. Zakłada się ograniczoną elastyczność przy przydzielaniu zasobów mocy, spektrum i przetwarzania (RF, pasmo podstawowe) i skromne możliwości współpracy między węzłami, aby zapewnić wykonalność i skalowalność pod względem kosztów / złożoności. Niemniej jednak, biorąc pod uwagę ekstremalny trend zagęszczania sieci, oczekuje się, że sieć przyszłości osiągnie wymagania dotyczące skalowania pojemności poprzez przyjęcie wysoce nieregularnych topologii, również obsługujących szereg hierarchicznych warstw komórkowych (to jest, mniejsze skupienia komórek w większych komórkach), sprawia, że "tradycyjne" planowanie sieci

komórkowych praktycznie nie ma znaczenia. Nowe złożone i nieprzewidywalne warunki interferencji ostatecznie podważyłyby wykonalność oczekiwanych poziomów ponownego wykorzystania przestrzennego. Ponadto obecne systemy komórkowe są projektowanymi sieciami ogólnego przeznaczenia. Ich słaba parametryzacja i ograniczone stopnie swobody stawiają znaczne ograniczenia w ich zdolności do dostosowania się do ogromnego dynamicznego zakresu aplikacji, usług i urządzeń oczekiwanych w 5G.

- Ograniczone ograniczenia architektury sieci radiowej: Istniejące i nadchodzące wydania sieci radiowych są ściśle związane z infrastrukturą sprzętową, narzucając poważne ograniczenia dotyczące możliwości rozbudowy sieci (koszty, przestoje) i ograniczając elastyczność w obsłudze wielu technologii radiowych. W przeszłości nie uznawano tego za poważną przeszkodę od rozpoczęcia badań nad nową technologią, standaryzacji jej, a ostatecznie jej wdrożenie trwało ponad dziesięć lat. Ewolucje w zakresie bezprzewodowej ewolucji przyspieszają: podczas gdy wciąż jesteśmy na początku wdrażania 4G, spodziewamy się pierwszych prób pilotażowych 5G w ramach Igrzysk Olimpijskich w Tokio w 2020 r. Ponadto szeroki zakres dynamiki wymagań 5G [8] będzie wymagał wiele rozwiązań konstrukcyjnych pod względem zasięgu, specyfikacji stacji bazowej, zużycia energii itp., zapewne oferowanych przez kilku dostawców. Zamknięte modele architektury sprzętowej wymagałyby zatem ogromnych nakładów kapitałowych i operacyjnych, aby móc wspierać interoperacyjność i rozszerzalność sieci.

- Ograniczenia łańcucha wartości sieci mobilnej: Z perspektywy modelowania biznesowego cały łańcuch sieci komórkowej (w tym widmo, fizyczna infrastruktura dostępu radiowego, łącze dosyłowe, sieć bazowa, zawartość, aplikacje, fakturowanie itp.) Był tradycyjnie posiadany i kontrolowany przez pojedynczą sieć. podmiot, operator sieci komórkowej (MNO). Ostatnio pojawiły się alternatywne koncepcje, a mianowicie: (i) operator wirtualnej sieci (MVNO), który jest właścicielem i kontroluje tylko część łańcucha (np. Sieć rdzeniową) i opiera się na umowach o poziomie usług (SLA) z operatorami sieci ruchomej w celu wykorzystania infrastruktury dostępu radiowego oraz (ii) dostawców usług OTT (over-the-top), którzy pobierają wartość z obecnego łańcucha MNO ("za darmo"). Niemniej jednak te modele biznesowe, w dużej mierze oparte na modelach twardych własności, nie są wystarczająco elastyczne, aby oferować nowe możliwości niszowe. Wynika to głównie z faktu, że koszty wdrożenia, zarządzania i zarządzania ogromną liczbą elementów infrastruktury wymaganej przez przyszłą sieć, przy jednoczesnym oferowaniu tanich usług użytkownikom końcowym, nie są dostępne dla obecnych operatorów (nie mówiąc już o obiecujących warunki generowania przychodów). Wspólną cechą wszystkich tradycyjnych modeli biznesowych jest ograniczony zasób / koszt dzielenie się między interesariuszami sieci. Aby przewyciężyć te ograniczenia, 5G będzie musiało podjąć wyzwanie "przemyslenia" projektu sieci bezprzewodowej pod kątem następujących trzech zmian paradygmatu:

1. Przemysł koncepcję systemu bezprzewodowego, wprowadzając nowe heterogeniczne środowisko UDN, w którym ogromna liczba numerów AN na jednostkę powierzchni jest wdrażana w heterogenicznych topologiach i technologiach, w tym:

- Sterowane przez operatora, bardzo zwarte, małe obiekty, oferujące wiele rzędów wielkości, zwiększające liczbę dostępnych elementów infrastruktury na urządzenie użytkownika.
- Uzupełniające sieci technologii dostępu radiowego (takie jak Wi-Fi, systemy anten kierunkowych - DAS) obsługiwane przez alternatywnych dostawców (np. W portach lotniczych, centrach handlowych, centrach konferencyjnych).
- Infrastruktura domowa wdrożona przez użytkownika, taka jak routery bezprzewodowe do dostępu do Internetu, femtomórk i bramy maszyna-maszyna.

- "Najbogatsze" urządzenia użytkowników (takie jak smartfony lub tablety) wyposażone w różne interfejsy bezprzewodowe i działające jako "prosumenci" dostępu do sieci w bardzo gęstych scenariuszach urządzeń (np. Audytorium, stadion, kino ).

Koncepcja heterogenicznej sieci UDN została zilustrowana dla różnych technologii dostępu radiowego, topologii i wielu schematów dostępu, w tym D2D, M2M itp. Podstawową ideą jest technologia komunikacji proksymalnej, tj. Wysoka dostępność łączności bezprzewodowej za pomocą dużej liczby węzłów w bliskim sąsiedztwie każdego węzła. Podstawowa topologia dostępu realizowana jest tutaj przez łącze "2-hopowe" (lub "2-warstwowe") składające się z "dostępu" przeskakującego łączącego każdy węzeł z bramą i drugim "backhaul". hop, łącząc bramę z głównym punktem dostępu do infrastruktury bezprzewodowej.

2. Ponownie przemyśleć projekt infrastruktury sieci radiowej i (sprzętowej) implementacji, poprzez zmętnienie funkcji sieci bezprzewodowej. Sieć bezprzewodowa Cloudification (WCN) może zapewniać na żądanie, opłacalne i zorientowane na usługi sieci zaprojektowane w locie, poprzez oddzielenie infrastruktury sprzętowej i obsługiwanych funkcji. Osiąga się to głównie poprzez wykorzystanie sprzętu ogólnego (towarowego) (np. Jednostek przetwarzania ogólnego lub GPP) i odpowiednich obiektów (np. Centrów danych IT) oraz poprzez wprowadzenie softwarizacji / wirtualizacji dostępu radiowego w celu wdrożenia w oprogramowaniu wszystkich funkcje systemu, w tym przetwarzanie pasma podstawowego, planowanie zasobów radiowych, routing sieci, itp. WCN pozwala na niezwykle szybkie aktualizacje technologii, wydajną konserwację, minimalne przerwy w świadczeniu usług oraz dynamiczną konfigurację i zarządzanie siecią na żądanie w czasie wykonywania, tworząc wrażenie "nieskończoności" skalowalność sieci "i" "elastyczna". Przedstawiono koncepcję systemu bezprzewodowej sieci bezprzewodowej. Podstawową ideą jest określenie dwóch substratów: fizycznego substratu, w którym znajdują się wszystkie zasoby fizyczne, oraz wirtualnego substratu, w którym wdrażane są wszystkie softwarowe funkcje sieciowe. Warstwa wirtualizacji konwergencji pośrednio realizuje odwzorowanie między zasobami fizycznymi i wirtualnymi (softwarowe funkcje sieciowe).

3. Zastanów się nad pojęciem "elementu zasobów", wprowadzając uogólnioną i rozszerzoną koncepcję zasobów, potencjalnie obejmującą każdą możliwą sieć komponent / element, taki jak anteny, stacje bazowe, łącza / zasoby łącza dosyłowego, pasma / porcje widmowe, przetwarzające jednostki przetwarzania itp., powodujące przerwanie sieci na różne podkomponenty i przekształcanie zasobów w "towary", które wprowadzają nowe modele własności i kontroli, tym samym rewolucjonizując obecny łańcuch Wartości Sieci Mobilnej. Katalizowany przez wirtualizację, redefiniując pojęcie własności zasobów i zasobów oraz dostępu, tworzy zupełnie nowy ekosystem, w którym możliwe jest zbudowanie sieci z towarów, dostosowanej do specyfiki każdej aplikacji, która ma być dostarczona.

#### 1.4 Krytyczne scenariusze użycia w 5G

Aby stworzyć fundamenty radykalnej i rewolucyjnej zmiany w obecnym myśleniu komórkowym wywołanym przez 5G, a jednocześnie wzmocnić jej wpływ i ułatwić jej przyjęcie, niezbędny jest wybór, identyfikacja i specyfikacja odpowiednich scenariuszy użycia. Należy również rozważyć aspekty ewolucyjne i śledzić elementy badań (w toku) prowadzone przez organy normalizacyjne, regulacje i fora technologiczne, takie jak 3GPP, ITU, NGMN itp. Zestaw dobrze zdefiniowanych scenariuszy, odzwierciedlający zarówno rewolucyjne trendy parametry krytyczne i ewolucyjne

aspekty ścieżki, powinny obejmować:

- Potencjalna topologia wdrożenia (np. Planowana, nieplanowana, na żądanie / sterowana zdarzeniami, deterministyczna / kanoniczna geometria, losowa);

- Pewne cechy geograficzne (np. Szeroki obszar na dużym obszarze, wewnątrz, hotspot);
- Istotne / krytyczne scenariusze profilu użytkownika (np. Stopa popytu, wrażliwe w czasie rzeczywistym,

kontrolować krytyczny, wysoki poziom niezawodności, mobilności).

Chociaż znaczna część prac badawczych została ostatnio poświęcona identyfikacji i specyfikacji najbardziej reprezentatywnych scenariuszy 5G, większość proponowanych podejść jest zgodna, mniej więcej, że najbardziej krytyczne wyzwania technologiczne i wyróżniki 5G powinny pojawić się w jednym z następujących trzech przypadków użycia (opisane również w podobny sposób w dokumencie rekomendacji ITU IMT 2020 i poza Vision)

#### **1.4.1 Zatłoczony lokalny dostęp**

Rozpowszechnienie mobilnego dostępu szerokopasmowego (najpierw dostarczone przez 3G-HSPA, a teraz przeniesione na kolejny poziom przez sieci 4G LTE-A), wspierane przez nadejście wysokiego poziomu urządzenia inteligentnego telefonu wraz z szybko rozwijającymi się serwisami społecznościowymi i usługami chmurowymi w świecie Internetu jeszcze bardziej kolidowały z konwergencją między komunikacją a sieciami komputerowymi. Pojawiła się nowa zmiana paradygmatu: upadek granic między treścią a medium komunikacyjnym. Przechodzimy teraz od scenariusza, w którym użytkownik zużywa statyczne informacje publikowane w Internecie (np. Przeglądanie stron internetowych, pobieranie plików), do nowej ery, w której informacje internetowe są tworzone dynamicznie i w dowolnym czasie przez użytkowników, którzy zarówno tworzą, jak i zużywają treści internetowe (np. poprzez przesyłanie filmów i zdjęć, publikowanie / interakcję z innymi osobami w sieciach społecznościowych). Prowadzi to do stale rosnących poziomów mobilnego przepływu danych, które dzisiejsze sieci wydają się niewystarczające do obsługi. Jeszcze trudniejsze jest wspieranie dużego dynamicznego zakresu czasowego i przestrzennego oraz zmienności danych mobilnych. W szczególności ekstremalne szczyty ruchu mobilnego mogą wystąpić w określonym miejscu (np. Przystanek autobusowy, stadion z meczem piłkarskim, specjalne wydarzenie rozrywkowe / happening) oraz w określonym czasie (podczas oczekiwania na autobus, kiedy cel jest punktowany, odliczanie Nowego Roku na Times Square w Nowym Jorku). Ogromny dostęp do lokalnego woluminu danych dla dynamicznego popytu tłumu jest kluczowym scenariuszem dla przyszłości sieci.

#### **1.4.2 Bursty IoT**

Transformacyjna wizja naszego społeczeństwa polega na łączeniu praktycznie wszystkich i wszystkich z Internetem ("Internet-of-Everything" lub IoE). Pojawiają się nowe zastosowania dotyczące komunikacji inicjowanej komputerowo, w tym inteligentne liczniki (czujniki / urządzenia wykonawcze), e-zdrowie (monitorowanie stanu zdrowia pacjentów) i inteligentny transport (śledzenie floty i zarządzanie ruchem). W odróżnieniu od powszechnie przyjętych zasad projektowania "konwencjonalnych" mobilnych sieci szerokopasmowych, które koncentrują się na zwiększeniu osiągniętej (szczytowej) szybkości transmisji danych na urządzenie, paradygmat komunikacji maszynowej jest przede wszystkim ukierunkowany na wsparcie masowo rosnących gęstości urządzeń generujących sporadycznie i sporadycznie / lub dane w boksie. Ta aktywność ruchu może tworzyć geograficznie i czasowo skoncentrowane serie danych, które napędzają obecne sieci do przeciążenia i przestoju. Przystosowywanie się do komunikacji IoT w technologii "burst" wiąże się z wyzwaniem, jakim jest obsługa bardzo zróżnicowanych wymagań dotyczących jakości usług oraz niestabilnych parametrów ruchu i użytkownika.

#### **1.4.3 Niezwykle niezawodna komunikacja o małym opóźnieniu**

Bezprzewodowe sterowanie przemysłowymi procesami produkcyjnymi lub produkcyjnymi (Cyber-Physical Systems in Industry 4.0), zdalna chirurgia medyczna, automatyzacja dystrybucji w inteligentnych sieciach, monitorowanie bezpieczeństwa transportu, autonomiczna jazda, gry w czasie rzeczywistym, wirtualna obecność i rozszerzona rzeczywistość to tylko kilka z nowe, wymagające aplikacje, które przyniosą przyszłe sieci daleko wykraczające poza postrzeganie "nieskończonej" pojemności i "nieskończonej" łączności w IoE. "Dotykowy Internet", jak się powszechnie nazywa, ma przynieść komunikację z prędkością ludzkich zmysłów i wprowadzić współprojektowanie systemów komunikacji, sterowania i obliczeń. W odróżnieniu od poprzednich dwóch scenariuszy, wyzwaniem w tym przypadku jest gruntowne przemyślenie sposobu, w jaki dotychczas zaprojektowano infrastrukturę bezprzewodową (zwykle koncentrując się na zwiększaniu przepustowości i mocy obliczeniowej o rząd wielkości) i kładąc nacisk na osiągnięcie "zero". opóźnienie, bardzo krótki czas odpowiedzi / reakcji (rzędu milisekundy), wysoka dostępność / niezawodność (druga przerwa w roku) oraz wysoki poziom bezpieczeństwa.

## **Nowy interfejs powietrzny 5G: wyzwania dla wydajnego współistnienia wielu usług dla 5G poniżej 6 GHz**

5G będzie musiało poradzić sobie z wysokim stopniem niejednorodności pod względem: (a) usług (mobilny internet szerokopasmowy, masowa komunikacja w zakresie urządzeń i komunikacji krytycznej z misją, usługi szerokopasmowe / multimediami i komunikacja samochodowa); (b) klasy urządzeń (czujniki niskiej klasy do tabletek high-end); (c) rodzaje rozmieszczenia (makra i małe komórki); (d) środowiska (o małej gęstości do bardzo gęstego miejskiego); oraz (e) poziomu mobilności (transport statyczny do szybkiego). W związku z tym należy wspierać różnorodne i często sprzeczne z kluczowymi wskaźnikami wydajności, takie jak wysoka przepustowość / użytkownicy, niskie opóźnienia, wysoka niezawodność, wszechobecny zasięg, wysoka mobilność, ogromna liczba urządzeń, niskie koszty / zużycie energii. 4G nie jest zaprojektowany tak, aby skutecznie spełniać tak wysoką różnorodność heterogeniczności. Ponadto posiadanie wielu technologii dostępu bezprzewodowego do obsługi wielu usług poniżej 6 GHz będzie zbyt kosztowne. Należy opracować nowy wielousługowy interfejs radiowy (AI) dla częstotliwości poniżej 6 GHz w ramach modułowej konstrukcji. Aby umożliwić systemowi dostosowanie się do oczekiwanej heterogeniczności, pożądanymi właściwościami są: elastyczność, skalowalność, wszechstronność, wydajność, odporność na przyszłość. W tym celu, techniczne komponenty sztucznej inteligencji (np. Elastyczny kształt fali i ramka, skalowalne procedury wielokrotnego dostępu, adaptacyjne schematy retransmisji, ulepszone schematy wielu anten z / bez współpracy, zaawansowane wykrywanie wielu użytkowników, koordynacja interferencji, wsparcie dla ultrasownych układów komórek zarządzanie zasobami radiowymi w wielu komórkach, urządzenie z urządzeniem) musi zostać opracowane i zintegrowane z ogólną strukturą sztucznej inteligencji, w której adaptacja do wyżej opisanych źródeł heterogeniczności zostanie osiągnięta i zweryfikowana za pomocą symulacji na poziomie systemu.

### **2.1 Wprowadzenie**

Technologia komórkowa była odnawiana co dziesięć lat od wprowadzenia GSM, które można uznać za głównego przedstawiciela drugiej generacji (2G) technologia komórkowa. W latach 90-tych GSM stało się powszechne i oferowało usługi głosowe na całym świecie. Wraz z rosnącym sukcesem sieci World Wide Web usługi transmisji danych stały się kluczowym czynnikiem napędowym nowej generacji technologii komórkowej. Około 2000 r. Wprowadzenie technologii komórkowej trzeciej generacji (3G) umożliwiło szybki rozwój mobilnych usług transmisji danych. Około roku 2010 wprowadzenie technologii 4G (LTE - Long-Term Evolution) przyniosło znaczące usprawnienia bezprzewodowego dostępu do Internetu szerokopasmowego do usług internetowych. Oczekuje się, że następną generacją nastąpi około 2020 roku. Istnieje pięć trendów, które przesuwały granice 4G, wskazując na potrzebę zastosowania technologii komórkowej piątej generacji:

- Trend 1 - Zwiększenie pojemności: przewiduje się, że zapotrzebowanie na dane bezprzewodowe znacząco wzrośnie, co spowoduje 1000-krotnie wyższe wolumeny danych mobilnych i 10-100-krotnie wyższe szybkości danych użytkowników końcowych.
- Trend 2 - Zwiększenie liczby podłączonych urządzeń: Przewiduje się, że liczba podłączonych urządzeń zwiększy się o 10-100, co oznacza, że do jednego punktu dostępowego musi być dostarczonych do 300 000 urządzeń.
- Trend 3 - Zwiększenie niezawodności: łączność bezprzewodowa zostanie zastosowana w nowych przypadkach użycia, które wymagają wyjątkowo niezawodnych połączeń (zwykle 99,999% dostępności



i niezawodności) oraz komunikacji o znaczeniu krytycznym, takich jak koordynacja pojazd-pojazd, krytyczna kontrola mocy siatka itp.

- Trend 4 - Zmniejszenie opóźnienia: Zdalna obecność i dotykowy Internet nakładają rygorystyczne ograniczenia opóźnień na połączenie od końca do końca, w tym część bezprzewodową. Prognozy wskazują, że opóźnienie powinno zostać zmniejszone pięciokrotnie, aby umożliwić takie usługi.
- Trend 5 - Zwiększenie wydajności: Wydajność w zakresie wykorzystania zasobów (np. Energia, widmo) staje się coraz bardziej wyraźna. Jest postrzegany jako nieodzowny składnik zdrowego / zrównoważonego rynku ICT / biznesu i środowiska.

Wyzwaniem dla 5G jest zatem nie tylko zwiększenie liczby użytkowników lub wydajności (jak to było dotychczas w przypadku poprzednich pokoleń), ale także opanowanie szerokiej gamy wymagań / warunków, jak zaobserwowano w powyższych tendencjach. W związku z tym istnieje potrzeba pracy w kierunku nowego interfejsu lotniczego 5G, który oferuje znacznie więcej niż szybszy wariant 4G. Interfejs radiowy 5G poniżej 6 GHz jest przewidziany, aby zapewnić następujące kluczowe cechy:

- elastyczność w zakresie obsługi szerokiej klasy usług wraz z powiązаныmi (szeroką klasą) kluczowymi wskaźnikami wydajności (KPI),
- skalowalność do obsługi dużej liczby urządzeń,
- wszechstronność do obsługi różnorodnych typów urządzeń i charakterystyki ruchu / transmisji,
- wydajność w celu spełnienia wymagań dotyczących zużycia energii i wykorzystania zasobów, oraz
- odporność na przyszłość, ułatwiająca integrację nowych funkcji.

Aby zrealizować tę wizję, należy realizować następujące cele:

- Współistnienie różnych usług, typów urządzeń i charakterystyki ruchu / transmisji w zespole za pomocą wysoce elastycznego, wszechstronnego i skalowalnego interfejsu powietrznego.
- Wszechobecny zasięg i duża pojemność wszędzie tam, gdzie jest to wymagane.
- Wysoka wydajność w zakresie zużycia energii i zasobów.
- Przyszłość w porównaniu z poprzednimi pokoleniami, ułatwiając proste wprowadzenie nowych funkcji.

- Ocena i walidacja opracowanych koncepcji za pomocą poziomu systemu symulacje i sprzętowe koncepcje wybranych elementów.

## **2.2 Usługi podstawowe i związane z nimi wskaźniki KPI**

### **2.2.1 Podstawowe usługi**

Kluczowe cechy interfejsu 5G (elastyczność, skalowalność, wszechstronność, wydajność i odporność na przyszłe zmiany) umożliwią nowe usługi w różnych obszarach zastosowań, np. Podłączonych urządzeniach, koordynacji pojazd-pojazd, krytycznej kontroli sieci energetycznej, zdalnej obecności, dotykowy Internet i więcej. Usługi te można pogrupować w następujących pięciu kategoriach o nazwie usługi 5G:

- Mobilna sieć szerokopasmowa (MBB2): 5G będzie musiało obsługiwać połączenia szerokopasmowe mobilne od dzisiaj, ale ze zwiększoną pojemnością, wydajnością i szybkością transmisji danych. Typowe

przypadki użycia odpowiadające MBB to strumieniowanie multimedialnych, usługi kombinowane, w tym Voice over IP (VoIP), jako jedna z usług, przeglądanie Internetu, wideokonferencje, pobieranie plików i przesyłanie do chmury, niektóre usługi związane z grami itp.

- Masowa komunikacja z maszyną (MMC3): MMC można uznać za jedną z części komunikacji typu komputerowego (MTC) [1-3]. InMMC Istnieje ogromna liczba czujników / mierników / urządzeń wykonawczych rozmieszczonych w dowolnym miejscu krajobrazu (np. Czujniki ognia w obszarach o dużym zagrożeniu pożarami lasów, czujniki do kontroli / zarządzania jakością wody, czujniki dymu, powiadomienia o awarii zasilania, powiadomienia o manipulacji , siłowniki do kontroli dostępu, temperatury, oświetlenia w budynkach lub sygnalizacji świetlnej w kontekście inteligentnych miast itp.) i które muszą mieć dostęp do sieci bezprzewodowej. W większości przypadków ruch MMC obejmuje stosunkowo małe pakiety na połączenie i dlatego wymaga niskiej przepustowości na urządzenie. Ponadto, typową cechą MMC jest to, że czujniki muszą być mocne, proste (to jest tanie), zdolne do pracy na zasadniczo długich żywotności baterii i osiągalne w trudnych warunkach pokrycia, np. W pomieszczeniach i piwnicach. Dziś jest to również znane jako "low-end Internet of Things (IoT)". Typowe przypadki użycia odpowiadające MMC to inteligentne pomiary, naturalne monitorowanie ekosystemów, zdalna konserwacja / kontrola, śledzenie / śledzenie floty / paczek itp.

- Komunikacja krytyczno-krytyczna (MCC4): MCC zawiera inny typ MTC, który implikuje łączność bezprzewodową, w której komunikaty o czujnikach muszą być przesyłane między odpowiednimi partnerami komunikacyjnymi, przy jednoczesnym spełnieniu jednego lub kilku z poniższych: bardzo krótkie czasy odpowiedzi, bardzo wysoka niezawodność i bardzo wysoka dostępność. Typowymi przypadkami użycia MCC są prywatne aplikacje bezpieczeństwa (np. Nadzór wideo i wykrywanie włamań), monitorowanie życiowych znaków, automatyzacja fabryki itd. Należy zauważyć, że MCC kieruje także komunikacją między osobami (P2P) w ramach np. klęsk żywiołowych i bezpieczeństwa publicznego.

- Usługi Broadcast / Multicast (BMS): BMS obejmuje jednoczesne dostarczanie treści w trybie "jeden do wielu" lub "wiele do wielu". Sprawna integracja BMS w 5G sprawi, że cenne zasoby widma będą dostępne dla innych (pojedynczych) usług. Typowymi przykładami BMS są nie tylko dobrze znane obudowy radia mobilnego i radia cyfrowego, ale także aktualizacje oprogramowania / oprogramowania sprzętowego / oprogramowania; Docelowymi przypadkami użycia mogą być dedykowane pokrycia "Stadium 2.0" lub "Inteligentne miasta" (np. podpis cyfrowy). W szczególnym przypadku dostarczania treści multimedialnych, BMS jest znany jako usługa transmisji multimedialnej / multimedialnej (MBMS).

- Komunikacja pojazd-pojazd i pojazd-infrastruktura (V2X): V2X oznacza bezpośrednią łączność bezprzewodową między pojazdami (pojazd-pojazd, V2V) lub między pojazdami a drogą (Vehicle-to-Infrastructure, V2I) . Dzisiaj, znany również jako Inteligentne Systemy Transportowe (ITS), V2X składa się z przypadków użycia, w których pozostałe cztery podstawowe usługi (MBB, MMC, MCC i BMS) obejmują węzły o dużych prędkościach. Typowymi przykładami są bezpieczeństwo na drodze (np. Unikanie kolizji), które jest kombinacją V2X z MCC, payas-you-drive, który jest połączeniem V2X z MMC i infotainment, który jest połączeniem V2X z BMS i / lub MBB.

## 2.2.2 Kluczowe wskaźniki wydajności

W odniesieniu do powyższych pięciu podstawowych usług określono następujące główne kluczowe wskaźniki efektywności dla 5G:

- KPI 0: Użytkownik doświadcza szybkości transmisji danych

- KPI 1: Natężenie ruchu (aby uzyskać wysoką pojemność systemu)
- KPI 2: Opóźnienie
- KPI 3: Pokrycie (w celu zapewnienia powszechnego dostępu)
- KPI 4: Mobilność
- KPI 5: gęstość połączenia
- KPI 6: Niezawodność / dostępność
- KPI 7: Zmniejszenie złożoności
- KPI 8: Efektywność energetyczna

Mimo że wiele kluczowych wskaźników wydajności jest zrozumiałych, dalsze wyjaśnienia znajdują się w następnym akapicie:

- KPI 0: szybkość transmisji danych doświadczana przez użytkownika bezpośrednio odnosi się do szerokości pasma lub wydajności widmowej przydzielonej do usługi. W większości przypadków przydzielona przepustowość usługi jest współdzielona z innymi użytkownikami, w zależności od obciążenia ruchem, a ułamek rzeczywiście przydzielonej przepustowości jest oceniany średnio przez cały czas trwania usługi, a nie natychmiast. Warto zauważyć, że istotne jest rozważenie KPI 0 w połączeniu z innymi wskaźnikami KPI, np. KPI 1, KPI 4 i KPI 5, ze względu na kompromisy między nimi. Oczekuje się, że faktyczne (zakres) wartości odpowiedniego wymagania silnie zależą będzie od konkretnego przypadku zastosowania.

- KPI 2: Opóźnienie definiuje się jak w białej księdze NGMN [5]. Dwa odrębne przypadki powinny być wzięty pod uwagę:

\* Opóźnienie E2E (bardziej związane z doświadczeniem użytkownika): Mierzy czas między przestaniem małego pakietu danych z warstwy aplikacji w węźle źródłowym a pomyślnym odbiorem w warstwie aplikacji w węźle docelowym plus równoważny czas potrzebny do przeniesienia odpowiedzi (jeśli przewidziano).

\* Opóźnienie płaszczyzny użytkownika (bardziej związane z projektem interfejsu lotniczego): Mierzy czas potrzebny na przestanie małego pakietu danych z terminala użytkownika do interfejsu warstwy 2 / warstwy 3 w docelowym węźle systemu 5G, wraz z równoważnym czasem potrzebnym do przeniesienia odpowiedzi z powrotem.

- KPI 3: zasięg może być zdefiniowany pod względem osiągalnego obszaru / odległości wokół / z nadajnika (-ów), w których jest nadal dostępny odpowiedni poziom czułości, w celu zagwarantowania, że pożądaný sygnał jest odbierany powyżej lub przynajmniej w punkcie odniesienia poziom czułości odbiornika (dotyczy zarówno Uplink (UL), jak i Downlink (DL)). W zależności od rozważanej usługi podstawowej, takiej jak np. MMC, zasięg może być również zdefiniowany jako Maksymalny Strata Sprężenia (MCL) dozwolony zarówno na UL, jak i na DL (przy założeniu minimum), pod względem różnicy między mocą transmisji nadajnik i referencyjny poziom czułości odbiornika (biorąc pod uwagę termiczną gęstość szumów, wielkość szumu odbiornika, margines interferencji, zajęte pasmo kanału, wymagany SINR i wzmocnienie przetwarzania odbiornika).

- KPI 6: można założyć, że dostępność jest równa (1 - prawdopodobieństwo zablokowania usługi), gdzie prawdopodobieństwo zablokowania usługi wynika z braku wystarczających zasobów, aby uzyskać dostęp, przyznać i zapewnić usługę, nawet w przypadku odpowiedniego zasięgu. Dodatkowo

można założyć, że niezawodność jest równa (1 - szybkość zrzutu usługi<sup>5</sup>), gdzie może wystąpić upuszczenie usługi, np. W przypadku dość poważnych zakłóceń wielokrotnych otrzymywanych razem z pożądanym sygnałem, nawet w przypadku odpowiedniego pokrycia. W zależności od rozważanej usługi podstawowej, niezawodność może być również zdefiniowana jako prawdopodobieństwo przeniesienia danego ładunku i otrzymania odpowiedniego potwierdzenia z pewnym wcześniej określonym wymaganiami niezawodności i opóźnienia.

- KPI 7 i 8: Te dwa wskaźniki KPI zostaną ocenione w porównaniu do obecnych standardów. Szacowana złożoność i zużycie energii powinny być porównywalne do dzisiejszych zastosowań metropolitalnych. Utrzymanie stałego zużycia energii przy jednoczesnej poprawie wydajności systemu oznacza potrzebę większej efektywności energetycznej.

Powyższe wskaźniki KPI można podzielić na dwie kategorie:

- KPI w zakresie User Experience: 0, 2, 36, 4, 6
- KPI wydajności systemu: 1, 3, 5, 7, 8

Kluczowe wskaźniki wydajności User Experience bezpośrednio wpływają na jakość usługi (QoS) i ogólną jakość obsługi (QoE) dla użytkownika danej usługi, podczas gdy KPI wydajności systemu odnoszą się głównie do wydajności świadczenia usług od operatora sieci mobilnej (MNO) perspektywicznie. Obydwie kategorie są równie ważne: niedopełnienie wymagań User Experience KPI prowadzi do braku klientów do świadczenia usługi (tj. Do osiągnięcia KPI wydajności systemu), mimo że usługa jest dostarczana sprawnie. Nieprzestrzeganie wskaźników KPI wydajności systemu prowadzi do braku operatorów sieci ruchomej świadczących usługi pomimo spełnienia oczekiwań użytkowników (tj. Spełnienia kluczowych wskaźników wydajności User Experience).

### **2.2.3 Znaczenie KPI dla usług podstawowych**

Bliższe zapoznanie się z podstawowymi usługami wprowadzonymi powyżej może pomóc w identyfikacji kluczowych wskaźników wydajności, które są dla nich najważniejsze. Im wyższy jest względny poziom wymagań odpowiedniego kluczowego wskaźnika wydajności na podstawie usługi bazowej, tym trudniejsze są odpowiednie wartości (zakres), a tym samym samo wymaganie dla tego wskaźnika KPI.

## **2.3 Przypadki użycia z literatury 5G**

Powszechnie uznaje się, że rozważenie każdej usługi osobno i budowa sieci 5G nie będzie zrównoważone ekonomicznie ani środowiskowo. Z tego powodu projekt sieci 5G musi zostać przeprowadzony w kierunku zestawu przypadków użycia które zazwyczaj łączą różne typy usług. W tej części krótko przedstawimy przypadki użycia 5G z literatury (głównie z projektu METIS, NGMN i 3GPP) i pokażemy, jak można zidentyfikować reprezentatywne przypadki użycia.

### **2.3.1 Przypadki użycia z projektów METIS / METIS-II**

Projekt METIS po raz pierwszy wprowadził 12 przypadków testowych (TC), w których przypadek testowy jest praktycznym scenariuszem sformułowanym z punktu widzenia użytkowników końcowych i zawiera zestaw założeń, ograniczeń i wymagań. Te przypadki testowe zostały wybrane tak, że w zasadzie próbują przestrzeni przyszłych aplikacji, co oznacza, że niektóre z nich mają aplikacje, które tradycyjnie nie były brane pod uwagę w badaniach systemów telekomunikacyjnych (stan na kwiecień 2013 r.). 12 przypadków testowych są następujące:

- TC1: biuro wirtualnej rzeczywistości

- TC2: Gęste miejskie społeczeństwo informacyjne
- TC3: Centrum handlowe
- TC4: Stadion
- TC5: Teleprotekcja w sieci inteligentnych sieci
- TC6: korek drogowy
- TC7: martwe pola
- TC8: Zdalne przetwarzanie w czasie rzeczywistym dla terminali mobilnych
- TC9: Festiwal na świeżym powietrzu
- TC10: komunikacja w sytuacjach awaryjnych
- TC11: Masowe rozmieszczenie czujników i elementów wykonawczych
- TC12: Efektywność i bezpieczeństwo ruchu

Następnie METIS wprowadził dziewięć przypadków użycia inspirowanych najnowszymi trendami technologicznymi i przyszłymi prognozami. Dziewięć przypadków użycia jest następujących:

- UC1: Gaming
- UC2: Maraton
- UC3: Media na żądanie
- UC4: Bezzałogowe statki powietrzne
- UC5: Zdalna interakcja dotykowa
- UC6: eHealth
- UC7: Bardzo niska sieć 5G
- UC8: Zdalne wykrywanie i kontrola samochodu
- UC9: Przemysł leśny na zdalnym sterowaniu

Projekt METIS-II wybrał podzbiór powyższych 21 przypadków użycia (zmiana nazwy przypadków testowych do użycia) dla TC1 (te same wymagania, które określono w METIS), TC2 (zaktualizowano wymagania METIS, aby zapewnić zgodność z wymaganiami NGMN w zakresie ponad 300 Mb / sw przypadku 95% użytkowników), TC11 (te same wymagania jak określono w METIS) i TC12 (zaktualizowane wymagania METIS w celu zapewnienia, że niski opóźnienie rozciąga się na V2I, dzięki czemu rozwiązania mogą pomóc w innych aplikacjach uMTC, oraz dodano możliwości MBB do rozrywki na pokładzie). METIS-II wprowadził jeden więcej przypadków użycia

- UC22: Dostęp szerokopasmowy wszędzie koncentrujący się na zasięgu na obszarach wiejskich i podmiejskich z wymaganiami NGMN 50 Mbps w DL i 25 Mbps w UL.

### **2.3.2 Przypadki użycia NGMN**

NGMN opracował 24 przypadki użycia dla 5G [5] jako reprezentatywne przykłady, które są pogrupowane w 8 rodzin przypadków użycia. Przypadki użycia i rodziny przypadków użycia służą jako dane wejściowe do ustawiania wymagań i definiowania elementów składowych architektury 5G.

Przypadki użycia nie są wyczerpujące, ale raczej traktowane jako narzędzie zapewniające, że poziom elastyczności wymagany w 5G jest dobrze uchwycony. Przypadki użycia 5G wymagają bardzo różnorodnych, a czasem ekstremalnych wymagań. Oczekuje się, że jedno rozwiązanie spełniające wszystkie ekstremalne wymagania w tym samym czasie może prowadzić do nadmiernej specyfikacji i wysokich kosztów, chyba że operatorowi sieci zaoferuje się wszechstronne ramy. Niemniej jednak przewiduje się, że kilka przypadków użycia będzie aktywnych jednocześnie w tej samej sieci operatora, co wymaga dużej elastyczności i skalowalności sieci 5G.

### **2.3.3 3GPP SMARTER SI przypadków użycia**

Studium wykonalności w zakresie nowych usług i rynków (LD\_SMARTER) realizowane przez 3GPP SA1 ma na celu zidentyfikowanie segmentów rynku i branż, których potrzeby powinny być zaspokojone przez 3GPP, oraz opracowanie przypadków użycia i wymagań, które będą wymagały ekosystemu 3GPP wspierać w przyszłości. Przedmiotem niniejszej pracy są przypadki użycia i odpowiednie wymagania, których nie można spełnić dzięki systemowi Evolved Packet System (EPS). Niemniej jednak, SMARTER obejmuje również analizę, które usługi i wymagania istniejące w istniejących systemach 3GPP muszą zostać uwzględnione i czy należy opracować mechanizmy awaryjne. Z powyższej tabeli można zauważyć, że 15 przypadków użycia SMARTER może zostać zmapowanych do odpowiednich przypadków użycia NGMN, których numeracja odnosi się do listy przypadków użycia NGMN, począwszy od przypadku użycia 1 (wideo wszechobecne) aż do użycia 24 (Usługi podobne do transmisji: lokalne, regionalne, krajowe). Pozostałe pięć przypadków użycia SMARTER nie jest objęte żadnym przypadkiem użycia NGMN. Te pięć przypadków użycia można scharakteryzować na dwa sposoby. Z jednej strony niektóre z nich są związane ze starszymi usługami świadczonymi przez wcześniejsze generacje / stare systemy (takie jak przypadki 4 i 16 zastosowania SMARTER). Z drugiej strony, niektóre inne odnoszą się do cech funkcjonalnych całego interfejsu lotniczego, takich jak elastyczność i skalowalność, architektura sieci (przecinanie sieci) i elastyczne trasowanie ruchu aplikacji (takie jak przypadki użycia SMARTER 2, 8 i 9). W ten sposób nie mapują bezpośrednio do usług podstawowych 5G.

### **2.4 Futerał 5G firmy FANTASTIC-5G**

Projekt FANTASTIC-5G wybrał podzbiór 5G przypadków użycia, które są w większości przypadków reprezentatywny i który może być zadowolony z widma poniżej 6 GHz (pierwotny główny punkt projektu).

#### **2.4.1 Wybrane przypadki użycia FANTASTIC-5G**

Aby zaprojektować interfejs powietrzny, który spełnia najbardziej rygorystyczne wymagania różnych podstawowych usług w realistycznych scenariuszach, należy wybrać podzbiór przypadków użycia, które są w większości reprezentatywne.

#### ***Krok 1: wybierz przypadki użycia nieobjęte LTE-LTE-Advanced lub nawet wcześniejszymi RAT-ami***

UC23 jest już objęty LTE i LTE-Advanced, w oparciu o jedną częstotliwość tryb sieciowy (SFN), natomiast UC24 wymaga dużej elastyczności i skalowalności, co oznacza, że należy uwzględnić nowe komponenty w porównaniu z dostępnymi w LTE i LTE-Advanced. W rzeczywistości, jak opisano w [5], podczas gdy personalizacja komunikacji doprowadzi do zmniejszenia zapotrzebowania na starszą transmisję, jaką stosuje się dzisiaj, na przykład telewizja, w pełni mobilne i połączone społeczeństwo będzie jednak potrzebowało wydajnego rozprowadzania informacji z jednego źródła do wielu miejsc docelowych. Usługi te mogą rozpowszechniać treści w taki sposób, jak dzisiaj (zwykle tylko DL), ale także udostępniać kanał zwrotny (w UL) dla usług interaktywnych lub informacji potwierdzających. Zarówno usługi czasu rzeczywistego, jak i usługi w czasie rzeczywistym powinny być możliwe i mogą mieć

szeroką dystrybucję lub koncentrować się w mniejszych regionach. Po tej refleksji, UC23 jest wykluczone.

### ***Krok 2: wybierz przypadki użycia nieobjęte wyższymi częstotliwościami (np. Powyżej 6 GHz) RAT***

W przypadku wyższych częstotliwości RAT (np. Powyżej 6 GHz), wiele RAT z potencjalnie różnymi interfejsami powietrznymi może się uzupełniać w celu zapewnienia dostępu szerokopasmowego w gęstych obszarach. Takie podejście pozwala na stosowanie wielu RAT-ów z uzupełniającymi przypadkami użycia, które można wykorzystać do zapewnienia przepustowości i wysokich szybkości transmisji danych w gęsto zaludnionych obszarach miejskich, a także wewnątrz budynków, w celu wsparcia kategorii mobilnych zastosowań szerokopasmowych. Aby zapewnić dostęp szerokopasmowy w gęstych obszarach o niższych częstotliwościach, RAT (poniżej 6 GHz) zdefiniowano przypadek użycia "Gęste społeczeństwo miejskie poniżej 6 GHz". Jak już wspomniano powyżej, ten przypadek użycia jest uważany za ważny dla zwiększenia pojemności w 5G i jest istotny dla usług szerokopasmowych 5G, w szczególności dla zakresu częstotliwości od 3 do 6 GHz.

### ***Krok 3: wybierz przypadki użycia, które obejmują i są bardziej wymagające niż inne przypadki użycia (pod względem kluczowych wskaźników wydajności i odpowiednich wymagań) w ramach tych samych podstawowych usług***

W ramach głównej usługi MBB, UC6 jest bardziej wymagający niż UC7, pod względem prawie wszystkich wskaźników KPI (od KPI 0 do KPI 6, z jedynymi wyjątkami KPI 7 i KPI 8) i odpowiadającymi im wymaganiami. Niemniej jednak, ponieważ KPI7 i KPI8 są ważnymi wskaźnikami dla operatorów sieci ruchomej, zwracamy szczególną uwagę na te wskaźniki KPI podczas pracy nad UC6. W ramach podstawowych usług MBB p V2X, UC8 może być uważany za bardziej wymagający niż UC9 i UC10, pod względem co najmniej KPI 3 i KPI 4, które są kluczowymi wskaźnikami KPI w przypadku zastosowań w pojazdach. Z perspektywy usług podstawowych V2X, UC8 jest bardziej wymagający niż UC11, ponieważ oprócz KPI 3 i KPI 4 (kluczowe w przypadku aplikacji dla pojazdów), ograniczenia związane z mobilnymi aplikacjami szerokopasmowymi muszą być rozważone w tym samym czasie i mogą być kwestionowane i zagrożone przez zasięg i mobilność. W ramach podstawowej usługi MMC, UC13 jest bardziej wymagający niż UC12, pod względem KPI 3, KPI 5, KPI 7 i KPI 8, co jest kluczowe w przypadku masywnego MTC o niskim / długim zasięgu / niskiej mocy. W ramach podstawowej usługi MCC, UC15 jest bardziej wymagający niż UC14, pod względem KPI 2, KPI 3 i KPI 6, które są kluczowymi wskaźnikami KPI w przypadku komunikacji o znaczeniu krytycznym. Z perspektywy usług podstawowych MCC, UC17 jest bardziej wymagający niż UC16, UC18, UC19, UC20, UC21 i UC22, ponieważ oprócz KPI 2, KPI 3 i KPI 6 (kluczowe w przypadku komunikacji o znaczeniu krytycznym), należy wziąć pod uwagę KPI 4 jednocześnie i mobilność zwykle stanowi znaczące wyzwanie w spełnianiu wymagań dotyczących zasięgu i niezawodności / dostępności. Po tych rozważaniach wykluczone są UC7, UC9, UC10, UC11, UC12, UC14, UC16, UC18, UC19, UC20, UC21, UC22. W oparciu o ogólne kryteria, wybrane przypadki użycia FANTASTIC-5G to UC3, UC6, UC8, UC13, UC15, UC17, UC24 (obecnie określane jako Przypadki użycia 1-7):

- Zastosuj wielkość 1: 50 Mb / s wszędzie w MBB
- Użyj Przypadek 2: Szybki pociąg MBB p V2X
- Użyj Przypadek 3: Sieci czujników MMC
- Użyj Przypadek 4: Dotykowy Internetowy MCK
- Użyj przypadku 5: Automatyczna kontrola ruchu / jazda MCCp V2X

- Use Case 6: Broadcast like services: lokalny, regionalny, krajowy BMS
- Użyj Przypadek 7: Gęste społeczeństwo miejskie poniżej 6 GHz MBB

Warto zauważyć, że siedem wybranych przypadków użycia reprezentuje wszystkie podstawowe usługi jako usługą pojedynczego klucza (MBB, MMC, MCC, BMS) lub jako połączenie dwóch podstawowych usług (V2X w połączeniu z MBB lub CMC). że przypadek NGMNuse "Gęste społeczeństwo miejskie" został zmodyfikowany do "gęstej społeczności miejskiej poniżej 6 GHz", koncentrując się na zakresie częstotliwości od 3 do 6 GHz. W związku z tym wartość NGMN z szybkością 300 Mb / s została zmniejszona do 50 Mb / s. Ważne jest również podkreślenie, że wybrane przypadki użycia według METIS II są w większości brane pod uwagę na powyższej liście wybranych przypadków użycia pod względem UC6, UC13 i UC17. Należy pamiętać, że UC4 (smart office) nie może być brane pod uwagę, ponieważ tak jest przypadek użycia, który jest zwykle pokryty (również) przez wyższe częstotliwości RAT.

## 2.5 Wyzwania dla projektu 5G poniżej 6 GHz

### 2.5.1 Metodologia projektowania

W oparciu o wybrane przypadki użycia i wymagania, zaprojektowano ogólne ramy interfejsu lotniczego. W tym celu najbardziej odpowiednie technologie wspomagające (np. Nowe przebiegi, kodowanie kanałów, struktura / numerologia ramek, urządzenie z urządzeniem (D2D), ulepszone MIMO, współpraca wielokomórkowa, kanały sterujące, procedury warstwy 1, itp.), które należy zidentyfikować i ocenić ich znaczenie dla osiągnięcia wymagań każdego przypadku użycia. Projekt ten musi być wykonany zarówno na poziomie łącza i systemu, jak opisano poniżej.

Aby osiągnąć to pojedyncze rozwiązanie, można przyjąć następujące trzyetapowe podejście:

- Etap 1: Dedykowane działania badawcze dotyczące komponentów specyficznych dla usługi (interfejsu lotniczego) dla każdej z pięciu podstawowych usług wymienionych w sekcji 2.2 muszą być wykonane. Na przykład kodowanie kanałów dla MMC, kształtu fali i ramki dla BMS, rozwiązania MIMO dla V2X, dostęp do kanału dla MMC, rozwiązania zarządzania zakłóceniami dla MBB itp. Jednocześnie należy opracować ramy interfejsu lotniczego dla każdej usługi, który definiuje funkcjonalne opisy komponentów i interfejsów pomiędzy nimi.
- Etap 2: Aby połączyć i zharmonizować pracę wykonywaną dla każdej usługi, można wprowadzić koncepcję sterowników integracji usług (SID). Jest jeden identyfikator SID dla każdej usługi podstawowej (która nosi tę samą nazwę wraz z usługą) oraz jeden identyfikator SID dla usług złożonych (zwany "Ogólnym identyfikatorem SID"). Główną rolą podstawowych identyfikatorów SID usług jest zbieranie różnych komponentów zoptymalizowanych dla każdej usługi i integracja ich z konkretnym interfejsem usługowym (obejmującym kształt fali, ramkę, sygnalizację kontrolną, kodowanie, modulację, procedury, schematy retransmisji, MIMO itd. .). Ta integracja zostanie przeprowadzona za pomocą interfejsu interfejsu lotniczego opracowanego na etapie 1. Należy pamiętać, że wszystkie komponenty interfejsu radiowego nie muszą być przeprojektowywane od zera. Wszystkie pozycje badawcze należy sprawdzić pod kątem odpowiedniego rozwiązania w LTE jako wzorca. Komponenty LTE zostaną ponownie użyte w razie potrzeby, ale ponieważ LTE jest a (nieelastyczny) niestandardowy projekt MBB, oczekuje się, że nie będzie to dotyczyć większości komponentów MCC, MMC, BMS i V2X.
- Etap 3: Kiedy rozwiązania interfejsu lotniczego dla każdej usługi zostaną zaprojektowane przez podstawowe identyfikatory SID, analiza wrażliwości musi zostać przeprowadzona przez wszystkie identyfikatory SID, aby zobaczyć, jak różne projekty muszą zostać dostosowane, aby były zgodne z więcej niż jedną usługą. W ten sposób można zbadać granice stopnia elastyczności rozwiązań



specyficznych dla danej usługi. Na przykład projekt kształtu fali i ramy dla MMC zostanie dostosowany do MCC poprzez przeprojektowanie / rekonfigurację, która często jest wyborem pomiędzy kilkoma konfiguracjami. W ten sposób możliwe jest połączenie w jedno rozwiązanie interfejsu lotniczego, które jest w stanie dostosować się do wymagań więcej niż jednej usługi. Dodatkowo oceniany będzie wpływ multipleksowania różnych usług, z których każda będzie zoptymalizowana dla konkretnej usługi w zakresie częstotliwości i / lub czasu.

Poniżej znajduje się bardziej szczegółowy opis identyfikatorów SID.

### **2.5.2 Sterowniki integracji usług**

Pięć z sześciu identyfikatorów SID jest bezpośrednio związanych z wyżej wymienionymi usługami podstawowymi: MBB, MMC, MCC, BMS i V2X. Każda z tych pięciu identyfikatorów tożsamości opracowuje interfejs radiowy specyficzny dla danej usługi, mając na uwadze odpowiednie cechy i potrzeby związane z usługami. W ten sposób staje się dostępny projekt struktury interfejsu lotniczego (etap 1). Ta struktura obejmuje (1) funkcjonalne opisy elementów interfejsu radiowego wymaganych do implementacji interfejsu powietrznego odpowiedniej usługi (np. Niektóre usługi potrzebują środków do sondowania UL, podczas gdy inne nie, niektóre usługi pozwalają na retransmisje Hybrid Automatic Repeat Request (HARQ), podczas gdy inne nie), (2) interfejsy między różnymi budynkami bloki (np. różne usługi będą wymagać różnych interakcji między warstwami) i (3) możliwość zastosowania do różnych ustawień wdrażania. W ramach każdego identyfikatora SID usługi podstawowej niektóre elementy będą (ponownie) zaprojektowane i zoptymalizowane dla tej usługi. Niektóre inne zostaną łatwo wykorzystane z istniejących technologii (pokazano ciemnoszare kule na rysunku 2.3). Następnie zostaną one wkomponowane w specyficzny dla usługi interfejs radiowy za pomocą struktury dla tej usługi (etap 2). Jeśli chodzi o szósty identyfikator SID, ma on na celu przeanalizowanie możliwości połączenia i harmonizacji rozwiązań specyficznych dla danej usługi w jeden interfejs powietrzny zapewniający wielousługowy interfejs powietrzny, który jednocześnie spełnia wymagania pięciu podstawowych usług. W tym celu szósty SID musi (1) zbadać zależności, (2) znaleźć synergie, (3) przeprowadzić analizę wrażliwości we współpracy z innymi identyfikatorami SID i (4) zapewnić ogólny projekt interfejsu lotniczego, spełniający wymagania wszystkie usługi pod uwagę. Szósty identyfikator SID jest jednym z kluczowych sposobów osiągnięcia pojedynczego wielousługowego interfejsu lotniczego. Rysunek 2.3 przedstawia rolę identyfikatorów SID w projekcie interfejsu lotniczego. Optymalnie, wynik tego działania jest bardzo elastyczny, wszechstronny i skalowalny interfejs interfejsu lotniczego umożliwiający obsługę wszystkich przewidzianych (a nawet jeszcze nie przewidziano) usługi z epoki 5G o najwyższej wydajności (pod względem widma i widma energii). Nawet jeśli pojedynczy interfejs multi-service air okaże się nieefektywny, zestaw Zoptymalizowane projekty interfejsów powietrza, z których każdy obejmuje podzbiór wyżej wymienionych usług, mogą być dostarczane dzięki identyfikatorom SID.

### **2.5.3 Wyzwania na poziomie łącza**

Projekt łącza obejmuje kilka elementów technicznych, takich jak modulacja i kodowanie, kształtu fali, ramki i kanału kontrolnego, procedury warstwy fizycznej itp. różnorodne i często sprzeczne ze sobą wymagania 5G, projekt linków opiera się na dwóch ogólnych pytaniach: (i) W jaki sposób nowe funkcje / usługi mogą być dodawane w późniejszych wersjach bez nakładania ciężkich ograniczeń i bez przeszkadzania starszym urządzeniom? oraz (ii) Jaka jest najlepsza równowaga między elastycznością a ustaloną wcześniejszą konfiguracją? Ze względu na szeroki zakres usług 5G, a także ważny cel umożliwienia przyszłej weryfikacji, projekt 5G od samego początku powinien cechować się wysokim stopniem elastyczności. Aby złagodzić obciążenia związane z kompatybilnością wsteczną, należy zaprojektować system z minimalną ilością stałych i predefiniowanych elementów. Dla najwyższej

elastyczności jedynym statycznym / predefiniowanym elementem jest Initial Access Channel (IAC). Każda kolejna pozycja powinna być otwarta na konfigurację między urządzeniem a siecią zgodnie z rzeczywistymi potrzebami, tym samym opierając się na danym zestawie opcji. W ten sposób nowe funkcje i usługi mogą zostać wdrożone w późniejszych wersjach bez wpływu na urządzenia starszego typu. Ponadto drugie pytanie stanowi wyzwanie dla obecnych systemów mobilnych, w których stosuje się raczej konfiguracje stałe (lub nieadaptujące się sam / adaptacyjne). Taka nieelastyczna koncepcja nie może w skuteczny sposób wspierać wielousługowego 5G, ponieważ może stanowić jedynie kompromis w zaspokajaniu potrzeb wszystkich przewidywanych usług. Tak więc opracowywane komponenty technologiczne wymagają elastycznych możliwości adaptacyjnych. Następujące komponenty techniczne mają kluczowe znaczenie z perspektywy poziomu łącza:

- Projektowanie kształtu fali: przebiegi z wieloma nośnymi z funkcjami filtrowania są uważane za kluczowe elementy zapewniające elastyczny projekt PHY wielousługowego interfejsu powietrznego, ponieważ pozwalają na podzielenie pasma systemu na osobne podpasma, których parametry PHY mogą być indywidualnie konfigurowane zgodnie z wymaganiami usługi. Kształty kandydatów można podzielić na dwie kategorie, a mianowicie

- \* filtrowany wielokamerowy podnośny, zawierający FBMC z sygnalizacją QAM / OQAM, OFDM (P-OFDM) w kształcie pulsu [13] i elastycznie skonfigurowany OFDM (FC-OFDM) i

- \* filtrowany wielokierunkowo pod pasmem, zawierający UF-OFDM i F-OFDM

Filtry wielokomorowe pod kontrolą inżynierską wykorzystują filtry prototypowe ze stromym zsuwaniem mocy w dziedzinie częstotliwości do kształtowania impulsów sygnałów podnośnych. Natomiast systemy filtrów wielokierunkowych z podziałem pod względem pasma (UF-OFDM i f-OFDM) wykorzystują filtrowanie FIR na grupę podnośnych po klasycznej modulacji OFDM, która koncepcyjnie zapewnia bliskie podobieństwo (w rzeczywistości rzeczywiste implementacje mogą zastosować funkcję filtrowania przed transformacją do dziedziny czasu w celu zmniejszenia złożoności).

- Procedury PHY związane z konkretnymi usługami: Aby zapewnić uzasadnione obciążenie dla 5G obsługującego jednocześnie wiele usług, w szczególności usługę MBB i usługę MMC z masowym dostępem, niezbędna jest również elastyczność w procedurach losowego dostępu. Istnieje wiele propozycji przebudowy procedury dostępu dla MMC Jednym z przykładów jest dostęp oparty na rywalizacji, wykorzystujący atestowane lub dwustopniowe formaty transmisji. Ten pierwszy jest zgodny z zasadą (zakodowanego) szczelinowego ALOHA (tj. Urządzenie bezpośrednio transmituje swoje dane w połączeniu z odpowiednią preambułą określającą używane zasoby), a zatem wymaga bardzo szczupłej kontroli, podczas gdy druga pozwala stacji bazowej w pewnym stopniu kontrolować transmisję (tj. urządzenie mające dane do przesłania inicjuje żądanie usługi w postaci preambuły UL) i wprowadza dłuższe opóźnienia, ale jest bardziej zasobooszczędne. Procedura jednorazowa ma następującą charakterystykę:

- \* Szybciej, jeśli się powiedzie, znacznie mniej sprzężeń DL

- \* Wysokie prawdopodobieństwo kolizji zmniejsza przepustowość

- \* Kodowany losowy dostęp w celu zwiększenia skalowalności

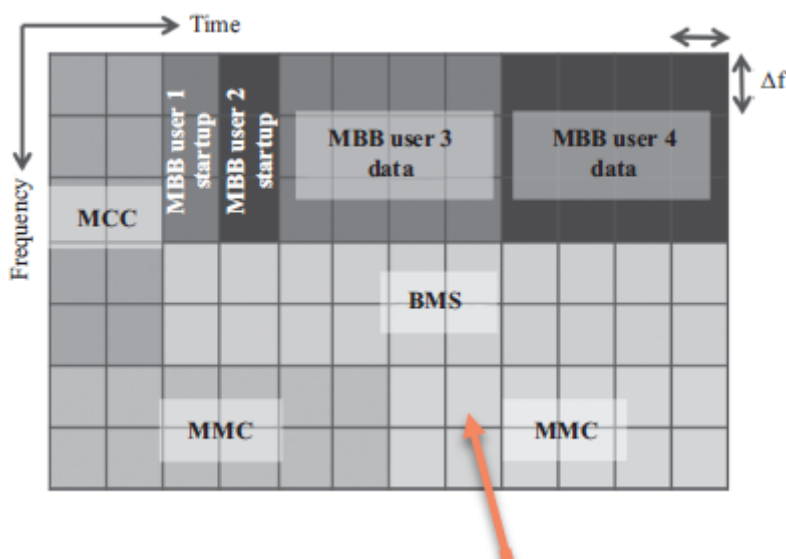
- \* Przewidywane rozwiązanie dla bardzo małych pakietów i małego obciążenia ruchem

- \* Wyjątkowe detektory przetwarzania sygnałów

Z drugiej strony mają dwa protokoły

- \* Dodatkowe opóźnienie
- \* W zależności od konfiguracji wymagana jest pewna ilość sprzężenia zwrotnego DL
- \* Zmniejszone prawdopodobieństwo kolizji dzięki nadprzykładowi obsłudze zleceń zwiększa przepustowość
- \* Przewidywane rozwiązanie dla większych pakietów i większego obciążenia ruchem
- Elastyczna konstrukcja ramy: Rama jest kluczowym elementem do agregowania specyficznej dla użytkownika tryby transmisji w jedno holistyczne rozwiązanie i jego dynamiczna rekonfiguracja jest kluczem do 5G. Konfiguracje te mogą się różnić pod względem np. Charakterystyk kształtu fali (zastosowany filtr, odstęp podnośnej), długości czasu transmisji (TTI) i innych elementów konstrukcji ramek, takich jak gęstość sygnału pilota i podział zasobów:
- \* MMC: umiarkowana przepustowość z dłuższymi TTI zoptymalizowanymi pod kątem niskich kosztów i lepszego zasięgu
- \* MCC: wyższa przepustowość i krótsze TTI zoptymalizowane pod kątem niskich opóźnień i rozsądnych kosztów ogólnych
- \* MBB: krótkie TTI, aby szybko rozpocząć sesję TCP, a następnie średni TTI dla zminimalizowanego narzutu
- \* BMS: długie TTI, aby zmaksymalizować zyski FEC z różnorodności czasowej

Zunifikowana koncepcja ramki została przedstawiona tu



#### 2.5.4 Wyzwania na poziomie systemu

Z perspektywy projektowania systemu i sieci, spełnienie wymagań różnych opisanych powyżej przypadków użycia stanowi wyzwanie dla następujących aspektów:

- Zarządzanie zasobami radiowymi (RRM): Efektywne planowanie rozważanych usług wymaga obsługi formatów planowania z różnymi TTI. Na przykład użytkownicy o ograniczonych opóźnieniach (np. MCC) wymagają krótkich TTI, natomiast użytkownicy MMC zaplanowani na wąską przepustowość są najskuteczniej obsługiwani z dłuższymi TTI.

- Zaawansowane opcje łączności: Aby spełnić rygorystyczne wymagania usług podstawowych 5G, należy wziąć pod uwagę zaawansowane opcje łączności, takie jak bezpośredni D2D [17], masowy dostęp i transmisję broadcast / multicast.

- Możliwości poprawy wydajności spektralnej: W przypadku usługi MBB wysokie oczekiwania na transmisję danych wymagają opracowania mechanizmów umożliwiających zwiększenie wydajności widmowej, takich jak zaawansowane odbiorniki, na przykład w przypadku nieortogonalnego dostępu wielokrotnego (NOMA) i zaawansowanych technik MIMO z lub bez współpracy między różne stacje bazowe [18]. Otwiera to cały szereg aspektów badawczych, aby rozwiązać problemy, kiedy próbuje wykorzystać potencjał dużej liczby strumieni i powiązanych pilotów i odbiorników. Inne aspekty, które wymagają rozwiązania, obejmują sprzężenie zwrotne stanu kanału i narzuty sygnalizacyjne dla właściwego wykorzystania MIMO.

Warto zauważyć, że istnieją inne wyzwania dla projektu systemu integrującego widmo poniżej 6 GHz z pasmami mmWave.

## **2.6 Ocena komponentów technicznych na poziomie systemu**

Aby osiągnąć ambitne cele, należy opracować wiele kluczowych komponentów technologii. Aby ocenić projekty, należy przeprowadzić symulacje na poziomie łącza i systemu Wybrane scenariusze i przypadki użycia dyktują modele środowiskowe i cele KPI. Komponenty technologii (procedury), które są opracowywane i wskazują obiecujące wzrosty wydajności, muszą zostać wdrożone za pośrednictwem określonych funkcji systemu. W pierwszej fazie podczas opracowywania komponentów technologicznych wykonywane są pierwsze oceny za pomocą symulacji na poziomie systemu. Wyniki te zostaną wykorzystane do dopracowania pojęć. Wreszcie, w drugiej fazie zostaną przeprowadzone kompleksowe symulacje na poziomie systemu wykorzystywane do analizy wykonalności projektów w osiąganiu celów i wskaźników KPI. Symulatory na poziomie systemu muszą mieć możliwość modelowania i symulacji pięciu podstawowych usług projektu. Pod tym względem architektura wysokiego poziomu uwzględnia:

- Modele środowiskowe
- Funkcje systemu
- Analiza

Odnośnie modeli środowiskowych wpływ ruchu, mobilności i warunków radiowych należy uwzględnić w symulatorach na poziomie systemu. W szczególności powinien być możliwe generowanie w symulatorach 1000 x więcej ruchu (np. w porównaniu do wersji bazowej 3GPP Release 12), użycie ogromnej liczby urządzeń, ultra-gęste infrastrukturę, komunikację D2D, wysoką mobilność itp. Pewne elementy muszą być rozwinięty, który kontroluje konfigurację topologii sieci, dystrybucję ruchu do komórek, dystrybucji komórek i użytkowników w kosmosie, dystrybucji ruchu w przestrzeni i czasie itd. Zgodność z testami porównawczymi i standardami 3GPP (Wydanie 12 oraz wybrane odpowiednie funkcje z wydania 13) ma kluczowe znaczenie dla zapewnienia prawidłowych porównań i ocen osiągnięć. Jeśli chodzi o cechy systemu, uwzględnia się wykorzystanie widma, np. W pasmach częstotliwości poniżej 6 GHz, wraz z zagęszczeniem sieci, wpływem modulacji, kodowaniem, MIMO itp. Ponadto odpowiednie abstrakcje warstw PHY i MAC w celu zdefiniowania systemu modele zachowania i aspekty złożoności ograniczonej złożoności muszą być wzięte pod uwagę. Na przykład aspekty warstwy PHY można modelować za pomocą narzędzi symulacyjnych, wykorzystując wstępnie obliczone krzywe odwzorowania e. Skuteczność widmowa w porównaniu ze wskaźnikami jakości sygnału, współczynnik błędów blokowych w funkcji stosunku sygnału do szumu i interferencji, fizyczna

szybkość transmisji danych w funkcji stosunku sygnału do szumu i interferencji, i tak dalej. Te tabele wyszukiwania są wypełniane wartościami uzyskanymi z wyników symulacji na poziomie łącza. Ponadto narzędzia symulacyjne zapewniają środki do analizy i dostarczania wyników związanych z docelowymi wskaźnikami KPI, w tym przepustowością, opóźnieniami, stratami pakietów, bitową stopą błędów, niskim zużyciem energii itp. Na koniec należy wspomnieć, że dobrze ugruntowane standardy i referencje są należy wziąć pod uwagę przy definiowaniu niezbędnych wymagań, modeli i technologii, które należy wykorzystać do pomyślnej realizacji i oceny bardzo elastycznego, wszechstronnego i skalowalnego interfejsu powietrznego dla przewidywanej wielości klas usług i typów urządzeń o rozszadnej złożoności i najwyższej wydajności dla pasm częstotliwości poniżej 6 GHz.

## **2.7 Wnioski**

W niniejszej części rozdziału przedstawiono bieżący krajobraz usług i przypadków użycia dla sieci 5G. Dokonuje również wyboru między tymi przypadkami użycia, aby można było wykonać elastyczną konstrukcję interfejsu dla widma poniżej 6 GHz. Wyzwania dotyczące tego projektu na łączach i poziomach systemu są następnie krótko omawiane wraz z ich tematami badawczymi i metodologią oceny dla odpowiednich elementów technicznych. W końcu zauważamy, że projekt nowego interfejsu lotniczego składa się w dużej mierze z tematów wymagających standaryzacji, np. Kształtu fali i ramki, procedur warstwy PHY, przetwarzania transceivera (zarządzanie zakłóceniami, odbiór wielu użytkowników itp.), Kodowania kanałów, kodowania adaptacyjnego i modulacji itp. Są to wszystkie tematy, które implikują interfejsy, komunikaty, sygnalizację, protokoły (wszystkie są standaryzowane w 3GPP) zamiast zastrzeżonych rozwiązań algorytmicznych, a ważnym wyzwaniem dla badaczy jest umożliwienie interakcji z organami normalizacyjnymi, w szczególności 3GPP, w odpowiednim czasie.

## **Nowe quasi-deterministyczne podejście do modelowania kanałów w pasmach fal milimetrowych**

### 3.1 Wprowadzenie

Rosnące zainteresowanie korzystaniem z pasm fal milimetrowych dla mobilnych sieci bezprzewodowych następnej generacji [1-5] prowadzi do wzrostu znaczenia badania właściwości kanału propagacyjnego i właściwego opracowania modelu kanału. Nowe systemy i standardy komunikacyjne wymagają modeli kanałów o falach milimetrowych mających zastosowanie w wielu przypadkach i środowiskach, a także w szerokim zakresie częstotliwości od 30 GHz do 90 GHz. Jednak pomimo dużej liczby eksperymentalnych kampanii pomiarowych i wyników w dzisiejszych czasach [6-11] dostępnych jest kilka praktycznych modeli kanałów fal milimetrowych. Jeden z wydanych modeli kanałów został opracowany dla standardu IEEE 802.11ad w paśmie 60 GHz (57-64 GHz). Oparty jest na pomiarach eksperymentalnych i badaniach ray tracingu i koncentruje się na ograniczonej liczbie scenariuszy wewnętrznych z parametryzacją site-specific. Najnowsze wyniki projektu METIS 2020 sugerują wykorzystanie różnych podejść do modelowania kanałów, w tym modeli stochastycznych (generycznych) i opartych na mapach (site-specific) z parametryzacją wyników pomiarów. Propagacja sygnału w pasmach niższych niż 6 GHz jest właściwie badana; wiele dokładnych i realistycznych podejść do modelowania stanowi podstawę zarówno do oceny powiązań, jak i poziomu systemu. Pasma milimetrowe musi zostać dokładnie zbadane pod kątem komunikacji bezprzewodowej, ponieważ 10-krotny wzrost częstotliwości nośnej i 100x wzrost szerokości pasma prowadzi do jakościowych zmian właściwości propagacyjnych. Po pierwsze, krótka długość fali powoduje znacznie wyższą stratę propagacji zgodnie z równaniem Friisa. Aby temu zaradzić, wymagane są anteny o wysokim zysku. Aby je wesprzeć, model kanału powinien uwzględniać właściwości przestrzenne promieni kanału w TX i RX, a także obejmować szeroki zakres technologii anten adaptacyjnych, od transmisji wiązki RF do przetwarzania pasma MIMO w paśmie podstawowym. Po drugie, jak potwierdzają liczne prace [7-9], kanał propagacji 60 GHz ma charakter quasi-optyczny. Frakcja mocy docierającej do odbiornika z powodu dyfrakcji i transmisji przez obiekty jest praktycznie nieprzydatna. Większość mocy transmisji jest propagowana pomiędzy nadajnikiem i odbiornikiem za pośrednictwem LOS oraz promieni odbitych pierwszego i drugiego rzędu. Aby ustanowić łącze komunikacyjne, muszą być używane sterowalne anteny kierunkowe, skierowane w kierunku ścieżki LOS (jeśli jest dostępna) i / lub jednej lub więcej najsilniejszych odbitych ścieżek. Dodatkową konsekwencją quasi-optycznej propagacji jest to, że model kanału powinien być w pełni trójwymiarowy, biorąc pod uwagę propagację sygnału w rzeczywistym środowisku. Przykładowo, ray tracing oparty na mapach może być skutecznym środkiem do prognozowania przestrzennych i czasowych właściwości kanału i może być wykorzystywany do wspomagania modelowania kanału. Po trzecie, należy zauważyć, że przy idealnych odbiciach każda ścieżka propagacji obejmowałaby tylko pojedynczy promień. Jednakże, jak wykazano w badaniach eksperymentalnych, każda odbita ścieżka faktycznie może składać się z szeregu promieni rozmieszczonych blisko siebie w czasie i domenach kątowych z powodu chropowatości i złożonej struktury powierzchni odbijających. W związku z tym podejście klastrowe ma bezpośrednie zastosowanie do modeli kanałów dla systemów wewnętrznych i zewnętrznych fal milimetrowych z każdym klastrem modelu odpowiadającym ścieżce odbitej LOS lub NLOS. Po czwarte, innym ważnym aspektem propagacji fal milimetrowych jest polaryzacja. Jak wykazały badania eksperymentalne, degradacja mocy jest głównie spowodowana niedopasowaniem polaryzacji między antenami. Współczynnik dyskryminacji crosspolaryzacyjnej (XPD) spowodowany odbiciami na ścieżce propagacji może wynosić około -15 do -20 dB. Na koniec, bezpośrednie pomiary eksperymentalne wykazały, że

nawet małe ruchy TX lub RX tylko na kilka centymetrów mogą znacząco zmienić funkcję transferu kanału w dziedzinie częstotliwości. Podsumowując, model kanału 3D dla pasm fal milimetrovych powinien w szczególności:

- Zapewnić dokładną charakterystykę czasoprzestrzeni w kanale propagacji (podstawowe wymaganie) dla głównych modeli użytkowania;
- Wsparcie formowania wiązki sterowanych anten kierunkowych po obu stronach TX i RX bez ograniczeń w technologii antenowej;
- Uwzględnienie charakterystyki polaryzacji anten i sygnałów;
- Obsługa niestacjonarnych charakterystyk kanału propagacji.

Podobne zadania zostały już rozwiązane podczas standardu IEEE 802.11ad (WiGig) opracował zestaw modeli kanałów wewnętrznych z dokładną parametryzacją modeli. Jednak dalszy rozwój zarówno technologii fal milimetrovych, jak i infrastruktury sieci mobilnych otwiera wiele nowych scenariuszy i przypadków użycia, zarówno wewnątrz, jak i na zewnątrz. W projekcie MiWEBA (Milimeter-Wave Backhaul and Access) heterogeniczna sieć małych komórek o milimetrovej fali jest uważana za podstawę dla przyszłych systemów komunikacyjnych 5G. Odpowiednie scenariusze dostępu, połączeń międzysieciowych i połączeń urządzenie-urządzenie (D2D) zostały opracowane dla scen plenerowych, kanionów ulicznych i hal hotelowych. Nadchodzący nowy standard IEEE 802.11ay (NG60), uważany za rozszerzenie istniejącego standardu 802.11ad, wprowadził nowe aplikacje i przypadki użycia, w tym rozładowanie mobilne, bezprzewodowe backhauling, systemy masowej dystrybucji wideo, wirtualną i rozszerzoną rzeczywistość itp. duża liczba scenariuszy i przypadków użycia sprawia, że tradycyjne podejścia (stochastyczne, oparte na geometrii, stochastyczne, site-specific deterministic, map-based) do modelowania kanałów są niepraktyczne ze względu na wymagania bardzo dużej ilości pomiarów eksperymentalnych lub rozległych symulacji ray tracingu dla każdego scenariusza. Potrzebne jest nowe podejście do modelowania kanałów w pasmach fal milimetrovych dla nowych środowisk wewnętrznych i zewnętrznych, w oparciu o eksperymenty i teoretyczne przybliżenia.

### **3.2 Podejście do modelowania kanałów Q-D**

Stan techniki do charakteryzowania kanałów komunikacji mobilnej obejmuje oddzielne opisy modeli utraty ścieżki i przestrzenno-czasowych charakterystyk kanału, zazwyczaj składających się z odpowiedzi impulsowych w kanale klastra i statystyki rozkładu kąтового. Najnowsze prace dotyczące modeli kanałów fal milimetrovych również stosują takie podejście i wykorzystują różne techniki analizy skupień do eksperymentalnego przetwarzania danych. Takie podejście sprawdza się jednak tylko w warunkach NLOS, co nie jest głównym przypadkiem użycia dla systemu komunikacji fal milimetrovych. Jednocześnie utrata propagacji cech fal milimetrovych prowadzi do osłabienia dalekich odbić i dominacji promieni bliskich bezpośredniej ścieżce. Wymaga to nowych podejść do scharakteryzowania niestacjonarnego kanału i efektów Dopplera. Aby zapewnić odpowiednie modelowanie wspomnianych wyżej aspektów propagacji kanału, w niniejszej pracy zaproponowano i rozwinięto podejście quasi-deterministyczne (Q-D) do modelowania kanałów. Należy zauważyć, że w tym rozdziale autorzy koncentrują się bardziej na samej metodologii niż na konkretnych modelach kanałów i ich parametrach. Jednak niektóre przykłady modeli kanałów Q-D zostały przedstawione tutaj dla wspólnych i najważniejszych scenariuszy. Od bardzo wczesnych prac eksperymentalnych i teoretycznych w obszarze fal milimetrovych przez pionierów Lebedeva i Bose'a ujawniono quasi-optyczną naturę propagacji sygnału fal milimetrovych. W celu praktycznej analizy systemów komunikacyjnych w tym paśmie można pominąć efekty dyfrakcji (oczywiście nadal istnieją i można je znaleźć za pomocą specjalnych pomiarów). Zgodnie z tym podejściem dla każdego ze scenariuszy

propagacji kanału, kilka najsilniejszych ścieżek propagacji (promienie, które wytwarzają istotną część odebranej użytecznej mocy sygnału) są określane jako pierwsze, a propagacja sygnału przez nie jest obliczana na podstawie geometrii rozmieszczenia, lokalizacji Stacje bazowe (BS) i sprzęt użytkownika (UE) w deterministyczny sposób. Moc sygnału dostarczana przez każdy z promieni jest obliczana zgodnie z teoretycznymi wzorami, biorąc pod uwagę straty przestrzeni swobodnej, odbicia, właściwości polaryzacji i efekty mobilności UE: przesunięcie dopplerowskie i przesunięcie użytkownika. Jednocześnie niektóre parametry w tych obliczeniach mogą być traktowane jako wartości losowe (np. Współczynniki odbicia) lub nawet jako procesy losowe (np. Ruch UE). Należy zauważyć, że liczba takich quasi-deterministycznych promieni (promienie D), które należy wziąć pod uwagę, może być różna, ponieważ w różnych scenariuszach liczba najsilniejszych promieni jest różna. Na przykład, dla scenariuszy otwartej przestrzeni zewnętrznej pokazano, że właściwości propagacji sygnału są w większości określone przez dwa promienie D - promień LOS i odbicie od ziemi. W przypadku scenariusza na zewnątrz kanionu ulicznego propagacja jest determinowana głównie przez LOS, ziemię i ściany odbijają promienie. W przypadku lobby hotelowego (kryty dostęp do dużej przestrzeni publicznej) należy wymodelować więcej promieni D. W tym scenariuszu proponuje się rozważenie wszystkich promieni z odbiciami do drugiego rzędu jako promieni D (w sposób podobny do przyjętego w metodycie oceny IEEE 802.11ad dla scenariuszy wewnętrznych). W rzeczywistym środowisku, oprócz silnych promieni D, wiele innych odbijanych fal dociera do odbiornika z różnych kierunków. Na przykład są samochody, drzewa, latarnie, ławki, daleko duże reflektory (budynki) itd. Wszystkie te promienie są uwzględniane w modelach kanału QD jako promienie wtórne lub przypadkowe (promienie R) i są opisane jako klastry losowe z określonymi parametry statystyczne wyodrębnione z dostępnych danych eksperymentalnych i szczegółowe modelowanie ray tracingu. Ta część koncentruje się na racjonalnym opisie nowego kanału metodologii modelowania i sposobie, w jaki została opracowana. Metodologia może być wykorzystana zarówno do analizy systemu i łącza na poziomie przedstawionego systemu komunikacji fal milimetrowych. Ponadto można opracować nowy model dla różnych scenariuszy przy użyciu podejścia Q-D, szczególnie w oparciu o nowe pomiary eksperymentalne. Metodologia Q-D może być używana do koordynowania kierunków kampanii pomiarowych, określania warunków eksperymentów i sprzętu do mierzenia określonych parametrów, które mogą być użyte do budowy odpowiednich modeli w ramach podejścia Q-D.

### **3.3 Pomiary eksperymentalne i interpretacja wyników dla kanałów fal milimetrowych**

Ta sekcja przedstawia eksperymentalne wyniki pomiarów dla pasma 60 GHz w środowisku zewnętrznym "w realnym świecie". Kampania pomiarowa została przeprowadzona przez Instytut Fraunhofera Heinricha Hertza (HHI) w kilku lokalizacjach w Berlinie w Niemczech oraz przez Intel / University of Nizhny Novgorod w Rosji. W poprzednich pracach wyniki podobnych pomiarów były tradycyjnie wykorzystywane do oceny wykładnika straty ścieżki propagacji i profilu opóźnienia mocy kanału. W niniejszej pracy pomiary eksperymentalne uzupełniono symulacją ray tracingu w zrekonstruowanym środowisku pomiarowym 3D. Jak pokażemy w następnych rozdziałach, to hybrydowe podejście do badania kanału, oparte zarówno na pomiarach jak i symulacjach, pozwala na dogłębną analizę mechanizmów propagacji fal milimetrowych w rzeczywistych niestacjonarnych środowiskach.

#### **3.3.1 Eksperymentalny opis pomiarów i wyniki**

##### ***Pomiary miejskich kanionów ulicznych***

Pomiary miejskich kanionów ulicznych przeprowadzono podczas godzin pracy w kanionie ulicy miejskiej, położony na Potsdamer StraÙe w centrum Berlina, Niemcy. Nowość tych eksperymentów polega na długoterminowym badaniu niestacjonarności środowiska, które charakteryzuje się

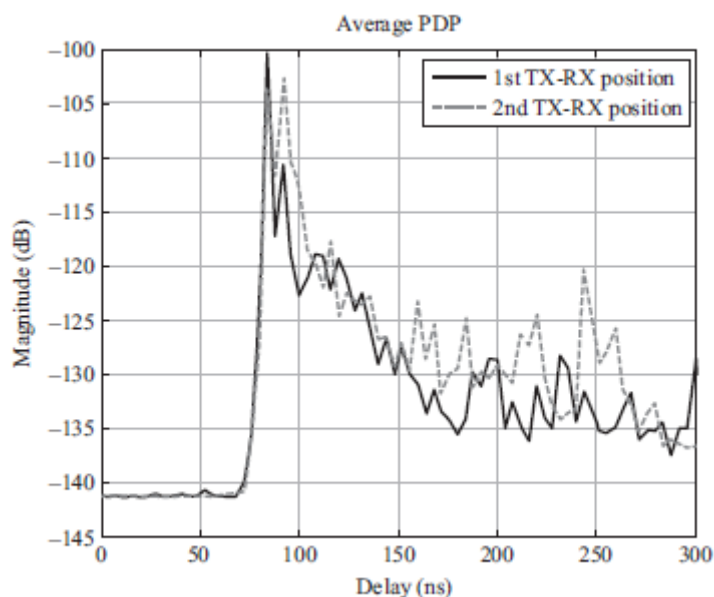


obecnością stosunkowo dużych obiektów statycznych, a także wielu dynamicznych (ruchomych) przeszkód i reflektorów typowych dla gęstego zewnętrznego obszaru miejskiego. Podczas pomiarów zastosowano anteny dookólne zarówno po stronie TX, jak i RX. Pomiar przeprowadzono głównie dla sondowania kanałów mmWave w szeregu stacjonarnych pozycji TX i RX w ustalonej odległości. Przeprowadzono także dalsze eksperymenty, ale z ruchomym RX i stacjonarnym TX. Antena TX miała wysokość 3,5 m nad ziemią, aby reprezentować typową stację bazową małej komórki, która jest dodawana do istniejących słupów lampy. RX został zamontowany na wózku mobilnym z typową wysokością anteny użytkownika 1,5 m. W pierwszym etapie wykonano wiele długoterminowych pomiarów dla stacjonarnych pozycji TX i RX oddzielonych od siebie 25 m. W drugim etapie przeprowadzono także pomiary mobilne, w których RX poruszał się ze stałą prędkością na chodniku w odległości 50 m na każdą stronę stacjonarnego TX. Sygnalizator kanału zastosowany w tej kampanii pomiarowej był oparty na specjalnie zbudowanej platformie FPGA z kluczowymi parametrami wymienionymi tu;

Wartość	parametru
Częstotliwość nośnika	60 GHz
Pasma	250 MHz
Moc wyjściowa	15 dBm
Antena zyskuje	2 dBi
Wzorzec anteny	Wszechkierunkowy (w płaszczyźnie poziomej) >? 30? połowa mocy wiązki (w elewacji)

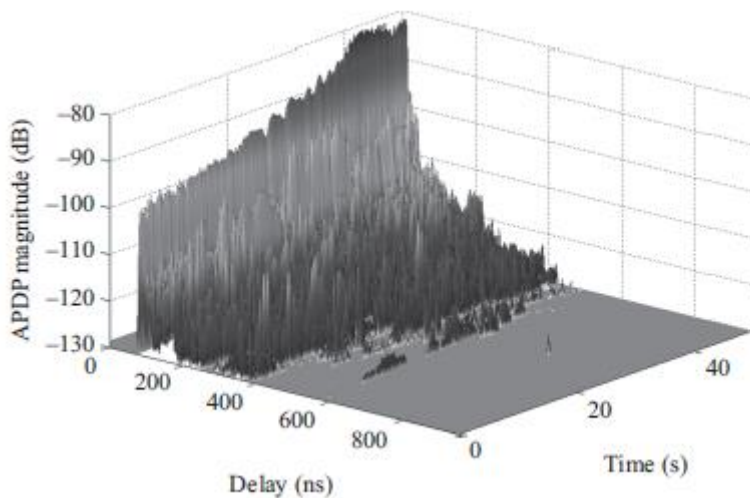
Wyjściowym sygnałem echosondy jest odpowiedź impulsowa kanału (CIR) dla każdej migawki pomiaru wykonanej co 800 ms. Podczas pierwszego etapu uzyskano 62.500 migawek CIR dla każdej statycznej pozycji pary TX i RX, co skutkuje pojedynczym okresem obserwacji wynoszącym 50 sekund.

Rysunek



pokazuje typową uśrednioną odpowiedź impulsową kanału uzyskaną z te statyczne pomiary. Opóźnienie najwyższego pików na figurze wynosi w przybliżeniu 83 ns, co odpowiada ścieżce liniowości

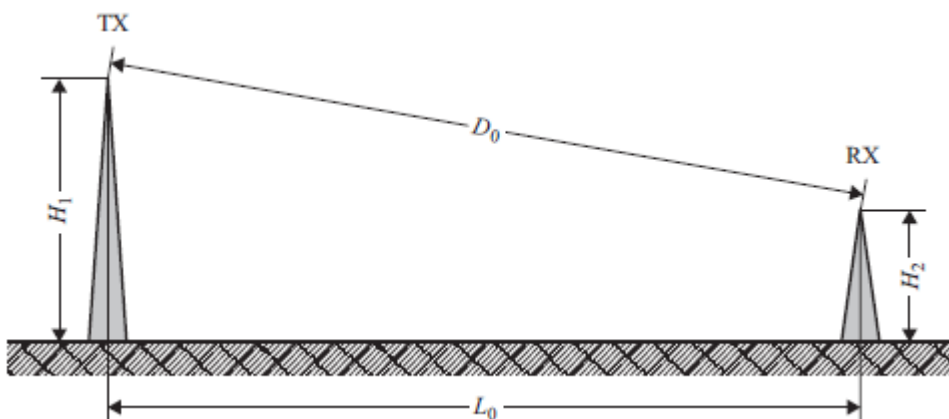
(LOS) między statycznym TX i RX oddzielnym 25 m od siebie. Z pomiarów CIR uzyskano profil opóźnienia zasilania kanału przez uśrednienie otrzymanych wartości mocy w czasie obserwacji 50 s (współczynnik uśredniania  $N = 62\ 500$ ). Ten średni profil opóźnienia zasilania (APDP) pokazano na rysunku 3.2 dla dwóch różnych pozycji pary TX i RX. Zmiana poziomu mocy pomiędzy dwoma położeniami odbiornika może być spowodowana przez jeden lub większą liczbę promieni (lub ścieżek) docierających do RX z różnych kątów przybycia, np. Pochodzących z różnych powierzchni odbijających z mniej więcej taką samą całkowitą długością ścieżki między TX a RX. Ze względu na ograniczoną szerokość pasma pomiarów (250 MHz) te wielościeżkowe komponenty nie mogą być jawnie rozdzielone i wytwarzają szybko zanikające efekty. Podczas drugiego etapu z ruchomym odbiornikiem wykonano szereg 40 serii pomiarowych z 62 500 zdjęciami. Na podstawie uzyskanego eksperymentu Wyniki APDP można obliczyć na podstawie pewnej liczby migawek, aby usunąć efekty zanikania. Przykładowy wynik takiego przebiegu pomiarowego pokazano na rysunku



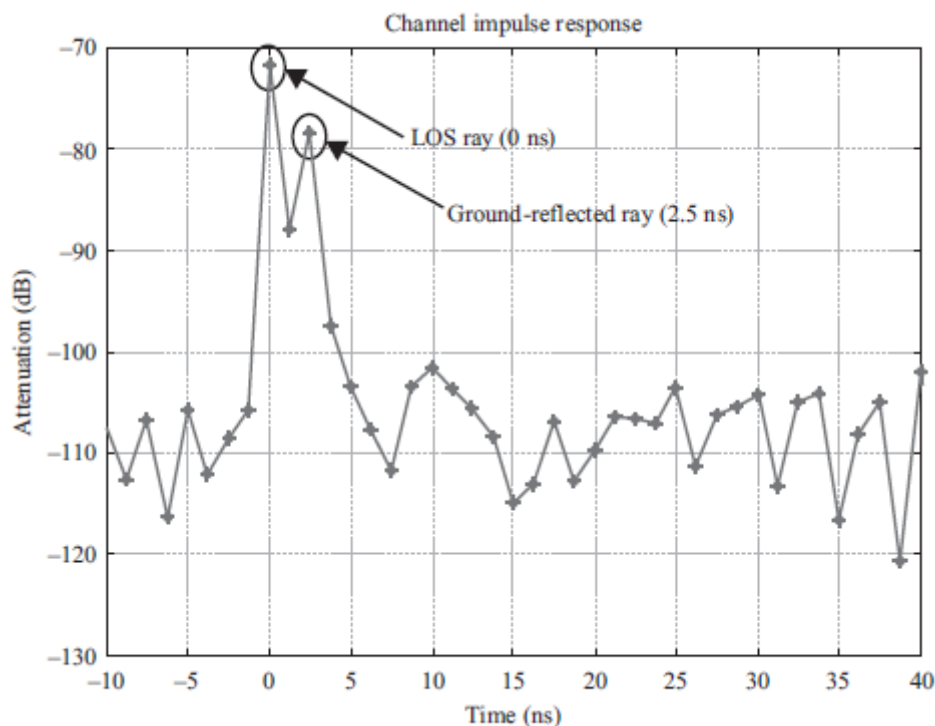
W tym pomiarze odbiornik został przesunięty w stronę nadajnika. Szybki efekt zanikania jest wyraźnie widoczny na komponentce LOS po lewej stronie wykresu.

### Pomiary w kampusie uniwersyteckim

Badanie w środowisku scenariuszowym na otwartej przestrzeni przeprowadzono w kampusie Uniwersytetu Niżnego Nowogrodu. Antena TX została zamontowana na przedsiionku budynku kampusu uniwersyteckiego na wysokości 6,2 m nad ziemią, aby reprezentować BS. Antena RX została zamontowana na ruchomej platformie o wysokości anteny wynoszącej 1,5 m, aby reprezentować mobilne UE. Schematyczna ilustracja tego scenariusza została przedstawiona na rysunku



gdzie zastosowano następujące główne parametry:  $D_0$  jest bezpośrednią odległością LOS między antenami TX i RX;  $L_0$  to pozioma odległość między nadajnikiem i odbiornikiem,  $H_1$  i  $H_2$  to odpowiednio wysokość nadajnika i odbiornika. Do eksperymentalnego pozyskiwania i przetwarzania danych wykorzystano specjalnie zaprojektowaną platformę pomiarową z kluczowymi parametrami technicznymi. W zależności od odległości  $L_0$  nadajnik był wyposażony w anteny o różnym wzmacnieniu. Dla odległości mniejszych niż 35 m prostokątna antena tubowa 14 x 18 mm osiągnięto 19,8 dBi, natomiast dla większych odległości zastosowano wysokorozdzielczą antenę obiektywu z dużą aperturą (100 mm) i zyskiem 34,5 dBi (patrz rys. 3.4). Odbiornik został wyposażony w antenę tubową o średnicy  $d = 14$  mm i zysku 12,3 dBi. Należy zauważyć, że szerokość pasma sygnału dźwiękowego użyta w tych eksperymentach była równa 800 MHz, co zapewniało maksymalną rozdzielczość czasową 1,25 ns. Zastosowanie anten kierunkowych zarówno w TX, jak i RX pomogło w odkryciu, że rozważany scenariusz eksperymentalny charakteryzował się tylko dwoma najsilniejszymi ścieżkami propagacji (lub promieniami). Pierwsza ścieżka odpowiada składnikowi LOS i drugiej ścieżce do odbicia od powierzchni gruntu (asfaltu). Wszystkie inne ścieżki propagacji wywołane odbiciami od otaczających obiektów były o ponad 15-20 dB niższe od tych dwóch najsilniejszych składników. Przykład mierzonej odpowiedzi impulsowej kanału z dwoma wyraźnie wyróżniającymi się pikami przedstawiono na rysunku

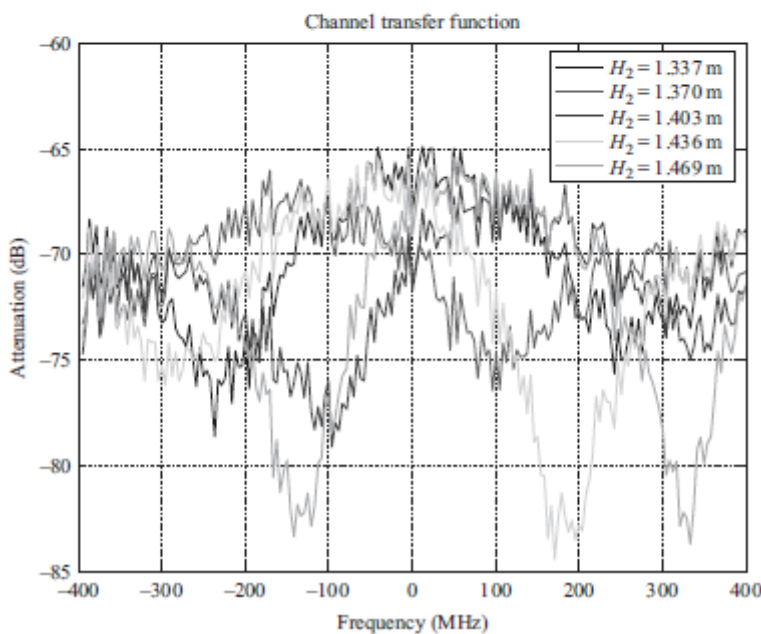
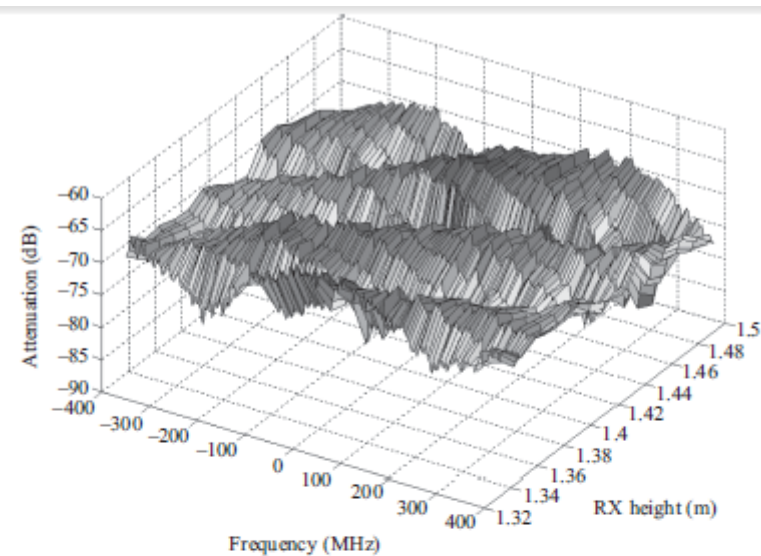


Piki odpowiadają LOS, a składniki odbijające światło od ziemi znajdują się 2,5 n od siebie.

Parametr	Wartość
Częstotliwość	60 GHz
Pasma	800 MHz
Moc wyjściowa	2,4 dBm
Czułość platformy	-75 dBm
Anteny TX	Obiektyw: $d = 100$ mm / Horn: 14 x 18 mm

Anteny TX	zysk 34,5 dBi / 19,8 dBi
Anteny TX HPBW	3° / 14° -18°
Antena RX	Okrągły horn d = 14 mm
Antena RX	zysk 12,3 dBi
Antena RX HPBW	30°

Różnica mocy między tymi dwoma wartościami szczytowymi jest w przybliżeniu równa współczynnikowi odbicia gruntu (-6 dB). Początkowo badano zależność funkcji przenoszenia kanału (CTF) od ruchu pionowego RX dla odległości  $L_0 = 30,6$  m. Na rysunku 3.7 przedstawiono szerokość pasma CTF o szerokości 800 MHz, gdy wysokość RX zmienia się z 1,34 m na 1,5 m. Kilka oddzielnych próbek CTF przedstawiono na rysunku

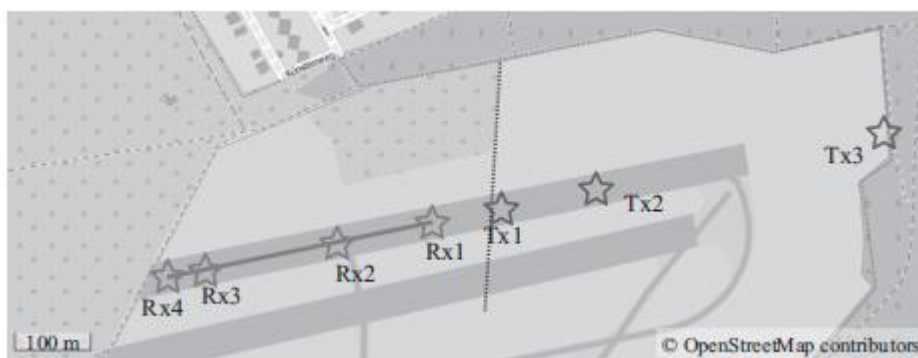


W tych eksperymentach zastosowano konfigurację polaryzacji anteny w poziomie / poziomie (H-H) zarówno w antenach TX, jak i RX. Wykonano również pomiary dla innych konfiguracji polaryzacji, ale użycie konfiguracji H-H pozwoliło nam wykluczyć dodatkowe efekty związane z kątem uderzenia

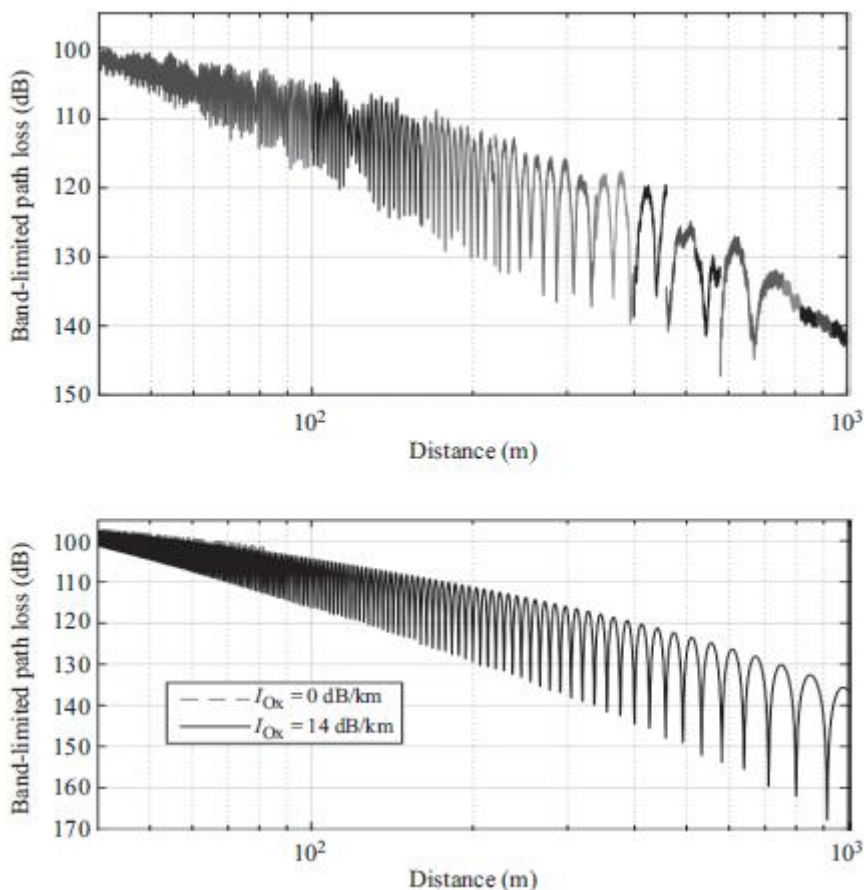
Brewstera. Jak pokazano na rysunkach, kanał ma selektywność wysokoczęstotliwościową w paśmie 800 MHz. Jednak najważniejszym wynikiem tych wyników są bardzo szybkie zmiany kanału milimetrowego, gdy zmienia się wysokość RX. Tylko RX poruszający się w kierunku pionowym o 2-3 cm powoduje znaczne zmiany kanału. Stosunek polaryzacji krzyżowej (XPR) mierzono również dla promienia odbitego od ziemi w rozważanym scenariuszu doświadczalnym. W tym celu w nadajniku i odbiorniku zastosowano ortogonalne konfiguracje polaryzacji anteny. Zgodnie z uzyskanymi wynikami eksperymentalnymi wartość XPR była mniejsza niż -25 dB. Szerokopasmowe pomiary eksperymentalne z antenami kierunkowymi ujawniły drobną strukturę kanału fal milimetrowych w otwartym środowisku otwartym. Zaobserwowano, że typowa mobilna antena użytkownika otrzyma co najmniej dwa promienie (bezpośredni promień LOS i promień odbitego podłoża) z raczej małym opóźnieniem czasowym (2,5 ns dla 30 m). Należy zauważyć, że wraz ze wzrostem odległości między miejscami TX i RX, opóźnienie czasowe i kąt przybycia między tymi promieniami maleją. Interferencja promieni bezpośrednich i odbitych od gruntu uzasadnia szybkie efekty zanikania obserwowane w eksperymentach z dookólną anteny na początku bieżącego odcinka i powinny być uwzględnione w metodologii modelowania kanałów.

### Badanie refleksji gruntu

Jak wykazano w poprzedniej sekcji, odbicie gruntu ma duży wpływ na propagację sygnału w środowisku zewnętrznym. Dawne lotnisko w Berlinie Gatow zapewnia płaską i jednorodną powierzchnię asfaltu dla nadajników i odbiorników do 1000 m, jak pokazano



Sygnalizator kanału był taki sam jak wcześniej opisany. Migawki kanału pomiarowego są pobierane w równych odległościach od RX. Dla każdej migawki (odpowiedź impulsowa kanału) obliczana jest chwilowa utrata ścieżki. Rysunek

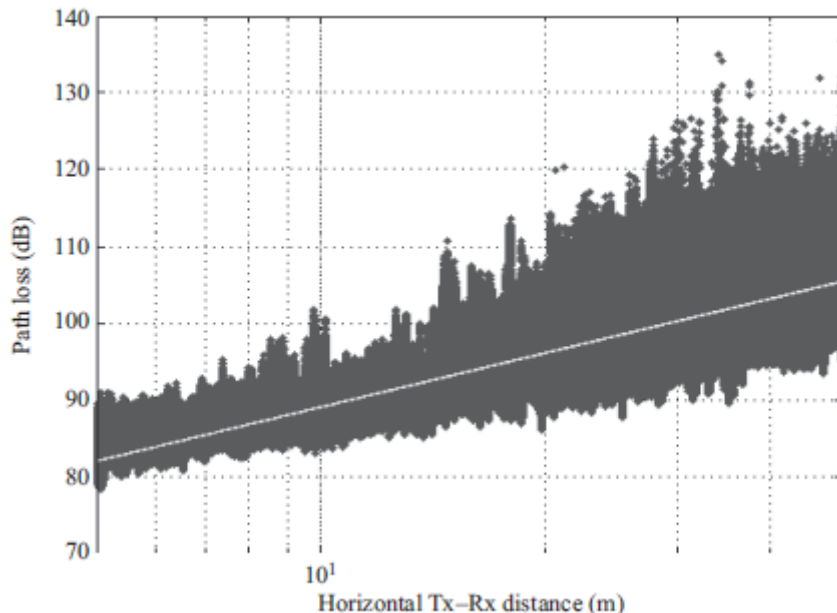


pokazuje zmierzoną stratę ścieżki w zakresie od 40 do 1000 m na asfalcie (nadajnik w odbiornikach Tx1, Tx2, Tx3; odbiornik przemieszczany między Rx1 i Rx4, polaryzacja pozioma). Filtr średniej ruchomej o szerokości 20 próbek został zastosowany w celu zmniejszenia efektów szumów. Odległość 60 m jest objęta jednym przebiegiem pomiarowym. Rysunek pokazuje zatem kombinację 16 kolejnych serii pomiarów. Niektóre artefakty można zaobserwować na przejściach między poszczególnymi przebiegami pomiarowymi. Może to być spowodowane niewielkimi zmianami wysokości anteny nad podłożem i błędami pozycjonowania, gdy przetwornik został przeniesiony z Tx1 na Tx2 i Tx3. Wyniki pomiarów polaryzacji pionowej nie są tutaj pokazane, ale wykazują takie samo zachowanie. Do modelowania propagacji stosuje się proste podejście dwupromowe. Linia drogi widzenia (LOS) i pojedyncza droga odbicia od podłoża są obliczane w sposób polarymetryczny. Właściwości odbicia gruntu są modelowane zgodnie z prawami odbicia Fresnela, a absorpcja tlenu jest brana pod uwagę. Dyskretna odpowiedź impulsowa jest dzielona i obliczana na ograniczoną pasmo kanału odpowiedź impulsową i obliczana jest utrata ścieżki, jak to było zrobione z danymi pomiarowymi. Rysunek 3.9 pokazuje utratę ścieżki w zakresie od 40 do 1000 m przy szerokości pasma 250 MHz dla szybkości absorpcji tlenu 0 dB i 14 dB na km. Ta ostatnia wartość dobrze pasuje do pomiaru i symulacji. Jest to zgodne z modelem MPM, który zapewnia tłumienie na poziomie 13,9 dB na km parametrów środowiskowych na dzień pomiaru. Porównanie zmierzonej i symulowanej utraty ścieżki pokazuje, że liczba i odległość zanikających otworów są ogólnie dobrze wyrównane. Jednak w pomiarach występują dodatkowe zakłócenia. Główną obserwacją z porównania wyników pomiarów z modelem z dwoma kanałami jest bardzo dobre wyrównanie modelu z danymi eksperymentalnymi na odległościach do 500 m. Brak efektów zanikania w odległości większej niż 500 m wymaga dalszych badań. Zanikanie może powodować znaczny wzrost utraty ścieżki w pewnych pozycjach, co wyraźnie widać na podstawie pomiarów i symulacji. W związku z tym efekt zaniku spowodowany odbiciem gruntu powinien być brany pod uwagę również w przypadku dalekodystansowych połączeń międzymiastowych.

### 3.3.2 Interpretacja wyników eksperymentalnych

#### *Pomiary miejskich kanionów ulicznych*

W badaniu strat ścieżek w kanionie ulicy miejskiej wykonano migawki pomiarów za pomocą anten dookólnych, jak opisano wcześniej. Przy obliczaniu utraty ścieżki na tej podstawie brane są pod uwagę wszystkie możliwe ścieżki propagacji między nadajnikiem i odbiornikiem. Na rysunku



strata ścieżki dla wszystkich odległości od 5 do 50 m jest przedstawiona jako wykres rozproszenia. Linia ciągła przedstawia dopasowanie liniowe najmniejszych kwadratów (LS) odpowiadające średniej utracie ścieżki. Parametry zidentyfikowane dla tego modelu utraty ścieżki można znaleźć tu

Parametr	Wartość
Odległość odniesienia $d_0$	5m
$L_{PL}(d_0)_{dB}$	82
n	2,36

Najmniejsze dopasowanie kwadratowe przeprowadzono na modelu utraty ścieżki ze stałą utratą ścieżki referencyjnej w odległości odniesienia:

$$L_{PL}(d_0)_{dB} = L_{PL}(d_0)_{dB} + 10n \log_{10}(d/d_0)$$

Uzyskana wartość strat ścieżki jest zbliżona do wykładnika straty ścieżki wolnej przestrzeni równej 2. Wskazuje to, że propagacja linii wzroku jest dominująca w tym środowisku. Wysoką wariancję utraty ścieżki można jednak zidentyfikować jako powodującą zanikanie z odbicia od podłoża i ściany bocznej

#### **Wykresy bitmapowe z opóźnieniem czasowym**

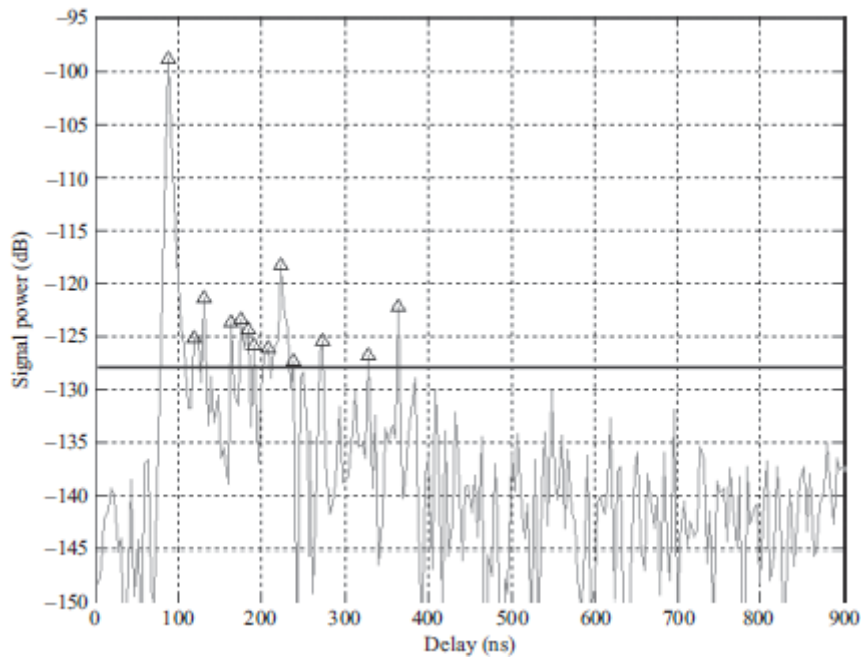
Przybliżenie utraty ścieżki w tabeli 3.4 może dać wstępny wgląd w cechy kanału milimetrowego. Niemniej jednak dogłębne zbadanie struktury kanału wymaga analizy odpowiedzi impulsowej kanału zarówno z perspektywy statystycznej, jak i deterministycznej. Aby ocenić właściwości kanału fal

milimetrowych, zmierzone odpowiedzi impulsowe kanału zostały przetworzone przez poniższy prosty algorytm progowej detekcji pików ,zgodnie z którym:

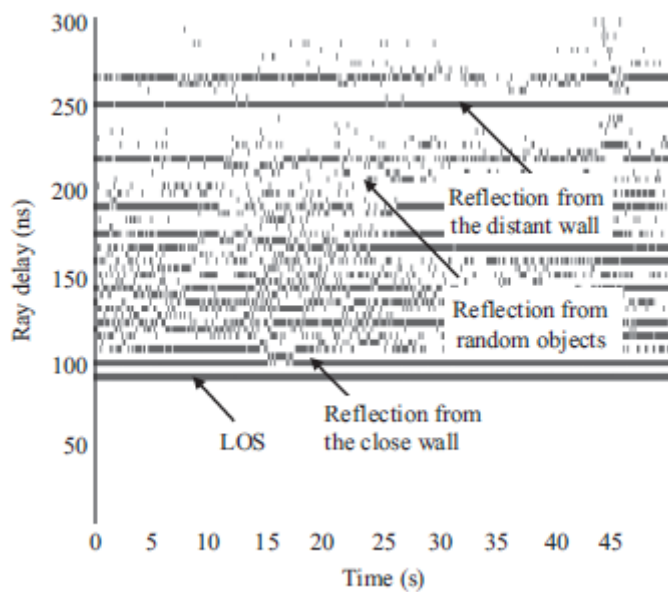
Punkt w profilu opóźnień mocy  $P(t_k)$  jest identyfikowany jako szczyt, jeżeli:

$$\begin{cases} P(t_{k-1}) < P(t_k) > P(t_{k+1}) \\ P(t_k) > P_{noise} + 10dB \end{cases}$$

Piki odpowiadające najsilniejszym promieniom lub wielościżkowym komponentom (MPC) rozdzielone przez echosondę pokazano na rysunku



Do analizy zachowania kanału w czasie, opóźnień wiązki a obserwacja zastosowano schemat bitmapowy. Przykład takiego diagramu przedstawiono na tu

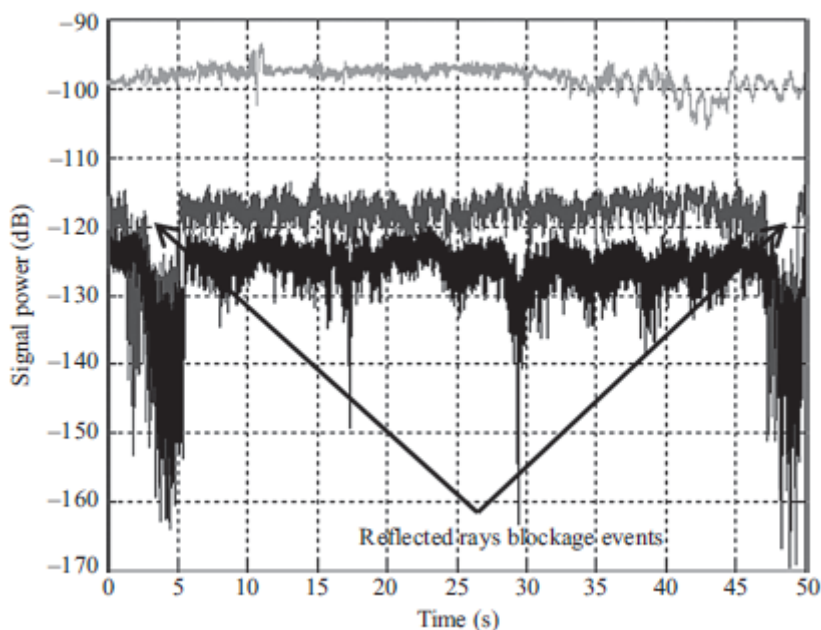




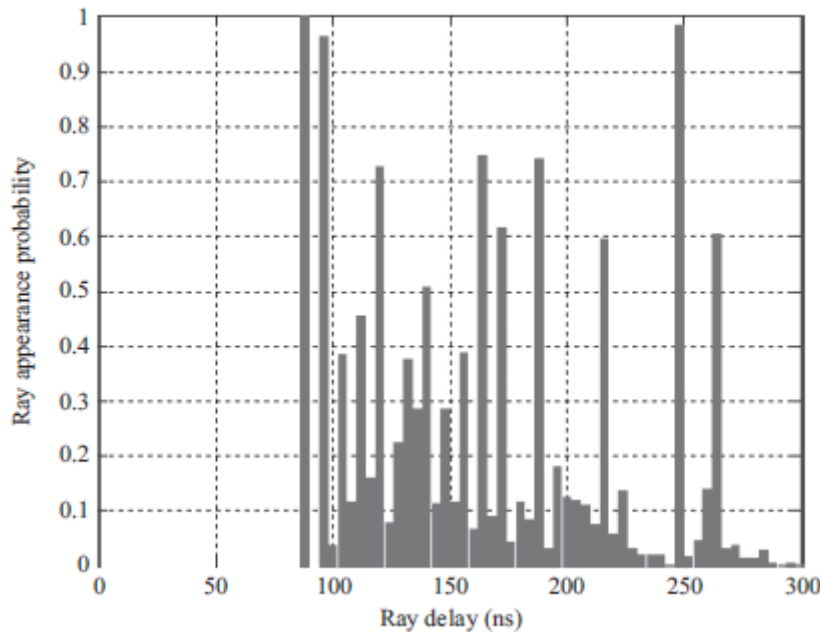
, gdzie pojawienie się / brak zidentyfikowanych najsilniejszych promieni są wskazane dla całkowity czas obserwacji 50 s. Warto zauważyć, że w każdym pomiarze promień LOS osiąga 83 ns, co odpowiada opóźnieniu spowodowanemu przez 25 m separację między TX i RX. Na ostatnim rysunku widać również, że istnieją inne bardzo stabilne promienie. Promienie te mogą być związane z odbiciami od dużych obiektów statycznych, takich jak ściany budynków i pawilony przystanków autobusowych. W tym samym czasie niektóre promienie pojawiają się losowo i znikają z powodu małej mocy i / lub blokady. Promienie te mogą być związane z odbiciami od odległych obiektów i blisko położonych elementów dynamicznego środowiska ulicznego. Dla ruchomych pomiarów z ruchomym RX wykonano 40 serii pomiarowych z 62 500 zdjęciami. W związku z tym odpowiedni zestaw danych obejmuje 2,5 mln CIR. Ponieważ RX został przesunięty o ponad 25 m podczas każdego biegu, uzyskano dokładne próbkowanie przestrzenne (odstęp 0,4 mm) na całej długości kanionu ulicznego.

### Klasyfikacja promieni

W dynamicznych środowiskach (ulicach miast, placach, centrach handlowych itp.) Nawet bezpośredni promień LOS i stałe odbite promienie są podatne na blokowanie. Przykład odbitych zdarzeń blokujących promieniowanie zarejestrowanych podczas eksperymentalnej kampanii pomiarowej w środowisku miejskiego kanionu ulicy można zobaczyć na wykresie mocy komponentów wielościeżkowych



Efekty blokowania mogą być analizowane statystycznie za pomocą wykresów bitmapowych z opóźnieniem czasowym danych pomiarowych uzyskanych z wielu eksperymentów. Procent "aktywności promieniowej" można oszacować na podstawie tych wykresów. Przyjmując własności ergodyczne procesu losowego blokowania, procent aktywności w czasie można wykorzystać jako oszacowanie prawdopodobieństwa blokady w zespole statystycznym. Rysunek poniższy pokazuje wykres słupkowy aktywności promieni w scenariuszu pomiarów Ulicznych Kanionów (pozycja blisko ściany). Rysunek



pozwała na klasyfikację promieni w trzech różnych grupach:

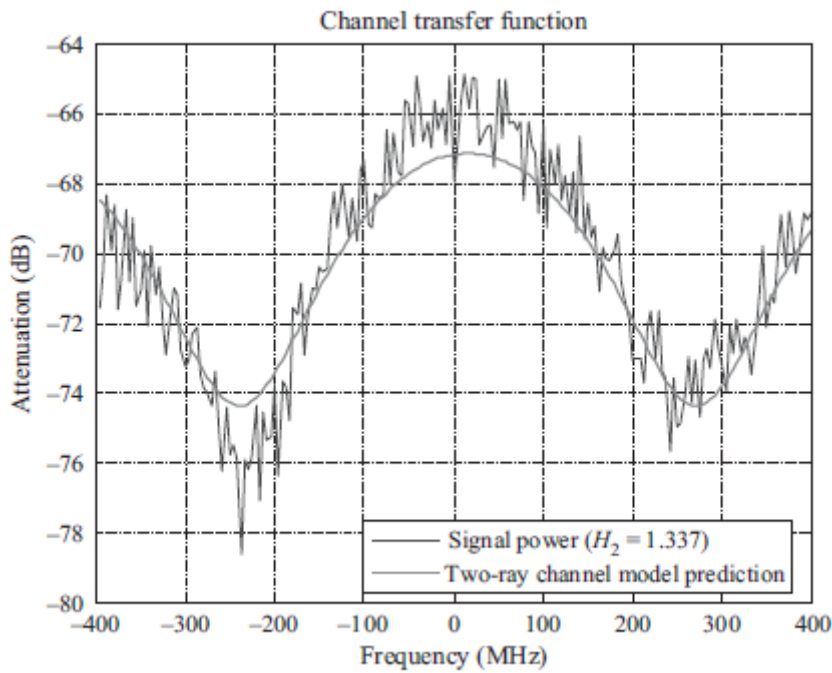
- Promieniami o aktywności powyżej 80% -90% są promienie D: silne i zawsze obecne, jeśli nie zablokowane. Procent zablokowania dla promieni D można oszacować na około 2% -4%
- Promieniami o aktywności wynoszącej około 40% -70% są promienie R: odbicia od odległych obiektów statycznych, słabsze i bardziej podatne na blokadę ze względu na dłuższy dystans podróży.
- Promienie o wartości aktywności poniżej 30% to promienie F: migające odbicie od przypadkowych poruszających się obiektów. Takie promienie nie są "zablokowane", "faktycznie" pojawiają się "" przez krótki czas.

Pierwszy typ promieni w znacznym stopniu przyczynia się do mocy sygnału, ponieważ jest obecny przez cały czas i zwykle można go wyraźnie zidentyfikować jako odbicia od ważnych obiektów makro scenariusza. Logiczne jest włączenie ich do modelu kanału jako deterministyczne (promienie D), wartości wyliczone jawnie. Element losowości, ważne dla modelowania kanału statystycznego, mogą być wprowadzane na poziomie wewnątrz klastra, poprzez dodanie losowego klastra rozkładającego się wykładniczo do głównego D-ray. Drugi rodzaj promieni (promienie R) to odbicia od przypadkowych obiektów lub obiektów, które nie są obowiązkowe w środowisku scenariusza. Tego rodzaju promienie mogą być włączone do modelu w klasyczny statystyczny sposób, jako promienie o parametrach (moc i opóźnienia) wybrane losowo zgodnie ze wstępnie zdefiniowanymi rozkładami. Trzeci rodzaj promieni może zostać wprowadzony do modelu w taki sam sposób, jak promienie R, ale z dodatkowym modelowaniem stochastycznym, aby uwzględnić prawdopodobieństwo pojawienia się i czasu trwania.

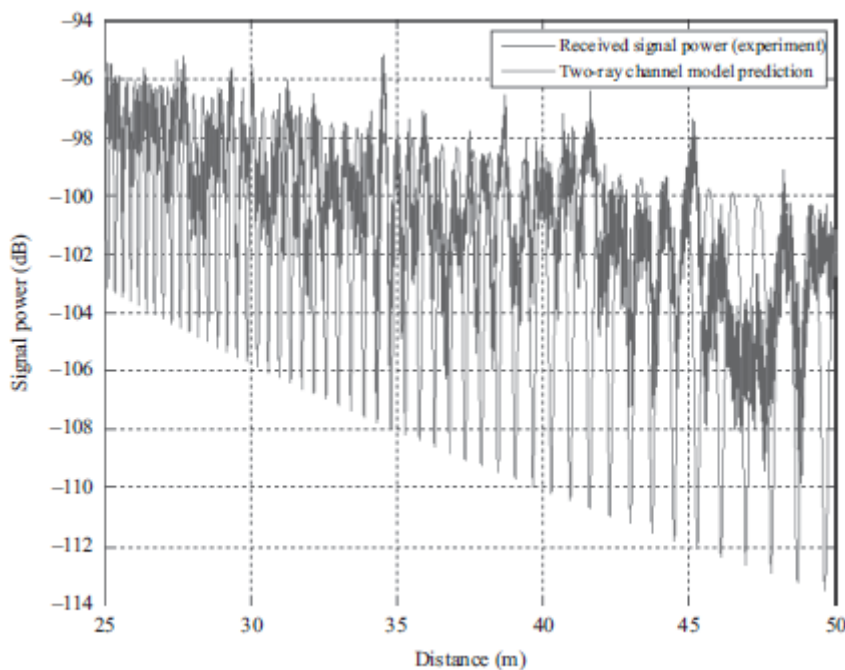
### Uprozczone przybliżenie dwóch promieni

W sekcji 3.3.2 wykazano, że model kanału dwuramiennego złożony z modelu LOS i odbicie od ziemi zapewniają bardzo dokładny opis kanału fal milimetrowych dla odległości do 500 m w "czystym" środowisku otwartym (pole lotniska). W tej sekcji pokażemy, że prosty model z dwoma promieniami można wykorzystać do przybliżonego opisu bardziej złożonych i wielodrożnych środowisk. Pierwszy

przykład, który pokazuje przewidywalną dokładność modelu dwukanałowego dla bardziej złożonego środowiska otwartego obszaru (kampusu uniwersyteckiego), przedstawiono tu



Na rysunku pokazano jedną z funkcji pomiaru transmitancji kanału (CTF) i odpowiadające jej obliczenia oparte na modelu dwupromowym. Odchylenie standardowe błędu wynosi tylko około 2 dB. Rysunek



przedstawia się sygnału szybkiego obniżania wartości dla wyników eksperymentalnych pomiarów w kanionie ulicy miejskiej w porównaniu z przybliżeniem modelu dwukanałowego. Na podstawie tych danych można zauważyć, że nawet w tym środowisku wielościeżkowym obserwujemy dobrą zgodność między modelem a danymi eksperymentalnymi. Należy zauważyć, że przy odległościach większych niż 25 m promień odbitego podłoża jest prawie tak silny, jak promień bezpośredni z powodu suwania się

względem powierzchni ulicy. Interferencja między dwoma promieniami D tworzy duże quasi-okresowe zanikające odstępów obserwowane na rysunku 3.16. Co więcej, głębokość zanikania można wykorzystać do przybliżonego oszacowania strat odbicia i mocy wszystkich innych promieni resztkowych. Bardzo dobra zgodność prostego modelu z dwoma promieniami z wynikami eksperymentalnymi w różnych środowiskach pomiarowych na zewnątrz dowodzi, że kanał milimetrový może być bardzo dobrze opisany przez kilka dominujących promieni. Wspólna analiza eksperymentalnych danych pomiarowych i ocen teoretycznych sugeruje przydatność do quasi-deterministycznego (Q-D) podejścia do modelowania kanałów milimetrových. Połączenie deterministycznego podejścia do oceny najsilniejszych parametrów promienia z opisem statystycznym losowych składników jest kluczem do spełnienia wymagań nowego modelu fal milimetrových.

### 3.4 Quasi-deterministyczne podejście do modelowania kanałów fal milimetrových

Wyniki eksperymentalne przedstawione w poprzednim rozdziale potwierdzają quasi-optyczny charakter propagacji fal milimetrových. Kanał komunikacji może być reprezentowany jako zestaw promieni o różnych opóźnieniach i ścieżkach propagacji. Co więcej, analiza danych eksperymentalnych wykazała, że generalnie promienie te można podzielić na trzy główne grupy: deterministyczne promienie D, losowe promienie R i migające promienie F. Parametry i właściwości promieni w każdej grupie można uzyskać na podstawie pomiarów eksperymentalnych i analizy ray-tracingu. Promienie deterministyczne (promienie D) w modelu odpowiadają podejściom opartym na mapowaniu kanałów, z jedną różnicą, że metodologia QD nie wymaga dokładnej mapy środowiska do rekonstrukcji, ponieważ tylko najsilniejsze i zawsze obecne promienie są uważane za D -rys w tej metodologii. Promienie R odpowiadają stochastycznemu podejściu do modelowania kanałów i biorą pod uwagę pozostałe promienie, które trudno jest zliczyć bezpośrednio ze względu na złożoność środowiska i czynniki losowe. Promienie F przedstawiają silne ścieżki, które pojawiają się przez krótki czas z powodu chwilowych odbić od przejeżdżających samochodów, otwierania drzwi i innych szybko poruszających się obiektów.

#### 3.4.1 Ogólna struktura modelu kanału Q-D

Zgodnie z dokumentem modelowania kanałów w standardzie 802.11ad ogólny opis odpowiedzi impulsowej kanału klastrowego (bez uwzględnienia polaryzacji) można podać:

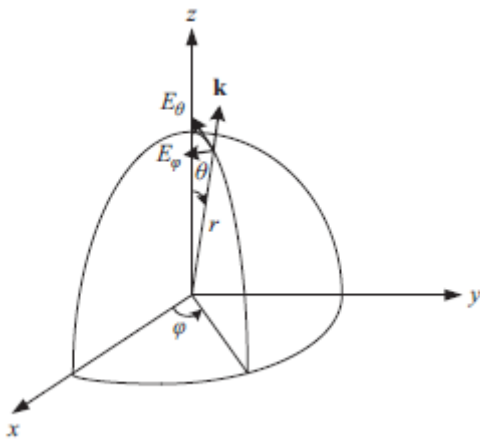
$$h(t, \varphi_{tx}, \theta_{tx}, \varphi_{rx}, \theta_{rx}) = \sum_i A^i C^i [t - T^i, \varphi_{tx} - \Phi_{tx}^i, \theta_{tx} - \Theta_{tx}^i, \varphi_{rx} - \Phi_{rx}^i, \theta_{rx} - \Theta_{rx}^i]$$

$$C^i(t, \varphi_{tx}, \theta_{tx}, \varphi_{rx}, \theta_{rx}) = \sum_k a^{i,k} \delta(t - \tau_k) \delta(\varphi_{tx} - \varphi_{tx}^{i,k}) \delta(\theta_{tx} - \theta_{tx}^{i,k})$$

$$\times \delta(\varphi_{rx} - \varphi_{rx}^{i,k}) \delta(\theta_{rx} - \theta_{rx}^{i,k})$$

gdzie  $h$  jest wygenerowaną całkowitą funkcją odpowiedzi impulsowej kanału,  $t$  oznacza aktualny czas;  $\varphi_{tx}, \theta_{tx}, \varphi_{rx}, \theta_{rx}$  są kątami azymutu i elewacji odpowiednio nadajnika i odbiornika;  $A^i$  i  $C^i$  są odpowiednio funkcją wzmocnienia i kanału odpowiedzi impulsowej dla  $i$ th klastra;  $\delta(\cdot)$  jest funkcją delta Diraca i kanału odpowiedzi impulsowej dla  $i$ -tego klastra;  $T^i, \Phi_{tx}^i, \Theta_{tx}^i, \Phi_{rx}^i, \Theta_{rx}^i$  i są współrzędnymi czasowo-kątowymi  $i$ -tego klastra;  $a^{i,k}$  to amplituda  $k$ -tego promienia  $i$ -tego klastra;  $\tau_k, \varphi_{tx}^{i,k}, \theta_{tx}^{i,k}, \varphi_{rx}^{i,k}, \theta_{rx}^{i,k}$  są względnymi współrzędnymi czasowo-kątowymi  $k$ -tego promienia  $i$ -tego klastra. Proponowany model kanału przyjmuje podejście grupowania z każdym klastrem składającym się z kilku promieni blisko rozmieszczonych w domenach czasowych i kątowych. W rzeczywistym środowisku czas i parametry kątowe różnych wiązek i promieni są funkcjami zmieniającymi się w czasie z powodu niestacjonarności. Jednak tempo tych zmian jest stosunkowo powolne. Model kanału fal milimetrových powinien również wspierać właściwości polaryzacyjne anten i ścieżek propagacji. Polaryzacja fal

elektromagnetycznych opisuje orientację pola elektrycznego  $E$  i orientację natężenia magnetycznego w przestrzeni i czasie. Wektor pola magnetycznego można zawsze jednoznacznie znaleźć, jeśli orientacja wektora pola elektrycznego  $E$  i kierunek propagacji są znane. Tak więc właściwości polaryzacji są zwykle opisane tylko dla wektora pola elektrycznego  $E$ . W obecnej metodologii modelowania kanałów będziemy podążać za podejściem do symulacji efektów polaryzacji, która została z powodzeniem zastosowana w metodologii modelowania kanałów 802.11ad Polaryzacja emitowanego sygnału radiowego jest określana przez właściwości polaryzacji anteny. W strefie pola dalekiego pola EM emitowanego przez antenę, wektor elektryczny  $E$  jest funkcją kierunku promieniowania (zdefiniowanego przez kąt azymutu  $\psi$  i kąt elewacji  $\theta$  w układzie współrzędnych odniesienia). Ilustrację transmitowanego wektora  $E$  w strefie dalekiego pola pokazano na rysunku



Wektor  $E$  jest prostopadły do kierunku propagacji  $k$  i może być rozłożony na dwa składowe ortogonalne:  $E_\theta$  i  $E_\phi$ , które należą odpowiednio do płaszczyzn stałych  $\psi$  i stałych kątów  $\theta$ . Znajomość  $E_\theta$  i  $E_\phi$  emitowanego sygnału (które mogą być funkcjami  $\psi$  i  $\theta$ ) w pełni opisuje charakterystykę polaryzacji anteny w strefie dalekiego pola. Kierunek polaryzacji może być opisany przez znormalizowany wektor  $e$ , tak zwany wektor Jonesa. Polaryzację fal można opisać za pomocą rachunku Jonesa wprowadzonego w optyce do opisu spolaryzowanego światła. W ogólnym przypadku wektor Jonesa składa się z dwóch komponentów pola elektrycznego fali EM. Wektor Jones  $e$  definiuje się jako znormalizowany dwuwymiarowy wektor pola elektrycznego  $E$ . Pierwszy element wektora Jonesa można zredukować do liczby rzeczywistej. Drugi element tego wektora jest złożony i, w ogólnym przypadku, określa różnicę faz między ortogonalnymi składowymi pola  $E$ . Z wybranymi bazami pola  $E$  (komponenty  $E_\theta$  i  $E_\phi$ ) dla TX i RX anteny, charakterystyka polaryzacji każdego promienia kanału propagacji może być opisana przez macierz polaryzacji kanałów  $H$ . W oparciu o eksperyment wyniki i teoretyczna analiza zjawiska, charakterystyka polaryzacji modelu zostały wprowadzone na poziomie klastrów, przy założeniu, że wszystkie promienie zawierające jeden klaster ma (w przybliżeniu) taką samą charakterystykę polaryzacji. W związku z tym, rozszerzenie struktury kanału dla wsparcia polaryzacji wymaga zmiany Współczynniki wzmocnienia współczynnika skalarnego  $A^{(i)}$  na  $2 \times 2$  klastyczne macierze polaryzacyjne  $H^{(i)}$ , a realizacja odpowiedzi impulsowej kanału jest opisana przez macierz  $h$ :

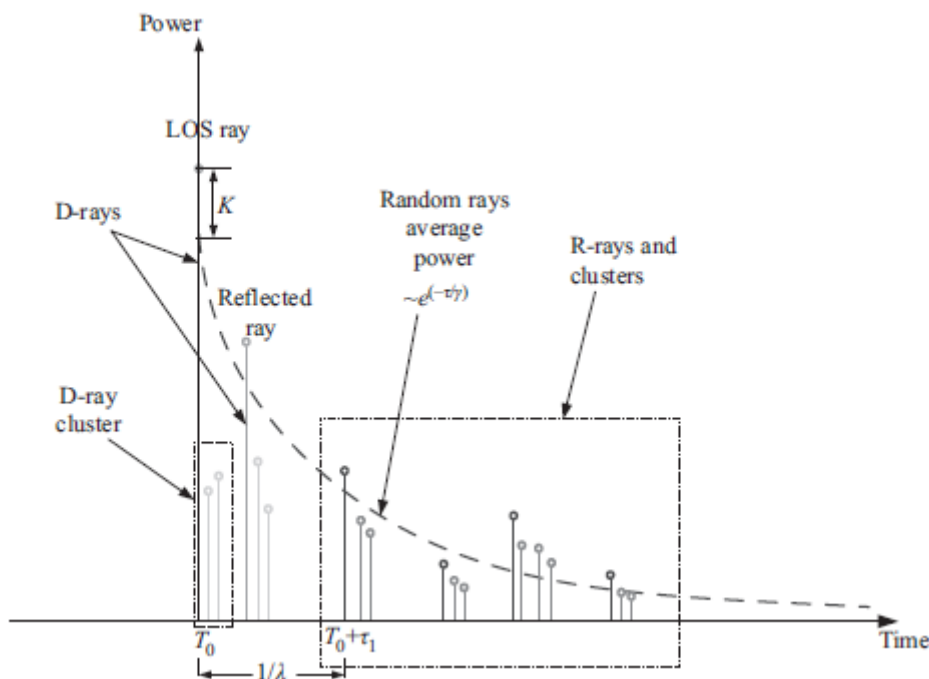
$$\mathbf{h}(t, \varphi_{tx}, \theta_{tx}, \varphi_{rx}, \theta_{rx}) = \sum_i \mathbf{H}^i C^i [t - T^i, \varphi_{tx} - \Phi_{tx}^i, \theta_{tx} - \Theta_{tx}^i, \varphi_{rx} - \Phi_{rx}^i, \theta_{rx} - \Theta_{rx}^i]$$

Ogólny opis odpowiedzi impulsowej kanału można uzyskać w bardziej specyficznej formie, zgodnie z zaproponowaną hybrydową, quasi-deterministyczną, stochastyczną metodologią. Odpowiedź impulsowa kanału może być reprezentowana jako suma klastrów odpowiadających promieniom  $D$  (bezpośrednie promieniowanie, odbicie gruntu, itp.), Skupienia odpowiadające promieniom  $R$  i

skupiskom odpowiadającym promieniom F. Każdy promień ma swoją własną macierz H, która charakteryzuje polaryzację, moc i fazy dwóch składników polaryzacji, opóźnienie promieniowania  $T^i$  i charakterystyki kątowe, takie jak kąty odjazdu (AoD)  $(\varphi_{dx}, \theta_{tx})$  i kąty przybycia (AoA)  $(\varphi_{rx}, \theta_{rx})$ . W takim przypadku równanie transmisji dla kanału pojedynczego może być zapisane jako:

$$y = G_{tx}(\varphi_{tx}, \theta_{tx})G_{rx}(\varphi_{rx}, \theta_{rx})e_{rx}^H H e_{tx} x$$

gdzie  $x$  i  $y$  są transmitowanymi i odbieranymi sygnałami,  $e_{tx}$  i  $e_{rx}$  są wektorami polaryzacji (Jones) odpowiednio dla anten TX i RX, a  $G_{tx}(\varphi, \theta)$  i  $G_{rx}(\varphi, \theta)$  są wzmocnieniami anten przy danym kancie współrzędne. W ogólnym przypadku  $G_{tx}$  i  $G_{rx}$  są różne dla różnych polaryzacji i powinny być reprezentowane jako wektory, podobnie jak  $e_{tx}$  i  $e_{rx}$ . W przypadku ścieżki sygnału LOS, macierz HLOS jest zbliżona do macierzy identyczności (składy nie-diagonalne mogą być niezerowe, ale znacznie mniejsze niż elementy diagonalne) pomnożone przez odpowiadający współczynnik wzmocnienia z powodu utraty ścieżki i rotacji fazy z powodu opóźnienia propagacji sygnału. Rozchodzenie się LOS nie zmienia charakterystyk polaryzacji sygnałów. Jednak charakterystyka polaryzacji sygnałów zmienia się po odbiciach. Zmiana właściwości polaryzacji po odbiciu jest określona przez rodzaj powierzchni i kąt padania. Tak więc charakterystyka polaryzacji może być różna dla różnych skupień, ale jest taka sama dla promieni zawierających jedną grupę. Rysunek



ilustruje przykład ogólnej struktury odpowiedzi na impulsy kanału: liczba klastrow wiązek D i klastrow wiązek R z rozkładającą się wykładniczo cieką strukturą wewnątrz klastra. Parametry wymagane do obliczenia modelu kanału zestawiono tu:

Parametr	Opis
$F_c$	Częstotliwość nośnika, Hz
$c$	Prędkość światła $c = 299\,792\,458$ m/s
$A_0$	Absorbpcja tlenu, dB/km (0,015 dB/m dla 60 GHz)
$H_{tx}$	Wysokość przetwornika powyżej poziomu gruntu, m

$H_{rx}$	Wysokość odbiornika nad ziemią, m
$L$	Odległość pozioma między TX a RX, m
$G_{tx}(\varphi, \theta)$	Wzmocnienie anteny TX przy $(\varphi, \theta)$ współrzędnych kątowych
$G_{rx}(\varphi, \theta)$	Wzmocnienie anteny RX przy $(\varphi, \theta)$ współrzędnych kątowych
$\epsilon_r$	Względna przenikalność elektryczna powierzchni gruntu w $F_c$
$\sigma$	Chropowatość powierzchni: odchylenie standardowe
$\xi$	Dyskryminacja krzyżowa, XPD

Ta notacja będzie używana dalej w opisie konkretnych scenariuszy.

### 3.4.2 Modelowanie promieni deterministycznych (wiązki D)

Promienie quasi-deterministyczne są obliczane jawnie zgodnie z parametrami scenariusza, geometrią i warunkami propagacji. Ważną częścią proponowanego podejścia Q-D do modelowania kanałów jest obliczenie odbitych parametrów promieni. Obliczenia są oparte na metodzie obrazów do określania charakterystyk kątowych odbić pierwszego i wyższego rzędu. Równanie Friisa określa stratę ścieżki P (w dB) na odległość d:

$$P = 20 \log_{10} \left( \frac{\lambda}{4\pi d} \right)$$

Równanie Fresnela służy do obliczenia mocy odbijanych promieni dla różnych komponentów polaryzacji:

$$R = 20 \log_{10} \left( \frac{\sin \phi - \sqrt{B}}{\sin \phi + \sqrt{B}} \right)$$

gdzie R jest współczynnikiem odbicia (stosunek mocy) powierzchni,  $\phi$  jest małym kątem padania a

$$B = \epsilon_r - \cos^2 \phi \quad \text{dla polaryzacji poziomej}$$

$$B = (\epsilon_r - \cos^2 \phi) / \epsilon_r^2 \quad \text{dla polaryzacji pionowej}$$

Jednak równanie Fresnela opisuje jedynie odbicie od idealnej powierzchni, podczas gdy w rzeczywistości większość powierzchni może być traktowana jako "szorstka" w porównaniu do długości fali sygnału zgodnie z kryterium szorstkości Rayleigha.

$$g = 4\pi\sigma_h \sin \phi / \lambda$$

gdzie  $\sigma_h$  jest standardowym odchyleniem wysokości powierzchni wokół lokalnej średniej wartości w obrębie pierwszej strefy Fresnela  $\lambda$  jest wolną przestrzenią fal,  $\phi$  jest małym kątem padania mierzonym względem stycznej do powierzchni. Powierzchnia może być uważana za gładką dla  $g < 0,3$ . Gdy powierzchnia jest zgrubna, odbitym sygnałem są dwa komponenty: element zwierciadlany, który jest spójny z sygnałem padającym, oraz element dyfuzyjny, który fluktuuje amplitudowo i fazowo z rozkładem Rayleigha. Na przykład asfalt o odchyleniach powierzchni około 0,5 mm można uznać za

gładką powierzchnię małym kątem padania  $10^\circ$  ( $g = 0,2$ ). Współczynnik odbicia składnika lustrzanego jest różny od współczynnika wyznaczonego przez równania Fresnela za pomocą mnożnika stałego:  $R_s = \rho_s R$ , gdzie

$$\rho_s = e^{-g^2/2}$$

Przez podstawienie w skali logarytmicznej, spadek mocy sygnału z powodu odbijającej powierzchni chropowatej jest opisana współczynnikiem F:

$$F = \frac{80}{\ln 10} (\pi \sigma_h \sin \phi / \lambda)^2$$

Promienie D są ściśle zależne od scenariusza, ale we wszystkich rozważanych scenariuszach występują dwa podstawowe promienie D: bezpośredni promień LOS i promień odbitego podłoża. Obliczenie tych dwóch podstawowych parametrów promieniowania będzie takie samo dla wszystkich scenariuszy.

### Promień bezpośredni

Bezpośredni promień LOS jest promieniem pomiędzy TX i RX. Jego parametry są obliczane za pomocą wolnej przestrzeni Friisa - równanie utraty ścieżki pustej przestrzeni (patrz: tutaj).

**Opóźnienie:** Opóźnienie bezpośredniego promienia jest obliczane z geometrii modelu:

$$\tau_D = d_D / c$$

$$d_D = \sqrt{L^2 + (H_{Tx} - H_{Rx})^2}$$

**Moc :** Moc promieniowania bezpośredniego obliczona jako utrata ścieżki wolnej przestrzeni z absorpcją tlenu:

$$P_D = 20 \log_{10} \left( \frac{\lambda}{4\pi d_D} \right) - A_0 d_D, \text{ in dB}$$

### Macierz kanału:

$$\mathbf{R} = \begin{bmatrix} 10^{P_D/20} & 0 \\ 0 & 10^{P_D/20} \end{bmatrix} e^{\frac{j2\pi d_D}{\lambda}}$$

**AoD :**  $0^\circ$  azymut i wysokość

**AoA :**  $0^\circ$  azymut i wysokość

### Promień odbity od ziemi

Odbite od ziemi promienie są obecne we wszystkich rozważanych scenariuszach. Jego parametry są obliczone na podstawie równania strat Friis'a i równania Fresnela uwzględniające odbicie i współczynnik rozproszenia powierzchni chropowatych. Zwróć uwagę, że elementy spolaryzowane poziomo i pionowo transmitowanego sygnału będą różnie odzwierciedlane, a tym samym macierzy kanału H powinny mieć różne przekątne

**Opóźnienie** Współczynnik odbicia od ziemi jest obliczany na podstawie geometrii modelu:



$$\tau_G = d_G/c$$

$$d_G = \sqrt{L^2 + (H_{tx} + H_{rx})^2}$$

**Moc** odbijająca od ziemi obliczona jako utrata ścieżki wolnej przestrzeni z absorpcją tlenu, z dodatkową stratą odbicia obliczoną na podstawie równań Fresnela. Strata odbicia R jest różna dla polaryzacji pionowych i poziomych

$$P_{\perp} = 20 \log_{10} \left( \frac{\lambda}{4\pi d_G} \right) - A_0 d_G + R_{\perp} + F; \quad P_{\parallel} = 20 \log_{10} \left( \frac{\lambda}{4\pi d_G} \right) - A_0 d_G + R_{\parallel} + F$$

$$F = \frac{80}{\ln 10} (\pi \sigma_h \sin \phi / \lambda)^2$$

$$R_{\perp} = 20 \log_{10} \left( \frac{\sin \phi - \sqrt{B_{\perp}}}{\sin \phi + \sqrt{B_{\perp}}} \right); \quad R_{\parallel} = 20 \log_{10} \left( \frac{\sin \phi - \sqrt{B_{\parallel}}}{\sin \phi + \sqrt{B_{\parallel}}} \right)$$

$$B_{\parallel} = \epsilon_r - \cos^2 \phi \quad \text{dla polaryzacji poziomej}$$

$$B_{\perp} = (\epsilon_r - \cos^2 \phi) / \epsilon_r^2 \quad \text{dla polaryzacji pionowej}$$

$$\text{gdzie } \tan \phi = \frac{H_{tx} + H_{rx}}{L}$$

**Macierz kanału :**

$$\mathbf{R} = \begin{bmatrix} 10^{P_{\perp}/20} & \xi \\ \xi & 10^{P_{\parallel}/20} \end{bmatrix} e^{\frac{j\pi d_G}{\lambda}}$$

**AoD :** Azymut  $0^\circ$

Wysokość :

$$\theta_{AoD} = \tan^{-1} \left( \frac{L}{H_{tx} - H_{rx}} \right) - \tan^{-1} \left( \frac{L}{H_{tx} + H_{rx}} \right)$$

**AoA :** Azymut  $0^\circ$

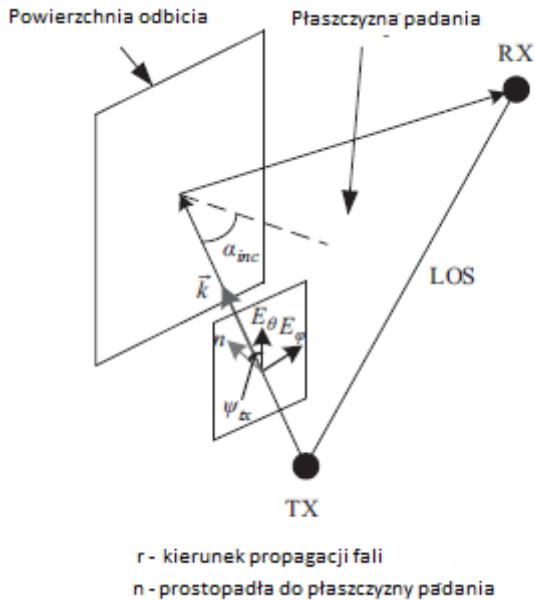
Wysokość :

$$\theta_{AoA} = \tan^{-1} \left( \frac{H_{tx} + H_{rx}}{L} \right) - \tan^{-1} \left( \frac{H_{tx} - H_{rx}}{L} \right)$$

### Polaryzacja dla promieni D

Zgodnie z proponowaną metodologią modelowania kanałów, wszystkie własności promieni quasideterministycznych są obliczane jawnie. Macierz kanału H zawiera całą charakterystykę

polaryzacji promienia i jest obliczana na podstawie odbicia od zdefiniowane na powierzchniach scenariusza. Wiadomo, że współczynniki odbicia są różne dla komponentów pola E zarówno równoległych, jak i prostopadłych do płaszczyzny padania i zależą od kąta padania. Teoretyczne sprzężenie między równoległymi i prostopadłymi składnikami odbitego sygnału wynosi zero dla płaskich interfejsów mediów (granice). Jednak ze względu na brak ideałów (chropowatość) powierzchni pewne sprzężenie zawsze istnieje w kanałach rzeczywistych. Przykład ścieżki odbitego sygnału pierwszego rzędu pokazano na rysunku



Macierz polaryzacji dla ścieżki sygnału odbijanego pierwszego rzędu może być znaleziona jako iloczyn macierzy, która obraca elementy wektora E z układu współrzędnych powiązanego z anteną TX do układu współrzędnych powiązanego z padającą płaszczyzną. Następnie stosuje się macierz odbicia R ze współczynnikami odbicia i współczynnikami sprzężenia crosspolaryzacyjnego, po czym następuje obrót do układu współrzędnych powiązanego z anteną RX. W ten sposób macierz propagacji kanału dla sygnałów odbitych pierwszego rzędu może być zdefiniowana jako:

$$\mathbf{H}_1 = \begin{bmatrix} \cos \psi_{rx} & \sin \psi_{rx} \\ -\sin \psi_{rx} & \cos \psi_{rx} \end{bmatrix} \begin{bmatrix} R_{\perp}(\alpha_{inc}) & \xi_1 \\ \xi_2 & R_{\parallel}(\alpha_{inc}) \end{bmatrix} \begin{bmatrix} \cos \psi_{tx} & \sin \psi_{tx} \\ -\sin \psi_{tx} & \cos \psi_{tx} \end{bmatrix}$$

Macierz najbardziej na lewo to ponowne obliczenie wektora polaryzacji z płaszczyzną padania do współrzędnych RX, prawą matrycą jest ponowne obliczenie wektora polaryzacji TX do płaszczyzny zapadalności. W centrum jest macierzą odbicia R, która zawiera współczynniki odbicia  $R_{\perp}$  i  $R_{\parallel}$  dla prostopadłych i równoległych komponentów pola elektrycznego  $E_{\perp}$  i  $E_{\parallel}$  odpowiednio. Elementy  $\xi_1$  i  $\xi_2$  w macierzy R są współczynnikami sprzężenia polaryzacji krzyżowej. Zwróć uwagę, że struktura macierzy H nie uwzględnia strat propagacji wzdłuż odpowiedniej ścieżki odbitej, która powinna być brana pod uwagę w ostatecznym modelu, ale nie wpływa na właściwości polaryzacji. Struktura macierzy polaryzacji H dla odbicia drugiego rzędu i zawiera dodatkowe macierze rotacji i odbicia:

$$\mathbf{H}_2 = \begin{bmatrix} \cos \psi_{rx} & \sin \psi_{rx} \\ -\sin \psi_{rx} & \cos \psi_{rx} \end{bmatrix} \begin{bmatrix} R_{\perp}(\alpha_{2inc}) & \xi_{21} \\ \xi_{22} & R_{\parallel}(\alpha_{2inc}) \end{bmatrix} \begin{bmatrix} \cos \psi_p & \sin \psi_p \\ -\sin \psi_p & \cos \psi_p \end{bmatrix} \\ \times \begin{bmatrix} R_{\perp}(\alpha_{1inc}) & \xi_{11} \\ \xi_{12} & R_{\parallel}(\alpha_{1inc}) \end{bmatrix} \begin{bmatrix} \cos \psi_{tx} & \sin \psi_{tx} \\ -\sin \psi_{tx} & \cos \psi_{tx} \end{bmatrix}$$

Należy zauważyć, że w ogólnym przypadku pierwsza padająca płaszczyzna nie pokrywa się z drugą padającą płaszczyzną. W związku z tym wymagane jest dodatkowe przeliczenie z układu współrzędnych powiązanego z pierwszą płaszczyzną padającą z układem współrzędnych związanym z drugą płaszczyzną padającą. Elementy niediagonalne macierzy kanału H reprezentują depolaryzację sygnału, tj. Sprzężenie między ortogonalnymi składowymi wektora E po stronie TX i RX. Istnieją dwa mechanizmy depolaryzacji: sprzężenie odbicia (sprzężenie pomiędzy równoległymi i prostopadłymi składowymi wektora E przy odbiciu) i sprzężenie geometryczne (sprzężenie ze względu na różne orientacje względne anten TX i RX). Można zauważyć, że proponowane podejście pozwala obliczyć oba mechanizmy, aby stworzyć dokładny model wpływu polaryzacji.

### 3.4.3 Modelowanie statystyczne losowych promieni (promienie R)

Aby uwzględnić liczbę promieni, których nie da się łatwo określić deterministycznie (odbicia od obiektów, które nie są w pełni określone w scenariuszu, obiekty o losowym lub nieznanym położeniu, obiekty o złożonej geometrii, odbicia wyższego rzędu, itp.) stosuje się podejście statystyczne w metodologii modelowania kanałów Q-D. Dokładniej, statystyczne podejście do impulsu kanału klastrowego stosuje się modelowanie odpowiedzi. Podejście statystyczne to podstawowy sposób generowania promieni R, który jest stosowany w metodologii modelowania kanałów Q-D. Klasy docierają w momentach  $T^i$  zgodnie z procesem Poissona i mają czasy między przybyciami, które są rozkładane wykładniczo. Amplitudy skupień  $A(T^i)$  są niezależnymi zmiennymi losowymi Rayleigha, a odpowiadające im fazy  $\Theta^i$  są niezależnymi, jednolitymi zmiennymi losowymi w  $[0, 2\pi]$ . Charakterystyka kąтова (AoD, AoA) promieni losowych zależy od scenariusza i zostanie zdefiniowana w odpowiednich sekcjach z opisem parametrów modeli kanałów. Przy takich założeniach, składowe chaosu losowego odpowiedzi impulsowej kanału są podane przez:

$$h_{cluster}(t) = \sum_{i=1}^{N_{cluster}} A(T^i) e^{j\Theta^i} C^i(t - T^i),$$

gdzie  $T^i$  jest czasem przybycia i-tego klastra mierzonym od czasu przybycia do promienia LOS,  $A(T^i)$ ,  $\Theta^i$  to amplituda i faza i-tego klastra.  $C^i$  jest intra-klastrem odpowiedzi impulsowej określonej poniżej. Promienie R są losowe, z Rayleighem rozłożone amplitudy i fazy losowe, z rozkładającym się wykładniczo profilem opóźnienia mocy. Całkowita moc promieni R jest określona przez współczynnik K w odniesieniu do mocy bezpośredniego składnika LOS:

$$\overline{P(T^i)} = P_0 e^{-T^i/\gamma} \\ \frac{P_{LOS}}{\sum \overline{P(T^i)}} = K$$

gdzie  $P_{LOS}$  jest mocą promienia LOS,  $P(T^i)$  to moc elementu klastra z czasem przybycia T, K jest stosunkiem LOS do wszystkich komponentów NLOS (współczynnik K), oraz  $\gamma$  oznacza stałą rozpadu

mocy dla klastrow. Opóźnienia klastra  $T^i$  są opisaną funkcją niezależnej funkcji gęstości prawdopodobieństwa między nadejściem sparametryzowanej przez współczynnik przybycia  $\lambda$ :

$$p(T^i | T^{i-1}) = \lambda e^{-\lambda(T^i - T^{i-1})}$$

Konkretne wartości parametrów PDP,  $\lambda$ ,  $K$  i  $\gamma$  są zdefiniowane w sekcjach modeli kanałów dla danego scenariusza.

### Polaryzacja dla promieni R

Losowe promienie (promienie R) są określone przez profil opóźnienia mocy i charakterystyki kątowe. Promienie R modelowały odbicia wyższego rzędu, odległe odbicia w środowisku otwartym i odbicia od różnych obiektów losowych pojawiły się w obszarze działania systemu komunikacyjnego. Są one generowane stochastycznie, tworząc te parametry jako wartości losowe z predefiniowanymi funkcjami rozkładu prawdopodobieństwa (PDF). W modelu kanału 802.11ad [12] opracowano przybliżenia statystyczne dla różnych typów odbijanych klastrow w oparciu o dostępne dane eksperymentalne i symulacje ray tracingu. Zaproponowano zestaw przybliżeń dla przekątnych i poza diagonalnych elementów odbić pierwszego i drugiego rzędu dla typowych warunków w pomieszczeniach (sala konferencyjna, kabina i salon) jako kombinacji logormalnych i jednolitych rozkładów. W Q-D modelu amplituda promienia aproksymowana jest przez rozkład Rayleigha (który jest bliski log-normal), aby prosta stała macryca polaryzacyjna  $H_p$  mogła być użyta do wprowadzenia właściwości polaryzacji do promieni R. Macierz polaryzacji kanałowej  $H_p$  dla promieni R jest określona przez

$$H_p = \begin{bmatrix} 1 & \pm 0.1 \\ 0.1 & \pm 1 \end{bmatrix}$$

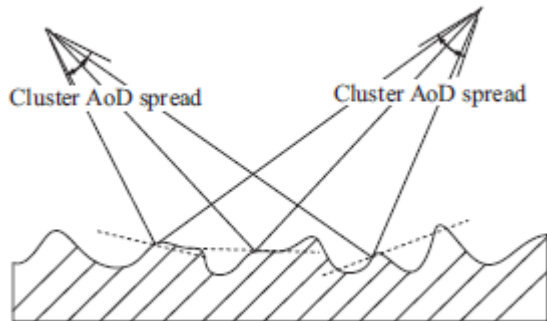
Wartości ze znakiem  $\pm$  zakłada się, że mają one losowy znak (+1 lub -1, na przykład) z równym prawdopodobieństwem, niezależnie od innych wartości. W przypadku promieni gromady głównego promienia losowego macierz polaryzacji jest taka sama jak w przypadku promieni losowych.

#### 3.4.4 Losowe odbicie Generowanie promieni R

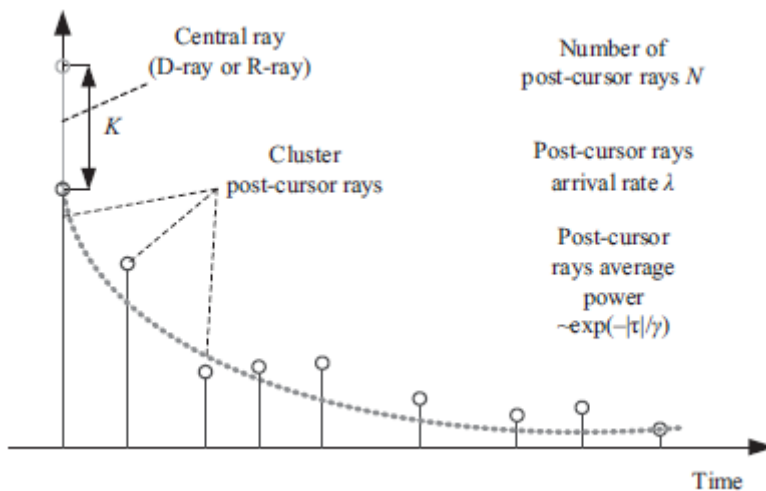
Syntetyczne przetwarzanie apertury wyników eksperymentalnych wykazało, że odbicia od różnych obiektów środowiskowych, takich jak drzewa, latarnie, przystanki autobusowe itp., można wyraźnie zidentyfikować (z dokładną oceną położenia reflektora) na podstawie danych eksperymentalnych. Takie promienie powinny być brane pod uwagę wraz z promieniami D, które pochodzą z dużych obiektów, ale określenie położenia każdego reflektora sprawia, że opis scenariusza jest złożony i konkretny. W związku z tym proponuje się generowanie tego rodzaju promieni (promienie R lub promienie F) jako odbicia od losowo umieszczonych obiektów sferycznych, które (w przeciwieństwie do obiektów płaskich) mogą tworzyć zwierciadlaną ścieżkę odbicia między dowolnymi dwoma punktami w przestrzeni 3D. Na razie, w oparciu o pomiary eksperymentalne, promienie R jako odbicia od przypadkowych obiektów zostały wprowadzone tylko w scenariuszu z kanionem ulicznym, oprócz statystycznie wygenerowanego profilu opóźnienia mocy promieni R. W ten sposób generowane są również promienie F, z różnicą czasu istnienia ścieżki w aplikacjach, w których analizowane są dłuższe okresy.

#### 3.4.5 Struktura wewnątrz klastra

Chropowatość powierzchni i obecność różnych nieregularnych obiektów na rozważanych powierzchniach odbijających i wewnątrz nich (cegły, okna, obramowania, włązy, tablice ogłoszeń na ścianach itp.) prowadzą do oddzielenia zwierciadlanego promienia odbicia od szeregu dodatkowych promieni o podobnych opóźnieniach i kątach: skupienie. Mechanizm powstawania skupisk na powierzchni o dużych nierównościach przedstawiono na rysunku



Można zauważyć, że odbicie od szorstkich powierzchni powoduje rozprzestrzenianie się sygnału pod kątem i opóźnieniem. Parametry wewnątrz klastra modelu kanału oszacowano na podstawie danych pomiarowych. Poszczególne promienie zostały zidentyfikowane w dziedzinie czasu i zmierzono charakterystyki statystyczne, w tym średnią liczbę promieni, szybkość przybycia promienia i czas zaniku mocy promieni. Na podstawie uzyskanych wyników podano model statystyczny dla parametrów domeny czasu klastra. Strukturę modelu przedstawiono schematycznie na rysunku



Dla każdego promienia bazowego strukturę wewnątrz klastra określa:

$$C(t) = \sum_{k=1}^{N_{intra}} \alpha(\tau_k) e^{j\theta_k} \delta(t - \tau_k),$$

gdzie  $N_{intra}$  jest liczbą składników wewnątrzgrupowych,  $\tau_k$  jest czasem przybycia k-ty składnik wewnątrz klastra mierzony od czasu przybycia podstawowego D-ray lub R-ray,  $\alpha(\tau_k)$ ,  $\theta(\tau_k)$  i  $\theta_k$  to amplituda, moc i faza k-tego wewnątrzgrupowego składnik. Komponenty wewnątrz klastra są losowe, z dystrybucją Rayleigha amplitudy i fazy losowe, z rozkładającym się wykładniczo profilem opóźnienia mocy. Całkowita moc składników wewnątrzgrupowych jest określona przez współczynnik K z w odniesieniu do podstawowej mocy promieni D lub R:

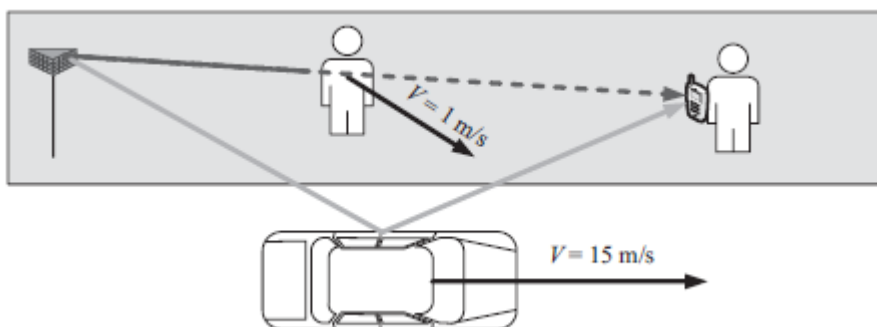
$$\overline{p(\tau_k)} = p_0 e^{-\tau_k/\gamma}$$

$$\frac{P_{base\ ray}}{\sum P(\tau_k)} = K$$

Generalnie generowanie struktury wewnątrz klastra jest bardzo podobne do generacji promieni, z wyjątkiem tego, że w przypadku generowania promieni R, promień LOS jest wykorzystywany jako czas i baza mocy, a do generowania struktury wewnątrz klastra wykorzystuje się do tego celu D-ray w kształcie klastra lub R-ray. Łącząc wszystkie promienie D, promienie R i ich odpowiednie komponenty struktury wewnątrz klastra da ostateczną odpowiedź impulsową kanału. Identyfikacja promieni wewnątrz klastra w domenie kątowej wymaga zwiększenia rozdzielczości kątowej za pomocą anten kierunkowych o bardzo dużym wzmocnieniu. Inne podejście polegało na zastosowaniu techniki "wirtualnej macierzy antenowej", w której nisko kierunkowy element anteny wykorzystywany jest do wykonywania pomiarów w wielu położeniach wzdłuż wirtualnej macierzy antenowej w celu utworzenia efektywnego otworu anteny, został wykorzystany w projekcie MEDIAN. Wyniki te przeanalizowano, w oparciu o zalecenie modelowania rozprzestrzeniania się kątów wewnątrz klastra dla kątów azymutu i kątów podniesienia dla nadajnika i odbiornika jako niezależnych normalnie rozdzielonych zmiennych losowych ze średnią zerową i RMS równej 50:  $N(0, 50)$ . Należy pamiętać, że rozsądne jest założenie, że różne typy klastrów mogą mieć wyróżniającą się strukturę wewnątrz klastra. Na przykład właściwości klastrów odbijanych od nawierzchni drogi mogą różnić się od właściwości klastrów odbitych od ceglanych ścian z powodu odmiennej struktury powierzchni materiału. Można również założyć, że właściwości klastrów odbijanych pierwszego i drugiego rzędu są różne, a klastry odbite drugiego rzędu mają większe spready w domenach skroniowych i kątowych. Wszystkie te efekty są rozumiane jako rozsądne. Ponieważ jednak liczba dostępnych wyników eksperymentalnych była ograniczona, opracowano wspólny model wewnątrz klastrowy dla wszystkich typów klastrów. Modyfikacje różnych modeli wewnątrz klastra dla różnych typów klastrów mogą być przedmiotem przyszłego kanału ulepszenia modelu.

### 3.4.6 Modelowanie zatorowe

Właściwa symulacja efektów blokowania jest ważną częścią modelowania kanałów fal milimetrowych. We wszystkich rozważanych scenariuszach, które obejmują symulacje łącza dostępu ścieżki propagacji sygnału są blokowane przez ludzi lub pojazdy przerywające promienie ze statycznymi pozycjami AP i UE, lub przez ruch UE w obszarach, w których zacienione są niektóre promienie. Konieczność wprowadzenia blokady do modelu kanału 3D potwierdzają pomiary eksperymentalne. Przeciwny, ale podobny efekt to pojawienie się nowych promieni na krótki czas - na przykład odbicia od przejeżdżających pojazdów i innych ruchomych odbić obiekty. Oba efekty można zaobserwować w wynikach eksperymentalnych pomiarów anteny dookólne. Rysunek



ilustruje mechanizm blokady promieni i wyglądu promienia. Po zdjęciu średni czas blokady i czas trwania flashowania odbicia można oszacować jako:

$$T_{\text{blockage}} \sim 0: 5 \text{ m (szerokość człowieka) / 1 m/s (średnia prędkość)} \sim 0,5 - 1 \text{ s}$$

$$T_{\text{flash}} \sim 4,5 \text{ m (długość samochodu) / 15 m/s (średnia prędkość)} \sim 0,2 - 0,3 \text{ s}$$

Analiza danych eksperymentalnych w wykresach bitmapowych z opóźnionym promieniowaniem czasowym daje w przybliżeniu te same wartości. Średni okres użytkowania (SP) typowej komunikacji w zakresie fal milimetrowych systemu są równe 1-3 ms (802.11ad). Oznacza to, że podczas blokady lub okresu flashowania mijają tysiące okresów serwisowych i dziesiątki tysięcy ramek. Symulacje na poziomie systemu rzadko obejmują więcej niż tysiąc klatek, więc Zablockowanie może być modelowane jako zdarzenie statyczne zamiast dynamicznego. Blokada jest określana raz na migawce kanału i pozostaje taka sama w czasie. W przypadku symulacji transmisji strumieniowej VoIP i wideo, które wymagają analizy dłuższych okresów, momenty blokowania można wprowadzić jako proces Poissona z odpowiednimi parametrami

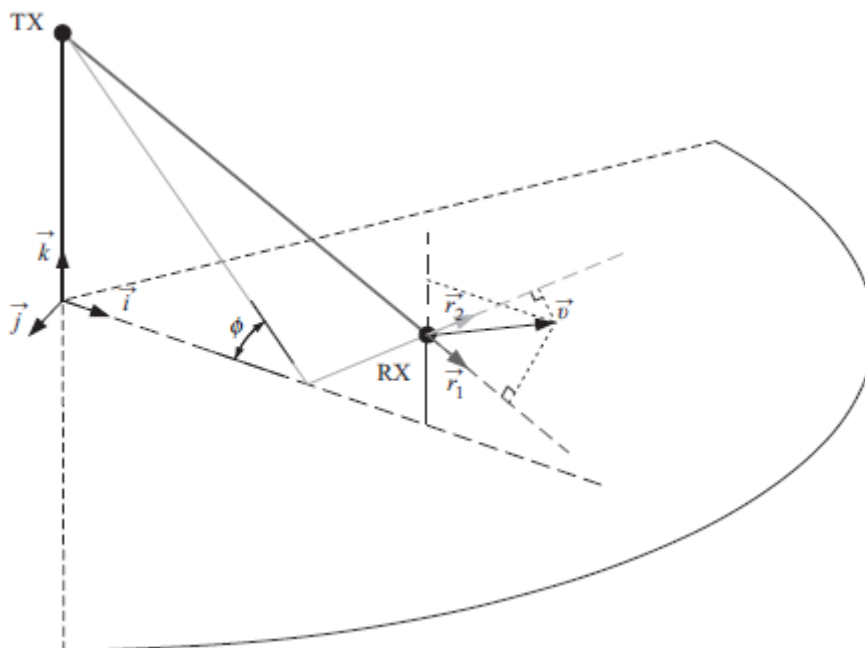
### 3.4.7 Modelowanie mobilności kanału i wpływ

Efekty mobilności w modelu kanału Q-D są opisane przez bezpośrednie wprowadzenie wektor prędkości dla każdego UE. W kanale wielościeżkowym ruch UE prowadzi do dodatkowego obrotu fazy dla każdej ścieżki propagacji, ale dla celów modelowania kanału, efekt ruchu może być wprowadzony dla promieni D i dla najsilniejszych promieni R w ten sam sposób. Dodatkowa rotacja faz dla i-ray wywołana przesunięciem częstotliwości Dopplera jest obliczana jako

$$\Delta\varphi_i(t) = 2\pi f_i^D t$$

$$f_i^D = (\mathbf{v}, \mathbf{r}_i) \frac{F_c}{c}$$

gdzie  $f_i^D$  jest przesunięciem częstotliwości dla i-tego promienia,  $\mathbf{v}$  jest chwilowym wektorem UE prędkości



,  $r_i$  jest wektorem jedności owego kierunku wiązki, a  $(\cdot)$  oznacza produkt skalarny. Wektor prędkości  $v$  może być reprezentowany przez sumę jego skalarnych składników:

$$v = v_x i + v_y j + v_z k$$

Poziome elementy wektora prędkości są specyficzne dla scenariusza. Dla scenariuszy bez preferowanego kierunku ruchu, takiego jak lobby hotelowe lub otwarty obszar, pozioma składowa prędkości może mieć równomiernie rozłożony kierunek i losową lub stałą wartość. W przypadku kanionu ulicznego ruch wzdłuż ulicy może być modelowany z predefiniowaną prędkością (3-5 km / h dla pieszych, 30-120 km / h dla pojazdów), a ruch po drugiej stronie ulicy może być modelowany jako zero średnia losowa Gaussian proces. Jak pokazano w sekcji 3.3.1, ruch pionowy mobilnego UE ma większy wpływ na kanał i powinien być rozpatrywany niezależnie. W ważnym przypadku, gdy mobilne UE jest w posiadaniu człowieka, różne modele ludzkiego chodu można zastosować do ruchu pionowego z  $(t)$ . W ramach metodologii Q-D ruch pionowy jest traktowany jako stacjonarny losowy proces Gaussa z funkcją korelacji równą:

$$K_z(\tau) = \sigma_z^2 e^{-\frac{\tau^2}{\tau_z^2}}$$

Parametry funkcji korelacji należy dostosować do rzeczywistego ruchu pieszego z prędkością 3-5 km / h. Dla rozważanych modeli kanałów dla pieszych zakładamy  $\sigma_z = 0,05$  m i  $\tau_z = 0,5$  s. Składnik pionowy  $v_z$  wektora prędkości  $v$  może być zdefiniowany przez ruch pionowy użytkownika jako pierwszej pochodnej. Znając wektor prędkości i promienie kąta przybycia (AoAs), wartości rotacji faz można obliczyć z powyższego równania i dodać do odpowiednich faz  $D$  i faz promieni  $R$ . Ruch UE może prowadzić do selektywnego czasowo zaniku powodowanego przez interferencję dwóch lub więcej promieni. Głębokość zanikania może być duża, zwłaszcza w przypadku LOS i zakłóceń odbitych od ziemi z dużą odległością przy stosunkowo niskiej przepustowości sygnału. Dla podróżujących w ruchu z prędkością 3 km / h, można obliczyć maksymalne przesunięcie Dopplera, które będzie wynosić około 170 Hz dla częstotliwości nośnej 60 GHz z typowym czasem koherencji kanału  $t_c$  równym  $1 / f_D = 6$  ms. Ale dla prędkości pojazdu 120 km / h maksymalne rozproszenie Dopplera osiągnie 6700 Hz z odpowiednim czasem koherencji  $\tau_c = 0,15$  ms. Aby ocenić potencjalny wpływ efektów mobilności na przyszłe systemy komunikacyjne 5 milimetrów, parametry standardu IEEE 802.11ad dla Wi-Fi fal milimetrowych są porównywane tu

Parametry systemowe IEEE 802.11ad	Parametry kanału milimetrowego
Czas trwania próbki 0,3-0,5 ns	Opóźnienie rozprzestrzeniania 3-5 ns
Szerokość pasma 2 160 MHz	Szerokość pasma koherencji 200-350 MHz
Symbol czas trwania 242 ns / 328 ns	Częstotliwość Dopplerowska 167-6,700

z typowymi parametrami kanału fal milimetrowych. Duża dostępna szerokość pasma (2160 MHz) jest kompensowana przez bardzo krótki okres symbolu: 242 ns dla trybu OFDM i 328 ns dla trybu SC FDE . Typowy kanał z falą milimetrową o czasie koherencji większym niż 0,15 ms w najgorszym przypadku wysokiej mobilności nie zostanie zmieniony w okresie symbolu. Jednocześnie, w typowym czasie trwania ramki wynoszącym około 0,5 ms, kanał pojazdu może podlegać znaczącym zmianom, które powinny być brane pod uwagę w przyszłych układach warstw PHY, harmonogramowaniu i alokacji zasobów czasowo-użytkownika. Tabela umożliwia również ocenę aspektów związanych z przepustowością projektu systemu. Propagacja wielościeżkowa z typowym rozkładem opóźnienia



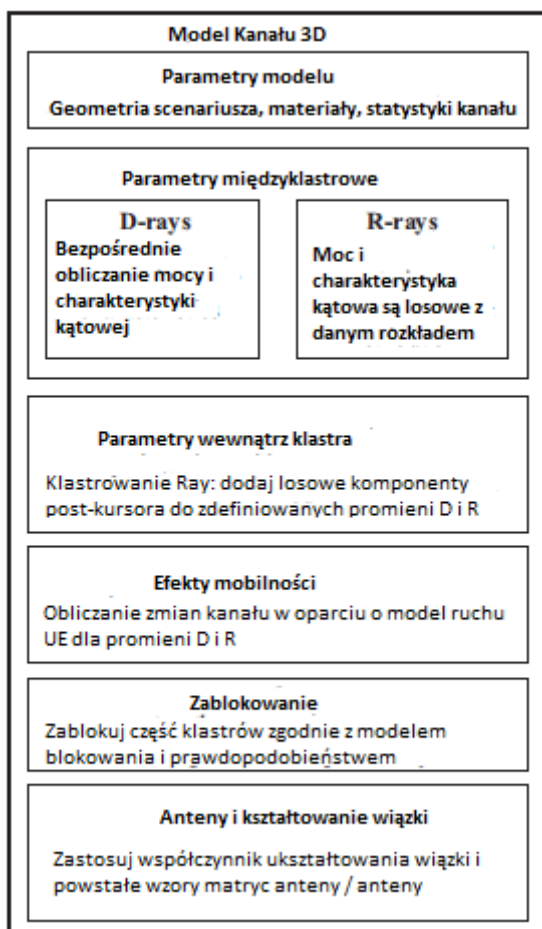
wynoszącym około 3-5 ns prowadzi do selektywności częstotliwościowej w kanale fal milimetrowych o szerokości pasma koherencji około 300 MHz. Tak więc, system komunikacji fal milimetrowych o szerokości pasma kanału mniejszej niż ta wartość może doświadczać głębokiego selektywnego w czasie płaskiego zanikania, co może prowadzić do awarii systemu.

### 3.5 Implementacja modeli kanałów Q-D

Generowanie trójwymiarowego modelu kanału fal milimetrowych zgodnie z metodologią Q-D składa się z następujących kroków:

- Definicja scenariusza i parametrów modelu.
- Obliczanie parametrów D-ray zgodnie z wybranym scenariuszem zalecenia.
- Obliczanie parametrów promieniowania R zgodnie z wybranym scenariuszem zalecenia.
- Zastosowanie blokady ścieżki zgodnie z wymogami scenariusza dla wybranych klastrów.
- Zastosowanie wzorców anten TX i RX oraz algorytmów kształtowania wiązki.
- Konwersja surowych danych odpowiedzi impulsowej kanału na dyskretny czas wymagane przez symulacje.

Przeływ generacji modelowania kanałów przedstawiono na rysunku



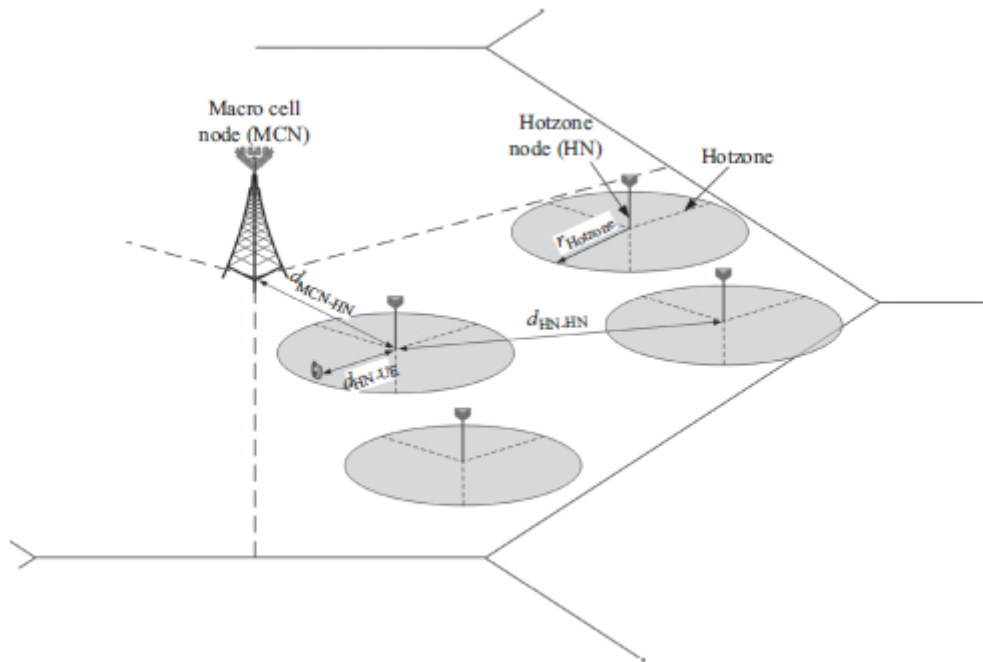
#### 3.5.1 Model kanału dostępu do kampusu uniwersyteckiego

Model kanału kampusu uniwersyteckiego przedstawia scenariusz z dużymi otwartymi obszarami z niskimi i zapasowymi budynkami, takimi jak kampus uniwersytecki, tereny parkowe, place miejskie.

### Modelowanie geometrii i wdrożenie UE

Geometria scenariusza Open Area jest oparta na heterogenicznym scenariuszu sieci 3GPP. Scenariusz ten reprezentuje rozmieszczenie z sześciokątnym układem dla węzłów makr LTE i pewną liczbą węzłów o niskiej mocy LTE równomiernie zrzuconych w strefach gorących obszaru geograficznego każdego z obszarów maczanych. Scenariusz zostaje poszerzony o dodanie fal milimetrowych do stref gorących. Biorąc pod uwagę zastosowanie anten kierunkowych kilka fal milimetrowych należy umieścić w gorącej strefie o różnych orientacjach. Parametry rozmieszczenia scenariusza otwartego obszaru zestawiono w tabeli i zilustrowano na rysunku

Parametr	Wartość
Odległość między lokalizacjami	500 m
Liczba komórek makr na witrynę	3
Liczba stref gorących na makro komórkę, $N_{HN}$	4
$d_{MCN-HN\_min}$	75 m
$d_{HN-HN\_min}$	40 m
Wysokość węzła cieplnego, $H_{HN}$	6 m
Upuszczenie UE	Występujące w grupach *
Liczba użytkowników na komórkę makr, $N_{users}$	>100
Fracja użytkowników hotspotów, $P^{Hotspot}$	9/10
$r_{Hotzone}$	40 m
$d_{HN-UE\_min}$	5 m
$d_{MCN-UE\_min}$	35 m
Wysokość UE, $H_{UE}$	1,5 m
Liczba BS fali milimetrowej na strefę hotzone	3
Materiał powierzchni	Asfalt
Powierzchnia $\epsilon_r$	4 + 0,2j
Chropowatość powierzchni (odchylenie standardowe)	3 mm



- Napraw całkowitą liczbę użytkowników,  $N_{users}$ , zrzuconych z każdego obszaru geograficznego makro.
- Losowo i jednolicie upuść skonfigurowaną liczbę węzłów hot-zone,  $N_{HN}$ , w każdym obszarze geograficznym makro (ten sam numer  $N_{HN}$  dla każdego makra obszar geograficzny).
- Losowo i jednolicie upuść użytkowników  $N_{users\_HN}$  w promieniu  $r_{Hotzone}$  każdego z nich węzeł hotzone, gdzie  $N_{users\_HN} [ bP^{Hotspot} N_{users} / N_{HN} ]$  i  $P^{Hotspot}$  to ułamek użytkowników hotspotów.
- Losowo i jednolicie upuść pozostałych użytkowników,  $N_{users} - N_{users\_HN} N_{HN}$  na cały makro obszar geograficzny danej makrokomórki (w tym obszar strefy gorącej).

#### Obliczenia promieni D i promieni R na otwartych obszarach

W kampusie uniwersyteckim (na zewnątrz większy obszar) wdrażanie scenariuszy UE ma tylko dwa dominujące promienie: promień odbitego światła i promień odbitego podłoża. Promienie te są liczone jako deterministyczne i jednoznacznie wyliczone zgodnie z sekcją 3.4.2. Oprócz głównych deterministycznych składników, promieni bezpośrednich i naziemnych, są przypadkowe elementy reprezentujące rozproszenie odbicia. Odbicie od odległych ścian, obiektów losowych i odbicie drugiego rzędu są brane pod uwagę jako elementy losowe, a ich parametry są zestawione w tabeli

Parametr	Wartość
Liczba klastrów, $N_{cluster}$	3
Przybycie klastra, $\lambda$	0,05 ns <sup>-1</sup>
Stała zaniku mocy klastra, $\gamma$	15 ns
Współczynnik K	6 dB
AOA	Wysokość: U [ $AOA_{ground}$ : $AOA_{LOS}$ ] Azymut: U [-60: 60°]
AOD	Wysokość : U [ $AOD_{ground}$ : $AOD_{LOS}$ ]

Azymut: U [-60: 60°]

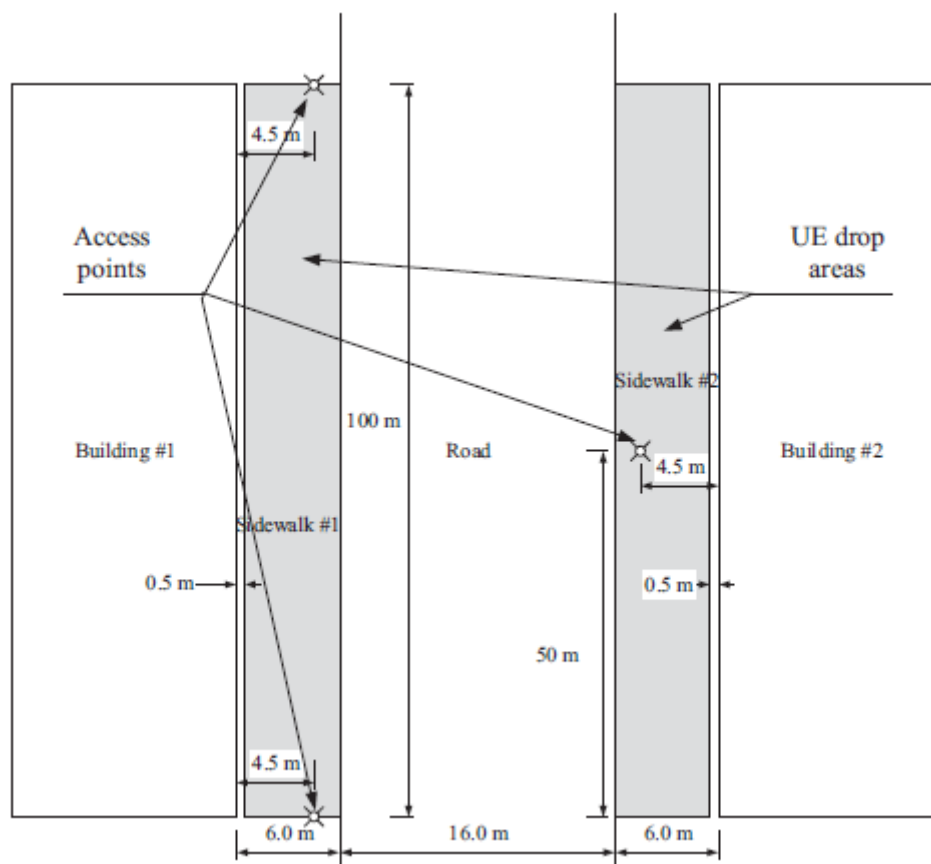
Zarówno promień D, jak i promień R w modelu kanału otwartego mają cienką strukturę klastrową, która dodaje promienie pozakursorowe do głównego komponentu D-ray i R-ray. Chociaż bezpośredni promień LOS może również mieć strukturę klastrową ze względu na zmiany ścieżki propagacji i częściowo zamknięty przez przeszkody strefy Fresnela, w proponowanym modelu bezpośredni promień nie ma klastrow.

### 3.5.2 Model kanału dostępu do kanionu ulicznego

Kanał uliczny (dostęp do ultra-wysokiej stopy w centrum miasta) model kanału reprezentuje typowy miejski scenariusz: ulice miasta z chodnikami dla pieszych wzdłuż wysokich długich budynków. W tym scenariuszu modeluje się łącze dostępu między punktami dostępowymi na słupach latarni i UE w rękach ludzkich.

#### Modelowanie geometrii i wdrożenie UE

Geometria scenariusza dostępu do kanionu ulicznego, który jest wykorzystywany do oceny parametrów modelu kanału za pomocą symulacji śledzenia promieni, jest pokazana na rysunku



Odpowiednie parametry liczbowe zestawiono w tabeli

Parametr	Wartość
Wysokość AP, $H_{tx}$	6 m
Wysokość UE, $H_{rx}$	1,5 m
Odległość AP od najbliższej ściany, $D_{tx}$	4,5 m

Szerokość chodnika	6 m
Szerokość drogi	16 m
Długość ulicy	100 m
Odległość AP-AP, ten sam bok	100 m
Odległość AP-AP, różne boki	50 m
Materiały drogowe i chodnikowe	Asfalt
Droga i chodnik $\epsilon_r$	4 + 0,2j
Szorstkość dróg i chodników $\sigma_g$ (odchylenie standardowe)	0,2 mm
Materiał ścian budowlanych	Beton
Ściany budynków $\epsilon_r$	6,25 + 0,3j
Chropowatość ścian budynków wynosi (odchylenie standardowe)	0,5 mm

### Obliczenia D-Ray i R-ray z kanionu na ulicy

W scenariuszu ulicznego kanionu (dostęp do ultra-wysokich stóp) instalacje UE pogrupowane są na stosunkowo wąskiej ścieżce, z dwoma dominującymi odbijanymi promieniami oprócz bezpośredniego: promieniem odbitym od ziemi i promieniem odbitym od ściany. Wszystkie te promienie są liczone jako deterministyczne i jawnie obliczane podczas modelowania kanałów. Parametry promienia odbitego od ściany są określone w ten sam sposób, co odbicie od ziemi, ale zamiast wysokości anteny  $H_{tx}$  i  $H_{rx}$  odległości pomiędzy antenami i najbliższą ścianą  $D_{tx}$  i  $D_{rx}$ . Oprócz deterministycznych składników, promieni bezpośrednich i odbitych, są przypadkowe elementy reprezentujące rozpraszanie odbicia. Odbicie od odległych ścian i odbicia drugiego rzędu są brane pod uwagę jako elementy losowe. Losowe składniki statystyk odpowiedzi impulsowej kanału pochodzą z kanionu ulicznego (ultraszybkie hot-spot) w modelowaniu w poprzednich sekcjach oraz z danych pomiarowych

Parametr	Wartość
Liczba klastrów, $N_{cluster}$	5
Szybkość przybycia klastra, $\lambda$	0,03 ns <sup>-1</sup>
Stała zaniku mocy klastra, $\gamma$	20 ns
Współczynnik promieniowania K	10 dB
AoA	Wysokość: U [-20: 20°] Azymut: U [-180: 180°]
AoD	Wysokość: U [-20: 20°] Azymut: U [-180: 180°]

Parametry klastra dla modelu D w kanionie ulicznym i promieni R pokazano w tabeli

Parametry	Wartość
Promień post-kursora K-factor, K	4
Czas rozpadu mocy po kursorze, $\gamma$	4,5 ns
Przybycie po kursie, $\lambda$	0,31 ns <sup>-1</sup>
Promień po kursie rozkładu amplitudy	Rayleigh
Liczba promieni posakresowych, N	4

### 3.5.3 Model kanału dostępu do lobby hotelowego

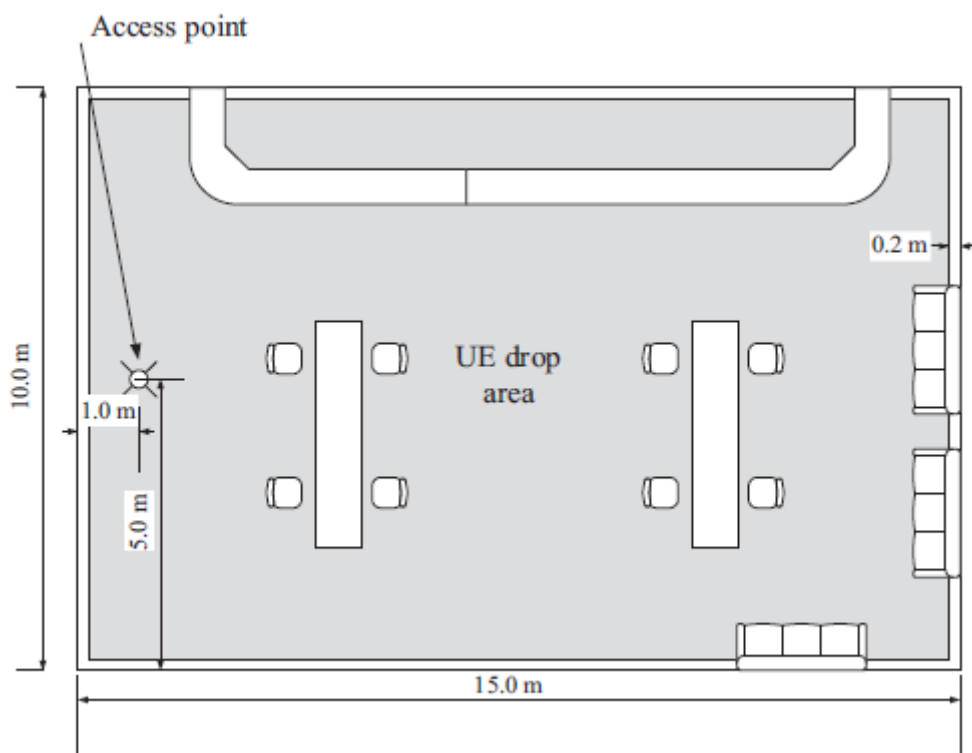
Model kanału hotelowego (kryty dostęp do dużego obszaru publicznego) reprezentuje typowy scenariusz wewnętrzny: dużą halę z wieloma użytkownikami wewnątrz. Podobne modele kanałów wewnętrznych zostały uwzględnione, z wykorzystaniem statystycznego podejścia do modelowania kanałów, odpowiedniego do symulacji warstwy łącza. Proponowane podejście quasi-deterministyczne opiera się na określonej lokalizacji UE i może być wykorzystywane również do symulacji na poziomie systemu.

#### Modelowanie geometrii i wdrożenie UE

Podstawowe parametry i geometrię zestawiono w tabeli

Parametry	Wartość
Wysokość AP, $H_x$	3,5 m
Pozycja AP	Środek najbliższej ściany
Wysokość UE, $H_{rx}$	1,5 m
Wysokość pomieszczenia	4 m
Szerokość pomieszczenia	10 m
Długość pokoju	15 m
Materiał podłogowy	Beton
Podłoga $\epsilon_{rf}$	4 + 0,2j
Chropowatość podłoża $\sigma_f$ (odchylenie standardowe)	0,1 mm
Ściany materiał	Beton
Ściany $\epsilon_{rw}$	4 + 0,2j
Ściany chropowatości $\sigma_w$ (odchylenie standardowe)	0,2 mm
Materiał sufitu	Płyta gipsowo-kartonowa
Sufit	6,25 + 0,3j
Chropowatość sufitu $\sigma_c$ (odchylenie standardowe)	0,2 mm

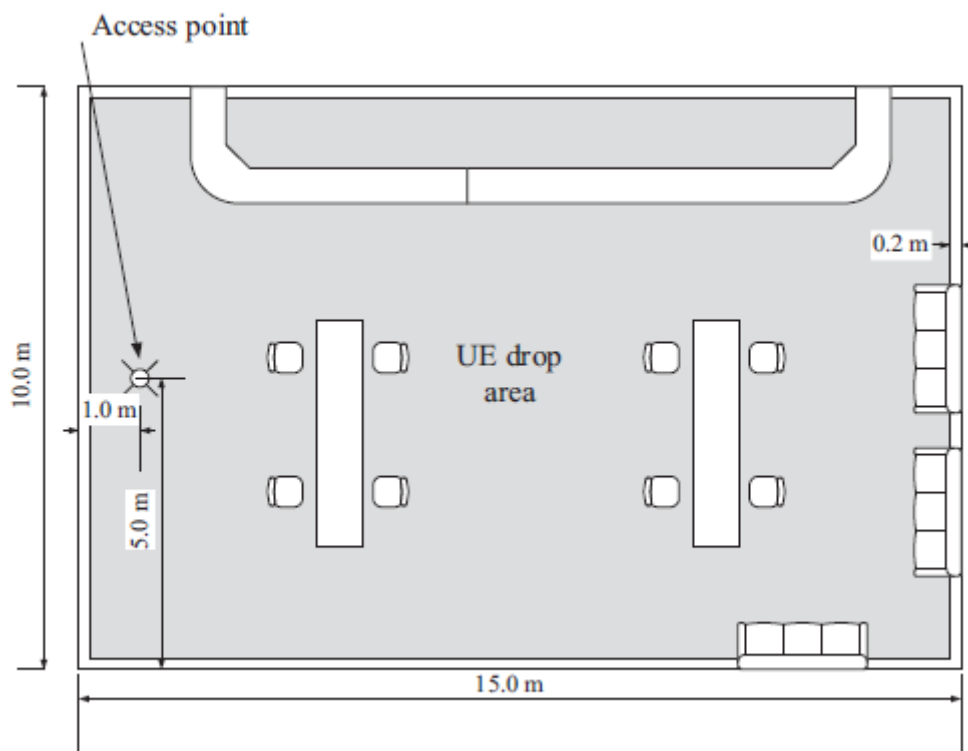
i zilustrowano



### Wyliczenia D i L promienie hotelu

Model kanału 3D do scenariusza lobby hotelowego powinien zawierać promienie odbicia do drugiego rzędu jako promienie D, obliczone na podstawie metody obrazów i równania Fresnela

Parametr	Wartość
Liczba klastrów, $N_{\text{cluster}}$	5
Wskaźnik przybycia klastra, $\lambda$	$0,01 \text{ ns}^{-1}$
Stała zaniku mocy klastra, $\gamma$	15 ns
Współczynnik promieniowania K	10 dB
AoA	Wysokość: U $[-80: 80^\circ]$ Azymut: U $[-180: 180^\circ]$
AoD	Wysokość : U $[-80: 80^\circ]$ Azymut: U $[-180: 180^\circ]$



Dokładne formuły wszystkich promieni D są zbyt długie i tak jest. Zaleca się stosowanie do obliczeń algorytmu śledzenia promieni. Oprócz deterministyczne odbicia od ścian, sufitu i podłogi, promienie R przedstawiają odbicia od innych obiektów w pomieszczeniu, które nie są wyraźnie opisane w scenariuszu. Parametry promieni R w modelu kanału lobby hotelowego są podsumowane w

Parametr	Wartość
Promień post-kursora K współczynnik, K	10 dB
Czas rozpadu mocy po kursorze, $\gamma$	4,5 ns
Przybycie po kursie, $\lambda$	0,31 ns? 1
Promień po kursie rozkładu amplitudy	Rayleigh
Liczba promieni następczych, N	6

Parametry wewnątrz klastra dla scenariuszy dostępu do pomieszczeń są pobierane bezpośrednio z odpowiedniego scenariusza wewnętrznego opracowanego i są oparte na pomiarach eksperymentalnych

### 3.5.4 Modele kanału dosyłowego

Schemat backhaulingu w scenariuszu powyżej dachu (ART) przedstawia łącze typu backhaul między dwoma punktami dostępowymi ART, zwykle z antenami o bardzo dużym zysku i wysokiej kierunkowości. Podobnie, backhaul na poziomie ulicy reprezentuje połączenie między dwoma punktami dostępowymi, zwykle umieszczonymi w latarniach w kanionie ulicy. Zastosowanie anten kierunkowych prowadzi do absolutnej dominacji bezpośredniego promienia LOS, podczas gdy inne promienie (które mogą występować w tym środowisku) są znacznie słabsze. Zatem model kanału dosyłowego backhaul ART i Street Street powinien brać pod uwagę tylko bezpośredni promień LOS.



Jednocześnie bezpośrednia ścieżka może być częściowo zatkana (przez obiekty znajdujące się w granicach pierwszej strefy Fresnela) lub może się różnić ze względu na niestabilność atmosfery (temperatura, gęstość powietrza, mgła i inne efekty pogodowe). Aby wprowadzić te ważne dla efektów scenariuszy backhaulingu, struktura klastra jest dodawana do promienia LOS. Parametry klastra przedstawiono w tabeli

Parametr	Wartość
Promień post-kursora K współczynnik, K	10 dB
Czas rozpadu mocy po kursorze, $\gamma$	10 ns
Przybycie po kursie, $\lambda$	0,03 ns? 1
Promień po kursie rozkładu amplitudy	Rayleigh
Liczba promieni pozakresowych, N	4

### 3.6 Wniosek

W tej części zaproponowano nową, quasi-deterministyczną metodologię modelowania kanałów fal milimetrowych. Metodologia opiera się na wielu wcześniejszych pracach eksperymentalnych i kilku specjalnych kampaniach pomiarowych przy 60 GHz podczas projektu MiWEBA w różnych scenariuszach i środowiskach. Ponieważ pasmo 60 GHz znajduje się dokładnie w środku zakresu częstotliwości fal milimetrowych 30-90 GHz, wyniki przedstawione w tej pracy można ekstrapolować w obu kierunkach, dzięki czemu opracowana metodologia ma zastosowanie do tego zakresu częstotliwości. Kampania pomiarowa w kanionie ulicznym została przeprowadzona przy użyciu anten dookólnych w połączeniu z symulacjami śledzenia promieni w rzeczywistym środowisku ulicznym w Berlinie. Wyniki pomiarów wykorzystano do opracowania metodyki i oceny parametrów modelu. Inne pomiary przeprowadzono w idealnym środowisku otwartym (stare lotnisko), aby badać LOS i odbijać promienie od ziemi z ustawieniami anten kierunkowych. Ostatnia kampania została przeprowadzona przy użyciu anten kierunkowych w bardziej realistycznym scenariuszu otwartego obszaru kampusu uniwersyteckiego. Oceniono parametry gruntowego odbicia, rozproszenia i polaryzacji krzyżowej. Przeprowadzono dodatkowe badania w celu doświadczalnego badania wpływu ruchu UE na funkcję przenoszenia kanału. W rezultacie opracowano nowe podejście Q-D do modelowania kanałów zewnętrznych na częstotliwości 60 GHz. Metodologia ta oparta jest na reprezentacji funkcji odpowiedzi impulsowej na fali milimetrowej przez kilka silnych scenariuszospecyficznych deterministycznych promieni (promienie D) i pewną liczbę stosunkowo niewielkich promieni losowych (promienie R), które reprezentują losowe lub niezdefiniowane komponenty środowiska. Proponowana metodologia pozwala stworzyć modele kanałów dla nowych scenariuszy, wykorzystując jedynie podstawowe założenia dotyczące geometrii środowiska, warunków środowiskowych i mobilności użytkowników. Elastyczna kombinacja podejść deterministycznych i statystycznych ma rozsądną złożoność i dokładność porównywalną z pełnymi specyficznymi dla miejsca modelami deterministycznymi. Jako przykład zastosowania metodologii Q-D opracowano modele kanałów dla otwartego obszaru (kampus uniwersytecki), kanionu ulicy i lobby hotelowego. Odpowiednie parametry modelu dla łączy dostępowych zostały wybrane w oparciu o pomiary eksperymentalne i symulacje ray tracingu. Wyraźne wprowadzenie deterministycznych promieni D i losowych promieni R w ramach podejścia do modelowania kanału Q-D pozwoliło na właściwy opis rzeczywistego dynamicznego środowiska zewnętrznego, biorąc pod uwagę efekty blokowania i mobilności. Wszelstronność metodologii Q-D pozwala na łatwe rozszerzenie tworzonych modeli kanałów na inne przypadki użycia w tym samym środowisku. Na przykład model kanału dostępu do ulicy można zmodyfikować dla D2D i połączeń

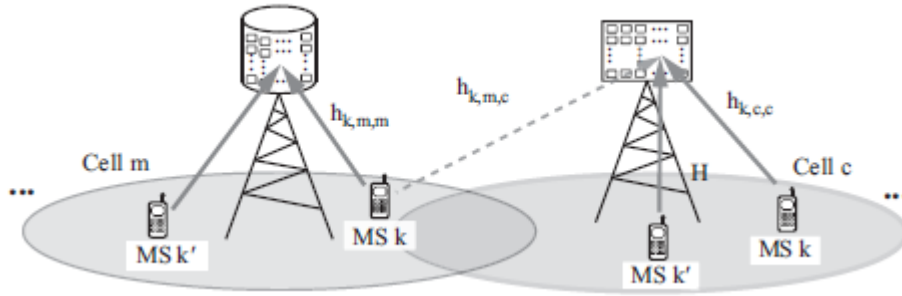
międzystanowych na poziomie ulicy. Na razie metodologia modelowania kanałów Q-D opracowana w ramach wspólnego projektu UE-Japonia w ramach 7PR MiWEBA została zaakceptowana jako podstawowe podejście do modelowania kanałów w nadchodzącym standardzie IEEE 802.11ay. Ten standard obejmuje nie tylko nowe scenariusze, ale także nowy model użycia, taki jak SU-MIMO i MU-MIMO. Problem ulepszenia modelu Q-D do wspierania trybów MIMO, zapewniający odpowiednią spójność przestrzenną i częstotliwościową, jest przedmiotem ostatnich prac, ale wymaga dodatkowych badań i ocen. Ponadto, modele kanałowe parametrów wewnątrz klastra powinny być udoskonalone dla wszystkich określonych scenariuszy i środowisk poprzez dodatkowe pomiary eksperymentalne przy użyciu sprzętu o wyższej rozdzielczości kątowej i czasowej.

## **Systemy antenowe o dużej skali**

Tu przedstawiamy jedną z najbardziej znanych technologii kandydujących 5G, czyli systemy antenowe o dużej skali (LSAS). Surowe wymagania dotyczące wydajności widmowej, narzucone zarówno przez rozwój nowych, wymagających dużej ilości danych aplikacji komunikacyjnych, jak i rosnącą liczbę urządzeń mobilnych, doprowadziły do zaprojektowania systemów komunikacyjnych w przeszłości, w teraźniejszości i najprawdopodobniej w dającej się przewidzieć przyszłości. Wprowadzenie systemów komunikacji z wieloma wejściami i wieloma wyjściami (MIMO) zrewolucjonizowało podejście do zwiększania prędkości transmisji danych. Rzeczywiście, systemy z wieloma antenami stanowią podstawę obecnych standardów komunikacji mobilnej, takich jak Long-Term Evolution (LTE). Oczekuje się jednak, że systemy komunikacyjne 5G zapewnią wysokie szybkości transmisji danych, ale powinny również to robić w sposób energooszczędny. Stworzyło to coraz większą dynamikę badań nad LSAS, gdzie redukcja kosztów i ulepszenia oferowane zarówno w sprzęcie, jak i komponentach przetwarzania sygnału stacji bazowych (BS), motywują dodanie większej liczby anten do dalszego zwiększania pojemności, przy jednoczesnym zmniejszeniu całkowitej mocy konsumpcja. LSAS, również powszechnie określane jako masywne systemy MIMO, ma na celu jednoczesną poprawę wydajności widmowej i energetycznej obecnych standardów mobilnych poprzez włączenie setek anten na stacjach bazowych (BS), aby obsługiwały dziesiątki użytkowników jednocześnie. Oczywiście, zatrudnienie tak dużej liczby anten mają znaczną liczbę implikacji na różnych poziomach projektowych od implikacji przepustowości kanału do implikacji sprzętowych i przetwarzania sygnału, takich jak projektowanie macierzy antenowych lub wymóg natychmiastowego informowania o stanie kanału (CSI) w celu osiągnięcia obiecanych ulepszeń. Ten szeroki zakres implikacji motywuje treść tego rozdziału, który koncentruje się na opisanu podstawowych cech LSAS i kładzie szczególny nacisk na najbardziej ważne praktyczne aspekty wdrażania. W szczególności sekcja 4.1 zawiera przegląd głównych zalet i wad LSAS: główne zalety obiecane przez duże systemy antenowe wraz z wyzwaniami związanymi z ich praktycznym wdrożeniem i działaniem. Następnie w Sekcji 4.2 opisano modele konwencjonalnie stosowane do charakteryzowania masowych systemów komunikacyjnych i podsumowano główne wyniki uzyskane w kampaniach pomiarowych. Wreszcie, sekcja 4.3 kończy tą część, opisując szereg istotnych aspektów, które należy uwzględnić przy optymalizacji i projektowaniu LSAS.

### **4.1 Obietnice i wyzwania LSAS**

Intuicyjnie zwiększając liczbę anten wykorzystywanych ogólnie do komunikacji poprawia się osiągalne wskaźniki ze względu na zwiększoną zyskowność zróżnicowania i multipleksowania. Jednakże korzyści przewidziane przez zastosowanie LSAS wykraczają poza powyższą koncepcję, wykorzystując szczególne cechy kanałów komunikacyjnych na dużą skalę, jak wyszczególniono poniżej. Aby ułatwić ekspozycję, zasady LSAS są zilustrowane w tym rozdziale przez skupienie się na przesyłaniu sygnałów ze stacji mobilnych (MS) w kierunku BS (etap łącza wstępującego), ponieważ przedstawione tu podstawowe koncepcje są bezpośrednio przetwarzalne dla łącza w dół. Rozważmy system komunikacji mobilnej złożony z komórek M, jak pokazano na rysunku 4.1.



W tym ogólnym systemie każda stacja BS jest wyposażona w  $N$  anten i jednocześnie obsługuje użytkowników  $K$  z jedną anteną. Wyróżniającą cechą LSAS w odniesieniu do obecnie wdrażanych systemów jest to, że  $N$  może być rzędu kilkudziesięciu lub kilkuset anten, tj. relacja  $N \gg K$  jest zwykle spełniona. Złożony wąskopasmowy model pasma częstotliwościowego nieskorelowanego kanału częstotliwości w powyższym systemie jest zazwyczaj zapisywany jako

$$\mathbf{h}_{k,c,m} = \sqrt{d_{k,c,m}} \mathbf{g}_{k,c,m}$$

gdzie  $\mathbf{h}_{k,c,m} \in \mathbb{C}^{N \times 1}$  reprezentuje kanał komunikacyjny między kodem użytkownika umieszczony w  $k$ -tej komórce i  $N$  antenach stacji bazowej znajdującej się w komórce  $m$ -tej. Ponadto  $d_{k,c,m}$  charakteryzuje składnik blaknięcia o dużej skali, który może obejmować zarówno utratę ścieżki, jak i zanikanie cienia, podczas gdy  $\mathbf{g}_{k,c,m} \in \mathbb{C}^{N \times 1}$  oznacza niskoskalowaną składową kanału zanikającego, kanału między  $k$ -tym użytkownikiem znajdującym się w  $c$ -tej komórce i stacją bazową która zapewnia zasięg użytkownikom w  $m$ -tej komórce. Powyższy model kanału zakłada, że komponent blaknący na dużą skalę jest identyczny dla wszystkich anten na stacji bazowej, co jest rozsądnym czynnikiem, gdy wymiary matrycy antenowej są niewielkie w porównaniu z odległością od użytkowników. Zauważmy, że chociaż zakładamy, że transmitowana jest pojedyncza nośna dla uproszczenia zapisu, model ten można łatwo rozszerzyć na szeroko stosowane ortogonalne zwielenokrotnianie z podziałem częstotliwości (OFDM). Rzeczywiście,  $\mathbf{h}_{k,c,m}$  dokładnie charakteryzuje wąskopasmowy kanał propagacji w danej podnośnej transmisji OFDM. Często przydatne jest zastosowanie złożonego kanału komunikacyjnego  $\mathbf{H}_{c,m} \in \mathbb{C}^{N \times K}$  między  $K$  MS w komórce  $c$  a  $m$ -tą stacją bazową, zdefiniowanym tak:

$$\mathbf{H}_{c,m} = (\mathbf{h}_{1,c,m}, \mathbf{h}_{2,c,m}, \dots, \mathbf{h}_{K,c,m}), \quad \forall c, m \in \{1, \dots, M\}.$$

Podobnie  $\mathbf{D}_{c,m} = \text{diag}(d_{1c,m}, d_{2c,m}, \dots, d_{Kc,m})$ ,  $\forall c, m \in \{1, \dots, M\}$  i  $\mathbf{G}_{c,m} = \text{diag}(g_{1c,m}, g_{2c,m}, \dots, g_{Kc,m})$ ,  $\forall c, m \in \{1, \dots, M\}$ . W całej tej sekcji zapisy  $\mathbf{G}_{c,m}$  są uważane za niezależne i identycznie rozdzielonymi zmiennymi losowymi spełniającymi  $\mathbb{E}[\mathbf{G}_{c,m} \mathbf{G}_{c,m}^H] = \mathbf{I}_N \quad \forall c, m \in \{1 \dots M\}$ . W poprzednim wyrażeniu  $\mathbb{E}[\cdot]$  to operator oczekiwania a  $\mathbf{I}_N$  reprezentuje  $N \times N$  macierz tożsamości. Zwróć uwagę, że na tym etapie nie ma żadnych założeń w konkretnym rozkładzie, z którego pochodzą współczynniki kanału. Sygnał odebrany przez  $c$ -ty BS,  $\mathbf{y}_c \in \mathbb{C}^{N \times 1}$ , przy danym użyciu kanału dla transmisji symbolu łączą zwrotnego, jest zwykle wyrażany jako

$$\mathbf{y}_c = \sum_{m=1}^M \sqrt{\rho} \mathbf{H}_{c,m} \mathbf{x}_m + \mathbf{n}.$$

gdzie  $x_m \in \mathbb{C}^{K \times 1}$  oznacza złożony sygnał transmitowany przez stacje MS w m-tej komórce, a  $n \in \mathbb{C}^{N \times 1}$  reprezentuje standardowy addytywny biały wektor szumu Gaussa (AWGN) z wpisami spełniającymi  $n_i \sim \mathcal{CN}(0,1)$ . Sygnał transmitowany przez k-tego użytkownika jest ograniczony, aby spełnić średnie ograniczenie mocy  $\mathbb{E}[|x_{m,k}|^2]$ . Co więcej,  $\rho$  charakteryzuje stosunek między średnią mocą transmisji a średnią mocą szumu. W poprzednich wyrażeniach  $\sim$  można odczytać jako "rozpowszechniany jako" a  $\mathcal{CN}(a,b)$  reprezentuje zmienną losową po złożonym rozkładzie normalnym ze średnią  $a$  i wariancją  $b$ . Koncepcje przedstawione w tej sekcji, o ile nie zaznaczono inaczej, wprowadza się rozważając system jednokomórkowy ( $M = 1$ ), gdy wystarczy ujawnić szczególny wgląd w LSAS, a subindeksy  $c$  i  $m$  są pominięte dla uproszczenia. Poniżej omawiamy kluczowe obietnice z teoretycznego studium LSAS, wraz z głównymi wyzwaniem w ich praktycznym wdrażaniu.

#### 4.1.1 Obietnice LSAS

##### 4.1.1.1 Korzystna propagacja i utwardzanie kanałów

Spektralne zyski wydajności oferowane przez LSAS zasadniczo opierają się na dwóch zjawiskach propagacji, które występują, gdy  $N$  rośnie duży: tak zwana korzystna propagacja i efekty umocnienia kanału. Wykazano, że w domenie asymptotycznej, gdy  $N$  rośnie w stosunku do  $K$ , LSAS z  $N \gg K$  doświadcza sprzyjających warunków propagacji, tzn.

$$\mathbf{h}_k^H \mathbf{h}_{k'} = \sqrt{d_k} \mathbf{g}_k^H \mathbf{g}_{k'} \sqrt{d_{k'}} \underset{N \gg K}{\approx} 0, \forall k, k' \in \{1, \dots, K\} \text{ and } k \neq k'.$$

i

$$\mathbf{h}_k^H \mathbf{h}_k = \sqrt{d_k} \mathbf{g}_k^H \mathbf{g}_k \sqrt{d_k} \underset{N \gg K}{\approx} N d_k, \forall k \in \{1, \dots, K\}.$$

W powyższym wyrażeniu  $(\cdot)^H$  oznacza hermitowski wektor lub macierz koniugatu. Zwróć uwagę, że powyższe dwa wyrażenia oznaczają, że:

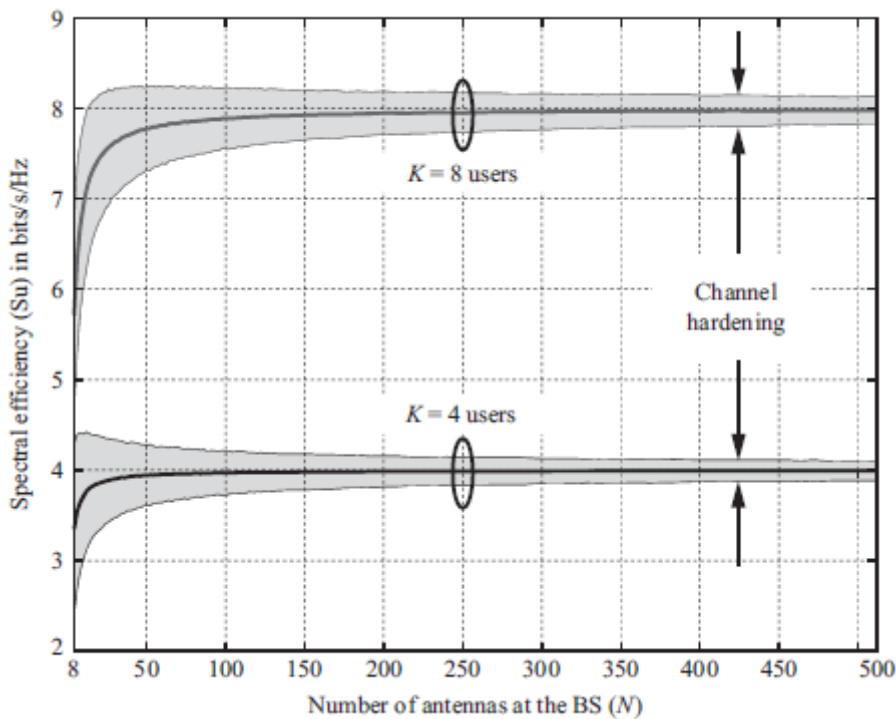
$$\mathbf{H}^H \mathbf{H} = \mathbf{D}^{1/2} \mathbf{G}^H \mathbf{G} \mathbf{D}^{1/2} \approx N \mathbf{D}.$$

Mówiąc wprost, korzystne warunki propagacji są odczuwane, gdy kanały komunikacyjne różnych użytkowników stają się ortogonalne zanikający składnik małej skali dla  $\mathbf{h}_k^H \mathbf{h}_{k'}$ . Fizycznie, ortogonalność kanału pojawia się w wyniku posiadania dużej liczby anten w BS i może być interpretowana w następujący sposób: im bliższy ortogonalny kanał użytkownika, tym większa różnica w kanałach komunikacyjnych różnych użytkowników. Ułatwia to przestrzenne rozdzielanie strumieni danych przesyłanych przez różnych użytkowników po stronie BS, ponieważ ich kanały stają się odrębne. Rzeczywiście, korzystne warunki propagacji mają bezpośredni wpływ na przepustowość łącza uplink w systemie MIMO dla wielu użytkowników, określanym jako  $C_u$ . W szczególności pojemność łącza w górę mierzona w bitach na sekundę (bitów / s) jest podana przez

$$\begin{aligned} C_u &= B \log_2 \det(\mathbf{I}_K + \rho \mathbf{H}^H \mathbf{H}) \approx B \log_2 \det(\mathbf{I}_K + N \rho \mathbf{D}) \\ &= B \sum_{k=1}^K \log_2(1 + N \rho d_k) \text{ bits/s,} \end{aligned}$$

gdzie  $B$  oznacza szerokość pasma systemowego w hercach (Hz), a  $\det(\cdot)$  jest operatorem wyznacznika. Powyższe wyrażenie uzyskuje się, zakładając, że LSAS BS ma doskonałą znajomość macierzy kanału  $\mathbf{H}$ . Warto odnotowania jest, że, ostatnia równość nie zależy od małych współczynników zanikania,  $g_k$ ,

które występują z powodu korzystnej propagacji warunki przedstawione .W prostych słowach, skutki zanikania na małą skalę uśredniono, co jest jedną z głównych motywacji wdrożenia LSAS w systemach 5G. Ostatnia równość pojemności łącza w górę również ujawnia jego bezpośrednią zależność od wielkoskalowych współczynników zanikania,  $d_k$ ; które zazwyczaj różnią się znacznie wolniej niż parametry blaknięcia na małą skalę . Istotnie, zmiany możliwych do osiągnięcia wskaźników w czasie są tradycyjnie uważane za jedyną funkcję współczynników zanikania na małą skalę. Ponieważ te parametry nie wpływają na ekspresję , osiągalne szybkości powodują zmniejszenie czasowych zmian w miarę wzrostu liczby anten BS, zjawisko określane jako umacnianie kanału. Innymi słowy, chwilowe osiągalne wskaźniki są zbliżone do osiągalnych współczynników ergodycznych . Ogólnie rzecz biorąc, zjawisko hartowania kanałów wskazuje, że wdrożenie LSAS w BS systemów komórkowych może ułatwić spójne dane użytkowników w długim okresie czasu, poprzez usunięcie różnic wprowadzanych przez efekty znikające na małą skalę, co poprawia jakość doświadczeń użytkowników. (QoE). Przeanalizujemy powyższą obserwację, korzystając z przykładu pokazanego na rysunku 4.2



Tutaj ewolucja średniej wydajności spektralnej,  $S_u = C_u / B$ , jest pokazana dla rosnącej liczby anten BS dla przypadków  $K = 4$  i  $K = 8$ . Jest to oznaczone pogrubionymi liniami ciągłymi na rysunku . Rysunek pokazuje również 90-procentowe przedziały ufności, które wiążą region zawierający 90% wartości chwilowej wydajności spektralnej. Są one oznaczone przez zacienione obszary wokół linii ciągłych dla każdego z  $K = \{4,8\}$  przypadków. W celu zilustrowania, rozważymy tutaj zanikanie Rayleigha, tj.  $i$ -te wejście wektora kanału następuje po  $h_{k,i} \sim \mathcal{CN}(0,1)$  i pomija efekt blaknięcia na dużą skalę ( $D = l_k$ ). Co więcej,  $\rho = 1/N$ ; tj. moc transmisji jest niezależna od liczby anten nadawczych  $N$ , ale wzrasta proporcjonalnie do liczby użytkowników  $K$ . Wyniki rysunku pokazują, że 90% przedziały ufności, tj. grubość zacienionych obszary są zmniejszane w miarę wzrostu  $N$  dla obu  $K = \{4,8\}$ , co oznacza, że użytkownik ocenia stawki różnią się mniej w odniesieniu do średniej, zgodnie z oczekiwaniami efektu utwardzania kanałów. Dodatkowo można zauważyć, że maksymalne średnie stawki wynoszą w przybliżeniu osiągnięte dla rozsądnych skończonych wartości  $N \approx 50$  dla przypadku  $K = 4$ , a  $N \approx 100$  anten jest wymaganych dla  $K = 8$ . Dzieje się tak, ponieważ potrzebna jest większa liczba anten do przybliżenia korzystnych warunków propagacji, gdy rośnie  $K$ , trend, który jest również spójny, gdy rozważane jest łącze w dół. Powyższe wyniki pokazują, że zarówno korzystna propagacja, jak i

umocnienie kanałów obiecanie przez LSAS mają wyraźny pozytywny wpływ na wskaźniki komunikacji. Jednak złożoność systemu pod względem sprzętu i oprogramowania powinna być również brana pod uwagę przy charakteryzowaniu korzyści osiągniętych przez LSAS. Na przykład znaczny wzrost liczby anten może ogromnie skomplikować przetwarzanie sygnału wymagane w BS dla transmisji symboli (wstępne kodowanie) i wykrywania, które są niezbędne dla doświadczania zarówno korzystnego propagowania, jak i utwardzania kanałów w praktycznych systemach. Okazuje się jednak, że najprostsze liniowe prekodery i detektory stają się optymalne z asymptotycznie dużą liczbą anten BS, co upraszcza działanie LSAS. Omawiamy ten aspekt w poniższej sekcji.

#### 4.1.1.2 Liniowe przetwarzanie sygnału i idealne prawo skalowania mocy

Przetwarzanie sygnału pasma podstawowego w łączu komunikacyjnym jest zwykle przeprowadzane w postaci wstępnego kodowania i wykrywania sygnału. Konkretnie, schematy wstępnego kodowania są wykorzystywane w BS podczas transmisji zstępującej. Mają one na celu kształtowanie transmitowanych symboli w BS zgodnie ze współczynnikami kanału, tak, że symbole informacyjne mogą być dokładnie wykrywane przez proste przetwarzanie w stwardnieniu rozsianym. Podobnie, schematy detekcji są wykorzystywane dla stacji BS podczas transmisji łącza uplink w celu określenia oryginalnych symboli przesyłanych przez terminale użytkowników. Powyższa procedura oferuje korzystny efekt przesunięcia złożoności przetwarzania sygnału do stacji bazowych, upraszczając w ten sposób złożoność na terminalach użytkownika, gdzie możliwości obliczeniowe i dostarczanie energii są bardziej ograniczone. Schematy kodowania wstępnego i wykrywania można podzielić na liniowe i nieliniowe w zależności od zastosowanych operacji przetwarzania sygnałów, co dodatkowo wpływa na ich działanie złożoność przetwarzania sygnału. Ogólnie rzecz biorąc, istnieje zasadniczy kompromis między nimi, tj. nieliniowe wstępne kodowanie i schematy detekcji zwykle oferują lepszą wydajność w porównaniu z liniowymi kosztami wymagającymi bardziej skomplikowanego przetwarzania sygnału. Jednak rzeczy zmieniają się w obszarze LSAS, gdzie nawet najprostsze liniowe programy do wstępnego kodowania i wykrywania stają się asymptotycznie optymalne. Pokazujemy to wyraźnie w poniższym, dla reprezentatywnego przypadku transmisji w górę. Poprzez zastosowanie liniowej operacji wykrywania na odebranym sygnale wyrażona przez macierz detekcji  $W \in \mathbb{C}^{N \times N}$ , sygnał wyjściowy może być wyrażony jako

$$\mathbf{r} = \mathbf{W}\mathbf{y} = \mathbf{W}(\sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{n}),$$

gdzie  $\mathbf{r} \in \mathbb{C}^{K \times 1}$  jest wektorem kolumnowym zawierającym szacunki przesyłanych symboli  $K$  przez terminale użytkownika z jedną anteną, tj.  $r_k$  reprezentuje estymację symbolu transmitowanego przez komunikator użytkownika w danym użyciu kanału. Zauważmy, że dla uproszczenia notacji przyjęliśmy scenariusz z pojedynczą komórką, tj.  $m = 1$ ). Rozważmy najprostszy przypadek, w którym  $W = H^H$ , który jest powszechnie określany jako detekcja dopasowanego filtrowania (MF) lub maksymalnego współczynnika łączenia (MRC). Następnie sygnał przetworzony, w sprzyjających warunkach propagacji, można rozszerzyć jako

$$\mathbf{r} = \mathbf{H}^H(\sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{n}) \approx N\sqrt{\rho}\mathbf{D}\mathbf{x} + \mathbf{H}^H\mathbf{n}.$$

Pierwszy wyraz w podsumowaniu końcowym,  $N\sqrt{\rho}\mathbf{D}\mathbf{x}$ , odpowiada sygnałowi składniak, podczas gdy drugi wyraz,  $\mathbf{H}^H\mathbf{n}$ , reprezentuje komponent szumowy powstały po wykryciu liniowym. Należy zauważyć, że  $H$  jest statystycznie niezależne od  $\mathbf{n}$ . Dlatego stosunek sygnału do szumu (SNR)  $k$ -tego użytkownika można wyrazić jako

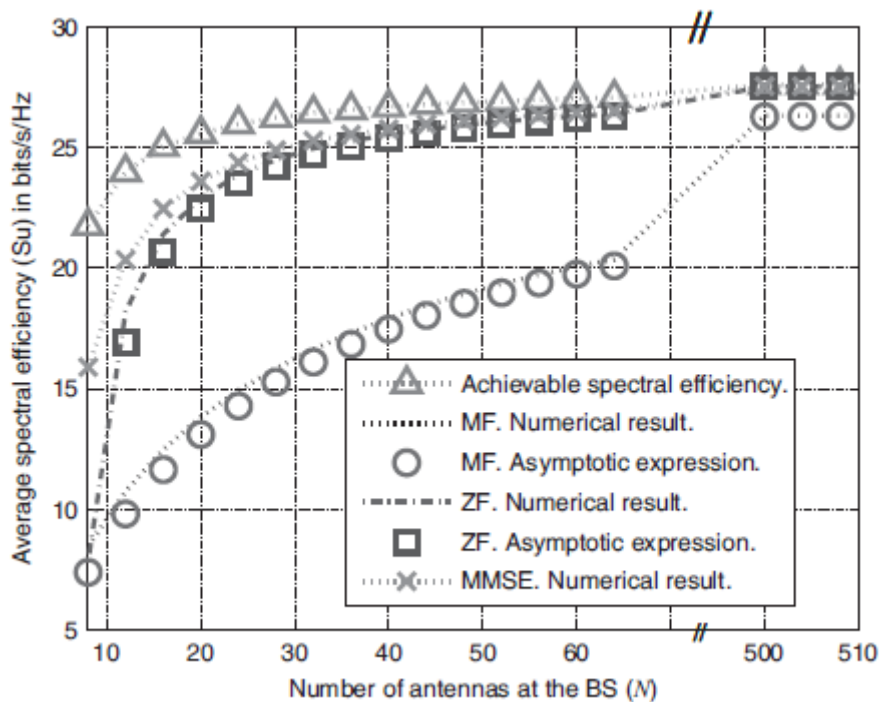
$$\text{SNR}_k = \frac{\mathbb{E}[|N\sqrt{\rho}d_k x_k|^2]}{\mathbb{E}[|\mathbf{h}_k^H \mathbf{n}|^2]} = \frac{N^2 \rho d_k^2 \mathbb{E}[|x_k|^2]}{d_k \mathbb{E}[|\mathbf{g}_k^H \mathbf{n}|^2]} = N \rho d_k,$$

gdzie oczekiwanie jest podejmowane na małej skali blaknięcie elementów  $g_k$ , transmitowane symbole  $x_k$  i komponent szumu  $n$ . Zwracamy uwagę, że jest to uzasadnione faktem, że statystyki dotyczące kanałów przestrzennych, które obejmują komponenty blaknięcie na dużą skalę, zmieniają się z mniejszą szybkością (kilka okresów spójności kanałów), gdy w porównaniu z poprzednimi czynnikami, które zmieniają się w tempie porównywalnym do wskaźnika symboli. Zauważ, że z ostatniej równości wynika z założenia, że  $\mathbb{E}[|x_k|^2]$ . oraz z korzystnej charakterystyki propagacji, omówione w poprzedniej sekcji. Rzeczywiście, dla korzystnych warunków propagacji, mamy  $\mathbb{E}[|\mathbf{g}_k^H \mathbf{n}|^2] = N$  dla jednostkowego szumu wariancji przyjętego w modelu. Powyższy SNR generuje szybkość łącza w górę dla użytkownika:

$$R_k = B \log_2(1 + \text{SNR}_k) = B \log_2(1 + N \rho d_k) \text{ bits/s.}$$

Należy zauważyć, że szybkość łącza w górę na użytkownika  $R_k$  z wykrywaniem MF jest identyczna z maksymalną osiągalną szybkością, jaką  $k$ -ty użytkownik doświadcza w korzystnych warunkach propagacji przedstawione. Innymi słowy, zastosowanie najprostszej strategii przetwarzania sygnału może teoretycznie osiągnąć maksymalne szybkości w korzystnych warunkach propagacji, które stanowią jeden z najbardziej znaczących aspektów LSAS. Z fizycznego punktu widzenia wynikająca intuicja jest taka, że duża liczba anten w BS umożliwia skoncentrowanie energii na użytkownikach. szczególne kierunki przestrzenne. Wynik ujawnia inny zauważalny aspekt LSAS: liniowy wzrost efektywnej wartości stosunku sygnał / szum SNR z liczbą anten w BS  $N$ , gdy do wykrywania wykorzystywany jest MF z doskonałym CSI. Innymi słowy, wymagana moc transmisji do uzyskania pewnej skali SNR odwrotnie proporcjonalnie do liczby anten  $N$  w BS, w korzystnych warunkach propagacji. Jest to określane jako idealne prawo skalowania mocy LSAS, które może być osiągnięte tylko wtedy, gdy idealna informacja o stanie kanału (CSI) jest dostępna w BS i bezpośrednio generalizuje inne schematy wstępnego i detekcyjnego]. Prawo dotyczące skalowania mocy ma również ważną implikację w projektowaniu sprzętu LSAS: ponieważ moc wykorzystywana do transmisji może zostać zmniejszona, aby osiągnąć wydajność systemu z mniejszą liczbą anten, transceivery mogą być zbudowane z częstotliwością radiową (RF) elementy o niższym zakresie dynamicznym, a zatem mogą być tańsze i bardziej energooszczędne. Co więcej, niezawodność LSAS jest również zwiększona w porównaniu z ich odpowiednikami w małej skali, ponieważ awaria jednego z wielu nadajników-odbiorników RF nie wpływa znacząco na wydajność całego systemu. Powyższe wyniki uzyskuje się jednak przy idealistycznych założeniach korzystnych warunków propagacji i doskonałej CSI w BS. W tym kontekście do tej pory nie określono wyraźnie wymaganej liczby anten, dla których korzystne warunki propagacji wchodzi w życie. Co więcej, dobrze wiadomo, że różne strategie wstępnego kodowania i wykrywania zapewniają różne osiągalne szybkości, gdy skończoną liczbę anten rozważa się w BS [19,20,23]. Aspekt ten można wyraźnie zaobserwować na rysunku 4.3,





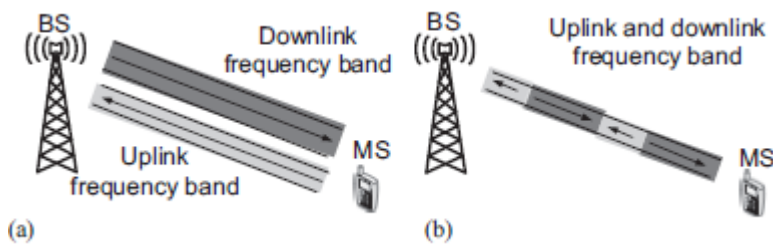
gdzie pokazano możliwe do osiągnięcia wskaźniki różnych liniowych schematów wykrywania w porównaniu z osiągalną wydajnością widmową kanału, dla rosnącej liczby anten BS  $N$ ,  $K = 8$  użytkowników i  $\rho = 10/N$ . Rozważane schematy wykrywania są już wprowadzone MRC ( $W = H^H$ ), detektor z wymuszeniem zera (ZF) ( $W = H^\dagger = (H^H H)^{-1} H$ ) i detektor minimalnego średniego kwadratu błędu (MMSE) ( $W = H^H H + \xi I_K)^{-1} H^H$ . W poprzednich wyrażeniach,  $(\cdot)^\dagger$  i  $(\cdot)^{-1}$  oznaczają odpowiednio pseudodwrotność i odwrotność macierzy, a  $\xi > 0$  jest stałą, która zasadniczo zależy od systemu SNR. Dodatkowo, wyniki numeryczne są porównywane z prostymi wyrażeniami asymptotycznymi. Wyniki pokazane na rysunku ilustrują, że liniowe schematy wstępnego kodowania wykonują w pobliżu maksymalnej sumarycznej wydajności widmowej, gdy  $N$  rośnie. Konkretnie, można zauważyć, że schematy detekcji ZF i MMSE zbliżają się w przybliżeniu do osiągalnej sumarycznej wydajności widmowej przy niższych wartościach  $N$  anten w BS, podczas gdy detekcja MF staje się bliska optimum przy wyższych wartościach  $N$ . Figura 4.3 pokazuje także, że aproksymacje asymptotyczne dokładnie scharakteryzować średnią wydajność widmową dla stosunkowo małej liczby anten BS; stała w teoretycznych wynikach uzyskanych dla LSAS. Warto zauważyć, że wyniki wyprowadzone w tej sekcji zakładają, że doskonała znajomość współczynników kanałowych jest dostępna w BS. Ponieważ zbieranie dokładnych danych CSI jest jednym z najważniejszych aspektów LSAS, poniżej opisujemy procedurę pozyskiwania CSI, jak również jej wpływ na wynikową wydajność.

## 4.1.2 Wyzwania LSAS

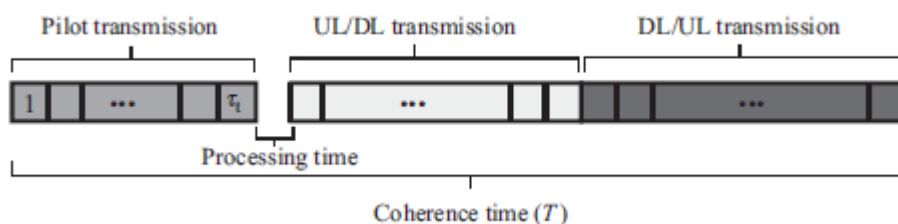
### 4.1.2.1. Uzyskiwanie informacji o stanie kanału w LSAS

Nabycie natychmiastowego CSI jest niezbędne w LSAS do wykorzystania nadmiernej liczby anten i osiągnięcia korzystnych warunków propagacji za pomocą technik liniowego wstępnego kodowania i wykrywania. Ta procedura nie była rozważana w poprzednich rozdziałach, w których z przyczyn ilustracyjnych przyjęto, że dokładny CSI jest dostępny w dużej BS bez ponoszenia kosztów. Jednak to idealistyczne założenie nie zachodzi w systemach rzeczywistych, ponieważ część czasu spójności kanału musi być wykorzystana do akwizycji CSI, a niedoskonałe szacunki kanału są nieuchronnie tworzone w wyniku tej procedury. Rzeczywiście, duża liczba anten rozmieszczonych w LSAS promuje zmianę filozofii w akwizycji CSI. Szczegółowo poniżej. W całej tej części rozdziału rozważamy blokowy model

kanалу, w którym kanał pozostaje stały podczas wielu okresów symboli, które pokrywają się z długością ramki i zmienia się niezależnie między ramkami. Ograniczona liczba symboli, dla których kanał może być uważany za stały,  $\tau = B_c T_c$ , określa czas, jaki BS może wydać na akwizycję CSI przed transmisją danych i odbiorem. W poprzednim wyrażeniu  $B_c$  oznacza szerokość pasma koherencji kanału a  $T_c$  oznacza czas koherencji kanału. Ograniczony czas koherencji jest powodem, dla którego większość badań LSAS skupiła się na transmisji TDD (ang. Time division duplex) z rysunku 4.4 (a)



,w której zarówno transmisja danych, jak i odbiór odbywają się w tym samym paśmie częstotliwości. TDD jest preferowany w przypadku LSAS, ponieważ długość interwału treningu  $\tau_t$  jest proporcjonalna do liczby aktywnych końcówek  $K$  i zamiast tego jest niezależna od dużej liczby anten rozmieszczonych w stacjach BS. Kontrastuje to z systemami dupleksowania z podziałem częstotliwości (FDD) z rysunku 4.4 (b), gdzie w bogatych warunkach rozpraszania długość treningu i sprzężenia wzrasta wraz z  $N$ . Dzięki wzajemności kanału łącza nadawczego / łącza w dół w transmisji TDD, kanał łącza w dół może być dokładnie oszacowany podczas akwizycji CSI łącza wstępującego, pod warunkiem, że końce komunikacji nadawania i odbioru są dokładnie skalibrowane. Bezsprzecznie, przy tym ostatnim założeniu wynika, że  $H^d = (H_u)^h$ , gdzie wskaźniki górne 'd' i 'u' oznaczają odpowiednio łącze "w dół" i "w górę". W związku z tym operacja TDD ułatwia akwizycję CSI w otwartej pętli, co czyni ją szczególnie atrakcyjną dla LSAS wdrożenia. Podczas gdy operacja FDD przyciągnęła również znaczne zainteresowanie badawcze w najnowszej literaturze LSAS, ze względu na jego przewagę koncentrujemy się na transmisji TDD w następujący sposób. W trybie TDD, odstęp czasu koherencji kanału  $T_c$ , który zasadniczo zależy od prędkości, z jaką poruszają się terminale użytkownika, jest podzielony na trzy fazy: szkolenie pilotów, transmisja danych i odbiór danych. Uproszczona ilustracja protokołu działania systemów TDD przedstawiona jest na rysunku 4.5.



Początkowo BS nabywa CSI za pośrednictwem sygnałów pilota łącza zwrotnego transmitowanych z terminali użytkownika w procesie, który zajmuje  $\tau_t \geq K$  kolejnych okresów symboli. Po pobraniu informacji pilota, stacja BS wymaga krótkiego czasu przetwarzania, w którym CSI jest obliczany na podstawie sygnałów pilota odebranych w poprzedniej fazie. Warunek  $\tau_t \geq K$  w procesie akwizycji CSI pozwala na ortogonalność pomiędzy sekwencją pilotującą  $k$ -tego użytkownika,

$$\theta_k = \sqrt{\rho_{tr}} [\theta_k(1), \dots, \theta_k(\tau_t)]^T$$

i sygnał pilota reszty użytkowników wewnątrzkomórkowych. W poprzednim wyrażeniu  $\rho_{tr} = \rho \tau_t$

reprezentuje efektywne szkolenie SNR, które zależy zarówno od czasu spędzonego na transmisji pilota łączy zwrotnego, jak i znormalizowanego SNR transmisji dla MS. Matematycznie, piloci powinni idealnie spełniać:

$$\theta_k^H \cdot \theta_{k'} = \begin{cases} 1, & \text{if } k = k', \\ 0, & \text{if } k \neq k'. \end{cases}$$

Sygnał odebrany w BS po fazie szkolenia pilotów  $Y_{tr} \in \mathbb{C}^{N \times K}$ , w pojedynczej komórce jest podana przez

$$Y_{tr} = \sqrt{\rho_{tr}} \mathbf{H} \Theta + \mathbf{N},$$

gdzie  $\Theta \in \mathbb{C}^{K \times K}$  jest definiowane jako  $\Theta = [\theta_1 \dots \theta_K]^T$  a  $\mathbf{N} \in \mathbb{C}^{N \times K}$  jest macierzą AWGN utworzoną przez niezależne i identyczną dystrybucję (i.i.d.) standardowych wejść Gaussa. Odebrane piloty są następnie przetwarzane przez skorelowanie ich z pilotami znanymi z BS. Wynikowy szacunek kanału  $\hat{\mathbf{H}} \in \mathbb{C}^{N \times K}$  można następnie wyrazić jako

$$\hat{\mathbf{H}} = Y_{tr} \Theta^H = (\sqrt{\rho_{tr}} \mathbf{H} \Theta + \mathbf{N}) \Theta^H = \sqrt{\rho_{tr}} \mathbf{H} + \mathbf{N} \Theta^H.$$

gdzie ostateczna równość wynika z  $\Theta \Theta^H = \mathbf{I}_K$ . Zauważ, że rozkład szumu i jego właściwości statystyczne nie są objęte powyższą operacją, ponieważ  $\Theta$  jest jednostkowe. Powyższa statystyka decyzyjna  $\hat{\mathbf{H}}$  może być następnie wykorzystana do oszacowania kanału komunikacji za pomocą tradycyjnych metod szacowania. Na przykład powszechnie stosowana ocena MMSE jest podana przez

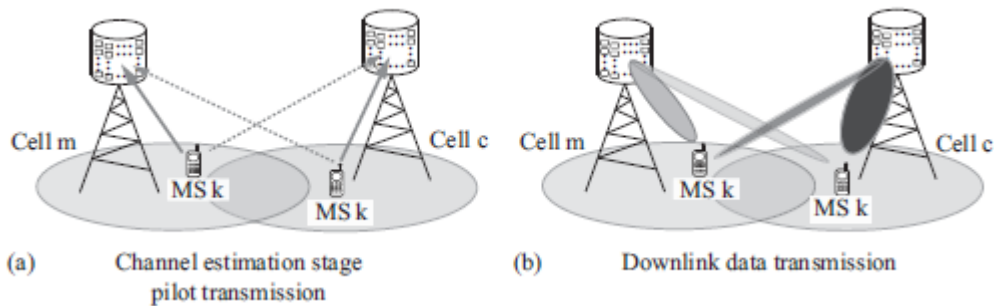
$$\hat{\mathbf{H}} = \left( \mathbf{H} + \frac{1}{\sqrt{\rho_{tr}}} \mathbf{N} \right) \hat{\mathbf{D}},$$

gdzie  $\hat{\mathbf{D}} = \left( \frac{1}{\sqrt{\rho_{tr}}} \mathbf{D}^{-1} + \mathbf{I}_K \right)^{-1}$  jest macierzą diagonalną. Należy zauważyć, że zastosowanie realistycznych niedoskonałych oszacowań CSI szkodzi idealnemu zachowaniu LSAS z perfekcyjnym CSI opisanym w sekcji 4.1.1, gdzie podkreślono, że skuteczny SNR doświadczył liniowego wzrostu z  $N$ . Zamiast tego można wykazać, że efektywny stosunek SNR rośnie proporcjonalnie do  $\sqrt{N}$  kiedy wynikowe oszacowanie MMSE stosuje się do liniowego wstępnego kodowania i wykrywania. Proces akwizycji kanału może ograniczyć ogólną wydajność praktycznego LSAS ze względu na konieczność zaspokojenia  $\tau_t \geq K$  w celu uzyskania precyzyjnego CSI wewnątrzkomórkowego. Na przykład, okres koherencji składający się z  $\tau \approx 200$  symboli. W przypadku scenariuszy z użytkownikami o wysokiej mobilności lub wysokimi częstotliwościami można spodziewać się 200 symboli. Ten skrócony okres koherencji może uniemożliwić przypisanie ortogonalnych sekwencji pilotowych wszystkim użytkownikom wewnątrzkomórkowym, wywołując znaczący błąd w estymowanym kanale, co z kolei wpływa na osiągalne szybkości transmisji i odbioru. Jednakże, gdy ortogonalność między każdą parą sekwencji pilota użytkownika jest zachowana, zastosowanie estymatora MMSE umożliwia uzyskanie jednego z najbardziej znaczących teoretycznych wyników LSAS: Zarówno szum termiczny, jak i niedoskonały CSI z powodu asymptotycznego zanikania kanału ( $N \rightarrow \infty$ ) w przypadku zastosowania liniowych prekoderów i detektorów. Intuicyjnie dzieje się tak, ponieważ zarówno błąd szumu, jak i błąd oszacowania kanału są niezależne od dostępnej estymacji kanału  $\hat{\mathbf{H}}$ . Innymi słowy, wydajność LSAS z pojedynczą komórką z prostymi schematami liniowego wstępnego kodowania i wykrywania jest asymptotycznie optymalna, nawet w przypadku niedoskonałego CSI. Naturalnie ten warunek nie

występuje w praktyce, ponieważ liczba anten jest skończona, a stacje BS są rozmieszczone w różnych komórkach. Dlatego naturalne jest pytanie, czy ten idealny warunek byłby nadal spełniony w scenariuszach wielokomórkowych z asymptotycznie dużą liczbą anten. Badamy ten aspekt w następujący sposób.

#### 4.1.2.2 LSAS w systemach wielokomórkowych: zanieczyszczenie pilotażowe

Czas koherencji zwykle nie jest wystarczająco długi, aby zapewnić ortogonalność między sekwencjami pilotowymi użytkowników znajdujących się w różnych komórkach. Zmusza to do użycia sekwencji pilotażowych, które generują tzw. efekt skażenia pilotażowego, który jest konwencjonalnie uważany za ostateczny limit wydajności LSAS. Zanieczyszczenia pilotażowe pojawiają się, gdy dane BS m-które idealnie odebrałyby informację o kanale dla użytkowników wewnątrz-komórek-uzyska niedokładne CSI podczas fazy treningu, ponieważ nie może w pełni zdekorelować sygnałów pilota generowanych przez użytkowników innych komórek, które dzielą tę samą sekwencję pilotującą. Rzeczywiście, najgorszy scenariusz, w którym identyczne sekwencje pilotujące są jednocześnie ponownie wykorzystywane we wszystkich komórkach, jest zwykle rozważany w literaturze. Fizycznie, efekt skażenia pilota powoduje, że zaatakowane stacje BS częściowo transmitują swoje sygnały do użytkowników mobilnych umieszczonych w innych komórkach, które dzielą tę samą sekwencję pilota. Innymi słowy, niedokładne oszacowanie kanału wewnątrzkomórkowego powoduje interferencję międzykomórkową, jak pokazano na figurze 4.6 (b) i opisano poniżej.



Po procedurze podobnej do wyprowadzenia metryki decyzyjne uzyskane na cth BS po korelacji pilota w układzie wielokomórkowym mogą być wyrażone tak :

$$\hat{\mathbf{H}}_{c,c} = \sqrt{\rho_{tr}} \mathbf{H}_{c,c} + \sqrt{\rho_{tr}} \sum_{m \neq c} \mathbf{H}_{m,c} + \mathbf{N}\Theta^H,$$

gdzie  $\mathbf{H}_{m,c}$  oznacza globalną macierz kanału połączenia w górę . Wynikowa ocena MMSE jest podana przez

$$\tilde{\mathbf{H}}_{c,c} = \left( \sum_{m=1}^M \mathbf{H}_{m,c} + \frac{1}{\sqrt{\rho_{tr}}} \mathbf{N}\Theta^H \right) \hat{\mathbf{D}}_c,$$

gdzie  $\hat{\mathbf{D}}_c \in \mathbb{C}^{K \times K}$  jest macierzą diagonalną k-diagonalnymi elementami podanymi przez

$$\hat{d}_k = d_{k,c,c} \left( \sum_{m=1}^M d_{k,m,c} + \frac{1}{\rho_{tr}} \right)^{-1}.$$

Tutaj  $d_{k,m,c}$  reprezentuje składnik blaknący o dużej skali pomiędzy k-tym użytkownikiem umieszczonym w c-tej komórce i N antenami BS umiejscowionymi w m-tej komórce. Należy zauważyć, że powyższe ) wyraźnie obejmuje kanały komunikacyjne z innych komórek  $m \neq c$ . Aby zilustrować spostrzeżenia dotyczące zanieczyszczenia pilota w łączu w górę, stosujemy najprostszą strategię wykrywania: MF. Ten liniowy detektor może osiągnąć asymptotyczne wartości sumaryczne w łączu zwrotnym LSAS w idealnych warunkach, jak pokazano w sekcji 4.1.1, a jego prostota przyczynia się do zrozumienia wpływu zanieczyszczenia pilota w scenariuszach wielokomórkowych. Rozważ system wielokomórkowy składający się z komórek M. Sygnał odebrany na c-ty BS w takim systemie jest podany przez

$$y_c = \sqrt{\rho} \mathbf{H}_{c,c} x_c + \sqrt{\rho} \sum_{m \neq c} \mathbf{H}_{m,c} x_m + \mathbf{n}_c.$$

Należy zauważyć, że wpływ interferencji międzykomórkowej jest wyraźnie pokazany ,ponieważ każdy BS odbiera informacje z terminali użytkownika wszystkich komórek M. Gdy MF stosuje się w odbiorze, wynikowy wektor decyzyjny na BS c-tej komórki można wyrazić jako

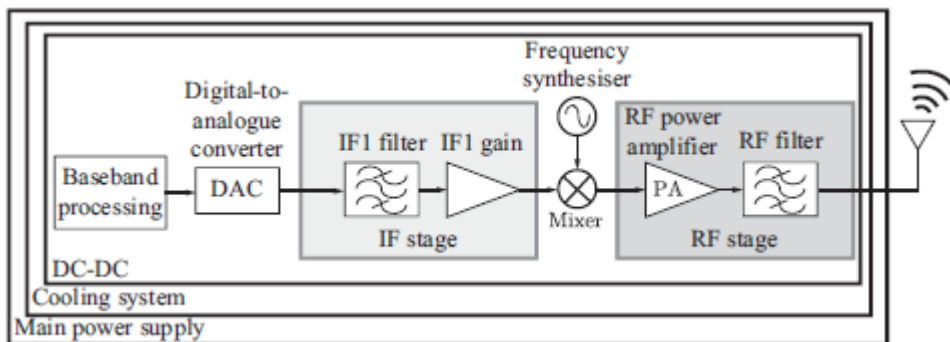
$$\begin{aligned} \mathbf{r}_c &= \tilde{\mathbf{H}}_{c,c}^H y_c = \tilde{\mathbf{H}}_{c,c}^H \left[ \sum_{m_2=1}^M \sqrt{\rho} \mathbf{H}_{m_2,c} x_{m_2} + \mathbf{n}_c \right] \\ &= \hat{\mathbf{D}}_c \left( \sum_{m_1=1}^M \mathbf{H}_{m_1,c} + \frac{1}{\sqrt{\rho_{\text{tr}}}} \mathbf{N}_c \right)^H \left( \sum_{m_2=1}^M \sqrt{\rho} \mathbf{H}_{m_2,c} x_{m_2} + \mathbf{n}_c \right) \\ &= \hat{\mathbf{D}}_c \left[ \sqrt{\rho} \sum_{m_1=m_2} \mathbf{H}_{m_1,c}^H \mathbf{H}_{m_2,c} x_{m_2} + \sum_{m_1=1}^M \mathbf{H}_{m_1,c}^H \mathbf{n}_c + \frac{1}{\sqrt{\rho_{\text{tr}}}} \mathbf{N}_c^H \mathbf{n}_c \right. \\ &\quad \left. + \frac{1}{\sqrt{\rho_{\text{tr}}}} \mathbf{N}_c^H \sum_{m_2=1}^M \sqrt{\rho} \mathbf{H}_{m_2,c} x_{m_2} + \sqrt{\rho} \sum_{m_1 \neq m_2} \mathbf{H}_{m_1,c}^H \mathbf{H}_{m_2,c} x_{m_2} \right], \end{aligned}$$

gdzie  $m_1, m_2 \in \{1, \dots, M\}$  wlicza indeksy komórek, a macrycę MF konstruuje się przy użyciu oszacowania kanału uzyskanego po oszacowaniu MMSE. Tak jak N rośnie duże, iloczyny w postaci  $\mathbf{H}_{m_1,c}^H \mathbf{n}_c$  można uznać za nieistotne ponieważ szum jest nieskorelowany z kanałem komunikacyjnym. Podobny argument dotyczy produktów  $\mathbf{N}_c^H \mathbf{H}_{m_2,c} x_{m_2}$ : Co więcej, iloczyny w postaci  $\mathbf{H}_{m_1,c}^H \mathbf{H}_{m_2,c} x_{m_2}, \forall m_1 \neq m_2$  są znacznie mniejsze niż iloczyny skorelowane ( $m_1 = m_2$ ) dla dużych N, które pozostają istotne. Efektem zanieczyszczenia pilotażowego jest spowodowany przez warunki interferencji w postaci  $\mathbf{H}_{m_1,c}^H \mathbf{H}_{m_2,c}, m_1 = m_2 \neq c$ , które nie znikają dla dużych N i są spowodowane przez stacje ruchome, które dzielą tą samą sekwencję pilota jak pokazano na rysunku 4.6 (a). Prowadzą one do częściowego ukształtowania wiązki sygnałów w kierunku stacji ruchomych znajdujących się w innych komórkach, jak pokazano na rysunku 4.6 (b). Badanie efektu skażenia pilotażowego przyciągnęło znaczną uwagę społeczności naukowej, ponieważ po raz pierwszy zidentyfikowano ją jako główne zjawisko ograniczania wydajności w LSAS. Istotnie, w ramach LSAS pojawiły się również różne strategie łagodzenia skażenia pilotażowego, wykorzystujące różne aspekty do rozróżniania zakłócających sygnałów międzykomórkowych z wewnątrzkomórkowego układu nerwowego. Na przykład, różnice między przestrzennymi macierzami kowariancji użytkowników wewnątrzkomórkowych i pozostałymi interferującymi stacjami ruchomymi dzielącymi identyczne sekwencje pilota mogą być wykorzystane do zmniejszenia skażenia pilota. Co więcej, efekt skażenia pilota można także zmniejszyć przez identyfikację różnic strat ścieżkowych pomiędzy sygnałami zakłócającymi z innych komórek i sygnałami wewnątrzkomórkowymi będącymi przedmiotem zainteresowania, przez niesynchronizowanie faz

pilota i fazy w połączeniach sąsiednich komórek lub przez projektowanie zaawansowanych strategii obejmujących oba łącza w dół. i szkolenie uplink. Chociaż szczegółowy opis powyższej kategorii technik wykracza poza zakres tego rozdziału, czytelnik jest kierowany do wyżej wymienionych prac do dalszej dyskusji. Powyższe prace pokazują, że istnieje szereg strategii mających na celu złagodzenie szkodliwego efektu zanieczyszczenia pilota. Istnieją jednak inne aspekty, które ograniczają najwyższą wydajność LSAS, co staramy się opisać poniżej.

#### 4.1.2.3 Ograniczenia sprzętowe w LSAS

Oczywiście, duża liczba anten wymaganych w LSAS pociąga za sobą znaczny wzrost wymaganych łańcuchów RF i ogólną złożoność sprzętową tych systemów w porównaniu z najnowszymi standardami. Obciążenie to promuje stosowanie tańszych komponentów sprzętowych, które mogą powodować zniekształcenia sygnału i niedoskonałości, zmniejszając w ten sposób zyski oferowane przez praktyczny LSAS. Wpływ zastosowania nieidealnego sprzętu na LSAS został włączony do ich analizy w literaturze. Niedoskonałości są zazwyczaj modelowane jako zakłócenia addytywne w modelu sygnału pasma podstawowego. Te zniekształcenia są wprowadzane przez włączenie dodatkowych źródeł szumów Gaussa w porównaniu do konwencjonalnego modelu systemu MIMO: jeden w nadajniku i dodatkowy w każdym odbiorniku. Źródła hałasu mają na celu modelowanie połączonego efektu szkodliwych efektów, takich jak brak równowagi I / Q nieidealnych mikserów, nieliniowości wzmacniaczy lub niedoskonałe filtry analogowe. Uproszczony schemat blokowy transmitującego łańcucha RF z typowym sprzętem składniki pokazano na rysunku 4.7.



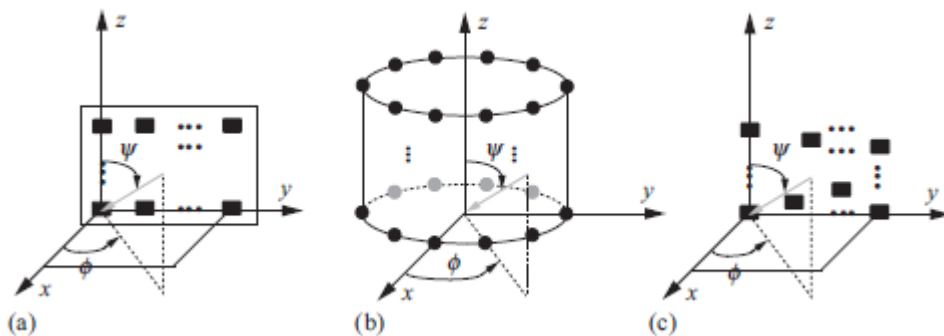
Model, który zawiera powyższy nieidealny sprzęt dla transmisji danych w górę w systemie dla pojedynczego użytkownika to

$$\mathbf{y} = \mathbf{h}(s + \eta_i^{\text{UT}}) + \eta_r^{\text{BS}} + \mathbf{n}',$$

gdzie zarówno uszkodzenia sprzętu na terminalu użytkownika,  $\eta^{\text{UT}}$ , jak i BS,  $\eta^{\text{BS}}$ , składają się z zmiennych losowych Gaussa, który jest motywowany przez addytywny efekt upośledzeń i centralne twierdzenie graniczne. W powyższym wyrażeniu  $s \in \mathbb{C}$  oznacza symbol transmitowany przez użytkownika, a  $\mathbf{n}'$  to wektor AWGN z niezależnymi wpisami spełniającymi  $n'_i \sim \mathcal{CN}(0, \sigma^2)$ . Co ciekawe, bardziej kompletny model pozwala na stwierdzenie, że ostateczna pojemność LSAS jest bardziej prawdopodobnie ograniczona przez awarie sprzętu na terminalach użytkownika, a nie na urządzeniach BS, który zamiast tego znika dla dużej liczby anten. To motywuje zatrudnienie tańszych komponentów w LSAS, co jest kluczowym aspektem z powodu dużej liczby komponentów RF, które należy wdrożyć. Więcej informacji na temat charakterystyki i wpływu awarii sprzętowych w LSAS, kierujemy zainteresowanego czytelnika do odpowiednich prac.

#### 4.1.2.4 Topologia macierzy anten i ograniczona przestrzeń fizyczna

Zwykle geometria macierzy antenowej ma znaczący wpływ na fizyczne wzmocnienie wiązki macierzy i ogólną wydajność macierzy. Zarówno wpływ geometrii macierzy, jak i ograniczenia posiadania ograniczonej przestrzeni fizycznej były do tego czasu ignorowane. Są to ważne aspekty, które należy uwzględnić przy projektowaniu realistycznego LSAS, ponieważ fizyczna przestrzeń, w której anteny muszą być przydzielone, mogą być rzeczywiście ograniczone. Rozważmy na przykład tablicę z  $N = 160$  elementów i centralną częstotliwością transmisji  $f = 2,6$  GHz, standardową częstotliwością dla obecnie wdrażanych systemów LTE. To skutkuje długością fali  $\lambda = c/f \approx 11,5$  cm, gdzie  $c = 3 \times 10^8$  m/s odnosi się do prędkości światła. Elementy w tablicach antenowych są zwykle umieszczane w odległości  $d \geq 0,5\lambda$ , która jest powszechnie określana jako odległość krytyczna]. Najwyraźniej umieszczenie anten wzdłuż prostej, jak w tradycyjnych jednoliniowych liniowych układach (ULA) wydaje się niepraktyczne w powyższym przykładzie, ponieważ rozmiar wynikowej macierzy wynosiłby  $d_t = N\lambda/2 \approx 9,2$  m, co może być trudne do dostosowania w obecnie stosowanych konstrukcjach anteny. Co więcej, ULA nie pozwalają na kierowanie wiązką w domenie elewacji, co może być konieczne do rozróżnienia pozycji użytkownika i do dokładnego ogniskowania przesyłanej energii. Zastosowanie innych topologii anteny, takich jak planarne, cylindryczne, sferyczne lub nawet nieregularne układy wydaje się bardziej odpowiednim wyborem dla LSAS, jak pokazano na rysunku 4.8.



Rzeczywiście, większość prototypów anten obecnie realizowanych dla LSAS przyjmuje jedną z tych geometrii z antenami podwójnie spolaryzowanymi. Zastosowanie podwójnie spolaryzowanych anten wydaje się niezbędne dla LSAS, ponieważ umożliwiają one podwojenie liczby portów antenowych bez zwiększania wzajemnego sprzężenia i fizycznej przestrzeni wymaganej dla macierzy antenowych. Na przykład wymiary powyższej macierzy antenowej z podwójnie spolaryzowanymi elementami antenowymi  $N = 160$  rozmieszczonymi w układzie prostokątnym z antenami  $N_{hor} = 20$  i antenami  $N_{ver} = 8$  zajmowałyby obszar w przybliżeniu równy  $115 \times 46$  cm<sup>2</sup>. Zauważ, że umieszczenie większej liczby anten w poziomie pozwala uzyskać lepszą rozdzielczość w domenie azymutalnej niż w wysokościowej. Niemniej jednak, w przypadku wielu scenariuszy wdrożenia LSAS, olbrzymia liczba przewidywanych anten może wymagać dopasowania w ograniczonej fizycznej przestrzeni dzisiejszych stacji BS. Dlatego mogą istnieć scenariusze, w których anteny muszą być umieszczone w odległości  $d \leq 0,5\lambda$ , aby spełnić określone fizyczne ograniczenia. W związku z tym, że liczba anten zwiększa się w stałej przestrzeni fizycznej, wymaga to zmniejszającego się rozdziału anteny. Prowadzi to dalej do wzrostu przestrzennej korelacji anten jako źródeł sygnału i do wzajemnego sprzężenia anten jako komponentów elektrycznych. W tym kontekście interesujące jest zbadanie kompromisu, który pojawia się w tych fizycznie ograniczonych scenariuszach. Oznacza to, że zwiększając  $N$ , liczba anten rośnie wraz ze wzrostem źródeł informacji, podczas gdy ich przestrzenne rozdzielanie zmniejsza pogarszanie efektów wzajemnego sprzężenia i korelacji przestrzennej

#### 4.1.2.5 Złożoność przetwarzania sygnału w LSAS

Pomimo faworyzowania technik liniowego przetwarzania sygnałów do transmisji i odbioru danych, zwiększone wymiary systemu przewidziane przez LSAS drastycznie rosną złożoność operacji przetwarzania sygnałów. Co więcej, pobór mocy cyfrowych procesorów sygnałowych (DSP) i interfejsów danych może być znaczący w LSAS. W tej sekcji omawiamy złożoność przetwarzania liniowego w LSAS, podkreślając najbardziej złożone operacje związane z procesem. Początkowy etap komunikacji LSAS polega na uzyskaniu CSI dla kolejnych procesów transmisji i wykrywania danych, jak pokazano na Rysunku 4.5. Jednak zastosowanie strategii szacowania kanału, takich jak MMSE konwencjonalnie rozważane w literaturze LSAS, może nałożyć znaczące obciążenie obliczeniowe ze względu na konieczność odwrócenia macryc o dużych rozmiarach. W tym kontekście zastosowanie przybliżeń wielomianowych do odwrotności macierzy stanowi odpowiedni kompromis między złożonością a jakością oceny w zależności od wybranego porządku wielomianowego. Po zakończeniu etapu szkoleniowego BS przechodzi do fazy transmisji danych i odbioru. W tym przypadku liczba szybkich transformacji Fouriera (FFT) i odwrotnych FFT, które muszą być obliczane na zasadzie symbol po symbolu, jest również znacznie zwiększona w odniesieniu do obecnych standardów. Stanowi to jedno z głównych obciążeń obliczeniowych, ponieważ muszą one być obliczane z szybkością symboli obejmującą:

$$C_{\text{FFT}} = \mathcal{O}(N\tau N_{\text{sub}} \log_2(N_{\text{sub}})) \text{flops},$$

gdzie  $\mathcal{O}$  oznacza operator Big-O,  $\tau$  oznacza liczbę zawartych symboli w czasie koherencji kanału  $T$  i  $N_{\text{sub}}$  oznaczają liczbę podnośnych w systemie przepustowości  $B$ . W odniesieniu do złożoności liniowych prekoderów i detektorów, główna wada strategii ZF i MMSE można znaleźć w inwersjach macierzy wymaganych do zbudowania macierzy wstępnego kodowania. Zwróć uwagę, że inwersje kanału muszą być wykonywane tylko na początku czasu koherencji kanału, a nie na podstawie symbol-symbol, ponieważ wynik pozostaje ważny, podczas gdy kanał pozostaje niezmienny. W szczególności, zakładając  $N \geq K$ , złożoność wstępnych obliczeń wymaganych do odbioru danych w określonym czasie koherencji jest określona przez

$$C_{\text{inv}} = \mathcal{O}(NK^2\tau_u N_{\text{blocks}}) \text{flops},$$

gdzie  $\tau_u < \tau$  oznacza liczbę symboli przeznaczonych do odbioru symbolu (łącze nadrzędne), a  $N_{\text{blocks}}$  oznacza całkowitą liczbę bloków częstotliwości z podobnym kanałem w całkowitej szerokości pasma  $B$ . Złożoność wielomianowa powyższych operacji z liczbą użytkowników  $K$  sprzyja wprowadzeniu metod przetwarzania sygnałów w celu zmniejszenia złożoności tego etapu. Obejmują one strategie, w których odwrotność macierzy jest przybliżana za pomocą rozszerzenia szeregów Neumanna i rozkładów wielomianowych. Co ważniejsze jednak, mnożenie macierzy-wektora wymiaru  $N$  muszą być wykonywane z prędkością symboli między odebrany (informacyjny) wektorem symboli a macrycą detekcji (wstępnego kodowania) do wytwarzania wynikowych decyzji (transmitowanych) symboli. W znacznym stopniu przyczyniają się one do ogólnej złożoności obliczeniowej LSAS [26, 65]. Ogólnie można stwierdzić, że LSAS nakłada rygorystyczne wymagania przetwarzania sygnałów w czasie rzeczywistym. Wykonanie powyższych operacji wydaje się wykonalne z wykorzystaniem obecnie dostępnych technologii [26]. Ponadto przewiduje się, że szybki rozwój technologii cyfrowego przetwarzania sygnałów jednocześnie zmniejszy zarówno ścisłe wymagania, jak i ekonomiczne koszty związane z możliwościami obliczeniowymi LSAS.

## 4.2 Modele kanałów LSAS i aspekty propagacji



Większość wyników teoretycznych opisanych w rozdziale 4.1 pochodzi z analizowania wydajności LSAS w ramach zestawu ograniczonego nieskorelowanego kanału modelem. Jak wyjaśniono powyżej, w praktyce istnieje wiele scenariuszy gdzie zazwyczaj strumienie przestrzenne są skorelowane. W związku z tym szereg teoretycznych i półempirycznych modeli kanałów jest dostępnych w literaturze, dla badania takich efektów. W tej sekcji omówiono modele powszechnie stosowane w celu dokładnego scharakteryzowania zachowania LSAS, aby ułatwić zainteresowanym czytelnikom modelowanie tych systemów. Ponadto opisano również wnioski wyciągnięte z kampanii pomiarowych i konkretne modele kanałów opracowane dla LSAS.

#### 4.2.1 Teoretyczne modele kanałów LSAS

Oparty na korelacji stochastyczny model kanału został szeroko zastosowany do teoretycznego scharakteryzowania zachowania LSAS ze względu na jego prostotę. W pierwszej kolejności koncentrujemy się na modelach kanałów opartych na korelacji i podkreślamy ich główne cechy i ograniczenia w modelowaniu kanałów komunikacyjnych LSAS. Aby zachować spójność z poprzednimi sekcjami, koncentrujemy się na modelu systemu uplink, w szczególności na scenariuszu pojedynczej komórki dla zapisu nutowego. Następnie koncentrujemy się na jednostronnym modelu korelacji, z korelacją narysowaną na BS, podczas gdy modele można rozszerzyć, aby rozwiązać problem korelacji anteny w MS. Rozważmy LSC z jedną komórką składającą się z BS z N antenami nadawczymi i K pojedynczych anten użytkowników. Model oparty na korelacji wyraża kanał łączący w górę między k-tym użytkownikiem a BS jako

$$\mathbf{h}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{g}_k,$$

gdzie  $\mathbf{R}_k \in \mathbb{C}^{N \times N}$ , jest hermetyczną deterministyczną macierzą, która charakteryzuje kanał macierzy korelacji k-tego użytkownika i, podobnie  $\mathbf{g}_k \in \mathbb{C}^{N \times 1} \sim \mathcal{CN}(0, \mathbf{I}_N)$  jest przypadkowym wektorem, którego zespolone współczynniki  $g_{k,n}$  reprezentują wzmocnienie kanału między n-tą anteną na BS a pojedynczą anteną k-tego użytkownika. Przez cały czas sekcja,  $\mathbf{g} \sim \mathcal{CN}(\mathbf{a}, \mathbf{B})$  oznacza, że losowy wektor  $\mathbf{g}$  spełnia złożony rozkład Gaussa ze średnią  $\mathbf{a}$  i macierzą kowariancji  $\mathbf{B}$ . Dlatego  $\mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{R}_k)$ . Powyższy model uważa, że nie ma korelacji przestrzennej wśród użytkowników, co jest rozsądnym założeniem, ponieważ zazwyczaj są one rozdzielone przez odległości większe niż  $\lambda$ . Ponadto należy zauważyć, że macierze kowariancji  $\mathbf{R}_k, \forall k \in \{1 \dots K\}$  zazwyczaj zmieniają się w tempie wolniejszym niż czas koherencji kanału. Powyższy model kanału umożliwi scharakteryzowanie różnorodnych efektów kanału, które mają wpływ na wydajność LSAS. Na przykład prosty model zanikania Rayleigha z utratą ścieżki wprowadzoną w Rozdziale 4.1 można wyrazić jako uszczegółowienie przez zezwolenie  $\mathbf{R}_k = \mathbf{D}$ . Zwróć uwagę, że ogólnie  $\text{rank}(\mathbf{R}_k) \neq N$ , gdzie  $\text{rank}(\cdot)$  odnosi się do rangi macierzy. Innymi słowy, model można wykorzystać do scharakteryzowania rzadkich kanałów komunikacyjnych. Równoważnie, reprezentacja kanału komunikacyjnego Karhunen-Loève (KL) może być wyrażona jako

$$\mathbf{h}_k = \mathbf{U}_k \sum_k \mathbf{g}_k,$$

gdzie  $\mathbf{U}_k \in \mathbb{C}^{N \times N}$  jest jednostkową macierzą z kolumnami, których wektory własne są macierzą kowariancji kanału  $\mathbf{R}_k = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{U}_k^H$  a  $\mathbf{\Sigma}_k \in \mathbb{C}^{N \times N}$  jest diagonalną macierzą zawierającą wartości własne  $\mathbf{R}_k$ . Równoważna reprezentacja wprowadzona była wielokrotnie wykorzystywana do analizy zachowania kanałów komunikacyjnych LSAS, ponieważ umożliwia wyraźne ukazanie niespójności kanału, która odnosi się do liczby zerowych wartości własnych  $\Sigma_k^{1/2}$ , i ułatwia opracowanie projektów macierzy antenowych o obniżonej częstotliwości radiowej zgodnie z opisem w Rozdziale 4.3. Fizycznie, obniżona pozycja w macierzy kowariancji  $\mathbf{R}_k$  może być wytwarzana przez różne efekty, takie jak znaczna

korelacja między antenami lub niewystarczające rozpraszanie, które są szczególnie ważne dla charakterystyki LSAS. Efekty te można scharakteryzować, stosując definicje macierzy kowariancji kanałów) o różnych stopniach złożoności, które są opisane poniżej. Spośród wielu istniejących modeli korelacji, tutaj skupiamy się na najbardziej odpowiednich modelach kanałów dla LSAS do tej pory.

#### 4.2.1.1 Wykładniczy model korelacji

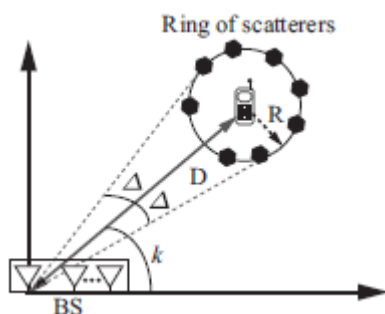
Wykładniczy model korelacji ma na celu scharakteryzowanie wpływu korelacji inter-antenowej, która indukuje podobieństwo w kanałach komunikacyjnych sąsiedniej anteny. Macierz kowariancji kanału użytkownika kilku w wykładniczym modelu korelacji jest podana przez

$$\mathbf{R}_k = \begin{pmatrix} 1 & t_k & t_k^2 & \dots & t_k^{N-1} \\ (t_k)^H & 1 & t_k & \dots & t_k^{N-2} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ (t_k^H)^{N-1} & \dots & (t_k^H)^2 & (t_k)^H & 1 \end{pmatrix},$$

gdzie  $t_k = |t_k|e^{j\phi_k} \in \mathbb{C}$  jest współczynnik korelacji dla  $k$  i przyjęto, że masowe BS zastępują liniowy układ (ULA) dla uproszczenia, choć ten model może być natychmiast uogólnić na inne struktury. Faza  $\phi_k$  z  $t_k$  jest zwykle ustawiana jako zero lub identyczna. Jednak nowsze prace definiują fazy  $\phi_k$  jako niezależne zmienne losowe, aby scharakteryzować różnice między kanałami różnych użytkowników. Wykładniczy model korelacji jest szczególnie istotny w przypadku LSAS o ograniczonej wielkości, ponieważ istnieje bezpośrednia zależność między odległością anteny a pozycjami macierzy korelacji kanałów. Ponieważ jednak ten prosty model nie jest w stanie dokładnie uchwycić innych ważnych efektów komunikacyjnych, takich jak ograniczone rozproszenie kanałów, ich zatrudnienie jest tradycyjnie ograniczone do poparcia dowodów pojęciowych.

#### 4.2.1.2 Model z jednym pierścieniem

Model jednopierścieniowy został konwencjonalnie zastosowany do charakteryzowania systemów komunikacji mobilnej, w których BS nie doświadcza znacznego lokalnego rozproszenia, podczas gdy terminale użytkownika są otoczone rozpraszaczami. Odpowiada to typowemu scenariuszowi, w którym stacje mobilne znajdują się na poziomie gruntu, podczas gdy stacja BS znajduje się na podwyższonej strukturze, jak ma to miejsce obecnie w przypadku typowych komórkowych stacji BS. Z tego powodu ten model został przyjęty w różnych pracach LSAS, takich jak. Jeden pierścień modelu lokalnego rozpraszająca może być fizycznie interpretować w sposób pokazany na rysunku 4,9 dla konkretnego przypadku Ula chociaż model ten może być z łatwością rozciągnięte na obejmują bardziej skomplikowane topologii anteny.



Niech użytkownik mobilny znajduje się przy kącie azymutu  $\kappa$  w odległości  $D$  od BS będzie otoczony pierścieniem rozpraszaczy nad kołowym obszarze o promieniu  $R$ . Następnie współczynnik korelacji między antenami  $i$  i  $j$  z ULA w BS jest podana dla  $k$ -tego użytkownika przez

$$\{\mathbf{R}_k\}_{i,j} = \frac{1}{2\Delta} \int_{-\Delta}^{\Delta} e^{j\frac{2\pi}{\lambda}d(i-j)\cos(\theta+\kappa)} d\theta,$$

gdzie  $\{X\}_{i,j}$  oznacza  $i, j$ -te elementy macierzy  $X$  i  $d$  jest separacją sąsiednich elementów antenowych w BS. Rozciągnięcie kątowe jest zdefiniowane jako  $\Delta \approx \arctan(R/D)$  a  $\kappa \in [0, 2\pi)$  jest kątem pomiędzy BS i kodem MS, jak pokazano na rysunku 4.9. Intuicyjnie, zredukowany  $D$  indukuje większą korelację kanału, ponieważ zmniejsza się zasięg kątów nadejścia uderzających sygnałów. Zauważ, że te założenia są tylko przybliżeniami dla LSAS, gdzie sferyczny kształt rzeczywistych fal uderzających może mieć wpływ na wydajność.

#### 4.2.1.3 Fizyczny lub skończony model kanału rozpraszającego

Ten model kanału dzieli domenę kątową na skończoną liczbę kierunków (wymiarów)  $P < N$ , które mogą odpowiadać liczbie aktywnych rozpraszaczy. Każdy z kierunków  $P$  od użytkownika  $k$  do BS charakteryzuje się parą kątów azymutu i wysokości przybycia. Angls of Arrival (AoA),  $\{\phi_p, \psi_p\}$ ,  $p \in \{1 \dots P\}$ . Tablica wektora sterującego,  $\mathbf{a}(\phi_p) \in \mathbb{C}^{N \times 1}$ , można wyrazić jako

$$\mathbf{a}(\phi_p) = \frac{1}{\sqrt{P}} \left[ e^{-j f_n(\phi_p, \psi_p)}, \dots, e^{-j f_N(\phi_p, \psi_p)} \right]^T.$$

Tutaj  $f_n(\phi_p, \psi_p)$  jest dowolną funkcją, która zależy od geometrii macierzy antenowej w BS. Na przykład, liniowa tablica o jednolitej odległości pomiędzy promieniającymi elementami następuje po

$$f_n(\phi_p) = \frac{2\pi(n-1)D}{\lambda} \cos(\phi_p)$$

a wektor sterowania ktor jest podany przez

$$\mathbf{a}_{ULA}(\phi_p) = \frac{1}{\sqrt{P}} \left[ 1, e^{-j2\pi\frac{d}{\lambda} \cos(\phi_p)}, \dots, e^{-j2\pi\frac{(N-1)d}{\lambda} \cos(\phi_p)} \right]^T$$

W powyższym przykładzie  $d$  jest odstępem między elementami anteny w nadajniku. Włączając fizyczny model rozproszenia do standardowego modelu, wynikowy wektor kanału od  $k$ -tej stacji ruchomej do BS,  $\mathbf{h}_k \in \mathbb{C}^{N \times 1}$ , jest podany przez

$$\mathbf{h}_k = \sum_{p=1}^P \mathbf{a}(\phi_p, \psi_p) a_{k,p},$$

gdzie  $a_{k,p}$  jest wzmocnieniem kanału między kodem a BS w wymiarze  $p$  i zazwyczaj zawiera szybkie efekty zanikania, cieniowania i utraty ścieżki. Dla uproszczenia przyjmuje się zwykle, że liczba ścieżek transmisji  $P$  jest identyczna dla wszystkich użytkowników  $k \in \{1 \dots K\}$ . Inne powszechne założenie polega na tym, że szybkie współczynniki zanikania zależą od kierunku fizycznego, podczas gdy czynniki blaknięcia na dużą skalę, takie jak utrata ścieżki, są niezależne od kierunku  $p$ . Fizyczny model kanału może być również wyrażony w postaci macierzy. W szczególności, pozwól, aby wektor złożony  $\mathbf{a}_k \in \mathbb{C}^{P \times 1}$  był zdefiniowany jako  $\mathbf{a}_k = [a_{k,1}, \dots, a_{k,P}]^T$ . Następnie kanał użytkownika  $k$ -tego użytkownika w fizycznym modelu kanału może być wyrażony jako

$$\mathbf{h}_k = \mathbf{A}_k \alpha_k,$$

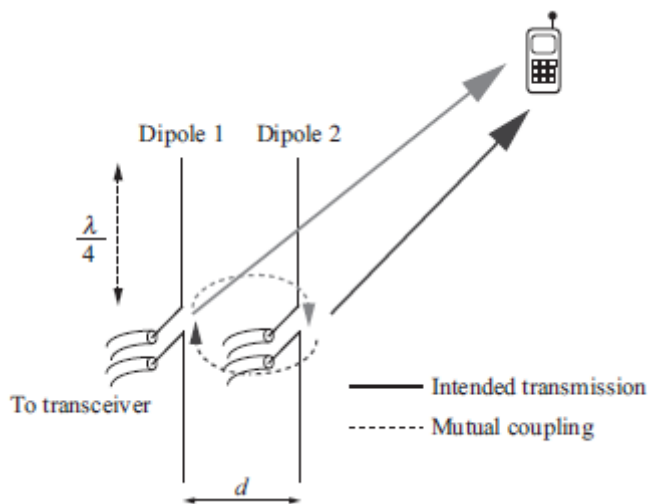
gdzie  $\mathbf{A} \in \mathbb{C}^{N \times P}$  jest macierzą pełnowartościową obejmującą wektory sterujące  $\mathbf{A} = [\mathbf{a}(\phi_1), \dots, \mathbf{a}(\phi_p)]$ . Należy zauważyć, że model ten mieści się w zakresie modeli stochastycznych opartych na korelacji z

$$\bar{\mathbf{R}}_k = [\sqrt{d_k} \mathbf{A}_k, \mathbf{0}_{N \times N-P}]$$

gdzie  $d_k$  odnosi się do blaknącego komponentu o dużej skali,  $k$ -tego użytkownika. Wiąże się to z tym zakresem  $(R_k) \leq P$ , która intuicyjnie może być interpretowana w następujący sposób: liczba niezależnych ścieżek transmisji między BS i  $k$ -tym użytkownikiem, która jest bezpośrednio związana z liczbą klastrów rozpraszających, określa maksymalną liczbę niezależnych strumieni danych wspierane przez LSAS. Ponadto, jak wyraźnie pokazano, modele zarówno fizyczne, jak i jedno-pierścieniowe umożliwiają scharakteryzowanie kątów przybycia do lub odejścia od układu antenowego. Ten aspekt w połączeniu z możliwością charakteryzowania ograniczonych środowisk rozpraszania promowały zastosowanie fizycznego modelu kanału do charakteryzowania zachowanie LSAS w różnych pracach

#### 4.2.2 Wzajemna charakterystyka sprzężenia

Duży wzrost liczby anten przewidzianych w LSAS implikuje, że przestrzeń przeznaczona na przydział anten musi również rosnąć, aby zapewnić nieskorelowaną pracę. Jednak fizyczne struktury anteny mogą podlegać ograniczonej przestrzeni, która może wpływać na osiągalne rozdzielanie anteny. Co więcej, przyszłe implementacje LSAS prawdopodobnie będą zawierać setki małych anten dipolowych i innych tanich komponentów. Powyższe może sprzyjać pojawianiu się wzajemnego sprzężenia pomiędzy elementami promieniującymi. Charakteryzuje się efektem wzajemnego sprzężenia energia pochłaniana przez anteny fizycznie blisko referencyjnej anteny operacyjnej, jak pokazano na rysunku 4.10.



Na etapie nadawania oznacza to, że część energii sygnału przeznaczonej dla konkretnej anteny wzbudza sąsiednie anteny w nadajniku, które w konsekwencji emitują część tej niezamierzonej energii. Dokładna charakterystyka efektów wzajemnego sprzężenia może być zapewniona przez analiza parametrów  $S$  systemu komunikacyjnego. To jest podejście przyjęte w celu modelowania efektu wzajemnego sprzężenia, gdzie parametry  $S$  są uzyskiwane za pomocą pełno falowych symulacji elektromagnetycznych. Zamiast tego prostsze podejście oparte jest na wykorzystaniu macierzy impedancji w celu scharakteryzowania efektów wzajemnego sprzężenia. W szczególności model ten wprowadza dodatkowa macierz do modelu kanału w celu uwzględnienia efektu wzajemnego

sprzężenia. Główną wadą tej alternatywy jest to, że wyniki tylko dotyczą anten o minimalnej rozpraszalności, takich jak długość połówki fali dipoli. Jednak jego bezpośredni związek z modelami obecnie zatrudnionymi w LSAS i możliwość użycia wyrażeń analitycznych do uzyskania impedancji matryc określonych typów anten ułatwiają korzystanie z tego modelu. W modelu macierzy sprzężeń wzajemnych kanał między k-tą pojedynczą anteną użytkownika i BS,  $\mathbf{h}_k \in \mathbb{C}^{N \times 1}$ , jest typowo wyrażony

$$\mathbf{h}_k = \mathbf{Z} \bar{\mathbf{R}}_k \mathbf{g}_k,$$

gdzie  $\bar{\mathbf{R}}_k$  i  $\mathbf{g}_k$  odpowiadają wprowadzonym w  $\mathbf{h}_k = \mathbf{R}_k^{\frac{1}{2}} \mathbf{g}_k$  i  $\mathbf{Z} \in \mathbb{C}^{N \times N}$  jest macierzą wprowadzoną do modelowania efektu wzajemnego sprzężenia. Wzorec wzajemnego sprzężenia jest rutynowo wyrażany jako

$$\mathbf{Z} = (\mathbf{Z}_A + \mathbf{Z}_L)(\Xi + \mathbf{Z}_L \mathbf{I}_N)^{-1}$$

gdzie  $Z_A$  jest impedancją anteny,  $Z_L$  jest impedancją obciążenia a  $\Xi$  macierzą daną przez

$$\Xi = \begin{pmatrix} Z_A & Z_m^{1,2} & Z_m^{1,3} & Z_m^{1,4} & \dots & Z_m^{1,N} \\ Z_m^{2,1} & Z_A & Z_m^{2,3} & Z_m^{2,4} & \dots & Z_m^{2,N} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ Z_m^{N,1} & Z_m^{N,2} & \dots & Z_m^{N,N-2} & Z_m^{N,N-1} & Z_A \end{pmatrix},$$

gdzie  $Z_m^{ij}$ ,  $i, j \in \{1 \dots N\}$  jest wzajemną impedancją między elementami anteny i i j. Typowe wartości impedancji obciążenia spełniają  $Z_L = 50 \Omega$ , w modelu wzajemnego sprzężenia, natomiast wartość  $Z_A = 73 + j42,5 \Omega$  jest często wykorzystywana do impedancji anteny dipolowej. Wzajemne impedancje można uzyskać, stosując wyrażenia w zamkniętej formie dostępne dla anten dipolowych o różnych położeniach względnych.

#### 4.2.3 Kampanie pomiarowe i modele kanałów specyficzne dla LSAS

Szczególne cechy LSAS mają znaczący wpływ na powiązane kanały komunikacyjne w porównaniu z tradycyjnymi konfiguracjami MIMO na małą skalę. Na przykład zamiast tradycyjnego planarnego czoła fali często zakłada się w małej skali systemy, sferyczne fale mogą być brane pod uwagę w tablicach LSAS motywowanych przez duży rozmiar fizyczny masywnych struktur anteny. Dodatkowo, klastry rozproszone można uznać za niestacjonarne zarówno w domenach fizycznych, jak i czasowych z podobnych powodów. Aspekty te motywują do rozwoju nowych modeli kanałów, które uwzględniają specyfikę LSAS. Powyższe cechy kanałów komunikacyjnych LSAS zaobserwowano w różnych przeprowadzonych kampaniach pomiarowych. W tym kontekście wyniki pomiarów wydają się potwierdzać ulepszenia separacji przestrzennej użytkownika, które można osiągnąć za pomocą LSAS. Innymi słowy, ortogonalność kanału wśród użytkowników, którzy mogliby również osiągnąć w rzeczywistych scenariuszach propagacji. Ogólnie rzecz biorąc, wyniki pomiarów wydają się wskazywać, że idealne osiągalne wartości teoretycznie obiecane w warunkach zanikania Rayleigha mogą być przybliżone w konkretnych warunkach pomiarowych. Jednak nie jest to ogólna cecha LSAS, ponieważ niektóre scenariusze, takie jak te utworzone przez blisko zlokalizowanych użytkowników doświadczających propagacji LOS, mogą spowodować znaczące straty w odniesieniu do idealnych osiągalnych stóp.

### 4.3 LSAS: projektowanie i optymalizacja systemu

#### 4.3.1 Zredukowane implementacje łańcuchów RF

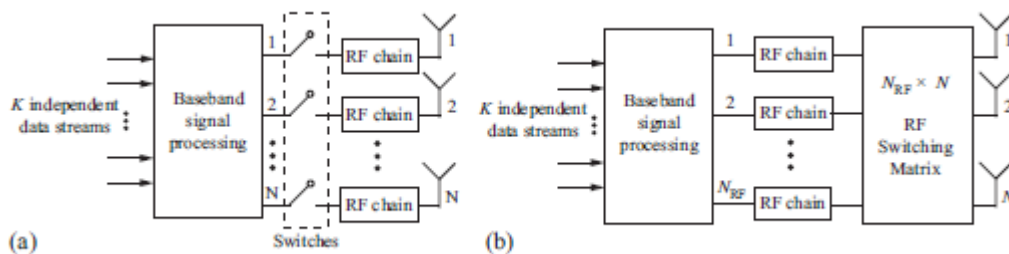
Pomimo tolerancji na tańsze komponenty RF, LSAS nadal wymaga dużej liczby komponentów sprzętowych, ponieważ ich teoretycznie obiecane korzyści zależą od wdrożenia jednego aktywnego łańcucha RF na antenę. Jednak każdy łańcuch RF ma znaczną liczbę elementów, które przyczyniają się do ogólnego zużycia energii. Na przykład, stopień transmisji obejmuje cyfrowe procesory sygnałowe, konwertery cyfrowo-analogowe, miksery, filtry, syntezytory częstotliwości i wzmacniacze mocy, które narzucają całkowitą moc zużywaną przez BS. Badanie i projektowanie konwencjonalnych transceiverów MIMO najczęściej zakłada, że liczba łańcuchów RF jest identyczna z liczbą anten BS, tj. Każda antena tablicy MIMO wykorzystuje swój dedykowany łańcuch RF. Chociaż to wysoce wydajne podejście zapewnia pełną elastyczność w przetwarzaniu sygnałów przesyłanych i odbieranych przez macierz anten, w szczególnym przypadku LSAS oznacza to, że znaczna liczba elementów sprzętowych musi być włączona do stacji BS. W tej części przedstawiamy szereg rozwiązań i praktycznych projektów omawianych w literaturze na temat uproszczenia sprzętu i ogólnego działania LSAS.

##### 4.3.1.1 Wybór anteny

Intuicyjne podejście do zmniejszenia złożoności sprzętu polega na zmniejszeniu liczby anten wykorzystywanych jednocześnie do nadawania i odbioru w BS. Takie schematy zwykle obejmują strategię wyboru anteny i polegają głównie na wyłączaniu anten, które nie wnoszą znaczącego wkładu w komunikację, na podstawie dostępnego CSI. Podobnie do odbieranego sygnału w dowolnym schemacie wyboru anteny z antenami  $L$  jednocześnie aktywnymi na BS można wyrazić jako

$$\bar{y}_c = \sum_{m=1}^M \sqrt{\rho} \mathbf{H}_{c,m}^{\mathcal{L}_m} \mathbf{x}_m + \mathbf{n}$$

gdzie  $\mathbf{H}_{c,m}^{\mathcal{L}_m} \in \mathbb{C}^{L \times K}$  oznacza wynikowy kanał po wybraniu  $L < N$  wierszy od  $\mathbf{H}_{c,m}$ , indeksowanych przez zbiory  $\mathcal{L}_m, m \in \{1, \dots, M\}$ . Oczywiście, zarówno liczba aktywnych anten  $L$  i specyficzne elementy anteny wybrane przez wskaźniki  $\mathcal{L}_m$  wpływają na najwyższą wydajność systemu, a jego staranna konstrukcja jest niezbędna w schematach wyboru anteny. W tym kontekście wdrożenie LSAS może sprzyjać dokonywaniu tego wyboru ze względu na dużą liczbę dostępnych anten. Dwie różne implementacje sprzętu do wyboru anteny pokazano na rys. 4.11 (a) i (b). Rysunek 4.11 (a) ilustruje elastyczny schemat wyboru anteny, w którym zaimplementowano  $N$  łańcuchów RF i podzbiór ich może zostać rozłączony. Zamiast tego, rysunek 4.11 (b) przedstawia implementację sprzętową, w której zaimplementowano zredukowaną liczbę  $N_{RF} < N$  łańcuchów RF i macierz przełączania RF określa aktywne anteny NRF.

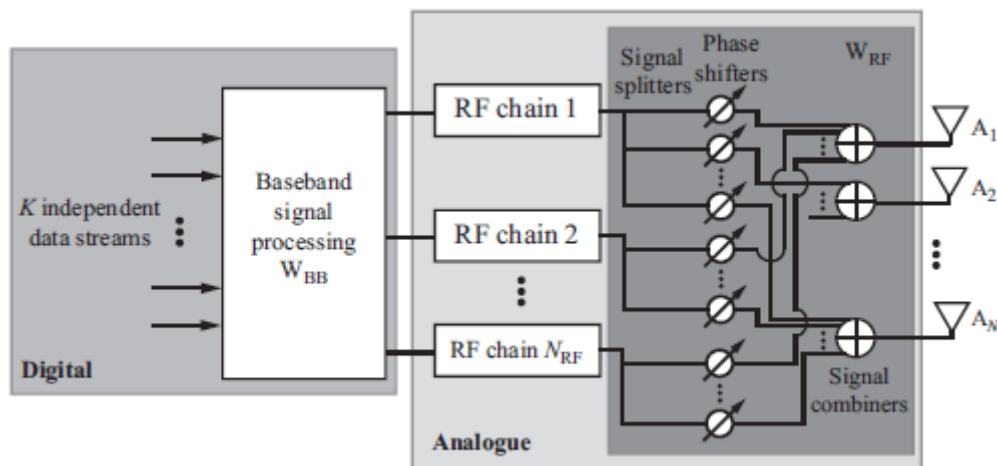


Należy pamiętać, że jedna architektura może być bardziej korzystna niż druga, w zależności od aktualnych wymagań systemowych. Pierwsze podejście ma tę wadę, że nadal wymaga rozmieszczenia dużej liczby łańcuchów RF, podczas gdy metoda przełączania RF może zniekształcać sygnały wyjściowe

i wprowadzać dodatkowe straty. Zastosowanie strategii wyboru anten do poprawy efektywności energetycznej LSAS zostało rozważone w kanałach teoretycznych i pomiarowych w, podczas gdy istnieje duża liczba podejść do wyboru anteny

#### 4.3.1.2 Hybrydowe analogowe i cyfrowe formowanie wiązki: redukcja liczby łańcuchów RF

Alternatywna strategia zmniejszania liczby łańcuchów RF bez wpływu na wydajność w LSAS, który wzbudził ostatnie zainteresowanie, wykorzystuje niespójność kanału W tym kontekście sparsność kanału może być interpretowana jako zestaw szczególnych właściwości, które powodują, że kanał komunikacji ma zmniejszoną liczbę stopni swobody (DoF). Na przykład, rozbieżność kanału może być promowana przez pojawienie się poważnej korelacji przestrzennej spowodowanej niewystarczającym odstępem między antenami w macierzy, lub ograniczoną liczbą fizycznych rozpraszaczy w kanale propagacji, jaka występuje w paśmie fal milimetrowych. W takich przypadkach można wykorzystać sparametryzowanie kanału w celu zmniejszenia liczby łańcuchów RF przez włączenie analogowych sieci formowania wiązki składających się z analogowych przesuwników fazowych i rozdzielaczy sygnału oraz kombinatorów sąsiadujących z przednimi końcami anteny. Podejścia te są powszechnie nazywane hybrydowymi analogowymi i cyfrowymi projektorami kierunkowymi, a ich analiza jest następująca. Chociaż w chwili tworzenia niniejszego rozdziału było to szczególnie żywe środowisko badawcze, tutaj omawiamy najważniejsze metodologie do tej pory. Architektura generycznego hybrydowego układu kształtowania wiązki pokazano na rysunku 4.12



Hybrydowy kształtowanie wiązki jest podzielony na dwa różne bloki: cyfrowy kształtownik wiązki podstawowej i analogowy kształtownik wiązki RF. Główną zaletą projektów hybrydowych kształtowania wiązki w porównaniu z tradycyjnymi schematami MIMO jest to, że kształtowanie wiązki pasma podstawowego może być zaprojektowane do działania na sygnałach o zredukowanej wymiarowości  $N_{RF} < N$ , co pozwala na zmniejszenie liczby łańcuchów RF  $N_{RF}$  w sieci architektura tablicowa, jak pokazano na rysunku. Wiąże się to z kosztem wykorzystania dodatkowych komponentów RF do podziału sygnałów wejściowych  $N_{RF} K$  na kształtowanie wiązki RF, z zastosowaniem arbitralnych przesunięć fazowych, które zależą od dostępnego CSI, i łączenia uzyskanych przesuniętych fazowo sygnałów dla wytworzenia wymaganych  $N$  wyjść. W scenariuszu jednokomórkowego łącza zwrotnego sygnał wyjściowy podczas stosowania formowania wiązki hybrydowej w odbiorniku można wyrazić jako

$$\mathbf{r} = \mathbf{W}\mathbf{y} = \mathbf{W}_{BB}\mathbf{W}_{RF}(\sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{n}).$$

gdzie  $W_{BB} \in \mathbb{C}^{N_{RX} \times N_{TX}}$  a  $W \in \mathbb{C}^{N \times N_{RF}}$  reprezentują pasmo podstawowe i analogowe macierze wstępnego kodowania, odpowiednio. Zwróć uwagę na zmniejszone wymiary  $W_{BB}$  w stosunku do  $W$ . Ponieważ analogowa macierz kształtująca wiązkę jest siecią przesuwającą fazę, pierwszy i drugi element  $W_{RF}$  przyjmuje postać

$$(W_{RF})_{i,l} = \frac{1}{\sqrt{N}} e^{j\varphi(i,l)},$$

, gdzie specjalne fazy przesuwu  $\varphi(i,l)$  są określane według dostępnego chwilowego lub długoterminowego (statystycznego) wskaźnika CSI. Projekt obu  $W_{RF}$  i  $W_{BB}$  dla LSAS jest rozważany w. Chociaż niektóre prace zakładają rozbudowane kanały rozproszone, duża liczba hybrydowych podejść opiera się na niespójności kanału w celu zmniejszenia liczby łańcuchów RF bez znaczącego wpływu na wydajność. Na przykład weź pod uwagę reprezentację kanału i jednego użytkownika dla uproszczenia. Rozproszony kanał komunikacji można zasadniczo opisać za pomocą podzbioru  $v_k \leq N$  wartości własnych i wektory własne i spełnia:

$$\mathbf{h}_k = U_k \sum_k^{1/2} \mathbf{g}_k \approx U_k^{(v_k)} \left( \sum_k^{1/2} \right)^{(v_k)} \mathbf{g}_k^{(v_k)}$$

gdzie  $\sum_k^{(v_k)} \in \mathbb{C}^{v_k \times v_k}$  reprezentuje diagonalną macierz zawierającą nieistotne wartości własne  $v_k$  macierzy kowariancji kanału  $R_k$  i kolumn  $U_k^{(v_k)} \in \mathbb{C}^{N \times v_k}$  są tworzone przez odpowiednie wektory własne. Tutaj,

$$\mathbf{g}_k \in \mathbb{C}^{v_k \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{v_k}).$$

W tym konkretnym przypadku do przybliżenia potrzebne byłyby tylko łańcuchy RF  $N_{RF} = v_k$  maksymalne osiągalne stawki, pod warunkiem że  $W_{RF} = U^{(v_k)}$ , który zależy tylko od statystyk kanału  $R_k$ . Co ciekawe, można wykazać, że w szczególnym przypadku asymptotycznie dużej ULA z krytycznym odstępem, i-ta kolumna optymalnej macierzy RF kształtującej wiązkę  $W_{RF}^{opt}$  odpowiada kolumnie dyskretnej macierzy przekształcenia Fouriera (DFT):

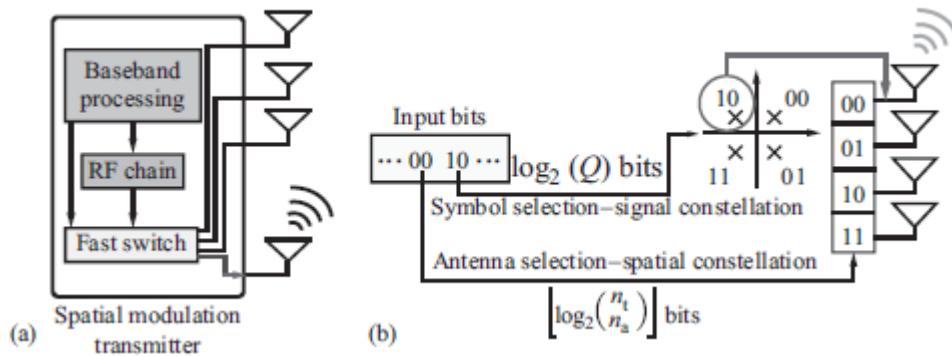
$$w_{RF,i} \approx \frac{1}{\sqrt{N}} \left[ 1, e^{-\frac{2\pi(i-1)l}{N}}, e^{-\frac{4\pi(i-1)l}{N}}, \dots, e^{-\frac{2\pi(i-1)(N-1)l}{N}} \right]^T.$$

Fizycznie można to interpretować jako podział przestrzeni fizycznej na  $N$  wiązek ortogonalnych za pomocą macierzy RF formującej wiązkę, w której jednak tylko podzestaw  $v_k$  z nich może być wykorzystany do transmisji danych ze względu na nieliczny charakter kanału. Odpowiednio, wybór przydatnych kierunków  $v_k$ , które przyczyniają się do propagowania sygnałów między BS a użytkownikami, jest dokonywany za pomocą wstępnego kodowania hybrydowego. Należy zauważyć, że zastosowano podobną koncepcję w celu zmniejszenia efektu zanieczyszczenia pilotażowego. Badania koncentrowały się również na zmniejszeniu złożoności w pełni połączonej analogowej strategii formowania wiązki przedstawionej w tej sekcji. W tym kontekście realistyczne podejście do zmniejszenia złożoności sprzętowej i przetwarzania sygnałów polega na wykorzystaniu korelacji międzykanałowej między antenami rozmieszczonymi blisko siebie i zaprojektowaniu nieoptymalnych podejść do zmniejszenia liczby połączeń i komponentów sprzętowych w analogowej sieci formowania wiązki. Należy jednak zauważyć, że niespójność nie jest ogólnie uważana za daną charakterystykę kanałów komunikacyjnych na częstotliwościach mikrofalowych i konieczne są dalsze badania w celu walidacji podejść opartych na rzadkościach.



### 4.3.1.3 Modulacja przestrzenna

Alternatywną strategią, która zainteresowała zmniejszenie liczby łańcuchów RF w LSAS, jest modulacja przestrzenna (SM), w której transmisja obejmuje tylko kilka (do jednego) łańcuchów RF, które są jednocześnie aktywne. Schemat blokowy nadajnika SM z pojedynczym łańcuchem RF przedstawiono na rysunku 4.13 (a)



Podstawowa idea SM polega na tym, że oprócz tradycyjnej modulacji, takiej jak kwadraturowa modulacja amplitudy (QAM), informacje, kodujące pewne dodatkowe informacje na temat indeksów anteny poprzez aktywację podzbioru anten, w oparciu o dane chwilowe. Generuje to trójwymiarowa konstelacja, jak zilustrowano na rysunku 4.13 (b), gdzie pierwsze dwa wymiary odpowiadają konwencjonalnym symbolom QAM, a trzecia jest określona przez wskaźniki anteny wykorzystywane do transmisji wybranych symboli modulacji. Matematycznie sygnał  $x_u \in \mathbb{C}^{n_t \times 1}$  przesyłany przez nadajnik SM z aktywnymi antenami z całości  $n_t$  anten przyjmuje postać:

$$x_u = [0 \quad \dots \quad s_{l_1}^q \quad \dots \quad s_{l_{n_a}}^q \quad \dots \quad 0]^T,$$

gdzie  $l_i \in \{1 \dots n_t\}$ ,  $i \in \{1, \dots, n_a\}$  oznacza aktywny indeks anteny,  $s^q$  reprezentuje  $q$ -ty symbol konstelacji nadawczej,  $(\cdot)^T$  oznacza transpozycję macierzy. Należy zauważyć, że zarówno aktywne wskaźniki anteny, jak i konwencjonalne symbole konstelacji muszą być oszacowane w odbiorze. SM umożliwia transmisję sumy całkowitej

$$B_{SM} = \lfloor \log_2 \left( \frac{n_t}{n_a} \right) \rfloor + \log_2(Q)$$

bitów na kanał, gdzie  $\lfloor \cdot \rfloor$  oznacza operator podłogi, a  $Q$  jest sygnałem konwencjonalnym kolejność konstelacji. Powoduje to zmniejszenie wydajności widmowej w stosunku do tradycyjnych systemów MIMO z  $B_{MIMO} = n_t \cdot \log_2(Q) > B_{SM}$ . Bity  $B_{SM}$  na kanał są przesyłane jednocześnie, z dedykowanym łańcuchem RF dla każdej z  $n_t$  anten nadawczych. Dlatego SM stanowi kompromis między wydajnością widmową, która jest poświęcana ze względu na zmniejszoną liczbę aktywnych anten, a złożonością sprzętową, która jest zmniejszona ze względu na potrzebę jedynie wymagającą zredukowanej liczby łańcuchów RF. Podczas gdy strategię SM można zastosować do łącza w dół w celu zmniejszenia złożoności sprzętu, są one również szczególnie odpowiednie do zwiększenia szybkości transmisji danych łącza w górę bez znaczącego zwiększenia złożoności zestawów słuchawkowych użytkowników w LSAS

### 4.3.2 Optymalizacja efektywności energetycznej

Oprócz powyższych i zgodnie z innymi dziedzinami w ogólnym sektorze ICT, coraz bardziej niepokoi nas, czy sieci bezprzewodowe są bardziej energooszczędne. W związku z tym analiza i optymalizacja wydajności energetycznej w LSAS przyciągnęły znaczące zainteresowanie badawcze. W związku z tym analiza i optymalizacja wydajności energetycznej w LSAS przyciągnęły znaczące zainteresowanie badawcze. Wydajność energetyczna,  $\varepsilon$ , mierzona w bitach na dżul (bity / J) jest określana jako

$$\varepsilon = \frac{R}{P_{\text{tot}}},$$

gdzie  $R$  odnosi się do metryki szybkości mierzonej w bitach na sekundę (bitów / s), zazwyczaj osiąganą stawkę sumy, a  $P_{\text{tot}}$  to moc pobierana przez składniki sprzętowe zaangażowane w łącze komunikacyjne. O ile początkowe wskaźniki efektywności energetycznej opierały się wyłącznie na zużyciu energii przez wzmacniacze mocy, duża liczba dodatkowych komponentów wymaganych w LSAS spowodowało, że to podejście było niewystarczające, ponieważ powinny zostać zaimplementowane  $N$  transmitujące łańcuchy RF. W związku z tym pojawiło się wiele wskaźników efektywności energetycznej, które uwzględniają zużycie energii zarówno sprzętu, jak i komponentów przetwarzania sygnału. Przyjmujemy tutaj jeden z pełniejszych wskaźników w literaturze. Koncentrując się na transmisji danych, metryka efektywności energetycznej wprowadzona w została dodatkowo rozszerzona na:

$$\varepsilon = \frac{R}{P_{\text{tot}}} = \frac{R(1 - \sigma_{\text{DC}})(1 - \sigma_{\text{MS}})(1 - \sigma_{\text{cool}})}{P_{\text{PA}} + P_{\text{RF}} + P_{\text{BB}}},$$

gdzie  $\sigma_{\text{DC}}$ ,  $\sigma_{\text{MS}}$  i  $\sigma_{\text{cool}}$  odnoszą się do strat poniesionych przez zasilacz DC-DC, odpowiednio głównego zasilania i strat aktywnego chłodzenia. Ponadto z odniesienie do struktury RF z rysunku 4.7,  $P_{\text{PA}} = P_t / \zeta$  odnosi się do średniej mocy zużyte przez wzmacniacze mocy o sprawności  $\zeta$ , aby wytworzyć sygnał o mocy wyjściowej  $P_t$  Watów, podczas gdy  $P_{\text{BB}}$  oznacza moc zużywaną przez pasmo podstawowe cyfrowe procesory sygnałowe i cyfrowe interfejsy danych.  $P_{\text{RF}}$  reprezentuje moc zużycie elementów elektronicznych łańcuchów RF z wyłączeniem mocy wzmacniacze, które można dalej wyrazić jako

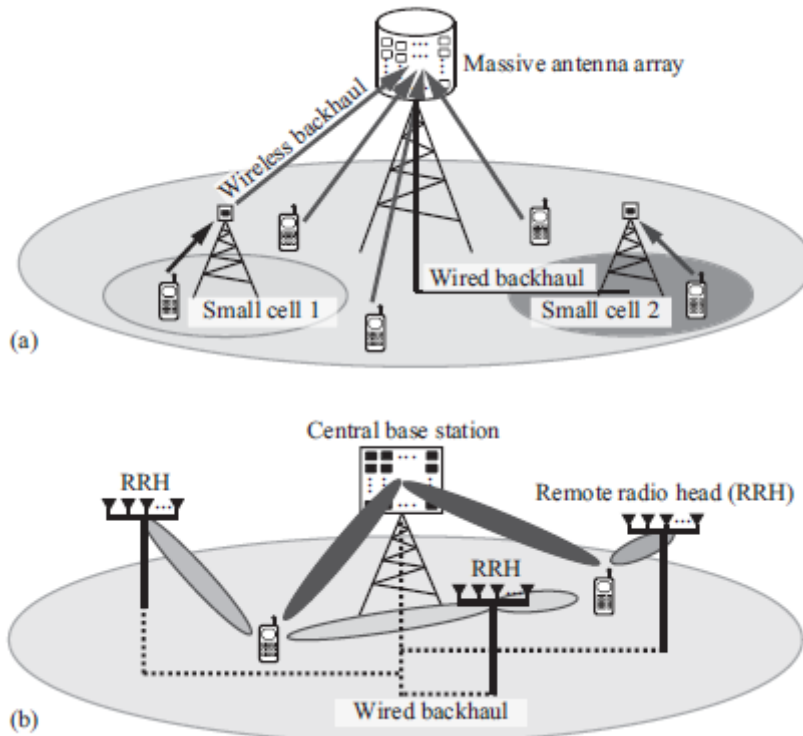
$$P_{\text{RF}} = N_{\text{RF}}(P_{\text{DAC}} + P_{\text{mix}} + P_{\text{filt}}) + P_{\text{syn}},$$

gdzie  $P_{\text{DAC}}$  i  $P_{\text{mix}}$  oznaczają moc wymaganą przez konwertery cyfrowo-analogowe (DAC) i miksery sygnałowe na każdy łańcuch RF. Liczba łańcuchów RF to  $N_{\text{RF}}$ . Ponadto,  $P_{\text{filt}}$  odpowiada zużyciu energii filtrów, a  $P_{\text{syn}}$  to moc pobierana przez syntezytor częstotliwości. Dodatkowo  $P_{\text{BB}} = p_c C$  oznacza zużycie energii procesora cyfrowego sygnału (DSP). Tutaj  $C$  reprezentuje średnią liczbę rzeczywistych operacji zmiennoprzecinkowych (flopów) na sekundę, a  $p_c$  oznacza dżule energii wymagane na rzeczywisty flop. Szczególną uwagę zwrócono na efektywność energetyczną LSAS w literaturze, wyraźnie biorąc pod uwagę pobór mocy komponentów RF. Szczególne problemy badane w kontekście energii efektywność obejmuje optymalizację efektywności energetycznej w zależności od liczby użytkowników  $K$  i  $N$  anten BS, rozwoju energooszczędnego strategii selekcji anten i porównania efektywności energetycznej małych komórek i LSAS, ponieważ obie są przekonującymi strategiami dla 5G. Należy zauważyć, że chociaż bezpośrednio porównywane w obu strategiach, mogą faktycznie współistnieć w przyszłych heterogenicznych sieciach (HetNets), jak opisano w sekcji 4.3.3.

#### 4.3.3 Architektury sieciowe i rozproszone LSAS

Zasady i praktyczne rozważania LSAS zostały wprowadzone w tej części poprzez rozważenie klasycznego jednorodnego scenariusza wielokomórkowego, w którym każda komórka LSAS składa się z dużej matrycy antenowej znajdującej się w jej centrum i służy użytkownikom, którzy wpadają w przypisany jej obszar zasięgu. Jednak przewiduje się, że LSAS będzie kompatybilny z wieloma

architekturami sieciowymi w celu dalszego wzmocnienia ich mocnych stron, o których krótko wspomniano poniżej. Jeden z czołowych scenariuszy 5G obejmuje podejście wielopoziomowe, w którym wiele małych punktów dostępu do komórek obejmuje niższy poziom, a makra BS z dużą liczbą anten stanowią wyższy poziom. Przykład takiej architektury HetNet pokazano na rysunku 4.14 (a),



gdzie można wyraźnie zauważyć, że niektórzy użytkownicy są obsługiwani przez ściśle zlokalizowane małe komórki ze zmniejszoną utratą ścieżki i macierzami anten o małej skali. Rysunek 4.14 (a) pokazuje również, że nadmiarową liczbę anten dostępnych w makro BS można następnie wykorzystać do utworzenia bezprzewodowego łącza dosyłowego do tych małych komórek, chociaż przewodowe rozwiązania typu backhaul są również silnymi kandydatami do realizacji 5G. Szczególnie istotne w tym kontekście jest opracowanie strategii wykorzystujących dużą liczbę anten do tłumienia interferencji między warstwami i poprawy osiągalnych prędkości sieci. Rzeczywiście, LSAS są także elastyczną alternatywą dla ustanawiania bezprzewodowych połączeń z oddalonymi głowicami radiowymi w HetNecie. Inny kwitnący obszar badań w tym kontekście bada wpływ alokacji anten LSAS w sposób rozproszony]. Ta architektura sieci, określana jako rozproszona LSAS, należy do szerszej kategorii wielobranżowych sieci współpracy MIMO. Ilustracyjny przykład rozproszonego LSAS pokazano na rysunku 4.14 (b), gdzie centralne BS jest bezpośrednio połączone z wieloma oddległymi głowicami radiowymi, które są fizycznie oddzielone. Jak wyraźnie pokazano na rysunku 4.14 (b), systemy te polegają na dzieleniu się informacjami, takimi jak dane użytkownika lub CSI pomiędzy stacjami BS i zdalnymi głowicami radiowymi w celu zwiększenia wydajności i różnorodności sieci, kosztem dodanych ogólnych kosztów pośrednich. Ogólnie rzecz biorąc, dyskusja nad rolą wielkoskalowych BS w wdrożeniach sieci wciąż znajduje się na wczesnym etapie i konieczne są dalsze badania w celu oceny ich potencjału i ich wydajności za pomocą technik dostępnych w obecnym standardzie LTE, takich jak rozszerzenie zasięgu komórki lub prawie puste podramki.

#### 4.4 Uwagi końcowe

W niniejszej części przedstawiono zasady systemów LSAS jako jeden z dominujących kandydaci do sieci komunikacyjnych 5G i nie tylko. Jest to niezwykle zalety LSAS, obiecane przez teorię komunikacji, które stanowią główny czynnik do przyjęcia systemów LSAS w przyszłych sieciach. Wdrażanie LSAS jest jednak przedmiotem poważnych wyzwań, które stawiają pytanie, czy obiecane korzyści mogą zostać zmaterializowane w praktyce iw jakim zakresie. Był to główny bodziec do praktycznego podejścia do wdrażania LSAS. Podczas gdy badania akademickie i przemysłowe w zakresie LSAS nabierają rozpędu na całym świecie, wiele problemów badawczych pozostaje otwartych do dnia dzisiejszego, jak wskazano w tym rozdziale, obiecując owocne pole badań w nadchodzących latach.

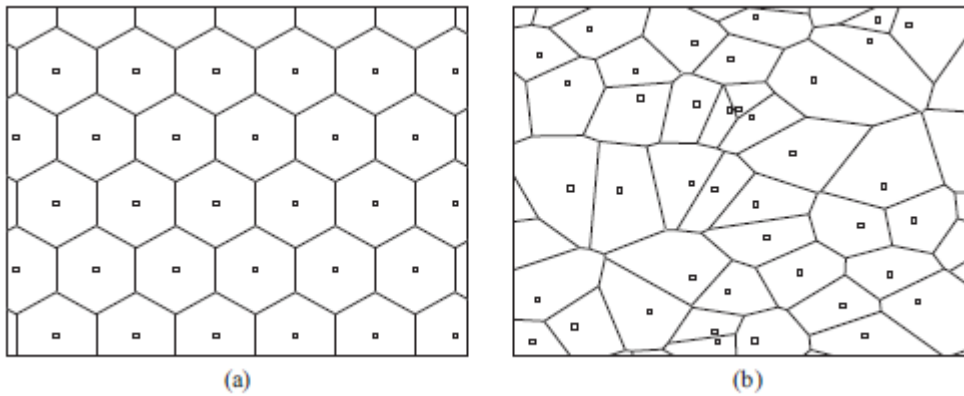
# **Skutki zagęszczania i losowości wdrażania infrastruktury w sieciach komórkowych**

## **5.1.1 Konsumpcja infrastruktury w przyszłych sieciach komórkowych**

Wykładniczy wzrost bezprzewodowego ruchu danych jasno pokazał, że obecne sieci komórkowe nie będą w stanie zapewnić akceptowalnej jakości usług w najbliższej przyszłości. Obserwacja ta skierowała uwagę branży bezprzewodowej i społeczności badawczej na zdefiniowanie, a ostatecznie na wdrożenie nowej (piątej) generacji sieci komórkowych, która jest w stanie obsługiwać wyjątkowo duże szybkości transmisji danych na łącze (bezprzewodowe), a także niezwykle dużą liczbę współbieżnych spinki do mankietów. W tym celu zidentyfikowano różne czynniki technologiczne, które można z grubsza podzielić na trzy główne kategorie:

- Zaawansowane techniki transmisji osiągające wysoką wydajność widmową, np. Masywny MIMO.
- Zwiększenie szerokości pasma transmisji, np. Wykorzystanie częstotliwości mmWave.
- Przeprojektowanie architektury systemu.

Dwie pierwsze kategorie były obecne w ewolucji sieci komórkowych od samego początku i można je postrzegać jako "standardowe" podejście do ulepszenia wydajności e (oczywiście nie oznacza to, że nie wiążą się z nimi żadne istotne wyzwania techniczne). Natomiast trzecia kategoria stanowi rewolucyjną (a nie ewolucyjną) zmianę zasadniczo stabilnego atrybutu sieci komórkowej, a mianowicie jej architektury. W szczególności tradycyjna koncepcja sieci komórkowej składającej się ze stacji bazowych rozmieszczonych na mniej lub bardziej rzadkim i regularnym wzorze staje się przestarzała. Najważniejszą nową koncepcją architektoniczną, częściowo już wdrożoną przez obecne sieci komórkowe, są "małe komórki", generowane przez rozmieszczanie niewielkich rozmiarów punktów dostępowych o małym zasięgu (AP), zasadniczo dostarczających "zastrzyki pojemności" do tradycyjnej sieci, wprowadzając w ten sposób pojęcie heterogenicznej sieci (HetNet). Najwyraźniej pojęcie zagęszczenia infrastruktury jest bezpośrednio związane z siecią małych komórek. Densyfikacja infrastruktury była naturalnie obecna w ewolucji sieci komórkowych, jednak oczekuje się, że jej rola w przyszłych sieciach będzie znacznie bardziej rozpowszechniona, nie tylko ograniczona do "stałych" AP, ale także obejmująca "oportunistyczne" punkty dostępowe odpowiadające nowym przewidywanym metody dostępu, takie jak komunikacja między urządzeniami. Terminy takie jak sieć "ultra-gęsta" lub "hiper-gęstość" stają się coraz bardziej popularne w literaturze, odnosząc się do bardzo wysokich poziomów gęstości infrastruktury (w punktach dostępowych na jednostkę powierzchni), w tym scenariuszy, w których gęstość infrastruktury w tej samej kolejności, co (lub nawet większe niż) gęstość sprzętu użytkownika (UE). To ekstremalne zagęszczenie infrastruktury powoduje powstanie nowego atrybutu systemowego, a mianowicie (quasi-) losowości pozycji AP, której nie było w poprzednich generacjach sieci komórkowych



. Ta przypadkowość jest wynikiem praktycznych trudności operatorów wdrażających infrastrukturę na arbitralnych stanowiskach, korzystania z urządzeń rozmieszczonych przez użytkowników oraz wykorzystywania UE jako dostawców treści

### 5.1.2 Analityczne modelowanie przyszłych sieci komórkowych

Oczywiste jest, że zagęszczenie i losowość wdrażania infrastruktury powoduje nowe środowisko operacyjne, w którym poziomy zakłóceń nie są już kontrolowane przez dokładne umieszczenie AP i mogą potencjalnie stać się bardzo dużymi w pewnych przypadkowych konfiguracjach rozmieszczenia. W związku z tym istotne jest zrozumienie technicznych implikacji tego nowego środowiska, jak również potencjału zagęszczania ekstremalnej infrastruktury jako sposobu na zaspokojenie przyszłych wymagań sieci. Oczywiście, dobrze sprawdzone metody inżynierskie analizy / projektowania systemu poprzez formułowanie i rozwiązywanie odpowiednich problemów optymalizacyjnych mogą być stosowane do tego celu, i to rzeczywiście zostało zrobione szeroko w literaturze. Metody te są wprawdzie wartościowe, ale mają poważne ograniczenia: są z natury ograniczone do skończonej (i zazwyczaj małej) liczby nadajników / odbiorników, a zatem nie mogą w wiarygodny sposób uchwycić skutków dużych i gęstych wdrożeń infrastruktury, krytycznych. atrybuty przyszłej sieci. Ponadto wyniki tych badań są zwykle w formie algorytmicznej, z ograniczonym analitycznym spostrzeżeniem. W kierunku analitycznego badania wydajności przyszłych ultra-gęstych sieci, biorąc wyraźnie pod uwagę losowość pozycji AP i wdrożeń na dużą skalę, naturalne jest rozważenie narzędzi z dziedziny (dwuwymiarowych) procesów losowych punktów, powszechnie zwanych geometria stochastyczna. Narzędzia do geometrii stochastycznej stosowano od dziesięcioleci do modelowania różnych systemów, jednak po raz pierwszy zastosowano je jako narzędzie do badania problemów z siecią komunikacyjną w 1996 roku. Jeśli chodzi o sieci bezprzewodowe / mobilne, geometria stochastyczna była pierwotnie (i nadal jest) intensywnie wykorzystywana w badaniach sieci ad hoc, co jest naturalne, ponieważ podstawową cechą tych sieci są losowe pozycje węzłów bezprzewodowych. Zastosowanie do sieci komórkowych pojawiło się dopiero niedawno w przełomowym dorobku, którego eleganckie wyniki i spostrzeżenia zainicjowały ogromne zainteresowanie badaczy komórkowych metodami geometrii stochastycznej, co zaowocowało ogromną ilością powiązanych dokumentów, które znacząco przyczyniły się do zrozumienia analizy wydajności systemów komórkowych. i projekt. Niedawno opublikowane znakomite książki i monografie dotyczące zastosowań geometrii stochastycznej w sieciach bezprzewodowych wskazują, że dziedzina osiągnęła już wysoki poziom dojrzałości, a (częściowe) zapoznanie się z pokrewnymi technikami i / lub wynikami wydaje się koniecznością dla sieci bezprzewodowych / komórkowych. inżynier systemowy / badacz.

### 5.1.3 Cel i organizacja sekcji

Cel tej sekcji jest dwójaki. Po pierwsze, aby scharakteryzować wydajność systemu w ramach gęstych i losowych wdrożeń infrastruktury, które pomogą uzyskać wgląd w projekt, a po drugie wykazać, w jaki sposób można zastosować narzędzia z geometrii stochastycznej w celu zbadania tych problemów. Liczne reprezentatywne scenariusze operacyjne i rozważane są wskaźniki wydajności, a niektóre z najważniejszych najnowszych wyników są przedstawione wystarczająco szczegółowo, aby można je było uznać za (techniczny) przewodnik dla nowych użytkowników w tej stosunkowo nowej dziedzinie badawczej. Należy zauważyć, że zawartość tej sekcji nie jest w żadnym wypadku wyczerpującym podsumowaniem wyników i technik, które pojawiły się w tej dziedzinie w ostatnich latach. Oczekuje się jednak, że zapewnia on dobry przegląd podstawowych metod analizy, które są powszechne w większości powiązanych prac. Sekcja jest zorganizowana w następujący sposób. Sekcja 5.2 zawiera przegląd / podsumowanie najważniejszych narzędzi z geometrii stochastycznej, które są przydatne do analizy sieci bezprzewodowych. Sekcja 5.3 przedstawia szczegółowo podstawowy model systemu, który zostanie rozważony w tej sekcji odpowiednio wskaźniki wydajności systemu. Obejmuje również dyskusję na ważny temat kryteriów stowarzyszenia AP-UE. Konkretnie scenariusze operacyjne są szczegółowo badane, biorąc pod uwagę stosunek sygnału do przekładni i stopę użytkownika jako metryki będące przedmiotem zainteresowania odpowiednio w sekcjach 5.4 i 5.5. Wreszcie sekcja 5.6 kończy.

Zapis:  $\mathbb{P}(\cdot)$  i  $\mathbb{P}(\cdot | \cdot)$  oznaczają odpowiednio miarę prawdopodobieństwa i warunkowego prawdopodobieństwa.  $\mathbb{E}(\cdot)$  i  $\mathbb{E}(\cdot | \cdot)$  oznaczają operatory oczekiwania i warunkowe oczekiwania, odpowiednio, a  $\mathbb{I}(\cdot)$  jest funkcją wskaźnika. Norma euklidesowa  $x \in \mathbb{R}^2$  oznaczono jako  $|x|$ , a obszar ograniczonego zestawu  $\mathcal{A} \subset \mathbb{R}^2$  jest oznaczone jako  $|\mathcal{A}|$ .

## 5.2 Przegląd jednorodnych procesów Poissona

W tej części przedstawiono krótki przegląd podstawowych właściwości przestrzennych, homogenicznych Poisson Point Processes (PPP), a także niektóre z najważniejszych i powszechnych narzędzi stosowanych w analizie systemów komórkowych opartych na modelu homogenicznego PPP. Prezentacja jest nieformalna w tym sensie, że nie przestrzega się matematycznie rygorystycznego opisu / wyprowadzania wyników, który wymaga znajomości teorii miary. Jednak w przypadku zastosowań inżynierskich rozpatrywanych w tym rozdziale ten poziom rygorystyczności nie jest konieczny, a czytelnik odwołuje się do doskonałych tekstów do bardziej zaawansowanego i dogłębnego traktowania tematu.

### 5.2.1 Podstawowe właściwości

Dwuwymiarowa przestrzeń euklidesowa  $\mathbb{R}^2$  zostanie uwzględniona w tej części. Proces punktowy  $\Phi = \{x_i\}_{i=1,2,\dots} \subset \mathbb{R}^2$  jest policzalnym zbiorem losowo umieszczonych punktów w  $\mathbb{R}^2$  bez punktów kumulacji, tj. liczba punktów zawartych w dowolnym ograniczonym podzbiore  $\mathbb{R}^2$  jest skończona. Z wielu typów procesów punktowych badanych w literaturze najczęściej spotykany jest zdecydowanie jednorodny PPP. A PPP jest jednoznacznie opisany pojedynczym parametrem  $\lambda > 0$ , określanym jako gęstość (lub intensywność) procesu, oraz ma następujące podstawowe właściwości statystyczne:

1. Liczba punktów  $N(\mathcal{A})$  w dowolnym zbiorze ograniczonym  $\mathcal{A} \subset \mathbb{R}^2$  jest zmienną losową Poissona o średniej  $\lambda |\mathcal{A}|$ , tj.

$$\mathbb{P}(N(\mathcal{A}) = k) = \frac{(\lambda |\mathcal{A}|)^k}{k!} e^{-\lambda |\mathcal{A}|}, \quad k = 0, 1, \dots,$$

gdzie  $|\mathcal{A}|$ , to obszar (miara Lebesgue'a) z  $\mathcal{A}$ . Należy zauważyć, że ta właściwość oznacza, że  $\lambda$  znajduje się w jednostkach średniej liczby punktów na jednostkę powierzchni, tj.  $\lambda$  zapewnia wskazanie, jak gęsto punkty PPP są rozmieszczone na płaszczyźnie.

2. Biorąc pod uwagę liczbę punktów  $N$  PPP w jakimkolwiek ograniczonym zbiorze  $\mathcal{A} \subset \mathbb{R}^2$  pozycje tych  $N$  punktów są niezależnie i równomiernie rozłożone w  $\mathcal{A}$ .

3. Dla dowolnych dwóch rozłącznych, ograniczonych zbiorów  $\mathcal{A} \subset \mathbb{R}^2$  i  $\mathcal{B} \subset \mathbb{R}^2$ ,  $\mathcal{A} \cap \mathcal{B} = \emptyset$ . Liczba punktów PPP w tych regionach, odpowiednio,  $N(\mathcal{A})$  i  $N(\mathcal{B})$  są niezależne.

Można wykazać, że powyższe właściwości implikują, że PPP jest prosty, tj. żadne dwa punkty  $\Phi$  nie są prawie takie same (choć mogą być dowolną odległością euklidesową). Właściwości 2 i 3 określa się jako kompletne (lub maksymalne) losowo. Dwie ważne cechy PPP, które wynikają z jej definicji, ale mają wielkie znaczenie i zasługują na wyraźne wymienienie, są następujące.

4. Superpozycja PPP: proces punktowy  $\Phi_1 \cup \Phi_2$ , generowany przez superpozycję dwóch niezależnych PPP  $\Phi_1$  i  $\Phi_2$  o gęstościach odpowiednio  $\lambda_1$  i  $\lambda_2$ , jest również PPP o gęstości  $\lambda_1 + \lambda_2$ .

5. Rozcieńczanie PPP: Niech  $\Phi = \{x_k\}$  oznacza PPP o gęstości  $\lambda$ . Proces punktowy  $\Phi' \subseteq \Phi$  wynikający z niezależnego usunięcia z punktu  $\Phi$  punktu  $x_k$  z prawdopodobieństwem  $1-p$  dla każdego  $x_k \in \Phi$ , jest również PPP o gęstości  $p\lambda$ .

Ważnym rozszerzeniem koncepcji procesów punktowych jest proces zaznaczonego punktu, który jest procesem punktowym  $\Phi$  z każdym punktem  $x \in \Phi$  związanym z losowym wektorem (znacznikiem)  $m \in \mathbb{R}^d$  gdzie  $d \geq 1$  to wymiar przestrzeni znaków. Wyznaczony proces punktowy jest użytecznym narzędziem, które pozwala na zwarte przedstawienie nie tylko (przypadkowych) pozycji zestawu elementów, ale także ich indywidualnych atrybutów. Na przykład, gdy  $\Phi$  reprezentuje pozycje AP sieci komórkowej, znacznik może reprezentować interesujące parametry, takie jak moc nadawania i liczba powiązanych (obsługiwanych) urządzeń użytkownika UE, których wartości są typowo losowe i (łącznie) zależą od lokalizacji wszystkie punkty  $\Phi$ , a także inne parametry zewnętrzne. Zauważ, że zaznaczony proces punktowy jest równoznaczny z nieoznaczonym procesem punktowym przez proste ustawienie wszystkich znaków równych dowolnej arbitralnej ustalonej wartości. Szczególnie interesujące dla celów analitycznych jest niezależnie oznaczony PPP, w którym znaki są niezależne i identycznie rozmieszczone, z znacznikiem  $m_x$  punktu  $x \in \Phi$  w zależności tylko od  $x$ ; to znaczy,

$$p(m_x | \Phi) = p(m_x | x), \text{ dla wszystkich } x \in \Phi;$$

gdzie  $p(m_x | \cdot)$  oznacza funkcję gęstości warunkowej prawdopodobieństwa  $m_x$  (założono, że istnieje).

### 5.2.2 Obliczanie funkcjonałów

Zazwyczaj chce się zbadać właściwości statystyczne funkcjonałów, których wartości zależą od punktów i znaczników punktu  $\Phi$  proces. Jest to na ogół trudne zadanie. Jednak w przypadku niezależnie oznaczonego PPP oczekiwaną wartość pewnych rodzajów funkcjonałów "rozdzielnych" można uzyskać w postaci integralnej, co w wielu przypadkach praktycznych interesów prowadzi do prostych zamkniętych wyrażeń. Zastanów się nad jakąkolwiek (mierzalną) funkcją  $f: \mathbb{R}^2 \times \mathbb{R}^d \mapsto [0, \infty]$  1P i funkcjonalny w postaci

$$f_s(\{(x, m_x)\}_{x \in \Phi \cap \mathcal{A}}) \triangleq \sum_{x \in \Phi \cap \mathcal{A}} f(x, m_x)$$

gdzie  $\Phi$  jest niezależnie oznaczonym PPP a  $\mathcal{A} \subseteq \mathbb{R}^2$ . Twierdzenie Campbella stwierdza, że:

$$\mathbb{E}(f_s) = \lambda \int_{\mathbb{R}^2 \cap \mathcal{A}} \mathbb{E}(f(x, m_x) | x) dx.$$

Jako aplikację należy rozważyć funkcję :



$$I = \sum_{x \in \Phi} P_x l(|x|)$$

gdzie  $\{P_x\}_{x \in \Phi}$  jest ograniczoną sekwencją liczb dodatnich, a  $l: [0, \infty) \mapsto [0, \infty)$  to monotonicznie malejąca funkcja. Ta funkcja może reprezentować całkowitą moc odebraną przez odbiornik znajdujący się w punkcie początkowym płaszczyzny z transmisji AP, których pozycje są rozdzielone jako PPP  $\Phi$  o gęstości  $\lambda$ , przy czym każde AP transmituje z losowy i.i.d. moc  $P_x$  i  $l(\cdot)$  jest deterministycznym modelem utraty ścieżki, który zależy tylko na odległość euklidesową między źródłem a miejscem docelowym. Bezpośrednia aplikacja ) pokazuje, że średnia moc odbierana jest równa:

$$\begin{aligned} \mathbb{E}(I) &= \lambda \int_{\mathbb{R}^2} \mathbb{E}(P_x |x| l(|x|)) dx \\ &= 2\pi\lambda \mathbb{E}(P_x) \int_0^\infty r l(r) dr \end{aligned}$$

gdzie druga równość zachodzi w przypadku, gdy  $P_x$  jest niezależny od  $x$  i przetacza na współrzędne biegunowe całki. Należy zauważyć, że  $\mathbb{E}(I)$  wymaga oszacowania całki, która w najgorszym przypadku może być oceniona numerycznie (jeśli istnieje). Mimo to, odpowiednia złożoność obliczeniowa jest znacznie mniejsza niż prosty, oparty na symulacji (Monte Carlo) oszacowanie funkcjonatu. Co ciekawe, dla wspólnego modelu utraty ścieżki  $l(|x|) = |x|^{-\alpha}$ ,  $\alpha > 0$  powyższa całka nie istnieje, tj. średnia odebrana moc jest nieograniczona. Jest to spowodowane faktem, że model PPP pozwala na realizację, gdzie istnieją punkty arbitralnie zbliżone do miejsca pochodzenia (lub jakiegokolwiek innego punktu w tym zakresie). Drugi typ funkcjonatów, dla których można znaleźć wyrażenia w postaci całkowej dla ich oczekiwań, ma postać:

$$f_p(\{(x, m_x)\}_{x \in \Phi \cap \mathcal{A}}) \stackrel{\Delta}{=} \prod_{x \in \Phi \cap \mathcal{A}} f(x, m_x)$$

z  $f(\cdot, \cdot)$  dowolną mierzalną funkcją o wartości dodatniej, jak te rozważane w twierdzeniu Campbella,  $\Phi$  niezależnie oznaczone PPP i  $\mathcal{A} \subseteq \mathbb{R}^2$ . To może być pokazane, że :

$$\mathbb{E}(f_p) = \exp\left(-\lambda \int_{\mathbb{R}^2 \cap \mathcal{A}} [1 - \mathbb{E}(f(x, m_x)|x)] dx\right)$$

Równanie to jest często określane jako funkcja generująca prawdopodobieństwo (PGFL) wzór dla PPP o gęstości  $\lambda$ . Formuła PGFL jest prawdopodobnie najbardziej potężnym narzędzie analityczne do PPP i jest rutynowo stosowane w prawie każdym urządzeniu bezprzewodowym przy badaniu sieci opartym na modelach PPP. Wiele przykładów jego zastosowania będzie pod warunkiem poniżej.

### 5.2.3 Kondycjonowanie, stacjonarność i ergodyczność

Inna ważna właściwość PPP  $\Phi$  dotyczy jej statystycznej charakterystyki, biorąc pod uwagę, że istnieje punkt  $x_0 \in \Phi$  dla niektórych arbitralnych  $x_0 \in \mathbb{R}^2$ . Na przykład, można zainteresować się badaniem właściwości sieci bezprzewodowej punktów dostępowych rozproszonych jako PPP, biorąc pod uwagę istnienie punktu dostępowego w określonej pozycji zainteresowania. Zauważ, że ta konfiguracja rozszerza dobrze znane pojęcie warunkowych, wektorowych zmiennych losowych do przestrzeni procesów punktowych. W zasadzie oczekuje się, że to uwarunkowanie spowoduje zmianę właściwości statystycznych  $\Phi$ , podobnie jak w przypadku "konwencjonalnych" zmiennych losowych. Co godne uwagi, nie jest tak w przypadku, gdy  $\Phi$  jest PPP. W szczególności można rygorystycznie wykazać w ramach teorii Palm, że właściwości statystyczne punktów warunkowego PPP, z wyłączeniem tego na

poziomie  $x_0$ , są takie same jak punktów bezwarunkowego PPP. Innymi słowy, aby wygenerować próbkę PPP zawierającą punkt przy  $x_0$ , należy pobrać próbkę z bezwarunkowego PPP i dodać dodatkowy punkt przy  $x_0$  (próbka bezwarunkowego PPP nie będzie zawierała  $x_0$  z prawdopodobieństwem 1). Ważnym zastosowaniem tej własności PPP jest rozszerzenie twierdzenia Campbella o funkcję o wartości dodatniej  $f$ , która zależy nie tylko od pozycji określonego punktu i jego znaku, ale także od pozycji wszystkich pozostałych punktów PPP. W szczególności zawiera:

$$\mathbb{E} \sum_{x \in \Phi \cap \mathcal{A}} f(x, m_x, \Phi \setminus \{x\}) = \lambda \int_{\mathcal{A}} \mathbb{E}(f(x, m_x, \Phi)) dx,$$

gdzie oczekiwanie wewnątrz całości jest w.r.t. statystyki  $m_x$  (podane  $x$ ) i  $\Phi$ . Dwie ostatnie ważne właściwości PPP to stacjonarność i ergodyczność. Stacjonarność oznacza, że właściwości statystyczne  $\Phi$  pozostają niezmiennicze gdy jego punkty są przesunięte przez wektor  $v \in \mathbb{R}^2$ . Z powodu stacjonarności wynika, że właściwości PPP uwarunkowane istnieniem punktu w punkcie  $x_0$  są niezależne od  $x_0$ . Jest to powód, dla którego  $x_0$  jest określane jako "typowy punkt", zwykle uważany za pochodzenie, tj.  $x_0 = (0,0)$ , dla wygody matematycznej. Można pokazać, że PPP jest również izotropowe, tj. Jego właściwości statystyczne pozostają niezmiennicze przez wspólną rotację jego punktów w  $\mathbb{R}^2$ . Ergodyczność oznacza, że dla dowolnej realizacji  $\phi$  PPP  $\Phi$ , przestrzenna średnia funkcjonalna postaci  $f(\phi \setminus \{x\})$ ,  $x \in \phi$  jest równa średniej zespołu tej funkcjonalności ocenianej w typowym punkcie, tj. :

$$\lim_{n \rightarrow \infty} \frac{1}{N_\phi(\mathcal{A}_n)} \sum_{x \in \phi \cap \mathcal{A}_n} f(\phi \setminus \{x\}) = \mathbb{E}(f(\Phi \setminus \{x_0\}))$$

gdzie  $\{\mathcal{A}_n\}$  jest dobrze zachowaną sekwencją wypukłych podzbiorów  $\mathbb{R}^2$  z  $\mathcal{A}_n \subset \mathcal{A}_{n+1}$ ,  $N_\phi(\mathcal{A})$  to liczba punktów  $\phi$  zawarta w  $\mathcal{A}$ ; a  $x_0$  jest (arbitralnie pozycjonowanym) typowym punktem. Jako przykład rozważ ponownie przypadek, w którym pozycje transmitujących punktów AP są dystrybuowane jako PPP. (Przestrzenna) średnia interferencja doświadczana w pozycjach AP dla każdej realizacji PPP jest równa (średniej) średniej interferencji doświadczanej przez AP umiejscowione w typowym punkcie. Nie można wystarczająco podkreślić znaczenia właściwości ergodyczności. Dla uproszczenia, ta sama notacja,  $\Phi$  zostaną użyte poniżej, aby określić zarówno PPP, jak i jego realizacje.

### 5.3 Model systemu i wskaźniki wydajności

Aby umożliwić łatwą analizę sieci komórkowych, należy wziąć pod uwagę pewne założenia dotyczące modelowania. Te założenia, które można uznać jak upraszczając w niektórych przypadkach, należy wziąć pod uwagę krytyczne aspekty (przyszłości) sieci, np. topologia systemu i warunki propagacji sygnału. Analiza tych modeli będzie, miejmy nadzieję, dostarczać wgląd w projekt systemu dla realistycznych wdrożeń systemu, nawet jeśli prognozy wydajności ilościowej mogą nie być bardzo dokładne. Oczywiście jest, że powszechna praktyka symulacji systemu jest jedyną metodą uzyskiwania dokładnych oszacowań wydajności, biorąc pod uwagę wiele parametrów i / lub upośledzeń, które trudno jest włączyć do badania analitycznego. Jednakże, jak zauważono, rosnąca złożoność przyszłej architektury sieci pod względem np. Nieregularności pozycji AP i wielu współistniejących poziomów, sprawia, że podejście symulacji jest niezwykle trudne, ponieważ zakres scenariuszy, które należy zbadać, staje się niezwykle duży. . Podejście do modelowania analitycznego pozwala na identyfikację trendów niektórych parametrów o znaczeniu krytycznym, ograniczając w ten sposób zakres scenariuszy, dla których należy wykonać symulacje wymagające obliczeń w ramach bardziej szczegółowego modelu systemu. Ta sekcja zawiera opis najczęstszych założeń stosowanych w badaniach sieci komórkowych w ramach geometrii stochastycznej. Ten model systemu będzie

stanowić podstawę do modelowania i analizowania różnych interesujących scenariuszy w dalszej części tego rozdziału. Ponadto w tej sekcji omówiono odpowiednie wskaźniki wydajności, które mierzą wydajność sieci, a także kryteria asocjacji AP-UE używane w analizie.

### 5.3.1 Model systemu

#### 5.3.1.1 Dystrybucja punktów dostępowych i użytkowników

Uwzględniona zostanie heterogeniczna sieć komórkowa (HetNet) złożona z wielu poziomów, z których każdy składa się z policzalnego zestawu punktów dostępowych (AP) rozmieszczonych na  $\mathbb{R}^2$ . Należy zwrócić uwagę, że rozpatrywana jest nieskończona liczba punktów dostępowych, które - choć nierealistyczne - są odpowiednie jako model dla dużych i / lub bardzo gęstych wdrożeń systemu. Punkty dostępowe należące do tej samej warstwy mają takie same właściwości, jak technologia dostępu radiowego i możliwości mocy nadawania. Tradycyjne badania (jednopoziomowych) sieci komórkowych uwzględniają sześciokątny wzór dla pozycji AP (stacji bazowych), który oprócz wymagającego intensywności obliczeniowej symulacje jako główne narzędzie oceny wydajności stały się mniej istotne do modelowania przyszłej sieci komórkowej. Ta ostatnia obserwacja opiera się na wprowadzenie z łatwością gęstych warstw małych komórek, tj. warstw składających się z wielu niskich kosztów wdrożone punkty dostępu, nałożone na te same obszary geograficzne i włączone w system w sposób oportunistyczny, co sprawia, że model pozycji regularnych dla pozycji AP jest nierealistyczny. Ta nieprawidłowość wykazała się również w pewnym stopniu do faktycznego rozmieszczenia makrokomórek. Naturalne jest zatem modelowanie pozycji AP każdego poziomu jako procesu losowego punktu, który, dla analitycznej zdolności krępowania, jest uważany za niezależny (od wszystkiego innego) PPP. W szczególności przypadkowe losowe pozycje AP  $i$ -tego poziomu sieci będą reprezentowane przez PPP  $\Phi_{a,i} \subset \mathbb{R}^2$  o gęstości  $\lambda_{a,i} > 0$ . W przypadku sieci jednopoziomowej indeks  $i$  zostanie usunięty z zapisu. Realizacja jedno-poziomowej sieci komórkowej pokazano na rysunku 5.1 (b). Należy zauważyć, że realizacja modelu PPP zwykle prowadzi do obszarów, na których gęstość punktów dostępowych jest lokalnie większa niż w innych obszarach, co przypomina istnienie gorących punktów w praktycznym wdrażaniu. Modelowanie pozycji UE jako kolejnego procesu punktowego w  $\mathbb{R}^2$  może być poparte podobnymi argumentami jak w przypadku AP, tym bardziej biorąc pod uwagę mobilność niektórych typów UE. W dalszej kolejności pozycje UE będą uważane za inne niezależne (od wszystkiego innego) PPP  $\Phi_u \subset \mathbb{R}^2$  gęstości  $\lambda_u > 0$ . Należy zauważyć, że niezależność pozycji APs i UE nie jest prawdziwa w przypadku praktycznych wdrożeń, gdzie oczekuje się, że obszary o lokalnie dużej gęstości UE będą uzupełnione lokalnie większą gęstością AP (hot) -spot obszary). Jednak to założenie dotyczące niezależności zostało użyte wyłącznie w literaturze w celu zapewnienia wygody analitycznej przy założeniu, że model może nie być wystarczająco dokładny w pewnych scenariuszach.

#### 5.3.1.2 Utrata ścieżki i zanikanie na małą skalę

Wszystkie nadajniki i odbiorniki w sieci są, dla uproszczenia, wyposażone w pojedynczą antenę wielokierunkową. Model utraty ścieżki rządzący modelem zmniejszenie mocy sygnału z nadajnika do odbiornika łącza bezprzewodowego jest uważane za funkcję zmniejszającą  $l: [0, \infty) \mapsto [0, \infty)$  odległości między dwoma węzłami. Najprostszym i najczęstszym modelem stosowanym w literaturze jest :

$$l(r) = r^{-\alpha},$$

gdzie  $r \geq 0$  to odległość łącza, a  $\alpha > 0$  to wykładnik strat ścieżki. Należy zauważyć, że jest to znormalizowana wersja modelu z  $l = 1$  dla  $r = 1$ , tj. rzeczywisty model utraty ścieżki ma postać  $(\beta r)^{-\alpha}$  z  $\beta \neq 1$ . Model znormalizowany jest zwykle stosowany, ponieważ jest matematycznie wygodny i nie wpływa na jakość analizy. Głównym zastrzeżeniem tego modelu jest to, że utrata ścieżki faktycznie staje się wzmocnieniem ścieżki dla  $r < 1$  i rośnie nieograniczona jako  $r \rightarrow 0$ ; co oczywiście jest nierealne.

Z tego powodu modele utraty ścieżki, takie jak  $l^{\circ} = (r + \epsilon)^{-\alpha}$ ,  $\epsilon > 0$  lub  $l(r) = \max(r^{-\alpha}, 1)$ , które eliminują tę kwestię, zostały zastosowane w literaturze. Jednakże, jak zostanie wykazane poniżej, prosty model daje rozsądne wyniki, gdy rozważa się schematy dekodowania pojedynczego użytkownika (gdzie zakłócenia są traktowane jako szum), a wykładnik straty ścieżki jest większy niż 2 (to ostatnie jest wymaganiem technicznym gwarantujące, że moc interferencyjna jest skończona z prawdopodobieństwem). Oprócz utraty ścieżki, na jakość łącza ma również wpływ niewielkie (szybkie) zanikanie z powodu (nie) spójnej superpozycji fal elektromagnetycznych docierających do odbiornika z wielu ścieżek. Znany model Rayleigha zastosowany w teorii komunikacji / informacji, w którym przy niewielkiej skali przyjmuje się wykładniczo rozłożoną zmienną losową średniej jednostkowej, jest również szeroko stosowany w badaniach systemów komórkowych. Przyjmuje się, że zanikanie na małą skalę jest niezależne między linkami. Rozszerzenia modelu Rayleigha, takie jak blaknięcie Nakagami i zanikanie logonormalne, można również znaleźć w literaturze, jednak kosztem nadmiernego komplikowania analizy bez oferowania istotnych spostrzeżeń.

Uwaga: W literaturze dotyczącej badań komórkowych opartych na strukturze geometrii stochastycznej uwzględniono wyłącznie zanikanie płaskie, tj. Dyspersja wielościeżkowa kanału jest znacznie mniejsza niż okres transmisji, lub, równoważnie, transmisje wąskopasmowe są brane pod uwagę, co nie jest zgodne z obecnym łączem szerokopasmowym sieci. Odpowiednie wyniki można zatem postrzegać jako przybliżenie wydajności w ramach kanałów selektywnych częstotliwościowo, lub dokładniej, jako wydajność transmisji ograniczonej do pojedynczego bloku zasobów doświadczającego płaskiego zanikania. W każdym przypadku rozszerzenie analizy na przypadek selektywności częstotliwości pozostaje otwartym tematem, który ma kluczowe znaczenie dla zbadania problemów, takich jak wykorzystanie wzmocnienia częstotliwości i wspólne planowanie urządzeń użytkownika na częstotliwościach.

### **5.3.1.3 Schemat wielokrotnego dostępu i dekodowanie sygnału**

Teoria informacji sugeruje, że ostateczne ograniczenia sieci komórkowej zstępującej wydajność pod względem szybkości transmisji (pojemności) osiąga się za pomocą schematu, w którym dane każdego UE są dostępne dla każdego AP, a transmisje są wykonywane wspólnie, w tym samym czasie i przy częstotliwościach. W przypadku łącza w górę, wspólne dekodowanie wszystkich UE powinno być wykonywane przez wszystkie punkty dostępowe systemu. Niestety, wydajność tych schematów jest bardzo trudna do scharakteryzowania w ogóle, a co najważniejsze, związana z tym złożoność kodowania / dekodowania i wymagania (na przykład nieskończona zdolność przenoszenia wstecz, techniki wykrywania wielu użytkowników itp.) są nierealistyczne. Konieczne jest zatem zastosowanie w analizie prostszych, ale bardziej praktycznych schematów transmisji (wielokrotnego dostępu). Większość prac w dziale ramowym geometrii stochastycznej przyjmuje schemat ortogonalnego wielokrotnego dostępu, w którym każde urządzenie UE jest obsługiwane przez pojedynczy AP (ten ostatni określony według pewnego kryterium stowarzyszenia AP-UE, które zostanie omówione później). Jest to również podejście stosowane przez obecne sieci komórkowe [23]. Prosty schemat time-division-multiple-access (TDMA) będzie rozpatrywany w dalszej części dla obsługi UE związanych z pewnym AP (dla połączenia wstępującego lub łącza downlink), chociaż każdy inny schemat ortogonalny, np. FDMA, CDMA, może być stosowane bez wpływu na analizę i wyniki. Żadne priorytety nie są rozpatrywane wśród UE (planowanie rutynowe), a każdy AP podejmuje niezależne decyzje planowania, tj. Nie zakłada się koordynacji między punktami dostępowymi. Dekodowanie pojedynczego użytkownika odbywa się w odbiornikach, tj. Interferencja jest traktowana jako szum. Jest to znowu suboptymalna, ale mimo to praktyczna metoda wykrywania.

### **5.3.1.3 Schemat wielokrotnego dostępu i dekodowanie sygnału**

Teoria informacji sugeruje, że ostateczne ograniczenia sieci komórkowej zstępującej wydajność pod względem szybkości transmisji (pojemności) osiąga się za pomocą schematu, w którym dane każdego UE są dostępne dla każdego AP, a transmisje są wykonywane wspólnie, w tym samym czasie i przy częstotliwościach. W przypadku łącza w górę, wspólne dekodowanie wszystkich UE powinno być wykonywane przez wszystkie punkty dostępowe systemu. Niestety, wydajność tych schematów jest bardzo trudna do scharakteryzowania w ogóle, a co najważniejsze, związana z tym złożoność kodowania / dekodowania i wymagania (np. nieskończona zdolność przenoszenia, techniki wykrywania wielu użytkowników itp.) Są nierealistyczne. Konieczne jest zatem zastosowanie w analizie prostszych, ale bardziej praktycznych schematów transmisji (wielokrotnego dostępu). Większość prac w szkielecie geometrii stochastycznej zakłada schemat ortogonalnego wielokrotnego dostępu, w którym każdy UE jest obsługiwany przez pojedynczy AP (ten ostatni określony przy pewnym kryterium stowarzyszenia AP-UE, który zostanie omówiony później). Jest to również podejście stosowane przez obecne sieci komórkowe. Prosty schemat time-division-multiple-access (TDMA) będzie rozpatrywany w dalszej części dla obsługi UE związanych z pewnym AP (dla połączenia wstępującego lub łącza downlink), chociaż każdy inny schemat ortogonalny, np. FDMA, CDMA, może być stosowane bez wpływu na analizę i wyniki. Żadne priorytety nie są rozpatrywane wśród UE (planowanie rutynowe), a każdy AP podejmuje niezależne decyzje planowania, tj. Nie zakłada się koordynacji między punktami dostępowymi. Dekodowanie pojedynczego użytkownika odbywa się w odbiornikach, tj. Interferencja jest traktowana jako szum. Jest to znowu suboptymalna, ale mimo to praktyczna metoda wykrywania. Biorąc pod uwagę uniwersalne ponowne wykorzystanie częstotliwości, tj. Wszystkie transmisje są wykonywane w tej samej szerokości pasma 3 i biorąc pod uwagę zagęszczanie infrastruktury sieci komórkowej, a także wiele aktywnych urządzeń użytkownika, jasne jest, że interferencja (interkom) staje się czynnikiem ograniczającym dla pomyślnych transmisji. Rozsądne jest zatem ignorowanie wpływu szumu cieplnego na wydajność systemu, tj. Uwzględnienie warunków operacyjnych o ograniczonej interferencji, co również upraszcza analizę. Jak wykazano w wielu badaniach, zaniedbanie hałasu termicznego ma marginalny wpływ na dokładność wyników. Wreszcie, zgodnie z ramową teorią informacji, rozważane są pełne bufory, tj. Dane są zawsze dostępne dla transmisji do określonego odbiorcy.

#### **5.3.1.4 Podsumowanie modelu systemu**

Podsumowując, wspólne założenia dotyczące modelu systemu we wszystkich przypadkach uwzględnionych w tej części (a także w większości pokrewnej literatury) są następujące, chyba że stwierdzono inaczej:

1. UE i poziomy AP są dystrybuowane zgodnie z niezależnymi PPP.
2. Model utraty ścieżki, ten sam wykładnik straty ścieżki  $a > 2$  dla wszystkich warstw.
3. Płaskie kanały zaniknięcia Rayleigha, niezależne od łącza AP-UE.
4. Pojedyncze, dookólne złącza antenowe.
5. Naprawiono moc nadawania dla punktów AP tego samego poziomu (ogólnie różne moce transmisji między poziomami).
6. TDMA na AP, brak koordynacji / współpracy między AP.
7. Dekodowanie pojedynczego użytkownika.
8. Powtórne wykorzystanie częstotliwości, ograniczenia operacyjne.
9. Transmisje z pełnym buforem.

### 5.3.2 Metryki wydajności systemu

Aby ocenić wydajność sieci komórkowej, odpowiednią wydajność należy zdefiniować metryki, które tradycyjnie są powiązane z koncepcjami pokrycia i przepustowości, odzwierciedlając zdolność sieci do zapewnienia (minimum jakość) usług na dużym obszarze geograficznym i dużych prędkości transmisji do dużych liczb użytkowników, odpowiednio. Mimo że istnieje wiele dyskusji na temat włączenia dodatkowych wskaźników odzwierciedlających inne cechy interesujące w przyszłości sieć, na przykład opóźnienie, podstawowa teoria nie jest dojrzała, dokonując analizy i projektowanie dostosowane jednoznacznie do tych koncepcji uciążliwe. W związku z tym konwencjonalne, oparte na stawkach wskaźniki zasięgu i pojemności pozostają tak popularne jak nigdy dotąd w literaturze komórkowej i są omówione poniżej.

#### 5.3.2.1 Stosunek sygnału do zakłóceń

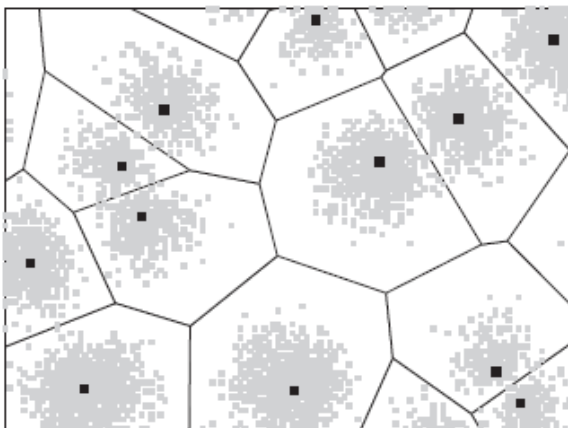
Naturalną metryką związaną z pojęciem pokrycia jest stosunek sygnału do interferencji (SIR) 4, który biorąc pod uwagę typowy odbiornik (UE dla łącza w dół, AP dla transmisji łącza uplink) znajduje się w.l.o.g. u źródła jest równy:

$$\text{SIR} = \frac{P_{x^*} g_{x^*} |x^*|^{-\alpha}}{\sum_{x \in \Phi \setminus \{x^*\}} P_x g_x |x|^{-\alpha}},$$

gdzie  $\Phi \subset \mathbb{R}^2$  jest zbiorem pozycji nadajników ( $\Phi = \cup_i \Phi_{a,i}$  lub  $\Phi = \Phi_u$  dla downlin i uplink, odpowiednio),  $P_x > 0$  i  $g_x > 0$  to moc nadawania i zanikania Rayleigha nadajnika zlokalizowanego odpowiednio przy  $x \in \Phi$ , a przy  $x^* \in \Phi$  jest umiejscowieniem obsługującego nadajnika. Oczywiście, zgodnie z opisanymi powyżej założeniami modelu systemu, SIR jest zmienną losową ze względu na losowość lokalizacji nadajników i zysków z zanikania kanałów. W celu scharakteryzowania zasięgu sieci, interesujące jest znalezienie prawdopodobieństwa zasięgu:

$$\mathbb{P}(\text{SIR} \geq \theta) = \mathbb{E}(\mathbb{I}(\text{SIR} \geq \theta)), \quad \theta > 0.$$

Prawdopodobieństwo pokrycia całkowicie charakteryzuje statystyczne właściwości SIR, ponieważ jest ono równe komplementarnej dystrybuantycznej funkcji (CCDF) SIR. SIR CDF, zwykle określane jako prawdopodobieństwo wyłączenia SIR, jest oczywiście równy  $\mathbb{P}(\text{SIR} \leq \theta) = 1 - \mathbb{P}(\text{SIR} \geq \theta)$ . Zwróć uwagę, że według właściwości ergodyczności PPP, oceniając prawdopodobieństwo zasięgu  $\mathbb{P}(\text{SIR} \geq \theta_0)$  dla niektórych  $\theta_0 > 0$  jest równoważne obliczeniu części (procent pokrycia)  $\mathbb{R}^2$ , dla której  $\text{SIR} \geq \theta_0$  dla dowolnej realizacji  $\Phi$ . Rysunek 5.2 pokazuje przykład sieci i jej obszaru pokrycia, dla którego  $\text{SIR} \geq 10$  dB.



### 5.3.2.2 Efektywność widmowa łącza

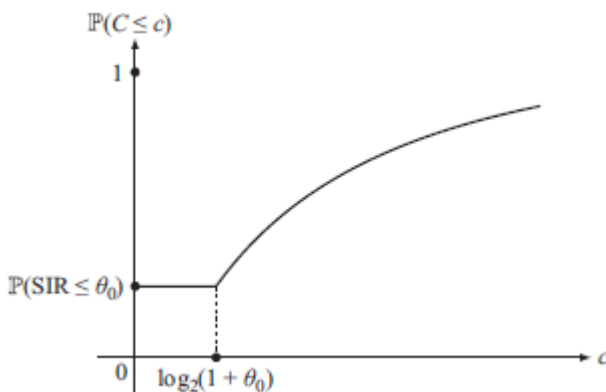
Pojęcie przepustowości sieci komórkowej jest bezpośrednio związane ze sprawnością spektralną w stosunku do łącza, która zapewnia górne ograniczenie szybkości w jednostkach bitów/s/ Hz, które AP może niezawodnie transmitować (przypadek łącza w dół) lub odbierać (łącze w górę). Przy założeniu sygnalizacji Gaussa dla wszystkich nadajników sprawność spektralna typowego połączenia pomiędzy odbiornikiem umieszczonym w punkcie początkowym a związanym z nim nadajnikiem jest bezpośrednio związana z odbiornikiem SIR zgodnie z dobrze znaną formułą pojemności

$$C(\text{SIR}) = \log_2(1+\text{SIR}) \mid (\text{SIR} \geq \theta_{\text{th}}),$$

gdzie  $\theta_{\text{th}} \geq 0$  to próg SIR, podyktowany wymaganiami aplikacji, systemu i synchronizacji, poniżej których rozważana jest wiarygodna komunikacja nieudane. Ponieważ SIR łącza komórkowego jest losowe, tak samo wydajność widmowa z CDF jest równa:

$$\mathbb{P}(C \leq c) = \begin{cases} \mathbb{P}(\text{SIR} \leq 2^c - 1) & c \geq \log_2(1 + \theta_{\text{th}}) \\ \mathbb{P}(\text{SIR} \leq \theta_0) & c < \log_2(1 + \theta_{\text{th}}). \end{cases}$$

Jakościowy przykład sprawności widmowej CDF przedstawiono na rysunku 5.3.



### 5.3.2.3 Wskaźnik UE

Efektywność widmowa łącza jest powszechnie przyjętą miarą w literaturze komórkowej ,tłumaczy możliwości SIR sieci na możliwości szybkości AP poprzez prostą transformację zmiennych. Jednakże, chociaż sprawność spektralna łącza jest ważną i użyteczną miarą / wskaźnikiem wydajności sieci , dostarcza jedynie częściowej informacji o rzeczywistej szybkości odczuwanej przez obsługiwane UE. Wynika to z tego, że szybkość UE zależy nie tylko od wydajności spektralnej łącza, ale także od współdzielenia zasobów między wieloma UE obsługiwanymi przez ten sam AP. Biorąc pod uwagę typowe UE znajdujące się w dowolnym położeniu na płaszczyźnie, rzeczywista stopa, której doświadczą, jest równa:

$$R \triangleq \frac{1}{1 + K_0} C,$$

w bitach/s/Hz, gdzie  $K_0 \geq 0$  oznacza liczbę dodatkowych UE obsługiwanych przez ten sam AP ponieważ typowe UE i C są wydajnością widmową łącza AP-UE, jak opisano powyżej. Termin  $(1/(1+K_0)) \leq 1$  będzie określany poniżej jako stopień wolności transmisji UE, zapożyczając terminologię z teoretycznej literatury komunikacji. Odzwierciedla to utratę stawki z tytułu TDMA, a jej wartość zależy od pozycje UE i AP, a także schemat asocjacji AP-UE.

### 5.3.3 Schematy asocjacji AP-UE

Różne schematy asocjacyjne AP-UE zostały zaproponowane / zbadane w literaturze. Biorąc pod uwagę typowy odbiornik (UE lub AP odpowiednio dla transmisji w łączy w dół lub w górę) i oznaczający pozycję skojarzonego (obsługującego) nadajnik jako  $x$ ; wszystkie systemy można traktować jako szczególne przypadki zasady stowarzyszenia ważonej ścieżki, dla której:

$$x^* = \arg \max_{x \in \Phi} w_x |x|^{-\alpha},$$

gdzie  $\Phi$  oznacza lokalizacje nadajników, a  $w_x \in \mathbb{R}$  oznacza wagę powiązania dla nadajnika ustawionego na  $x \in \Phi$ . Ogólnie, wagi asocjacji są zmiennymi losowymi dystrybuowanymi według pewnego wspólnego pliku pdf, z różnymi odbiornikami "obserwującymi" różne realizacje tych ciężarów. Najbardziej praktyczne opcje dla odważników są następujące.

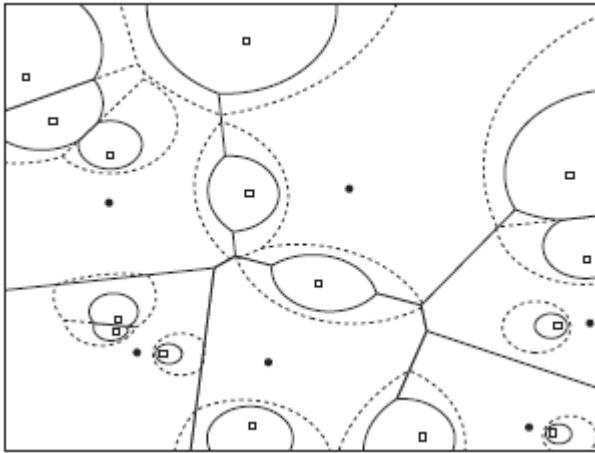
- $w_x = 1$ : powiązanie geometryczne (najbliższy nadajnik)
- $w_x = P_x$ : maksymalne średnie otrzymane skojarzenie mocy
- $w_x = P_x g_x$ : maksymalna chwilowa moc odbioru mocy

Dla pierwszych dwóch przypadków ten sam  $w_x$  jest obserwowany przez wszystkie AP, podczas gdy ostatni przypadek powoduje różne wartości  $w_x$  z powodu niezależnych zysków zanikania między połączeniami AP-UE. Ważną cechą procedury stowarzyszenia z ważoną ścieżką-utrata jest to, że może być ona realizowana wyłącznie po stronie UE, mierząc moc dedykowanych sygnałów pilotujących (referencyjnych), jak ma to również miejsce w praktyce. Zwróć uwagę, że geometryczne i maksymalne średnie otrzymane skojarzenie mocy spowoduje inne skojarzenia tylko w przypadku, gdy rozważane jest połączenie w dół sieci wielowarstwowej z różnymi mocami transmisji między poziomami. W przypadku połączeń jednokrotnego połączenia w dół i połączenia w górę te dwie reguły asocjacyjne są zasadniczo takie same. Ciekawa interpretacja maksymalnego chwilowego powiązania mocy otrzymanej można uzyskać, badając wynikowy SIR, który jest równy:

$$\frac{\max_{x \in \Phi} P_x g_x |x|^{-\alpha}}{\sum_{y \in \Phi \setminus \{\arg \max_{x \in \Phi} P_x g_x |x|^{-\alpha}\}} P_y g_y |y|^{-\alpha}} = \max_{x \in \Phi} \frac{P_x g_x}{\sum_{y \in \Phi \setminus \{x\}} P_y g_y |y|^{-\alpha}},$$

pokazując, że ta reguła asocjacji jest równoważna z maksymalną chwilową SIR stowarzyszenie. Powyższe zasady asocjacji uwzględniają tylko otrzymaną moc do robienia decyzje stowarzyszeniowe. Jednak takie podejście może być wysoce nieoptymalne, gdy metryka odsetek jest stopą UE, która, oprócz mocy odbieranej, zależy w dużym stopniu na obciążeniu ruchem odczuwanym przez punkt dostępowy. Ta kwestia jest szczególnie widoczna w dolnym łączy dwupoziomowego HetNet złożonego z makro i mikro-AP ze znacznie różnymi możliwościami mocy nadawania. Ze względu na dużą moc nadawania różnica, maksymalna reguła asocjacyjna SIR spowoduje, że większość UE będzie się łączyć makro-poziom z AP dla mikropierścierów obsługujących tylko te nieliczne UE, które się znajdują w pobliżu. Przykład wynikowych regionów asocjacyjnych pokazano na rysunku 5.4 (linie ciągłe).





Oczywiście korzystne byłoby bardziej równomierne rozłożenie obciążenia między poziomami, nawet jeśli spowoduje to, że niektóre UE będą miały mniejszy SIR niż to możliwe. Jednak decyzje o stowarzyszeniu AP-UE, które uwzględniają także równowagę ruchu, mogą być podejmowane tylko przez stronę sieci, ponieważ informacje o obciążeniu ruchem są nieznane dla UE (i zwykle nie są dostarczane przez sieć). Z drugiej strony, ogólnie pożądanym jest wdrażanie decyzji o stowarzyszeniach po stronie UE (przynajmniej dla początkowa procedura przyłączenia do sieci) ze względu na złożoność i skalowanie. W tym celu w literaturze zaproponowano rozszerzenie maksymalnego chwilowego powiązania mocy otrzymanej, gdzie ciężary odpowiadające każdemu AP należącemu do określonego poziomu są mnożone przez zależny od poziomu współczynnik skalowania (odchylenie), tj.  $w_x = P_x = P_i$  jest zastąpione przez  $B_i P_i$  dla AP umiejscowionego w  $x \in \Phi_{a,i}$ ; gdzie  $B_i \geq 1$  jest ustalonym wspólnym odchyleniem dla ciężarów AP należących do poziomu  $i$  i zwiększając odchylenie odpowiadające poziomom z mniejszym. Zwiększając odchylenie odpowiadające poziomom o mniejszej mocy, regiony asocjacyjne między poziomami stają się bardziej zrównoważone, jak pokazano na przykładzie z rysunku 5.4 (linie przerywane), uzyskując w ten sposób bardziej równomierną dystrybucję UE między poziomami.

## 5.4 Analiza wydajności SIR

W tej sekcji opisano wydajność systemu w.r.t. do SIR i powiązać spektralne wskaźniki efektywności. Rozważono różne scenariusze operacyjne, które różnią się pod względem obciążenia UE, kierunku nadawania (łączy w górę lub w dół) i kryterium asocjacji AP. Derywacje są przedstawione dość dużą ilością szczegółów, aby wyraźnie pokazać, w jaki sposób narzędzia analityczne przedstawione w punkcie 5.2 mogą być stosowane do analizy systemów komórkowych. Szczególny nacisk położono na to, w jaki sposób wydajność systemu zależy od gęstości punktów dostępowych i UE oraz czy (i jak) zagęszczenie infrastruktury może być korzystne.

### 5.4.1 W pełni obciążona sieć: połączenie w dół

#### 5.4.1.1 Stowarzyszenie AP z pojedynczą warstwą

Sieć jednopoziomową rozpatruje się w oparciu o takie same założenia dotyczące modelu systemu, jak te opisane w pkt 5.3. Przyjmuje się najbliższe powiązanie AP, przy czym wszystkie AP transmitują w każdej chwili ze stałą mocą  $P$  (niezależnie od pozycji AP). To ostatnie przypuszczenie sugeruje, że każdy AP jest powiązany z co najmniej jednym UE, co oznacza scenariusz wdrożenia infrastruktury, gdzie  $\lambda_a \ll \lambda_u$ , określaną dalej jako w pełni obciążona sieć. Interesuje nas uzyskanie prawdopodobieństwa zasięgu SIR dla typowego UE umieszczonego na początku i powiązanego z jego najbliższym AP umiejscowionym przy  $x^* \in \Phi_a$ . Uwarunkowane na pozycji  $x$  i oznaczane przez

$$I \triangleq \sum_{x \in \Phi_a \setminus \{x^*\}} P g_x |x|^{-\alpha}$$

a moc interferencji, doświadczana dystrybucja SIR przez typowe UE jest równa:

$$\begin{aligned} \mathbb{P}(\text{SIR} \geq \theta | x^*) &= \mathbb{P}\left(\frac{P g_{x^*} |x^*|^{-\alpha}}{I} \geq \theta \mid x^*\right) \\ &= \mathbb{E}(\mathbb{P}(g_{x^*} \geq \theta |x^*|^\alpha I/P \mid x^*, I)) \\ &\stackrel{(a)}{=} \mathbb{E}\left(e^{-\theta |x^*|^\alpha I/P} \mid x^*\right) \\ &= \mathbb{E}\left(\prod_{x \in \Phi_a \setminus \{x^*\}} e^{-\theta |x^*|^\alpha g_x |x|^{-\alpha}} \mid x^*\right) \\ &\stackrel{(b)}{=} \mathbb{E}\left(\prod_{x \in \Phi_a \setminus B(0, |x^*|)} e^{-\theta |x^*|^\alpha g_x |x|^{-\alpha}} \mid x^*\right) \\ &\stackrel{(c)}{=} \exp\left(-\lambda_a \int_{\mathbb{R}^2 \setminus B(0, |x^*|)} \left[1 - \frac{1}{1 + \theta |x^*|^\alpha |x|^{-\alpha}}\right] dx\right), \end{aligned}$$

dla dowolnego  $\theta > 0$ , gdzie (a) jest spowodowane rozkładem  $g_x$  z wykładniczą średnią jednostkową, (b) następuje przez odnotowanie, że  $\Phi_a \setminus \{x^*\} = \Phi_a \setminus B(0, |x^*|)$ , gdzie  $B(x_0, R) \subset \mathbb{R}^2$  oznacza kulę wyśrodkowaną przy  $x_0, R) \in \mathbb{R}^2$  o promieniu  $R > 0$  (strefa ochronna dla transmisji w dół), oraz (c) jest bezpośrednim zastosowaniem formuły PGFL, gdzie znaki są równe  $m_x = \theta |x^*|^\alpha$ ,  $g_x$ ,  $x \in \Phi_a$ . Przekształcenie ostatniej całki we współrzędne biegunowe i po jakiejś algebrze można wykazać, że:

$$\mathbb{P}(\text{SIR} \geq \theta | x^*) = \exp(-\pi \lambda_a |x^*|^2 \rho(\theta, \alpha)),$$

gdzie  $\rho(\theta, \alpha) \triangleq \theta^{2/\alpha} \int_{\theta^{-2/\alpha}}^{\infty} (du / (1 + u^{\alpha/2})) = (2\theta / (\alpha - 2)) {}_2F_1(1; 1 - (2/\alpha); 2 - (2/\alpha); -\theta)$ .

, a  ${}_2F_1(\cdot)$  to funkcja hipergeometryczna Gaussa. To względnie proste wyrażenie weryfikuje intuicję, że SIR podany  $x^*$  (stochastycznie) ulega degradacji wraz ze wzrostem gęstości AP i / lub zwiększeniem odległości od  $x^*$ : Jednak warunkowy rozkład SIR ma niewielką wartość praktyczną dla scharakteryzowania wydajności systemu. W celu uzyskania (bezwarunkowego) prawdopodobieństwa zasięgu typowego UE, wyrażenie musi zostać uśrednione dla rozkładu  $|x^*|$ . Ten ostatni można uzyskać, zauważając, że:

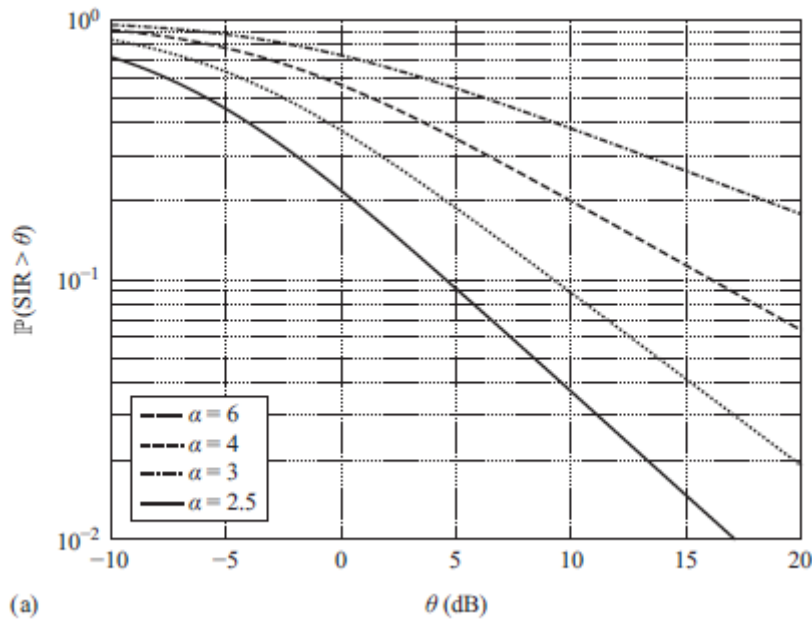
$$\begin{aligned} \mathbb{P}(|x^*| > r) &= \mathbb{P}(N(\mathcal{B}(0, r)) = 0) \\ &= e^{-\pi \lambda_a r^2}, \quad 0 \leq r < \infty, \end{aligned}$$

gdzie  $N(\mathcal{A})$  to liczba punktów  $\Phi_a$  zawartych w  $\mathcal{A} \subset \mathbb{R}^2$  a druga równość wynika z właściwości 1 PPP. To wyrażenie weryfikuje intuicję, że gęstsze rozmieszczenie AP skutkuje mniejszą odległością od powiązanego punktu AP (w sensie statystycznym). Prawdopodobieństwo zasięgu można teraz uzyskać jako prostej formuły:

$$\mathbb{P}(\text{SIR} \geq \theta) = \frac{1}{1 + \rho(\theta, \alpha)}.$$

Najważniejszym aspektem jest sparametryzowanie prawdopodobieństwa zasięgu tylko przez wykładnik strat ścieżki  $\alpha$  i nie zależy od gęstości punktów AP. Ten ostatni wynik jest dość nieoczekiwany, ponieważ zarówno moc odbierana z powiązanego punktu AP, jak i moc interferencji, zależą od  $l_a$ . Nieformalnie można stwierdzić, że zmiana wartości  $l_a$  wpływa również na licznik i mianownik SIR, co powoduje stałe prawdopodobieństwo pokrycia. Ta niezależność od  $l_a$  jest bardzo ważnym wynikiem, ponieważ sugeruje ona, że w rozważanym modelu systemu, arbitralnie duże przestrzenne poziomy ponownego wykorzystania mogą być osiągnięte przez zagęszczenie infrastruktury bez żadnego wpływu na postrzegany SIR, stanowiąc tym samym silny argument na korzyść (ultra) gęste sieci.

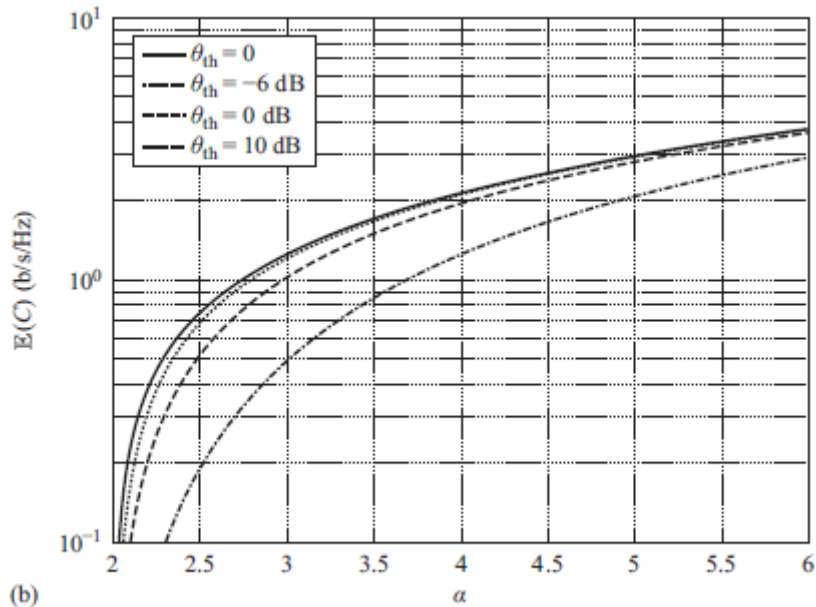
Rysunek 5.6 (a) pokazuje prawdopodobieństwo pokrycia dla różnych wartości  $\alpha$ .



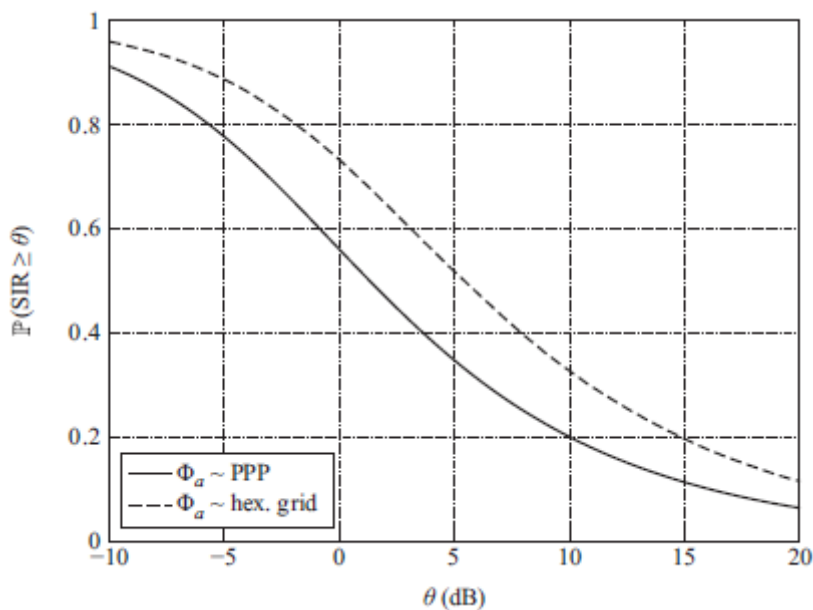
Można zauważyć, że działanie przy dużym wykładniku stratności ścieżki jest korzystne, ponieważ w tym przypadku interferencja jest generowana zasadniczo tylko z ograniczonej liczby pobliskich AP ze względu na zwiększoną utratę ścieżki. Średnia sprawność spektralna łącza  $\mathbb{E}(C)$  pod różnymi progami SIR  $\theta$ -tego, otrzymana numerycznie jako

$$\mathbb{E}(C) = \int_0^\infty \mathbb{P}(C > c) dc$$

używając wyrażenia rozkładu pokazano na rysunku 5.6 (b).



Zgodnie z trendem prawdopodobieństwa zasięgu,  $\mathbb{E}(C)$  wzrasta z  $\alpha$ . Zgodnie z oczekiwaniami, większe wartości  $\theta$ -tego przekładają się na mniejszą średnią wydajność widmową. Aby zrozumieć konsekwencje losowego wdrożenia infrastruktury w porównaniu z tradycyjnym okresowym (sześciokątnym) wzorcem wdrażania, prawdopodobieństwo zasięgu SIR w modelu PPP i sześciokątny wzór dla  $\Phi_a$  są porównywane na rysunku 5.7.



Wydajność układu heksagonalnego uzyskuje się za pomocą symulacji Monte Carlo, przy typowym UE równomiernie rozłożonym w komórce losowej i we wszystkich innych założeniach modelu systemu, tak samo jak w modelu PPP. Dla sześciokątnego wzoru AP zastosowano gęstość 1, chociaż symulacje sugerują, że tak jak w przypadku PPP, prawdopodobieństwo zasięgu jest niezależne od gęstości AP. Widać, że sześciokątny wzór zapewnia lepsze pokrycie przez margines 4-5 dB. Jest to oczekiwane, ponieważ okresowy wzór AP ogranicza liczbę elementów zakłócających, które są bliskie typowemu UE, podczas gdy w modelu PPP może znajdować się wiele obiektów zakłócających w bliskim sąsiedztwie.

Ponieważ regularne / okresowe wdrażanie gęstej infrastruktury jest raczej niemożliwe, taka utrata wydajności jest kosztem, z którym muszą sobie poradzić projektanci systemów.

#### 5.4.1.2 Wiele poziomów - maksymalne średnie skojarzenie mocy otrzymanej

Analizę poprzedniej sekcji można bezpośrednio rozszerzyć na przypadek sieci HetNet składającej się z  $K \geq 2$  poziomów, w których obecnie występuje termin "w pełni obciążona sieć" odpowiada warunkowi  $\lambda_u \gg \sum_i \lambda_{a,i}$ ; Wszystkie AP warstwy T transmitują ze stałą mocą  $P_k > 0$ , podczas gdy moc sygnału pilota równa się  $B_k P_k$ ; gdzie  $B_k \geq 1$  oznacza odchylenie mocy pilota od poziomu k; dobierane odpowiednio do celów wyładunku, jak omówiono w punkcie 5.3.3. UE jest powiązane z AP z maksymalną obciążoną średnią otrzymaną mocą pilota. Przy dokładnie takich samych założeniach dotyczących modelu systemu i przy zastosowaniu podobnej metodologii, jak w przypadku pojedynczej warstwy, prawdopodobieństwo pokrycia typowego UE może być równe :

$$\mathbb{P}(\text{SIR} \geq \theta) = \sum_{k=1}^K \left( \sum_{l=1}^K \frac{\lambda_{a,l}}{\lambda_{a,k}} \left( \frac{P_l B_l}{P_k B_k} \right)^{2/\alpha} (1 + \rho(\theta B_k / B_l, \alpha)) \right)^{-1} ;$$

z  $\rho(\cdot, \cdot)$ . Oczywiście pokrywają się dla  $K = 1$ :

Co ciekawe, ustawiając  $B_k = B_l$  dla wszystkich  $k, l$ ), daje takie samo wyrażenie dla prawdopodobieństwa pokrycia, tj. gdy nie wprowadzono odchylenia w procedurze asocjacji AP-UE, wydajność SIR w HetNet jest taka sama jak pojedynczej sieć warstwowa. W szczególności wydajność HetNet staje się niezależna od liczby, gęstości i mocy nadawania poziomów, co sugeruje, że ponowne wykorzystanie przestrzenne może być dowolnie zwiększone przez wprowadzenie do sieciowych punktów dostępowych należących do istniejących lub nowych poziomów, zasadniczo w sposób losowy. Należy zauważyć, że przy nierównomiernych odchyleniach, prawdopodobieństwo zasięgu SIR zmniejsza się z przypadku równych błędów, ponieważ niektóre UE będą kojarzone z punktami dostępowymi, które nie zapewniają maksymalnej mocy odebranej dla transmisji danych. Obserwacji tej nie należy jednak uważać za silny argument przeciwko wprowadzaniu uprzedzeń, ponieważ degradacja SIR może być mniejsza w porównaniu z korzyściami, jakie daje bardziej wyrównane stowarzyszenie UE między poziomami.

#### 5.4.1.3 Maksymalne skojarzenie SIR

Wydajność SIR sieci pobierającej w ramach maksymalnego powiązania SIR UE-AP jest rozpatrywana w niniejszym podrozdziale. Dokładne wyrażenia analityczne dla prawdopodobieństwa zasięgu uzyskano w literaturze, są one jednak nadmiernie skomplikowane i nie będą tutaj prezentowane. Zamiast tego, jest bardziej pouczające, zarówno z technicznego punktu widzenia, jak i w celu uzyskania pewnych spostrzeżeń, do rozważenia górnej granicy prawdopodobieństwa zasięgu najpierw uzyskanego w. Biorąc pod uwagę najpierw sieć z warstwą pojedynczą i oznaczającą jako SIR(x) SIR z typowe UE zlokalizowane w punkcie początkowym, gdy jest powiązane z AP przy  $x \in \Phi_a$ , prawdopodobieństwo zasięgu maksymalnego schematu asocjacji SIR może być ograniczone jako

$$\begin{aligned}
\mathbb{P}\left(\max_{x \in \Phi} \text{SIR}(x) \geq \theta\right) &= \mathbb{P}\left(\bigcup_{x \in \Phi} \{\text{SIR}(x) \geq \theta\}\right) \\
&\leq \sum_{x \in \Phi} \mathbb{P}(\text{SIR}(x) \geq \theta) \\
&= \mathbb{E}\left(\sum_{x \in \Phi} \mathbb{I}(\text{SIR}(x) \geq \theta)\right) \\
&= \lambda_a \int_{\mathbb{R}^2} \mathbb{P}(\text{SIR}(x) \geq \theta) dx,
\end{aligned}$$

gdzie ostatnia równość została uzyskana dzięki zastosowaniu twierdzenia Campbella. Uwaga po prawej stronie, że ta górna granica ma fizyczne znaczenie, mianowicie jest równa średniej liczbie punktów dostępowych, które są w stanie zapewnić typowe UE i SIR większe niż  $\theta$ . Stosując podobną procedurę jak przy uzyskiwaniu, prawdopodobieństwo pokrycia przez punkt AP ustawione na  $x \in \mathbb{R}^2$  jest równe

$$\mathbb{P}(\text{SIR}(x) \geq \theta) = \exp\left(-\pi \lambda_a |x|^2 \kappa(\alpha) \theta^{2/\alpha}\right)$$

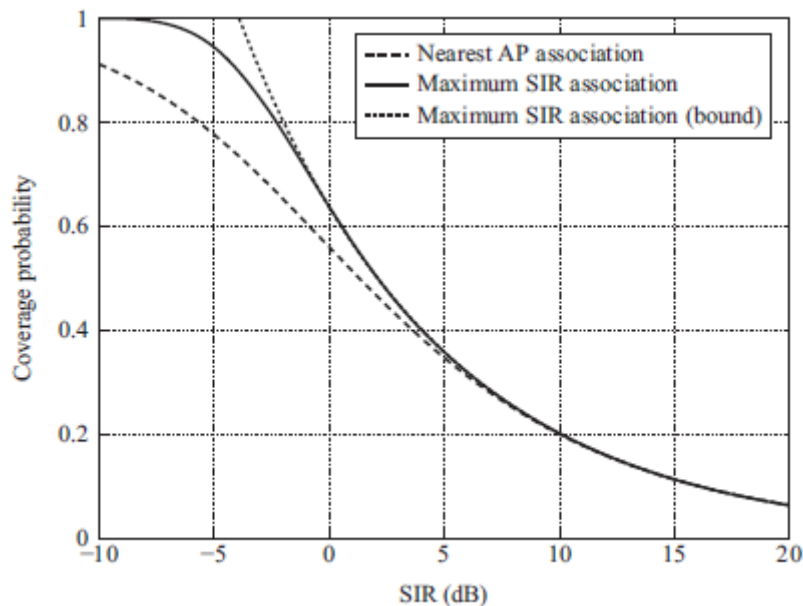
gdzie

$$\kappa(\alpha) \triangleq (2\pi/\alpha) / \sin(2\pi/\alpha)$$

Zastępując w powyższym równaniu, skutkuje

$$\mathbb{P}\left(\max_{x \in \Phi} \text{SIR}(x) \geq \theta\right) \leq \frac{1}{\kappa(\alpha) \theta^{2/\alpha}}.$$

Należy zauważyć, że to proste powiązanie jest niezależne od gęstości punktów dostępowych. Można wykazać, że ta niezależność gęstości jest prawdziwa również dla dokładnego prawdopodobieństwa zasięgu maksymalnego schematu asocjacji SIR, podobnie jak w przypadku prawdopodobieństwa zasięgu z najbliższym powiązaniem AP. Dokładne prawdopodobieństwo zasięgu z maksymalnym powiązaniem SIR przedstawiono na Rysunku 5.8 (uzyskanym przy pomocy symulacji Monte Carlo)



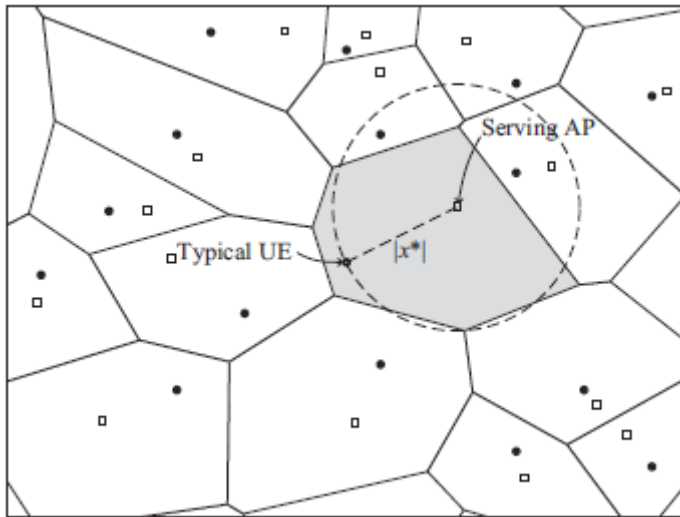
wraz z górną granicą i prawdopodobieństwem zasięgu najbliższej reguły asocjacji AP określonej przez). Zgodnie z oczekiwaniami, maksymalny schemat asocjacji SIR przewyższa najbliższe powiązanie AP, szczególnie w przypadku niskich SIR, gdzie punkt dostępowy, który zapewnia największy SIR, nie zawsze jest najbliższy. W przypadku dużych SIR (większych niż 5 dB) oba systemy mają taką samą wydajność jak AP maksymalizujący SIR zbiega się w większości przypadków z najbliższym. Należy zauważyć, że górna granica zapewnia doskonałą wskazanie dokładnego prawdopodobieństwa zakrycia dla wartości SIR większych niż około  $<2$  dB. W rzeczywistości górne ograniczenie jest rzeczywiście równe prawdopodobieństwu pokrycia dla wartości SIR większych lub równych 0 dB. Wynika to z obserwacji, że jeśli istnieje punkt dostępowy, który może zapewnić SIR większy lub równy 0 dB, to jest unikalny, renderując zdarzenia  $\{SIR(x) \geq \theta\}_{x \in \Phi}$  wzajemnie się wykluczają dla każdego  $\theta \geq 0$  dB. Analiza sieci jednowarstwowej może zostać bezpośrednio rozszerzona na przypadek wielowarstwowy. Okazuje się, że górna granica prawdopodobieństwa zasięgu w tym przypadku jest równa wyrażeniu w przypadku pojedynczej warstwy, tj. W ramach maksymalnego schematu asocjacji SIR, wydajność sieci pod względem SIR jest niezmienna do liczby warstw i poziomu transmisji mocy, podobne do obserwacji pod maksymalnym średnim skojarzeniem mocy otrzymanej z równymi uprzedzeniami.

#### 5.4.2 W pełni obciążona sieć: łącze w górę

W przeciwieństwie do mnóstwa prac w ramach geometrii stochastycznej, które badają wydajność łącza w dół sieci komórkowej, wydajność łącza w górę jest znacznie mniej zwracana, chociaż może się wydawać od pierwszego wejrzenia jako "podwójne" badanie dla łącza w dół. dochodzenie. Jest to spowodowane nie tylko tym, że łącze odbiorcze jest uważane za bardziej krytyczne dla wydajności, ale również z powodu trudności analitycznych związanych z konfiguracją łącza wstępującego. Trudności te można zrozumieć, biorąc pod uwagę konfigurację równoległą do rozważanej w sekcji 5.4.1.1, w której urządzenia UE łącza zwrotnego kojarzą się z ich najbliższym AP i transmitują z tą samą mocą stałą. Próba wykonania tych samych kroków, co w analizie połączenia, napotyka na krytyczne wyzwania wynikające z następujących faktów.

1. Pozycje zakłócających UE nie są dystrybuowane jako PPP.
2. "Strefa wartownicza", w obrębie której powiązany AP jest wolny od zakłócających UE, to jego komórka (Voronoi), która nie ma prostej reprezentacji matematycznej.

Pierwszy fakt można zrozumieć, zauważając, że lokalizacja zakłócających UE jest procesem punktowym generowanym z  $\Phi_u$  przez rzucenie wszystkich punktów należących do tej samej komórki AP, z wyjątkiem losowo wybranej. Jednak komórki AP nie są niezależne i dlatego zdarzenia dwóch punktów  $F_u$  zlokalizowanych w dwóch różnych (i rozłącznych) komórkach nie są niezależne. Jest to oczywiście sprzeczne z podstawowymi właściwościami PPP. Drugi fakt jest oczywisty po prostej kontroli realizacji rozważanego ustawienia (jak pokazano na rysunku).



Należy zauważyć, że w przeciwieństwie do przypadku łącza pobierającego, geometryczna strefa ochronna łącza w górę nie gwarantuje, że wszystkie zakłócenia są w większej odległości od określonego AP niż powiązane z nim UE. Chociaż proces punktowy zakłócających UE może być w zasadzie przybliżony jako PPP o gęstości równej  $\lambda$ ; nieregularność strefy ochronnej nie pozwala na dalszy postęp analityczny. Aby uzyskać analityczną charakterystykę wydajności łącza w górę, zaproponowano subtelną, ale krytyczną modyfikację modelu systemu. W szczególności, ustalone moc mocy transmisji jest obniżane na rzecz schematu sterowania mocą dla każdego łącza, gdzie moc nadawania każdego UE łącza zwrotnego jest taka, że średnia moc odebrana w powiązonym AP jest równa pewnej uprzednio zdefiniowanej wartości  $P^* > 0$ ; tj. dla odległości łącza  $r > 0$ ; UE łącza zwrotnego transmituje z mocą energii  $P^* r^\alpha$ . Uzasadnieniem tego podejścia jest to, że gwarantuje to, że średnia moc odebrana w punkcie AP wygenerowanym z powiązanego urządzenia użytkownika UE jest większa niż moc interferencji generowana z dowolnego innego UE, nawet w przypadku, gdy ta ostatnia ma miejsce być bliżej geometrycznie (sprawdź to twierdzenie). Innymi słowami, kontrola mocy łącza w górę generuje strefę ochronną w domenie mocy, która różni się od naturalnej geometrycznej strefy ochronnej łącza zwrotnego. Biorąc pod uwagę łącze wysyłające typowego UE, z początkiem płaszczyzny przesuniętej do jej obsługującego położenia AP, SIR połączenia wysyłającego jest równe:

$$SIR = \frac{P^* g_x}{\sum_{x \in \tilde{\Phi}_u} \mathbb{I}(P_x |x|^{-\alpha} < P^*) P_x g_x |x|^{-\alpha}}$$

gdzie  $\tilde{\Phi}_u \subset \Phi_u$  jest punktowym procesem zakłócania UE,  $g_x$  jest zanikającym wzmocnieniem

typowe łącze  $\{g_x\}_{x \in \tilde{\Phi}_u}$  to zakłócające kanały zyski, a  $\{P_x\}_{x \in \tilde{\Phi}_u}$  są mocami nadawania zakłócających UE,  $g_x$  jest zanikającym wzmocnieniem typowego łącza. Wykorzystanie funkcji wskaźnika w mianowniku formuły SIR, chociaż jest zbędne, ponieważ trywialnie ocenia jedność dla wszystkich  $x \in \tilde{\Phi}_u$ , jest wygodny dla następujących obliczeń, ponieważ wyraźnie reprezentuje skuteczną strefę



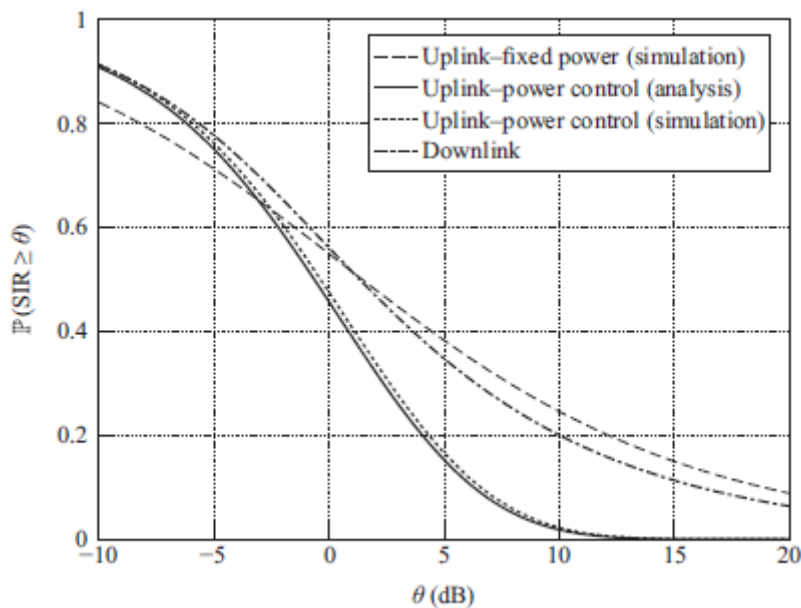
ochronną obsługującego AP. Zgodnie z założeniami (a)  $\tilde{\Phi}_u$  rozprawdzanego jako PPP o gęstości  $\lambda_a$  i (b)  $\{P_x\}_{x \in \tilde{\Phi}_u}$  są równe i.i.d. zmienne losowe, wyprowadzenie prawdopodobieństwa zasięgu łącza w górę przebiega podobnie do przypadku łącza pobierającego, prowadząc do:

$$\begin{aligned} \mathbb{P}(\text{SIR} \geq \theta) &= \mathbb{E} \left( \prod_{x \in \tilde{\Phi}_u} e^{-\frac{\theta}{P^*} \mathbb{I}(P_x |x|^{-\alpha} < P^*) P_x g_x |x|^{-\alpha}} \right) \\ &\stackrel{(a)}{=} \exp \left( -2\pi\lambda_a \mathbb{E} \left( \int_{(P_x/P^*)^{-1/\alpha}}^{\infty} \frac{\theta r^{-\alpha} P_x/P^*}{1 + \theta r^{-\alpha} P_x/P^*} r dr \right) \right) \\ &\stackrel{(b)}{=} \exp \left( \pi\lambda_a \mathbb{E} \left( \left( \frac{P_x}{P^*} \right)^{2/\alpha} \right) \rho(\theta, \alpha) \right), \end{aligned}$$

gdzie (a) jest bezpośrednim zastosowaniem formuły PGFL, po której następuje oczekiwanie ponad  $g_x$ ; przesunięcie oczekiwania ponad  $P_x$  z całki, przełączenie na współrzędne biegunowe i uwzględnienie specyfikacji strefy ochronnej i (b) następuje poprzez zmianę zmiennej całkującej na  $u = r^2(\theta P/P^*)^{-2/\alpha}$  Ilość  $\rho(\cdot, \cdot)$  została zdefiniowana. Łatwo pokazać, że  $\mathbb{E}((P_x/P^*)^{2/\alpha}) = (1/\pi\lambda_a)$ , co ostatecznie prowadzi do:

$$\mathbb{P}(\text{SIR} \geq \theta) = e^{-\rho(\theta, \alpha)}.$$

To proste wyrażenie pokazuje, że wydajność SIR w łączu w górę z najbliższym AP stowarzyszenie jest niezależne od  $P$  i gęstości infrastruktury  $\lambda_a$ , dokładnie tak, jak było . Sprawa wydajności łącza w dół. Porównanie z wyrażeniem SIR łącza w dół pokazuje, że SIR w łączu w dół stochastycznie dominuje w SIR łącza zwrotnego z kontrolą mocy. Wyrażenia te są porównywane na rysunku 5.10 wraz z symulacją Monte Carlo zarówno dla łącza stałego, jak i kontrolowanego pod względem schematów mocy



Można zauważyć, że symulacja dla kontrolowanego pod kątem łącza nadawczego pasuje do analitycznej ekspresji (5.17), co sugeruje, że przybliżenia zastosowane w jej wyprowadzeniu są dokładne. Porównanie z wydajnością łącza w dół pokazuje, że łącza w górę i łącza w dół mają podobne

tendencje, chociaż schemat kontroli łączy w górę jest znacznie gorszy przy wysokich SIR-ach. W przeciwieństwie do tego, schemat połączenia ładującego o stałej mocy przypomina bliższą wydajność łączy nadawczego pokazującą degradację przy niskich SIR z powodu obecności elementów zakłócających będących bliżej punktu AP niż powiązanych z nim UE.

### 5.4.3 Lekko obciążona sieć

Poprzednie sekcje uważały środowisko operacyjne, w którym znajduje się każdy punkt dostępowy jest związany z co najmniej jednym UE, skutecznie przyjmując  $\lambda_u \gg \sum_i \lambda_{a,i}$  z rzeczywistą wartością  $\lambda_u$  nieistotną. Jednak w praktycznych warunkach operacyjnych niektóre z AP będą powiązane z brakiem UE. Oznacza to, że liczba osób zakłócających jest faktycznie mniejsza niż liczba rozważana w poprzednich sekcjach. Równoważnie, gęstość procesu punktowego reprezentującego pozycje zakłócające jest mniejsza niż  $\lambda_a$  w przypadku jednopoziomowego łączy w dół i podobnie w innych przypadkach. Oczywiście jest, że wartość  $\lambda_u$  odgrywa tutaj ważną rolę i interesujące jest zbadanie, w jaki sposób wpływa ona na właściwości pozycji elementów zakłócających, a także na wydajność zasięgu SIR sieci. W tej sekcji rozważane jest to pytanie w przypadku sieci jednopoziomowej sieci łączy w dół z analizą sieci wielowarstwowych i / lub łączy zwrotnych zgodnie z tymi samymi zasadami.

#### 5.4.3.1 Prawdopodobieństwo aktywności AP

Naturalnym punktem wyjścia do tego badania jest określenie statystyk z  $K \geq 0$ ; reprezentujący liczbę UE powiązanych z losowym AP. Należy zauważyć, że  $K = K(\Phi_u \Phi_a, \{w_x\}_{x \in \Phi_a})$ , tj. liczba UE powiązanych z losowym AP zależy od pozycji AP i UE, a także od (losowych) wag używanych przez schemat asocjacji. Niestety wyrażenie dla  $\mathbb{P}(K = k, k = 0, 1, \dots)$  nie jest dostępne, nawet dla najprostszego, najbliższego schematu asocjacji AP ( $\{w_x\}_{x \in \Phi_a}$ ). Jednakże można uzyskać opis pierwszych momentów rozkładu  $K$ , które dają pewien wgląd w to, jak  $K$  zależy od gęstości AP i UE. W szczególności, w łagodnym stanie, w którym  $\mathbb{E}(w_x^{2/a})$  jest skończony, pokazano, że:

$$\mathbb{E}(K) = \frac{\lambda_u}{\lambda_a},$$

co pokazuje, że średnia  $K$  zależy tylko od stosunku gęstości UE i AP, a nie od ich rzeczywistych wartości. Należy zauważyć, że wynik ten dotyczy każdego schematu asocjacji z ważoną ścieżką-utrata, w tym schematów opisanych w sekcji 5.3.3. Interesujące jest to, że  $\mathbb{E}(K)$  jest taki sam z  $w_x = 1$  (najbliższe skojarzenie AP),  $w_x = p_{g_x}$  (maksymalne chwilowe otrzymane wzmocnienie mocy), lub, o ile to możliwe, z dowolną inną losową wagą asocjacyjną  $w_x$ . Wyrażenie powyższe sugeruje, że zwiększenie  $\lambda_u$  nad  $\lambda_a$  zwiększa prawdopodobieństwo AP powiązanego z co najmniej jednym UE, co jest oczywiście intuicyjne. Zostało to dokładnie wykazane, gdzie prawdopodobieństwo aktywności AP,  $\mathbb{P}(K > 0) = 1 - \mathbb{P}(K = 0)$  było ograniczone jako

$$\frac{\lambda_u}{\lambda_u + \lambda_a} < \mathbb{P}(K > 0) \leq 1 - e^{-\lambda_u/\lambda_a}.$$

Dolna granica jest nieosiągalna, natomiast górna granica może zostać osiągnięta poprzez odpowiedni dobór rozkładu masy asocjacji. Przybliżone wyrażenie prawdopodobieństwa aktywności AP zaproponowano jako

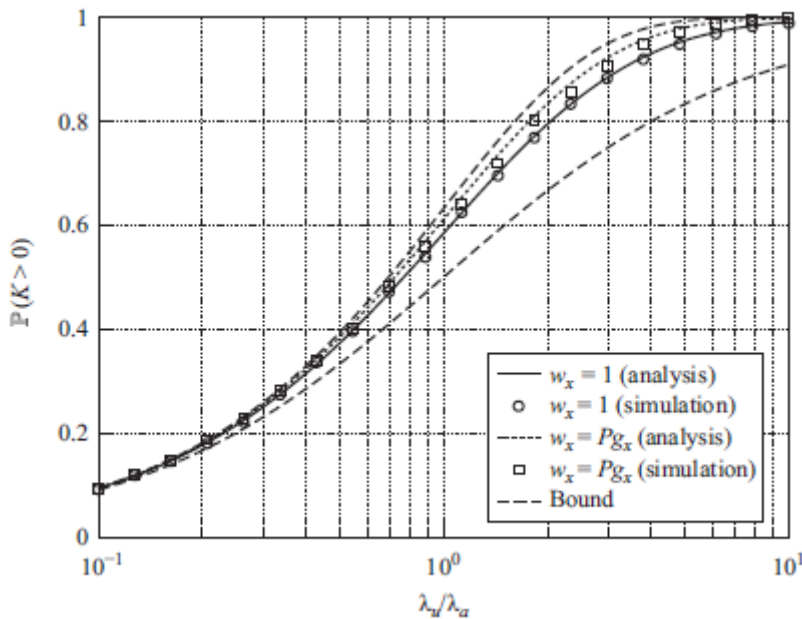
$$\mathbb{P}(K > 0) \approx 1 - \left(1 - \frac{\lambda_u}{\xi \lambda_a}\right)^{-\xi},$$

gdzie

$$\xi \triangleq (7/2) \mathbb{E}(w_x^{2/\alpha}) \mathbb{E}(w_x^{-2/\alpha})$$

Zauważ, że dwa równania implikują, że  $\mathbb{P}(K > 0) \approx \mathbb{E}(K)$  kiedy  $\lambda_u \ll \lambda_a$ .

Rysunek 5.11



pokazuje prawdopodobieństwo aktywności AP dla najbliższego skojarzenia AP ( $w_x = 1$ ) i maksymalne chwilowe otrzymane powiązanie mocy ( $w_x = P_{g_x}$ ) uzyskane przez symulację Monte Carlo wraz z ich przybliżonymi wyrażeniami) w funkcji średniego obciążenia UE losowej komórki,  $\lambda_u = \lambda_a$ . Wyznaczono również granice. Można zauważyć, że oba schematy asocjacyjne dają podobne, choć odmienne, prawdopodobieństwa działania, z przybliżonym wyrażeniem zapewniającym bardzo dobre dopasowanie w obu przypadkach, szczególnie dla najbliższego schematu stowarzyszenia AP. Należy zauważyć, że w środowisku operacyjnym z  $\lambda_a = \lambda_u$ , tj. w sieci ultra gęstej, prawdopodobieństwo aktywności wynosi tylko około 0,6, tj. Około 40% punktów AP nie transmituje, co jest znaczącym odchyleniem od przyjętego założenia sieci z pełnym obciążeniem w sekcji 5.4.1.1. Obserwacja ta zdecydowanie sugeruje, że wydajność systemu w ramach dużej gęstości infrastruktury powinna zostać odpowiednio oceniona.

### 5.4.3.2 Prawdopodobieństwo pokrycia

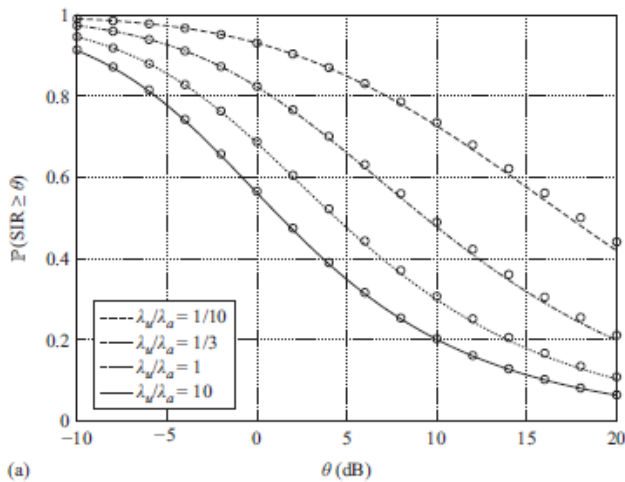
Po uzyskaniu (przybliżonego) wyrażenia dla prawdopodobieństwa aktywności AP, Problem oceny wydajności systemu można teraz zaatakować. Niech  $\bar{\Phi}_a \subseteq \Phi_a$  oznaczają położenia transmitujących punktów AP, tj. punktów AP powiązanych z co najmniej jednym UE. SIR doświadczony przez typowe UE znajdujące się w miejscu pochodzenia jest taki sam, z sumą w mianowniku wykonaną dla  $\bar{\Phi}_a \setminus \{x^*\}$  zamiast  $\Phi_a \setminus \{x^*\}$ . Aby określić rozkład SIR, statystyki  $\bar{\Phi}_a$  a muszą być określone, który jednak jest

trudnym zadaniem ze względu na korelację między zdarzeniami dwóch AP związanych z (co najmniej) jednym UE. Aby pokonać tę trudność, naturalnym podejściem jest założenie, że wydarzenia stowarzyszeniowe są niezależne i rozważmy  $\tilde{\Phi}_a$  jako punktowy proces wynikający z przerzedzania  $\Phi_a$  mianowicie przez niezależnie usuwa każdy punkt  $x \in \Phi_a$  z prawdopodobieństwem  $\mathbb{P}(K = 0)$ . Z właściwości 5 PPP (patrz sekcja 5.2) wynika, że  $\tilde{\Phi}_a$  jest PPP gęstością  $\mathbb{P}(K > 0)\lambda_a$ .

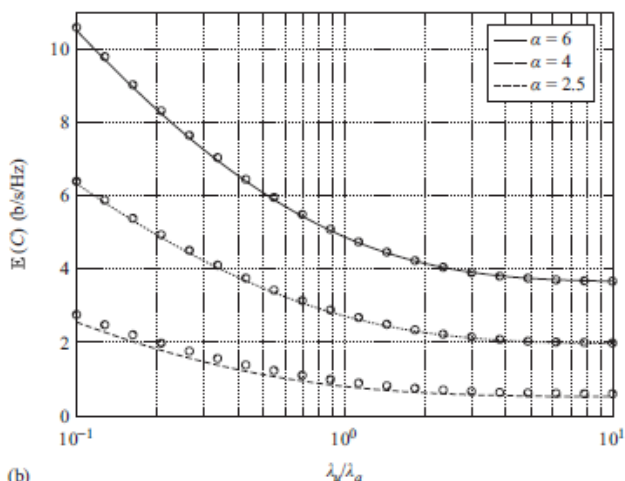
W ramach tego przybliżenia statystyk  $\tilde{\Phi}_a$ , można tu zastosować dokładnie te same procedury z rozdziału 5.4.1.1, używając  $\mathbb{P}(K > 0)\lambda_a$  zamiast  $\lambda_a$  dla gęstości zakłócających punktów AP. Na przykład, po tych samych krokach co w wyprowadzaniu , prawdopodobieństwo pokrycia sieci komórkowej zstępującej z skończonym  $\lambda_u$  i najbliższym powiązaniem AP może być (w przybliżeniu) równe (:

$$\mathbb{P}(\text{SIR} \geq \theta) = \frac{1}{1 + \mathbb{P}(K > 0)\rho(\theta, \alpha)}.$$

Zastąpienie  $\mathbb{P}(K > 0)$  przybliżonym wyrażeniem pokazuje, że prawdopodobieństwo pokrycia jest malejącą funkcją  $\lambda_u / \lambda_a$  zgodnie z oczekiwaniami, ze względu na zwiększone poziomy interferencji generowane przez większą liczbę aktywnych punktów AP. Stanowi to kolejny argument przemawiający za zagęszczeniem sieci, tj. zwiększenie wartości  $\lambda_a$  dla stałej wartości  $\lambda_u$  nie tylko pozwala na arbitralnie duże poziomy ponownego wykorzystania przestrzeni (jak argumentowano w sekcji 5.4.1.1), ale także z natury zwiększa prawdopodobieństwo pokrycia bez odwoływania się do innych metod, takich jak jako rozszerzenie pasma i zaawansowane techniki transmisji. Rysunek 5.12 (a)



(a)



(b)

przedstawia prawdopodobieństwo zasięgu łącza w dół z najbliższym powiązaniem AP w funkcji  $\lambda_u / \lambda_a$ . Podano zarówno dokładne prawdopodobieństwo (uzyskane za pomocą symulacji Monte Carlo), jak i analityczne przybliżenie. Można zauważyć, że przybliżone wyrażenie zapewnia doskonałe dopasowanie rzeczywistego prawdopodobieństwa dla wszystkich zakresów wartości  $\lambda_u / \lambda_a$ , co sugeruje, że przybliżenie  $\Phi_a$  jako PPP jest dokładne. Jak wspomniano powyżej, wydajność systemu jest korzystna z mniejszych wartości  $\lambda_u = \lambda_a$ , a korzyść jest określana ilościowo na tej figurze, np. Zmniejszenie  $\lambda_u = \lambda_a$  z 10 do 1 zapewnia wzmocnienie SIR o około 3 dB. Rysunek 5.12 (b) pokazuje średnią efektywność widmową łącza  $\mathbb{E}(C)$  uzyskaną numerycznie w oparciu o przybliżone wyrażenie analityczne dla prawdopodobieństwa zasięgu wraz z symulacjami Monte Carlo. Można zauważyć, że dla wartości  $\lambda_u / \lambda_a$  około 5,  $\mathbb{E}(C)$  osiąga plateau, co sugeruje, że dla tych wartości obciążenia UE bardzo niewiele z AP jest nieaktywnych. Oznacza to, że wyniki uzyskane przy założeniu pełnego obciążenia sieci są rzeczywiście dokładne dla  $\lambda_u \approx 5\lambda_a$  lub więcej. Mniejsze wartości obciążenia UE odpowiadają zwiększonej wydajności widmowej, przy czym najbardziej znaczące ulepszenia pokazano w reżimie  $\lambda_u \leq \lambda_a$  (ultra gęstej sieci) i dla większych wartości  $a$ . Zauważ, że symulacje wydają się być lekko stronnice w kierunku większych wartości niż te przewidywane przez analizę. Ta rozbieżność wynika nie tylko z aproksymacji zastosowanych w analizie, ale także z powodu praktycznych ograniczeń konfiguracji symulacji, które muszą koniecznie uwzględniać skończony podzbiór  $\mathbb{R}^2$  (podczas gdy model rozciąga się na całą płaszczyznę). Uwaga: W przypadku nieaktywnych punktów dostępowych pojęcie maksymalnego powiązania SIR nie jest jasne. Wynika to z tego, że poziom zakłóceń odbierany przez pewne UE zależy od aktywnych AP, które z kolei zależą od decyzji o połączeniu innych UE. Oznacza to, że działanie lekko obciążonych sieci przy maksymalnym powiązaniu mocy otrzymanej musi być zaatakowane w inny sposób niż ten rozważany w sekcji 5.4.1.3, w którym wykorzystano równowagę z maksymalnym powiązaniem SIR. Chociaż analizę można zasadniczo przeprowadzić przy użyciu technik omówionych do tej pory, wydaje się, że nie można uzyskać prostych, zamkniętych wyrażeń formularzy, które dostarczyłyby dodatkowych wglądów w numeryczną ocenę wydajności.

## 5.5 Analiza wydajności użytkowników

W tej sekcji przedstawiono wydajność systemu w odniesieniu do stopy UE,  $R$ , zgodnie z definicją w Sekcja 5.3. Omówiono zarówno analizę wydajności, jak i aspekty projektowania systemu. Zostanie omówione tylko najbliższe stowarzyszenie AP. Dla uproszczenia, analizy i wyników dla pojedynczej sieci zostaną przedstawione.

### 5.5.1 Stopnie rozkładu swobody

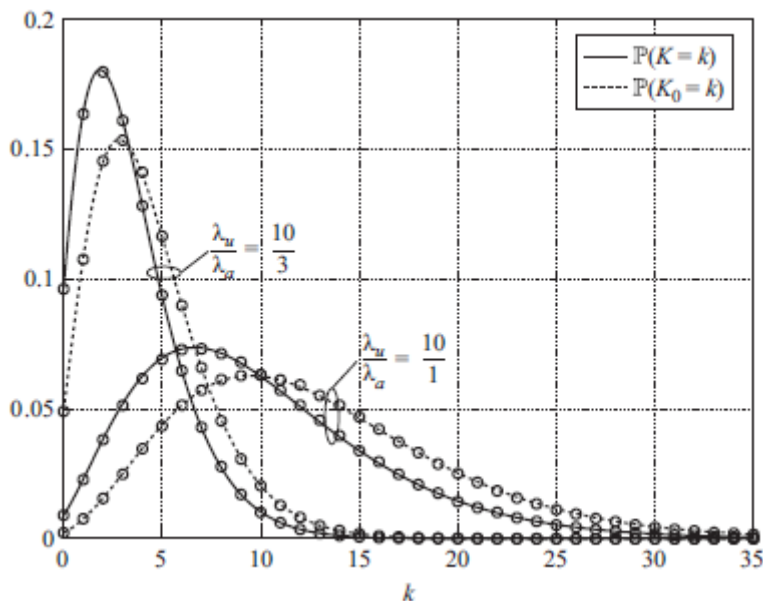
Nowym elementem wprowadzonym przez definicję  $R$  w porównaniu z innymi parametrami wydajności jest ilość  $K_0$ , tj. liczba dodatkowych UE związanych z AP obsługującym typowe UE, które określa stopnie swobody UE,  $1/(1+K_0)$ . Biorąc pod uwagę typowe UE zlokalizowane u źródła,  $K_0$  może być równoważnie postrzegana jako liczba elementów  $\Phi_u$  zawartych w komórce generowanej przez punkty  $\Phi_a$ , które obejmują pochodzenie. Na pierwszy rzut oka może się wydawać, że (częściowa) charakterystyka  $K_0$  została już przeprowadzona w sekcji 5.4.3, gdzie przedstawiono niektóre statystyczne właściwości  $K$ , liczbę UE powiązanych z losowym AP. Jednak  $K_0$  i  $K$  nie są identycznie rozmieszczone, stąd różnica w zapisie. Wynika to z faktu, że właściwości statystyczne komórki zawierającej początek (lub dowolny inny punkt w tej sprawie) różnią się od właściwości komórki losowej (bez dodatkowych wymagań nałożonych, na przykład na uwzględnienie określonych punktów). Jako ilustrację tej różnicy, niech  $|\mathcal{B}_0| > 0$  i  $|\mathcal{B}| > 0$  oznaczają obszary komórki zawierającej początek i dowolną inną komórkę losową. Można pokazać, że  $\mathbb{E}(\mathcal{B}_0) > \mathbb{E}(\mathcal{B})$ , z której wynika, że:

$$\begin{aligned}
\mathbb{E}(K_0) &= \mathbb{E}(\mathbb{E}(K_0|\mathcal{C}_0)) \\
&\stackrel{(a)}{=} \mathbb{E}(\lambda_a|\mathcal{C}_0) \\
&> \lambda_a \mathbb{E}(|\mathcal{C}|) \\
&= \mathbb{E}(K),
\end{aligned}$$

gdzie (a) wynika z Właściwości 1 z  $\Phi_a$ . Dlatego też liczba urządzeń użytkownika powiązanych z komórką, w tym miejsce pochodzenia, jest większa średnio niż liczba urządzeń UE obsługiwanych przez dowolną inną komórkę losową. Ten pozorny "paradoks" może być postrzegany jako przejaw dwuwymiarowego przypadku dobrze znanego "paradoksu oczekiwania", występującego w badaniu jednowymiarowych procesów punktowych Poissona. Podobnie jak w przypadku  $K$ , dokładne wyrażenia w zamkniętej formie dla rozkładu  $K_0$  nie są dostępne, co czyni niezbędne przybliżenia. Dobre przybliżenie rozkładów  $K_0$  i  $K$ , w zależności tylko od stosunku  $\lambda_u/\lambda_a$ , podano w jako

$$\mathbb{P}(\mathcal{K} = k) = \frac{(7/2)^\eta \Gamma(k + \eta) (\lambda_u/\lambda_a)^k}{\Gamma(\eta) k! (\lambda_u/\lambda_a + 7/2)^{k+\eta}}, \quad k = 0, 10, \dots,$$

gdzie  $\eta = 7/2$  dla  $\mathcal{K} = K$  a  $\eta = 9/2$  dla  $\mathcal{K} = K_0$  i  $\Gamma(\cdot)$  oznacza funkcję Gamma. Ocena średniej  $K_0$  daje  $\mathbb{E}(K_0) = (9/7)(\lambda_u/\lambda_a)$  tj. średnia  $K_0$  jest około 1:28 razy większa od średniej  $K$ . Dokładność jest pokazana na rysunku 5.13,



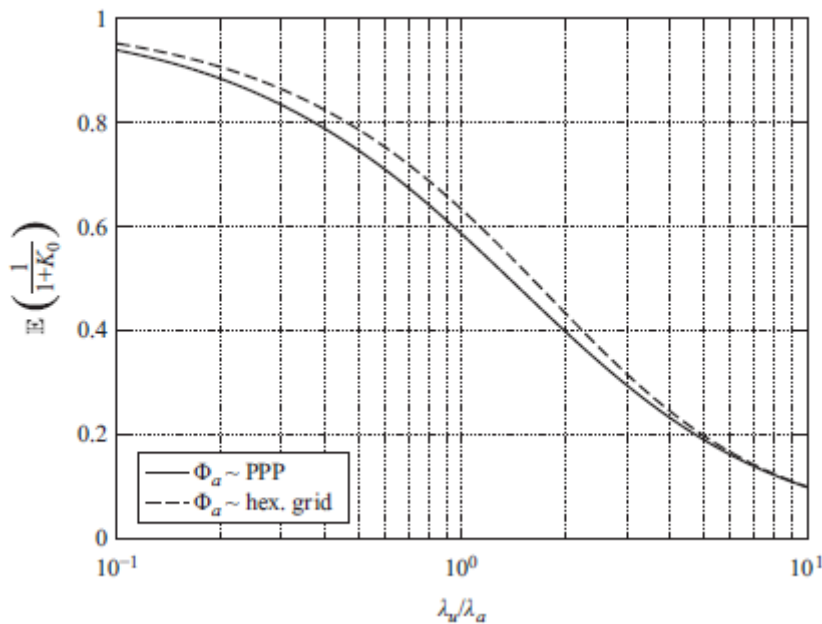
gdzie obserwuje się doskonałe dopasowanie z rozkładu empirycznego uzyskany za pomocą symulacji Monte Carlo dla dwóch wartości testowych  $\lambda_u/\lambda_a$ . Dowody eksperymentalne sugerują, że ten poziom dokładności dotyczy dowolnej wartości  $\lambda_u/\lambda_a$ . Przy dostępnym rozkładzie  $K_0$  rozkład całkowicie stopni swobody  $(1/1+K_0)$ . Naturalną metryką pozwalającą uzyskać wgląd w to, jak stopnie swobody zależą od parametrów systemu, jest ich średnia  $\mathbb{E}(1/1+K_0)$ , która okazuje się być dostępna w formie zamkniętej (nie ma potrzeby wykonywania bezpośredniej oceny numerycznej z wykorzystaniem rozkładu  $K_0$ ). Pokazano, że

$$\mathbb{E}\left(\frac{1}{1+K_0}\right) = \frac{\lambda_a}{\lambda_u} \mathbb{P}(K > 0),$$

gdzie  $\mathbb{P}(K > 0)$  jest (w przybliżeniu) podane przez przy  $\zeta = 7/2$ . Interesujące jest porównanie tej wartości z odpowiednią dla przypadku, w którym  $\Phi_a$  jest dystrybuowane jako okresowa sześciokątna siatka o tej samej gęstości  $\lambda_a$ . Oznaczając jako  $\bar{K}_0$  liczbę UE obsługiwanych przez ten sam punkt dostępowy, co typowy UE w ramach zwykłego wdrożenia AP, łatwo jest pokazać, że  $\bar{K}_0$  jest zmienną losową Poissona o średniej  $\lambda_u / \lambda_a$  i wynika z tego, że:

$$\begin{aligned} \mathbb{E}\left(\frac{1}{1+\bar{K}_0}\right) &= \sum_{k=0}^{\infty} \frac{1}{1+k} \frac{(\lambda_u/\lambda_a)^k}{k!} e^{-\lambda_u/\lambda_a} \\ &= \frac{\lambda_a}{\lambda_u} \left(1 - e^{-\lambda_u/\lambda_a}\right) \\ &\stackrel{(a)}{\geq} \mathbb{E}\left(\frac{1}{1+K_0}\right), \end{aligned}$$

Pokazuje to, że losowość wdrożenia infrastruktury powoduje pogorszenie wydajności UE w kategoriach średnich stopni swobody, oprócz utraty pod względem wydajności SIR, co wykazano w sekcji. 5.4.1.1. Ta degradacja jest określona ilościowo na rysunku 5.14, gdzie można zauważyć, że jest raczej ograniczona i najbardziej zauważalna w reżimie  $\lambda_u < 2\lambda_a$ .



### 5.5.2 Częstotliwość użytkowników CDF i problemy z projektowaniem systemu

Po ustaleniu statystyk  $K_0$  problem statystycznej charakterystyki (typowa) szybkość UE  $R$  może być teraz zaatakowana. Wynika to bezpośrednio z definicji  $R$ , że jego CDF jest równy:

$$\mathbb{P}(R \leq r) = \mathbb{E}\left(\mathbb{P}\left(\frac{1}{1+K_0} \log_2(1 + \text{SIR}) \mathbb{I}(\text{SIR} \geq \theta_{\text{th}}) \leq r \mid K_0\right)\right)$$

dla  $r > 0$ . Aby kontynuować, należy określić  $\mathbb{P}(\text{SIR} \leq \theta \mid K_0)$ , tj. CDF z SIR kondycjonowanego na  $K_0$ . Należy zauważyć, że w analizie tradycyjnego modelu rozmieszczenia w układzie heksagonalnym AP nie stanowi to problemu, ponieważ SIR i  $K_0$  są niezależne, tj.  $\mathbb{P}(\text{SIR} \leq \theta \mid K_0) = \mathbb{P}(\text{SIR} \leq \theta)$ . Jednak nie jest tak w przypadku, gdy  $\Phi_a$  jest PPP (lub innym procesem losowym). Można to zrozumieć intuicyjnie, zauważając, że większe wartości  $K_0$  sugerują większy obszar obsługującego AP, co z kolei implikuje, że zakłócające AP są położone dalej i odpowiednio większy SIR, tj. SIR i  $K_0$  są dodatnio skorelowane. Niestety, określenie  $\mathbb{P}(\text{SIR} \leq \theta \mid K_0)$  wydaje się być trudnym problemem, a ignorowanie zależności między SIR a  $K_0$  wydaje się być jedynym sposobem postępowania. Pozwala to na użycie wyrażenia zamkniętej formy (krajńowych) rozkładów SIR i  $K_0$  co skutkuje:

$$\mathbb{P}(R \leq r) = \sum_{k=0}^{\infty} \mathbb{P}(K_0 = k)(1 - F(k, r))$$

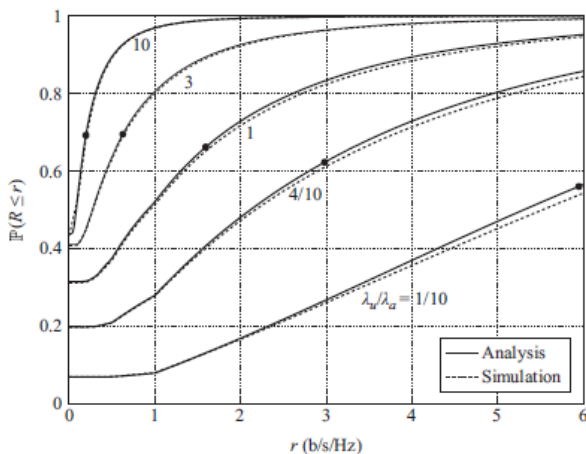
gdzie

$$F(k, r) \triangleq \begin{cases} \frac{1}{1 + \mathbb{P}(K > 0)\rho(2^{r(1+k)} - 1, \alpha)}, & r > \frac{1}{1+k} \log_2(1 + \theta_{\text{th}}) \\ \frac{1}{1 + \mathbb{P}(K > 0)\rho(\theta_{\text{th}}, \alpha)}, & r \leq \frac{1}{1+k} \log_2(1 + \theta_{\text{th}}), \end{cases}$$

z rozkładami  $K_0$  i  $K$  jak w (5.23). To wyrażenie dla stawki UE CDF można łatwo ocenić numerycznie (tylko umiarkowana liczba warunków sumy są wymagane dla wysokiej dokładności). Przyjęto również niezależność SIR i  $K_0$  pozwala na wyrażenie integralne dla  $\mathbb{E}(R)$ , bez wyraźnego użycia pełnej dystrybucji jak

$$\begin{aligned} \mathbb{E}(R) &= \mathbb{E}\left(\frac{1}{1 + K_0}\right) \mathbb{E}(\log_2(1 + \text{SIR})\mathbb{I}(\text{SIR} \geq \theta_{\text{th}})) \\ &= \frac{\lambda_a \mathbb{P}(K > 0)}{\lambda_u} \left( F(0, 0) + \int_{\log_2(1 + \theta_{\text{th}})}^{\infty} F(0, x) dx \right) \end{aligned}$$

Wskaźnik CDF przedstawiono na rysunku 5.15,





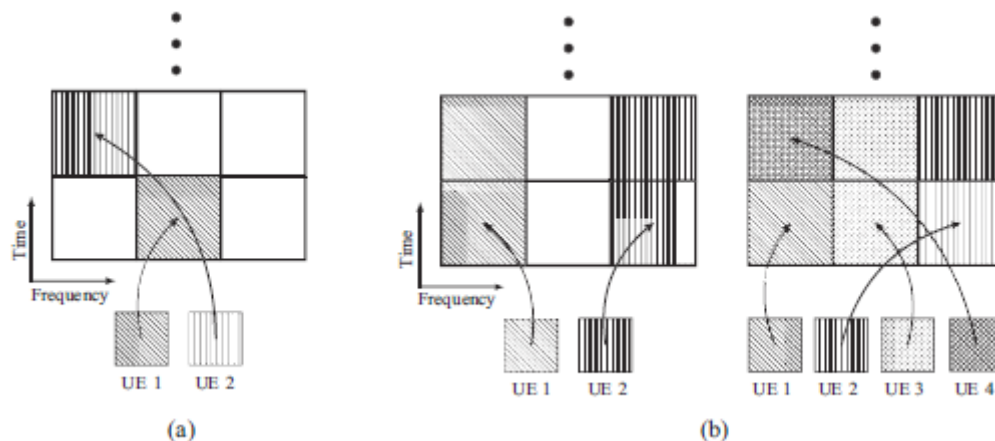
oszacowano za pomocą symulacji Monte Carlo dla różnych wartości  $\lambda_u / \lambda_a$ , w zakresie od 10 (ciężka sieć obciążona) do 1 / 10 (ultra gęstej sieci). Przybliżony wyraz wykazuje bardzo dobrą dokładność, co sugeruje, że zależność  $K_0$  i SIR jest rzeczywiście bardzo mała. Większe natężenie ruchu powoduje obniżone szybkości zgodnie z oczekiwaniami, podczas gdy ultra-gęsta sieć zapewnia znaczące wzrosty wydajności, wykorzystując zwiększoną swobodę dostępną ze względu na mniejszą liczbę UE powiązanych z tym samym AP. Należy zauważyć, że wydajność sieci jest wysoce niezrównoważona, ponieważ większość UE doświadcza szybkości mniejszych niż średnia (odpowiednio około 70% i 55% w przypadku ciężko obciążonych i bardzo gęstych sieci) i znaczna część UE faktycznie osiąga zerową prędkość z powodu przerwy SIR ( tj.  $SIR < \theta$ -te), szczególnie w przypadku bardzo obciążonych sieci.

### 5.5.3 Projektowanie systemu

Teoretyczna ekspresja CDF może być oczywiście wykorzystana nie tylko do wydajnego przewidywania wydajności systemu, ale, co ważniejsze, do uzyskania wglądu w projektowanie systemu. Na przykład projektant systemu ukierunkowany na pewną jakość usług (QoS) dla UE (wyrażony jako pewna funkcja CDF z R) i  $\lambda_u$  ustalony i znany, może wykorzystywać powyższe formuły do skutecznego określania (minimalnej) gęstości infrastruktury  $\lambda_a$  musi zapewnić tę QoS bez uciekania się do intensywnych komputerowo symulacji komputerowych (przykład zostanie omówiony później). Jest to ważny problem z projektowaniem sieci, biorąc pod uwagę obecny trend zagęszczania sieci. W tym miejscu pojawia się pytanie, czy (i ile) minimum  $\lambda_a$  wymagania mogą być rozluźnione w bardziej skomplikowanym modelu systemu niż rozważany do tej pory, biorąc pod uwagę, na przykład, zaawansowane techniki transmisji osiągające wyższą wydajność widmową. Odpowiedź na to pytanie wymaga oczywiście charakterystyki wydajności systemu w ramach tych technik. W tym celu pojawiły się ostatnio różne prace, rozszerzające dotychczasowy model systemu systemowego, obejmujący np. Formowanie wiązki, łączenie maksymalnego współczynnika i transmisję kooperacyjną z wielu punktów dostępowych. Zgodnie z oczekiwaniami, metody te zapewniają (statystyczny) wzrost SIR, co odzwierciedla się jako wzrost wydajności spektralnej łącza i stopy UE. Korzyści są kwantyfikowane za pomocą wyrażeń analitycznych, które nieuchronnie są bardziej skomplikowane niż te przedstawione w tej części. Jednak innym krytycznym parametrem określającym szybkość UE, która w ogóle nie jest uwzględniana przez te podejścia, są stopnie swobody  $1/(1+K_0)$ ; która zależy wyłącznie od schematu wielokrotnego dostępu. Mimo że założony dotychczas prosty TDMA jest całkowicie wiarygodny i pozwala na prostą analizę, można przypuszczać, że (łącznie) planowanie w czasie i częstotliwości może zwiększyć wydajność, przynajmniej w pewnych warunkach operacyjnych. Zostanie to zbadane w dalszej części, odpowiednio modyfikując model systemu i wprowadzając schematy szeregowania, które wykorzystują wymiary czasu i częstotliwości.

#### 5.5.3.1 Planowanie zstępujące w czasie i częstotliwości

Przepustowość łącza w dół zostanie uznana za podzieloną na  $N \geq 1$  subkanały (SC) o równej przepustowości, z których każda może być użyta do transmisji do pojedynczego urządzenia UE na AP w określonym przedziale czasowym (wszystkie przedziały czasowe mają równą długość). Liczba N jest zmienną wolną, która pozostaje stała, gdy jego (optymalna) wartość została określona na etapie projektowania systemu. Aby wykorzystać dostępne częstotliwości, w zaproponowano proste, analitycznie uległe rozszerzenie podstawowego schematu TDMA, który można nieformalnie opisać jako wprowadzenie elementu "przeskoku częstotliwości (FH)" do TDMA. W szczególności, każde UE jest podawane sekwencyjnie przez skojarzony z nim punkt dostępowy, tak jak w przypadku TDMA, jednakże transmisja do danego UE odbywa się na pojedynczym, losowo wybranym SC z całej dostępnej N. Przykład tej procedury szeregowania, określanej dalej jako "TDMA-FH", jest przedstawiony na Figurze 5.16 (a) dla  $N = 3$  i AP obsługujący  $K = 2$  UE.



Należy zauważyć, że TDMA-FH wprowadza dwa efekty na osiągniętą szybkość UE, w porównaniu ze schematem TDMA:

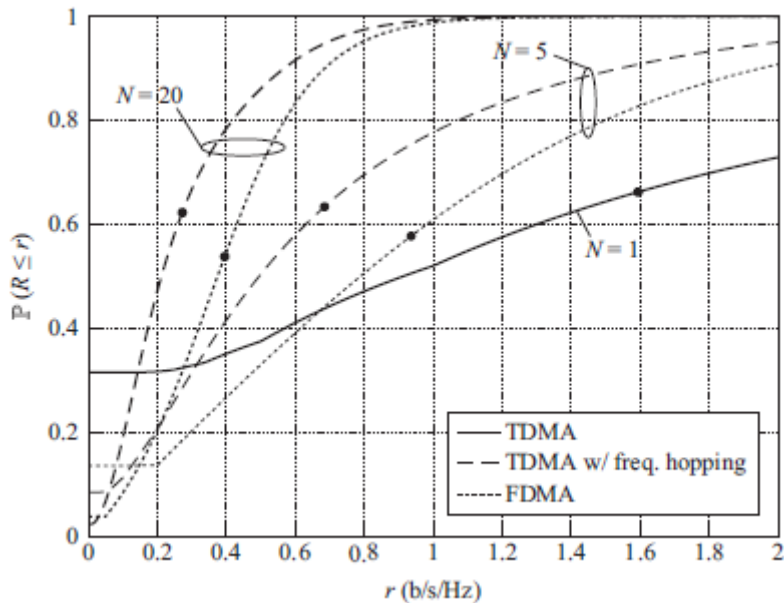
- Stopnie swobody każdego UE są zmniejszone o współczynnik  $1/N$ , tj. wyrażenie szybkości UE staje się  $R = \log_2(1 + SIR) / N(1 + K_0)$ .
- SIR jest zwiększony, ponieważ część (aktywnych) AP nie transmituje w tym samym czasie częstotliwość używana przez punkt dostępowy obsługujący typowe UE.

Łatwo zauważyć, że gęstość zakłócających AP wykorzystujących TDMA-FH staje się równa  $(1/N) \mathbb{P} \cdot (K > 0) \lambda_b$ . Przybliżanie ich pozycji jako PPP, analiza wydajności TDMA-FH jest jedynie drobną modyfikacją odpowiadającą jej w przypadku TDMA.

Zaproponowano inne podejście w próbach skuteczniejszego wykorzystania przepustowości poprzez umożliwienie wielu równoczesnych transmisji w ramach jednego przedziału czasowego. Podstawową zasadą tego schematu szeregowania jest to, że urządzenia użytkownika UE są równomiernie i losowo rozmieszczone na SC (dozwolone jest tylko pojedyncze SC dla każdego UE), i w przypadku, gdy wiele UE jest przypisanych do tego samego SC, wykonywane jest współdzielenie czasu. Przykład tego schematu planowania, określanego poniżej jako "FDMA", przedstawiono na rysunku 5.16 (b). W porównaniu z TDMA-FH, FDMA osiąga lepsze wykorzystanie przepustowości systemu, odzwierciedlone przez wyrażenie szybkości UE, które staje się:

$$R = \frac{1}{N(1 + L_0)} \log_2(1 + SIR),$$

gdzie  $0 \leq L_0 \leq [K_0/N] \leq K_0$  jest liczbą dodatkowych UEs dzielących to samo SC co typowe UE. Skutecznie, FDMA zwiększa stopnie swobody w porównaniu z TDMA-FH. Z drugiej strony, bardziej wydajne wykorzystanie pasma FDMA powoduje zwiększenie gęstości zakłóceń w porównaniu z TDMA-FH. Analiza wyników dla FDMA jest bardziej zaangażowana w porównaniu z TDMA-FH, jednak jest zgodna z tymi samymi zasadami, co we wszystkich wcześniej zbadanych przypadkach. CDF R dostarczonego przez TDMA, TDMA-FH i FDMA pokazano na rysunku 5.17



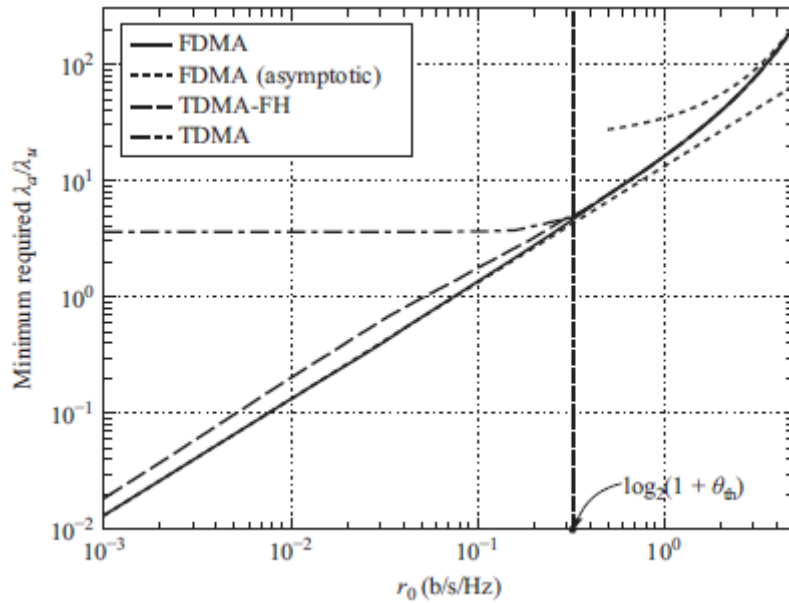
dla  $N = 5$  i  $N = 20$  (oczywiście TDMA można postrzegać jako specjalny przypadek pozostałych dwóch schematów z  $N = 1$ ). Obciążenie UE odpowiadające  $\lambda_u / \lambda_a = 1$  jest rozważane z tendencjami wydajności podobnymi do innych wartości  $\lambda_u / \lambda_a$ . Można zauważyć, że TDMA-FH znacznie poprawia wydajność w porównaniu z TDMA w reżimie o niskiej szybkości (urządzenia krawędziowe UE), ponieważ jego właściwe zmniejszenie interferencji skutkuje zmniejszonym prawdopodobieństwem, że urządzenie UE doświadcza awarii SIR. Zgodnie z oczekiwaniami zwiększenie  $N$  dodatkowo zwiększa ten efekt. Jednocześnie wyniki z TDMA-FH pogarszają się w reżimie o wyższych stawkach (jak również pod względem średniej stopy UE) w porównaniu z TDMA z powodu zmniejszenia dostępnych stopni swobody transmisji. Wydajność FDMA jest zgodna z tymi samymi trendami. Jak można się było spodziewać, FDMA zapewnia (nieco) gorszą wydajność w reżimie niskiej stawki w porównaniu z TDMA-FH ze względu na zwiększoną interferencję wytwarzaną przez transmisje współbieżne, podczas gdy jest w stanie zapewnić lepszą wydajność przy wyższych prędkościach ze względu na zwiększony stopień wolności.

### 5.5.3.2 Wymagania dotyczące gęstości infrastruktury

Aby zobaczyć potencjał planowania w częstotliwości, należy rozważyć problem określenia minimalnej wartości  $\lambda_a / \lambda_u$ , tj. gęstości AP znormalizowanej przez gęstość UE, niezbędną do zapewnienia szybkości UE  $R \geq r_0 > 0$  z prawdopodobieństwem  $0 < \varepsilon < 1$  (ekwiwalentnie,  $\varepsilon 100\%$  UEs doświadcza szybkości równych lub większych niż  $r_0$ ). Ta minimalna wartość  $\lambda_a / \lambda_u$  może być określona (liczbowo) dla przypadku TDMA poprzez bezpośrednie wykorzystanie przedstawionych wcześniej formuł wydajności. Dla przypadków TDMA-FH i FDMA, minimalny parametr  $\lambda_a / \lambda_u$  zależy od parametru  $N$ , który może być również zoptymalizowany numerycznie. Dla przypadku FDMA, zamknięta forma górna granica minimalnych wymagań (pod optymalnym wyborem  $N$ ) można uzyskać, która asymptotycznie skaluje się jako

$$\lambda_a / \lambda_u \leq \begin{cases} \frac{(1 - \varepsilon) \rho(\theta_{th}, \alpha)}{\varepsilon \log_2(1 + \theta_{th})} r_0, & r_0 \ll \log_2(1 + \theta_{th}) \\ \frac{(1 - \varepsilon)}{\varepsilon} \kappa(\alpha) 2^{2r_0/\alpha}, & r_0 \gg \log_2(1 + \theta_{th}), \end{cases}$$

tj. liniowo i wykładniczo z  $r_0$  odpowiednio dla małych i dużych  $r_0$ . Rysunek 5.18



przedstawia minimalne wartości  $\lambda_a/\lambda_u$  z TDMA, TDMA-FH i FDMA dla różnych  $r_0$  z  $\varepsilon = 0,1$  i  $\theta_{ty} = -6$  dB (próg SIR dla LTE). Przedstawiono asymptotyczne górne ograniczenie dla przypadku FDMA. Można zauważyć, że logarytm wartości szybkości  $\log_2(1+\theta_{ty})$  jest "punktem przejściowym" określającym, jak minimalne wymagania dotyczące gęstości systemu są skalowane. W szczególności dla wartości  $r_0 > \log_2(1+\theta_{ty})$ , wszystkie trzy schematy zapewniają taką samą wydajność, jak się okazuje, że optymalny  $N$  wynosi 1 dla obu TDMA-FH i FDMA. Z drugiej strony dla wartości  $r_0 < \log_2(1+\theta_{ty})$  partycjonowanie pasma jest korzystne i faktycznie pozwala na dramatyczny (rzędu rzędów) spadek wymagań infrastrukturalnych w porównaniu z TDMA. FDMA ma minimalne wymagania gęstości, a następnie dokładnie TDMA-FH. Wyniki te wskazują, co następuje:

1. W przypadku małych wymagań stawka prosty schemat TDMA jest wysoce nieefektywny i wymaga specjalnej uwagi przy interpretacji analiz wydajności uzyskanych zgodnie z założeniem TDMA.
2. W przypadku wymagań o dużej stawce gęstość infrastruktury musi wzrastać wykładniczo, tzn. konieczne jest bardzo duże zagęszczenie sieci. Jednak wymagania te mogą być w praktyce niewykonalne, powodując zastosowanie zaawansowanych technik transmisji i / lub zwiększenia dostępnej przepustowości niezbędnej w przyszłych sieciach komórkowych.

# Łącza bezprzewodowe urządzenie-urządzenie (D2D) do komunikacji między maszynami (M2M)

## Abstrakcja

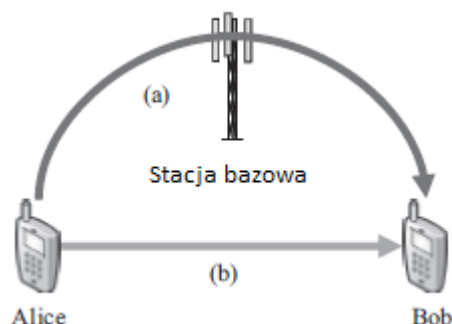
Komunikacja typu Device-to-Device (D2D) będzie odgrywać ważną rolę w sieciach komórkowych piątej generacji (5G), zwiększając przestrzenne ponowne wykorzystywanie zasobów widma i umożliwiając połączenia komunikacyjne o niskim opóźnieniu. D2D składa się z dwóch podstawowych elementów: odkrywania bliskości i bezpośredniej komunikacji między użytkownikami w pobliżu. Innym wyłaniającym się trendem w bezprzewodowych systemach komórkowych jest komunikacja typu Machine-to-Machine (M2M), często charakteryzująca się stałą, niską szybkością transmisji. W tej części motywujemy synergii między D2D i M2M, a obecne technologie, które umożliwiają komunikację M2M-via-D2D, działają jako podkład dla zwykłych transmisji komórkowych

## 6.1 Wprowadzenie

Spośród wielu technologii bezprzewodowych rozwijanych w kierunku 5G, bezprzewodowe sieci komórkowe obejmują dwa trendy technologiczne: komunikację między urządzeniami (D2D) i komunikację między maszynami (M2M). Trendy te różnią się od głównego nurtu technologii, odziedziczonego po poprzednich generacjach bezprzewodowych, których celem jest osiągnięcie wyższych szybkości transmisji danych i lepszego zasięgu. Komunikacja D2D umożliwia bezpośrednie połączenia między urządzeniami bezprzewodowymi wykorzystującymi to samo spektrum i interfejsem radiowym wykorzystywanym do komunikacji komórkowej. Drugi trend komunikacji M2M, znany również pod nazwą Machine-Type Communications (MTC), ma na celu dołączenie dużej liczby urządzeń niskiej mocy o niskim poborze mocy, nazywanych urządzeniami typu maszynowego (MTD) do sieci komórkowej. Istnieje wiele czynników, które wymagają większej liczby podłączonych MTD: inteligentnej sieci, monitorowania środowiska na dużą skalę i struktury, monitorowania zasobów i zdrowia itd.

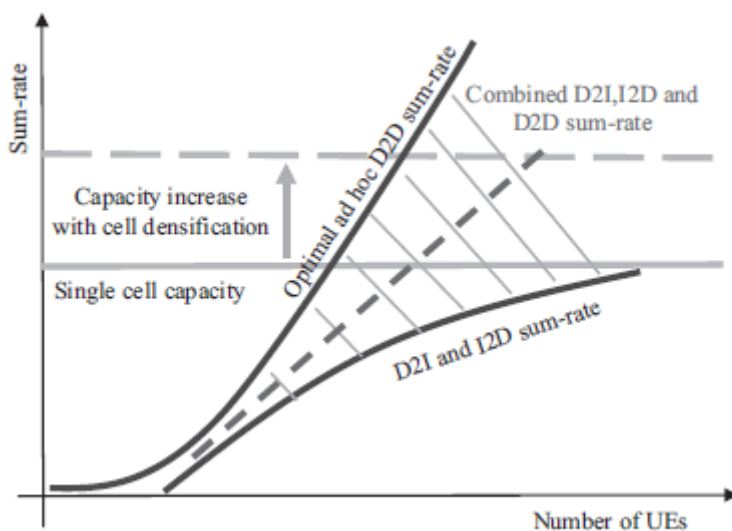
### 6.1.1 Kluczowe pojęcia w komunikacji D2D

Kluczową koncepcję komunikacji urządzenie-urządzenie można podsumować w następujący sposób. Zastanówmy się nad dwoma użytkownikami Alice i Bobem, z których każdy ma skojarzone urządzenie komórkowe, którzy chcą się ze sobą komunikować za pomocą zasobów sieci komórkowej, jak pokazano na rysunku 6.1.

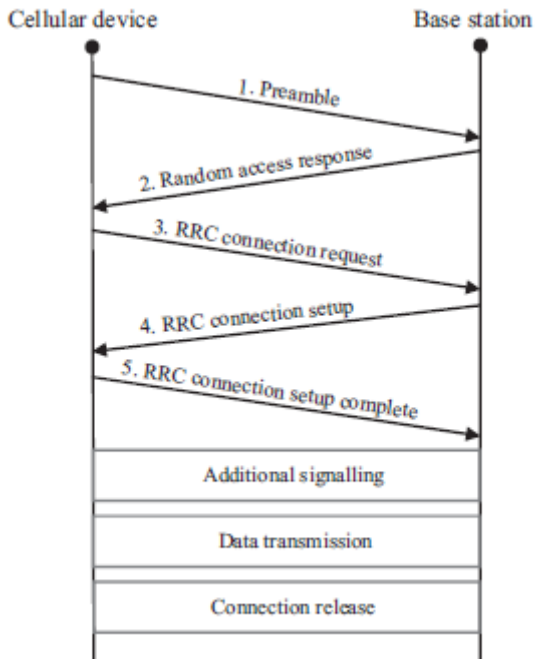


W tradycyjnej sieci komórkowej, takiej jak GSM, UTM i LTE-A (przed wydaniem 12), ta komunikacja odbywałaby się w infrastrukturze sieci komórkowej. W przeciwieństwie do komunikacji D2D, urządzenia komunikują się bezpośrednio, co jest możliwe tylko wtedy, gdy dwa urządzenia są w pobliżu

wzajemnie. Aby ułatwić wyjaśnienie kluczowych pojęć, zakładamy, że znajdujemy się w sieci komórkowej LTE-A. W sieci komórkowej ustanowienie łączności D2D między proksymalnymi urządzeniami komórkowymi jest wspomagane przez sieć; i oczekuje się, że odegra ważną rolę w systemach 5G. Motywacja do wprowadzenia komunikacji D2D w kontekście komórkowym wynika z (i) potrzeby poprawy wykorzystania zasobów sieciowych, tj. Sprawności widmowej obszaru oraz (ii) umożliwienia nowych usług komórkowych wymagających bezpośredniej komunikacji między urządzeniami. Potencjalny wzrost efektywności wykorzystania widma wynika z intuicji że całkowita przepustowość sieci może zostać zwiększona, jeżeli bezpośrednia łączność między pobliskimi urządzeniami jest dozwolona w komórce. Jest to również motywacja do wprowadzania małych komórek, czasami określanych jako zagęszczanie komórek; podczas gdy D2D przyjmuje tę koncepcję dalej poprzez zagęszczenie łącza. Rysunek 6.2 ilustruje potencjalne korzyści z zezwalania na połączenia D2D w sieci komórkowej, w porównaniu ze wzrostem pojemności po prostu wprowadzając mniejsze komórki.



Co więcej, wprowadzenie komunikacji D2D otwiera możliwość rozładowania infrastruktury komórkowej, gdy tylko identyczna treść musi być obsługiwana w tym samym zlokalizowanym obszarze, a także możliwość posiadania wiarygodnego łącza równorzędnego w licencjonowanym widmie do udostępniania np. treści multimedialnych. W takich przypadkach wolumeny danych są stosunkowo duże i dlatego warto skorzystać z wysokiej jakości łączy D2D. W przypadku połączeń przechodzących przez infrastrukturę komórkową komunikacja może odbywać się tylko wtedy, gdy oba urządzenia komunikacyjne są podłączone do sieci i są przełączane w stan RRC\_Connected. Ustalenie tego połączenia odbywa się z wykorzystaniem komórkowego protokołu Access\_Reservation i dodatkowej wymiany sygnalizacji wymaganej do przełączenia urządzenia ze stanu RRC\_Idle na RRC\_Connected. Ilustrujemy ten protokół na rysunku 6.3.

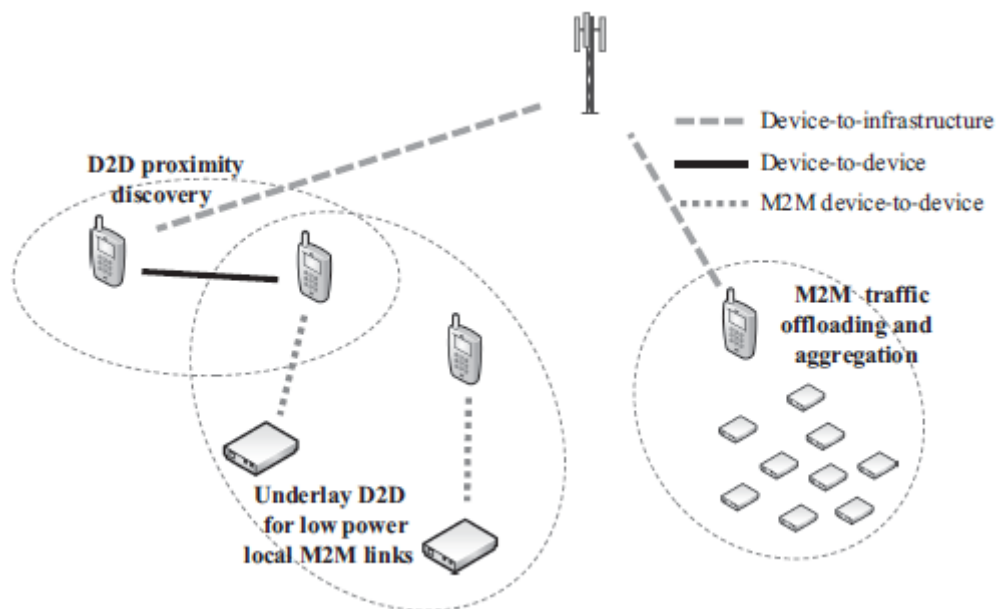


Po zakończeniu tej procedury przez oba urządzenia, połączenie komunikacyjne może zostać ustanowione między dwoma urządzeniami komórkowymi za pośrednictwem infrastruktury komórkowej. W przypadku D2D przed komunikacją konieczne jest, aby urządzenia wykryły, czy znajdują się blisko siebie. Wyszukiwanie zbliżeniowe może być wspierane przez sieć, w którym to przypadku urządzenia muszą najpierw zostać podłączone do sieci (za pomocą opisanej procedury dostępu na Rysunku 6.3). Z drugiej strony, jeśli brak jest obsługi sieci lub urządzenia znajdują się poza zasięgiem sieci komórkowej, odkrywanie bliskości odbywa się w sposób ad hoc. Bez względu na podejście stosowane do wykrywania bliskości, urządzenia nawiązują bezpośrednią komunikację, jeśli są blisko siebie. Jeśli nie są one proksymalne i znajdują się w zasięgu sieci komórkowej, komunikacja odbywa się za pośrednictwem infrastruktury sieci komórkowej. W przypadku, gdy nie ma zasięgu sieci komórkowej, jedyną opcją uzyskania łączności jest przekazywanie wielu przeskoków obsługiwane przez inne urządzenia komórkowe.

### 6.1.2 W jaki sposób D2D może pomóc MTC?

Podstawową cechą korzystania z D2D dla MTC jest wymaganie dla łącza D2D o niskiej częstotliwości. Na przykład rozważ czujniki na ciele jako urządzenia MTC: każde urządzenie wymaga niskiej częstotliwości i bardzo niskiej mocy transmisji, niewystarczającej do połączenia bezpośrednio ze stacją bazową (BS). W takim przypadku urządzenie MTC wykorzystuje łącze D2-D do smartfonu, które dodatkowo przekazuje ruch MTC do stacji BS. Ogólnie rzecz biorąc, ten typ D2Drelayed MTC może być używany w innych aplikacjach MTC, które nie mają wysokich wymagań dotyczących opóźnień. Obserwujemy, że pożądane jest przenoszenie ruchu maszynowego nie poprzez bezpośrednie połączenia z BS, ale przez pobliski przekaźnik. Dlatego w tym rozdziale omawiamy możliwości otwierane przez łącza D2D do obsługi MTD, dla których motywacja jest dwójaka: (1) Gdy MTD ma możliwość połączenia się z infrastrukturą za pośrednictwem pobliskiego urządzenia, wtedy może obniżyć moc nadawczą; (2) mogą istnieć potencjalnie tysiące MTD połączonych z tą samą komórką, tak że zarządzanie wieloma indywidualnymi połączeniami może spowodować duży nadmiar i nieefektywność w systemie komórkowym. Jeśli wiele przekaźników MTD jest połączonych z węzłem przekazującym, a węzeł przekaźnikowy uzyskuje dostęp do BS w imieniu tych urządzeń, wówczas obciążenie zarządzania połączeniem w stacji BS może drastycznie zmniejszyć się. Krytycy mogą wskazać

dwa argumenty przeciwko używaniu przekazywania dla ruchu maszynowego. Po pierwsze, przekazywanie zwiększa opóźnienie połączenia; jednakże istnieje wiele różnych aplikacji maszynowych i nie wszystkie z nich wymagają opóźnień w milisekundach. Przykłady obejmują liczniki ciepła, czujniki zdrowia itp., a przekazywanie może być ograniczone do ruchu o wyższej tolerancji opóźnienia. Drugi argument jest taki, że wprowadzenie przekaźników infrastruktury jest zbyt kosztowne, a to ustawia motywację do przekazywania ruchu MTC przez inne, bardziej wydajne urządzenie, takie jak smartfon, za pomocą łącza D2D. Inną ważną obserwacją, którą opiszemy w dalszej części rozdziału, jest to, że komunikacja generowana podczas procedury wykrywania bliskości może być również sklasyfikowana jako ruch podobny do MTC. Schemat rozważany w tym rozdziale pokazano na rysunku 6.4. Różnica w stosunku do powszechnego stosowania transmisji D2D polega na tym, że tutaj łącze D2D służy do wysyłania z niską, stałą stawką, a celem jest uzyskanie pewnej gwarancji prawdopodobieństwa wyłączenia. Ostatnio zbadano kilka potencjalnych synergii między D2D i M2M, mających na celu zwiększenie ponownego wykorzystania, zasięgu i pokrycia spektralnego. W tych synergjach badano w kontekście operacji grupowych, tj. Urządzenia komórkowego z obsługą D2D, które działa jako głowica dla grupy MTD. W łączu w górę głowica klastra agreguje i przekazuje zebrane żądania, pakiety danych i informacje o stanie z urządzeń MTC do połączonego BS 3G. W łączu "w dół" serwer nagłówek przekazuje komunikaty zarządzania i pakiety danych z BS do MTD w grupie. Wykorzystanie komunikacji D2D zmniejsza przeciążenie sygnalizacji na interfejsie radiowym i obciążenie zarządzania siecią, co jest zgodne z duchem zarządzania grupowego zdefiniowanego przez 3GPP. Rozważono, w jaki sposób bezpośrednia komunikacja w obrębie grupy urządzeń może poprawić wydajność konwencjonalnego systemu komórkowego, wykorzystując wybór między bezpośrednim D2D i przekazywaniem infrastruktury. W, wpływ MTD na sieć komórkową jest łagodzony przez losowe rozmieszczanie kolektorów danych w celu gromadzenia ruchu z MTD. Wreszcie, w protokole trunkingowym i agregacji był zaproponowany w taki sam sposób, jak przedstawiono na rysunku 6.4.



Standaryzacja traktuje D2D pod nazwą Proximity Services (ProSe), gdzie w kilku innych przypadkach wykorzystuje ona przypadek użycia, w którym urządzenie komórkowe poprawia zasięg, działając jako przekaźnik w imieniu jednego lub więcej innych urządzeń komórkowych, które znajdują się poza zasięgiem sieci. Z punktu widzenia architektury, przekazywanie D2D w ProSe jest określone jako rozszerzenie zasięgu i jest specyficzne dla przypadków użycia bezpieczeństwa publicznego]. Raport] poza



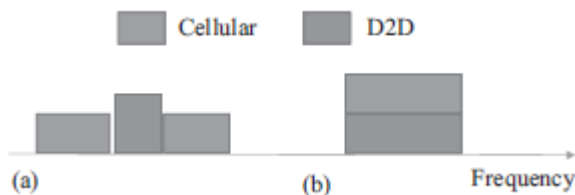
zapewnieniem aspektów projektowych dla warstw fizycznych i górnych, stwierdza również, że jednym z wymogów jest wsparcie dużej liczby współbieżnie korzystających z ProSe użytkowników.

### 6.1.3 Organizacja części

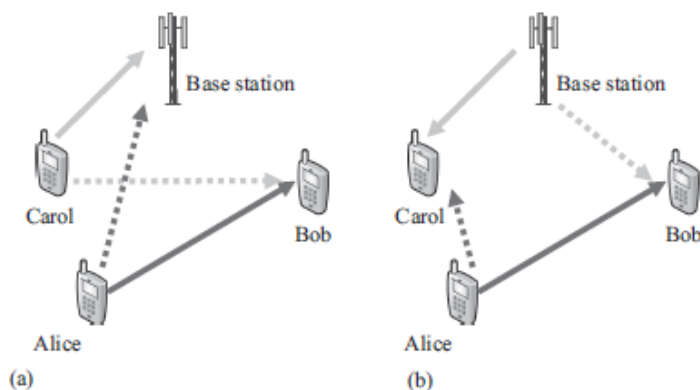
W sekcji 6.2 pokazujemy najnowsze wyniki, które umożliwiają operację podkładania, jednocześnie oferując zerową awarię dla łącza będącego w zasięgu. W sekcji 6.3 przedstawiamy kluczowe ogólne pojęcia odkrywania bliskości i pokazemy, jak zastosować technikę podkładania w celu wdrożenia wykrywania. W sekcji 6.4 przedstawiamy i omawiamy dwie topologie komunikacji podkładowej i uzyskujemy karę mocy transmisji wymaganą do osiągnięcia takiej samej wydajności jak w przypadku bez operacji podkładania. Rozdział kończy się podsumowaniem prezentowanych wyników.

## 6.2 Podkładanie linków D2D

łącza D2D, z perspektywy sieci komórkowej, mogą wykorzystywać pasmo i / lub pozapasmowych zasobów widma



Zasoby wewnątrzpasmo zazwyczaj należą do widma objętego licencją na dostęp komórkowy; podczas gdy zasoby poza pasmem odpowiadają pasmom nielicencjonowanym, takim jak pasma ISM. W zasobach pozapasmowych nie występuje zakłócenie działania sieci komórkowej oraz w przypadku, gdy nie ma centralnego podmiotu odpowiedzialnego za kontrolowanie wykorzystania widma, nie ma również koordynacja interferencji. W zasobach wewnątrz pasma dostęp do kanału można klasyfikować jako nakładka lub podkład. W dostępie do nakładek łącza komórkowe i D2D mają dedykowane i ortogonalne zasoby; podczas dostępu do podkładu, łącza komórkowe i D2D dzielą te same zasoby radiowe. W tym rozdziale skupiamy się na wykorzystaniu dostępu do podkładu D2D. W przypadku operacji podkładania łącza D2D wprowadzają zakłócenia i są zakłócone przez sąsiednie węzły sieci. Zakładając, że łącza komórkowe i łącza D2D są na dwóch osobnych poziomach, interferencję można zaklasyfikować jako: (1) międzywarstwową, która odnosi się do sytuacji, w których agresor (np. Użytkownik D2D) i ofiara (np. użytkownik komórkowy) zakłóceń należą do różnych warstwa sieci; (2) współdzielenie, w którym agresor (np. Użytkownik D2D) i ofiara (np. sąsiadujący użytkownik D2D) należą do tej samej warstwy sieci. Komunikacja podkładania może odbywać się zarówno w łączu komórkowym, jak i w łączu w dół, co przedstawiono na Rysunku 6.6.



Kiedy podkład jest wykonywany w zasobach połączenia ładującego, wówczas interferencja D2D występuje na BS; podczas gdy podłoże korzysta z zasobów łącza pobierającego, interferencja występuje u użytkownika komórkowego. Tradycyjnie BS lepiej radzi sobie z interferencją niż użytkownik komórkowy, co tłumaczy, dlaczego większość prac w literaturze opiera się na podkładzie łącza w górę. W przeciwieństwie do tego omawiamy podstawy połączeń D2D w zasobach łącza pobierającego, motywowane wprowadzeniem urządzeń komórkowych z zaawansowanymi możliwościami dekodowania wielu użytkowników, takimi jak kolejne usuwanie zakłóceń i wspólne dekodowanie.

### 6.2.1 Model sygnału podpodłogowego

W rozważanym ustawieniu podkładu zstępującego na rysunku 6.6, Carol i Bob otrzymują jednocześnie transmisje z BS i Alicji. Złożony sygnał  $y$  odebrane w odbiorniku (Boab lub Carol) to

$$y = h_A x_A + h_B x_B + Z$$

gdzie  $h_B$  i  $h_A$  oznaczają odpowiednio kompleksowe wzmocnienia kanałów między BS, Alice i odbiornikiem, podczas gdy  $Z$  jest szumem Gaussa o wartości zespolonej z wariancją  $E[|Z|^2] = \sigma^2 \cdot x_B + x_A$  są podane przez okrągły zerowy średni łańcuchowy sygnał Gaussa transmitowany przez BS i Alice, gdzie odpowiednie wariancje są  $E[|x_A|^2] = P_A$  i  $E[|x_B|^2]$ . Dla uproszczenia zakłada się, że moc nadawcza Alice jest stała i zwymiarowana, aby osiągnąć cel systemowy. Zakłada się, że wszystkie łącza komunikacyjne są scharakteryzowane przez blaknięcie Rayleigha, to znaczy, że zanikanie kanału nie zmienia się w obrębie epoki szeregowania.  $g_i = |h_i|^2$  to wzmocnienie obwiedni kanału, które w przypadku zanikania Rayleigh następuje po rozkładzie wykładniczym. Związana z tym funkcja gęstości prawdopodobieństwa (PDF),  $f_g(x)$ , jest podana przez,

$$f_g(x) = \frac{1}{\bar{g}} \exp\left(-\frac{x}{\bar{g}}\right), \text{ with } \bar{g} = 1$$

Cała komunikacja odbywa się za pomocą przyzwyczajenia przepustowości dla pojedynczego użytkownika, od punktu do punktu Gaussowskie słownictwo i chwilowa osiągalna szybkość  $R_i$  jest podana przez asymptotyczną zdolność Shannon w AWGN,  $R_i = C(x) = W \log_2(1 + x)$ , gdzie  $W$  jest przepustowością kanału. Docelowy stosunek sygnału do szumu (SNR)  $\Gamma_x$ ; dla danego docelowa szybkość  $R_x$  jest zdefiniowany jako  $\Gamma_x = C^{-1}(R_x) = 2^{R_x/W} - 1$ .

### 6.2.2 Optymalne dekodowanie podkładania

Jak pokazano w (6.1), dwu-użytkownik Gaussian Multiple Access Channel (MAC) tworzony jest zarówno na odbiornikach Boba, jak i Carol. W standardowym założeniu MAC jest aby odbiornik pomyślnie zdekodował sygnały wszystkich nadajników. Oznaczając współczynniki sygnałów obecnych jako  $R_B$ , dla BS i  $R_A$ , dla Alice, następnie MAC określa nierówności

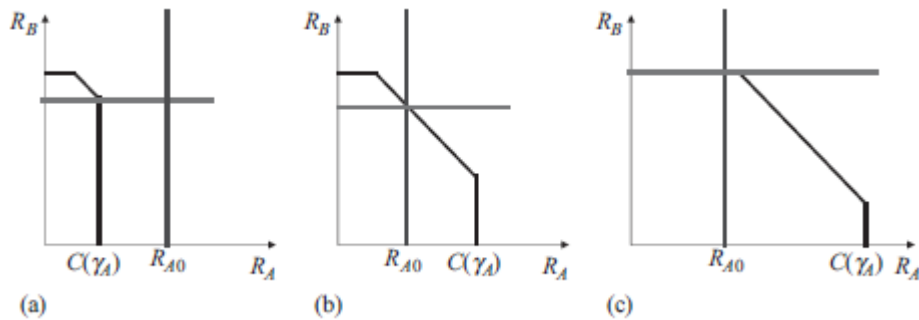
$$R_A \leq C(\gamma_A)$$

$$R_B \leq C(\gamma_B)$$

$$R_A + R_B \leq C(\gamma_A + \gamma_B)$$

gdzie  $\gamma_A$  i  $\gamma_B$  oznaczają stosunek sygnału do szumu odpowiadający sygnałom Alice i BS w odbiorniku. W naszym scenariuszu MTC zakładamy, że Alicja transmituje ze stałą stawką  $R_{A,0}$ . Dlatego analiza naszego modelu różni się od klasycznego traktowania MAC, ponieważ  $R_{A,0}$  jest stałe i musimy przeanalizować

dekodowalność  $R_B$ , niezależnie od tego, czy  $R_{A,0}$  jest zdekodowalna. Na rysunku 6.7 przedstawiamy trzy reżimy robocze wspólnego dekodera.



Przyjmujemy teraz jako punkt widzenia odbiornik Carol. Jeśli  $\gamma_A < \Gamma_{A,0}$ , wówczas pierwsza nierówność jest naruszona, tak że sygnał z Alice traktowany jest jako szum, a maksymalna osiągalna szybkość łączy w dół to

$$R_B \leq C\left(\frac{\gamma_B}{1 + \gamma_A}\right)$$

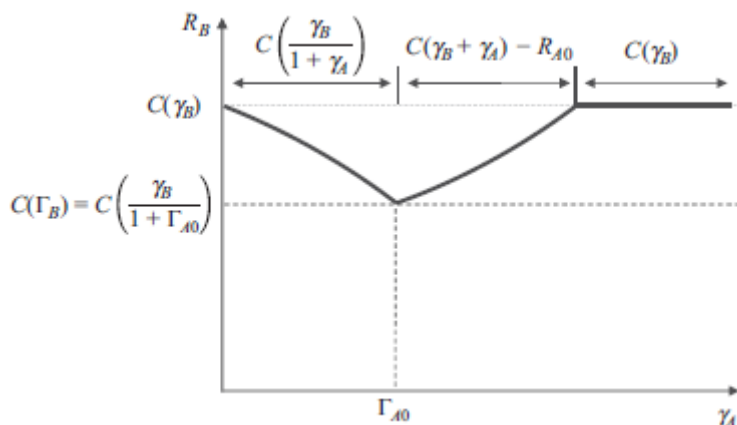
Z drugiej strony, gdy  $\gamma_A \geq \Gamma_{A,0}$  to są dwie opcje: (a) albo kontynuuj traktowanie sygnału z Alice jako szum użycie powyższego równania, w celu określenia maksymalnego  $R_B$  lub (b) stosować nierówności MAC co prowadzi do:

$$R_B \leq \min\{C(\gamma_B), C(\gamma_A + \gamma_B) - R_{A,0}\} = \min\{C(\gamma_B), C(\gamma_A + \gamma_B) - C(\Gamma_{A,0})\}$$

Należy jednak pamiętać, że jeśli dla danego  $\gamma_A \geq \Gamma_{A,0}$  wyrażenie prowadzi do wyższej niż  $R_B$  w porównaniu do powyższego, wtedy odbiorca powinien traktować sygnał z Alicji jako szum (ale zobaczymy, że tak nie jest). Z lekkim nadużywaniem notacji, możemy zdefiniować  $R_B(\gamma_A)$  jako maksymalny osiągalny  $R_B$  dla danego  $\gamma_A$  który można następnie zapisać zwięźle jako

$$R_B(\gamma_A) = \begin{cases} C\left(\frac{\gamma_B}{1 + \gamma_A}\right) \\ \max\left\{C\left(\frac{\gamma_B}{1 + \gamma_A}\right), \min\{C(\gamma_B), C(\gamma_A + \gamma_B) - C(\Gamma_{A,0})\}\right\} \end{cases}$$

Rysunek 6.8 pokazuje,



jak ta funkcja wygląda. Gdy  $\gamma_A < \Gamma_{A,0}$ ,  $R_B(\gamma_A)$  maleje, ponieważ większy  $\gamma_A$  oznacza większy szum, odpowiadający reżimowi przedstawiony na rysunku 6.7 (a). Po  $\gamma_A$  osiąga  $\Gamma_{A,0}$   $R_B(\gamma_A)$  zaczyna wzrastać, co odpowiada reżimowi przedstawionemu na rysunku 6.7 (b). Wreszcie, kiedy połączenie między Alice i Carol stają się zbyt silne  $\gamma_A \geq \Gamma_{A,0}$ , szybkość połączenia w dół osiąga swoje maksymalna możliwa wartość  $R_B(\gamma_A) = C(\gamma_B)$  w przypadku zniknięcia zakłócenia transmisji Alice, odpowiadającego reżimowi przedstawionemu na rys. 6.7 (c). Możemy teraz podać następujące informacje:

Lemat 6.2.1 Niech będzie transmisja o stałej stopie  $R_{A,0}$ : Niech B zna stawkę  $R_{A,0}$  i realizacja SNR  $\gamma_B$  w odbiorniku, ale nie  $\gamma_A$ . Następnie maksymalna szybkość transmisji łącza nadawczego, która jest zawsze dekodowalna przez odbiornik, wynosi  $R_B = C(\Gamma_B)$  gdzie

$$\Gamma_B = \frac{\gamma_B}{1 + \Gamma_{A,0}}$$

i

$$\Gamma_{A,0} = C^{-1}(R_{A,0}).$$

Dowód. Dowód opiera się na znalezieniu punktu przełączania pomiędzy region, w którym transmisja z B jest dekodowana przez traktowanie transmisji od A jako szum, do regionu, w którym oba sygnały są wspólnie dekodowane, jak przedstawiono w Rysunek 6.8. Maksymalna transmisja łącza nadawczego  $R_B = C(\Gamma_B)$  odpowiada wtedy ten sam punkt przełączania, który występuje, gdy  $\gamma_A = \Gamma_{A,0}$ . Innymi słowy, przy dostosowywaniu prędkości sygnału z punktu B do zdefiniowanego górnego ograniczenia, transmisja B zawsze może być dekodowana, niezależnie od transmisji z dekodowania A. Ciekawym efektem ubocznym tego górnego ograniczenia jest to, w jaki sposób wpływa on na wyłączenie transmisji urządzenia ogłaszającego. Kiedy stosujemy górną granicę prędkości łącza nadawczego, wszystkie nierówności MAC są przestrzegane, tj. Sygnał z B jest zawsze dekodowalny. Wtedy dekodowalność  $R_{A,0}$  będzie zależęć wyłącznie od tego, czy  $R_{A,0} \leq C(\Gamma_A)$ , tj. prawdopodobieństwo transmisji z A jest dekodowane przez  $\Pr\{\gamma_A \geq \Gamma_A\}$ .

Wniosek 6.2.2 . Niech będzie istniała transmisja o stopie  $R_{A,0}$  i niech B zna stawkę  $R_{A,0}$  i SNR  $\gamma_B$ , i wybierz jego szybkość tak, aby  $R_B = C(\Gamma_B)$ . Wtedy dekodowalność  $R_{A,0}$  będzie zależęć wyłącznie od  $R_{A,0} \leq C(\gamma_A)$

Dowód. Z Lematu 6.2.1 wynika, że transmisja z B zawsze jest dekodowalna. Dlatego istnieją dwa scenariusze dekodowania: (i) B jest dekodowany, traktując transmisję Alice jako szum umożliwiając dekodowanie transmisji Alice w obecności szumu wtedy i tylko wtedy, gdy  $R_{A,0} \leq C(\Gamma_{A,0})$ ; (ii) transmisja Alice jest dekodowana wspólnie z B; co jest prawdą tylko wtedy, gdy wszystkie nierówności utrzymują

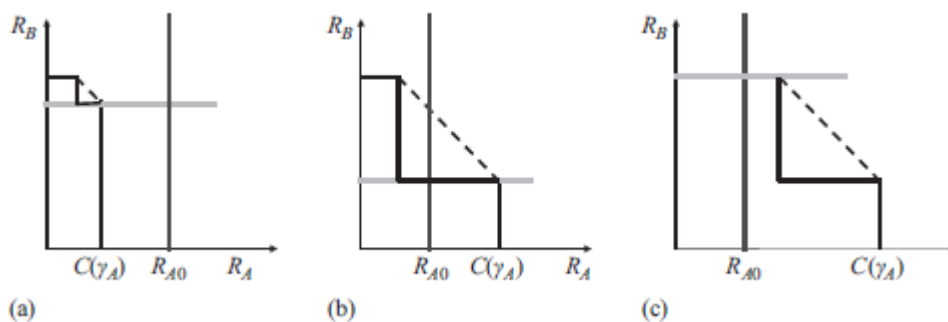
się. Pierwsza nierówność,  $R_{A,0} \leq C(\gamma_A)$  musi umożliwiać wspólne dekodowanie. Druga nierówność utrzymuje się, ponieważ  $R_B \leq C(\Gamma_B) \leq C(\gamma_B)$  Wreszcie trzecia nierówność zawiera także:

$$\begin{aligned} R_{A,0} + R_B &= C(\Gamma_{A,0}) + C(\Gamma_B) \\ &= C(\Gamma_{A,0}) + C(\gamma_B + \Gamma_{A,0}) - C(\Gamma_{A,0}) \\ &= C(\gamma_B + \Gamma_{A,0}) \leq C(\gamma_B + \gamma_A) \end{aligned}$$

gdzie ostatnia nierówność pochodzi od  $\Gamma_{A,0} \leq \gamma_A$ , jedno z wymagań stanu dekodowania ma miejsce. Zauważamy, że przedstawiony schemat komunikacji opiera się na konfiguracji teoretycznej i książkach kodowych. Dlatego, aby wprowadzić tę koncepcję w życie, można zastosować techniki transmisji, które są odpowiednie do dekodowania wielu użytkowników. W poniższych sekcjach skorzystamy z tego wyniku, aby umożliwić działanie podkładania.

### 6.2.3 Nieoptymalne dekodowanie podkładania

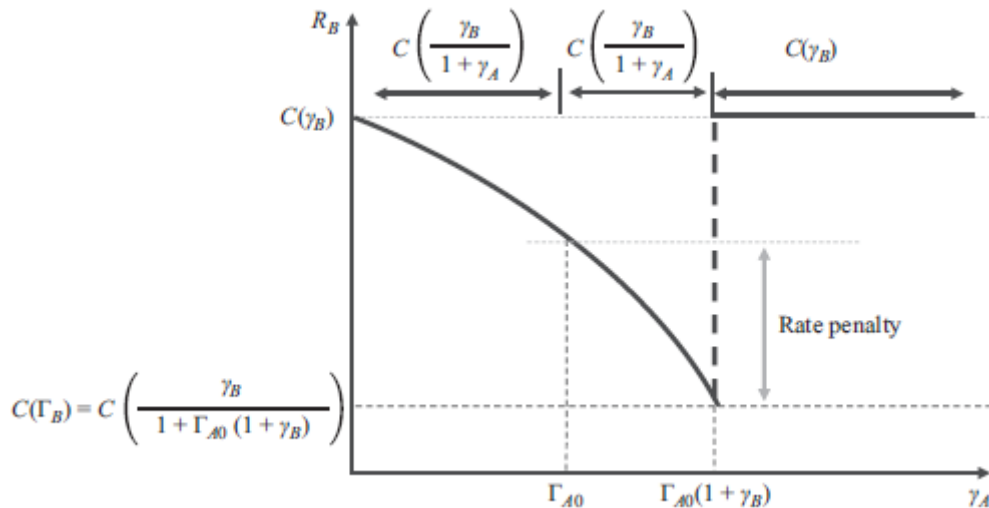
Rozważamy teraz przypadek, w którym odbiornik stosuje dekodowanie pojedynczego użytkownika (SD) w każdym etapie dekodowania, traktując jeszcze nieodtworzonego użytkownika jako szum. Teraz pokażemy jak wybrać maksymalną stawkę downlink dla danego  $\gamma_B$ , która gwarantuje brak przerwy. Przypomnijmy, że istnieją trzy regiony operacyjne odbiornika JD. Po zastosowaniu SD dostępne są tylko dwie opcje dekodowania, przedstawione na rysunku 6.9.



Jeśli połączenie między Alicją a odbiorcą:

- jest słaby, następnie  $\gamma_A$  jest niski i sygnał z B powinien zostać zdekodowany przez potraktowanie sygnał od Alice jako szum. Reżim ten przedstawiono na rys. 6.9 (a) i (b).
- jest bardzo silny, biorąc pod uwagę, że szybkość transmisji Alice jest ustalona na  $R_{A,0}$ ; następnie sygnał z Alice jest dekodowany, odejmowany, a sygnał "czysty" B jest dekodowany z maksymalną możliwą szybkością  $C(\gamma_B)$ . Można pokazać, że sygnał z Alice jest dekodowalny, a zatem może być traktowany jako silny, gdy  $\gamma_A = \Gamma_{A,0}(1 + \gamma_B)$  Ten reżim jest przedstawiony na rysunku 6.9 (c).

Następnie możemy zdefiniować  $R_B(\gamma_A)$  co ilustruje rysunek 6.10,



jako maksymalny osiągalne  $R_B$  dla danego  $\gamma_A$ , które można zapisać jako

$$R_B(\gamma_A) = \begin{cases} C\left(\frac{\gamma_B}{1 + \gamma_A}\right) & \gamma_A < \Gamma_{A0}(1 + \gamma_B) \\ C(\gamma_B) & \gamma_A \geq \Gamma_{A0}(1 + \gamma_B) \end{cases}$$

Twierdzenie 6.2.3 Niech będzie stała transmisja ze stopą  $R_A$  i niech odbiornik stosuje kolejne dekodowanie pojedynczego użytkownika. Jeśli B zna chwilową wartość SNR  $\gamma_B$ , ale nie  $\gamma_A$ , to maksymalna prędkość transmisji łączy w dół, która jest zawsze dekodowalna przez odbiornik jest  $R_B = C(\Gamma_B^{SD})$

gdzie

$$\Gamma_B^{SD} = \frac{\gamma_B}{1 + \Gamma_{A0}(1 + \gamma_B)}$$

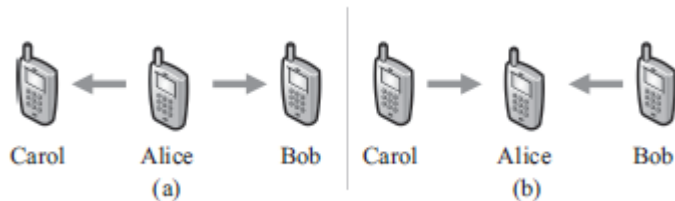
Dowód. Dowód, opiera się na znalezieniu punktu przełączania między regionem, w którym transmisja z B jest dekodowana, traktując transmisję z A jako szum, do regionu, w którym sygnał z A może zostać zdekodowany, a następnie odjęty od złożony sygnał, tak że sygnał z B może być dekodowany tylko w obecności szumu. Maksymalna transmisja w dół jest  $R_B = C(\Gamma_B^{SD})$ , następnie odpowiada temu samemu punktowi przełączania, który występuje, gdy  $\gamma_A = \Gamma_{A0}$

Oczywiście, zerowy wskaźnik przestoju w ustawieniu SD jest znacznie niższy niż w modelu Ustawienie JD.

### 6.3 Wykrywanie bliskich odpowiedzi

Przed ustanowieniem łączy D2D, sieć musi najpierw zostać poinformowana, jeśli tak urządzenia chcą się komunikować i są blisko siebie. W a Sieć komórkowa 3GPP, tę informację można wywnioskować z komplementarnej mechanizmy wykrywania zbliżeniowego zachodzące zarówno w Evolved Packet Core (EPC), jak iw sieci dostępu radiowego (E-UTRAN) . W EPC-oznaczone jako "EPC-poziom Discovery " - bliskość między urządzeniami można wyodrębnić z okresowych aktualizacji lokalizacji urządzenia, podczas gdy zainteresowanie nawiązaniem bezpośredniej komunikacji między urządzeniami może być wyzwalone przez urządzenia lub od strony sieci. Ta ostatnia może być uruchomiona z monitorowania

trwającej komunikacji między urządzeniami odbywające się za pośrednictwem infrastruktury sieciowej. Z drugiej strony, w E-UTRAN-u oznaczonym jako "bezpośrednie odkrycie" - bliskość jest ustalana przez bezpośrednią komunikację między zainteresowanymi urządzeniami. Istnieją obecnie dwa bezpośrednie protokoły wykrywania zbliżeniowego w 3GPP: Model A ("Jestem tutaj") i Model B ("Kto tam?"). W Modelu A protokoły wykrywania, jak pokazano na Rysunku 6.11 (a),

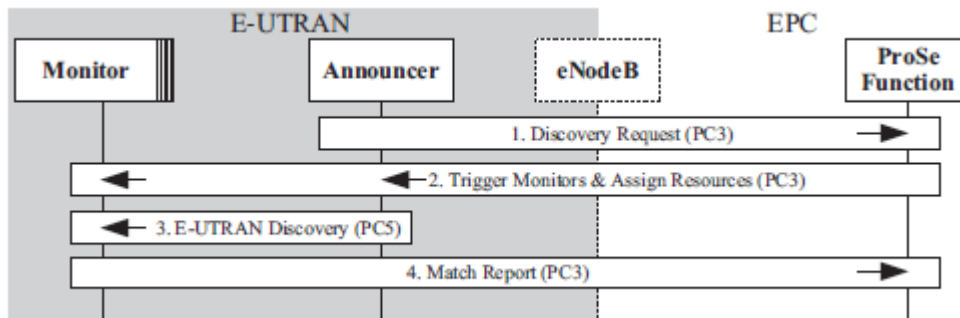


urządzenia komórkowe mogą pełnić rolę spikera lub monitora. Urządzenia w roli spikera nadają komunikat ogłaszający informacje, które mogą być interesujące dla monitorów w jego bliskim regionie. Z drugiej strony, urządzenia w roli monitora nasłuchują pewnych interesujących informacji przesyłanych przez spikerów w ich pobliżu. Transmisja spikera jest zwymiarowana za pomocą stałej mocy i szybkość, z jaką wykrywają to urządzenia w obszarze bliskim wysokiego prawdopodobieństwa. W protokole wykrywania modelu B, jak pokazano na rys. 6.11 (b), urządzenia komórkowe mogą pełnić rolę odkrywcy lub odkrywającego. W tym modelu odkrywca otrzymuje informacje z pobliskich odkryć. W przeciwieństwie do Modelu A, ten protokół jest oparty na agregacji transmisji z pobliskiego urządzenia komórkowe, które niekoniecznie są znane a priori przez sieć znajdująca się w pobliżu regionu.<sup>2</sup> W pozostałej części tej sekcji skoncentrujemy naszą analizę na ustawieniu protokołu wykrywania modelu A. Podejście zarządzania zasobami radiowymi w przypadku Modelu A polega na zaplanowaniu dedykowanych zasobów dla transmisji komunikatora, tj. zasobów ortogonalnych do tych przypisanych do innych aktywnych połączeń komórkowych. W przeciwieństwie do tego rozważamy przypadek, w którym transmisja spikera ma być podstawą transmisji ciągłego łącza w dół do monitorów. Prowadzi to do utworzenia kanału wielodostępowego, złożonego przez komunikator i transmisje BS. Występowanie ogniw komunikacyjnych typu D2D, choć wcześniej rozważane w literaturze, nie było dotąd brane pod uwagę w otoczeniu odkrywania bliskości do czasu naszej ostatniej pracy. Główną przeszkodą dla odkrywania bliskich elementów jest to, że stan kanału między spikerem a urządzeniem monitorującym jest nieznan, co powoduje, że nieznan jest poziom zakłóceń transmisji łącza w dół, a zatem uniemożliwia eNodeB radzenie sobie z nim. Jednakże, jak omówiono w oraz w rozdziale 6.2, wykazaliśmy, że jeśli sygnał zakłócający wykorzystuje stałą częstotliwość, a odbiornik łącza odbiorczego wykorzystuje dekodowanie wielu użytkowników, wtedy możliwe jest odbieranie transmisji łącza nadawczego z zerową awarią, bez konieczności BS, aby mieć wszelkie wcześniejsze informacje o stanie kanału Channel na stronie Nadajnik (CSIT) łącza interferującego. Sygnały zastosowane w procesie wykrywania wykorzystują stałą szybkość, co sprawia, że są idealnymi kandydatami do podkładania, gwarantując zerową awarię połączenia w dół. W tej części przeformułujemy wyniki w sekcji 6.2 pod względem adaptacji mocy, tak aby w urządzeniach monitorujących osiągnięto transmisję zstępującą w dół z zerową amplitudą. Analiza i wyniki liczbowe pokazują, że operacja podkładania jest rzeczywiście możliwa, ale kosztem zwiększonego zużycia energii łącza w dół, które zwiększa się z prędkością transmisji podkładania.

### 6.3.1 Przegląd wykrywania bliskości sterowany siecią

W protokole wykrywania modelu A zaangażowane są trzy podmioty sieciowe, mianowicie spiker, monitor i funkcja ProSe (znajdujące się w infrastrukturze sieciowej w EPC). Funkcja ProSe jest odpowiedzialna za kontrolę dostępu do ProSe, koordynację wykrywania bliskości i monitorowanie ciągłych połączeń komunikacyjnych D2D. Rysunek 6.12 przedstawia cztery zasadnicze kroki, które

podejmuje konferansjer, aby ukończyć protokół wykrywania bliskości, po autoryzacji do używania ProSe.

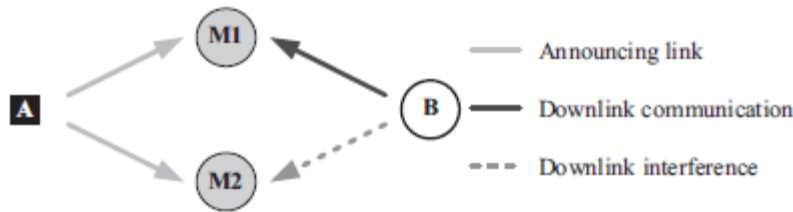


W kroku Discovery Request konferansjer wysyła żądanie wykrywania do funkcji ProSe. Jeśli to żądanie zostanie zaakceptowane, etap Trigger Monitors & Assign Resources ma miejsce, gdy funkcja ProSe uruchamia monitory w regionie bliskim spikera, przypisując i informując interweniujące urządzenia (spiker i monitory), które zasoby systemowe zostały przydzielone do transmisji komunikat komunikatora, tzn. zasoby radiowe, na których będzie odbywał się interfejs PC5. W etapie E-UTRAN Discovery konferansjer przesyła swój komunikat do monitorów za pośrednictwem interfejsu PC5 w E-UTRAN. Na koniec, w kroku Raport dopasowania, monitory, które mogły pomyślnie zdekodować komunikat kontrahenta, przekazują odebrane informacje do funkcji ProSe w celu potwierdzenia tożsamości spikera. Po tym ostatnim kroku można ustalić bezpośrednie łącze komunikacyjne D2D. W pozostałej części skupiamy się na trzecim kroku protokołu wykrywania. Mianowicie, przy przesyłaniu komunikatu konferenckiego stanowiącego podstawę transmisji w dół skierowanej do jednego z monitorów.

### 6.3.2 Model systemu

Koncentrujemy naszą analizę na sieci komórkowej złożonej ze stacji bazowej i wielu urządzeń komórkowych. Zestaw tych urządzeń pełni rolę konferansjera, podczas gdy pozostałe urządzenia przejmują rolę monitorującą, słuchając komunikatu konferansjera w zasobach przydzielonych do tego celu przez sieć. Zakładamy, że spikerowie są wystarczająco rozproszeni w przestrzeni tak, że ich obszary bliskości nie pokrywają się<sup>3</sup> i że sieć przypisuje różne zasoby każdemu spikerowi. Przyjmuje się, że sieć ma wiedzę na temat pozycji każdego urządzenia za pomocą okresowych aktualizacji lokalizacji urządzenia, nie znając warunków kanału między dowolnymi dwoma urządzeniami przed procedurą wykrywania bliskości. Dlatego sieć wie, które monitory będą potencjalnie znajdować się w bliskim regionie.<sup>4</sup> Przyjmujemy, że w pobliżu obszaru spikera A znajdują się pobliskie monitory N. Dalej zakładamy, że każdy z tych monitorów ma dedykowaną ciągłą transmisję w dół od stacji bazowej B. Wyznaczamy powiązany kanał łącza nadawczego i jego docelowy odbiornik przez ten sam indeks dolny, np. I-tym odbiornikiem docelowym jest i-to urządzenie. Zakładamy, że te transmisje w łączy w dół mają stałą szybkość RB; do spełnienia z zerową awarią. Założone stałe założenie stawki łącza upraszczającego upraszcza wprowadzenie naszej koncepcji sterowania zasilaniem, ale zauważamy, że możliwe jest rozszerzenie adaptacyjnego ustawienia szybkości łącza nadawczego. Wreszcie, zakładamy system komórkowy oparty na OFDMA, gdzie (i) synchronizacja między urządzeniami sieciowymi jest zapewniona przez cykliczny prefiks w każdej transmisji; (ii) urządzenia są w stanie odbierać i dekodować wszystkie transmisje w podnośne łącza w dół; i (iii) urządzenia okresowo raportują do kanału B jakości dla wszystkich kanałów łącza pobierania, które są szacowane na podstawie sygnałów pilota kanału łącza nadawczego. Jeśli założymy, że konferansjer transmituje w kanale w górę, jak zilustrowano na rysunku 6.13,





następnie sygnał zespolony  $y_{k,i}$  odebrany w  $k$ -tym kanale w  $i$ -tym urządzeniu to

$$y_{k,i} = h_{A,i}x_A + h_{B,i}x_B + Z$$

gdzie  $h_{B,i}$  i  $h_{A,i}$  oznaczają, odpowiednio, złożony zysk kanału między nadajnikami B i A i  $i$ -tym odbiornikiem, podczas gdy  $Z$  jest szumem Gaussa o wartościach zespolonych  $E[|Z|^2] = \sigma^2$  i  $x_B$  i  $x_A$  są podane przez okrągły zerowy średni zespolony sygnał Gaussa transmitowany przez A i B, gdzie odpowiednie wariancje są  $E[|x_A|^2] = P_A$  i  $E[|x_B|^2] = P_B$ . Przyjmuje się, że  $P_A$  jest stałe i zwymiarowane w taki sposób, że monitory w promieniu regionu zbliżeniowego są w stanie wykryć transmisję spikerów z bardzo dużym prawdopodobieństwem. Jak wcześniej wspomniano, zakłada się, że transmisja w łączu w dół ma stałą częstotliwość  $R_B$ , która ma być zerowa. Aby spełnić to wymaganie,  $P_B$  jest dostosowywany za pomocą schematu sterowania mocą opisanego w sekcji 6.3.3, który radzi sobie z warunkami kanału łączu w dół i interferencją z transmisji komunikatora z podkładaniem. Schemat podstawowy zakłada ustawienie ortogonalne, w którym kanały łączu downlink mają tylko transmisję łączu downlink, tak, że odebrany sygnał z  $i$ -tego urządzenia jest określony przez

$$y_{j,i} = h_{B,i}x_B + Z \quad z \quad j \neq k$$

Zakładamy, że B ma stałą szybkość transmisji w dół i jest w stanie dostosować swoją moc transmisji w oparciu o informacje o stanie kanału dostępne w nadajniku (CSIT) z B -  $M_i$  linki. Dalej zakładamy, że brak jakiegokolwiek informacji o stanie kanału na temat łączu A -  $M_i$ , ponieważ zakłada się, że nie nastąpiła żadna komunikacja między A i  $M_i$  przed transmisją spikera. Z drugiej strony zakłada się, że B zna stałą szybkość transmisji powiązaną z transmisją z bliskości, oznaczoną jako  $R_A$ .

### 6.3.3 Analiza

W naszej analizie skupiamy się na wydatkach mocy transmisji łączu w dół wymaganych do umożliwienia działania podkładu protokołu bezpośredniego wykrywania zbliżeniowego. Wyniki obliczono jako zero-outage w Lemacie 6.2.1 do ustawienia sterowania mocą, zakładając że każda transmisja łączu w dół ma stałą stawkę  $R_B$ . Zaczynamy od analizy przypadku gdzie jest jedno urządzenie monitorujące, a następnie rozszerzyć analizę do wielu urządzeń monitorujących. Na koniec, rozszerzamy analizę na przypadek, w którym sieć wybiera jeden z dostępnych kanałów zstępujących, które mają być poddane bezpośredniemu protokołowi odkrywania bliskości, w oparciu o minimalną wymaganą moc transmisji łączu nadawczego.

#### 6.3.3.1 Pojedyncze urządzenie monitorujące

Zaczynamy od wprowadzenia mechanizmu kontroli mocy dla transmisji w dół, zakładając pojedynczy monitor w regionie bliskim. Definiujemy natychmiastową SNR transmisję w łączu w dół

$$\gamma_B = \frac{P_B g d_B^{-\alpha}}{\sigma^2}$$

gdzie  $d_B^{-\alpha}$  jest utratą ścieżki,  $\sigma^2$  jest mocą szumu,  $g = |h|^2$  jest realizacją wzmocnienia kanału łącza nadawczego i  $P_B$  mocą transmisji łącza downlink. Tak więc pojemność takiego otrzymanego kanału Gaussa to

$$\begin{aligned} R_B &= W \log_2 \left( 1 + \frac{\gamma_B}{1 + \Gamma_A} \right) \\ &= W \log_2 \left( 1 + \frac{P_B g d_B^{-\alpha}}{\sigma^2} \cdot \frac{1}{1 + \Gamma_A} \right) \end{aligned}$$

gdzie  $W$  jest przepustowością przypisanych zasobów. Termin  $1/(1 + \Gamma_A)$  konta dla stawki kary wynikającej z transmisji komunikatora z podkładem, aby uzyskać gwarancję osiągnięcia zerowej przerwy w łączu "w dół", jak podano w Lemacie 6.2.1. Wymagana moc  $P_B$  do osiągnięcia stałej stopy  $R_B$  jest następnie podawana przez

$$P_B = \frac{1}{g} \cdot K$$

gdzie

$$K = \frac{\sigma^2(1 + \Gamma_A)}{d_B^{-\alpha}} \cdot \left[ 2^{R_B/W} - 1 \right]$$

Aby poradzić sobie z głębokimi zanikami, wprowadzono skróconą strategię inwersji kanałów. W tym ustawieniu B będzie próbować transmisji tylko wtedy, gdy głębokość zanikania przekroczy próg odcięcia  $\mu$  tj. gdy  $g > \mu$  zdarzenie wyłączenia tego łącza jest następujące:

$$Pr\{g < \mu\} = \int_0^{\mu} f_g(x) dx = 1 - \exp(-\mu/\bar{g})$$

gdzie  $\mu = -\bar{g} \log(1 - Pr\{g < \mu\})$ . Zauważamy, że zawsze występuje transmisja spikera, chociaż transmisja w dół może nie nastąpić z powodu strategii skracania mocy. Innymi słowy, transmisja spikera jest dekodowalna z prawdopodobieństwem  $Pr(\gamma_A \geq \Gamma_A)$ . Średnia moc nadawania wymagana do wsparcia  $R_B$ , to:

$$E[P_B] = \int_{\mu}^{\infty} \frac{1}{g} \cdot K f_g(g) dg = E_1\left(\frac{\mu}{\bar{g}}\right) \cdot K$$

gdzie  $E(\cdot)$  jest wykładniczą funkcją całkową.

### 6.3.3.2 Wiele urządzeń monitorujących

Gdy istnieje wiele urządzeń monitorujących w obszarze bliskości, które odbierają ruch emisji z BS, to  $P_B$  musi zostać obliczone dla najgorszego łącza. Dzieje się tak, ponieważ  $R_B$  musi być dekodowalny na wszystkich monitorach w regionie bliskim. Wymagana moc staje się:

$$P_B = \frac{1}{\min(g_1, \dots, g_N)} \cdot K$$

gdzie  $g_i$  jest wzmocnieniem kanału łączą w dół i-tego monitora i gdzie, dla analizy ustępliwości, zakłada się, że  $K$  jest identyczne dla wszystkich monitorów. Aby obliczyć wartość odcięcia, wprowadzamy rozkład statystyk rzędu  $n$ :

$$f_n(x) = Nf_g(x) \binom{N-1}{n-1} F(x)^{n-1} (1-F(x))^{N-n}$$

z którego chcemy  $n = 1$ , rozkład najgorszego kanału:

$$f_1(x) = \frac{N}{\bar{g}} \exp\left(-\frac{Nx}{\bar{g}}\right)$$

Zdarzenie wyłączenia jest następnie definiowane jako

$$\begin{aligned} Pr\{\min(g_1, \dots, g_N) < \mu\} &= \int_0^\mu f_1(x) dx \\ &= 1 - \exp(-N\mu/\bar{g}) \end{aligned}$$

ai powiązany próg odcięcia jako

$$\mu = -\frac{\bar{g}}{N} \log(1 - Pr\{\min(g_1, \dots, g_N) < \mu\})$$

Średnie wymagane  $P_B$  w tym ustawieniu jest następnie podawane przez

$$E[P_B|N] = E_1\left(\frac{N\mu}{\bar{g}}\right) \cdot N \cdot K$$

### 6.3.3.3 Wybór kanału zstępującego do podkładania

Teraz rozważamy przypadek, w którym istnieje  $M$  kanałów downlink, z których sieć może wybrać jedną, która ma być poddana. W tym ustawieniu kanał wybrany dla operacji podkładania to taka, która wymaga najmniej zużytej mocy. Zakładamy, że każdy kanał łączą w dół ma różne wzmocnienia dla każdego monitora. Załóżmy, że z monitorów  $N$ , jeden z najgorszym wzmocnieniem kanału w  $i$ -tym kanale  $j$  dany przez  $u_i = \min(g_{1,i}, \dots, g_{N,i})$ , gdzie  $g_{n,i}$  odpowiada wzmocnieniu kanału łączą nadawczego z kanału  $i$ -tego w urządzeniu  $n$ . Następnie kanał, który doprowadzi do najmniejszego nakładu energii jest jednym z zyskiem danego kanału przez  $u_{\max} = \max(u_1, \dots, u_M)$ . Innymi słowy, przy założeniu, że monitor z najgorszych warunków w każdym kanale jest czynnikiem ograniczającym, a następnie wybór najlepszego kanału w tych warunkach jest modelowane przez rozkład statystyczny zamów maksymalnej wartości. Odpowiednia funkcja gęstości,  $f_M(x)$  jest dany przez,

$$f_M(x) = Mf_1(x)[F_1(x)]^{M-1}$$

a funkcja skumulowana  $F_M(x)$  jest podawana przez

$$F_M(x) = [F_1(x)]^M = \left[1 - \exp\left(-\frac{Nx}{\bar{g}}\right)\right]^M$$

Zdarzenie wyłączenia jest następnie definiowane przez

$$Pr\{u_{max} < \mu\} = F_M(\mu) = \left(1 - \exp\left(-\frac{N\mu}{\bar{g}}\right)\right)^M$$

gdzie

$$\mu = -\frac{\bar{g}}{N} \log(1 - Pr\{u_{max} < \mu\}^{1/M})$$

W tych warunkach średnia moc zużywana jest przez

$$E[P_B|M] = K \int_{\mu}^{\infty} \frac{1}{x} f_M(x) dx$$

które można obliczyć za pomocą integracji numerycznej.

Stawka za każde wykorzystane zasoby,  $S$ , koncentruje się na najgorszym kanale, jest podana przez,

$$S = R_B \cdot Pr\{u_{max} \geq \mu\} + R_A \cdot Pr\{\gamma_A \geq \Gamma_A\}.$$

Jako miarę efektywności energetycznej bierzemy pod uwagę wymaganą ilość energii wymaganej do przekazania trochę informacji,  $\psi$  podane przez,

$$\psi = E\left[\frac{P_B}{S}\right] = \frac{K}{S} \int_{\mu}^{\infty} \frac{1}{x} f_N(x) dx$$

które można obliczyć za pomocą integracji numerycznej.

#### 6.3.3.4 Ortogonalna kontrola mocy łącza w dół

Teraz rozważamy linię podstawową jako ustawienie, w którym spiker i link w dół transmisje są przydzielane zasoby ortogonalne przez sieć. Innymi słowy, transmisja w dół i transmisje z komunikatora występują w dwóch identycznych kanałach ortogonalnych. Począwszy od pojemności asymptotycznej kanału Gaussa na  $i$ -tym urządzeniu:

$$R_B = W \log_2 \left(1 + \frac{P_B g_i d_B^{-\alpha}}{\sigma^2}\right)$$

wymagana moc pobierana w dół  $P_B$  w celu osiągnięcia stałej stopy  $R_B$  jest następnie podawana przez

$$P_B = \frac{1}{g_i} \cdot \frac{\sigma^2 [2^{R_B/W} - 1]}{r^{-\alpha}}$$

Aby zapewnić rzetelne porównanie, skupiamy się na mocy wymaganej do obsługi najgorszego kanału zstępującego. Średnia moc nadawania wymagana do utrzymania  $R_B$ , jest wtedy:

$$E[P_B] = \int_{\mu}^{\infty} \frac{1}{x} \cdot \frac{\sigma^2 [2^{R_B/W} - 1]}{d_B^{-\alpha}} f_M(x) dx$$

które można obliczyć numerycznie. Stopa procentowa na każdy wykorzystany zasób  $S$ , od najgorszego kanału i kanału przeznaczanego na transmisję urządzenia przekazującego, podana jest przez

$$S = \frac{R_B \cdot Pr\{u_{max} \geq \mu\} + R_A \cdot Pr\{\gamma_A \geq \Gamma_A\}}{2}$$

Zauważ, że dzielenie przez 2 pochodzi od komunikatora, a transmisje w dół odbywają się w dwóch identycznych kanałach ortogonalnych. Związana z tym efektywność energetyczna,  $\psi$ , jest podana przez,

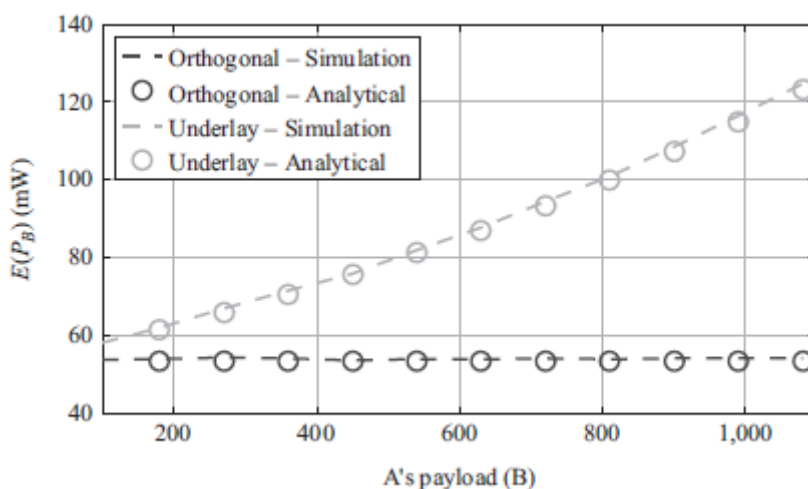
$$\psi = E \left[ \frac{P_B}{S} \right] = \frac{1}{S} \int_{\mu}^{\infty} \frac{1}{x} \cdot \frac{\sigma^2 [2^{R_B/W} - 1]}{d_B^{-\alpha}} f_M(x) dx.$$

### 6.3.4 Wyniki liczbowe i dyskusja

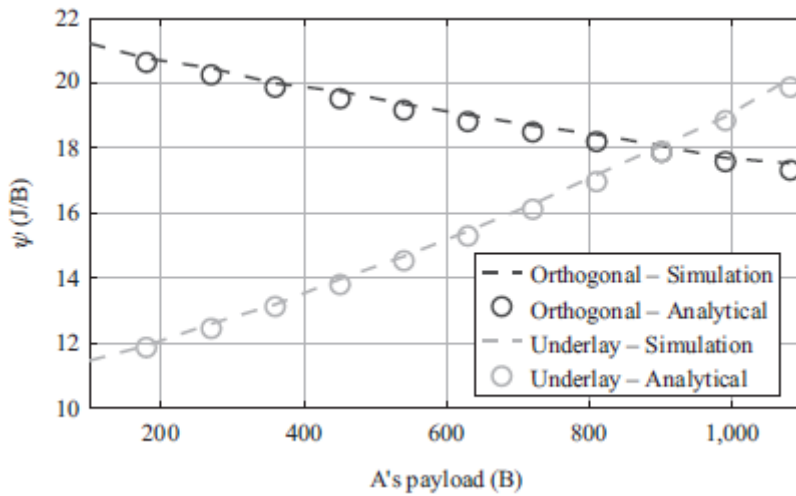
W tej sekcji przedstawiamy wyniki liczbowe odpowiadające wydajności wykrywania zbliżeniowego w ustawieniu podkładania, pod względem wydatkowanej mocy  $E[P_B]$  i efektywności energetycznej  $\psi$  w zależności od ładowności spikera (obliczonej jako  $R_A \cdot W \cdot T_A$ , gdzie  $T_A$  jest czas trwania transmisji spikera). Te wyniki liczbowe uzyskano przez ocenę odpowiednich wyrażeń analitycznych, uzyskanych w poprzedniej sekcji i za pomocą stochastycznych symulacji Monte Carlo. Tabela 6.1 zawiera listę odpowiednich założonych parametrów systemu.

Parametr	Wartość	Parametr	Wartość
$W$	180 kHz	$T_A$	5 ms
$d_B$	200 m	$d_A$	20 m
$P_A$	20 dBm	$\sigma^2$	-97 dBm
$\alpha$	4	$Pr\{\gamma_A \geq \Gamma_A\}$	0.99
$R_B$	5 B/s <sup>2</sup>	$Pr\{u_{max} \geq \mu\}$	0.99
$N, M$	20	A's payload	[100,1,100] (B)

Zauważmy najpierw, że suma osiągnięta w ustawieniu podkładania będzie zawsze wyższa niż w ustawieniu zasobów ortogonalnych Ta wyższa stawka sumy wiąże się z kosztami dodatkowych wydatków związanych z transmisją w dół w porównaniu z ustawieniem zasobów ortogonalnych, jak pokazano na rysunku 6.14.



W szczególności, w ustawieniu podkładania, im wyższa jest szybkość wywoławcza (i odpowiadająca jej ładowność), tym wyższa będzie wymagana moc transmisji łącza nadawczego. Z drugiej strony, z perspektywy efektywności energetycznej, na Rysunku 6.15



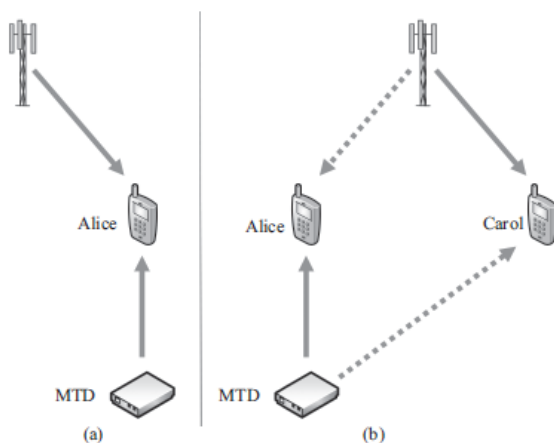
zauważamy, że dla niższych ładunków spikera i pomimo wyższej wymaganej mocy łącza w dół, schemat podkładania jest bardziej energooszczędny. Jest tak w tym przypadku, ponieważ kara wprowadzona wyrazem  $(1 + \Gamma_A)$  dotyczy ładunków, które są wystarczająco niskie, aby nie zrównoważyć w sposób znaczący przyrostów sumy w ustawieniu podkładania.

### 6.3.5 Uwagi

Wykrywanie zbliżeniowe jest istotnym czynnikiem ułatwiającym komunikację D2D. Pokazaliśmy, że wiarygodne wykrywanie bliskich podkładów jest wykonalne, osiągając wyższe sumy niż w ustawieniu zasobów ortogonalnych, kosztem większej ilości zużytej mocy łącza pobierającego. Z drugiej strony zaobserwowaliśmy, że w przypadku niewielkich ładunków komunikatora odkrycie podkładu jest dwa razy bardziej energooszczędne w porównaniu z ustawieniem prostopadłym.

### 6.4 Podkład bezpośredniej komunikacji

Bezpośrednia komunikacja pomiędzy dwoma urządzeniami ma miejsce, gdy są one względem siebie proksymalne, co można określić za pomocą opisanej wcześniej procedury. Rozważamy teraz przypadek, w którym te bezpośrednie połączenia opierają się na transmisji zstępującej. Rozważmy dwie topologie podkładania przedstawione na rysunku 6.16,



gdzie różnica występuje w urządzeniu, które działa jako zamierzony odbiornik transmisji w dół. Aby włączyć te dwie topologie komunikacji podkładania, wykorzystujemy wynik zerowego zaniku łącza w dół dla wspólnego dekodowania, jako prezentowana w punkcie 6.2. Oznaczamy topologię na rysunku 6.16 (a) i (b) odpowiednio jako topologie A i B. Rozpoczynamy od analizy topologii A, pod względem wymaganej mocy transmisji łącza nadawczego, aby osiągnąć taką samą wydajność transmisji, jak w przypadku bez przekładania podkładu. Celem jest zapewnienie, że transmisja w łączu w dół ma zerową awarię, podczas gdy transmisja w złączu odbiorczym nie wpływa na transmisję MTD. Te dwa warunki są zapewnione przez zastosowanie wyników w Lemacie 6.2.1 i Następstwie 6.2.2.

#### 6.4.1 Analiza topologii A

W tej topologii sieci, jak pokazano na Rysunku 6.16 (a), Alicja odbiera ruch zstępujący z BS i z MTD. Aby uprościć notację, oznaczamy odpowiednio BS i MTD jako B i M. Związek pomiędzy MTD i Alice charakteryzują się stałą szybkością i mocą, oznaczane odpowiednio jako  $R_M$  i  $P_M$ : MTD transmituje do Alicji podczas transmisji łącza w dół B, tak, że odbiornik Alice obserwuje kanał wielokrotnego dostępu (MAC) z dwoma nadajnikami:

$$y_A = h_{B,A}x_{B,A} + h_{M,A}x_M + Z$$

gdzie  $h_{B,A}$  i  $h_{M,A}$  to złożone zyski kanałów między BS i Alice, oraz między MTD i Alice, odpowiednio.  $x_B$  i  $x_M$  są podane przez okrągłe zero-średnie kompleksowe sygnały Gaussa transmitowane przez węzły B i M, tak, że odpowiednie wariancje to  $E[|x_B|^2] = P_B$  i  $E[|x_M|^2] = P_M$ . Tutaj  $P_B$  i  $P_M$  reprezentują poziomy mocy stałej używane odpowiednio przez B i M. Wreszcie,  $Z$  jest szumem Gaussa o wartościach zespolonych  $E[|Z|^2] = \sigma^2$ . Poniżej określimy mechanizm sterowania mocą dla transmisji w dół, która umożliwi transmisję łącza nadawczego do osiągnięcia takiego samego poziomu jak w przypadku bez transmisji podkładania. Definiowanie chwilowej wartości SNR transmisji w łączu w dół:

$$\gamma_B = \frac{P_B g d_B^{-\alpha}}{\sigma^2}$$

gdzie  $d_B^{-\alpha}$  to straty strat ścieżek,  $\sigma^2$  to moc szumów,  $g = |h|^2$  jest łączem w dół realizacją wzmocnienia kanału i  $P_B$  mocy transmisji łącza downlink. Biorąc pod uwagę górną granicę współczynnika dekodowania zera w punkcie zerowym, możemy wyrazić zerową stopę połączenia zstępującego jako

$$R_B = W \log_2 \left( 1 + \frac{P_B g d_B^{-\alpha}}{\sigma^2} \cdot \frac{1}{1 + \Gamma_A} \right)$$

które można przekształcić w:

$$P_B = \frac{1}{g} \cdot \frac{\sigma^2 (1 + \Gamma_A)}{d_B^{-\alpha}} \cdot \left[ 2^{R_B/W} - 1 \right]$$

Teraz zauważając, że osiągalna szybkość łącza nadawczego  $R_B$ , gdy nie ma podkładania obecny jest

$$\dot{R}_B = W \log_2 \left( 1 + \frac{\dot{P}_B g d_B^{-\alpha}}{\sigma^2} \right)$$

gdzie  $P_B$  jest oryginalną mocą transmisji w dół. Następnie:

$$P_B = \frac{1}{g} \cdot \frac{\sigma^2(1 + \Gamma_A)}{d_B^{-\alpha}} \cdot \left[ 2^{W \log_2 \left( 1 + \frac{P_B g d_B^{-\alpha}}{\sigma^2} \right) / W} - 1 \right]$$

$$= (1 + \Gamma_A) \dot{P}_B.$$

Dodatkowa moc pobierana w dół wymagana do zapewnienia, że transmisja łącza w dół jest w stanie osiągnąć tę samą szybkość łącza w dół, jest proporcjonalna do  $(1 + \Gamma_A) = 2^{R_A/W}$ . Innymi słowy, wymagana moc wzrośnie wykładniczo wraz z prędkością transmisji podkładania. Ten schemat podkładania jest zatem odpowiedni dla aplikacji o niskiej szybkości, zwykle związanych z MTC.

#### 6.4.2 Analiza topologii B

Rozważamy teraz topologię sieci przedstawioną na rysunku 6.16 (b), gdzie łącza komunikacyjne między BS i Carol oraz między MTD i Alice występują jednocześnie w szczelinie czasowej. Podobnie jak w poprzedniej topologii, związek między MTD i Alice charakteryzuje się stałą stawką i mocą, oznaczane odpowiednio jako  $R_M$  i  $P_M$ : podczas gdy związek między BS a Carol charakteryzuje się zmienną szybkością i mocą stałą, zwane dalej, odpowiednio, jako  $R_B$  i  $P_B$ . MTD transmituje do Alicji podczas transmisji B łącza w dół, tak że odbiorniki Alice i Carol obserwują kanał wielokrotnego dostępu (MAC) z dwoma nadajnikami:

$$y_A = h_{M,A}x_M + h_{B,A}x_B + Z, \quad y_C = h_{M,C}x_M + h_{B,C}x_B + Z$$

gdzie  $h_{B,A}$ ;  $h_{B,C}$ ;  $h_{M,C}$  i  $h^{M,A}$  oznaczają złożone zyski kanałów na danym kanale między odpowiednimi węzłami.  $x_B$  i  $x_M$  są odpowiednio sygnałami transmitowanymi przez B i MTD, tak że  $E[|x_B|^2] = P_B$  i  $E[|x_M|^2] = P_M$ .  $Z$  to zespolony szum Gaussa z wariancją  $\sigma^2$ : Przypomnijmy, że chcemy zapewnić, że transmisja w dół nie wpływa na transmisję podkładową MTD. Jest to ułatwione przez wynik zerowej przerwy w stosowności 6.2.2, która wymaga dekodowania transmisji w dół w odbiorniku Alice. Dlatego też wymagana moc transmisji łącza nadawczego jest obliczana na podstawie najgorszego łącza między BS a Alice lub Carol. Wymagana moc transmisji staje się wtedy obliczana na podstawie najgorszego powiązania między BS a Alice lub Carol. Wtedy wymagana moc transmisji staje się

$$P_B = \frac{1}{\min(g_A, g_C)} \cdot \frac{\sigma^2(1 + \Gamma_A)}{d^{-\alpha}} \cdot \left[ 2^{R_B/W} - 1 \right]$$

gdzie odpowiednio  $g_A$  i  $g_C$ , kanał łącza w dół zyskuje w kierunku Alice i Carol. W tym scenariuszu istnieje możliwość, że najgorsze wzmocnienie kanału łącza jest zbyt niskie, prowadząc do nadmiernej dużej ilości mocy, aby osiągnąć wymaganą szybkość łącza. Aby sobie z tym poradzić przy tych głębokich zanikach kanału przyjmujemy schemat sterowania mocą oparty na obciążeniu tekście odwrócenie kanału]. W tym ustawieniu stacja BS będzie próbować transmisji tylko wtedy, gdy głębokość zanikania jest powyżej progu odcięcia  $m$ ; tj. gdy  $\min(g_A, g_C) > \mu$ . Do konta dla wartości odcięcia wprowadzamy rozkład statystyk  $n$ -tego porządku:

$$f_n(x) = 2f_g(x) \binom{2-1}{n-1} F(x)^{n-1} (1 - F(x))^{2-n}$$

z którego chcemy  $n = 1$ , rozkład najgorszego kanału:



$$f_1(x) = \frac{2}{\bar{g}} \exp\left(-\frac{2x}{\bar{g}}\right)$$

Zdarzenie wyłączenia jest następnie definiowane jako

$$\begin{aligned} Pr\{\min(g_A, g_C) < \mu\} &= \int_0^\mu f_1(x) dx \\ &= 1 - \exp(-2\mu/\bar{g}) \end{aligned}$$

i powiązany próg odcięcia jako

$$\mu = -\frac{\bar{g}}{2} \log(1 - Pr\{\min(g_A, g_C) < \mu\})$$

Zakładamy, że w przypadku awarii, transmisja w dół nie następuje ani w Alice, ani w Carol. Teraz zauważając, że osiągalna szybkość pobierania  $R_B$ , gdy nie ma podkładu, jest,

$$\dot{R}_B = W \log_2 \left( 1 + \frac{\dot{P}_B g_C d_B^{-\alpha}}{\sigma^2} \right)$$

gdzie  $P_B$  jest oryginalną mocą transmisji w dół. Następnie:

$$\begin{aligned} P_B &= \frac{1}{\min(g_A, g_C)} \cdot \frac{\sigma^2(1 + \Gamma_A)}{d_B^{-\alpha}} \cdot \left[ 2^{W \log_2 \left( 1 + \frac{\dot{P}_B g_C d_B^{-\alpha}}{\sigma^2} \right) / W} - 1 \right] \\ &= (1 + \Gamma_A) \cdot \frac{g_C}{\min(g_A, g_C)} \cdot \dot{P}_B. \end{aligned}$$

Dodatkowa moc wymagana do spełnienia tej samej stawki połączenia z dostawą, jak w przypadku bez podkładania, zależy od najgorszego wykonania kanału. Przeciętnie moc potrzebna dodatkowa moc jest następnie przez

$$\begin{aligned} E \left[ \frac{P_B}{\dot{P}} \right] &= \int_0^\infty \int_\mu^\infty (1 + \Gamma_A) \frac{y}{x} f_1(x) f_g(y) dx dy \\ &= (1 + \Gamma_A) \cdot 2 \cdot E_1(2\mu) \end{aligned}$$

gdzie  $E_1(x)$  jest wykładniczą funkcją całkową. Podobnie jak w topologii A, wymagana dodatkowa moc downlink jest proporcjonalna do  $2^{RA/W}$ . Dlatego ten schemat podkładania jest również odpowiedni dla aplikacji o niskiej szybkości, takich jak te związane z MTC.

# **Buforowanie w dużych sieciach bezprzewodowych**

## **7.1 Wprowadzenie**

Ponad miliard abonentów sieci bezprzewodowych dziś i trzykrotne potrojenie tej liczby w ciągu następnych pięciu lat sprawiają, że branża ICT staje w obliczu lawiny zapotrzebowania na ruch. Bezprecedensowy wzrost natężenia ruchu w sieciach bezprzewodowych to efekt uboczny rewolucji informatycznej, której doświadczamy, napędzany przez rozprzestrzenianie się inteligentnych urządzeń mobilnych i chciwych aplikacji wideo. Kwartalny raport z monitoringu Internetu firmy Akamai ujawnia, że ruch sieciowy osiągnął skalę exabyte i stale rośnie. Z raportów firmy Cisco wynika, że ruch mobilny ma coraz większy udział w rynku, a wkrótce osiąga 60%. Dlatego dużym wyzwaniem dla nadchodzących sieci bezprzewodowych piątej generacji (5G) jest sprostanie rosnącemu zapotrzebowaniu na ruch sieciowy. Ustalony fundamentalny wynik stwierdza, że długa multi-bezprzewodowa komunikacja nie jest skalowana [2], tj. Maksymalna wspólna częstotliwość ruchu dla wszystkich strumieni jest odwrotnie proporcjonalna do średniej liczby przeskoków. W rezultacie proponowane techniki radzenia sobie z wyzwaniami komunikacyjnymi przyszłych sieci koncentrują się głównie na dalszym zwiększaniu przepustowości łącza i zagęszczaniu sieci bezprzewodowej za pomocą większej liczby punktów dostępu (AP) lub stacji bazowych (BS), które są połączone z Internetem przewodami. Zaproponowano i wdrożono kilka zaawansowanych technologii transmisji w celu zwiększenia wydajności sieci bezprzewodowych, w tym agregacji nośnej, wielomeny (MIMO) i komunikacji kooperatywnej, jako skoordynowanego szeregowania. Pomimo ciągłego postępu, wszystkie te nowoczesne technologie warstwy fizycznej (PHY) nie pozwalają na znaczne ulepszenia, ponieważ osiągają teoretyczne granice. Sieci ultra-gęste (UDN) i sieci heterogeniczne (HetNets) są zaprojektowane tak, aby zapewnić znaczący skok wydajności sieci poprzez wdrażanie heterogenicznej infrastruktury, np. Zdalnych głowic radiowych (RRH) i węzłów o niskiej mocy, takich jak pikokomórki, małe komórki, femt AP i przekaźniki. Zagęszczenie sieci poprzez zainstalowanie większej ilości infrastruktury ma dwie zasadnicze zalety. Po pierwsze, przybliżając sieć użytkownikom końcowym, jakość łącza i zasięg mogą zostać zwiększone ze względu na zmniejszenie odległości. Po drugie, duża liczba komórek pozwala na bardziej wydajne ponowne wykorzystanie widma, a zatem większe szybkości transmisji danych. Niemniej jednak, w gęstych sieciach o dużej skali, łączność typu backhaul, pojemność i niezawodność mogą stać się wąskim gardłem wydajności, nie wspominając nawet o dodatkowych kosztach związanych z ich wdrożeniem i konserwacją. Dwie dekady temu World Wide Web doświadczał podobnego wyzwania. Zgodnie ze starszym modelem klient-serwer każda przeglądarka na świecie musiała łączyć się z tym samym serwerem zdalnym. Suma ruchu tych połączeń typu end-to-end spowodowała katastrofalne zwiększenie ruchu w backhaulie internetowym. Rozwiązaniem tego problemu było buforowanie sieci; technika zapożyczona z procesorów obliczeniowych i systemów operacyjnych, która wykorzystuje czasowe wraz z przestrzennym sąsiedztwem żądań użytkowników, aby poprawić postrzeganą przez użytkowników jakość usług (QoS) i wskaźniki wydajności sieci. Wykorzystanie buforowania sieci prowadzi do rozprzestrzeniania się sieci dostarczania treści (CDN), odpowiedzialnych za replikację zawartości w całym Internecie dzisiaj i przechwytywanie ponad 60% ruchu kierowanego na zdalne serwery, eliminując w ten sposób znaczną część ruchu dosyłowego. Z motywacją wynikającą z sukcesu CDN-ów został niedawno zaproponowany aby używać buforowania w celu poprawy stabilności sieci bezprzewodowych. Przejście z sieci CDN do sieci bezprzewodowych nie tworzy po prostu "bezprzewodowej sieci CDN". Pojawiają się nowe paradygmaty replikacji treści, nowe systemy obiecują poprawić efektywność poprzez połączenie buforowania z właściwościami warstwy fizycznej (PHY), podczas gdy nowe wyzwania są narzucane przez mobilność, geometryczne właściwości propagacji sygnału bezprzewodowego i zakłócenia. Wiele długotrwałych modeli i związana z nimi konwencjonalna mądrość mogą wymagać przeglądu w

nadchodzących latach. Tu zbadamy fundamentalne teoretyczne podstawy sieci bezprzewodowych z możliwościami buforowania. Dokładniej:

1. Prezentujemy narzędzie analityczne do uproszczenia obliczeń wymaganych do określenia optymalnej replikacji treści w sieci bezprzewodowej. To narzędzie służy do wyprowadzania asymptotycznych praw skalowania wydajności backhaul dla sieci bezprzewodowych multi-hop. Przedstawione wyniki uwzględniają (i) liczbę węzłów sieci  $N$ , (ii) liczbę treści  $M$  oraz (iii) wielkość pamięci podręcznej  $K$  jako kluczowe parametry "rozmiaru" systemu, które zwiększają się arbitralnie w różnych proporcjach. Analizując różne przypadki, wyciągamy cenną intuicję z zalet buforowania dla zrównoważonego rozwoju sieci bezprzewodowych.
2. Badamy wydajność gęstych sieci komórkowych z małymi komórkami z włączoną pamięcią podręczną i dostarczamy użytecznych informacji na temat tego, jak wartości operacyjne systemu, topologia sieci i zakłócenia wpływają na wydajność sieci. Korzystając z narzędzi z geometrii stochastycznej do modelowania rozkładu węzłów, podajemy wytyczne dotyczące umieszczania najpopularniejszych treści, np. Na urządzeniu mobilnym lub w AP / BS. Analizowany jest także efekt korelacji przestrzennej w żądaniach treści.
3. Zakończymy podsumowaniem ciekawych kierunków badań na przyszłość, które mogą odgrywać kluczową rolę w rozprzestrzenianiu pamięci podręcznej jako techniki sieci bezprzewodowej.

## **7.2 Zrównoważone, wielopoziomowe sieci bezprzewodowe z buforowaniem**

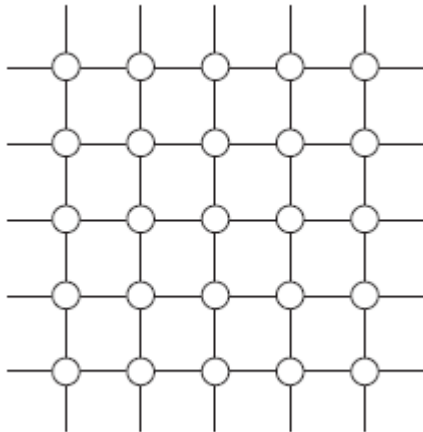
Trwałość bezprzewodowych sieci wielopasmowych można scharakteryzować poprzez badanie prawa skalowania pojemności sieci, która rośnie. W swojej doniosłej pracy Gupta i Kumar badali zachowanie asymptotyczne sieci bezprzewodowych typu multi-hop gdy komunikacja odbywa się między  $N$  parami losowo wybranych węzłów. W takim scenariuszu maksymalna stawka danych to  $O(1/\sqrt{N})$  w związku z tym, że sieć rośnie  $N \rightarrow \infty$ , prędkość transmisji danych znika do zera. To stwierdzenie jest sprzeczne z zasadą trwałości komunikacji międzyprzedmiotowej. Co więcej, wykazano, że prawo  $O(1/\sqrt{N})$  powstało z rozważań geometrycznych transmisji bezprzewodowych 2D, w związku z czym niemożliwe jest naruszenie pod klasycznym modelem pary losowo komunikującej się. W tej sekcji powracamy do tematu zrównoważonego rozwoju bezprzewodowego multi-hopu sieci z dodanym elementem buforowania. Każdy węzeł w sieci reprezentuje użytkownika, który jest zainteresowany określoną treścią (zamiast komunikowania się z innym węzłem). Zawartość może być replikowana w pamięciach podręcznych sieci, a zatem użytkownik może wygodnie pobrać ją z pobliskiej pamięci podręcznej, zmniejszając w ten sposób liczbę przeszukiwanych przeskoków. Wyprowadzone asymptotyczne prawa dotyczące skalowania pojemności sugerują, że buforowanie ma potężny wpływ na zrównoważenie sieci bezprzewodowych i istnieją ciekawe reżimy, w których łamana jest prawa  $O(1/\sqrt{N})$ . Podstawowymi parametrami wielkości są liczba węzłów  $N$  oraz liczba treści  $M$ , a zatem prawa skalowania będą zależą od tego, jak będą one wzrastać do nieskończoności. Innym kluczowym parametrem jest rozmiar pamięci podręcznej  $K$ , który przedstawia liczbę treści, które mogą być buforowane w każdym węźle. Zabieranie  $K$  do nieskończoności reprezentuje interesujący reżim, który odzwierciedla sieci, w których węzły są uaktualniane, ponieważ pamięć masowa staje się zasobna i niedroga. Innym czynnikiem wpływającym na popularność jest popularność treści. Ze skończoną popularnością, popularne pliki są żądane wiele razy, a więc mniejsze  $K$  mają większy wpływ na wydajność systemu. Zgodnie z przewidywaniami, w naszej analizie skończoność popularności ma decydujący wpływ na trwałość sieci bezprzewodowych. Przedstawiona analiza stanowi podsumowanie i obejmuje rozwiązanie ciężkiej, kombinatorycznej optymalizacji, która przechwytuje najlepszy sposób replikowania treści z różnych popularnych stron i pobierania ich za pośrednictwem tras sieciowych. Na

szczęście, poprzez złagodzenie problemu do wypukłej optymalizacji i zaokrąglenia rozwiązania, można uzyskać optymalną wydajność zamówienia.

## 7.2.1 Model systemu

### 7.2.1.1 Model sieci

Sieć jest modelowana przez topologię sieci, jak na rysunku 7.1,



która rejestruje podstawowe cechy sieci bezprzewodowych (i) wielozakresowej komunikacji krótkiego zasięgu są preferowane i (ii) średnica sieci skaluje się jako  $\sqrt{N}$ . Formalnie rozważamy zbiór  $N$  węzłów

indeksowanych przez  $n \in \mathcal{N} \triangleq \{1, 2, \dots, N\}$

ułożone na  $\sqrt{N} \times \sqrt{N}$  kwadratowej siatce na płaszczyźnie. Każdy węzeł jest połączony przez niepokierunkowane łącza z czterema sąsiadami, które znajdują się obok niego w tym samym wierszu lub kolumnie. Utrzymując stałą gęstość węzłów i zwiększając rozmiar sieci  $N$ , otrzymujemy sieć skalowania, ze skalowaniem przepustowości losowej pary jako  $O(1/\sqrt{N})$

### 7.2.1.2 Żądania treści

Węzły generują żądania do katalogu treści indeksowanego za pomocą  $m \in \mathcal{M} \triangleq \{1, 2, \dots, M\}$ . Zwykle każdy węzeł żąda treści  $m$  z kursem  $\lambda_n^m$ . Takie rozważanie doprowadziłoby analizę zrównoważonego rozwoju do tworzenia regionu zdolności (który jest zbiorem wszystkich wektorów szybkościowych  $[\lambda_n^m]$  które są trwałe), które jednak wyraża się za pomocą zestawu nierówności i nie zapewni prostego wyrażenie dla praw skalujących. Aby uprościć ekspozycję i uzyskać wynik ilościowy, zakładamy symetryczny przypadek, w którym wszystkie węzły mają  $\lambda_n^m = \lambda^m, \forall n$ . Dokładniej rzecz biorąc, rozważamy niezależny model odniesienia (IRM), zgodnie z którym wnioski o plik  $m$  wynikają z niezależnego procesu Poissona o intensywności  $\lambda p_m$ , gdzie  $p_m$  reprezentuje dystrybucję popularności plików - ta ostatnia jest przyjęta jako niezmienna w czasie. Na koniec ustalamy  $\lambda = 1$  i badamy wymaganą przepustowość łącza, aby wesprzeć komunikację. Jest to odwrotność klasycznego podejścia, w którym ustalamy pojemność łącza i badamy skalowanie maksymalnej możliwej do obsłużenia przepustowości. Na przykład w sieci, w której pod stałymi łączami przepustowość skaluje się jako  $O(1/\sqrt{N})$  nasz model będzie generować wymagane skalowanie przepustowości łącza  $\Theta(\sqrt{N})$  w celu utrzymania stałej przepustowości  $\lambda = 1$ . Bardziej ogólnie, otrzymane prawa skalowania dla wymaganej przepustowości łącza mogą być odwrócone, odzwierciedlają prawa dotyczące skalowania przepustowości. W związku z tym, aby osiągnąć zrównoważony rozwój, potrzebujemy łącza o zdolności do bycia tak małym, jak to możliwe, najlepiej, aby było  $O(1)$ .

### 7.2.1.3 Model buforowania

Każdy węzeł  $n$  jest wyposażony w pamięć podręczną, której zawartość oznaczona jest zbiorem  $\mathcal{B}_n$ , a podzbiór  $\mathcal{M}$ . Jeśli żądanie w węźle  $n$  dotyczy pliku  $m$ , który znajduje się w  $\mathcal{B}_n$ , to jest obsługiwane lokalnie. Ze względu na ograniczoną pamięć podręcznej  $m$  często nie będzie dostępne w  $\mathcal{B}_n$ , zatem węzeł  $n$  będzie musiał żądać  $m$  przez sieć od jakiegoś innego węzła  $w$ , który utrzymuje  $m$  w swojej pamięci podręcznej. Oznaczamy przez  $[\mathcal{R}_{n,k}]$  zestaw tras łączących węzły z pobliskimi pamięciami podręcznymi. Wybór zbiorów  $[\mathcal{B}_n]$  i  $[\mathcal{R}_{n,k}]$  ma zasadniczy wpływ na ładowanie łącza sieciowego. Niech ponadto  $K$  będzie pojemnością pamięci podręcznej węzłów mierzoną liczbą plików, które może przechowywać. Oznacza to, że wszystkie pliki  $M$  mają ten sam rozmiar (jednostkę), umieszczając ograniczenie na liczności zawartości pamięci podręcznej  $|\mathcal{B}_n| \leq K$ . Ważnymi parametrami wielkości systemu są:  $N, M, K$ . Ciekawym reżimem, który będziemy badać jest  $K < M$ ?  $KN$ : (7.1) Pierwsza nierówność oznacza, że rozmiar pamięci podręcznej  $K$  nie wystarcza, aby zmieścić wszystkie pliki, a zatem każdy węzeł musi dokonać wyboru plików do pamięci podręcznej. Druga nierówność wymaga, aby całkowita pojemność pamięci podręcznej sieci  $KN$  (sumująca wszystkie indywidualne pamięci podręczne) była wystarczająca do przechowywania wszystkich plików co najmniej raz, a tym samym można obsłużyć każde żądanie zawartości.

### 7.2.2 Optymalizacja pojemności multihop: relaksacja i zaokrąglanie

Ze względu na ograniczenia miejsca docelowego  $|\mathcal{B}_n| \leq K$ , struktura problemów z optymalizacją pamięci podręcznej jest z natury kombinatoryczna, a zatem są trudne do rozwiązania. W tej sekcji (i) przedstawiamy istotne problemy związane z optymalizacją praw do skalowania pojemności z buforowaniem, (ii) proponujemy rozluźniony problem oparty na gęstości plików, który jest łatwy do rozwiązania, i w ten sposób (iii) ustanawiamy możliwe rozwiązanie, które jest ta sama kolejność z pierwotnym problemem kombinatorycznym.

#### 7.2.2.1 Problem wspólnego routingu replikacji

Aby uzyskać prawidłowe prawo skalowania wymaganej przepustowości łącza, konieczne jest wybranie najlepszego miejsca i najlepszego routingu, który minimalizuje obciążenie w najgorszym ogniwie w sieci. Niech  $C_\ell$  będzie obciążeniem ruchem przenoszonym przez łącze  $\ell$ . Sieć działa stabilnie, nie odrzucając żądań, tylko jeśli przepustowość łącza "przekracza"  $C_\ell$  dla wszystkich  $\ell$ . Link łądzący  $C_\ell$  można skorygować poprzez (i) buforowanie popularnej treści (poprzez wybór miejsca  $[\mathcal{B}_n]$ ) i (ii) równoważenie obciążenia w zestawie ścieżek routingu  $[\mathcal{R}_{n,k}]$  z węzła buforowania  $n$  do węzła żądania  $k$ . Stąd właściwe prawo skalowania wymaganych mocy ogniw  $C^*$  jest identyfikowany jako wartość następującej optymalizacji:

**Problem 7.1**  $C^* = \text{Minimize}_{[\mathcal{B}_n], [\mathcal{R}_{n,m}]} \max_\ell C_\ell$ , s.t.

1.  $|\mathcal{B}_n| \leq K, \quad \forall n$  ograniczenie pamięci podręcznej
2.  $\sum_n \mathbf{1}_{\{m \in \mathcal{B}_n\}} \geq 1, \quad \forall m \in \mathcal{M}$  wszystkie pliki są buforowane
3.  $[\mathcal{R}_{n,m}]$  są możliwe trasy

Ta wspólna optymalizacja okazuje się być trudnym problemem kombinatorycznym, nie nadającym się do łatwego do obliczenia rozwiązania. Dlatego korzystamy z uproszczeń i przybliżeń, które zapewniają rozwiązanie optymalne pod względem zamówienia, tj. którego wartość funkcji obiektywu znajduje się w stałej, optymalnej, trudnej do obliczenia min maks  $C_\ell$ . Pierwszym krokiem, który zachowuje optymalność zamówienia rozwiązania jest relaksacja celu średniego ruchu łącza  $\text{avg}_\ell C_\ell$  następnie

zmienne trasujące  $[\mathcal{R}_{n,k}]$  może być ustalone na najkrótsze ścieżki bez utraty optymalności. Jednak decydujący etap polega na zerwaniu sprzężenia między poszczególnymi pamięciami podręcznymi  $[\mathcal{B}_n]$ . Aby osiągnąć to oddzielenie, wprowadzamy nowe pojęcie gęstości replikacji treści  $m$ .

### 7.2.2.2 Rozluźniony problem oparty na gęstości

Biorąc pod uwagę pozycję  $[\mathcal{B}_n]$ . Rozważ częstotliwość występowania każdego pliku  $m$  w pamięciach podręcznych lub gęstość replikacji  $d_m$  jako ułamek węzłów przechowujących plik  $m$  w sieci:

$$d_m = \frac{1}{N} \sum_{n \in \mathcal{N}} 1_{\{m \in \mathcal{B}_n\}}.$$

W oparciu o te dane definiujemy prostszy problem:

**Problem 7.2**  $C = \text{Minimize}_{\{d_m\}} \sum_{m \in \mathcal{M}} \left( \frac{1}{\sqrt{d_m}} - 1 \right) p_m$ , s.t.

1. Dla dowolnego  $m \in \mathcal{M}$ ,  $1/N \leq d_m \leq 1$ ,

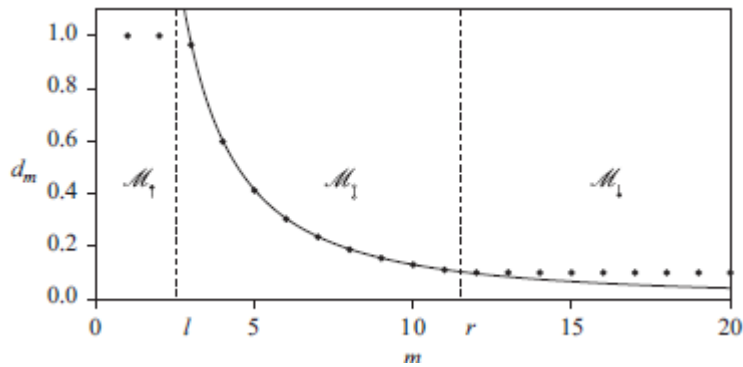
2.  $\sum_{m \in \mathcal{M}} d_m \leq K$ .

W powyższym przykładzie optymalizujemy gęstość  $d_m$ , wyrażającą ułamek pamięci podręcznych zawierający plik  $m$ . W celu  $1/\sqrt{d_m}-1$  przybliża (w kolejności) średnią liczbę przeskoków z losowego węzła do pamięci podręcznej zawierającej  $m$ . Ważona prawdopodobieństwem  $p_m$  żądań na  $m$ , suma wyraża średnie obciążenie łącza na żądanie. Dodatkowo ograniczenie  $\sum_{m \in \mathcal{M}} d_m \leq K$  odzwierciedla kolejną relaksację, w której ograniczenie rozmiaru pamięci podręcznej jest spełnione tylko w średniej w sieci. Zauważ, że usunęliśmy pojedyncze ograniczenia pamięci podręcznej, a możliwe rozwiązanie Problemu 7.2 może zapewnić gęstość plików, która odpowiada buforowaniu ponad pliki  $K$  w jednym węźle. Oczywiście jest, że każde możliwe rozwiązanie problemu 7.1 daje gęstość plików  $\{d_m\}$ , co jest możliwe w przypadku problemu 7.2, ale nie na odwrót, stąd też formuła oparta na gęstości jest zrelaksowaną wersją pierwotnego problemu min-max i mamy  $C = O(C^*)$ . W szczególności, Problem 7.2 jest wypukły, a jego unikalne rozwiązanie można znaleźć za pomocą warunków Karush-Kuhn-Tucker (KKT). Jeśli chodzi o ograniczenia dotyczące wartości minimalnej i maksymalnej  $d_m$ , jeden z nich może być równy lub żaden. Rozdziela to  $\mathcal{M}$  na trzy podzbiory, "z góry skrócony"  $\mathcal{M}_\uparrow = \{m : d_m = 1\}$  zawierający pliki przechowywane we wszystkich węzłach, „z dołu obcinane”  $\mathcal{M}_\downarrow = \{m : d_m = 1/N\}$  zawierający pliki przechowywane tylko w jednym węźle i uzupełniające "nieskrócone"  $\mathcal{M}_\updownarrow = \mathcal{M} \setminus (\mathcal{M}_\uparrow \cup \mathcal{M}_\downarrow)$  plików o  $1/N < d_m < 1$ . Rozmieszczając  $p_m$  w porządku malejącym, stają się partycjami  $\mathcal{M}_\uparrow = \{1, 2, \dots, l-1\}$ ,  $\mathcal{M}_\updownarrow = \{l, l+1, \dots, r-1\}$  i  $\mathcal{M}_\downarrow = \{r, r+2, \dots, M\}$

$l$  i  $r$  są liczbami całkowitymi z  $1 \leq l \leq r \leq M + 1$ . Rozwiązanie  $d_m$  jest równe:

$$d_m = \begin{cases} 1, & m \in \mathcal{M}_\uparrow, \\ \frac{K - l + 1 - \frac{M-r+1}{N} \frac{2}{3} p_m}{\sum_{j \in \mathcal{M}_\updownarrow} \frac{2}{3} p_j}, & m \in \mathcal{M}_\updownarrow, \\ 1/N, & m \in \mathcal{M}_\downarrow. \end{cases}$$

Rysunek 7.2 ilustruje takie przykładowe rozwiązanie, obrazujące gęstość  $d_m$ , indeksy  $l$  i  $r$ , a także zestawy  $\mathcal{M}_\uparrow$ ,  $\mathcal{M}_\updownarrow$ ,  $\mathcal{M}_\downarrow$  kiedy popularne pliki podążają za prawem Zipfa (patrz punkt 7.2.3).

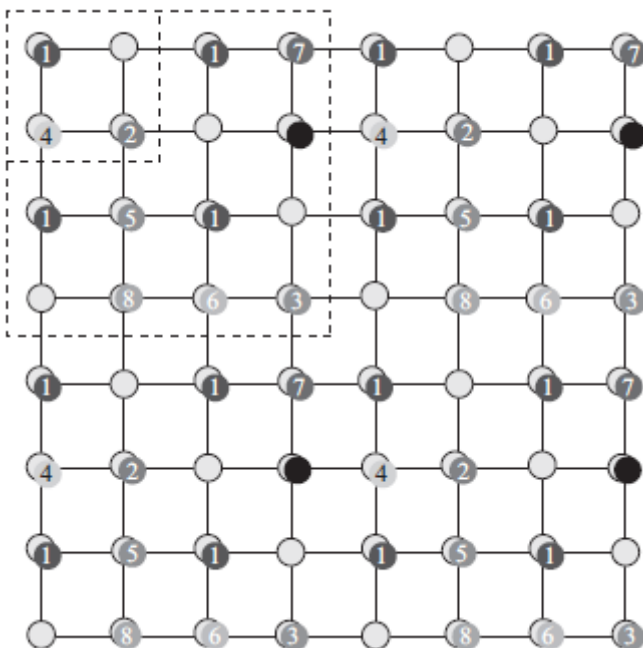


### 7.2.2.3 Gęstość dyskretna (zaokrąglenie)

Rozwiązanie nie jest bezpośrednio odwzorowane na możliwe rozwiązanie problemu 7.1, ale możemy zbudować jeden przy użyciu procesu dwuetapowego, (i) pierwsze zaokrąglenie od  $[d_m]$  do  $[d_m^o]$  oraz (ii) drugie umieszczenie treści symetrycznie w sieci zgodnie z  $[d_m^o]$  tak, że ograniczenia  $[\mathfrak{B}_n] \leq K$  są spełnione. W przypadku (i) po prostu definiujemy  $d_m^o \triangleq 4^{-v_m^o}$  zaokrąglone do największej potęgi mniejszej lub równej  $[d_m]$

$$d_m^o \triangleq \max\{4^{-i} : 4^{-i} \leq d_m, i \in \{0, 1, \dots, v\}\}$$

Następnie dla (ii), podajemy algorytm alokacji plików  $\mathcal{M}$  w pamięci podręcznej  $[\mathfrak{B}_n]$  biorąc pod uwagę gęstości replikacji  $d_m^o$ . Algorytm można wyjaśnić na podstawie średniej rysunku 7.3.



Zaczynamy od szarego pliku, który ma  $d_0^o = 1$ , ten plik jest po prostu buforowany wszędzie. Następnie dla pliku o  $d_0^o = 1/4$ , skupiamy się na podsiatce 2 x 2 węzłów (każda taka podsiatka wystarcza, ale pomaga naprawić początek jako odniesienie). W tym podkatalogu próbujemy wypełnić przekątną, co w tym przypadku jest osiągnięte poprzez umieszczenie pliku 1 na współrzędnej (1,1) (górny lewy węzeł w siatce). W ostatnim kroku tego pliku umieszczamy repliki, umieszczając podkatalog w dowolnym miejscu w sieci. W rezultacie plik 1 jest replikowany z gęstością 1/4, zgodnie z zaleceniami rozwiązania.

Następnie dla plików o gęstości  $1/16$  powiększamy podsiatkę do  $4 \times 4$ . Ogólnie podsiatka ma rozmiar  $2^v_m \times 2^v_m$  i jest dopasowany do wszystkich rozważanych podgrup. Następnie wstawiamy nowe pliki zaczynające się od przekątnej, w szczególności pliki 2 i 3 w przykładzie. Następnie pliki 4, 5, 6 są wypełnione drugą przekątną, która znajduje się poniżej pierwszej, a 7 uzupełnia drugą przekątną poprzez zawijanie. na współrzędnej (1, 4). Podczas wypełniania wybieramy tylko te węzły, które mają mniej plików niż maksymalne. Na przykład, kiedy wypełniamy obiekt 2 w podsiatce  $4 \times 4$ , pomijamy węzeł (1,1) ponieważ ten węzeł ma już dwa pliki (plik 0 i plik 1), a plik 2 umieszczamy w węźle (2,2), który do tej pory miał tylko jeden plik (plik 0). Gdzie umieścić plik 10 z  $d_{10} = 1/64$ ? Zastanowilibyśmy się nad podsiatką  $8 \times 8$  (tj. cała siatka), pierwsze trzy przekątne są całkowicie wypełnione, a zatem umieścimy je w pierwszym otwartym miejscu w czwartej przekątnej, to znaczy w węźle (4,1). Korzystając z tego miejsca docelowego, ostatecznie uzyskujemy wykonalne rozwiązanie problemu 7.1 o wartości  $C^\circ = \Omega(C)$ .

### Twierdzenie 7.1 (Rozłożenie relaksacji i zaokrąglenia optymalizuje obciążenie ogniw).

Tam istnieją stałe dodatnie  $a$ ,  $b$ , które zależą od rozkładu  $[p_m]$  i pojemność pamięci podręcznej  $K$ , taka, że  $C^* \leq C^\circ aC^* + b$ :

Ponadto jest to  $\Theta(C^*) = \Theta(C^\circ) = \Theta(C)$ .

Podsumowując, możemy wyprowadzić prawa skalowania sieci bezprzewodowych z buforowaniem  $C^*$  bezpośrednio za pomocą zrelaksowanego rozwiązania problemu 7.2, zwracamy uwagę, że ponieważ  $a$ ,  $b$  powyżej zależy od  $K$ , skalowanie praw w odniesieniu do  $K$  musi uważnie rozważyć tę zależność.

### 7.2.3 Prawa asymptotyczne dla popularności Zipfa

Aby zbadać skalowanie  $C$ , przełączamy się z arbitralnej popularności na prawo Zipfa, co zaobserwowano, że dobrze modeluje popularność ruchu na stronie www i inne rodzaje usług. To jest zdefiniowane w następujący sposób:

$$p_m = \frac{1}{H_\tau(M)} m^{-\tau};$$

gdzie  $H_\tau(n) \triangleq \sum_{j=1}^n j^{-\tau}$  jest skróconą funkcją zeta szacowaną w  $\tau$  (zwaną również  $n$ -tą uogólnioną liczbą harmoniczną ułożoną w porządku  $\tau$ ), a  $\tau$  jest parametrem rozkładu, dostosowującym tempo spadku popularności za pomocą  $m$ . Jeśli chodzi o parametr  $\tau$ , dla ruchu sieciowego zgłaszane są wartości  $\leq 1$ . Interesujące jest to, że Padmanabhan i Qiu mierzą  $\tau = 1$  w proxy w porównaniu z 1.4-1.6 na "zajętym" serwerze internetowym. Analizując zapisy z przeglądania na urządzeniach mobilnych, Yamakami podaje wartości w przedziale 1-1,5, natomiast w materiale wideo generowanym przez użytkownika (UGC), pasuje popularność do kombinacji Zipf i Exponential Cut-off, gdzie komponent Zipf ma  $\tau$  w przedziałach około 0,98-1,47 i 0,45-1,09 w dwóch różnych scenariuszach. Ostatnio popularność treści w systemach P2P została dopasowana do  $\tau = 0.95$  oraz w systemie wideo na żądanie na 0.70. Ponieważ nie ma rozstrzygającej odpowiedzi na temat wartości  $\tau$ , rozważamy wszystkie możliwości w naszym dochodzeniu. Podajemy przybliżenie dla  $H_\tau(n)$  przez ograniczenie sumy: dla  $n \geq m \geq 0$ ,



$$\int_m^n (x+1)^{-\tau} dx \leq H_\tau(n) - H_\tau(m) \leq 1 + \int_{m+1}^n x^{-\tau} dx, \Rightarrow$$

$$\begin{cases} \frac{(n+1)^{1-\tau} - (m+1)^{1-\tau}}{1-\tau} \leq H_\tau(n) - H_\tau(m) \leq \frac{n^{1-\tau} - (m+1)^{1-\tau}}{1-\tau} + 1, & \text{if } \tau \neq 1 \\ \ln \frac{n+1}{m+1} \leq H_\tau(n) - H_\tau(m) \leq \ln \frac{n+1}{m+2}, & \text{if } \tau = 1 \end{cases} \quad (7.5)$$

Zastąpienie rozwiązania i podłączenie dystrybucji Zipf do celu problemu 7.2, powoduje, że:

$$C \triangleq \sum_{m \in \mathcal{M}} \left( d_m^{-\frac{1}{2}} - 1 \right) p_m = C_{\circlearrowleft} + C_{\bullet} - \sum_{j=1}^M p_m.$$

gdzie  $\sum_{j=1}^M p_m = O(1)$  (ponieważ zawsze leży w  $[0,1]$  i

$$C_{\circlearrowleft} \triangleq \sum_{m \in \mathcal{M}_1} \frac{p_m}{\sqrt{d_m}} \stackrel{(1.2),(1.4)}{=} \frac{\left[ H_{\frac{2}{3}}(r-1) - H_{\frac{2}{3}}(l-1) \right]^{\frac{3}{2}}}{\sqrt{K-l+1 - \frac{M-r+1}{N}} H_\tau(M)},$$

$$C_{\bullet} \triangleq \sum_{m \in \mathcal{M}_2} \frac{p_m}{\sqrt{d_m}} \stackrel{(1.2),(1.4)}{=} \sqrt{N} \frac{H_\tau(M) - H_\tau(r-1)}{H_\tau(M)}.$$

(7.7 ; 7.8)

Aby analitycznie obliczyć prawo C, musimy przeanalizować l i r. Zauważ, że skoro M skaluje do nieskończoności, l; r może również skalować do nieskończoności, lub nie, w zależności od rzeczywistego rozwiązania. Ponadto, obserwując, że powyższe wyrażenia zależą od  $H_\tau$ ;  $H_{2/3}$ , spodziewamy się pojawienia różnych przypadków ze względu na formę (7.5). Te w terminach są mapowane na skalowanie praw poprzez (7.7) i (7.8). Wyniki przedstawiono są tu

$M$	$\tau$	$M$ finite	$N \rightarrow \infty$ , then		
			$M \rightarrow \infty$	$M \sim KN$ , hence $M = \Theta(N)$	
			$KN - M = o(1)$	$KN - M = O(1)$	
$C$	$\tau < 1$	$\Theta(1)$	$\Theta(\sqrt{M})$	$\Theta(\sqrt{N})$	$\Theta(\sqrt{N})$
	$1 < \tau < \frac{3}{2}$	$\Theta(1)$	$\Theta(M^{\frac{3}{2}-\tau})$	$\Theta\left(\frac{\sqrt{N}}{(KN-M)^{\tau-1}}\right)$	$\Theta(\sqrt{N})$
	$\tau > \frac{3}{2}$	$\Theta(1)$	$\Theta(1)$	$\Theta\left(\frac{\sqrt{N}}{(KN-M)^{\frac{3\tau-1}{2\tau}}}\right)$	$\Theta(\sqrt{N})$

Przypadki  $\tau = 1$  i  $\tau = 3/2$  są pomijane, aby uniknąć bałaganu; są one podobne do przypadków  $\tau < 1$  i  $\tau > 3/2$  dla czynnika logarytmicznego.

#### 7.2.4 Skalowanie praw dla stałej wielkości pamięci podręcznej K

Tabela powyższa pokazuje, w jaki sposób rozwiązanie problemu 7.2 skaluje się z parametrami rozmiaru systemu  $N$  (liczba węzłów / użytkowników) i  $M$  (rozmiar katalogu);  $K$  (rozmiar pamięci podręcznej) w tej tabeli jest stałą. Z twierdzenia 7.1, skalowanie  $C$  ma również zastosowanie do wymaganej wydajności łącza dla utrzymania jednolitej stawki żądania  $\lambda = 1$ . Na przykład skalowanie  $\Theta(1)$  oznacza, że sieć może utrzymać wydajność jednostki, nawet jeśli jest to łącze zdolne (wystarczająco duże) stałe; jest to pożądany przypadek. Z drugiej strony  $\Theta(\sqrt{M})$  oznacza, że w celu utrzymania przepustowości jednostki pojemność łącza musi wzrosnąć proporcjonalnie do pierwiastka kwadratowego z katalogu plików. Ponieważ w praktyce systemy  $M$  powinny rosnać wraz z rozmiarem sieci  $N$ , podczas gdy przepustowość łącza ma spełniać pewne ograniczenia (takie jak limit Shannon), oznacza to, że sieć nie będzie mogła być zrównoważona zgodnie z takim prawem. Przypomnijmy, że możemy również odwrócić skalowanie, aby zobaczyć, jak przepustowość będzie skalowana w sieci o stałych mocach łącza. Na przykład, jeśli zmienimy prawo  $\Theta(\sqrt{M})$  daje  $\lambda = \Theta(1/\sqrt{M})$ , którym jest skalowanie Gupta-Kumara dla losowo komunikujących się par w sieci skalującej. To prawo pojawia się w naszej analizie za każdym razem, gdy zdolność replikacji  $KN - M = \Theta(1)$  jest niska, tzn. prawie wszystkie pamięci podręczne są używane do przechowywania każdego pliku raz - patrz ostatnia kolumna powyższej tabeli. Ponieważ  $K$  jest stała, wyniki są uzyskiwane dla różnych przypadków, w których skala  $N$  i  $M$  jest nieskończona. Druga kolumna odpowiada przypadkowi, w którym tylko skale  $N$  i  $M$  są stałe w takim przypadku system jest trwały (wymagane łącze wydajności  $\Theta(1)$ ). Jednak w systemach praktycznych katalog plików rośnie wraz z populacją użytkowników. Trzecia kolumna odnosi się do przypadku, w którym przyjmujemy ograniczenia w kolejności, najpierw  $N \rightarrow \infty$ , a następnie  $M \rightarrow \infty$ , tzn. mamy  $M = o(N)$ , a katalog rośnie podliniowo do rozmiaru sieci. Prawa skalowania w tym przypadku odnoszą się do rozmiaru katalogu plików, który jest nieznaczną poprawą w stosunku do skalowania Gupta-Kumara. Uzyskano różne skalowanie dla różnych wartości  $\tau$  gdzie większe wartości ustawiamy  $M = \alpha KN$  dla pewnej pozytywnej  $\alpha < 1$ , a następnie przyjmujemy granicę  $N \rightarrow \infty$ ; Prawa skalowania w tym przypadku również zależą od  $KN - M$ , tj. Ile slotów replikacji pozostało w pamięciach podręcznych po tym, jak przechowujemy każdy plik jeden raz. Są dwa przypadki, (a)  $KN - M = \omega(1)$  i (b)  $KN - M = O(1)$ . W pierwszym przypadku, pomimo  $M = \alpha KN$ , jest wystarczająco dużo miejsc do replikacji, aby uzyskać skalowania, które reprezentują ulepszenia w skalowaniu Gupta-Kumara, o ile  $\tau > 1$ . W drugim przypadku, uzyskujemy skalowanie Gupta-Kumara niezależnie od wartości  $\tau$ .

### 7.2.5 Skalowanie rozmiaru pamięci podręcznej $K \rightarrow \infty$

Ponieważ pamięć staje się tańsza i tańsza, możemy przewidzieć scenariusz, w którym rozmiar pamięci podręcznej na węzeł  $K$  również skaluje się do nieskończoności. W tym miejscu krótko omówimy wynikające z tego prawa dotyczące skal przedstawione w Tabeli 7.2.

		“High” $K$	“Low” $K$
$C$	$\tau < 1$	$O\left(\sqrt{\frac{M}{K}}\right)$	$\Theta(\sqrt{N})$
	$1 < \tau < \frac{3}{2}$	$O\left(\frac{M^{\frac{1}{2}-\tau}}{\sqrt{K}}\right)$	$\Theta\left(\frac{\sqrt{N}}{(KN - M)^{\tau-1}}\right)$
	$\tau > \frac{3}{2}$	$O(1)$	$\Theta\left(\sqrt{\frac{N}{KN - M}}\right)$

Biorąc pod uwagę, jak  $N$ ;  $M$ ;  $K$  rośnie, możemy uczyć się systemu w różnych reżimach działania, co komplikuje ekspozycję skalowania praw. Z tego powodu skupiamy się tutaj na dwóch konkretnych systemach interesów. Porównujemy całkowitą pamięć sieciową  $KN$  z katalogiem plików  $M$  i podzielimy

analizę na dwa przypadki (i)  $KN = \Omega(M)$  i (ii)  $KN = O(M)$ . Pierwszy przypadek nazywa się "High" K, a drugi "Low" K:

#### **7.2.5.1 " High " K**

Najbardziej interesującym reżimem do zbadania dla idealnego zrównoważenia jest ten z  $C = O(1)$ . Jak pokazują formuły, aby ograniczyć C, najtrudniejszy przypadek dotyczy  $\tau < 1$ : pojemność pamięci podręcznej węzła K powinna być skalowana tak szybko, jak zawartość woluminu M. W przypadku półproduktu  $1 < \tau < 3/2$ , pojemność węzła K musi być skalowana za pomocą M, ale wolniejsza, w linii sublinearnej moc. Przypadek  $\tau > 3/2$  jest całkiem interesujący, ponieważ  $C = O(1)$  zawsze jest prawdziwe.

#### **7.2.5.2 "Low" K**

Reżim ten charakteryzuje się jako niezrównoważony, ponieważ wzrastają węzły sieci pamięci podręcznej K i / lub liczba węzłów N względem objętości treści M nie jest wystarczająca do utrzymania C na niskim poziomie. Po pierwsze, należy pamiętać, że gdy pojemność pamięci replikacji poza przechowywanie kopii pierwotnej jest ograniczona, tj.  $KN - M = O(1)$ , C skaluje się tak szybko jak  $\sqrt{N}$ . Ponadto należy zauważyć, że dla  $\tau > 1$ ,  $\sqrt{N}$  jest zmniejszany o pojemność  $KN - M$  dostępne do replikacji poza kopią podstawową przy mocy równej  $\sqrt{KN-M}$  lub  $(KN-M)^{\tau-1}$  dla  $\tau > 3/2$  lub  $1 < \tau < 3/2$ , odpowiednio, określa to zysk z dodania dodatkowej pojemności w sieci poza tą wymaganą do przechowywania kopii pierwotnej na plik.

#### **7.2.6 Dyskusja na temat trwałości bezprzewodowych sieci pamięci podręcznych**

Przedstawiliśmy wspólny problem dostarczania i replikacji, który prowadzi do scharakteryzowania trwałości wieloskokowych sieci bezprzewodowych z buforowaniem. Formułując dokładny problem kombinatoryczny w kwadratowych sieciach kratownicowych, rozluźniliśmy go do prostego problemu opartego na gęstości, którego analiza pozwoliła wyprowadzić prawa skalowania wymaganej przepustowości łącza. Kluczowymi czynnikami wpływającymi na stabilność sieci są: (a) parametr prawa zasilania w zakresie popularności plików t oraz (b) względne skalowanie liczby plików w porównaniu do węzłów sieci. Podczas gdy nasze dotychczasowe leczenie cechuje zalety buforowania w sieciach wielopasmowych, inna linia pracy bada prawa skalowania buforowania w bezprzewodowym medium transmisyjnym za pomocą kodowania. Następnie kontynuujemy naszą ekspozycję przybliżając topologię sieci bezprzewodowych, biorąc pod uwagę wykorzystanie buforowania w małych komórkach przy współpracy urządzenia z urządzeniem

#### **7.3 Sieci małych komórek z włączoną pamięcią podręczną**

Pamięć podręczna i dystrybucja zawartości w sieciach bezprzewodowych została uznana za obiecujące rozwiązanie do rozładowywania ruchu w celu zwiększenia wydajności i jakości usług (QoE) poprzez wykorzystanie popularności treści i korelacji zapytań czasowo-przestrzennych. Koncepcja buforowania na krawędziach sieci dostępu radiowego, określana tutaj jako buforowanie małych komórek (SC), jest napędzana przez przyrost przepustowości i opóźnienia, od wprowadzania funkcji buforowania do SBS i zmniejszania wykorzystania backhaul. Z włączoną pamięcią podręczną, małe sieci komórkowe (SCN) zostały zaproponowane jako potencjalne rozwiązanie w celu złagodzenia problemu wąskich gardeł backhaul. Główną ideą jest wprowadzenie pojemności pamięci podręcznej w małych komórkowych stacjach bazowych (SBS) w celu wstępnego pobrania popularnych treści poza godzinami szczytu, zanim zostaną one zamówione przez lokalnych użytkowników. Buforowanie w sieciach bezprzewodowych wykorzystuje również wysoki stopień asynchronicznego ponownego wykorzystania treści przez aplikacje zorientowane na informację, takie jak wideo na żądanie (VoD), sieci społecznościowe i

udostępnianie treści. Gdy użytkownicy proszą o popularną zawartość już zapisaną w pamięci podręcznej w lokalnych SBS, opóźnienie usługi jest znacznie ograniczone, ponieważ nie musi przechodzić przez backhaul, aby pobrać zawartość ze zdalnych serwerów. Poprawiona efektywność energetyczna jest również ważną zaletą buforowania małych komórek głównie z powodu wielokrotnego pobierania tej samej treści z rdzenia sieci do lokalnych SBS są unikane. Innym podejściem do ponownego wykorzystania treści jest wykorzystanie dostępnej przestrzeni dyskowej urządzeń mobilnych. Popularne treści są zapisywane w pamięci podręcznej urządzeń i dystrybuowane do innych urządzeń w pobliżu za pośrednictwem komunikacji D2D. Ji i in. [34] omawiają przerywanie przepustowości i przerw w sieciach buforujących D2D przy użyciu modelu protokołu dla planowania przestrzennego współistniejących połączeń D2D. Golrezaei i inni badali zachowanie skalowania sieci D2D i podali wyrażenia w postaci zamkniętej dla optymalnej odległości współpracy jako funkcji parametrów modelu systemu. Kang i współpracownicy rozważają losową strategię buforowania dla stochastycznie rozproszonych urządzeń pamięci podręcznej, aby obsługiwać pobliskie urządzenia użytkowników i proponuje algorytm, aby znaleźć optymalne buforujące prawdopodobieństwa zawartości w bibliotece żądań.

### 7.3.1 SCN z włączoną pamięcią podręczną z lokalną korelacją interesów użytkownika

Wcześniejsze badania badające problemy z buforowaniem sieci bezprzewodowych często zakładają podobny wzorzec popularności dla wszystkich użytkowników w systemie. Uwzględnianie różnych poziomów zainteresowania w stosunku do żądanej treści Optymalne decyzje dotyczące buforowania są następnie podejmowane w każdym SBS przy użyciu rozproszonego uczenia się, aby zminimalizować całkowite opóźnienie usługi. Interesuje nas tutaj przypadek, w którym decyzje o buforowaniu podejmowane są w SBS w oparciu o wzorce lokalnych użytkowników, a nie o globalną popularność treści, ponieważ w SCN z włączoną pamięcią podręczną każdy SBS zwykle obsługuje niewielką liczbę użytkowników. W szczególności, każdy SBS tworzy swoją lokalną bibliotekę treści regularnych, pobierając regularnie pliki żądane przez użytkowników w zakresie próbkowania w zakresie próbkowania, a następnie określając zbiór plików, które mają być buforowane, z uwagi na ograniczoną pojemność pamięci podręcznej. Używamy tradycyjnej strategii "buforowania najpopularniejszych plików", aby przeanalizować wpływ korelacji odsetka użytkownika na prawdopodobieństwo trafienia / pominięcia pamięci podręcznej. Wyniki można uogólnić na bardziej zaawansowane strategie buforowania. Przyjmujemy dystrybucję podobną do Zipf dla lokalnej regularnej popularności zawartości w tym samym regionie, a my uzyskujemy prawdopodobieństwo usługi pamięci podręcznej jako funkcję maksymalnej odległości próbkowania. Uważamy również, że istnieje koszt pobierania próbek zależny od odległości i badamy kompromis między prawdopodobieństwem usługi buforowania a kosztem pobierania próbek.

#### 7.3.1.1 Model systemu

Rozważamy model SCN z pamięcią podręczną z ograniczoną przestrzenią dyskową w każdym SBS, obsługując jego pobliskich użytkowników w pewnej odległości dyktowanej przez ograniczenie mocy transmisji. Modelujemy dystrybucję SBS na dwuwymiarowej płaszczyźnie euklidesowej  $\mathbb{R}^2$  przez jednorodny proces Poissona (PPP)  $\Phi_u = \{Y_i, i \in \mathbb{N}^+\}$  z intensywnością  $\lambda_c$ , gdzie  $Y_i$  oznacza pozycję i-tego SBSa. Obsługiwani użytkownicy są dystrybuowani zgodnie z innym niezależnym jednorodnym PPP  $\Phi_u = \{x_j, j \in \mathbb{N}^+\}$  z intensywnością  $\lambda_u \geq \lambda_c$ , gdzie  $x_j$  oznacza pozycję j-tego użytkownika. Wprowadzamy pojęcie "wzoru lokalnych zainteresowań użytkowników", które powstaje, gdy każdy SBS pobiera regularnie żadaną treść wszystkich użytkowników w swoim obszarze próbkowania o promieniu  $R_c$ . Pamięci podręczne małych komórek decydują, które pliki mają być buforowane na podstawie przyjętej polityki buforowania. Zakładając, że istnieje maksymalna odległość, którą może obsłużyć SBS, oznaczona przez  $R_v$ , dysk wyśrodkowany na SBS z promieniem  $R_v$  może być postrzegany jako region

usług tego SBS. Zwiększając zakres próbkowania, SBS mogą mieć lepszą znajomość lokalnego wzoru zainteresowania użytkownika, ale z powodu ograniczonego przechowywania pamięci podręcznej mogą być przechowywane tylko najbardziej popularne pliki. W przypadku losowego użytkownika, który jest próbkowany, prawdopodobieństwo odnalezienia jego zwykłej zawartości buforowanej na swoich obejmujących SBS zmniejszy się wraz z rozmiarem regionu próbkowania. Ograniczenie zakresu usług daje także górną granicę prawdopodobieństwa, że dowolny użytkownik będzie obsługiwany przez małe pamięci podręczne. Oznaczamy to prawdopodobieństwo usługi przez  $P_{sv}$ . Zwiększenie zakresu próbkowania prowadzi również do wyższych kosztów poznania wzoru żądania lokalnego w SBS. Załóżmy, że każdy użytkownik ma bibliotekę rozmiaru  $J$ , która zawiera regularnie żądane pliki. Dla uproszczenia zakładamy, że wszystkie pliki mają równy rozmiar jednostki. Każdy SBS ma ograniczony rozmiar pamięci podręcznej oznaczony przez  $M$ , który jest maksymalną liczbą plików, które można przechowywać. Oznaczając  $N$  liczbę użytkowników w obszarze próbkowania SBS, z teorii procesów punktowych, mamy  $\mathbb{E}[N] = \lambda \pi R^2 \rho$ . Gdy użytkownicy z tego samego regionu próbkowania żądają podobnej treści, różni użytkownicy mogą nakładać się na pliki w swoich zwykłych bibliotekach treści, dlatego ogólny rozmiar biblioteki treści nie zwiększy się liniowo wraz z liczbą użytkowników w tym obszarze. Uwzględniając uwarunkowania posiadania  $N$  użytkowników w obszarze próbkowania typowego SBS, oznaczamy  $\ell = \{c_1, \dots, c_S\}$  lokalną bibliotekę treści  $N$  użytkowników z  $S$  reprezentującym rozmiar biblioteki. Oznaczmy przez  $g: \mathbb{N}^+ \rightarrow \mathbb{N}^+$  funkcję mapowania od liczby  $N$  użytkowników do lokalnej biblioteki rozmiaru  $S$ . W rzeczywistości  $S = g(N)$  może nauczyć się numerycznie na SBSach i przybliżoną funkcję można znaleźć poprzez dopasowanie danych. Tutaj zakładamy, że  $g$  jest funkcją odcinkową i następuje po  $S = \min(JN, \lceil J(1 + \mu \log N) \rceil)$ , gdzie  $\mu$  jest stałym czynnikiem charakteryzującym poziom podobieństwa lokalnego zainteresowania użytkownika. Dla danej lokalnej biblioteki treści regularnych, przyjmujemy, że dystrybucja popularności plików w niej działa zgodnie z prawem Zipfa-Mandelbrota, tzn. dla  $i$ -tego najpopularniejszego pliku mamy jego żądanie prawdopodobieństwa jako

$$p_i = \frac{\Omega}{(q + i)^\gamma}$$

gdzie

$$\Omega = \left( \sum_{n=1}^S (q + n)^\gamma \right)^{-1}$$

jest stałą normalizacyjną,  $q$  jest parametrem przesunięcia, co jest związane z przesunięciem w oparciu o dystrybucję Zipf, a  $\gamma$  jest kształtem parametr, który określa poziom koncentracji (korelacji) popularności treści. Podobnie jak funkcja  $g$ , można nauczyć się prawdziwej dystrybucji popularności pobieranie próbek i sortowanie popularności wszystkich żądanych plików w lokalnym regionie. Zauważ, że  $\mu$  w funkcji  $g$  definiuje jak podobne są zwykłe biblioteki treści użytkowników w tym samym obszarze, i  $\gamma$  w dystrybucji popularności określa względną rozbieżność popularności treści w lokalnej bibliotece. Połączenie tych dwóch czynników daje pełną charakteryzację poziomu korelacji lokalnego zainteresowania użytkowników.

### 7.3.1.2 Analiza wydajności

Analiza prawdopodobieństwa Usługi. Uważamy, że obowiązują zasady "buforowanie najpopularniejszych plików" i założmy, że jednostki pamięci podręcznej w SBS mają identyczny i ograniczony rozmiar pamięci jako  $M$  plików. Załóżmy, że mamy SBS u źródła z  $N$  użytkownikami w regionie pobierania próbek o promieniu  $R$ . Kiedy  $M \geq S$ , wszystkie pliki w lokalnej bibliotece treści

mogą być przechowywane w SBS. Gdy  $M < S$ , ze względu na ograniczone przechowywanie w pamięci podręcznej tylko  $M$  najbardziej popularnych plików w regionie próbkowania może być buforowany. Prawdopodobieństwo pominięcia pamięci podręcznej  $P_m$ , które jest prawdopodobne, że plik losowy w lokalnej bibliotece treści regularnych  $\ell$  nie jest buforowany, można podać jako

$$P_m = \sum_{i=M+1}^S p_i = \sum_{i=M+1}^S \frac{1}{(q+i)^\gamma \sum_{n=1}^S (q+n)^\gamma}.$$

Oczywiście,  $P_m$  wzrasta wraz z  $S$ , co weryfikuje intuicja, że większy rozmiar biblioteki treści prowadzi do większego prawdopodobieństwa utraty pamięci podręcznej, gdy pojemność pamięci podręcznej jest skończona. Zakładając, że użytkownicy spoza zakresu próbkowania SBS mają prawie pomijalne prawdopodobieństwo znalezienia ich regularnie żądanej treści w tym małym buforze podręcznym komórki, prawdopodobieństwo  $P_{sv}$  jest podawane przez prawdopodobieństwo przecięcia trzech zdarzeń:

- $\mathcal{E}_1$  = regularna biblioteka treści użytkownika, od którego pobierane są próbki przez co najmniej jeden SBS,
- $\mathcal{E}_2$  = żądana zawartość jest buforowana,
- $\mathcal{E}_3$  = użytkownika w zakresie usług SBS pobierania próbek.

Co więcej, mamy  $\mathcal{E}_1 \subseteq \mathcal{E}_3$  jeśli  $R_p \leq R_v$  i  $\mathcal{E}_3 \subset \mathcal{E}_1$  w przeciwnym razie. W przypadku kiedy  $R_p \leq R_v$ ,  $P_{sv}$  może być podane przez

$$P_{sv} = \mathbb{P} [\text{użytkownik próbkowany} \cap \text{zawartość buforowana}]$$

$$= 1 - \mathbb{P} [\text{nie będący próbkowanym}]$$

$$= 1 - \sum_{k=1}^{\infty} \mathbb{P} [\text{próbkowany przez } k \text{ cacherów}] \cdot \mathbb{P} [\text{brak pamięci podręcznej } k]$$

$$= 1 - e^{-\lambda_c \pi R_p^2} - \sum_{k=1}^{\infty} \frac{(\lambda_c \pi R_p^2)^k}{k!} \exp(-\lambda_c \pi R_p^2) \prod_{j=1}^k P_m^j$$

gdzie  $P_m^j$  jest prawdopodobieństwem pominięcia pamięci podręcznej  $j$ -tego SBS. Kiedy różne małe komórki mają różny poziomy korelacji żądań, co odpowiada różnym  $\gamma$  w rozkładzie Zipfa, prawdopodobieństwo pominięcia pamięci podręcznej można podać jako

$$P_m^j = \sum_{i=M+1}^S \frac{1}{(q+i)^{\gamma_j} \sum_{n=1}^S (q+n)^{\gamma_j}}$$

gdzie  $\gamma_j$  jest parametrem kształtu rozkładu popularności treści w próbce regionu  $j$ -tego SBS. Najwyraźniej  $P_{sv}$  wzrasta monotonicznie o  $R_p$ ; co oznacza, że większa odległość próbkowania daje większe prawdopodobieństwo przypadkowego użytkownika, aby znaleźć jego regularnie żadaną treść buforowaną w pokrywającym SBSie. W przypadku, gdy  $R_p > R_v$  prawdopodobieństwo usługi jest podane przez

$P_{sv} = \mathbb{P}$  [użytkownik w zakresie usług \ zawartość jest buforowana]

$$= 1 - e^{-\lambda_c \pi R_v^2} - \sum_{k=1}^{\infty} \frac{(\lambda_c \pi R_v^2)^k}{k!} \exp(-\lambda_c \pi R_v^2) \prod_{j=1}^k P_m^j.$$

W przeciwieństwie do poprzedniego przypadku, dla ustalonego  $R_v$  prawdopodobieństwo usługi maleje z  $R_p$ ; co jest łatwe do zrozumienia, ponieważ większy obszar próbkowania odpowiada wyższemu prawdopodobieństwu pominięcia pamięci podręcznej dla użytkowników próbkowanych, a liczba użytkowników, którzy mogą być obsłużeni nie zwiększa się wraz z zasięgiem próbkowania ze względu na zakres usług. Łącząc te dwa przypadki mamy:

$P_{sv} = \mathbb{P}$ [user w zakresie usług \ zawartość jest buforowana]

$$= 1 - e^{-\lambda_c \pi \min \{R_p, R_v\}^2} - \sum_{k=1}^{\infty} \frac{(\lambda_c \pi \min \{R_p, R_v\}^2)^k}{k!} \prod_{j=1}^k P_m^j.$$

Obserwujemy, że dla danego ograniczenia w zakresie usług SBS, usługa prawdopodobieństwo wzrasta z  $R_p$ , dopóki nie osiągnie limitu zakresu usług  $R_v$ ; potem to maleje monotonicznie o  $R_p$ .

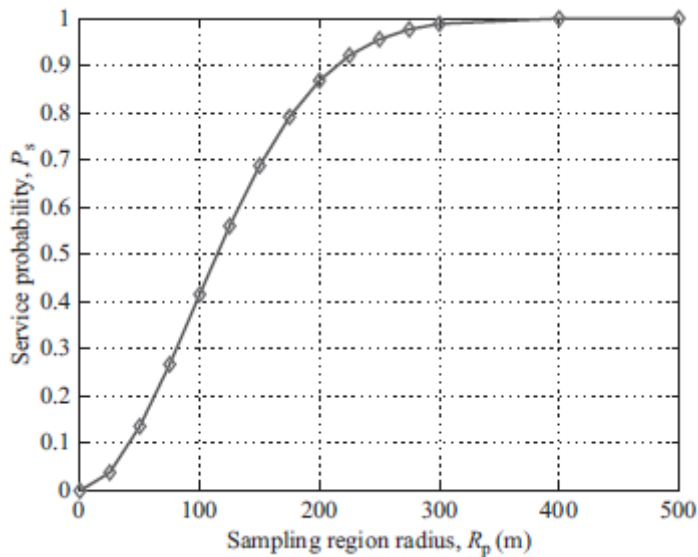
Analiza Kosztów Pobierania Próbek. Regularne pobieranie próbek umożliwia uczenie się SBSom w interesie lokalnych użytkowników, a tym samym może być usługą oferowaną przez małych cacherów komórkowych lepiej dostosowaną do wzoru żądania lokalnego. Z drugiej strony, pobór mocy w procedurze pobierania próbek musi być brany pod uwagę w odniesieniu do energii względem wydajności. Oznaczenie przez  $c_s$  kosztów pobierania próbek typowego SBS w miejscu pochodzenia z promieniem obszaru próbkowania  $R_p$ ; wprowadzamy funkcję kosztów do pobierania próbek, która zależy od odległości od próbkowanych użytkowników do SBS, tj.:

$$C_s = \sum_{j \in \Phi_u \cap \mathcal{B}(0, R_p)} f(\|x_j - Y_0\|) \cdot \eta$$

gdzie  $\|x_j - Y_0\|$  jest odległością euklidesową między j-tym użytkownikiem w obszarze próbkowania i typowym SBS w punkcie początkowym,  $f$  jest zależną od odległości funkcją kosztów, a  $\eta$  jest stałym współczynnikiem kosztu. Dla wygody ekspozycji wybieramy  $\eta = 1$  i model odwrotnej ścieżki dla  $f$  as  $f(x) = x^\alpha$ , gdzie  $\alpha$  jest wykładnikiem ścieżki strat. Z twierdzenia Campbella mamy średni koszt przy uśrednianiu różnych realizacji PPP jako

$$\begin{aligned}
\mathbb{E}[C_s] &= \mathbb{E} \left[ \sum_{j \in \Phi_u \cap \mathcal{B}(0, R_p)} \|x_j - Y_0\|^\alpha \right] \\
&= 2\pi\lambda_u \int_0^{R_p} r^\alpha \cdot r dr \\
&= \frac{2\pi\lambda_u}{\alpha + 2} (R_p^{\alpha+2} - 1).
\end{aligned} \tag{7.16}$$

Widzimy, że koszt pobierania próbek wzrasta liniowo wraz z gęstością użytkownika i podąża za prawem mocy rzędu  $\alpha + 2$  z promieniem obszaru próbkowania. Zwróć uwagę, że przy gęstych instalacjach sieci większa odległość próbkowania zapewnia lepsze poznanie schematu żądania; jednak koszt pobierania próbek może być zbyt wysoki i należy go wziąć pod uwagę przy projektowaniu optymalnej odległości próbkowania. Optymalny zakres próbkowania w ramach ograniczenia kosztów. Rysunek 7.4 pokazuje przykład prawdopodobieństwa usługi jako funkcji promienia regionu próbkowania, gdy nie ma ograniczeń w zakresie usług.



Widzimy, że zwiększający się zakres próbkowania daje większe prawdopodobieństwo obsługi, ale przyrost ma tendencję do nasycania się, gdy  $R_v$  jest względnie duży. Biorąc pod uwagę ograniczony zakres usług, ignorując koszt pobierania próbek, optymalny rozmiar obszaru próbkowania jest równy  $R_v$ , ponieważ daje on najwyższe prawdopodobieństwo obsługi. Teraz szukamy optymalnego zakresu próbkowania, który zapewnia wystarczająco dobre prawdopodobieństwo obsługi, a jednocześnie zapewnia, że koszt pobierania próbek jest poniżej danego ograniczenia. Problem sformułujemy w

$$R_p^* = \arg \max_{R_p} P_{sv}$$

Z zastrzeżeniem:

$$R_p \in [0, R_v],$$

$$C_s(R_p^*) < C_{\max}, \tag{7.18; 7.19}$$



gdzie  $C_{\max}$  to wstępnie zdefiniowane ograniczenie kosztów. Pierwszy warunek wynika z faktu, że gdy  $R_p > R_v$ , prawdopodobieństwo usługi maleje monotonicznie o  $R_p$ . Drugi warunek wynika z ograniczenia kosztów pobierania próbek. Za pomocą (7.16) możemy wyprowadzić ograniczenie na  $R_p$ , aby spełnić (7.19) jako

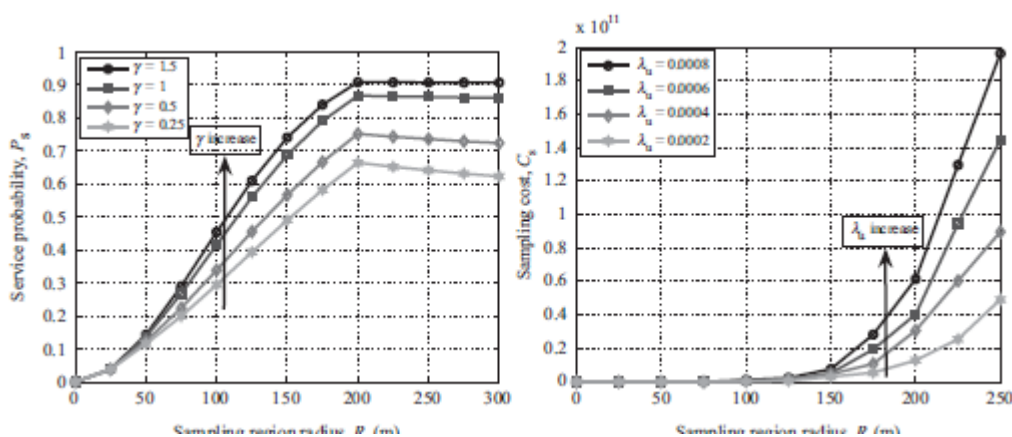
$$R_p^* < \left( \frac{C_{\max}(\alpha + 2)}{2\pi\lambda_u + 1} \right)^{\frac{1}{\alpha+2}}.$$

Ponieważ  $P_{sv}$  jest monotonicznie rosnącą funkcją  $R_p$  dla  $R_p \in [0, R_v]$ , łącząc te dwa ograniczenia mamy optymalną odległość próbkowania podaną przez

$$R_p^* = \min \left( R_v, \left( \frac{C_{\max}(\alpha + 2)}{2\pi\lambda_u + 1} \right)^{\frac{1}{\alpha+2}} \right).$$

### 7.3.1.3 Wyniki numeryczne

Zatwierdzamy naszą analizę sieci małych komórek z włączoną pamięcią podręczną i próbkowania lokalnego użytkownika, i pokazujemy kompromis między prawdopodobieństwem usługi pamięci podręcznej małych komórek a kosztem próbkowania dla różnych poziomów korelacji lokalnego użytkownika i gęstości użytkownika. Ustawiliśmy  $\lambda_c = 2 \times 10^{-5}$  SBS/m<sup>2</sup> i  $\lambda_u = 8 \times 10^{-4}$  użytkowników / m<sup>2</sup> jako gęstość SBS i użytkowników do obsłużenia. Obszar usług każdego SBS ma promień  $R_v = 200$  m. Standardowa wielkość biblioteki treści dla każdego użytkownika wynosi  $J = 10$ . Maksymalne przechowywanie pamięci podręcznej SBS wynosi  $M = 50$ . Używamy  $S = \min(JN, \lceil J(1 + 3 \log N) \rceil)$  jako funkcja odwzorowania od liczby użytkowników do ogólnego rozmiaru lokalnej biblioteki treści. Zakłada się, że dystrybucja popularności będzie zgodna z prawem Zipfa-Mandelbrota z  $q = 0$ . Przyjmujemy ten sam parametr kształtu dla rozkładu popularności treści w obszarze próbkowania wszystkich SBS, co oznacza, że  $\gamma_j = \gamma$  dla wszystkich  $j$ . Zauważ, że biorąc pod uwagę jednolite lub różne parametry kształtu dla rozkładu popularności, nie zmienia się ogólny trend, w jaki odległość próbkowania wpływa na prawdopodobieństwo usługi  $P_{sv}$ . Wyniki liczbowe przedstawiono z różnymi wartościami parametru kształtu  $\gamma \in [0,25, 1,5]$  które odpowiadają różnym poziomom korelacji między użytkownikami, w celu sprawdzenia, jak wydajność usługi buforowania skaluje się z lokalnym wzorcem zainteresowań użytkownika Na rysunku 7.5 (po lewej)



wykreślamy prawdopodobieństwo usługi  $P_{sv}$  jako funkcję zakresu próbkowania  $R_p$  dla  $\gamma = 0.25, 0.5, 1$  i  $1: 5$ , odpowiednio. Widzimy, że  $P_{sv}$  zwiększa się z  $R_p$ , zanim osiągnie  $R_v = 200$  m, a następnie maleje, gdy  $R_p$  rośnie. Jednak szybkość spadku jest stosunkowo niewielka w porównaniu do wzrostu w  $R_p \in$

[0,R<sub>v</sub>] ,szczególnie dla wyższych wartości  $\gamma$ . Wynika to z faktu, że wyższy  $\gamma$  odpowiada wyższej koncentracji popularności treści. Chociaż zwiększenie biblioteki rozmiar zwiększa liczbę plików, które nie są buforowane, najpopularniejsze są pliki buforowane z większym prawdopodobieństwem, co nadal prowadzi do mniejszego prawdopodobieństwa utraty pamięci podręcznej w regionie pobierania próbek. Rysunek 7.5 (po prawej) pokazuje symulowany koszt próbkowania na SBS jako funkcję maksymalnej odległości próbkowania z przyjętą funkcją kosztów zależną od odległości. Wyniki symulacji są uzyskiwane w różnych ustawieniach gęstości użytkownika i przy uśrednianiu ponad 5000 realizacji PPP. Wskazuje to na znaczny wzrost kosztów lokalnego pobierania próbek podczas zwiększania odległości próbkowania  $R_p$ . W scenariuszu o dużym zagęszczeniu użytkowników problem z kosztami zyskuje na znaczeniu, co wskazuje na konieczność uwzględnienia kosztu próbkowania przy wyszukiwaniu optymalnej odległości próbkowania. Podsumowując, zbadaliśmy SCN z obsługą pamięci podręcznej z próbkowaniem lokalnego użytkownika, zarządzanym przez SBS do nauki wzorców lokalnych. Głównym celem tego artykułu jest idea lokalnego regularnego pobierania próbek treści i wpływu odległości próbkowania na wydajność usługi pamięci podręcznej, co będzie korzystne dla projektowania procedury próbkowania SCN z włączoną pamięcią podręczną.

### 7.3.2 Buforowanie D2D a buforowanie małych komórek

W tej sekcji przeanalizujemy następujące pytanie: biorąc pod uwagę możliwości przechowywania obu SBS i urządzenia mobilnego, gdzie skuteczniejsze jest buforowanie treści? Innymi słowy, chcielibyśmy porównać (w miarę możliwości) buforowanie małych komórek z buforowaniem D2D. Chociaż takie porównanie może być sztuczne lub częściowo sprawiedliwe, ponieważ jest szczególnie wrażliwe na modele i założenia, niektóre pierwsze wnioski mogą być surowe. Na pierwszy rzut oka buforowanie D2D i buforowanie SC mogą przynieść porównywalne korzyści pod względem przepustowości sieci, wydajności spektralnej obszaru i efektywności energetycznej. Jednak, są one zupełnie różne w następujących aspektach:

- **Pojemność pamięci podręcznej.** Obiekty buforowania wdrożone w SBS mogą mieć bardzo duże możliwości przechowywania dzięki niskim kosztom jednostek pamięci masowej. W przeciwieństwie do tego urządzenia użytkownika, takie jak telefony komórkowe i tablety, mają stosunkowo małą pojemność i mogą obsługiwać jedynie niewielką liczbę żądań generowanych przez urządzenia znajdujące się w ich pobliżu.
- **Wysyłaj moc i zasięg.** Urządzenia użytkownika zwykle przesyłają z dużo mniejszą mocą (na przykład 23 dBm) niż SBS (na przykład 24-30 dBm), co z kolei odpowiada mniejszemu zakresowi pokrycia. W rezultacie prawdopodobieństwa sukcesu przesyłania wspomaganego pamięcią podręczną w powyższych dwóch przypadkach mogą być różne.
- **Gęstość żądań pamięci podręcznej.** Komunikacja D2D zwykle obejmuje małe odległości transmisji. W sieciach o dużej gęstości użytkowników, w przypadku buforowania D2D, więcej równoczesnych łączy może współistnieć w tym samym regionie, dzieląc te same zasoby widma w porównaniu z przypadkiem buforowania SC. Co więcej, szczególnym przypadkiem w buforowaniu D2D jest sytuacja, gdy użytkownik znajdzie żądany plik przechowywany w swoim własnym urządzeniu. W takim przypadku żądanie użytkownika jest spełnione bez żadnych opóźnień i kosztów.
- **Zużycie energii.** Oprócz poboru mocy urządzeń lub SBS, koszt energii z pobierania treści przez backhaul (w buforowaniu SC) może być znaczny. Gdy żądanie użytkownika nie może być spełnione lokalnie, poprzez komunikację D2D lub z pamięci podręcznych SBS, żądana treść zostanie pobrana z sieci rdzeniowej przez backhaul małej komórki. Ogólnie rzecz biorąc, pobór mocy w trybie backhaul jest znacznie wyższy niż koszt mocy transmisji.

Uwzględnimy wyżej wymienione różnice i porównamy wydajność buforowania w urządzeniach użytkowników (tj. buforowanie D2D) i buforowanie w SBS (tj. buforowanie SC) w przypadkowych topologiach. Udostępniamy wyrazy analityczne dla kluczowych wskaźników wydajności, w tym prawdopodobieństwo trafienia w pamięci podręcznej, gęstość żądań z pamięci podręcznej i średnie zużycie energii. Nasze wyniki pokazują, że wydajność sieci z włączoną pamięcią podręczną z buforowaniem D2D lub buforowaniem małych komórek zależy w dużym stopniu od gęstości użytkownika i dystrybucja popularności treści.

### 7.3.2.1 Model sieci

Uważamy, że sieć małych komórek (SCN), w której SBS są rozmieszczone w dwuwymiarowej płaszczyźnie euklidesowej  $\mathbb{R}^2$ , zgodnie z homogenicznym PPP  $\Phi_s$  o natężeniu  $\lambda_s$ . Użytkownicy mobilni są dystrybuowani według innej niezależnej homogenicznej PPP  $\Phi_u$  o intensywności  $\lambda_u$ . Możliwości buforowania można włączyć na urządzeniach użytkowników w celu potencjalnej komunikacji D2D, określanej jako buforowanie D2D, lub instalując jednostki pamięci w SBS, ukształtowane jako buforowanie SC. Każdy użytkownik mobilny wysyła losowe żądanie z prawdopodobieństwem  $p \in [0,1]$ ?. W rezultacie aktywni użytkownicy, którzy mają być obsługiwani, tworzą jednorodną PPP  $\Phi_u^r$  z intensywnością  $p\lambda_u$  (niezależne przeredzanie). Nieaktywni użytkownicy tworzą kolejny jednorodny PPP  $\Phi_u^t$  intensywnością  $(1-p)\lambda_u$ . Mogą one służyć jako potencjalne nadajniki D2D w przypadku trybu buforowania D2D lub pozostać bez dźwięku, jeśli komunikacja D2D nie jest włączona. W zależności od tego, czy pamięć podręczna jest włączona na urządzeniach lub na krawędzi / SBS, gdy aktywny użytkownik żąda pliku, mogą wystąpić następujące przypadki:

- tylko z buforowaniem D2D, jeśli żądany plik nie jest buforowany we własnym urządzeniu, użytkownik wyszukuje plik w urządzeniach znajdujących się w jego pobliżu w pewnej odległości. Jeśli istnieje co najmniej jeden potencjalny nadajnik D2D, który zażądał zbuforowanego pliku, plik jest przesyłany od najbliższego. W przeciwnym razie użytkownik łączy się z najbliższym SBS, aby pobrać plik z sieci rdzeniowej przez backhaul;
- tylko z buforowaniem SC, aktywny użytkownik zawsze łączy się z najbliższym SBS. Jeśli to jest powiązany SBS ma buforowany plik wewnątrz, SBS przesyła plik bezpośrednio do użytkownika. W przeciwnym razie plik jest pobierany z sieci rdzeniowej przez backhaul, a następnie przesyłany do użytkownika.

Zakładamy współdzielenie widma pomiędzy transmisjami współbieżnymi w takiej sieci, tj. zarówno łącza komunikacyjne D2D, jak i małe komórki odbierają zakłócenia od współistniejących nadajników. Uważamy, że biblioteka treści skończonych  $\mathcal{F} = \{f_1, \dots, f_N\}$  na żądanie użytkownika, gdzie  $f_i$  jest  $i$ -tym najpopularniejszym plikiem, a  $N$  jest rozmiarem biblioteki. Zakłada się, że wszystkie pliki są mają równy rozmiar, który jest znormalizowany do jednego. Używamy standardowego prawa Zipf dla rozkładu popularności, co oznacza, że prawdopodobieństwo żądania  $i$ -tej jest najbardziej popularne plik to

$$p_i = \frac{\Omega}{i^\gamma},$$

gdzie  $\Omega = \left( \sum_{j=1}^N j^{-\gamma} \right)^{-1}$

jest współczynnikiem normalizacji, a  $\gamma$  jest parametrem kształtu prawa Zipfa, które definiuje poziom korelacji żądań użytkowników. Wysokie wartości  $\gamma$  oznaczają, że większość żądań jest generowana z kilku najpopularniejszych plików. Dla użytkownika wykonując losowe żądanie,  $p_i$  może być postrzegane

jako prawdopodobieństwo, że żądany plik to  $f_i$ . Biorąc pod uwagę znajomość dystrybucji popularności treści, w zależności od tego, czy jest stosowane buforowanie D2D, czy buforowanie SC, stosujemy następujące zasady buforowania:

- tylko z buforowaniem D2D, ponieważ losowy aktywny użytkownik najprawdopodobniej będzie miał wiele potencjalnych nadajników D2D, każde urządzenie będzie niezależnie buforować pliki z zastrzeżeniem ograniczonego pojemnościowo magazynu zgodnie ze wspólnym rozkładem prawdopodobieństwa, w celu zwiększenia różnorodności treści w ramach wyszukiwania odległości. Optymalne prawdopodobieństwa zapisywania w pamięci podręcznej są określane przez minimalizację średniego prawdopodobieństwa pominięcia pamięci podręcznej;
- tylko z buforowaniem SC, ponieważ aktywny użytkownik zawsze łączy się z najbliższym SBS, nie ma nakładających się obszarów pokrycia różnych SBS. Dlatego stosujemy konwencjonalną "" pamięć podręczna najpopularniejszych treści "(MPC), co oznacza, że wszystkie SBS buforują te same najpopularniejsze pliki w ramach ich pamięci podręcznej.

Buforowanie D2D z rozmieszczeniem probabilistycznej pamięci podręcznej. Oznaczmy przez  $M_d$  rozmiar pamięci podręcznej urządzeń użytkownika, gdzie przy losowej zasadzie buforowania D2D, każdy użytkownik przechowuje  $f_i$  z prawdopodobieństwem  $q_i$ . Niech  $R_d$  będzie maksymalną odległością wyszukiwania / wykrywania urządzenia użytkownika dla ustanowienia komunikacji D2D. Prawdopodobieństwo, że żadne potencjalne przetworniki D2D nie zostaną znalezione, gdy jest żądane, jest równoważne z prawdopodobieństwem braku punktów z  $\Phi_u^t$  z niezależnym prawdopodobieństwem przeredzenia  $q_i$  w wyszukiwaniu obszar (puste prawdopodobieństwo). Niewłaściwe prawdopodobieństwo  $f_i$  znalezienia się w pamięci podręcznej w odległości wykrywania  $R_d$  jest podane przez

$$p_m^i(R_d) = e^{-\pi(1-\rho)\lambda_u q_i R_d^2}$$

Optymalny wektor prawdopodobieństwa buforowania  $q = [q_1; \dots; q_N]$  dla zminimalizowania średniego prawdopodobieństwa pominięcia pamięci podręcznej przez  $\mathcal{F}$  można znaleźć rozwiązując następujący problem optymalizacji:

$$\min_q \sum_{i=1}^N p_i e^{-\pi(1-\rho)\lambda_u q_i R_d^2}$$

z zastrzeżeniem:

$$\sum_{i=1}^N q_i - M_d \leq 0$$

$$q_i \in [0, 1].$$

Korzystamy z algorytmu optymalnego poszukiwania podwójnego rozwiązania (ODSA) aby znaleźć optymalne  $q$ . W pozostałej części artykułu zakłada się, że wszystkie wyniki związane z buforowaniem D2D są uzyskiwane przy użyciu optymalnych prawdopodobieństw buforowania.

SC Caching z polityką MP. Oznaczmy przez  $M_s$  pojemność bufora podręcznego SBS. Zgodnie z polityką buforowania MPC, tylko pliki o kolejności popularności  $i \in [1, M_s]$  byłby buforowany w każdym SBS.

### 7.3.2.2 Analiza wydajności

Potencjalne korzyści z buforowania treści bezprzewodowych są wychwytywane głównie przez prawdopodobieństwo trafienia w pamięci podręcznej, co daje możliwość obsługi żądań użytkowników i dostarczania treści bez konieczności pobierania jej z sieci rdzeniowej. Ponadto istnieje potencjał zysku z ponownego wykorzystania przestrzennego poprzez ustanowienie opartych na bliskości komunikacji z pamięcią podręczną, dzielących to samo widmo. W tej sekcji przedstawiamy wyniki analityczne dotyczące kilku kluczowych wskaźników wydajności dla sieci komórkowej z pamięcią podręczną z buforowaniem D2D lub buforowaniem SC. Prawdopodobieństwo trafienia w pamięci podręcznej to prawdopodobieństwo, że losowy aktywny użytkownik znajdzie żądany plik w lokalnych pamięciach podręcznych.

### **Buforowanie D2D**

W przypadku buforowania D2D żądanie trafienia w pamięci podręcznej może wystąpić w dwóch przypadkach:

- użytkownik żądający pliku może go znaleźć we własnej pamięci podręcznej, którą nazywamy "Własny wniosek";
- gdy żądany plik nie jest buforowany we własnym urządzeniu, użytkownik znajduje go w pamięci podręcznej w pobliżu potencjalnych nadajników D2D w odległości  $R_d$ .

Oznaczając przez  $p_{\text{self}}^d$  samo-żądanie prawdopodobieństwa losowego użytkownika, mamy:

$$p_{\text{self}}^d = \sum_{i=1}^N p_i q_i. \quad (7.26)$$

Prawdopodobieństwo, że aktywny użytkownik jest obsługiwany przez pobliski potencjalny nadajnik D2D jest dany przez

$$p_r^d = \sum_{i=1}^N p_i (1 - q_i) \left( 1 - e^{-\pi(1-\rho)\lambda_u q_i R_d^2} \right) \quad (7.27)$$

gdzie  $p_{\text{hit},i}^d = 1 - e^{-\pi(1-\rho)\lambda_u q_i R_d^2}$  jest prawdopodobieństwem posiadania co najmniej jednego potencjalnego nadajnika D2D w odległości  $R_d$  z plikiem z pamięci podręcznej. Dlatego prawdopodobieństwo trafienia w pamięć podręczną jest sumą (7.26) i (7.27), podaną przez

$$p_{\text{hit}}^d = p_{\text{self}}^d + p_r^d = 1 - \sum_{i=1}^N p_i (1 - q_i) e^{-\pi(1-\rho)\lambda_u q_i R_d^2}.$$

**Buforowanie SC**. Prawdopodobieństwo trafienia w pamięci podręcznej w przypadku buforowania SC jest po prostu prawdopodobieństwo, że użytkownik znajdzie żądany plik przechowywany w pamięci podręcznej skojarzonego z nim pliku SBS. Zgodnie z polityką buforowania MPC mamy:

$$p_{\text{hit}}^s = \sum_{i=1}^{M_s} p_i$$

Gdy aktywny użytkownik znajdzie żądany plik w pamięci podręcznej w pobliskich urządzeniach

lub najbliższy SBS, nie ma gwarancji, że przesyłanie pliku z asystą pamięci podręcznej zakończy się powodzeniem. Obliczamy tutaj prawdopodobieństwo sukcesu typowej kondycjonowania transmisji z pamięcią podręczną przy posiadaniu odbiornika u źródła. Dla danej realizacji sieci zakładamy, że  $K$  ustanowiło połączenia komunikacyjne wspomagane pamięcią podręczną z  $\mathbb{E}[K] \leq \rho\lambda_u$ . W przypadku losowego łącza  $i \in [1, K]$ . Stosunek sygnału do zakłóceń z szumem (SINR) jest określany przez

$$\text{SINR}_i = \frac{P_i |h_{i,i}|^2 d_{i,i}^{-\alpha}}{\sigma^2 + \sum_{j \in T \setminus \{i\}} P_j |h_{j,i}|^2 d_{j,i}^{-\alpha}}$$

gdzie  $P_i = \{P_d, P_s\}$  nadaje moc nadajnika D2D lub SBS, w zależności od tego, czy funkcje buforowania są włączone na urządzeniach mobilnych czy w SBS;  $h_{j,i}$  oznacza kanał o małej skali zanikający od nadajnika  $j$  do odbiornika  $i$ , który następuje po  $\ell\mathcal{N}(0,1)$  (zanik Rayleigha);  $d_{j,i}$  oznacza odległość między nadajnikiem  $j$  a odbiornikiem;  $\sigma^2$  oznacza moc szumów tła. Zauważ, że interferencja pochodzi nie tylko z transmisji wspomaganych pamięcią podręczną, ale również z linków użytkownika SBS, gdy żądany plik użytkownika nie jest lokalnie buforowany. Tak więc w przypadku buforowania D2D należy do zestawu aktywnych nadajników  $T \subseteq \{\Phi_s \cup \Phi_u\}$ . W przypadku buforowania SC,  $T \subseteq \Phi_s$  ponieważ aktywni użytkownicy mogą być obsługiwani tylko przez SBS. Zakładamy sieć ograniczoną interferencyjnie, w której szum termiczny tła jest pomijalny, a metryka będąca przedmiotem zainteresowania to stosunek sygnału do sprzężenia (SIR).

**D2D Caching.** Dla losowego aktywnego użytkownika żądającego pliku, z prawdopodobieństwem  $p_r^d$ , żądany plik nie jest buforowany we własnym urządzeniu, ale w jego urządzeniu w pobliżu. Dlatego gęstość łączy komunikacyjnych z pamięcią podręczną to  $\rho\lambda_u p_r^d$ . Pamiętaj, że wielu użytkowników może znaleźć ten sam najbliższy nadajnik D2D. W tym przypadku tylko jeden użytkownik może połączyć się z tym urządzeniem, inni muszą szukać innego nadajnika D2D. Oznaczenie przez  $\Phi_t^d$  zestawu aktywnych nadajników D2D; Chociaż wynikowy zestaw nie jest rozprawdzany zgodnie z jednorodnym PPP, gęstość  $\Phi_t^d$  jest podany przez

$$\lambda_t^d = \rho\lambda_u p_r^d.$$

Zestaw użytkowników, którzy nie mogą znaleźć swoich żądanych plików w lokalnych pamięciach podręcznych, będzie obsługiwany przez SBS. Gęstość użytkowników obsługiwanych przez SBS to

$$\lambda_r^s = \rho\lambda_u(1 - p_{\text{hit}}^d)$$

. Według najbliższego stowarzyszonym SBSem generowana jest tesselacja Poissona-Voronoi. Prawdopodobieństwo pustki typowej komórki Voronoi można określić w przybliżeniu jako

$$p_{\text{void}} \simeq \left(1 + \frac{\lambda_r^s}{3.5\lambda_s}\right)^{-3.5}.$$

Gęstość aktywnych SBS jest zatem określana przez

$$\begin{aligned} \lambda_t^s &= \lambda_s(1 - p_{\text{void}}) \\ &\simeq \lambda_s \left(1 - \left(1 + \frac{\rho\lambda_u(1 - p_{\text{hit}}^d)}{3.5\lambda_s}\right)^{-3.5}\right). \end{aligned} \quad (7.32)$$

Uwarunkowanie posiadania typowego odbiornika D2D w miejscu jego pochodzenia nadajnik przy odległości  $d_x$  i przyjmując jednorodną PPP dla obu aktywnych D2D nadajniki  $\Phi_t^d$  i aktywne SBS  $\Phi_t^s$ , prawdopodobieństwo powodzenia jest podane jako

$$\begin{aligned} P_{\text{suc}}^d &= \mathbb{P} \left[ \frac{P_d |h_{i,i}|^2 d_x^{-\alpha}}{\sum_{j \in \Phi_t^d \setminus \{i\}} P_d |h_{j,i}|^2 d_{j,i}^{-\alpha} + \sum_{k \in \Phi_t^s} P_s |h_{k,i}|^2 d_{k,i}^{-\alpha}} > \theta \right] \\ &= \mathbb{E} \left[ \mathcal{L}_{I_d}(\theta d_x^\alpha) \cdot \mathcal{L}_{I_s} \left( \theta \frac{P_s}{P_d} d_x^\alpha \right) \right] \\ &= \mathbb{E} \left[ \exp \left( -\frac{\pi d_x^2 \theta^\alpha}{\text{sinc}(2/\alpha)} \left( \lambda_t^d + \lambda_t^s (P_s/P_d)^{\frac{2}{\alpha}} \right) \right) \right] \end{aligned}$$

gdzie  $\ell_{I_x} = \mathbb{E}[\exp(-sI_x)]$  jest przekształceniem Laplace'a interferencji  $I_x$ , a  $\theta$  jest progami SIR dla pomyślnej transmisji D2D. Oczekiwanie jest ponad dystrybucję  $d_x$  i przez bibliotekę treści  $\mathcal{F}$ . Kiedy nastąpi trafienie w pamięci podręcznej, dystrybucja odległości D2D  $d_x$  zależy od kolejności popularności żądanego pliku. Jeśli typowy użytkownik zażąda  $f_i$ , warunkując posiadanie co najmniej jednego potencjalnego nadajnika D2D w odległości  $R_d$ , to pdf odległości D2D jest podana przez

$$f_{d_x}^i(r) = \begin{cases} \frac{2\pi(1-\rho)\lambda_u q_i r}{1 - e^{-\pi(1-\rho)\lambda_u q_i R_d^2}} e^{-\pi(1-\rho)\lambda_u q_i r^2} & 0 \leq r \leq R_d \\ 0 & r > R_d. \end{cases}$$

Następnie przybliżone prawdopodobieństwo sukcesu, gdy trafienie w pamięci podręcznej pliku  $f_i$  ma miejsce, jest podane przez

$$\begin{aligned} P_{\text{suc},i}^d &= \int_0^{R_d} \exp \left[ -\frac{\pi r^2 \theta^\alpha}{\text{sinc}(2/\alpha)} \left( \rho \lambda_u P_r^d + \lambda_t^s (P_s/P_d)^{\frac{2}{\alpha}} \right) \right] \\ &\quad \cdot \frac{2\pi(1-\rho)\lambda_u q_i r}{1 - e^{-\pi(1-\rho)\lambda_u q_i R_d^2}} e^{-\pi(1-\rho)\lambda_u q_i r^2} dr, \end{aligned} \tag{7.36}$$

gdzie  $pdr$  i  $Ist$  podane są odpowiednio w (7.27) i (7.32).

**SC Caching.** W przypadku buforowania SC użytkownicy są zawsze połączeni z najbliższymi SBS w przypadku trafień pamięci podręcznej i zdarzeń pamięci podręcznej. Oznaczmy przez  $\tilde{\Phi}_t^s$  zestaw aktywnych SBS, podobnie, używając prawdopodobieństwa pustki komórek Voronoi, gęstość aktywnych SBS jest określona przez

$$\tilde{\lambda}_t^s \simeq \lambda_s \left( 1 - \left( 1 + \frac{\rho \lambda_u}{3.5 \lambda_s} \right)^{-3.5} \right) \tag{7.37}$$

Uwarunkowanie posiadania typowego odbiornika na początku z powiązaniem z nim SBS odległość  $d_s$  i przy użyciu najbliższego stowarzyszenia SBS, mamy pdf  $d_s$  podany jako

$$f_{d_s}(r) = 2\pi\lambda_s r \cdot e^{-\pi\lambda_s r^2}.$$

Dla danego progu SIR  $\theta$ , prawdopodobieństwo powodzenia transmisji małych komórek wspomaganych pamięcią podręczną jest podawane przez

$$\begin{aligned} p_{\text{suc}}^s &= \mathbb{P} \left[ \frac{P_s |h_{i,i}|^2 d_s^{-\alpha}}{\sum_{k \in \Phi_t^s \setminus \{i\}} P_s |h_{k,i}|^2 d_{k,i}^{-\alpha}} > \theta \right] \\ &= \int_0^\infty 2\pi\lambda_s r \cdot e^{-\pi\lambda_s r^2} \exp \left( -\frac{\pi\tilde{\lambda}_t^s r^2 \theta^{\frac{2}{\alpha}}}{\text{sinc}(2/\alpha)} \right) dr, \end{aligned} \quad (7.39)$$

gdzie  $\tilde{\lambda}_t^s$  podano w (7.37).

### Gęstość żądań pamięci podręcznej

Żądanie użytkownika mówi się, że jest "obsługiwane", jeśli żądany plik znajduje się w lokalnych pamięciach podręcznych i jeśli transmisja pliku zakończyła się powodzeniem. W oparciu o powyższe wyniki obliczamy gęstość żądań obsługiwanych w pamięci podręcznej, która jest średnią liczbą żądań, które mogą być z powodzeniem i jednocześnie obsługiwane przez lokalną pamięć podręczną na jednostkę powierzchni.

**D2D Caching.** W przypadku buforowania D2D losowe żądanie użytkownika może być obsługiwane przez żądanie własne lub przez proksymalną komunikację D2D. Oznaczmy przez  $\mu_{\text{suc}}^d$  gęstość żądań pamięci podręcznej, mamy:

$$\mu_{\text{suc}}^d = \rho\lambda_u \left( p_{\text{self}}^d + \sum_{i=1}^N p_i (1 - q_i) p_{\text{hit},i}^d p_{\text{suc},i}^d \right)$$

gdzie  $p_{\text{suc},i}^d$  jest podana w (7.36).

**SC Caching.** W przypadku buforowania SC maksymalna liczba transmisji wspomaganych pamięcią podręczną w danym przedziale czasowym jest ograniczona przez gęstość SBS. Prawdopodobieństwo, że SBS ma co najmniej jednego aktywnego użytkownika w swojej komórce żądającej plików przechowywane w pamięci podręcznej jest podane przez

$$p_t^s = 1 - \left( 1 + \frac{\rho\lambda_u p_{\text{hit}}^s}{3.5\lambda_s} \right)^{-3.5}.$$

Gęstość transmisji wspomaganej przez pamięć podręczną to  $\lambda_s p_t^s$ . Następnie gęstość obsługi pamięci podręcznej żądań, która jest gęstością pomyślnej transmisji małych komórek obsługiwanych, jest podana przez

$$\mu_{\text{suc}}^s = \lambda_s p_t^s p_{\text{suc}}^s = \lambda_s \left( 1 - \left( 1 + \frac{\rho\lambda_u p_{\text{hit}}^s}{3.5\lambda_s} \right)^{-3.5} \right) p_{\text{suc}}^s$$

gdzie  $p_{\text{suc}}^s$  jest podany w (7.39).



### 7.3.2.3 Zużycie energii

W przypadku losowego żądania użytkownika, w zdarzeniu trafienia w pamięci podręcznej, zużyta moc dostarczania treści zawiera tylko moc nadania nadajnika D2D lub powiązanego SBS. W przypadku braku pamięci podręcznej żądany plik jest najpierw pobierany z sieci rdzeniowej przez backhaul, a następnie przesyłany z najbliższego SBS do użytkownika. Tak więc dodatkowa energia zużywana jest w backhaul. Oznaczmy  $P_b$  zapotrzebowanie mocy backhaul wymagane do obsługi żądania użytkownika w jednym SBS; badamy poniżej pobór mocy na żądanie użytkownika z buforowaniem D2D lub buforowaniem SC. W przypadku pamięci podręcznej D2D pamiętaj, że w przypadku wystąpienia żądania własnego, żadna energia nie jest zużywana na żądanie. Losowe żądanie użytkownika ma prawdopodobieństwo, że  $p_r^d$  będzie obsługiwane przez pobliski nadajnik D2D, a prawdopodobieństwo  $1 - p_{hit}^d$  trafi do najbliższego SBS. Moc zużywana na żądanie użytkownika jest podana jako

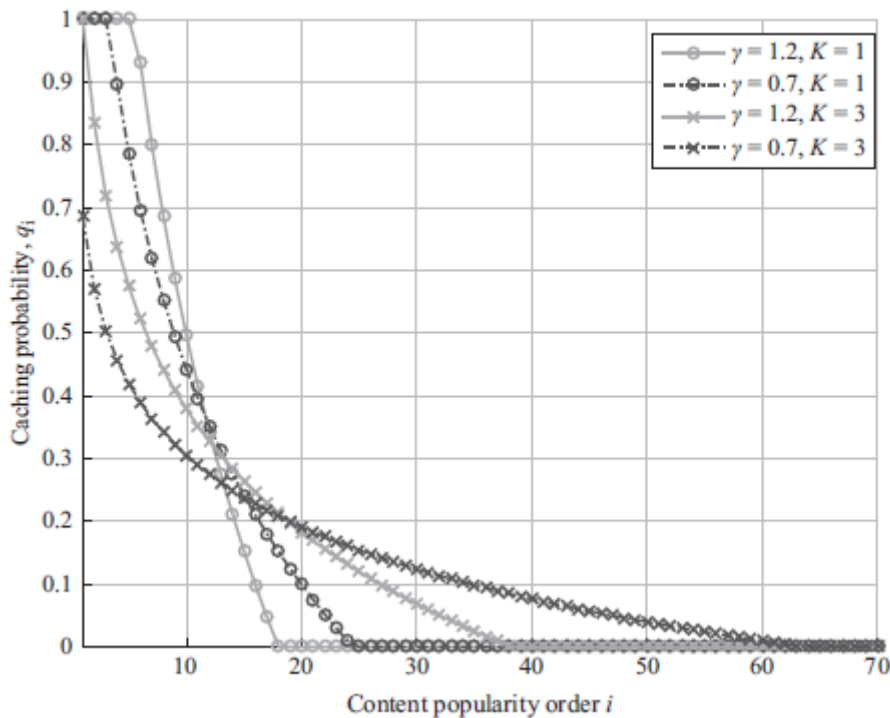
$$P_{avg}^d = p_r^d P_d + (1 - p_{hit}^d)(P_s + P_b).$$

Dla przypadku z buforowaniem SC, gdy wystąpi losowe żądanie użytkownika, transmisja moc najbliższego SBS jest zawsze zużywana zarówno w przypadku trafień pamięci podręcznej, jak i pamięci podręcznej. Moc backhaul jest dodatkowo zużywana z prawdopodobieństwem  $1 - p_{hit}^s$ . Tak więc mamy:

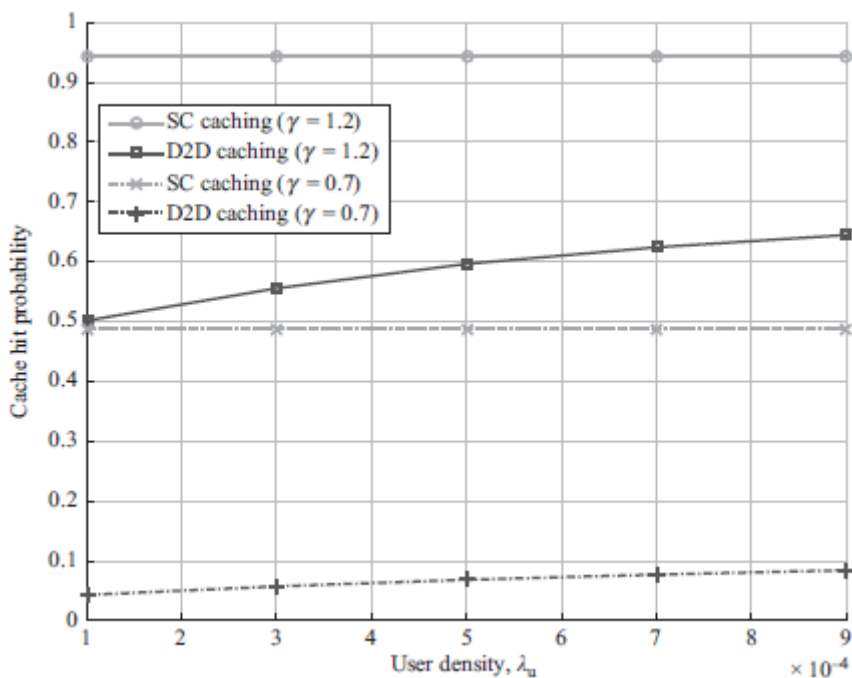
$$P_{avg}^s = P_s + (1 - p_{hit}^s)P_b.$$

### 7.3.2.4 Wyniki numeryczne

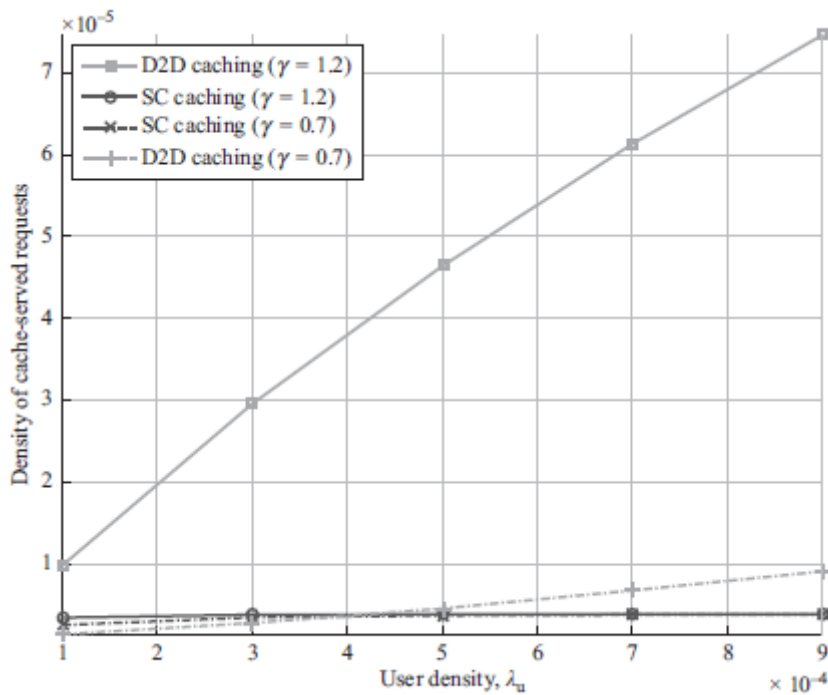
Dla oceny numerycznej ustawiamy  $\lambda_s = 10^{-5}$  SBS / m<sup>2</sup> i  $\lambda_u = K \times 10^{-4}$  użytkowników / m<sup>2</sup>, gdzie K jest współczynnikiem proporcji. Wybieramy  $\rho = 0,2$ , co oznacza, że 20% użytkowników poprosi o plik w bibliotece zawartości  $\mathcal{F}$ . Rozmiar pamięci podręcznej w SBS i na urządzeniu użytkownika to odpowiednio  $M_s = 10^4$  i  $M_d = 10$ , odpowiednio. Biblioteka treści ma rozmiar  $N = 10^5$ , a dystrybucja popularności podana jest przez prawo Zipfa z parametrem kształtu  $\gamma = 0,7$  dla przypadku o małej skośności popularności i  $\gamma = 1,2$  dla drugiego przypadku. Odległość wyszukiwania urządzenia użytkownika do ustalenia łącza D2D wynosi  $R_d = 75$  m. Moc nadawania SBS i urządzenia użytkownika wynosi odpowiednio  $P_s = 100$  mW i  $P_d = 2$  mW, odpowiednio. Zużycie mocy w SBS na żądanie użytkownika wynosi  $P_b = 10$  W. Docelowy SIR dla pomyślnych transmisji jest wybierany jako  $\theta = 0$  dB. Prezentujemy wyniki liczbowe dla  $K \in [1,9]$ , w celu porównania wydajności pomiędzy buforowaniem na urządzeniach użytkownika i buforowaniem w SBS dla różnych użytkowników reżimy gęstości. Na rysunku 7.6



wykreślamy optymalne prawdopodobieństwo buforowania  $q_i$  pliku  $f_i$ ,  $i \in [1, N]$ , uzyskane przez rozwiązanie problemu optymalizacji. Porównując wyniki z  $\gamma = 1,2$  i z  $\gamma = 0,7$ , widzimy to dla niższych; użytkownicy mają tendencję do buforowania większej liczby różnych plików, ponieważ żądania użytkowników są bardziej zróżnicowane ze względu na niski poziom koncentracji popularności. Obserwujemy ten sam trend, gdy gęstość użytkownika jest wyższa, np.  $K = 3$ : Jest to uzasadnione, ponieważ przy większej gęstości użytkownika prawdopodobieństwo ustanowienia komunikacji D2D jest wyższe. Tak więc, więcej różnych plików powinno być buforowanych w urządzeniach użytkowników w celu obsłużenia większej liczby żądań przez transmisję D2D wspomaganą pamięcią podręczną. Rysunek 7.7 pokazuje



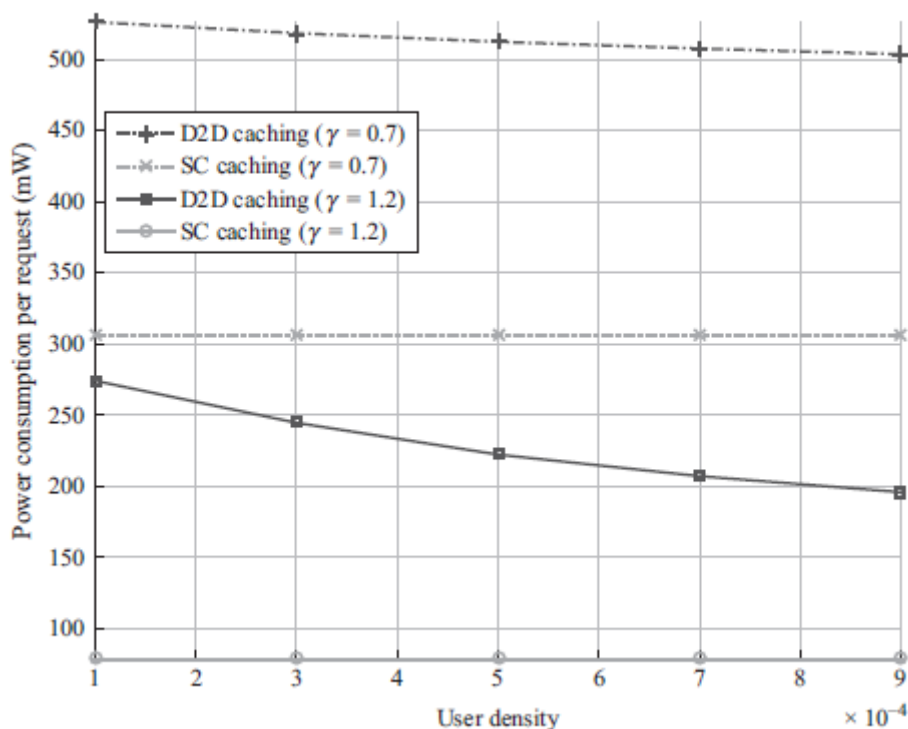
prawdopodobieństwo trafienia w pamięci podręcznej dla przypadku z buforowaniem D2D i z buforowaniem SC, odpowiednio. Zgodnie z oczekiwaniami prawdopodobieństwo trafienia w pamięci podręcznej dla obu przypadków jest wyższe przy niższych wartościach  $\lambda_s$ . Co więcej, widzimy, że buforowanie w SBS powoduje znacznie wyższe prawdopodobieństwo trafienia w pamięci podręcznej niż buforowanie na urządzeniach mobilnych, w wyniku większej pojemności pamięci podręcznej. Gęstość żądań obsługiwanych w pamięci podręcznej określa, ile żądań można obsłużyć jednocześnie, używając lokalnych pamięci podręcznych. Z rysunku 7.8 wynika, że



buforowanie D2D przewyższa buforowanie SC dla wyższych wartości  $\gamma$ , szczególnie w przypadku systemu o dużej gęstości użytkowników. W przypadku z niższym  $\gamma$ ; wydajność buforowania SC jest nieco lepsza w systemie rzadkich użytkowników, podczas gdy wydajność buforowania D2D przewyższa buforowanie SC po zwiększeniu gęstości użytkownika. Zaletą buforowania D2D wynika głównie z dwóch powodów:

- gdy gęstość użytkownika jest wysoka, istnieje liczba potencjalnych nadajników D2D ogólnie większy niż liczba SBS, co pozwala na większą równoczesność łącza transmisyjne wspomagane pamięcią podręczną;
- Samodzielne żądanie buforowania D2D daje możliwość obsługi dużej liczby żądań użytkowników, gdy popularność treści jest silnie skoncentrowana.

Rysunek 7.9 przedstawia



średnie zużycie energii na żądanie użytkownika, pokazując, że buforowanie SC jest prawdopodobnie bardziej energooszczędne niż w przypadku buforowania D2D, co wynika głównie z wysokiego prawdopodobieństwa trafienia w pamięci podręcznej. W porównaniu do mocy pobieranej do pobierania treści przez backhaul, moc nadawania zarówno urządzeń użytkowników, jak i SBS jest stosunkowo niska. Dlatego im większe jest prawdopodobieństwo, że lokalnie obsłuży się żądanie użytkownika, potrzebna jest mniejsza moc. Zauważamy również, że zużycie energii na żądanie użytkownika z buforowaniem D2D maleje wraz z gęstością użytkownika. Dzieje się tak dlatego, że liczba potencjalnych nadajników D2D w odległości wykrywania od użytkownika zwiększa się, gdy zwiększa się gęstość użytkownika, co daje większe prawdopodobieństwo obsługi żądania użytkownika przez transmisję D2D wspomaganą pamięcią podręczną. Łącząc te wyniki, otrzymujemy następujące wiadomości na wynos:

- w sieciach o wysokim zagęszczeniu użytkowników, buforowanie D2D daje możliwość obsługi większej liczby żądań użytkowników jednocześnie za pomocą D2D wspomaganego pamięcią podręczną na krótkich dystansach komunikacja;
- Buforowanie krawędzi (SBS) skutkuje znacznie wyższym prawdopodobieństwem trafienia w pamięci podręcznej, ponieważ jednostki pamięci w SBS mają znacznie większą pojemność niż na urządzeniach mobilnych. W rezultacie buforowanie SC jest także bardziej energooszczędne, ponieważ w trybie domyślnym zużywa się średnio mniej energii w celu pobrania pliku z sieci rdzeniowej.

### 7.3.2.5 Dyskusja

Porównaliśmy wydajność buforowania D2D i buforowania małych komórek w przestrzennej sieci bezprzewodowej. Głównym przesłaniem na wynos jest to, że w sieciach bezprzewodowych o wysokim zagęszczeniu użytkowników korzystne jest, jeśli chodzi o przestrzenne ponowne wykorzystanie zasobów komunikacyjnych w celu wykonania buforowania na urządzeniach mobilnych i korzystania z proksymalnej komunikacji D2D. Z drugiej strony, rozmieszczanie jednostek pamięci o dużej pojemności do buforowania na krawędzi (tj. w małych komórkach) jest bardziej korzystne pod względem

wydajności energetycznej, w wyniku wysokiego prawdopodobieństwa trafienia w pamięć podręczną w takim przypadku.

#### **7.4 Wyzwania badawcze dotyczące buforowania bezprzewodowego**

Tu podaliśmy matematyczne dowody korzyści z buforowania dla działania przyszłych sieci bezprzewodowych. Uzupełniamy rozdział krótko podsumowując najtrudniejsze problemy i najciekawsze przyszłe kierunki badań w zakresie buforowania w sieciach bezprzewodowych.

##### **7.4.1 Modelowanie żądań w sieciach bezprzewodowych**

Standardowym modelem do analizy wydajności buforowania sieci jest Niezależny Model Referencyjny (IRM): zawartość  $m$  jest wymagana zgodnie z niezależnym procesem Poissona z prędkością  $\lambda p_m$ ; gdzie  $p_m$  odnosi się do popularności treści, wzorowanej na prawie mocy, tj.  $p_m \propto m^{-\tau}$ . Ten dobrze ugruntowany model rozwija się dzięki swojej prostocie; ma tylko dwa parametry,  $\lambda$  aby kontrolować szybkość żądań, i  $\tau$  aby kontrolować skos popularności. IRM zakłada, że popularność treści jest statyczna, co oczywiście nie jest prawdą. Najświeższe wiadomości i odcinki seriali telewizyjnych są przykładami efemerycznych treści o szybko zmieniającej się popularności; pojawiają się, stają się coraz bardziej popularne i stopniowo stają się niepopularne. Ostatnie prace proponowały nowatorskie narzędzia matematyczne do modelowania popularności treści w różnym czasie. Badania uwzględniają duże próbki aplikacji YouTube i VoD i odkrywają: popularność treści jest zmienna w czasie, a ponadto ma stały wpływ na analizę wydajność buforowania. W niejednorodnym modelu Poissona proponowanym w każdej treści jest powiązany strzał, którego czas trwania odzwierciedla długość życia treści, a którego wysokość oznacza jego natychmiastową popularność. Model ten nosi nazwę Model Szumu Śrutowego (SNM), odzwierciedlający szum Poissona z elektroniki. Podczas gdy kształt pulsu nie jest istotny, badanie wykazuje silne korelacje między popularnością i czasem trwania; podobno popularne treści rozwijają się dłużej. Wreszcie, model oparty na klasach może wygodnie wychwycić korelacje czasowo-przestrzenne, jednocześnie umożliwiając analityczną łatwopalność. Użytkownicy mobilni to gruntownie pobierający treści efemeryczne, dlatego oczekuje się, że podobna analiza dotycząca treści bezprzewodowych wykaże silniejsze efekty. Dodatkowo, zmiany czasu wymagają nieustannej oceny popularności, co jest trudniejsze w sieciach bezprzewodowych, ponieważ pamięci podręczne są używane przez mniejsze populacje i stąd "widzą" sekwencję żądania sparser. W rzeczywistości koncepcja zmieniającej się w czasie popularności nie jest zupełnie nowa dla inżynierów pracujących z pamięcią podręczną. Starsze analizy systemów buforowania odwołują się do tego terminu "lokalizacja czasowa", gdzie pojawiają się impulsy żądań dla pewnych treści. Obecnie internetowe systemy buforowania działają w oparciu o strategię dynamicznej eksmisji, takie jak Least-Recently-Used (LRU) i Least-Frequently-Used (LFU), których celem jest uchwycenie tymczasowej lokalizacji w sposób heurystyczny. Wydajność LRU w modelu SNM jest analizowana w. Ostatnie badanie analizuje model SNM i zapewnia optymalną politykę wspólnego buforowania i szacowania popularności. Wynik pokazuje, że oszacowanie częstotliwości (podobne do LFU) jest optymalne dla zawartości istniejącej w systemie przez pewien czas i otrzymało już kilka żądań. Ciągłe szacowanie popularności jest trudniejsze w systemach bezprzewodowych, w których żądania ludności są podzielone na wiele pamięci podręcznych, a te, które docierają do dowolnej pamięci podręcznej, są mniejsze. Ten rozwój motywuje nowe techniki buforowania, które wykorzystują metodologie uczenia się, aby dokładnie śledzić ewolucję popularności treści w czasie.

##### **7.4.2 Kodowane buforowanie dla eksploatacji medium transmisyjnego**

Podstawowa praca Maddah-Ali i Niesen zaproponowała wykorzystanie buforowania w celu poprawy wydajności transmisji bezprzewodowych w komórce. Podczas gdy tradycyjne techniki przechowują całą zawartość w terminalach mobilnych w celu wyeliminowania retransmisji popularnych treści,

kodowana technika buforowania przechowuje zakodowane kombinacje części treści. Następnie BS może nadawać zakodowaną sekwencję wiadomości, tak aby wszystkie odbiorniki w grupie mogły łączyć ją z fragmentami pamięci podręcznej, aby dekodować zawartość. Pokazano, że generuje bloki zasobów równe  $N(1-K/M)/(1+NK/M)$ , gdzie  $N$  to liczba użytkowników,  $K$  rozmiar pamięci podręcznej, a  $M$  rozmiar katalogu. Jeśli zatem część pamięci  $K/M$  z możliwością zapisu jest ustalona, to wymagana liczba bloków zasobów nie zwiększa się wraz z liczbą użytkowników  $N$ , można to sprawdzić, przyjmując limit  $N \rightarrow \infty$ , w którym powyższa wielkość zbiega się z stałą. Innymi słowy, w scenariuszu bezprzewodowej BS z  $N$  użytkownikami oglądającymi poszczególne treści, wynik pokazuje, że możemy zwiększyć  $N$  w nieskończoność przy zachowaniu stałej jakości usługi. Ta rewolucyjna idea zgromadziła wiele wysiłków badawczych, takich jak sieci device-to-device (D2D), niejednorodne treści popularnonaukowe, internetowe strategie buforowania, multi-serwery, wiele bibliotek i połączenie z CSI. Z punktu widzenia realizacji, obiecujące kierunki badań obejmują rozszerzenia do przechwytywania aspektów systemowych, takich jak (i) skośność popularności, (ii) żądania asynchroniczne, (iii) skończone długości kodu oraz (iv) rozmiary buforów podręcznych, które skalują wolniej niż  $M$ . Zakładając, że te praktyczne wyzwania zostały rozwiązane, buforowanie dla systemu bezprzewodowego zostaną powiązane z zastosowanymi technikami warstwy fizycznej w BS i podręcznej.

#### 7.4.3 Wspólne modele buforowania

W kontekście heterogenicznych sieci bezprzewodowych użytkownik może łączyć się z siecią w różny sposób w celu uzyskania żądanej treści. Na przykład współczesne smartfony otrzymują sygnał o więcej niż 10 BS jednocześnie. W przyszłych zagęszczonych sieciach komórkowych telefon komórkowy będzie połączony z kilkoma femto-, pico- lub nano-komórkami. Można to wykorzystać przez wspólne techniki buforowania. Ponieważ użytkownik może pobrać żadaną treść z wielu punktów końcowych sieci, sąsiednie pamięci podręczne powinny współpracować i unikać wielokrotnego przechowywania tych samych obiektów. Większość optymalizacji umieszczania treści w pamięci podręcznej sieci bezprzewodowej sprowadza się do zestawu zasłania problem w dwuczściowym wykresie łączącym użytkowników z dostępnymi pamięciami podręcznymi. Dlatego znalezienie zawartości do zapisania w pamięci podręcznej jest trudnym problemem, nawet jeśli popularność jest znana. Możliwe jest złagodzenie problemu w celu wypukłej optymalizacji za pomocą rozproszonych kodów pamięci, w których każda pamięć podręczna przechowuje zakodowane kombinacje zawartości. Pomysły te prowadzą do kilku interesujących algorytmów w literaturze o buforowaniu wielodostępowym.

#### 7.4.4 Zaawansowane modele przestrzenne

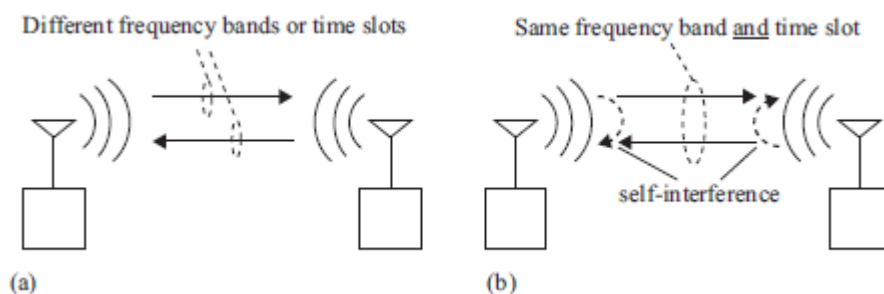
Geometria stochastyczna i teoria procesów punktowych odegrały znaczącą rolę w określeniu wydajności sieci bezprzewodowych na dużą skalę i zrozumieniu wpływu interferencji i losowej topologii na różne metryki, takie jak zasięg, obszar wydajności widmowej i prawdopodobieństwo trafienia. PPP było szeroko stosowane do modelowania podmioty sieciowe, głównie ze względu na jej podatność. Chociaż PPP zapewnia dobre aproksymacja zakłóceń związanych z odpychającymi procesami punktowymi, topologią sieci, w szczególności sieci bezprzewodowe z włączoną pamięcią podręczną, mogą obejmować inne złożone interakcje, topologiczne struktury i korelacje między jednostkami sieciowymi. Poznanie nowych procesów punktowych, które równoważą kompromis między podatnością i praktycznością, do modelowania sieci bezprzewodowych z włączoną pamięcią podręczną jest trudnym, ale istotnym kierunkiem badań. W związku z tym podjęto kilka pierwszych prób przeanalizowania pamięci podręcznej SC i D2D przy użyciu procesów dwumianowych i procesów w procesach klastrowych

# Pełny duplex

## 8.1 Wprowadzenie

Ewolucja bezprzewodowych sieci komunikacyjnych przy stale rosnącym ruchu danych wymaga ulepszeń interfejsu lotniczego w celu wydajniejszego wykorzystania skończonych zasobów radiowych. Transmisja full-duplex (FD) jest uważana za obiecującą technikę interfejsu dla przyszłych systemów bezprzewodowych, ponieważ zajmuje się kluczowymi zagadnieniami, takimi jak sprawność widmowa, pojemność łącza i opóźnienia. Dlatego operatorzy sieci bezprzewodowych obejmują FD jako potencjalnego kandydata technologii dla systemów 5G we wspólnej wizji 5G. FD, lub bardziej bezpośrednio w pasmie FD, koncepcja transmisji różni się zasadniczo od konwencjonalnych schematów duplexu, tj. Dupleksowania z podziałem częstotliwości (FDD) i dupleksowania z podziałem czasu (TDD), które są stosowane w starszych bezprzewodowych systemach komunikacyjnych. Podczas gdy konwencjonalne schematy duplexu oparte są na rozdzielaniu sygnałów nadawczych i odbiorczych w dziedzinie częstotliwości lub czasu, koncepcja FD umożliwia nadajnikowi-odbiornikowi bezprzewodowego transmitowanie i odbieranie jednocześnie z tą samą częstotliwością nośną.

Rysunek 8.1



ilustruje podstawową różnicę między FD a konwencjonalnymi transmisjami half-duplex. Koncepcja FD nakłada wysokie wymagania na implementację urządzenia nadawczo-odbiorczego ze względu na równoczesną transmisję i odbiór w tym samym paśmie częstotliwości, ale może przynieść znaczne ulepszenia jako powrót. Do niedawna uważano, że wykorzystanie transmisji FD w bezprzewodowych systemach komunikacyjnych jest niewykonalne ze względu na zjawisko samo-interferencji, które występuje jako część transmitowanego wycieku sygnału lub łączy się z odbiornikiem umieszczonym w pobliżu. Samointerferencja pogarsza wydajność odbiornika, chyba że można go znacznie ograniczyć. W ostatnich latach komunikacja FD przyciągnęła znaczną uwagę, ponieważ wiele grup w przemyśle i środowiskach akademickich zaproponowało różne projekty budowy transceiverów FD. Proponowane rozwiązania obejmują różne projekty anten, projekty fal radiowych (RF) i analogowych obwodów pasma podstawowego, cyfrowe algorytmy pasma podstawowego i ich kombinacje w celu uzyskania wysokiej zdolności do anulowania interferencji z różnymi typami urządzeń radiowych. Opublikowane wyniki kilku zespołów badawczych wskazują, że technologia bezprzewodowa FD zbliża się do wymagań praktycznych systemów. Włączenie urządzeń bezprzewodowych do pracy w trybie FD oferuje możliwość znacznej poprawy działania systemów bezprzewodowych. FD może podwoić pojemność łącza dwukierunkowego w porównaniu do transmisji półdupleksowej. Może umożliwić większą elastyczność w wykorzystaniu widma. Te same zasoby częstotliwości mogą być używane do transmisji jedno- lub dwukierunkowej. FD może dalej uzupełniać dotychczasowe półdupleksowe systemy oparte na TDD lub FDD. Poza wydajnością spektralną i warstwą fizyczną, koncepcje FD mogą być korzystnie wykorzystane w wyższych warstwach, takich jak warstwa o średnim dostępie. Transmisja FD może zmniejszyć opóźnienia interfejsu powietrznego dzięki jednoczesnemu odbiorze informacji zwrotnej (kanały sterujące, sygnalizacja związana z protokołem korekcji błędów itp.) I ułatwiają ulepszone

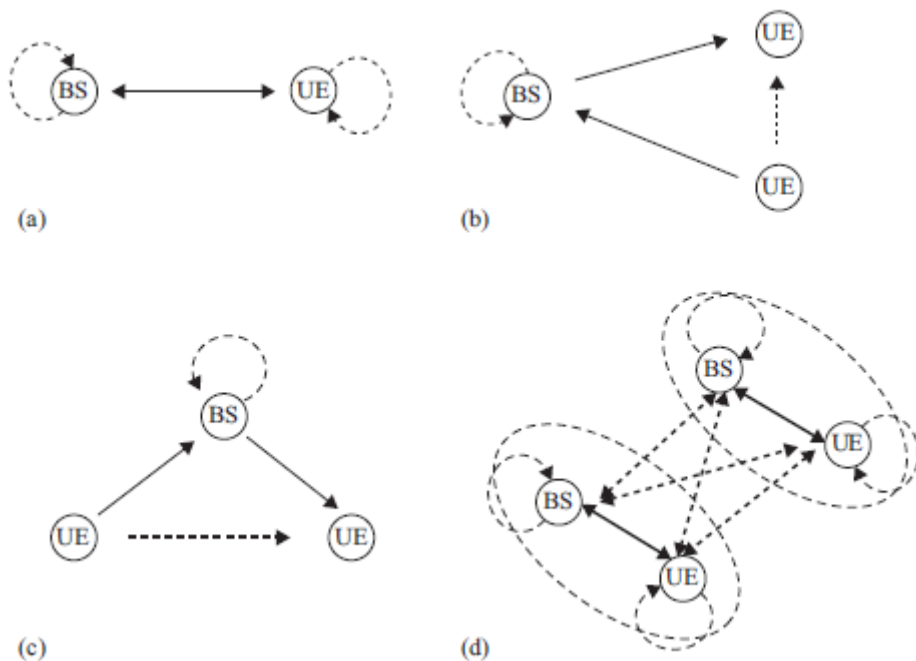
mechanizmy wykrywania / unikania kolizji w sieciach opartych na rywalizacji. Zaproponowano również zastosowanie transceiverów FD w celu poprawienia tajności warstwy fizycznej w komunikacji bezprzewodowej. Technologia bezprzewodowa FD w paśmie ma długą historię, a koncepcja została już wdrożona już od lat 40. XX wieku w projektowaniu systemów radarowych z ciągłą falą (CW). Systemy bezprzewodowej komunikacji naziemnej, takie jak komórkowe urządzenia mobilne i Wi-Fi, jeszcze nie wykorzystwały w dużej mierze transmisji FD. Pierwszym znanym zastosowaniem FD do komunikacji bezprzewodowej było przekazywanie, w którym zastosowano wzmacniacze wewnątrz pasma w celu zwiększenia zasięgu przez odbiór, wzmacnianie i ponowne przesyłanie sygnału bezprzewodowego w tym samym paśmie częstotliwości. Jednak zainteresowanie zastosowaniem FD do przyszłej ewolucji systemów bezprzewodowych znacznie wzrosło w ciągu ostatnich kilku lat i obecnie jest uważane za potencjalny komponent technologiczny dla systemów radiowych 5G. Proponowane przypadki użycia proponowane dla FD w literaturze obejmują przekazywanie, samo-backhauling, mikrofalowe łącza dosyłowe, sieci mesh, kognitywne radio, punkty dostępowe / stacje bazowe i urządzenia w Wi-Fi i systemach mobilnych. Systemy krótkiego zasięgu (np. femto lub piko komórki w mobilnych systemach bezprzewodowych i systemach Wi-Fi) są szczególnie obiecującymi środowiskami operacyjnymi dla FD, ponieważ wymagania dotyczące samo-zakłóceń są znacznie bardziej zrelaksowane w porównaniu do systemów wielkogabarytowych. Ułatwia to wdrażanie technologii FD w kompaktowych atrakcyjnych pod względem komercyjnym urządzeniach bezprzewodowych. Tu omówiono transceiver FD i rozwiązania systemowe dla sieci bezprzewodowych 5G. Koncentruje się na rozwiązaniach transceivera FD dla urządzeń bezprzewodowych o zwartej budowie, takich jak urządzenia podręczne i małe punkty dostępowe / stacje bazowe (np. Femto-komórka). Koncentrujemy się na rozwiązaniach transmisji FD w małych systemach radiokomunikacyjnych. W 8.2 przedstawiono wyzwania związane z wdrażaniem transceivera FD i transmisją FD w małych systemach radiokomunikacyjnych. Omówiono opcje architektury transceivera FD i rozwiązania transceiverów dla urządzeń kompaktowych, w tym przykłady wdrożonych prototypów sprzętowych opracowanych w projekcie badawczym DUPLO UE 7PR w 8.3. W 8.4 omówiono aspekty systemu FD i potencjalne zyski w zakresie wydajności wykorzystania transmisji FD w małych systemach radiowych. Ostatecznie wnioski i propozycje dalszych prac znajdują się w sekcji 8.5.

## **8.2 Transmisja bezprzewodowa FD w scenariuszach 5G i wyzwania**

### **8.2.1 Scenariusze wdrażania pełnego duplexu**

Zagęszczenie wdrożeń sieci jest trendem w 4G, a rozwój kontynuowany jest przy 5G. Małe komórki poprawiają przestrzenne ponowne wykorzystanie zasobów widmowych i ułatwiają niskie moce transmisji nawet przy wysokich przepływnościach. Systemy małej powierzchni są atrakcyjnymi obszarami zastosowań dla transmisji FD, ponieważ wymagania dotyczące redukcji zakłóceń własnych są znacznie bardziej rozluźnione w porównaniu do makrokomórek. Ułatwia to wdrażanie technologii FD w kompaktowych atrakcyjnych handlowo urządzeniach bezprzewodowych. Rysunek 8.2

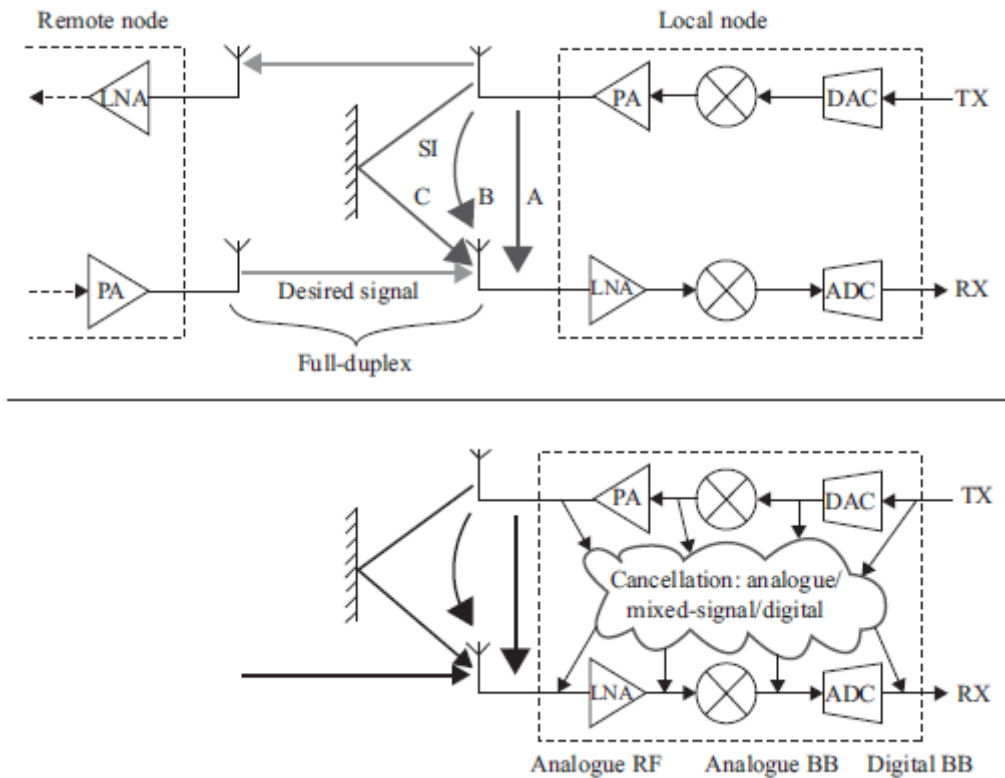




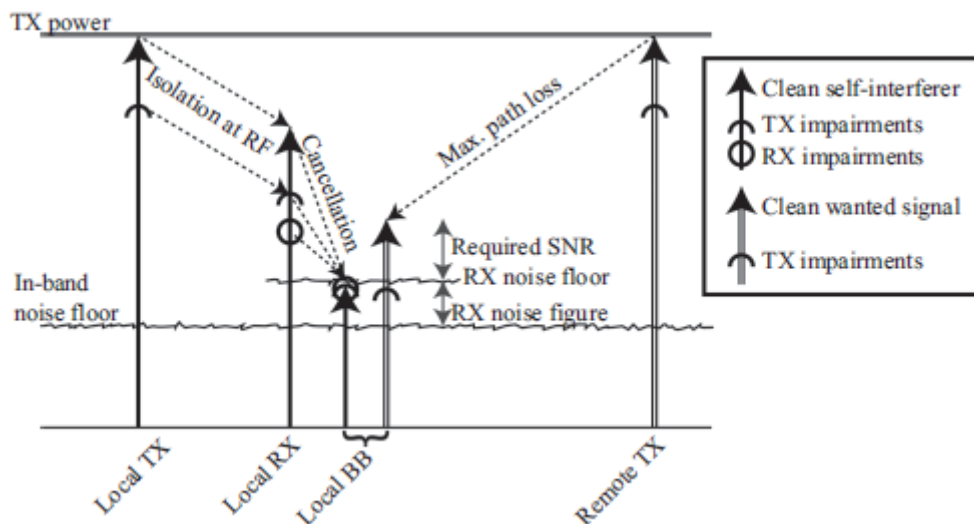
pokazuje podstawowe konfiguracje sieci związane z potencjalnymi przypadkami użycia FD i ilustruje wyzwania związane z implementacją FD. Pierwsze zestawienie na rysunku 8.2 (a) to połączenie typu punkt-punkt między dwoma węzłami, typami stacji bazowej (BS) lub sprzętu użytkownika (UE) lub obydwo. Typowym przykładem zastosowania może być np. łącze typu backhaul lub pojedyncze połączenie BS-UE w izolowanej małej komórce. Również połączenie Wi-Fi w jednej komórce należy do tej kategorii. W tym zestawie szczególnym wyzwaniem związanym z transmisją FD jest samo-interferencja (SI) w każdym węźle. Zasadniczo poziom SI powinien zostać zredukowany do poziomu szumów, aby nie obniżyć pożądanej jakości odbioru sygnału. Idealnie, w tym przypadku, transmisja FD może podwoić pojemność łącza przez połączenie półdupleksowe. Kolejny potencjalny przypadek użycia dla transmisji FD pokazano na rysunku 8.2 (b), gdzie ten sam zasób widmowy jest jednocześnie wykorzystywany dla dwóch połączeń; tj. BS transmituje sygnał do jednego UE podczas odbierania innego sygnału z innego UE. W tym przypadku tylko stacja BS musi być urządzeniem obsługującym FD, podczas gdy urządzenia użytkownika UE mogą być urządzeniami typu half-duplex. Dodatkowym wyzwaniem jest interferencja między węzłami (między wysyłającym i odbierającym UE). W scenariuszu przekazywania pokazanym na rysunku 8.2 (c) węzeł przekaźnikowy musi być urządzeniem obsługującym FD. Bezpośrednie połączenie (UE z UE) może być uznawane za zakłócenie lub pożądany kanał sygnału w zależności od schematu przekazywania. W przypadku wielu komórek operacja FD tworzy ścieżki interferencyjne między wszystkimi aktywnymi węzłami, jak pokazano na rysunku 8.2 (d). Tak więc, na poziomie sieci, wyzwanie wdrożenia systemu FD polega nie tylko na tym, jak zrealizować kompaktowe urządzenie FD z wysoką zdolnością do samoczynnego zakłócenia interferencji (SIC), ale także jak walczyć z dodatkową interferencją między węzłami spowodowaną transmisją odbywającą się jednocześnie z recepcją. Co więcej, wzmocnienie poziomu sieci transmisji FD zależy od wielu aspektów, takich jak asymetria ruchu, architektura sieci, warunki propagacji itp. Transmisja FD w różnych topologiach sieci omówiono bardziej szczegółowo w sekcji 8.4.

### 8.2.2. Budżet łącza pełnego duplexu i wyzwania nadajnika/odbiornika

Głównym wyzwaniem dla warstwy fizycznej w sieci bezprzewodowej FD są zakłócenia własne: ponieważ transmisja i odbiór występują jednocześnie na tej samej częstotliwości, przesyłany sygnał z łatwością wycieka na ścieżkę odbiornika, tworząc bardzo silną interferencję wewnątrz pasma. Rysunek 8.3



pokazuje połączenie FD z lokalnym i zdalnym węzłem nadawczo-odbiorczym. Każdy węzeł ma jedną antenę dla nadajnika i jedną dla odbiornika. W tym łańcuchu trzy źródła SI ograniczają prawidłowy odbiór sygnału przychodzącego ze zdalnego węzła przez lokalny RX. Po pierwsze, wyciek może nastąpić na chipie lub na pokładzie (typ A). Taka bezpośrednia rozmowa kwalifikacyjna prawdopodobnie nastąpi przy gęstej integracji. Po drugie, może wystąpić wyciek liniowy lub rozlewanie między dwiema antenami (typ B). Ten SI jest ogólnie zmniejszone poprzez zastosowanie struktur anteny o określonych wzorach promieniowania lub polaryzacjach. Na koniec, sygnały TX mogą być odbijane na pobliskich obiektach z powrotem do anteny odbiorczej (typ C). Takie odbicia wielościeżkowe zwykle skutkują SI zależnym od częstotliwości. Sygnał samozszyszczenia sprzężony ze ścieżką odbiornika składa się z wielu komponentów: nadawany sygnał jest uszkodzony z powodu zakłóceń nadajnika, takich jak nieliniowość, szum fazowy i kwantowy. Niektóre z tych produktów ubocznych są hałaśliwe, inne są deterministyczne. Ten przesyłany sygnał, w tym jego produkty uboczne, jest sprzężony z odbiornikiem różnymi ścieżkami, jak pokazano na rysunku 8.3. Aby uzyskać bardzo wysokie całkowite odrzucanie zakłóceń własnych, anulowanie samo-przeszkadzania musi być realizowane konstruktywnie w wielu etapach w całym łańcuchu nadajnika-odbiornika FD (izolacja w części anteny i dalsze anulowanie w RF i przetwarzanie pasma podstawowego), jak również zilustrowano, na rysunku 8.3. Ponadto, techniki te powinny być samodopasowujące się w celu utrzymania wymagań w różnych warunkach i środowiskach propagacji. Rysunek 8.4



pokazuje, w jaki sposób lokalne i zdalne sygnały nadajnika są przesyłane do lokalnego pasma podstawowego odbiornika, gdzie oba węzły działają z jednakową średnią mocą nadawania. Sygnał lokalnie nadawany (Local TX) składa się z czystego sygnału i jego produktów ubocznych z powodu zakłóceń nadajnika (półokręgu). Izolacja anteny będzie tłumić samo-interferencje połączone z odbiornikiem (Local RX), wraz z uszkodzeniami nadajnika. Dodatkowe produkty uboczne będą powstawać na tym dużym sygnale z powodu upośledzenia odbiornika (kółko). Techniki anulowania są wymagane w celu dalszego zmniejszenia samo-interferencji i jej produktów ubocznych w kierunku cyfrowego pasma podstawowego odbiornika (lokalny BB). Jeśli wszystkie elementy samo-interferencji zostaną stłumione, aby znaleźć się poniżej poziomu szumu odbiornika, czułość odbiornika i tolerowana utrata ścieżki dla łącza jest równa jego odpowiednikowi półduplexu, a stosunek sygnału do szumu (SNR) jest utrzymywany w celu odbierania sygnału ze zdalnego węzła (Remote TX). Aby oszacować budżet łącza, wymagane odrzucenie zakłóceń samoobsługowych jest obliczane dla typowego połączenia bezprzewodowego o małej odległości z pasmem kanału 10 MHz i luźnym szumem odbiornika o wartości 14 dB. Następnie poziom szumu odbiornika wewnątrz pasma jest równy:

$$-174 \text{ dBm/Hz} + 10^* \log(10\text{MHz}) + 14 \text{ dB} = -90 \text{ dBm}$$

Przy realistycznej średniej sile transmisji wynoszącej 10 dBm wymagane odrzucenie wynosi zatem co najmniej 100 dB, aby zapobiec utracie czułości w porównaniu z połączeniem półduplexowym. W praktycznych rozwiązaniach antenowych [3] efektywna izolacja jest ograniczona do około 40 dB z powodu odbić od otoczenia. Dlatego wymagane jest dodatkowe 60 dB anulowania, aby osiągnąć docelowe odrzucenie 100 dB i tłumić samo-interferencję poniżej poziomu szumu odbiornika. W przypadku nadajnika-odbiornika FD z pojedynczą portem wymagane jest całkowite odrzucenie zakłóceń samoczynnych z RF i obwodami anulowania pasma podstawowego.

### 8.2.3 Wyzwania związane z projektowaniem protokołu MAC

Aby w pełni wykorzystać potencjał transmisji FD w sieciach bezprzewodowych, projektowanie protokołów kontroli dostępu średniozaawansowanego (MAC) i wyższej warstwy jest również ważne, jak w przypadku warstwy fizycznej. Transmisja FD może znacząco zwiększyć ogólną przepustowość sieci bezprzewodowej, wykraczając poza zwykłe podwojenie sprawności widmowej łącza punkt-punkt. Wynika to z faktu, że wykorzystanie węzłów FD w sieci umożliwia wykrywanie kolizji w czasie rzeczywistym. Gdy system bezprzewodowy działa w trybie halfduplex, nie może jednocześnie wysyłać i wykrywać kolizji. Stąd, gdy zdarza się kolizja, cały pakiet zostaje utracony. W związku z tym powszechnie używany jest nośnik-sensor, wielokrotny dostęp z unikaniem kolizji (CSMA / CA). W tym

przypadku nadawca transmituje tylko wtedy, gdy wyczuje bezczynny kanał. W nadajniku-odbiorniku z funkcją FD to ograniczenie jest usuwane, co daje możliwość projektowania algorytmów MAC za pomocą wykrywania kolizji w czasie rzeczywistym. Zmniejsza to czas wykrywania kolizji, a tym samym zwiększa przepustowość. Co więcej, transmisja FD pomaga rozwiązać niektóre ważne problemy w bezprzewodowych sieciach ad hoc, takie jak problem ukrytego węzła, duże opóźnienie między końcami i utrata przepustowości z powodu przeciążenia. W sieciach opartych na infrastrukturze, takich jak sieci komórkowe, planowanie (np. Alokacja zasobów i energii) jest ważne dla optymalizacji wydajności. Równoczesna transmisja w górę i w dół w trybie FD powoduje dodatkowe zakłócenia wewnątrzkomórkowe. Zakłócenia interkomu w gęstych sieciach mogą być również poważniejsze. Dlatego, aby osiągnąć potencjalny zysk trybu FD, niezbędny jest inteligentny program szeregujący do parowania użytkowników łącza w dół i łącza zwrotnego z odpowiednimi mocami transmisji. Ze względu na kombinatoryczny charakter parowania wielu użytkowników łącza w górę i łącza w dół na każdej podnośnej (w systemie opartym na OFDM), a także na trudności optymalnej alokacji mocy do każdej podnośnej, alokacja zasobów w systemach FD jest bardzo trudna. Wymagana jest również estymacja interferencyjnych kanałów międzywęzłowych. Kolejnym wyzwaniem przy projektowaniu i wdrażaniu wydajnego FD MAC jest rozwiązanie problemu asymetrycznego przepływu danych w sieci o symetrycznych przepustowościach łącza. Ruch w komunikacji multimedialnej, takich jak dane i wideo, jest generalnie asymetryczny w łączu w górę i w dół. FD MAC musi zaproponować rozwiązania, które pozwolą korzystać z operacji FD z asymetrycznym ruchem danych.

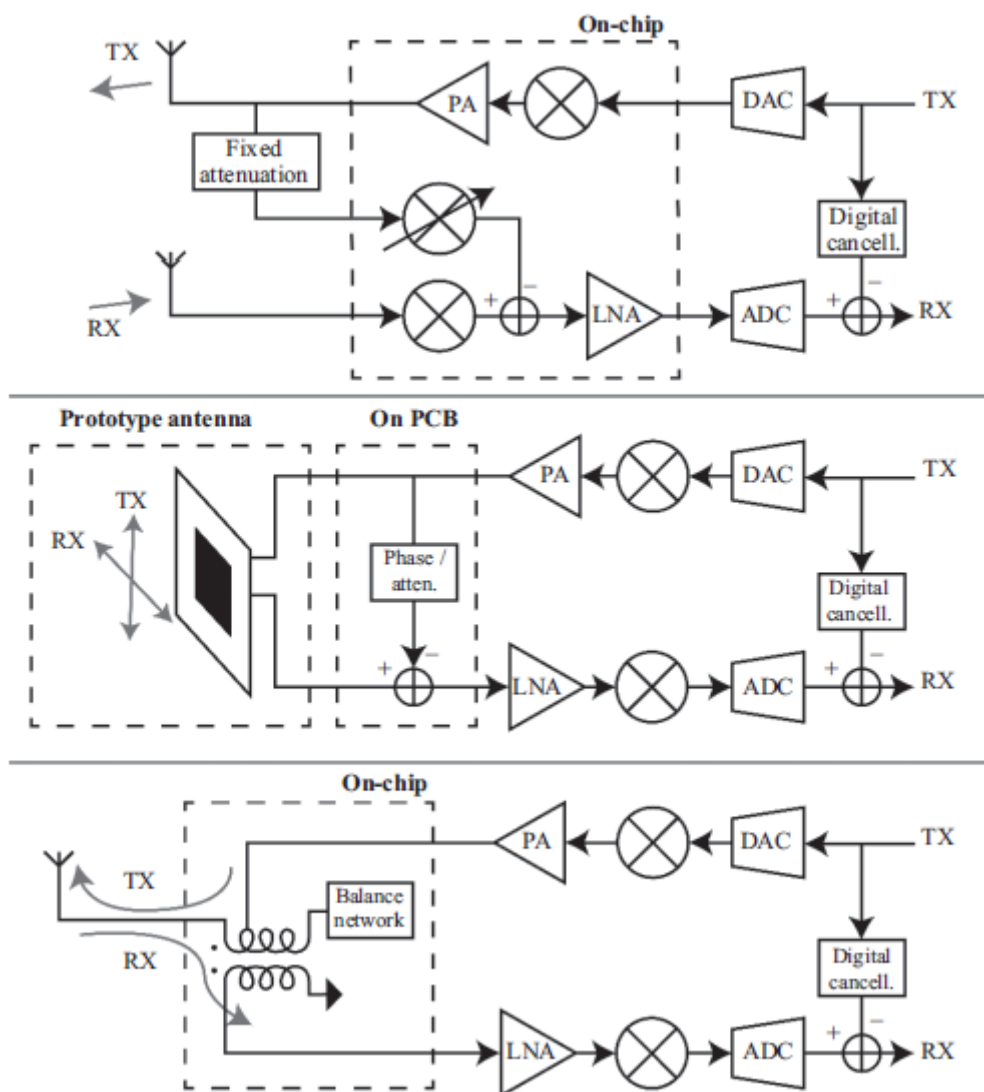
- Ponadto, aby uwzględnić technologię FD w przyszłych standardach komórkowych, a także odkrywać zastosowania tej technologii w obecnej infrastrukturze bezprzewodowej, należy odpowiednio zaadresować i odpowiednio zarządzać współistnieniem węzłów FD i półduplexu. W tym celu protokół MAC musi zachować sprawiedliwość wśród węzłów przy przydzielaniu możliwości dostępu do medium.

### **8.3 Nadajnik/odbiornik z pełnym dupleksem**

#### **8.3.1 Ogólne**

W ostatnich latach przeprowadzono coraz więcej badań w celu rozwiązania problemu z samo-interferencją w transceiverach FD; badano różne techniki, aby zapobiec i wyeliminować wyciek samoistny z TX do RX. Aby zapobiec bezpośredniemu sprzężeniu w antenie FD, zaproponowano struktury z wieloma antenami, aby wykorzystać odstęp i położenie elementów anteny, aby utworzyć izolację w strukturze anteny. Takie struktury anteny mogą osiągnąć izolację do 40 dB, ale zapobiegać gęstej integracji ze względu na wymagane fizyczne odległości między elementami anteny. Inne techniki głównie anulują samointerferencję w ścieżce odbiornika w oparciu o znajomość transmitowanego sygnału. To anulowanie może być wykonywane w różnych lokalizacjach w urządzeniu nadawczo-odbiorczym, mianowicie w cyfrowym paśmie podstawowym, analogowym paśmie podstawowym i / lub analogowym kanale RF. Eliminacja cyfrowa wykorzystuje zalety przetwarzania cyfrowego, ale nie kompensuje szumów i zniekształceń generowanych w transceiverze analogowym i nie może zapobiec nasyceniu odbiornika z powodu silnych sygnałów zakłóceń samoczynnych. Analogowe i RF anulowanie umożliwiają anulowanie szumów i zniekształceń wywołanych przez nadajnik, a także złagodzenie wymagań liniowości odbiornika, ponieważ samo-interferencja jest redukowana na początku w łańcuchu odbiorczym. Zaproponowano kilka architektur anulowania sygnałów mieszanych tak, że cyfrowy sygnał nadawczy jest używany jako odniesienie, aby skompensować samo-interferencję na wejściu RF odbiornika. Ta architektura wymaga dedykowanego konwertera, który w praktyce wprowadza własny szum i zniekształcenia oraz ogranicza efektywność anulowania. Aby uzyskać całkowite odrzucenie zakłóceń własnych o wartości przekraczającej 100 dB, wymagana jest kombinacja technik samowykształcenia i eliminacji zakłóceń. Najnowszy projekt nadajnika-odbiornika został opisany w oferowaniu atrakcyjnej wydajności FD z izolacją z samo-interferencją przekraczającą 100 dB

na szerokości pasma 80 MHz. Ta konstrukcja wykorzystuje pojedynczą antenę portową w połączeniu z cyrkulatorem częstotliwości radiowej i stosuje zarówno anulowanie cyfrowe, jak i analogowe. Ta architektura jest wykorzystywana w komercyjnych produktach Kumu Networks<sup>TM</sup> do kierowania na bezprzewodowy backhauling i przekazywanie w ustalonym miejscu. Konstrukcja transceivera nie jest jednak odpowiednia dla radiotelefonów krytycznych o współczynniku kształtu stosowanych w urządzeniach mobilnych i punktach dostępu do małych komórek. Analogowa struktura anulująca implementuje sieć fizycznych linii opóźniających, które nie mogą być znacząco skalowane dla danej częstotliwości roboczej. Dlatego też inne przenośne urządzenia nadawczo-odbiorcze są wymagane w przenośnych aplikacjach 5G. Prezentujemy różne potencjalne rozwiązania transceivera, które wspierają integrację w kompaktowych i przenośnych radiotelefonach. Sekcja 8.3.2 omawia kilka rozwiązań RF / analogowych i strukturę anteny wspierających pracę FD, a rozdział 8.3.3 przedstawia rozwiązania z zakresu cyfrowej eliminacji, które uzupełniają anulowanie układu samohamowności. W przypadku typowego scenariusza krótkiego lub wewnętrznego całkowite odrzucenie zakłóceń własnych powinno wynosić około 100 dB w paśmie 10 MHz (sekcja 8.2.2). Przedstawione rozwiązania eliminujące analog / RF eliminują bezpośredni przeciek (przesłuchy w nadajniku-odbiorniku i bezpośrednie sprzężenie w antenie), podczas gdy eliminacja cyfrowa obejmuje również odbicia wielościeżkowe. Odbite sygnały ulegają znacznemu opóźnieniu i powodują samo-interferencje zależne od częstotliwości. Jeśli chodzi o docelowy scenariusz krótkiego lub wewnętrznego, odbicie wielościeżkowe jest o około 40-65 dB mniejsze od bezpośredniego przecieku, to samo anulowanie zakłóceń powinno wynosić około 50 dB. Sekcja 8.3.4 integruje rozwiązania analogowe / RF i cyfrowe, a proof-of-concept potwierdza działanie systemu FD. Wreszcie, Rozdział 8.3.5 omawia zastosowanie transmisji FD do systemów wielu wejść wielokrotnego wyjścia (MIMO). Przykłady realizacji w tym rozdziale oparte są na prototypach sprzętowych opracowanych w projekcie badawczym DUPLO UE 7PR. 8.3.2 Rozwiązania RF / analogowe i antenowe dla urządzeń kompaktowych Urządzenia o współczynniku kształtu Koncepcja FD jest sprzeczna z gęsto zintegrowanymi radiotelefonami, ponieważ integracja promuje wyciek sygnału w obwodach nadajnika-odbiornika i między elementami anteny. W tej sekcji omówiono zakłócające architektury analogowe, które pozwalają zminimalizować wyciek z samo-interferencji, przy jednoczesnym utrzymaniu gęstej (ko-) integracji w kompaktowych radiotelefonach dla aplikacji mobilnych i dostępu małego obszaru. Trzy architektury przedstawione na rysunku 8.5

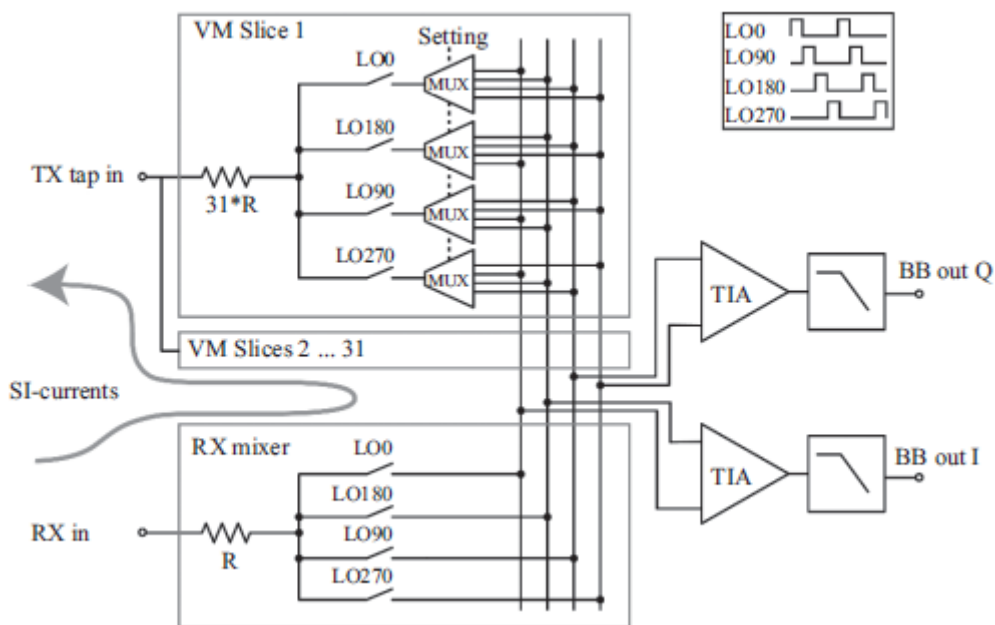


są uważane za minimalizujące bezpośredni wyciek z nadajnika do jego własnego odbiornika. Pierwsza architektura to kompletny analogowy front-end, który implementuje nowy mechanizm anulowania interferencji, w tym dynamiczną konwersję w dół. Część analogowa oznaczona linią przerywaną została prototypowana w technologii CMOS. Wspiera współ-integrację ze starszymi systemami, umożliwia niską produkcję i działa przy niskiej mocy. Druga architektura obejmuje pojedynczą strukturę anteny z natywną izolacją pomiędzy jej portami i aktywną siecią anulowania w portach antenowych. Ta struktura anteny jest zwarta i, w połączeniu z siecią eliminującą, oferuje solidną izolację z samo-interferencją w zmieniających się warunkach działania. Trzecia architektura implementuje innowacyjny duplexer, który umożliwia wykorzystanie starszej jednoportowej anteny do komunikacji FD. Ten duplexer jest prototypowany w technologii SoI CMOS, jak pokazano linią przerywaną. Te trzy architektury mają na celu integrację w kompaktowych radiotelefonach i oferują określone korzyści i cechy, które można wykorzystać w różnych aplikacjach. Dlatego nie są one przeznaczone do połączenia.

### 8.3.2.1 Front-end z automatyczną eliminacją zakłóceń

W kompaktowych urządzeniach radiowych FD przestrzeń kosmiczna nie pozwala na zaawansowane rozwiązania antenowe zapewnić wysoką izolację SI. Ponadto, zmienne środowisko bliskiego pola (np. Osoba trzymająca urządzenie) ciągle zmienia wielkość i fazę samo-przenikającej się, wymagającej

adaptacyjnego rozwiązania. Eksperymenty wykorzystujące dwie skrzyżowane anteny dipolowe WLAN jako prostej anteny FD wykazały, że izolacja 25 dB jest rozsądnym założeniem. Pozostawia to miejsce na poprawę 20-40 dB w domenie analogowej, zanim samo wąskie gardło stanie się samo-zakłóceniem spowodowanym przez odbicie na pobliskich obiektach. Biorąc pod uwagę ten scenariusz, przedstawiono innowacyjną i nowatorską architekturę front-end. Ten projekt zapewnia anulowanie przez odjęcie samo-interferencji na analogowym paśmie podstawowym odbiornika w oparciu o przesyłany sygnał RF. Ta topologia nadaje się głównie do aktywnej stymulacji izolacji umiarkowanego rozwiązania anteny, np. Gdy antena zapewnia umiarkowaną izolację 15-20 dB, do 30-45 dB, zanim odbicia staną się dominującą interferencją. Odbicia te są następnie anulowane w cyfrowym pasie podstawowym przez drugie anulowanie ścieżka, która uwzględnia opóźnienie w domenie cyfrowej. Mocną stroną tej architektury jest to, że tłumienie, przesunięcie fazowe i konwersja w dół mogą być łączone w jeden składnik funkcjonalny: mikser wektora modulatora. Jednakże, ponieważ punkt anulowania znajduje się w analogowym przetwarzaniu pasma podstawowego, silne samo-przeszkadzanie jest w pełni obecne na wejściu LNA. Aby zapobiec utopieniu słabego leżącego pod spodem pożądanego sygnału, wymagana jest wysoka liniowość w pierwszych etapach odbierania. Można to osiągnąć, stosując pierwszą architekturę mixera zarówno dla głównego miksera, jak i modulatora wektorowego. Symulacje obwodów potwierdziły wyjątkowy potencjał tej architektury i wskazały, że wady można rozwiązać za pomocą odpowiednich topologii obwodów. Rysunek 8.6

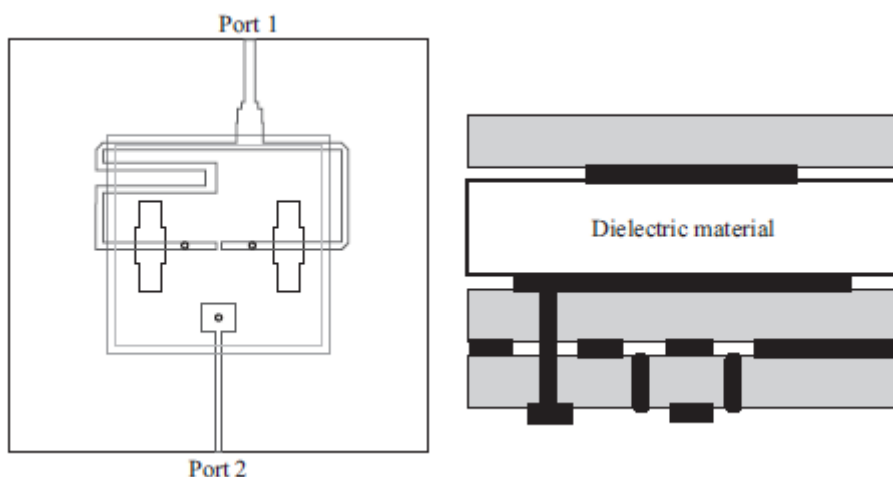


pokazuje uproszczony schemat odbiornika z włączeniem miksera miksującego w dół. Zasadniczo, zarówno VM, jak i główny mikser RX są cztero-fazowo przełączanymi mikserami rezystorowymi, które utrzymują wysoką liniowość przy dużych mocach samo-interferencji. Modulator wektorowy jest plastrowaną wersją głównego miksera RX, zawierający multipleksery w każdym wycinku, które kierują czterofazowy prąd wyjściowy do odpowiednich faz pasma podstawowego. Ilość wycinków VM określa liczbę punktów konstelacji fazy / amplitudy, które można pokryć, a tym samym ilość anulowania, którą można osiągnąć dzięki efektom kwantyzacji. W tym celu stosuje się 31 plastrów umożliwiających 27,1 dB anulowanie SI. Prototypowy front-end pokrywa obszar 2 mm<sup>2</sup> w CMOS 65 nm i został sprawdzony na podstawie pomiarów. Wydajność z automatycznym usuwaniem zakłóceń mierzono dwiema skrzyżowanymi antenami dipolowymi WLAN (izolacja mierzona ~ 30 dB) i wielotonowym sygnałem

podobnym do 802.11a przy szerokości pasma 16,25 MHz, co daje całkowitą wartość anulowania ponad 45 dB. Badano także odporność na szumy fazowe w trybie FD przy użyciu tego samego syntezyatora częstotliwości konwersji w górę i w dół. Taka odporność jest wymagana, ponieważ eksperymenty pokazują, że różne syntezyatory generujące nieskorelowany szum fazowy -38 dBc w pasmie będącym przedmiotem zainteresowania indukują łączny poziom szumów wynoszący -35 dBc, ograniczając anulowanie. Pomiary te wykazały bardzo wysoką odporność na szumy fazowe w FD.

### 8.3.2.2 Antena podwójnie spolaryzowana i usuwanie aktywnych analogów

Konstrukcja struktury anteny FD jest ukierunkowana na wysoką wewnętrzną izolację i niską krzyżową polaryzację pomiędzy portami anteny nadawczej i odbiorczej. Taka izolacja jest tradycyjnie osiągana dzięki elementom wielokanałowym, stosującym osłony radiacyjne i / lub destrukcyjne wzorce promieniowania, aby zapobiec wyciekowi z transmisji do anteny odbiorczej (-ych). W przypadku celowania w strukturę anteny o małym współczynniku kształtu atrakcyjna jest alternatywna architektura zbudowana na pojedynczym elemencie anteny o ortogonalnej polaryzacji. Taka struktura anteny została przedstawiona na rysunku 8.5, gdzie nadajnik i odbiornik są połączone z odpowiednimi płaszczyznami polaryzacji przez dwa porty antenowe. Spełnienie zarówno wymagań izolacji, jak i polaryzacji krzyżowej. Jednocześnie jest to bardzo trudne, szczególnie w przypadku kompaktowych rozwiązań antenowych, w których oba gniazda wzbudzenia są blisko rozmieszczone, a pasożytnicze połączenie między nimi jest wysokie. Te projekty łączą różne mechanizmy wzbudzenia, takie jak sieci sprzężone z gniazdem i wielopunktowe sieci zasilające w strukturę anteny w stosie, aby uzyskać atrakcyjne wartości izolacji. Skumulowana struktura składa się z dwóch kwadratowych łąt i symetrycznej sieci wzbudzenia. Stosowana jest również polaryzacja dwu liniowa, ponieważ polaryzacja liniowa zasadniczo nie jest zależna od częstotliwości pracy (w przeciwieństwie do anteny kołowo spolaryzowanej). Aby umożliwić niskoprofilową, tanią i łatwą integrację, te projekty są wdrażane przy użyciu technologii mikropaskowej. W przypadku anteny mikropaskowej podwójną spolaryzowaną liniowo operację uzyskuje się za pomocą pary kanałów sond, aby odpowiednio wzbudzić dwa prostopadłe tryby podstawowe z pojedynczej wiązki promieniowania. Dyskryminacja między dwiema polaryzacjami mogłaby jednak zostać osłabiona przez niepożądane tryby wyższego rzędu, które mogą powodować polaryzację krzyżową. Aby powstrzymać polaryzację krzyżową w każdym porcie zasilającym podwójnie spolaryzowanej anteny mikropaskowej, poprawiając w ten sposób poziom izolacji pomiędzy portami, struktura anteny zawiera symetryczną sieć wzbudzenia z podwójnym zasilaniem z różnicą fazową wynoszącą  $180^\circ$  i para szczelin sprzęgających kryteriami optymalizacyjnymi były zmniejszenie współczynnika kształtu, zwiększenie izolacji z automatyczną separacją i poprawienie dopasowania impedancji. Uzyskana geometria anteny drugiego prototypu o wartości  $60 \times 60 \times 8$  mm jest zilustrowane na rysunku 8.7





i składa się z dwóch kwadratowych łat i trzech warstw podłoża o tej samej grubości i stałej dielektrycznej. Płaszczyzna pozioma (PORT 1) realizuje podziałkę  $1/2$ , a polaryzacja pionowa (PORT 2) jest wzbudzana przez linię mikropaskową o mocy 50 W i wewnętrzną przepustowość. Do oddzielenia dwóch kwadratowych łat dodano sztywny materiał dielektryczny o stałej dielektrycznej 2,5. Podane parametry są bardzo atrakcyjne dla FD, oferując ponad 50 dB izolacji anteny, 6,5 dB wzmocnienia anteny i 75% wydajności anteny przy szerokości pasma 10 MHz. Te pomiary są również wykonywane przy szerokości pasma 80 MHz, co wskazuje na niewielki spadek wydajności. Pomiary wskazują jednak na znaczną degradację działania anteny FD, gdy przedmioty są umieszczane lub przemieszczane w pobliżu. Aby utrzymać izolację samouszkodzającą w antenie (przed LNA), zaleca się wdrożenie aktywnej sieci anulowania działającej pomiędzy portami antenowymi, jak pokazano na rysunku 8.5. Ta aktywna sieć usuwania RF wykorzystuje tłumioną i przesuniętą w fazie kopię transmitowanego sygnału RF, aby anulować samo-interferencję w ścieżce odbiornika przed LNA. Wyniki przedstawione w [8] ilustrują ulepszone tłumienie samo-interferencji oprócz izolacji anteny, skutkując anulowaniem 62 dB. Prototyp anteny w połączeniu z aktywną siecią usuwania został włączony do prototypu transceivera FD opisanego w sekcji 8.3.4 i został sprawdzony w różnych warunkach działania.

### 8.3.2.3 Duplexer wagi elektrycznej

Rozmiar anteny jest jednym z głównych wąskich gardeł dla kompaktowej integracji FD radia. Technika antenowa opisana w poprzedniej sekcji wykorzystuje pojedynczy element anteny, unikając fizycznej odległości między tradycyjną strukturą multiantenny. Ten pojedynczy element anteny jest jednak specyficzny dla operacji FD i ma dwa porty. Podczas kierowania na gęstszą integrację i kointegrację z dotychczasowe systemy, przy użyciu konwencjonalnej anteny z półki, są bardziej atrakcyjne. Takie anteny jednoportowe są dostępne w bardzo małych rozmiarach (np. Komponenty SMD) i są już zaimplementowane w starszych systemach. Dzieląc tę antenę do następnej operacji pomiędzy transceiverem FD a dotychczasowym systemem, można osiągnąć wysoką wydajność implementacji podczas projektowania wielostandardowego radia. Używanie takiej konwencjonalnej anteny do działania FD oznacza, że nie ma izolacji anteny, i że duplexer jest wymagany, aby zapobiec sprzężeniu sygnałów nadajnika z jego własnym odbiornikiem. Ten duplexer łączy transceiver FD z anteną pojedynczego portu, jak pokazano na rysunku 8.5. Fale powierzchniowo-akustycznej (SAW) - typowe duplexery są powszechnie stosowane w systemach FDD, aby zapobiec przedostawaniu się sygnału nadajnika do odbiornika za pomocą filtrowania. Jednak takich duplexerów nie można stosować w FD, ponieważ nie ma separacji częstotliwości. Inną opcją jest użycie cyrkulatora do kierowania sygnałów nadajnika i odbiornika przez wspólną antenę. Takie pompy cyrkulacyjne zapewniają jedynie umiarkowaną izolację około 20 dB i wymagają dodatkowych pętli anulowania przy częstotliwości radiowej, aby były odpowiednie dla FD. Podczas kierowania na implementację w radiach kompaktowych połączony obieg cyrkulatora i złożony obwód RF są raczej nieatrakcyjne. Zbadano alternatywną topologię duplexu, a mianowicie duplexer bilansu elektrycznego (EBD). Początkowo proponowano koncepcję EBD w celu uzyskania przestrajalnych filtrów duplexera dla FDD, ale jest ona również bardzo atrakcyjna dla FD ze względu na jej potencjał do wydajnej implementacji w technologii CMOS, jej wysoki współczynnik izolacji, aby zapobiec wyciekom w kierunku własnego odbiornika i jego elastyczności strojenia. Zasada działania EBD została przedstawiona na rysunku 8.5; składa się głównie z transformatora hybrydowego i sieci balansowej. Sieć bilansu elektrycznego jest zasadniczo przestrajalną sztuczną impedancją obciążenia, która jest dostosowana do impedancji anteny tak, że przesyłane sygnały nie są propagowane przez transformator do jej własnego odbiornika. Sieć balansowa jest przestrajana w celu obsługi wykorzystania szerokiej gamy typów anten (z różnymi impedancjami) i śledzenia ewentualnych zmian impedancji anteny z powodu, na przykład, dynamicznych warunków działania lub poruszających się obiektów w pobliżu anteny. EBD to obwód RF, który wykonuje samoczynne usuwanie zakłóceń bardzo blisko anteny. Chroni to pierwsze elementy

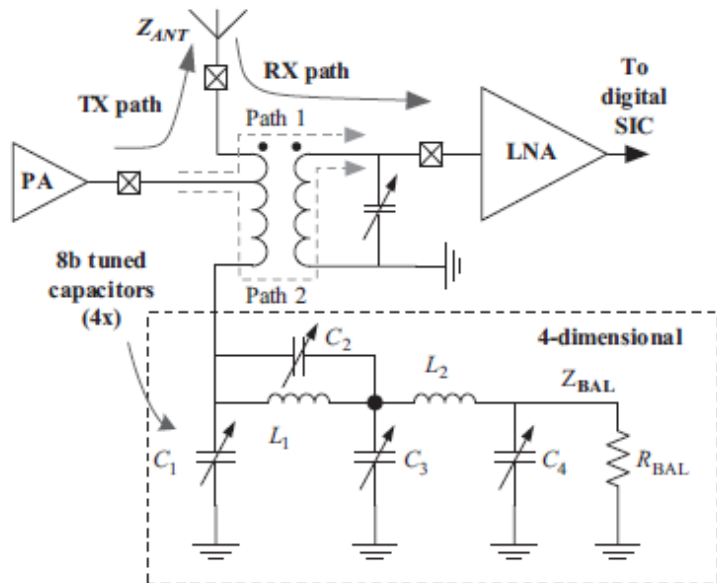
odbiornika przed nasyceniem z powodu samo-interferencji i zniekształceń spowodowanych przez upośledzenia nadajnika (np. Nieliniowość, szum fazowy) w anulowaniu. Ponadto, ze względu na czysto pasywną strukturę EBD, wszelkie produkty szumowe i nieliniowe generowane w nadajniku są również anulowane przez sam obwód. Niedawno wprowadzono topologię EBD i zatwierdzono prototyp dla aplikacji FD. Prezentowany prototyp EBD zawiera ośmiokątny transformator, sieć balansową RC z przestrajalną pojemnością i oporem oraz bufor różnicowy odbiornika w technologii 180 nm UMC CMOS. Układ obejmuje obszar 1,5 x 1,5 mm i jest zawarty w prototypie urządzenia nadawczo-odbiorczego FD opisanym w rozdziale 8.3.4. Opracowano inteligentny algorytm strojenia [28], aby szybko dostroić i śledzić optymalną konfigurację sieci wagi. Pomiar walidacyjne wskazywały na kilka istotnych problemów, o których mowa w:

- Utrata implementacji w ścieżce RF nadajnika i odbiornika o wartości około 6 dB: mimo że strata ta wydaje się być dość wysoka, jest porównywalna ze starszymi systemami FDD.
- Duża zależność impedancji anteny od częstotliwości i warunków działania

(np. przemieszczanie obiektów w pobliżu anteny); Niezbędne jest dynamiczne (ponowne) dostrojenie sieci bilansowej.

- Ograniczona szerokość izolacji: EBD oferuje wyjątkowo wysoki pik izolacji > 90 dB w stanie równowagi, ale jego szczytowy kształt widmowy ogranicza średnia izolacja na szerszej szerokości pasma. Ten szczytowy kształt jest spowodowany przez zmienność impedancji anteny na częstotliwościach (in-band). Ponieważ sieć balansowa nie może być zaprojektowana do śledzenia impedancji anteny nad chwilową przepustowością, konieczne jest nowatorskie podejście do projektowania w celu zwiększenia szerokości pasma operacyjnego.
- Zniekształcenia w sieci bilansowej przenikają do odbiornika. Transformator EBD nie eliminuje zniekształceń generowanych w sieci równoważonej w celu wycieku w kierunku odbiornika. To bezpośrednio ogranicza moc nadawania, aby zapobiec nieliniowości generowanej w sieci równoważonej. Odkrycie to wskazuje na ścisłe wymaganie liniowości w sieci bilansowej, aby wspierać tranzytowe moce większe niż -10 dBm.

Ten prototyp został przeprojektowany w celu zwiększenia wydajności FD. Prototyp ten został zaimplementowany w CMOS o grubości 180 nm (Sol). Alternatywną topologię sieci bilansowej, jak pokazano na rysunku 8.8,

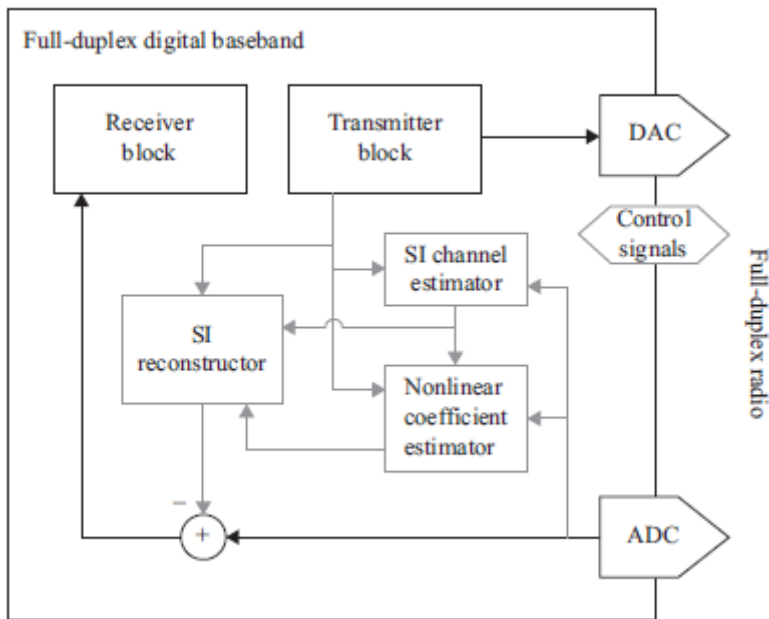


zastosowano w celu osiągnięcia wysokiej liniowości i swobody strojenia przy zachowaniu wydajności implementacji. Czerowymiarowa zdolność dostrajania w sieci równoważającej umożliwia dostrojenie rzeczywistej i wyobrażonej impedancji na częstotliwości i wykupienie przepustowości za pomocą średniego SIC, podczas gdy sieć R / C pierwszego prototypu obsługuje tylko strojenie pojedynczego piku izolacji na całej częstotliwości. Należy zauważyć, że przeciwstawiając się pierwszemu prototypowi, preferowana jest topologia jedno-końcowa po stronie odbiornika transformatora hybrydowego, aby uniknąć SI trybu wspólnego i umożliwić działanie o wysokiej mocy.

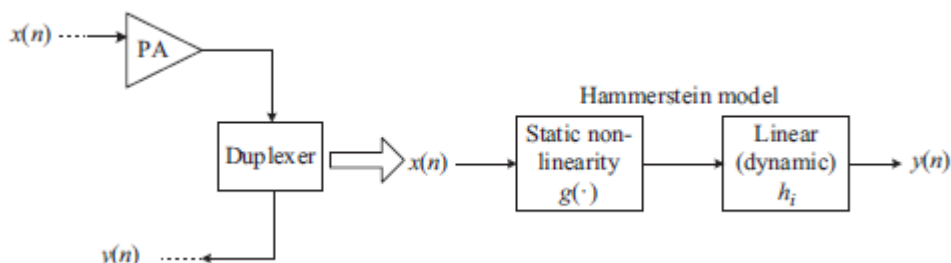
### 8.3.3 Cyfrowe rozwiązania pasma podstawowego dla transceivera pełnego duplexu

Problem anulowania samo-interferencji w domenie cyfrowej przypomina usuwanie echa w modemach cyfrowych linii abonenckich (DSL), przekaźnikach komórkowych i wzmacniaczach cyfrowej transmisji wideo (DVB). Stąd te same zasady stosowane w tłumikach echa modemu DSL mogą być stosowane w systemach radiowych FD w celu złagodzenia samo-interferencji. W transmisji bezprzewodowej FD wymagania są bardziej restrykcyjne i wymagające. W porównaniu z tłumikami echa DSL, szerokie pasmo częstotliwości i nieliniowe efekty spowodowane zaburzeniami RF i analogowymi powodują, że samozakłócenie w kompaktowych radiotelefonach FD jest bardziej złożone i trudniejsze do naśladowania. Ogólnie rzecz biorąc, przekaźniki mogą cieszyć się względnie wysoką izolacją na poziomie anteny, wykorzystując ich większy współczynnik kształtu i możliwość zastosowania anten kierunkowych ze względu na ich specyficzne scenariusze zastosowania. Z drugiej strony kanał samo-interferencji w przenośnych i mobilnych radiotelefonach FD może być zmienny ze względu na dynamiczne środowisko, w którym wykorzystywane są urządzenia, podczas gdy kanał echa w DSL i pętla interferencyjna (kanał samo-zakłóceń) w przekaźnikach a wzmacniacze są mniej lub bardziej statyczne. W odbiorniku FD, anulowanie pre-ADC, w tym izolacja ścieżki odbioru / odbioru (z powodu zastosowania oddzielnych anten do nadawania i odbioru lub przez zastosowanie obwodów izolacyjnych) i anulowanie RF / analogowe SI, usuwają tylko część samoczynnej interferencji. Obecne rozwiązania analogowe oferują do 60 dB możliwości anulowania SI. Ponieważ całkowite zapotrzebowanie na anulowanie SI może być większe niż 100 dB, wymagane są techniki anulowania RF / analogowe i cyfrowe SI. Oznacza to, że pozostały SI musi zostać anulowany przy cyfrowym paśmie podstawowym. Resztkowa moc SI odbierana w paśmie podstawowym może być do 60 dB wyższa niż

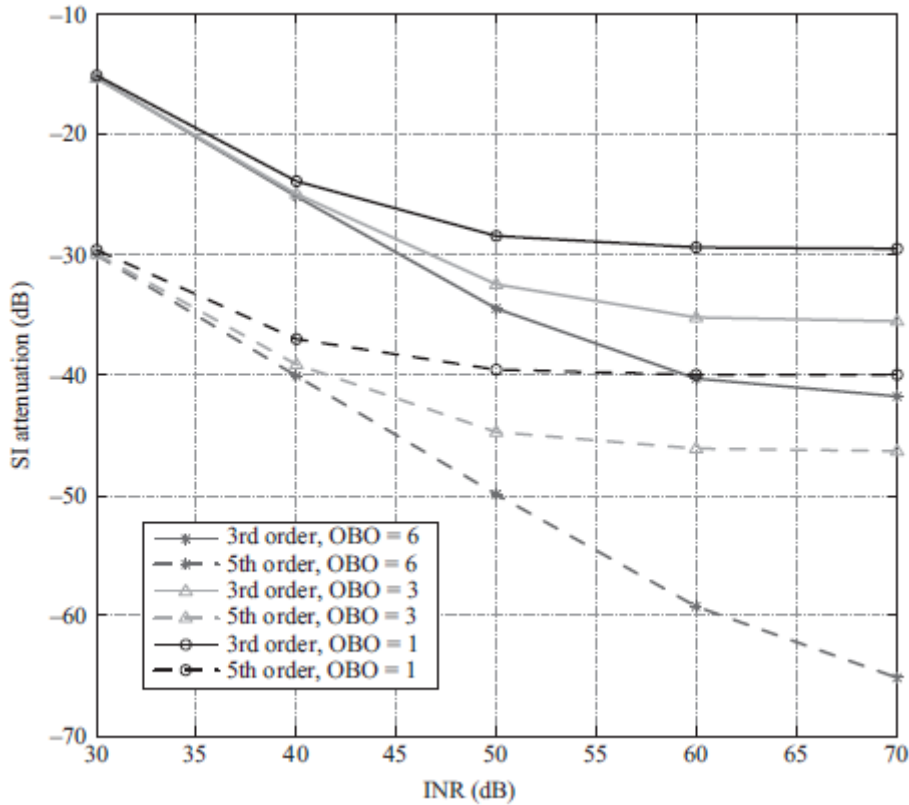
poziom szumu, a sygnał SI zawiera nieliniowe komponenty utworzone na wzmacniaczu mocy, wzmacniaczu lownowym itp. Ogólny schemat blokowy dla cyfrowej rezygnacji jest wyświetlany na Rysunku 8.9.



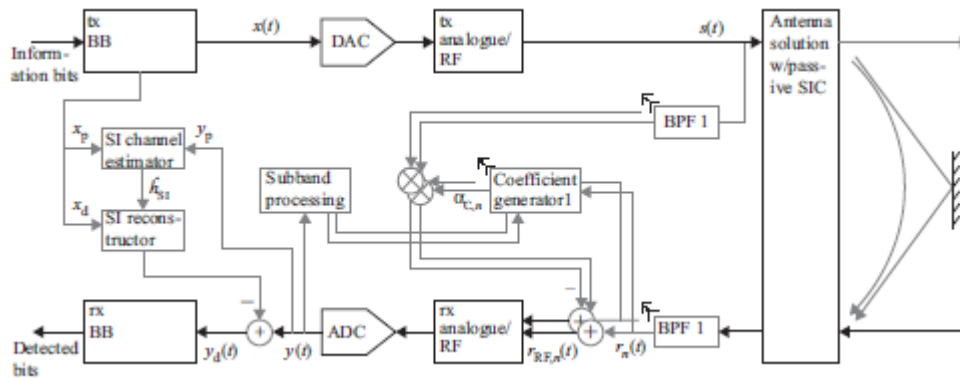
Znane transmitowane dane są wykorzystywane do oszacowania kanału SI, którego dokładność zależy od stosunku samo-zakłóceń do szumu. W rzeczywistym scenariuszu kanał SI zawiera wielość i może dynamicznie się zmieniać. Korzystając z oszacowanych komponentów liniowych i nieliniowych, szacuje się całkowitą interferencję własną od odebranego sygnału. Maksymalne osiągalne anulowanie samo-interferencji zależy od dokładności oszacowania kanału i dokładności modeli nieliniowych. Ponadto skuteczność anulowania zależy od anulowania pre-ADC, ponieważ może to spowodować podniesienie poziomu hałasu. Jedną z głównych wad nadajników wpływających na wydajność anulowania SI jest nieliniowość wzmacniacza mocy. Aby wyeliminować komponenty interferencyjne spowodowane nieliniowością, potrzebny jest model dla nieliniowości. Jednym z powszechnie stosowanych modeli nadajników jest model Hammersteina, będący kaskadą statycznej nieliniowej części, po której następuje filtr liniowy. Jest to powszechnie stosowana metoda modelowania PA z pamięcią i naturalny wybór do modelowania kaskady PA i selektywnego kanału selektywnej interferencji częstotliwości, jak pokazano na rysunku 8.10.



Wydajność samozasilania opartego na modelu Hammersteina została zademonstrowana na podstawie wyników symulacji na rysunku 8.11.



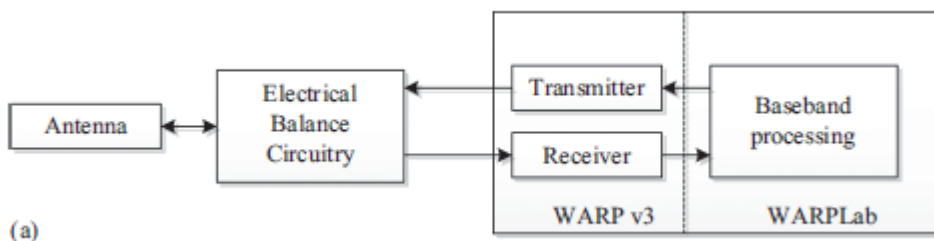
Sygnal jest sygnałem OFDM z 52 podnośnymi. Wzmacniacz mocy (PA) w nadajniku został zmodyfikowany przy użyciu modelu Modified-Saleh I. Parametr OBO na rysunku 8.11 wskazuje wartość back-off wzmacniacza mocy użytej w symulacjach. Back-off definiuje się tu jako różnicę między poziomem nasycenia modelu PA i szczytową mocą sygnału wejściowego PA. Parametr INR to stosunek mocy samo-zakłóceń-tonoise przed anulowaniem pasma podstawowego. Duplexer na Rysunku 8.10 jest elektronicznym obwodem eliminującym zakłócenia samoobsługowe oparte na równowadze elektrycznej. W odbiorniku sygnał używany do anulowania SI jest generowany przy użyciu modelu Hammersteina. Statyczna nieliniowość jest modelowana jako funkcje wielomianowe trzeciego i piątego rzędu, a część liniowa jest filtrem FIR. Współczynniki filtra FIR są szacowane przy użyciu z góry określonego sygnału treningowego. W praktycznym urządzeniu nadawczo-odbiorczym FD, izolacja SI dla sygnału szerokopasmowego nie jest jednolita w interesującym paśmie. Na przykład, układ tłumienia samo-zakłóceń analogowych opartych na równowadze elektrycznej jest dostrajany z określoną częstotliwością (zwykle w środku pasma), a następnie poziom izolacji pogarsza się, gdy idzie się dalej od dostrojonego punktu. Różnica ta może osiągnąć 15 dB lub więcej. Selektywność częstotliwościowa izolacji SI zwiększy złożoność estymacji parametrów modelu Hammersteina. Na przykład w przedstawionych powyżej symulacjach długość FIR wynosiła 52 razy. Jedno rozwiązanie łagodzące skutki selektywnego odbioru częstotliwościowego izolacji ścieżki może być nierównomiernym anulowaniem SI w paśmie, tj. pasmo może być podzielone na kilka podpasm, a następnie anulowanie SI jest przeprowadzane w każdym podpaśmie niezależnym od innych podpasm, jak przedstawiono na rysunku 8.12.



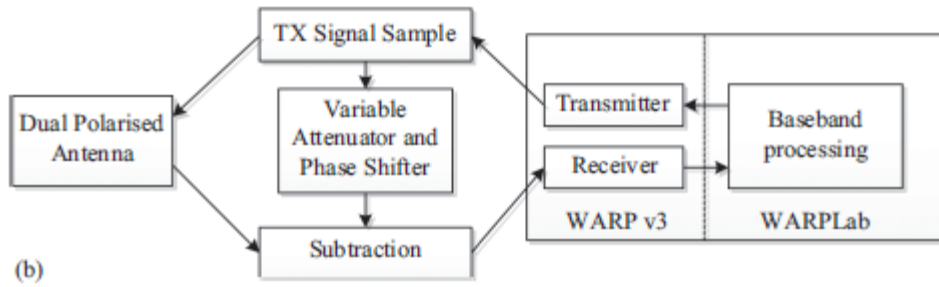
Inne nieidealne właściwości zmniejszające wydajność usuwania SI obejmują nierównowagę IQ, przetworniki cyfrowo-analogowe i analogowo-cyfrowe oraz szum fazowy. Powszechne jest to, że ich wpływ na odbierany sygnał zakłóceń własnych nie może być oszacowany za pomocą prostego estymatora kanału zastosowanego w estymacji kanału między węzłami. Jednak wszystkie zjawiska są dobrze zrozumiałe i zaproponowano wiele technik w literaturze otwartej, aby oszacować i zrekomensować je, patrz np. I odnośniki w nich.

### 8.3.4 Prototypowe projektowanie i testowanie transceivera

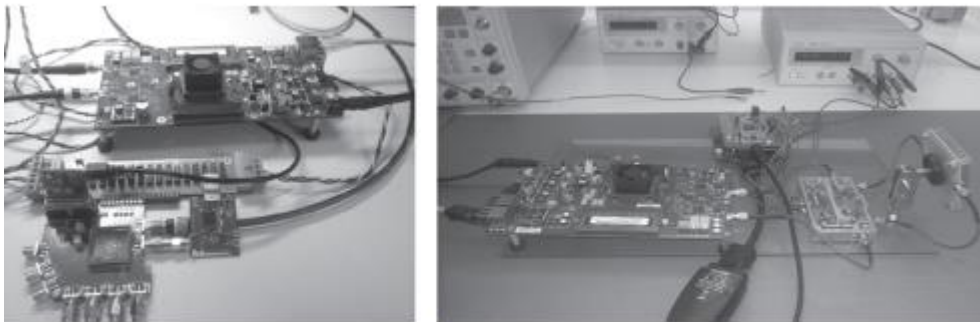
Wdrożenie prototypowych transceiverów umożliwia ocenę FD komunikacja w praktycznych środowiskach operacyjnych. W tym celu analogowe oraz antenowe anulowanie zintegrowano w dwóch różnych prototypowych transceiverów FD. Pierwszy prototyp, jednoportowy węzeł radiowy anteny, przedstawiono na rysunku 8.13 (a).



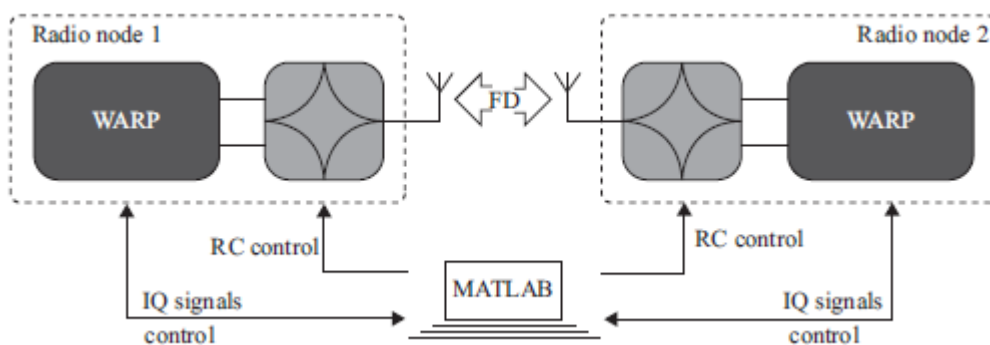
Transceiver łączy w sobie duplexer bilansu elektrycznego i algorytmy usuwania pasma podstawowego. Płyta radiowa WARP v3 i framework WARPLab 7 zostały użyte do integracji. Obwód bilansu elektrycznego łączy się między portami RF urządzenia WARP a anteną z pojedynczym portem i izoluje odbierane sygnały z przesyłanych sygnałów. Funkcja pasma podstawowego, w tym sygnały testowe i cyfrowe anulowanie pasma podstawowego, została zaimplementowana w programie MATLAB. Sygnał OFDM z 64 podnośnymi i sygnałem 20 MHz szerokości pasma została wykorzystana jako przebieg sygnału testowego. W demonstratorze zastosowano dwa liniowe schematy cyfrowej eliminacji zakłóceń, oparte na domenie czasu i dziedzinie częstotliwości. Drugim zaimplementowanym prototypem FD jest dwuportowy węzeł radiowy anteny. Ten transceiver FD integruje podwójnie spolaryzowaną antenę i sieć aktywnego anulowania opisaną wcześniej w części 8.3.2. Architektura nadajnika-odbiornika radiowego została przedstawiona na rysunku 8.13 (b).



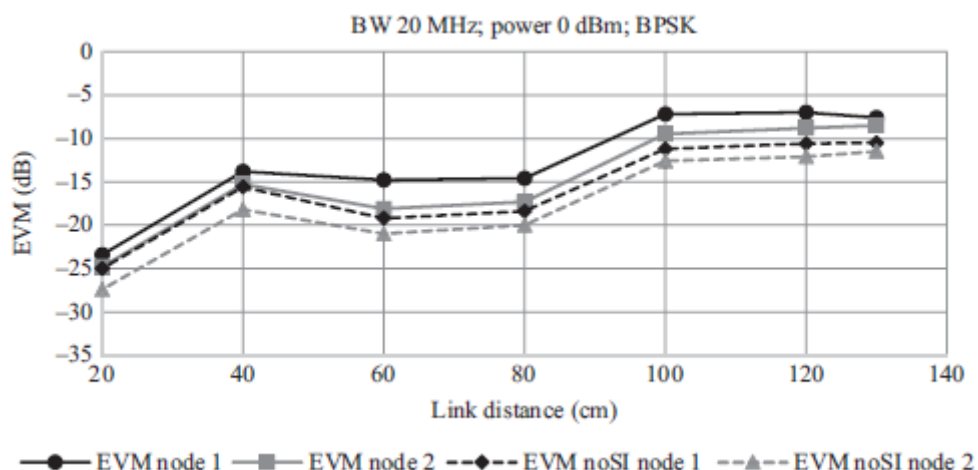
Podobnie jak w przypadku węzła radiowego z pojedynczą anteną, do integracji została użyta płyta radiowa WARP v3 i framework WARPLab. Pasma podstawowe FD i blok eliminacji cyfrowej są również takie same jak w prototypie pojedynczej anteny. Aktywna sieć anulowania jest dostrajana za pomocą algorytmu gradientowego spadku, który jest zaimplementowany w zewnętrznym mikrokontrolerze. Zaimplementowane prototypy sprzętu nadawczo-odbiorczego dla anteny jednoportowej i dwuportowej anteny przedstawiono odpowiednio na rysunku 8.14 (a) i (b).



Platforma oceniająca FD opiera się na dwóch węzłach radiowych FD, które komunikują się za pomocą bezprzewodowego łącza punkt-punkt FD, jak pokazano na rysunku 8.15

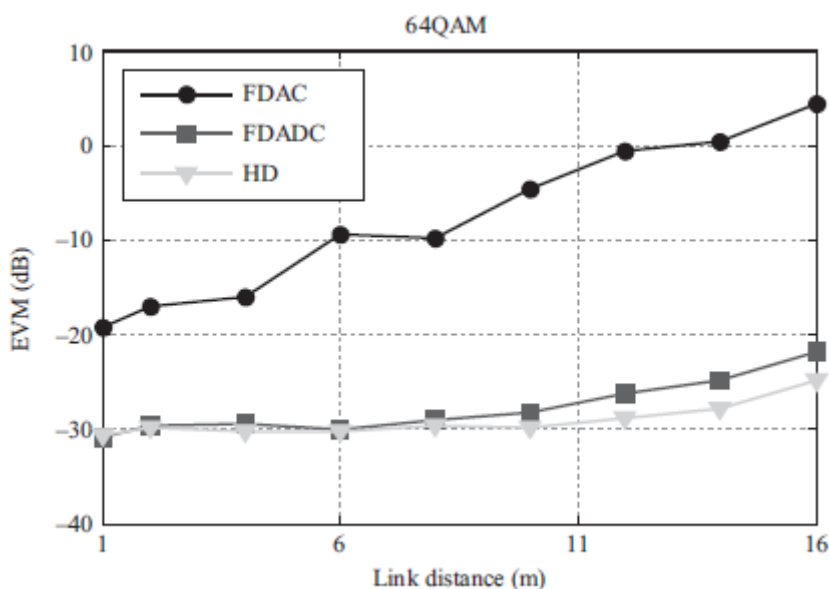


dla rozwiązania z anteną jednoportową. Pomiary oceny przeprowadzono w nieekranowanym otwartym laboratorium środowisko z ludźmi i obiektami zbliżającymi się do węzłów radiowych FD. Kanaly w paśmie ISM 2,4 GHz zostały wykorzystane i zatwierdzono różne moce transmisji i schematy modulacji. Rysunek 8.16



ilustruje wydajność systemu z rozwiązaniem anteny jednoportowej na różnych odległościach łącza, gdy transmitowany jest sygnał BPSK z mocą wyjściową WARP 0 dBm. Wartość wektora błędu (EVM) konstelacji po demodulacji została zmierzona w obu węzłach radiowych, gdy węzły działają w FD iw półduplexie (brak SI). Wyniki pokazują, że wydajność EVM podczas operacji FD ściśle odpowiada wydajności trybu półduplex. Oznacza to, że SI jest wystarczająco zredukowany przez rozwiązania SIC, aby utrzymać SNR w odbiorniku. Główne ograniczenia związane z odległością łącza wynikają z problemów liniowości pierwszego prototypu EBD omówionego już w części 8.3.2. Całkowite wyeliminowanie zakłóceń samoczynnych zapewniane przez jednoportowy węzeł radiowy anteny wynosi 65 dB w paśmie 20 MHz.

Rysunek 8.17



ilustruje wydajność systemu z rozwiązaniem anteny dwuportowej w różnych odległościach łącza, gdy sygnał 64 QAM jest transmitowany z mocą nadawczą WARP 0 dBm. W tym przypadku FDAC na rysunku 8.17 reprezentuje EVM, gdy stosuje się tylko anulowanie analogowe w celu zredukowania SI (analogowa SIC wynosi 60 dB w 20 MHz BW), FDADC reprezentuje EVM, gdy stosowane są zarówno analogowe, jak i cyfrowe anulowanie (cyfrowe anulowanie zapewnia 30 dB dodatkowego SIC), a HD to wydajność EVM łącza półduplexowego. Uzyskane wyniki pokazują, w jaki sposób łącze FD zapewnia



podobną wydajność do połączenia half-duplex, a jedynie niewielka degradacja EVM wynosząca 2% jest obserwowana przy maksymalnej odległości łącza wynoszącej 16 metrów. Podsumowując, przeprowadzone pomiary wydajności wskazują, że osiągalna zdolność do samoczynnej eliminacji zakłóceń wynosi od 65 dB do 90 dB w paśmie 20 MHz, z rozwiązaniem anteny jednoportowej i dwuportowej anteny. Mimo że nie udało się osiągnąć docelowej eliminacji interferencji 100 dB za pomocą tych prototypów sprzętowych, wyniki są zachęcające, ponieważ istnieją już zidentyfikowane rozwiązania pozwalające przewyżnić ograniczenia wynikające z pierwszych prototypów implementacji.

### **8.3.5 Systemy z wieloma dupleksami z wieloma antenami**

Wiadomo, że osiągnięcie przestrzennego wzmocnienia kanału bezprzewodowego za pomocą wielu technik antenowych w systemach MIMO jest jednym z najważniejszych przełomów w technologii bezprzewodowej w ciągu ostatnich kilku dekad. W związku z tym dobrze wiadomo, że jeśli FD będzie szeroko stosowany, musi działać dobrze w systemach MIMO. Jednak rozszerzenie schematu FD, a dokładniej anulowanie samo-interferencji, na konfigurację MIMO, nie jest bynajmniej proste. Faktycznie operacja FD i współistnienie z technologią MIMO jest obecnie znana jako ważna bariera w popularyzacji technologii FD. Problem w projektowaniu i implementacji systemu FD MIMO jest wielozakresowy, ponieważ każdy sygnał transmisyjny koliduje ze wszystkimi lokalnymi odbieranymi łańcuchami antenowymi / RF, jak pokazano na rysunku 8.18, gdzie zakłada się, że każda antena może nadawać i odbierać w tym samym czasie (pojedynczy praca w trybie pełnego duplexu anteny) za pomocą cyrkulatora lub obwodu duplexera. Wykazano, że sprawność widmowa FD MIMO jest jednakowo lepsza niż jego zoptymalizowany odpowiednik półduplexowy i prawie się podwaja, gdy działa w ramach ograniczeń zakresu dynamicznego. Podobne podejście zastosowano w celu uzyskania osiągalnych szybkości przekaźnika FD MIMO opartego na dekodowaniu i przekazywaniu oraz techniki zwiększania dolnej granicy dla szybkości transmisji. Gdy stosuje się wiele anten w transceiverze FD, jednym ze sposobów tłumienia samo-interferencji byłoby zastosowanie dodatkowych przestrzennych stopni swobody, tj. Poprzez zastosowanie technik takich jak formowanie wiązki, sterowanie null i wstępne kodowanie w celu tłumienia samo-interferencji. Jedną z pierwszych propozycji systemu FD MIMO wymaga anten 4M do zbudowania radiotelefonu MIMO z anteną FD M, a nawet wtedy nie zapewnia wymaganej eliminacji samo-zakłóceń dla systemów Wi-Fi (pasmo 20 MHz), aby osiągnąć oczekiwane podwojenie wydajność. W kontekście przekaźników FD MIMO omówiono kilka analiz i zawarte w nich odwołania). Z drugiej strony, zakładając, że w obwodach RF i analogowym osiąga się pewien poziom eliminacji samoczynnej interferencji FD MIMO, rozwiązania w cyfrowym paśmie podstawowym w celu stłumienia pozostałych resztkowych samo-interferencji i jej nieliniowego komponentu są wspólnym podejściem w analizie takie systemy, w których motywuje się rozwój niezawodnych technik przetwarzania sygnałów w celu złagodzenia pozostałościowej samo-interferencji. Nieliniowe modelowanie i anulowanie samo-interferencji z uwzględnieniem wpływu nierównowagi IQ i nieliniowości wzmacniacza mocy i pamięci zostało zbadane w [48], podczas gdy algorytmy wstępnego kodowania w cyfrowym paśmie podstawowym zostały użyte w celu złagodzenia pozostałościowej samo-interferencji dla FD System MIMO. Analizy te pokazują, że oprócz anulowania samo-interferencji w obwodach RF i analogowych, zastosowanie algorytmów wstępnego kodowania w cyfrowym paśmie podstawowym może pomóc w poprawie wydajności systemu. Wydajność tych rozwiązań, podobnie jak w przypadku ich pojedynczych anten, zależy od dokładności estymacji kanału samo-interferencyjnego MIMO, a także od tego, jak dokładnie estymowane i modelowane są nieliniowe komponenty samo-interferencji. W podejściu alternatywnym, wstępne kodowanie było stosowane w rodzaju mieszanego podejścia sygnałowego w celu anulowania samo-interferencji FD MIMO, podczas gdy kształtowanie wiązki jest wykorzystywane w transmisji sygnału. W większości wyżej wymienionych technik zakłada się, że już samo anulowanie samo-interferencji przy

częstotliwości radiowej jest już osiągnięte, co w połączeniu z rozsądną izolacją nadawczo-odbiorczą, może zlikwidować znaczną ilość samo-interferencji. Inne podejścia proponują zastosowanie dodatkowych łańcuchów odbierania / nadawania w celu odtworzenia komponentu samo-interferencji, co nie jest uzasadnione ze względu na dodatkowy koszt / złożoność projektu / implementacji przekaźnika MIMO. Niewiele trwa dyskusja na temat praktycznych rozwiązań w celu anulowania / łagodzenia samo-interferencji z wielu nadajników systemu MIMO na poziomie RF / analogowym. Prostym podejściem byłoby rozszerzenie eliminacji samo-interferencji w rozwiązaniach z pojedynczym wejściem z jednym wyjściem do systemu MIMO za pomocą dedykowanego obwodu anulowania dla każdego sygnału nadawczego w każdym łańcuchu odbiorczym nadajnika-odbiornika MIMO. To jednak wykładniczo zwiększa złożoność systemu nadawczo-odbiorczego FD MIMO. Niedawno zaproponowano praktyczną metodę pokonania wyzwania polegającego na anulowaniu samo-interferencji pomiędzy wszystkimi łańcuchami nadawczo-odbiorczymi FD MIMO. Zamiast wprowadzenia oddzielnej kopii obwodu anulowania dla każdej pary łańcuchów nadajnika-odbiornika MIMO i algorytmu anulowania dla każdej pary łańcuchów, które doświadczają tak zwanej przenikania, zastosowano tylko jeden obwód anulowania na łańcuch odbiorczy. Ponadto założono, że łańcuchy MIMO są kolokowane i mają podobne środowisko, tak że kanały samo-interferencji ze wszystkich łańcuchów transmisyjnych do każdego łańcucha odbiorczego są wspólne. Co więcej, stosuje się sposób szacowania kanałów samointerferencji ze wszystkich łańcuchów transmisji do każdego z łańcuchów odbiorczych jednocześnie, co poprawia ogólną wydajność, która w przeciwnym razie ulegałaby degradacji liniowej z liczbą anten MIMO na każdym końcu łącza. Prace zostały prototypowane, a wydajność proponowanych programów została oceniona przy użyciu gotowych radiotelefonów i sprzętu testowego. Pozostaje pytanie, czy założony wspólny kanał dla M łańcuchów transmisyjnych dla każdego łańcucha odbiorczego jest zawsze dobrym oszacowaniem kanałów samo-interferencji MIMO. Ponadto anulowanie samo-interferencji z wykorzystaniem opóźnienia i tłumienia w domenie RF ma swoje własne ograniczenia w odtwarzaniu dużych opóźnień. Ogólnie rzecz biorąc, anulowanie samo-interferencji w systemach MIMO nadal wymaga starannego rozważenia i nowatorskich podejść. Ponadto, aby zwiększyć zaufanie, rozwiązania te muszą zostać wdrożone na prawdziwych urządzeniach nadawczo-odbiorczych, a wydajność należy porównać z systemami MIMO pod względem osiągalnej przepustowości w różnych kanałach. Tymczasem rozwiązania cyfrowe mogą pomóc w poprawie ogólnej eliminacji zakłóceń w działaniu systemów FD MIMO

## **8.4 Transmisja full-duplex w sieciach bezprzewodowych**

### **8.4.1 Ogólne**

Transmisja FD jest atrakcyjną technologią interfejsu lotniczego w celu poprawy wydajności systemu w przyszłych systemach bezprzewodowych i ostatnio zyskała dużą uwagę ze strony środowisk akademickich i przemysłu. Ta sekcja przedstawia niektóre wyniki na poziomie systemu lub sieci uzyskane w projekcie DUPLO, a także wskazuje na znaczące wyniki badań uzyskane w innych pracach badawczych. W sieciach infrastrukturalnych, takich jak sieci komórkowe, planowanie odgrywa kluczową rolę w zbieraniu korzyści transmisji FD. Jednoczesna transmisja w górę i w dół wprowadzona w trybie FD powoduje dodatkowe zakłócenia wewnątrzkomórkowe i międzykomórkowe. Dlatego, aby osiągnąć potencjalny zysk trybu FD, niezbędny jest inteligentny program szeregujący do parowania użytkowników łącza w dół i łącza zwrotnego z odpowiednimi mocami transmisji. Ze względu na kombinatoryjny charakter parowania wielu użytkowników łącza zwrotnego i łącza pobierającego na każdej podnośnej, a także trudność optymalnego przydziału mocy do każdej podnośnej, alokacja zasobów w systemach FD w celu maksymalizacji wydajności sieci jest bardzo trudna. Ponadto, chociaż moc nadawczą użytkowników łącza zwrotnego można regulować, zakłada się, że moc nadawcza stacji bazowych jest stała. W mobilnych sieciach ad-hoc transmisja FD pomaga rozwiązywać problemy z

ukrytymi węzłami i zmniejszać kolizje, a tym samym zapewnia środki do zwiększania przepustowości sieci. W idealnym scenariuszu dwukierunkowego przekazywania FD, w którym samo-interferencja w węźle przekaźnikowym jest doskonale anulowana, wykazano, że systemy przekaźników FD generalnie osiągają łączność ergodyczną między dwoma końcami dwukrotnie większą niż systemy duplexowe. Ograniczona zdolność do samoczynnej eliminacji zakłóceń w węźle przekazującym zmniejszy wzmacnienie FD. Jednakże, np. W przypadku przekaźników infrastruktury, samoczynna interferencja może zostać skutecznie anulowana za pomocą rozwiązań antenowych. W przekaźnikach wieloprogramowych i mobilnych sieciach ad hoc główne wyzwania związane z transmisją FD są związane ze zwiększoną interferencją między węzłami, ponieważ wszystkie aktywne węzły transmitują jednocześnie. Jednym z rozwiązań jest zastosowanie anten kierunkowych w celu zmniejszenia interferencji między węzłami. Również wydajny protokół rutowania może być ważnym elementem zwiększającym przepustowość od końca do końca. Proponowane są różne rozproszone protokoły MAC dla transmisji FD, w tym Contraflow i FD-MAC. Zgłoszone zwiększenie przepustowości protokołu FD-MAC przez transmisję półduplexową wynosi 70%. Innym protokołem MAC zaproponowanym dla radia FD jest Janus. Protokół ten jest oparty na scentralizowanym mechanizmie szeregowania i odnotowano osiągnięcie 2,5-krotności przepustowości opartego na systemie półduplexowym opartego na CSMA / CA.

#### **8.4.2 Transmisja full-duplex w małych systemach radiokomunikacyjnych**

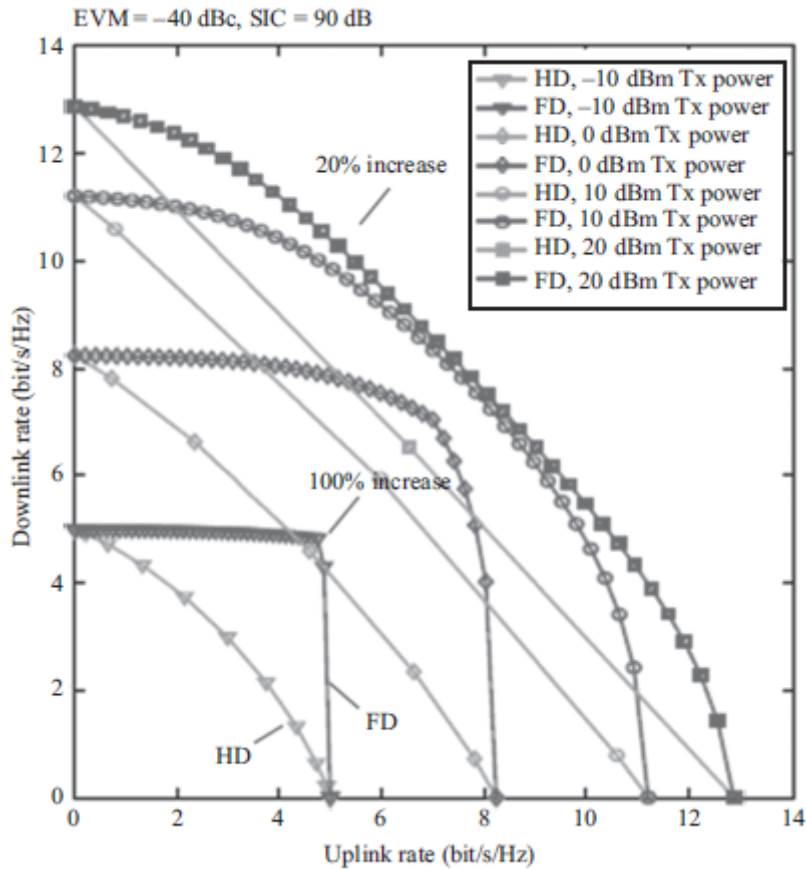
W tej sekcji omówiono potencjalne przypadki użycia transmisji FD w radiu małej sieci systemów komunikacji i ocenia osiągalne zyski wydajności w FD transmisja przez półduplex (HD) w różnych konfiguracjach sieci. Omawiane przykłady są oparte na badaniach wydajności systemu przeprowadzonych w projekt DUPLO. Scenariusze omówione w tej sekcji wymieniono poniżej. To jest warto zauważyć, że wydajność i algorytmy i protokoły związane z ocenione przypadki badania są traktowane w skrócie w tej sekcji.

1. Łącze punkt-punkt jest podstawową jednostką i podstawowym elementem wielu sieci bezprzewodowych. Na przykład w bezprzewodowej sieci lokalnej (WLAN), każda transmisja pomiędzy urządzeniami użytkownika i ich punktem dostępu (AP) może być traktowana jako połączenie punkt-punkt podczas okresu transmisji. Kolejnym przykładem może być bezprzewodowy backhauling w małych wdrożeniach komórkowych na platformach mobilnych, takich jak pociągi, samochody lub samoloty.
2. Samodzielne małe komórki LTE [83-86] ze stacją bazową FD obsługującą wiele urządzeń UE, które działają (a) w trybie półduplexu lub (b) w trybie pełnego duplexu. Typową realizacją tego zastosowania może być mała komórka umieszczona w domu na przykład w obszarze mieszkalnym.
3. Wiele małych komórek LTE [83-86] ze stacją bazową FD obsługujących UE, które działają (a) w trybie półduplexu lub (b) w trybie pełnego duplexu. Typową realizacją tego przypadku może być wdrożenie niewielkich komórek w przestrzeni biurowej, na arenie sportowej lub w centrum handlowym, aby wymienić tylko kilka.

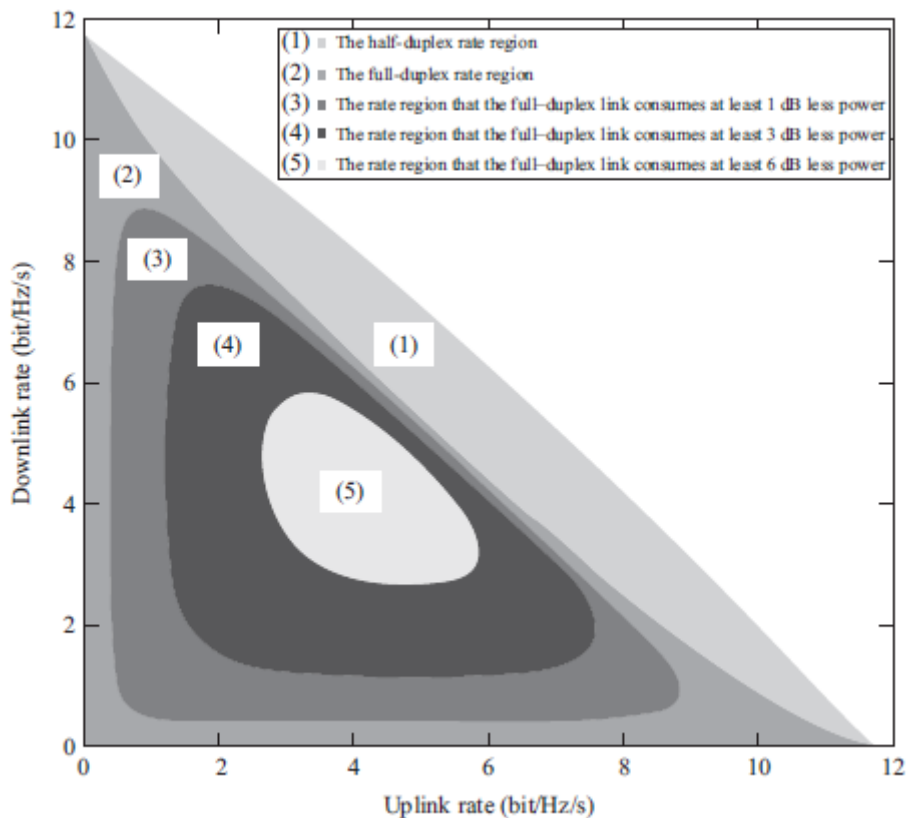
##### **8.4.2.1 Łącze punkt-punkt**

Połączenie punkt-punkt między dwoma węzłami stanowi podstawowy przypadek użycia dla transmisji FD. Pierwsze wyniki dotyczące wydajności łącza FD są oparte na porównaniu osiągalnego obszaru częstotliwości i sumarycznej wydajności energetycznej dla dwukierunkowego połączenia FD lub półduplexa połączenie punkt-punkt z wykorzystaniem przebiegu OFDM wspomaganego prefiksem cyklicznym (CP) do transmisji sygnału. Wpływ niedoskonałości urządzenia nadawczo-odbiorczego na wydajność łącza FD został uwzględniony w modelu systemu. Nie-idealność jest określana ilościowo

jako poziom wektora błędu (EVM) i jest modelowana jako szum Gaussa dodany do oryginalnego sygnału przez nadajnik. Zakładany transceiver FD tłumi samo-interferencję poprzez trzystopniowy proces. Etapy to izolacja anteny, anulowanie RF i cyfrowe anulowanie pasma podstawowego. Rysunek 8.19



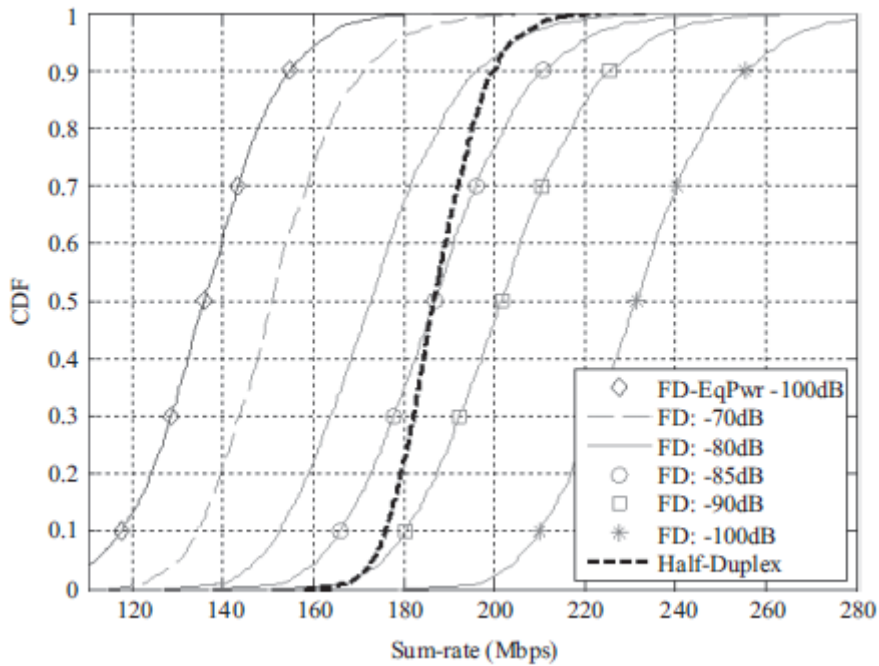
pokazuje osiągalny obszar częstotliwości dla połączeń półdupleksowych i FD o różnych poziomach mocy transmisji, przy założeniu, że odległość łącza między węzłami wynosi 20 m. Wyniki wskazują, że przy krótkich odległościach transmisji maksymalna stawka sumowa osiągnięta przez łącze FD jest prawie dwukrotnie większa niż w przypadku łącza półdupleksowego, gdy moc nadawania wynosi -10 dBm. Gdy moc nadawania zostanie zwiększona do 20 dBm, maksymalna poprawa szybkości sumowania zostanie zmniejszona do około 20%. Wydajność energetyczna jest również ważna do oceny w porównawczych ramach komunikacji FD i half-duplex. W tym celu, zużyta moc transmisji, aby osiągnąć pewien obszar szybkości, jest następnie oceniana. Rysunek 8.20



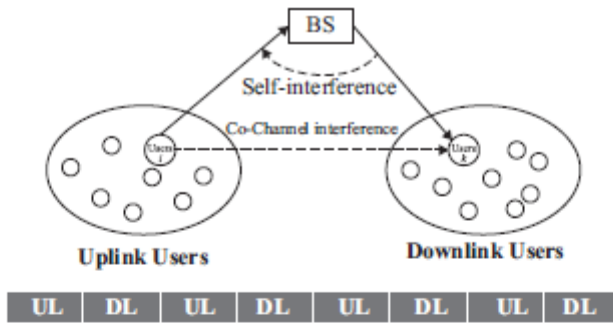
pokazuje możliwe do uzyskania kontury obszaru prędkości z różnymi wartościami poboru mocy. Można wywnioskować, że łącze FD może osiągnąć ponad 75% swojego regionu szybkości z 1 dB mniejszym zużyciem energii niż łącze półdupleksowe. Ponadto, w obszarze wokół pary szybkości łącza nadawczego-łącza zwrotnego {4,5, 4,5} bit / Hz / s, przewaga efektywności energetycznej jest większa niż 6 dB.

#### 8.4.2.2 Jednoosobowe, pełnodupleksowe, małe komórki

W pierwszym przypadku badania w tej sekcji rozważana jest pojedyncza komórka o promieniu 200 m, gdzie lokalizacje użytkowników są losowo generowane i równomiernie rozmieszczone w komórce. Stacja bazowa FD ma maksymalną moc nadawczą 30 dBm, a każdy z użytkowników ma moc nadawania 23 dBm. Urządzenia użytkownika działają w trybie half-duplex. Całkowita liczba użytkowników wynosi 20, z 10 użytkownikami łącza w dół i 10 użytkownikami łącza w górę. Stacja bazowa wykorzystuje iteracyjny algorytm oparty na teorii gier, aby wspólnie optymalizować alokację mocy łącza w górę i podnośną łącza pobierającego oraz alokację mocy w celu maksymalizacji sumarycznej szybkości systemu. Rysunek 8.21



pokazuje skumulowaną funkcję rozkładu (CDF) sumy stawek dla systemu FD, z różnymi wartościami anulowania interferencji. Jako punkty odniesienia dla porównania stosuje się FD z równym rozdziałem mocy i półduplexem (TDD) z optymalną podnośną i przydziałem mocy. W tym przypadku można stwierdzić, że przy SIC 85 dB, wydajność systemu FD równa się wydajności systemu HD, ale z SIC 110 dB, obserwuje się 40% wzrost wydajności. Co więcej, wyniki pokazują, że proponowany algorytm znacząco poprawił wydajność FD (o około 70%) w porównaniu do równego podziału mocy zasobów, podkreślając znaczenie schematu zarządzania zasobami radiowymi w zbieraniu zysków transmisji FD. Inny przypadek badania z charakterystyką pikokomórki (promień komórki 40 m, moc nadawania BS 23 dBm, moc nadawania UE 24 dBm) i inteligentna strategia szeregowania i strategia alokacji mocy w stacji bazowej wywnioskowały, że samo-interferencja musi anulować co najmniej 80 dB, aby uzyskać wzrost wydajności z FD ponad transmisją half-duplex. Algorytm przydzielania zasobów jest rozszerzony do FD-BS z wieloma scenariuszami FD-UE, gdzie podnośne są przydzielane wyłącznie do użytkowników, aby zapewnić ortogonalność wśród użytkowników i wyeliminować interferencję między użytkownikami. Co więcej, wydajność systemu FD jest oceniana przy różnych ustawieniach systemu, takich jak obszar zasięgu komórki i maksymalna moc transmisji węzłów. Ocena na poziomie systemu wykazała wyższość techniki FD nad HD jest bardziej widoczny w przypadku komórek o małym zasięgu i niskich węzły mocy transmisyjnej. Na przykład z maksymalną mocą transmisji 5 dBm BS i UE, 39%, 50% i 70% zyski mogą być osiągnięte przez FD z odpowiednio 90, 95 i 100 dB współczynnikami eliminacji zakłóceń. Dzięki odpowiedniemu przydzieleniu zasobów radiowych można uzyskać znaczne zwiększenie wydajności widmowej za pomocą techniki FD z obecnie możliwą eliminacją interferencji możliwości. W zależności od lokalizacji użytkowników mobilnych, kanałów propagacji, możliwości anulowania autowysłania w BS, mocy transmisji telefonu komórkowego użytkownicy i BS itp., czasami lepiej jest używać trybu transmisji HD zamiast transmisji FD. Model systemu pokazano na rysunku 8.22.



(a) Half-duplex TDD



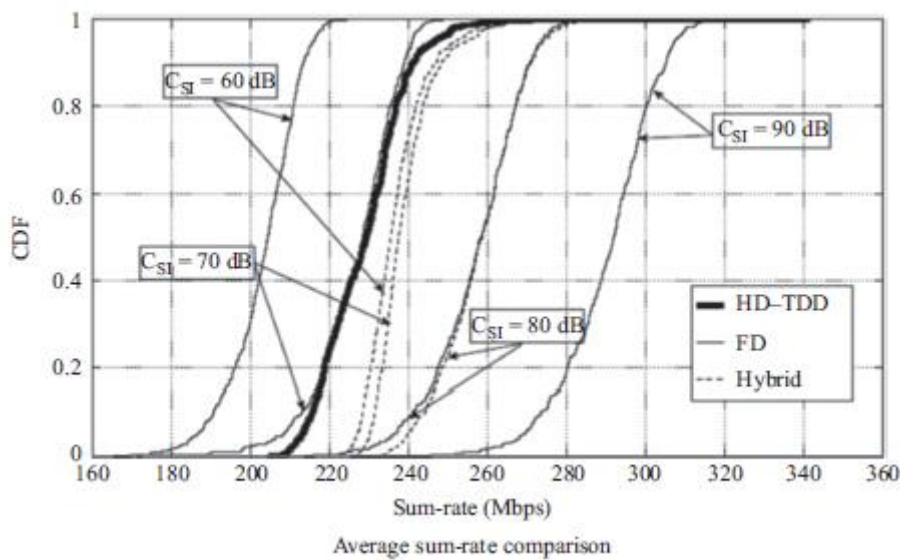
(b) Full-duplex Hybrid TDD

UL: Uplink      DL: Downlink      FD: Uplink + Downlink

### Simulation parameters

Parameter	Settings
Cell radius	40 m
Number of subcarriers	$N = 1024$
$ \mathcal{X}^{UL} ,  \mathcal{K}^{DL} $	[10, 10]
Bandwidth	10 MHz
Maximum BS power	$P_T = 24$ dBm
Maximum user power	$P_i = 23$ dBm
Thermal noise density	-174 dBm/Hz
Noise figure	BS: 13 dB, User: 9 dB
Path loss (dB) between BS and users ( $d$ in km)	LOS: $103.8 + 20.9 \log_{10} d$ NLOS: $145.4 + 37.5 \log_{10} d$
Path loss (dB) between users ( $d$ in km)	LOS: $98.45 + 20 \log_{10} d, d \leq 50$ m NLOS: $175.78 + 40 \log_{10} d, d \leq 50$ m
Shadowing standard deviation	LOS: 10 dB NLOS: 12 dB

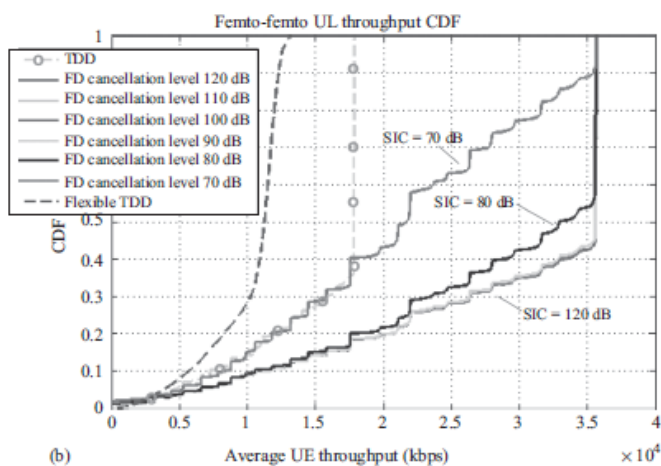
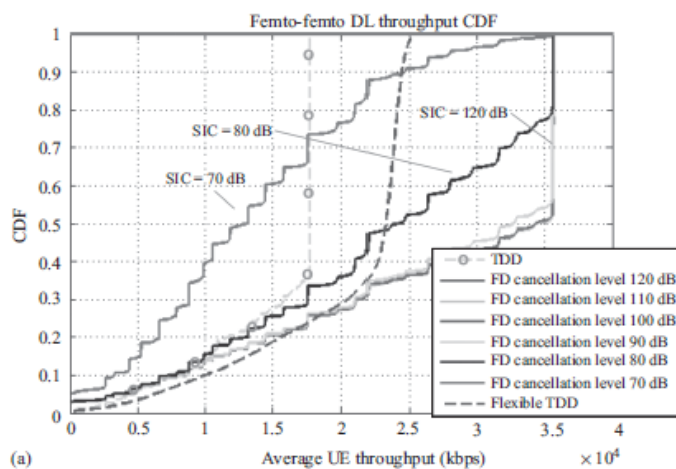
Na rysunku 8.23



pokazano, że przy wysokich wartościach eliminacji zakłóceń własnych, planowanie FD przewyższa harmonogramowanie HD-TDD, a hybrydowe planowanie przełącza do planowania FD, więc korzystne jest przydzielanie każdej szczeliny czasowej do równoczesnej transmisji łącza uplink i łącza w dół. Z drugiej strony, przy niskich wartościach eliminacji zakłóceń własnych, planowanie HD-TDD przewyższa harmonogramowanie FD, a planowanie hybrydowe rozpoczyna przełączanie się do planowania HD-TDD, więc korzystne jest przydzielanie przedziałów czasowych transmisji łącza uplink lub łącza nadawczego. Badania wydajności systemu z częściowymi informacjami o stanie kanału pokazują, jak średnia wydajność wydajności systemu FD (łącze wstępujące plus łącze w dół) przy różnych błędach oszacowania kanału powoduje pogorszenie wydajności wraz ze wzrostem błęd oszacowania. Można łatwo obniżyć wydajność systemu o 10% -20% przy umiarkowanych błędach oszacowania kanału, podkreślając w ten sposób znaczenie wysokiej jakości estymatorów kanału z funkcją FD.

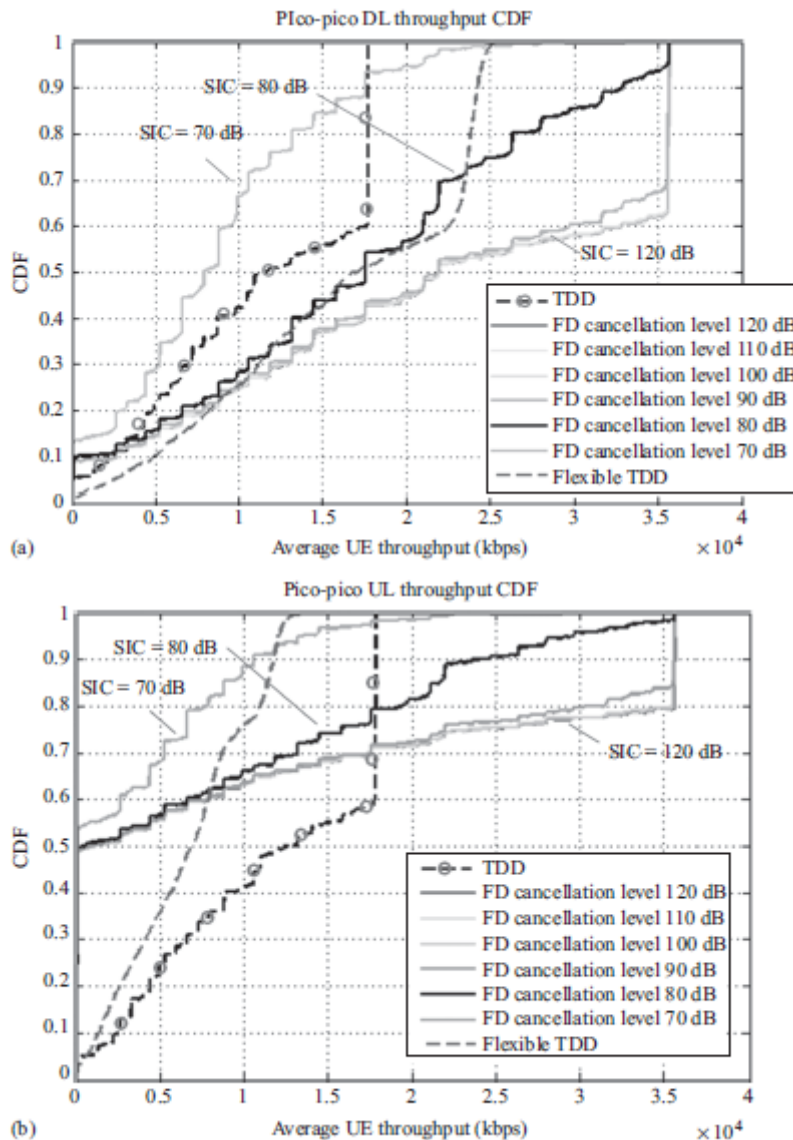
### 8.4.2.3 Wiele półdupleksowych małych komórek

W tej sekcji przedstawiono wyniki pierwszej symulacji systemu dla operacji FD w środowisku wielu małych komórek LTE, przy założeniu, że istnieje tylko jeden sprzęt użytkownika FD (UE) na komórkę. Dlatego skupiamy się na badaniu wpływu interferencji między komórkami spowodowanych transmisją FD. W rezultacie zapewnia się osiągalną przepustowość systemu z różnymi wartościami eliminacji zakłóceń własnych. Tryb półdupleksowy TDD jest używany jako schemat odniesienia w porównaniach. Planowanie użytkownika i kontrola mocy nie są uwzględniane w symulacjach. Rozważane są dwa typy scenariuszy wdrażania wielu komórek, a mianowicie scenariusze wewnętrzne i zewnętrzne. Scenariusze wdrożenia są zgodne z modelami femto i picocell 3GPP. W przypadku scenariuszy wewnętrznych i zewnętrznych średnia przepustowość UE zapewnia kierunki DL i UL. Rysunek 8.24 (a) i (b)



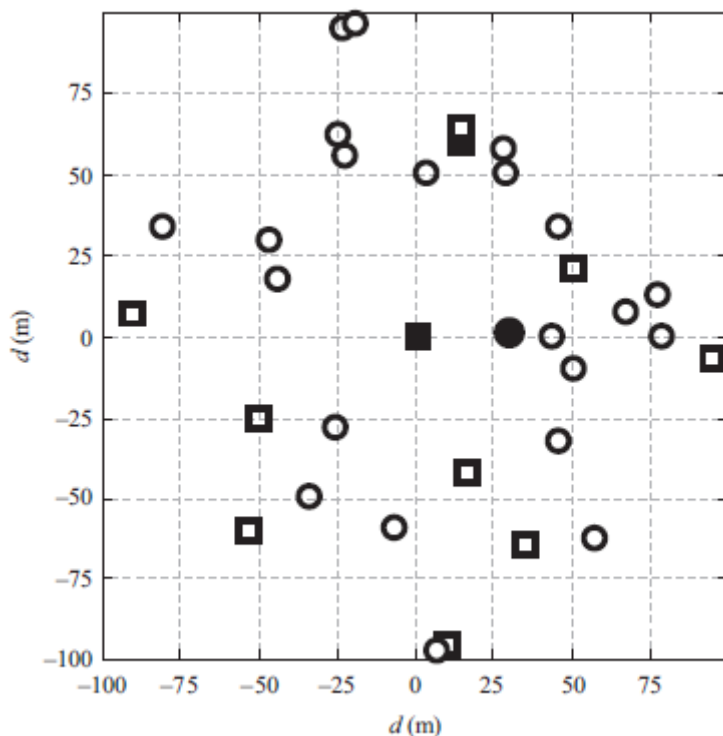


przedstawiają CDF średniej przepustowości UE w DL i UL scenariuszy wewnętrznych, odpowiednio. Z tych liczb, maksimum przepustowości wdrożenia FD jest dwa razy większa niż w przypadku TDD tryb, zgodnie z oczekiwaniami. Warto zauważyć, że schemat FD nie tylko przewyższa TDD w reżimie niskiego SINR, ale także gwarantuje maksymalną osiągalną przepustowość dla około 50% FD UE, jeśli poziom anulowania samo-interferencji wynosi co najmniej 80 dB. Porównując pokrycie FD i TDD w tym konkretnym scenariuszu, ten pierwszy jest tylko nieco gorszy niż ten ostatni w DL, podczas gdy oba są równe w UL. Podczas gdy elastyczny schemat TDD przewyższa standardowy TDD w DL, działa on gorzej w UL. We wszystkich przypadkach, z wyłączeniem FD z samoczynną eliminacją zakłóceń 70 dB, stosowany schemat modulacji i kodowania (MCS) ogranicza osiągalną maksymalną przepustowość. Rysunek 8.25 (a) i (b)

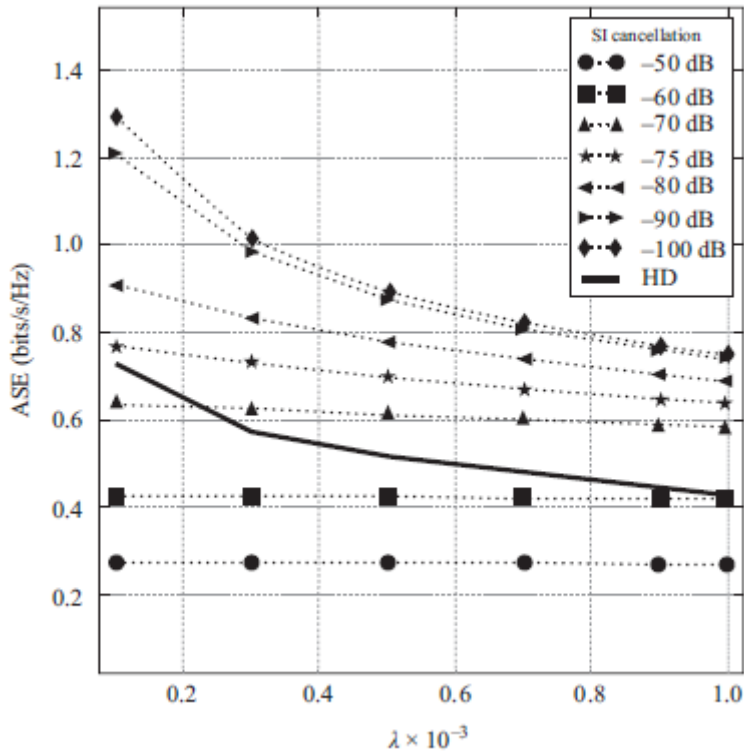


ilustruje, odpowiednio, średnią przepustowość UE w DL i UL scenariuszy zewnętrznych. W DL tych scenariuszy schemat FD z samoczynną eliminacją zakłóceń o wartości co najmniej 80 dB osiąga wyższą przepustowość niż standardowy TDD, jednocześnie zapewniając prawie podobny zasięg (niskie wartości CDF). Co więcej, około 40% urządzeń FD UE osiąga maksymalną przepustowość w kierunku DL. W UL, podczas gdy prawie 30% odbiorników FD UE z samoczynną eliminacją zakłóceń wynoszącą co najmniej 90 dB osiąga przepustowość wyższą niż przepustowość użytkowników TDD, tylko około

20% z nich działa również, gdy poziom anulowania wynosi 80 dB. Zasięg jest w rzeczywistości głównym ograniczeniem tego scenariusza: tylko 5% TDD UE jest w stanie wyłączonym w przeciwieństwie do prawie 50% napędów FD. Powodem jest dodatkowa interferencja między komórkami w rozmieszczeniach FD, a dokładniej DL Zakłócenia krzyżowe UL i słaba izolacja między zewnętrznymi stacjami BS. Na podstawie tych wyników oczywiste jest, że w scenariuszach wewnętrznych z dobrą izolacją pomiędzy stacjami BS, pełna konfiguracja duplexu może przewyższać schematy TDD pod względem średniej przepustowości UE, zapewniając jednakowy zasięg, o ile poziom anulowania samo-interferencji wynosi co najmniej 80 dB. Ponadto, gdy izolacja między stacjami BS jest słaba, zmniejsza się zarówno zasięg, jak i osiągalne szybkości transmisji danych dla urządzeń UE z obsługą FD z powodu wysokiej interferencji krzyżowej DL do UL. Jednak wyniki te są ograniczone do scenariuszy z pojedynczym UE na stronę komórki i potrzebne są dalsze badania, aby potwierdzić, czy nadal utrzymują wdrożenia dla wielu użytkowników (więcej niż jeden UE na komórkę). Z drugiej strony, wyniki te wyraźnie wskazują na potrzebę technik łagodzenia zakłóceń urządzeń w celu zmniejszenia interferencji DL do UL i w rezultacie umożliwienia komunikacji w pełnym duplexie w scenariuszach ze słabą izolacją między węzłami, zwłaszcza BS z wysoką mocą transmisji. Na przykład systemy kontroli mocy i usuwania interferencji międzykomórkowej lub schematy koordynacji są obiecujące jako rozwiązania do zarządzania zakłóceniami. W innym scenariuszu z niewielkim badaniem komórkowym rozważa się przypadkową topologię sieci, w której UE i małe komórki BS są równomiernie rozproszone w obszarze sieciowym, patrz Rysunek 8.26.



Urządzenia są w stanie transmitować i odbierać jednocześnie kosztem samowyzwalania. Nawet przy zastosowaniu zaawansowanych technik eliminacji zakłóceń i tłumienia anteny nadal pozostaje poziom samozszyszczenia, który można modelować jako zmienną losową, która pozwala na emulację różnych konfiguracji (nie) na linii wzroku. Zasadniczo ta struktura analityczna oparta jest na koncepcjach geometrii stochastycznej. Topologia sieci jest modelowana za pomocą procesu punkowego Poissona, natomiast łącza bezprzewodowe śledzą złożony shadowing Log-Normal (LN) i blaknięcie modelu Nakagami-m. Wydajność tej sieci ocenia się pod względem wydajności widmowej, a także prawdopodobieństwa wyłączenia. Rysunek 8.27



pokazuje, że pod względem średniej wydajności widmowej (ASE) schematy FD stają się bardziej atrakcyjne i lepsze od schematu HD, mimo że doświadczają dodatkowych zakłóceń. Na przykład, dla poziomów eliminacji interferencji 75 dB (oznaczonych jako 75 dB na Rysunku 8.27), schemat FD prezentuje o 50% wyższą wydajność widmową niż HD. Jeszcze większe zyski można uzyskać, jeśli anulowanie samointerferencji zostanie dodatkowo poprawione.

### 8.5 Podsumowanie

Technologia FD zyskała wiele uwagi w ciągu ostatnich kilku lat. Uwaga została przyspieszona, ponieważ opracowywane są nowe i interesujące techniki, które próbują rozwiązać jeden z głównych problemów komunikacji FD: samo-interferencję w węzłach radiowych. Najbardziej obiecujące techniki pochodzą od grup badawczych, które rozważają podejście systemowe, w którym samo-interferencje są odrzucane i anulowane w różnych miejscach w radiu, tj. w antenie oraz w domenie RF i cyfrowej. Zainteresowanie zastosowaniem FD do przyszłej ewolucji systemów bezprzewodowych znacznie wzrosło. Bezprzewodowa transmisja FD jest uważana za obiecującą technikę interfejsu bezprzewodowego dla systemów bezprzewodowych 5G, ponieważ zajmuje się kluczowymi obszarami, w których potrzebne są ulepszenia systemu, takimi jak sprawność widmowa, pojemność łącza i opóźnienie sieci. Aby uzyskać całkowite odrzucenie zakłóceń własnych o wartości przekraczającej 100 dB, wymagana jest kombinacja technik samowykształcenia i eliminacji zakłóceń. Chociaż cel ten jest trudny, najnowszy projekt nadajnika-odbiornika pokazuje już, że jest on możliwy do osiągnięcia nawet w przypadku rozwiązania z anteną jednoportową. Osiągnięcie tego w gęsto zintegrowanym radiu stanowi kolejne wyzwanie. Cel ten jest ważny, ponieważ umożliwiłby otwieranie aplikacji technologii FD na urządzenia kompaktowe. Rozwiązania urządzeń nadawczo-odbiorczych omówione w tym rozdziale są potencjalnymi kandydatami na urządzenia kompaktowe FD, chociaż ich pierwsze prototypy wciąż pozostają w tyle za docelową wartością 100 dB. W związku z tym przyszłe możliwości i potrzeby rozwojowe są związane z poprawą wydajności (zdolność do eliminacji interferencji własnej i przepustowość) i ulepszeniami projektu (poprawiona liniowość, rozwiązania architektoniczne

redukujące szum fazowy, ulepszone algorytmy pasma podstawowego). Inteligentne telefony i inne urządzenia radiowe obsługują wiele protokołów radiowych i pasm częstotliwości. Dlatego, z praktycznego punktu widzenia bezprzewodowego nadawczo-odbiorczego, ważne jest, aby w konstrukcji transceivera FD uwzględnić integrację z innymi protokołami radiowymi w tym samym urządzeniu radiowym. Konstrukcja nadawczo-odbiorcza obsługująca elastyczne drukowanie dwustronne jest również potencjalnym kierunkiem przyszłych badań. Szansę dla FD w większych współczynnikach kształtu należy również rozważyć w odniesieniu do przyszłego 5G, na przykład z podłączonymi samochodami. Ponadto rozwój technik FD dla systemów z wieloma antenami jest kolejnym potencjalnym kierunkiem badań. Potencjalne przypadki użycia dla FD w sieciach 5G to FD BS obsługujące sprzęt użytkownika HD, urządzenie FD UE, połączenia urządzenie-urządzenie, przekazywanie, transmisja w sieciach typu mesh i (samo) backhauling. Badania na poziomie systemu wskazują, że w systemach o małej powierzchni transmisja FD może zapewnić wzrost wydajności ponad półdupleks, nawet przy umiarkowanych poziomach anulowania interferencji w transceiverze FD. W związku z tym nie jest konieczne projektowanie operacji FD w oparciu o budżet łącza najgorszego przypadku w celu uzyskania korzyści na poziomie systemu bezprzewodowego. Rozwiązanie na poziomie systemu może być przystosowane do zmiany trybu pracy pomiędzy transmisją półdupleksową a transmisją FD. Naturalnie, wysoka zdolność do anulowania interferencji w transceiverze FD jest korzystna w maksymalizowaniu wzmocnienia poziomu systemu transmisji FD. Zarządzanie zasobami radiowymi jest kluczowym czynnikiem, który należy wziąć pod uwagę, gdy chcemy efektywnie wykorzystać transmisję FD w sieciach bezprzewodowych. Wraz z wprowadzeniem sieci 5G nastąpi znaczny wzrost liczby małych komórek. Zatem koordynacja, samoorganizacja, elastyczne i wydajne wykorzystanie zasobów widmowych, planowanie i zarządzanie zasobami radiowymi w ogóle są aspektami, które będą fundamentalne dla następnej generacji komunikacji bezprzewodowej. Na przykład wpływ ograniczeń sygnalizacji i niedoskonałych informacji o stanie kanału (np. W algorytmach planowania) na wydajności systemu, rozwiązania do walki z interferencją między komórkami i współistnienie z systemami half-duplex wymagają dalszych badań, aby uzyskać więcej kompleksowe spojrzenie na to, jak najlepiej byłoby wdrożyć FD w przyszłych systemach 5G. Również protokoły MAC i rozwiązania routingu mogą zostać zoptymalizowane w celu poprawy ich wydajności w praktycznych mobilnych sieciach ad hoc. Ogólnie rzecz biorąc, wyniki omówione w tym rozdziale są zachęcające i otwierają drogę do wprowadzenia transmisji FD jako technologii interfejsu lotniczego dla przyszłych systemów 5G.

# Odłączony dostęp do łącza uplink i downlink w heterogenicznych sieciach

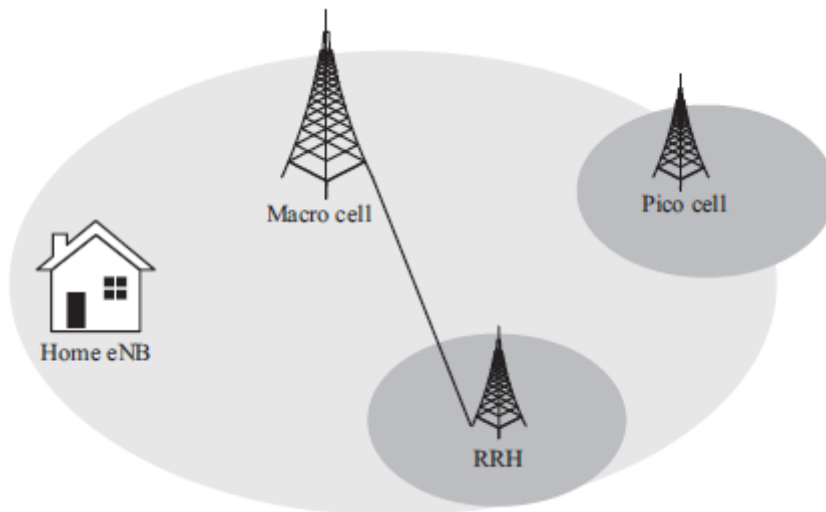
## 9.1 Wprowadzenie

Koncentracja danych UE nie tylko wymagają większej przepustowości od sieci bezprzewodowych, ale także zorientowanej na usługi jakości usług (QoS) i porównywalnej jakości doświadczenia (QoE) zarówno w łączu w dół (DL), jak i w łączu w górę (UL). Wraz z wprowadzeniem Internet of Things (IoT), komunikacji między maszynami, usługami w chmurze i powszechnie używanymi mediami społecznościowymi, UE i urządzenia są generalnie coraz bardziej generatorami treści niż przedtem. W konsekwencji nowe generacje komunikacji mobilnej muszą opracować strategie, które poprawią ogólne wrażenia urządzenia. Dodanie inteligencji do warstw Media Access Control (MAC) i Radio Resource Control (RRC) wykazało znaczną poprawę pod względem szybkości transmisji danych i uczciwości. Jednak niestała QoE UE w obszarze komórki motywuje do zwiększenia liczby obsługujących eNB przez dodanie komórek o krótszym zasięgu, zbliżając sieć do UE. Ta zmiana w rozmieszczeniu pozwala na dalszą poprawę pojemności sieci, umożliwiając lepsze równoważenie obciążenia między komórkami i eliminując dziury zasięgu. Jednak ten nowy paradygmat projektowania systemu wiąże się z poważnymi wyzwaniami otwartymi, aby zapewnić poprawne działanie, takie jak: poprawa backhaul, mobilność i zarządzanie zakłóceniami, powiązanie z komórkami i relacja UL / DL. Jedną z kluczowych kwestii przy przechodzeniu w kierunku sieci użytkownika lub usługi jest zapewnienie sieci mobilnej wystarczającej elastyczności, aby wybrać komórkę obsługującą, która lepiej odpowiada wymaganiom urządzenia lub usługi. W tym kierunku, łączenie w dół i rozłączanie wstępujące (DUDe) idzie o krok dalej i pozwala użytkownikowi niezależnie przesyłać i odbierać do różnych stacji bazowych i odbierać z nich. Zasadniczo DUDe przełamuje twarde i klasyczne ograniczenia selekcji komórek w oparciu o moc odbieraną przez downlink i zapewnia sieci swobodę kojarzenia użytkowników z różnymi eNB w DL i UL. Ta część opisuje szczegółowo technikę DUDe. Po pierwsze, sekcja 9.2 omawia główne wyzwania, przed którymi stoi HetNets w zakresie planowania radiowego i zarządzania zakłóceniami, oraz dokonuje przeglądu potencjalnych rozwiązań, aby rozwiązać te problemy. Następnie odłączanie UL i DL jest uznawane za nową technikę, która może skutecznie poprawić wydajność HetNet, jak pokazano w części 9.4, a także może mieć pozytywny wpływ na wydajność innych technologii dostępu radiowego, takich jak Carrier Aggregation (CA), transmisja Coordinated Multipoint i odbiór (CoMP) lub falę milimetrową (sekcja 9.5). Umożliwione architektury sieci dostępu radiowego omówiono w Rozdziale 9.6, a możliwości i wyzwania wyższej warstwy omówiono w Rozdziale 9.7.

## 9.2 HetNets wyzwania w sieci 4G

LTE-A musi zmierzyć się ze zmienną QoE w komórkach z powodu różnicy w wydajności widmowej w zależności od pozycji UE. Dodawanie widma lub poprawa adaptacji łącza zapewnia szybszą łączność, jednak nie ma jednorodnej wydajności faktycznie spełnione. Jednym z wyzwań, z którym muszą się zmierzyć 4G i przyszłe technologie, takie jak nadchodzący 5G, są nowe wymagania dotyczące wydajności spektralnej obszaru i dystrybucji użytkowników. Dalsze ulepszenia w celu osiągnięcia uczciwego działania na użytkownika wzdłuż komórki są możliwe dzięki zwiększeniu gęstości rozmieszczenia węzłów eNB. Obecnie sieci komórkowe zmieniają się z jednopoziomowego, jednorodnego podejścia sieciowego do wielopoziomowego heterogenicznego, tak zwanego HetNets. W ostatnich latach stało się popularnym podejściem jako wydajne i skalowalne rozwiązanie poprawiające przepustowość sieci w newralgicznych punktach; jest to również realne rozwiązanie, które ma poprawić sprawiedliwość, ponieważ sieć zbliża się do UE. HetNet to wdrożenie sieciowe składające się z Macro Cells lub MCells i nałożonych węzłów niskiej mocy, takich jak komórki piko, femto, RRH lub przekaźniki, określane ogólnie jako małe komórki lub SCell. MCells to eNB o dużej mocy,

zwykle zlokalizowane wzdłuż obszaru geograficznego, z takimi celami jak maksymalizacja zasięgu lub redukcja zakłóceń. Pęcherzyki są odpowiedzialne za wyeliminowanie otworów ochronnych lub poprawę wydajności poprzez rozładowywanie MCells w obszarach o gorącym punkcie. Rysunek 9.1 pokazuje typowe wdrożenie HetNet.

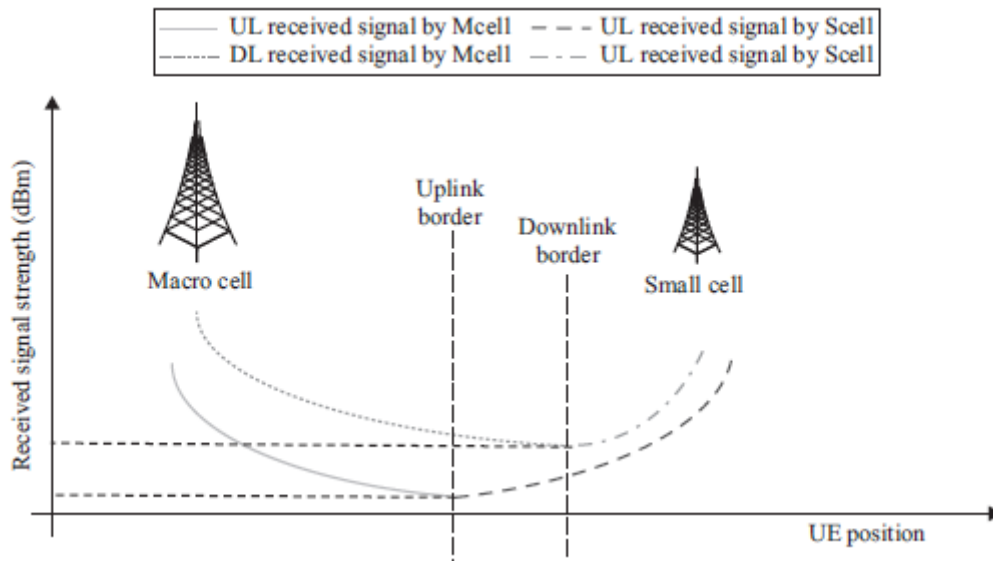


W oparciu o rozmieszczenie częstotliwości powstają dwa rozwiązania: współkanałowe rozmieszczenie, w którym zarówno MCells, jak i SCell mają wspólne pasmo częstotliwości; i dedykowane wdrożenie, w którym każdy typ komórek transmituje na różnych częstotliwościach. Podczas gdy pierwsza pozwala maksymalizować wykorzystanie i pojemność widma, zakłócenia mogą stać się głównym problemem w projektowaniu systemu. Z drugiej strony, dedykowane wdrożenia mogą rozwiązać problem interferencji, kosztem nadmiernego zaopatrywania SCell w zasoby częstotliwości. W tym sensie ten nowy paradygmat wdrażania sieci potencjalnie zwiększa wydajność widmową; jednakże wprowadzono znaczące wyzwania w celu zapewnienia prawidłowego działania HetNets.

### 9.2.1 Wyzwania związane z planowaniem radiowym w HetNecie

Tendencje hiperrozrostu komórek z wieloma poziomami wymagają różnych podejść w planowaniu i projektowaniu sieci, aby spełnić podstawowe cele maksymalizacji stopy użytkowników i wzmocnienia między innymi spektralnej efektywności uczciwego użytkownika. Oczwistym wyzwaniem w planowaniu sieci HetNets jest wdrożenie częstotliwości: jeśli maksymalizacja pojemności jest głównym czynnikiem napędowym lub dostęp do widma jest ograniczony, częstotliwości MCell powinny być ponownie wykorzystane przez inne poziomy. W rozmieszczeniach międzykanałowych interferencja między komórkami jest ważnym ograniczeniem. Zarówno w UL, jak i DL, występuje większa liczba węzłów zakłócających w porównaniu do klasycznych wdrożeń MCell; duża liczba ogniw SCell wykorzystujących te same nośniki może generować duże zakłócenia. W takich przypadkach należy rozważyć zaawansowane procedury zarządzania zasobami radiowymi, aby zminimalizować generowane zakłócenia. Z drugiej strony, dedykowane wdrożenia są atrakcyjne przy dużych przepustowościach, co pozwoli lepiej kontrolować zakłócenia. Z definicji ogniwa są węzłami małej mocy. Bardzo niska moc nadawania DL może powodować niski zasięg SCell, zapewniając bardzo niewielką migrację użytkownika z MCell. Z drugiej strony, wyższe moce nadawania zwiększają rozmiar SCell, co może zwiększać obciążenie SCell. Różnice w gęstości transmisji DL pomiędzy różnymi węzłami i wybór komórki w oparciu o moc odbieraną DL powodują problemy z niezbilansowaniem między połączeniami UL i DL. Problem braku równowagi DL / UL został rozpoznany przez 3GPP. Mówi się, że UE znajduje się w tej sytuacji, jeśli najlepsza obsługująca komórka z perspektywy mocy odbieranej przez UL i najlepsza komórka obsługująca DL są różne. Optymalne połączenie UE w UL jest z SCell,

ponieważ jest bliżej, ale w DL będzie podłączone do MCell. Klasyczne reguły asocjacji komórek mają ogromny wpływ na nierównowagę obciążenia i wydajność UL. Rysunek 9.2 pokazuje graficzny przykład problemu niezbilansowania HetNet.



Muszą być zapewnione SCell z energooszczędnym i niedrogim backhaułem, co dość często okazało się trudne w wdrożeniach SCells. Brak sprzężenia zwrotnego o dużej wydajności może ograniczyć dostępność SCell, co może osłabić potencjalne ulepszenia wynikające z zagęszczania komórek. W związku z tym procedury stowarzyszenia komórek powinny uwzględniać dostępność takich zasobów, które zapewniają użytkownikowi odpowiednią jakość doświadczenia.

### 9.2.2 Strategie usprawnień w HetNecie

Wybór komórki oparty na mocy odbieranej sygnалу odniesienia (RSRP) skutkuje problemami z niezbilansowaniem, ponieważ zasięg MCell jest znacznie większy niż zasięg SCell. Jedną ze strategii, która może rozwiązać ten problem i przynieść trochę uczciwości do UL, jest rozszerzenie zakresu komórek (RE), które umożliwia UE skojarzenie z komórkami, które nie zapewniają najwyższego DL RSRP, zwiększając obszar pokrycia SCell. Zakres SCell jest rozszerzany przez dodanie przesunięcia wyboru komórki do RSRP mierzonego od SCell, a urządzenia użytkownika UE w zakresie rozszerzonego obszaru są powiązane z SCell zarówno w UL jak i DL. W tej sytuacji UL UE jest dość poprawione, podczas gdy DL jest powiązane w sposób nieoptymalny, ponieważ maksymalna otrzymana moc jest nadal zapewniona przez MCell. Chociaż rozszerzanie zakresu węzłów o małej mocy częściowo kompensuje problemy z zakłóceniami UL, interferencja DL na RE UE jest znacznie zwiększona. Badania wykazały, że stosowanie wysokich przesunięć (większych niż 3-6 dB) zwiększa poziom interferencji DL. Z tego powodu zbadano także mechanizmy eICIC w tym kontekście. Techniki te są w większości odmianami tego samego pomysłu: wyciszanie ramek i skoordynowane planowanie. Rzeczywiście, 3GPP wprowadziło możliwość użycia niemal pustych podramek (ABS) od czasu wydania LTE-A 10. W końcu technika rozciągania zasięgu jest ograniczona do umiarkowanych wartości przesunięcia i dostosowanie nie jest trywialne dla bardzo niejednorodnych śladów pokrycia, ponieważ wystąpi w ultra -dla wdrożeń. Ponadto obecne prace badawcze zakładają powolne aktualizacje (skala sekund) w celu koordynacji zakłóceń. Powodem jest zwiększenie RTT (z powodu opóźnień sygnalizacji X2), ale co ważniejsze, potrzeba zagwarantowania stabilności systemu. Jednym z głównych celów projektu HetNets jest wykorzystanie zasobów radiowych w sieciach MCells i SCell w celu uzyskania przepustowości na użytkownika i pojemności systemu podobnej do idealnych wdrożeń sieci dosyłowej. 3GPP wprowadził

koncepcję podwójnej łączności w Wydaniu 12, w którym użytkownik zużywa zasoby radiowe dostarczane przez co najmniej dwa różne punkty sieciowe. Podwójna łączność jest jednym z potencjalnych rozwiązań 3GPP, które poprawiają wydajność użytkownika, łącząc zalety zasięgu MCell i pojemności SCell. W scenariuszu DL, w którym MCells współużytkują zasoby z innymi komórkami, gdzie proponowane jest okno CA w celu określenia, czy urządzenia UE z obsługą CA powinny zostać włączone do ośrodka CA wewnątrz ośrodka. Rozważane jest dedykowane rozmieszczenie częstotliwości. Korzyści z agregowania zasobów z obu komórek są weryfikowane dla różnych wzorców ruchu, a także dla różnych sytuacji obciążenia. Poprawa wydajności w UE przy użyciu współdzielonych zasobów stanowi wyraźną wskazówkę, że techniki współpracy stają się obowiązkowe, aby zmaksymalizować wykorzystanie zasobów i sprostać wymaganiom stawianym innowacyjnym i bardziej wymagającym aplikacjom. Kolejnym krokiem w optymalizacji HetNet jest relacja między UL i DL oraz to, w jaki sposób zasady powiązania wpływają na wydajność obu łącz. Zarówno odłączenie mocy UL / DL, jak i obciążenie MCell / SCell i brak równowagi mocy powoduje rozłączenie obu ogniw, co jest szczególnie korzystne w heterogenicznych rozmieszczeniach w kanałach międzykanałowych. Podstawowy schemat odsprężenia pomiędzy SCell i MCell pokazano na rysunku 9.3. W wydaniu 12, 3GPP dostarczył wstępną ocenę wydajności HetNet, gdy zawiera podział UL i DL, wyniki pokazują poprawę szczególnie na krawędzi komórki dla scenariuszy niskiego i średniego obciążenia. W literaturze ostatnio zajęto się problemem nierównowagi sił i ładunków, a można zidentyfikować kilka istotnych odniesień. Wyniki pod względem zysku, jaki można uzyskać na podstawie przepustowości UL, są bardzo obiecujące. Wyniki pokazują ten sam trend między analizą geometrii stochastycznej a rzeczywistymi danymi eksperymentalnymi. Wreszcie, Singh przyczynia się do tematu z analizą UL SINR i rozkładów szybkości w zależności od zasad asocjacji, biorąc pod uwagę parametry projektu kontroli mocy UL. Wyniki pokazują, że minimalne powiązanie utraty ścieżki prowadzi do identycznego rozkładu obciążenia we wszystkich komórkach, co jest również optymalne pod względem szybkości, niezależnie od parametrów sterowania mocą. Gdy maksymalne pokrycie złącza UL i DL musi być zmaksymalizowane, optymalne rozwiązanie to odłączone połączenie. Jest to korzystne, ponieważ zmniejsza nierównowagę QoS między obydwoma linkami.

### 9.2.3 Oddzielenie jako rozwiązanie

Obecne technologie, które odnoszą się do głównych wyzwań HetNet w kontekście planowania radiowego zarządzania zakłóceniami można podzielić na trzy główne grupy:

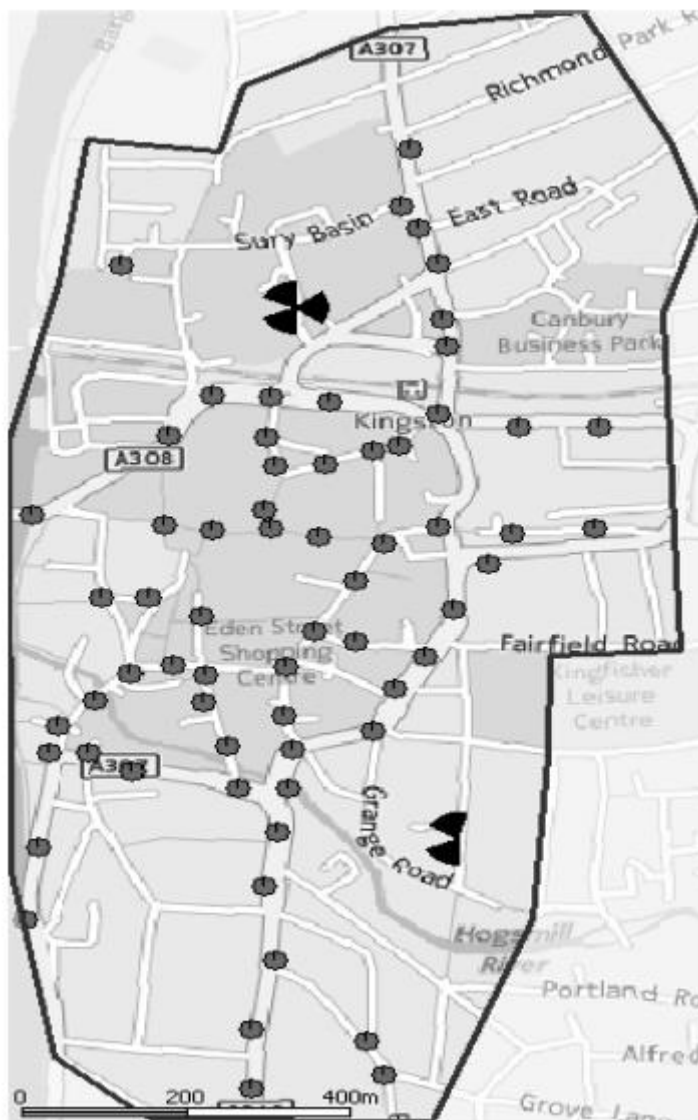
- Podwójna łączność. W literaturze zweryfikowano usprawnienia UL w dedykowanym wdrożeniu. Jednakże nie przeprowadzono jeszcze badań, które obejmowałyby podwójną łączność jako rozwiązanie problemu braku równowagi UL / DL, zasady asocjacji zaproponowane w stanie techniki są oparte na DL RSRP lub RSRQ.
- Komórka RE z funkcją EICIC. W literaturze zweryfikowano poprawę UL w rozmieszczeniach w kanałach. Strategia ta pomaga zmniejszyć nierównowagę UL / DL, podczas gdy poprawa UL jest zmaksymalizowana, zwiększa się również interferencja między komórkami DL. Dlatego technikom RE zawsze muszą towarzyszyć rozwiązania eICIC dla DL, takie jak ABS. Niemniej jednak technika RE jest ograniczona do umiarkowanych wartości przesunięcia ze względu na ostrą interferencję w DL. Reguły wyboru komórki są oparte na DL RSRP i RSRQ z dodanym offsetem RE.
- Odłączanie UL / DL. Istnieją zweryfikowane ulepszenia UL w rozmieszczeniach dwukanałowych za pomocą DUE. Odłączone zasady powiązań w UL i DL mogą zaspokoić różne wymagania obu łącz i z kolei mogą skutecznie rozwiązać problem braku równowagi UL / DL pod względem zasięgu, obciążenia i zakłóceń.



W oparciu o to porównanie oczywiste jest, że DUDe zapewnia korzyści z posiadania bardzo wysokich przesunięć RE w UL bez efektów zakłóceń w DL, ponieważ oba łącza są rozdzielone i połączone z najlepszą obsługującą komórką.

### 9.3 Ustawienia symulacji

W tej sekcji przedstawiono konfigurację symulacyjną użytą w ocenie wydajności w rozdziale 9.4. Konfiguracja symulacji wynika w dużej mierze ze scenariusza symulacji opartego na sieci testowej Vodafone LTE SCell w obszarze Londyn pokazanym na rysunku 9.4.



Sieć testowa obejmuje obszar o długości około jednego kilometra kwadratowego i składa się z dwóch sektorowych miejsc Makroekonomicznych i sześćdziesięciu czterech Gwiazd SCell zobrazowanych odpowiednio czarnymi kształtami i małymi okręgami. Dystrybucja użytkowników opiera się na danych o ruchu wyodrębnionych z sieci testów na żywo. Dodatkowo wykorzystywany jest model przewidywania ścieżki śledzenia wysokiej rozdzielczości 3D, który uwzględnia dane dotyczące bałaganu, terenu i budynku. Gwarantuje to realistyczny i dokładny model propagacji. Kontrola mocy UL jest oparta na algorytmie frakcyjnej kompensacji stratkowej w otwartej pętli, określonej przez 3GPP.

## 9.4 Ocena wydajności

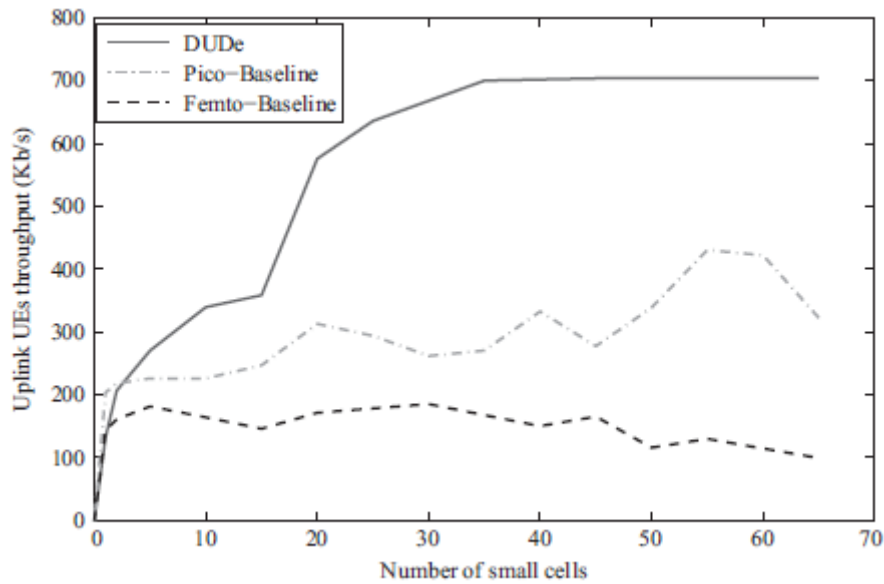
Jak wyjaśniono w Części 9.1, DUDe uważa się za ważną zmianę paradygmatu z podejścia skoncentrowanego na komórce używanego do tej pory w projektowaniu sieci komórkowych na bardziej elastyczne podejście zorientowane na urządzenie, które ma być jednym z głównych elementów składowych przyszłej sieci 5G. DUDe może zaoferować znaczne korzyści w zakresie UL pod względem zasięgu, przepustowości, niezawodności, równoważenie obciążenia i zachowanie interferencyjne. Poprawa wydajności UL stała się coraz ważniejsza ze względu na pojawienie się i wykładniczy wzrost IoT, gdzie ruch jest często skoncentrowany na UL, a także rosnąca popularność symetrycznych aplikacji ruchu, takich jak sieci społecznościowe, rozmowy wideo i gry wideo w czasie rzeczywistym. W tej sekcji wyniki symulacji, które potwierdzają korzyści pod względem wcześniej wymienionych kryteriów, zostaną przedstawione i szczegółowo omówione. Najprostsza forma DUDe jest rozważana w tej sekcji, w której powiązanie UL jest oparte na odebranej mocy UL, podczas gdy powiązanie DL jest oparte na mocy odebranej przez DL. DUDe jest porównywany z konwencjonalną operacją LTE, w której zarówno skojarzenia UL, jak i DL są sprzężone i oparte na mocy odbieranej przez DL. Rozważa się dwa przypadki bazowe LTE, w których SCell traktowane są jako komórki Pico i femtokomórki, a te dwa przypadki są określane odpowiednio jako Pico-Baseline i Femto-Baseline. Moc nadawania makr, Pico, femto wynosi odpowiednio 46, 30, 20 dBm. Następnie zostanie przedstawiony zestaw wyników ilustrujący wzrost wydajności DUDe.

### 9.4.1 Pokrycie i pojemność

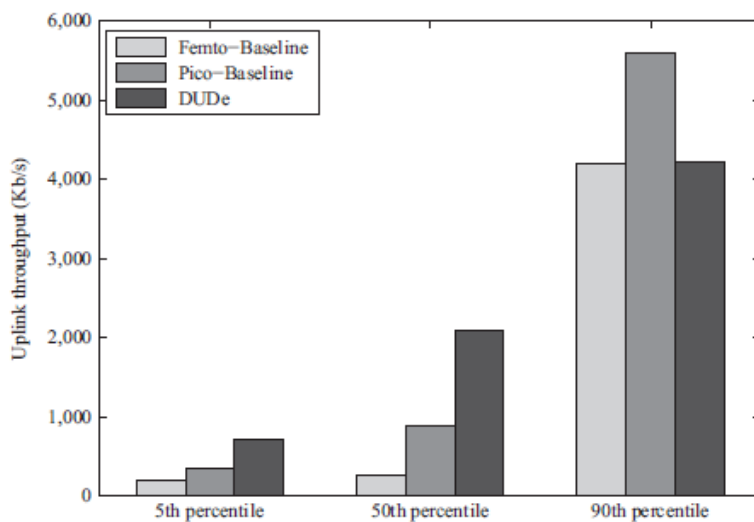
W sieci HetNet pokrycia UL i DL są całkiem różne, jak to omówiono w Części 9.2 i pokazano na rysunku 9.2, dlatego oparcie połączeń UL i DL na tym samym kryterium, które jest mocą odbieraną przez DL, jest wysoce nieoptymalne. W tej części zilustrowano ogromną różnicę między pokryciem UL w przypadkach bazowych LTE a DUDe, a wynikające z nich przyrosty wydajności są podświetlone. Rysunek 9.5 ilustruje pokrycie UL warstw Mcell i SCell dla trzech przypadków w porównaniu, w których pokrycie Mcell i SCell są wyświetlane odpowiednio w kolorze czarnym i szarym.



Wykazano, że zasięg UL Scells jest bardzo mały w przypadkach Femto-Baseline i Pico-Baseline. Jednak w przypadku DUDe Scells mają znacznie większy zasięg UL, który jest znacznie mniej zdominowany przez Mcells niż w przypadkach bazowych LTE. Efekt ten powoduje bardziej równomierną dystrybucję UE między węzłami, co z kolei prowadzi do znacznie bardziej efektywnego wykorzystania zasobów, co zostanie wykazane w następujących wynikach. Jednorodna pokrycia ma duży wpływ na przepustowość, a zwłaszcza przepustowość UE na krawędzi komórek, co jest reprezentowane przez przepustowość piątego percentyla. Rysunek 9.6 pokazuje wpływ zwiększenia liczby Scellów w symulowanym scenariuszu na przepustowość piątego percentyla.



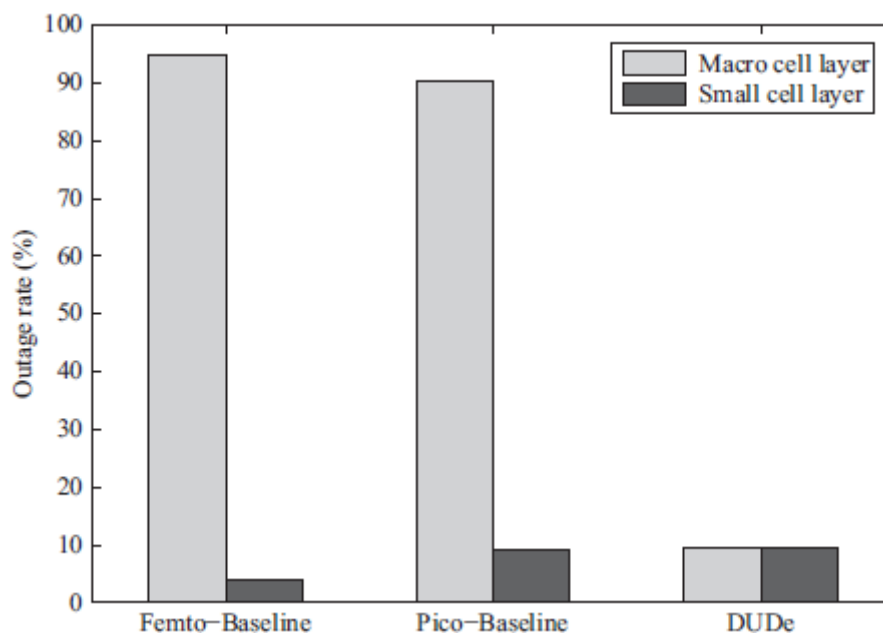
W przypadku DUDe można zauważyć, że przepustowość piątego centyla poprawia się szybko wraz z liczbą Scellów, jest to częściowo spowodowane zwiększonym zasięgiem UL Scells w DUDe, jak pokazano na rysunku 9.5. Gdy liczba Scellów wzrasta, przepustowość piątego percentyla UE zaczyna nasycać się, gdy te urządzenia użytkownika stają się ograniczone przez jakość kanału i moc nadawania, a zyski zaczynają być bardziej wyraźne w 50. i 90. percentylu. Patrząc na przypadki Femto-Baseline i Pico-Baseline, można zauważyć, że dodanie Scells ma niewielki wpływ na piątą percentylową przepustowość ze względu na bardzo ograniczony zasięg Scells w obu przypadkach, co czyni je bardziej efektywnymi dla 50. i 90. percentyla UEs. Ponadto w tych dwóch przypadkach przepustowość zmienia się wraz ze wzrostem liczby Scellów. Jest to spowodowane wysoką interferencją urządzenia z efektami Scells UE tworzą krawędzie UE komórek Mcell, ponieważ te UE są bliżej Scellów, więc cierpią na wysoki poziom interferencji. Efekt ten jest bardziej podkreślony w przypadku Femto-Baseline, ponieważ przepustowość piątego percentyla zaczyna maleć po pewnym punkcie, podczas gdy w przypadku DUDe przepustowość wzrasta w bardziej stabilny sposób, ponieważ UE zawsze łączą się z węzłem, do którego mają najlepsze UL kanał, który gwarantuje niższy poziom zakłóceń. Ten wpływ na interferencję zostanie omówiony w dalszej części. Rysunek 9.7 pokazuje wyniki przepustowości 5, 50 i 90 percentyla dla tych trzech przypadków.



piąty percentyl UL przepustowości w DUDe jest zwiększony o więcej niż 200% i 100% w porównaniu do odpowiednio linii bazowej Femto i Pico-Baseline. Podobnie, wydajność 50. percentyla jest poprawiona w przypadku DUDe o 600% i 100% w porównaniu do linii bazowej Femto i linii bazowej Pico. Zyski są wynikiem lepszego pokrycia Scells w przypadku DUDe, co skutkuje lepszą dystrybucją UE między węzłami, co daje bardziej efektywne wykorzystanie zasobów. Ponadto fakt, że UE łączy się z węzłem, do którego mają najlepszy kanał UL, skutkuje poprawą stosunku sygnału do zakłóceń i szumów (SINR), który pozwala UE na stosowanie wyższego schematu modulacji i kodowania, i z kolei uzyskanie lepszego wykorzystanie zasobów i większa przepustowość. Jeśli chodzi o przepustowość na poziomie 90. centymetra, DUDe nie osiąga najwyższej przepustowości, co ma sens, ponieważ DUDe ma na celu poprawę efektu równoważenia obciążenia, który w naturalny sposób spowoduje zmniejszenie szczytowej szybkości przesyłania danych. Dodatkowo można zauważyć, że Pico-Baseline osiąga najwyższą przepustowość na poziomie 90. percentyla, co wydaje się sprzeczne z intuicją, ponieważ można by oczekiwać, że linia bazowa femto byłaby tą, która osiąga najwyższą szybkość transmisji danych szczytowych. Jednakże, patrząc na 98. percentyl przepustowości, przepustowość osiąga 15 Mb / s i 10 Mb / s odpowiednio w przypadkach Femto-Baseline i Pico-Baseline. Pokazuje to, że wpływ Scells w przypadku Femto-Baseline jest ograniczony do małej liczby UE.

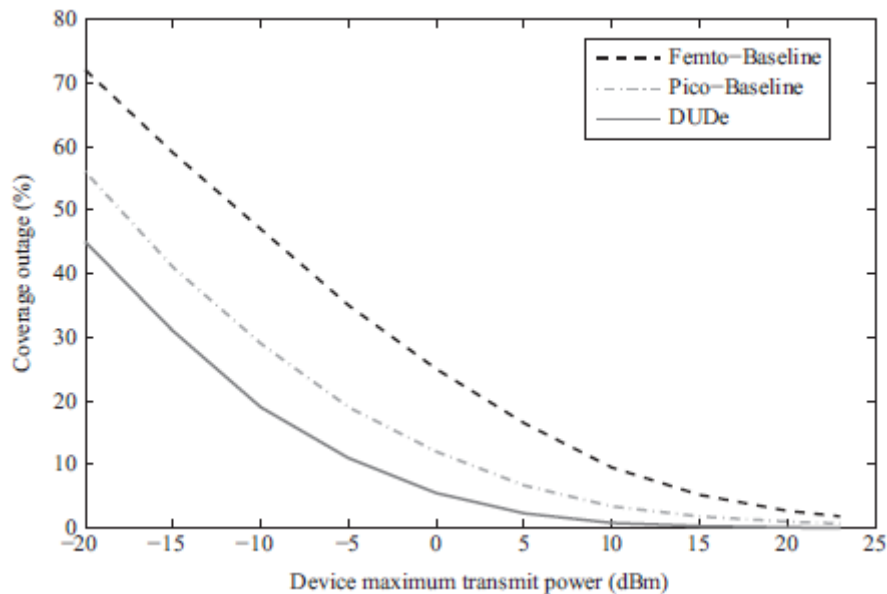
#### 9.4.2 Niezawodność i równoważenie obciążenia

Niezawodność staje się jednym z najważniejszych wymagań przyszłych systemów komórkowych ze względu na rozprzestrzenianie się Internetu Rzeczy, która jest parasolem dla wielu aplikacji, takich jak sieci czujników, automatyzacja fabryki i wiele innych. Na niezawodność systemów bezprzewodowych może mieć wpływ kilka czynników, w tym przeciążenie i utrata zasięgu, w których urządzenie mogło zostać odłączone z powodu braku dostępnych zasobów lub z powodu bardzo niskiej jakości łącza, wynikającej z braku zasięgu lub podlegania wysoki poziom zakłóceń. DUDe odgrywa kluczową rolę w poprawianiu niezawodności UL poprzez redukcję wskaźnika awarii, co zostanie pokazane poniżej. Rysunek 9.8 ilustruje średnią awaryjność dla warstw Mcell i Scell dla trzech przypadków, w których symulowany jest scenariusz o dużym natężeniu ruchu poprzez ustawienie minimalnej żądanej przepustowości dla każdego UE na 1 Mb / s.

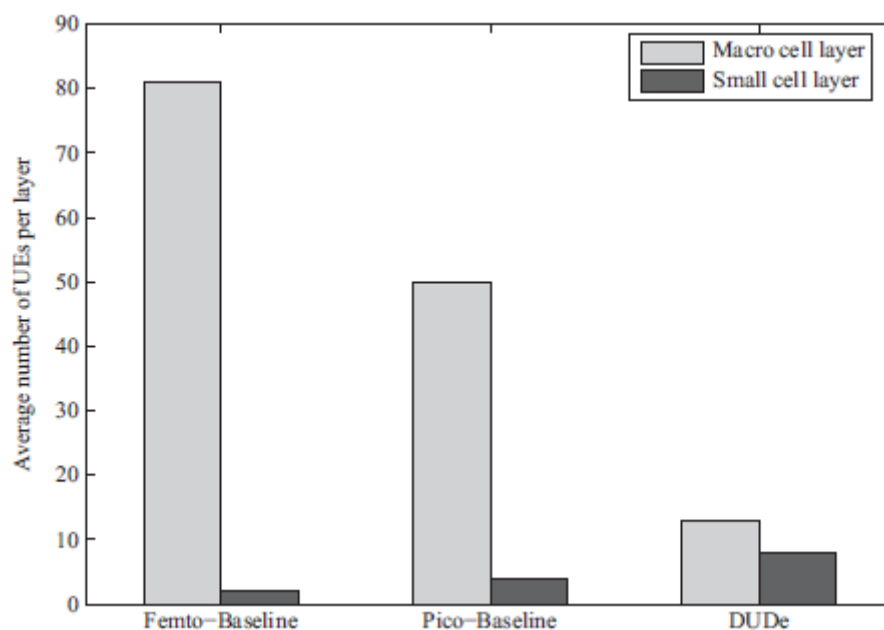


Przerwa definiowana jest jako procent urządzeń UE, które nie mogą osiągnąć przepustowości minimalnej 1 Mb / s. Ponieważ scenariusz ten uważa się za scenariusz dużego natężenia ruchu, wymaga

on bardzo wydajnego wykorzystania zasobów w celu spełnienia wymagań UE o wysokiej przepustowości. Jak można zauważyć na rysunku, warstwa Macro ma bardzo wysoką szybkość wyłączenia w przypadkach bazowych LTE, co tłumaczy się tym, że warstwa Macro jest bardzo przeciążona w UL i MCells nie mogą obsługiwać wszystkich UE z wymaganą przepustowością co skutkuje dużą liczbą porzuconych UE. Jednakże, w przypadku DUDe, urządzenia UE są rozmieszczone bardziej równomiernie wśród węzłów, tak więc szybkość wyłączenia może być drastycznie zmniejszona do mniej niż 10% na warstwach MCell i SCell. Innym trendem, który wpływa na niezawodność, jest przerwa spowodowana utratą zasięgu lub słabymi kanałami. Rysunek 9.9 pokazuje wskaźnik przestoju w stosunku do maksymalnej mocy transmisji UE.



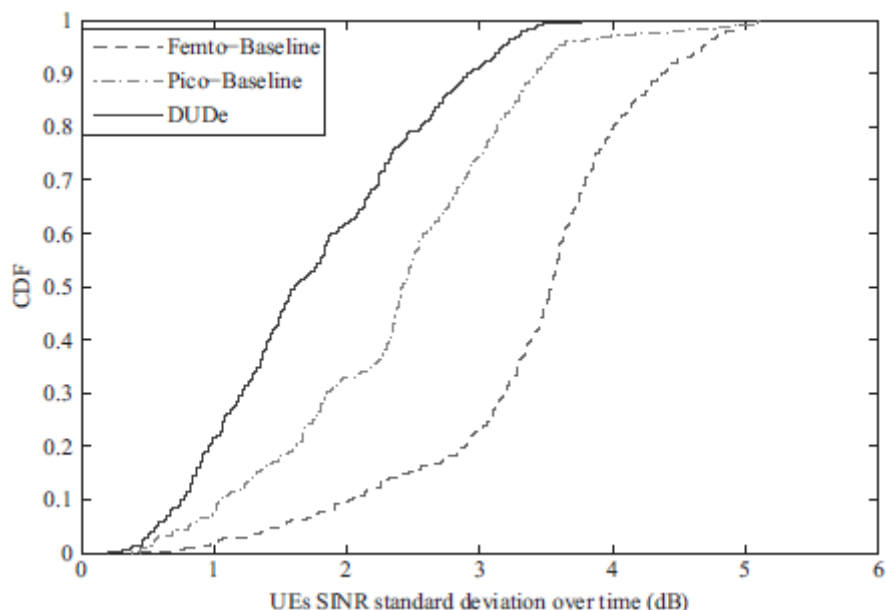
Przerwa w tym miejscu jest określona przez ułamek UE, którego jakość sygnału UL jest niższa niż wymagana do uzyskania najniższego modulacji i schematu kodowania (MCS) LTE, a zatem jest odrzucana. Wynik ten pokazuje, że dla tej samej maksymalnej mocy nadawczej UE DUDe może zmniejszyć częstotliwość wyłączenia UL nawet o 25% w porównaniu do linii bazowej LTE. Alternatywnie, przy tej samej szybkości wyłączenia, maksymalna moc nadawania może zostać zmniejszona o maksymalnie 10 dB za pomocą DUDe. Te wyniki są bardzo istotne dla aplikacji IoT, w których albo niezawodność łącza, albo żywotność baterii albo oba mają ogromne znaczenie i, jak widać na rysunku, można je drastycznie poprawić za pomocą DUDe. Obciążenie na danym BS w UL może być inne niż obciążenie w DL na tym samym BS. Oznacza to, że nie jest optymalne, jeśli chodzi o równoważenie obciążenia, aby ten sam zestaw urządzeń UE był połączony z tym samym BS w UL i DL. Ponadto, DUDe zostało pokazane w poprzednich wynikach, aby poprawić zasięg UL dla SCell, co skutkuje znacznie lepszą dystrybucją UE wśród różnych warstw sieci komórkowej. Ten efekt równoważenia obciążenia jest przedstawiony na rysunku 9.10, gdzie średnia liczba urządzeń UE na komórkę w każdej warstwie jest pokazana dla trzech porównywanych przypadków.



DUDe skutkuje znacznie bardziej równomiernym rozkładem UE między różnymi typami węzłów, co bezpośrednio przekłada się na radykalnie poprawioną wydajność widmową i redukcję przestojów, jak pokazano w poprzednich wynikach.

### 9.4.3 Zachowanie interferencyjne

Zakłócenia są jednym z głównych czynników ograniczających w sieciach komórkowych. DUDe ma zdolność do zmniejszenia zakłóceń UL z powodu wielu uzupełniających się efektów. Po pierwsze i w wyniku połączenia UL, które ma tendencję do łączenia UE w UL z ich najbliższym węzłem lub węzłem, do którego mają najlepszą moc odbieraną UL, może to być tłumaczone na zmniejszenie mocy transmisji UE, jak pokazano na rysunku 9.9. Ma to wpływ na zmniejszenie zakłóceń UL dla innych stacji bazowych, co jest dość znaczące, zwłaszcza dla urządzeń UE o niskim SINR UL. Po drugie, DUDe zapewnia możliwość niezależnego wyboru powiązania, które minimalizuje zakłócenia w UE, jak również BS. Interferencja UL zależy od położenia UE względem zakłócającego BS, zastosowanego algorytmu regulacji mocy UL i ciężaru wstępnego kodowania UL. Przeciwnie, zakłócenie DL w kierunku pewnego UE zależy od mocy nadawania BS, wiązki sygnału DL i odległości od różnych stacji BS. Ponadto, UL i DL mają różne zachowania związane z ruchem i szeregowaniem, które można uznać za prawie niezależne. Ze wszystkich powyższych powodów, odsprężone powiązanie UL i DL, które pozwala UE lub BS szukać najlepszego połączenia UL lub DL, które zmniejsza zakłócenia na odpowiednim łączu, ma wiele sensu i oczekuje się, że przewyższy konwencjonalne sprężone skojarzenie. Oprócz zmniejszenia średniej interferencji, DUDe pozwala na zmniejszenie odchylenia standardowego SINR UL UE w czasie, jak pokazano na rysunku 9.11.



Zmniejszenie wariancji SINR oznacza, że kanał jest bardziej stabilny i przewidywalny, co ma znaczący wpływ na zmniejszenie złożoności funkcji zarządzania zasobami radiowymi (RRM) i samoorganizującej się sieci (SON).

## 9.5 Interoperacyjność DUDe z funkcjami 4G i 5G

Wdrożenie niezawodnej, wydajnej sieci o wysokiej prędkości spektralnej wymaga włączenia szeregu innowacyjnych funkcji, pod warunkiem, że rozwiązania na poziomie łącza ewoluowały w pobliżu mocy ograniczonej Shannon dzięki zaawansowanym programom modulacji i kodowania (MCS). Biorąc to pod uwagę, 4G i nadchodzące 5G muszą oferować pionierskie rozwiązania lub ulepszone wersje wcześniejszych wersji w zarządzaniu spektrum i komunikacji kooperatywnej. W szczególności interoperacyjność DUDe z innymi technologiami dostępu radiowego może doprowadzić do ogólnej poprawy tych funkcji. Zmniejszona zmienność interferencji, zwiększona elastyczność sieci lub zmniejszona moc nadawania to niektóre z zalet, które pomagają w pełni wykorzystać techniki zarządzania radiem.

### 9.5.1 Międzygrupowa agregacja nośna

CA zapewnia wielkie korzyści, gdy jest przeprowadzany w HetNets. W szczególności uwzględnienie CA w takim kontekście uznano za możliwy do zrealizowania sposób wielostanowiskowej funkcji alokacji zasobów radiowych, która została po raz pierwszy wprowadzona w wydaniu 11. Ponadto CA ogólnie (wiele lub ta sama agregacja lokacji) umożliwia zwiększenie przepustowości poprzez zwiększenie dostępnej przepustowości i obsługuje techniki zarządzania mobilnością i zakłóceniami. Kilka badań skupiło się szczególnie na implementacji CA w UL, gdzie możliwości energetyczne UE stanowią najbardziej ograniczające ograniczenie. Zauważono, że potencjalne zyski transmisji CA są silnie powiązane z wymaganą mocą, co w dużej mierze zależy od alokacji przepustowości i ścieżkę dostępu do UE. Powiązanie komórki określa, który eNB obsługuje UE. Węzeł eNB znajdujący się bliżej UE oznacza, że potrzeba mniej mocy do transmisji danych w przydzielonej przepustowości. Aby zapewnić poprawne działanie zagregowanych transmisji szerokości pasma w UL, istotne jest uwzględnienie maksymalnej mocy transmisji UE. W wdrożeniach tylko dla MCell, urządzenia UE krawędzi komórki są mniej prawdopodobne, aby transmitować w CA; jednak w heterogenicznych wdrożeniach odległość do węzła eNB jest krótsza, biorąc pod uwagę większą gęstość komórek. W takim kontekście, jeżeli UE są

powiązane w oparciu o DL RSRP, transmisja UL CA będzie wysoce ograniczona. Strategie odsprężania są bardziej łagodne w przyjmowaniu zagregowanych transmisji, głównie ze względu na poprawę dostępności energii UL przez stowarzyszenia niezwiązane z wielkością produkcji. Jest to ważne, ponieważ CA ma być stosowany zarówno w UL jak i DL, a przy tradycyjnych regułach mocy skojarzonej DL, zastosowanie CA w UL jest poważnie uwarunkowane brakiem dostępności mocy.

### **9.5.2 Kooperatywa wielopunktowa**

Współpraca stacji bazowej, w postaci transmisji Coordinated Multi-Point (CoMP) lub odbiór zyskał popularność w kontekście HetNets jako środek do zwiększenia osiągalnej przepustowości UE. Węzły eNB w tym samym klastrze komunikują się za pośrednictwem łączy dosyłowych (to znaczy za pośrednictwem interfejsu X2) w celu minimalizacji zakłóceń międzykomórkowych i wykorzystania zalet rozproszonych systemów antenowych. W rzeczywistości interwencja w ramach klastra współpracy może zostać skutecznie anulowana. Ten poziom koordynacji i współpracy może być realizowany zarówno w UL jak i DL, a realizacja takiej koordynacji silnie zależy od dostępności wystarczającej przepustowości łącza dosyłowego, najpierw do obsługi UE w klastrze komórkowym, a po drugie do komunikacji z innymi komórkami w klastrze współpracy. Ta zależność wsteczna może być bardzo ograniczona w sytuacjach dużego obciążenia i ograniczonych przepustowości łącza. Zwiększona elastyczność zapewniana przez odłączone powiązania UL i DL zapewnia korzyści przy wyborze skoordynowanych transmisji UL lub DL. W szczególności, nie ma potrzeby jednoczesnego połączenia UL i DL z całym współpracującym zestawem stacji bazowych, a UE może mieć nierówne łącza aktywne UL i DL (jak w przypadku CA). To elastyczne powiązanie wewnątrz klastra i współdziałanie DUDe z CoMP idzie o krok dalej w sieci zorientowanej na urządzenie, ponieważ UE może niezależnie wybrać liczbę i pozycję komórek obsługujących DL i UL, zgodnie z kilkoma parametrami wejściowymi, jako backhaul. pojemność, ograniczenie mocy, maksymalizacja przepustowości, między innymi.

### **9.5.3 Fala milimetrowa**

Rosnące zapotrzebowanie ruchu sieciowego doprowadziło do kilku trendów w sieciach komórkowych, w tym do zagęszczenia sieci i pojawienia się heterogenicznych sieci, w których SCell odgrywają kluczową rolę w zapewnieniu przepustowości w hotspotach i gęstych obszarach miejskich. Innym podejściem było zbadanie nowych nieużywanych pasm częstotliwości, które pomogłyby zaspokoić prognozowane przyszłe potrzeby w zakresie ruchu. Popularne częstotliwości sub-6 GHz stają się coraz rzadsze i nie będą już w stanie radzić sobie z rosnącym ruchem sieciowym. Rozwiązaniem tego problemu jest uciekanie się do wyższych częstotliwości w pasmach fal milimetrowych (20 - 100 GHz), w których znaczna część widma jest niewykorzystana lub całkowicie niewykorzystana. Kilka GHz dostępnego widma obiecuje spektakularny wzrost wydajności, który kwalifikuje technologie fal milimetrowych jako jeden z głównych elementów przyszłych sieci 5G. W rzeczywistości pasma fal milimetrowych nie zastąpią pasm sub-6 GHz, przynajmniej początkowo, gdzie sub-6 GHz będzie nadal potrzebne do zapewnienia zasięgu i wszechobecnej obsługi, ponieważ częstotliwości fal milimetrowych mają słabe właściwości propagacyjne i są dość wrażliwe na blokady. Istnienie sieci fal milimetrowych jako nakładki na tradycyjne sieci sub-6 GHz wymagałoby zmiany w konwencjonalnych technikach łączenia komórek. DUDe ma odegrać kluczową rolę w sieciach fal milimetrowych z kilku powodów. Ostatnie badania dotyczące ekspozycji na pole elektromagnetyczne wykazały, że aby być w zgodzie z obowiązkowymi granicami ekspozycji na częstotliwościach powyżej 6 GHz, maksymalna moc nadawania urządzeń musiałaby zostać zmniejszona o kilka dB poniżej konwencjonalnych poziomów mocy w obecnych sieciach komórkowych. Ma to znaczący wpływ na UL; ponieważ zasięg UL zależy od mocy nadawczej urządzeń, a ponieważ mmWaves mają niekorzystne właściwości propagacyjne, zmniejszenie mocy transmisji urządzeń może spowodować poważne pogorszenie jakości łącza w UL. Możliwym rozwiązaniem tego problemu byłoby odseparowanie UL i DL, gdzie dla niektórych UE UL



będzie obsługiwany przez MCH UHF o lepszym budżecie łącza, podczas gdy DL będzie obsługiwany z SCells mmWave. To znaczy, podczas gdy poprzednio DUDe omawiano w kontekście łączenia UE z MCells w DL i SCells w UL, dla mWaves może okazać się użyta przeciwna strategia. Ponadto oczekuje się, że ogniwa mmWave mają bardzo ograniczony obszar pokrycia, w związku z czym konwencjonalne techniki łączenia oparte na mocy odbieranej przez DL pozostawiłyby ogniwa mmWave w niewielkim stopniu niewykorzystywane, biorąc pod uwagę ogromną ilość widma dostępnego w pasmach fal milimetrowych. Rozgałęzienie SCC lub rozszerzenie zakresu to technika wprowadzona w wydaniu 10 LTE, w której skutecznie stosuje się do SCells polaryzację selekcji komórek w celu zwiększenia zasięgu SCell i poprawy ich efektu rozładowania. Podobne podejście można zastosować w przypadku fal milimetrowych, aby rozwiązać wspomniany powyżej problem, stosując agresywne wartości rozszerzania zakresu w celu przyciągnięcia jak największej liczby UE, w celu wykorzystania dużych fragmentów widmowych dostępnych na częstotliwościach fal milimetrowych. DUDe będzie odgrywać ważną rolę w tej konfiguracji, ponieważ pozwoli ustawić różne wartości rozszerzenia zasięgu dla UL i DL, aby spełnić wymagania obu łączy. Na przykład, jeśli UL musi być wysoce niezawodny, podczas gdy DL może być mniej niezawodny, ale bardziej wymagający pod względem przepustowości, można zastosować wysoką wartość rozszerzenia DL SCell, nie stosując żadnego zakresu rozszerzenia w UL. Dlatego DUDe oferuje elastyczność, aby zaspokoić różne potrzeby UL i DL w scenariuszach fal milimetrowych.

#### **9.5.4 Różne techniki duplexowania**

DUDe może działać zarówno z FDD jak i TDD, z różnymi implikacjami od a poziom systemu i punkt widzenia widma. TDD zapewnia znacznie większą elastyczność w handlu zasobami DL i UL w porównaniu do FDD. W przypadku DUDe wykazano, że potrzeba mniej zasobów UL, aby uzyskać tę samą wartość UL w porównaniu z operacją sprzężoną. Może to prowadzić do tego, że DUDe ma pozytywny wpływ na wskaźnik DL, umożliwiając DL korzystanie z większej ilości zasobów za pośrednictwem dynamicznego TDD. Kolejną zaletą TDD jest możliwość estymacji kanału DL za pomocą sygnałów referencyjnych UL. Jest to bardzo ważne szczególnie dla kanałów o dużej wymiarowości, jak w przypadku masywnego MIMO. Jednakże, gdy stosuje się DUDe, ta wzajemność jest przerywana, ponieważ transmisje DL i UL pochodzą i kończą się na różnych BS. Znaczna część istniejącego widma to sparowane widmo FDD, dlatego z tych dwóch powodów masywny MIMO może wymagać wsparcia bez wzajemności kanału. W perspektywie średnio- i długoterminowej DUDe wraz z różnymi nowymi trendami technologicznymi może doprowadzić do ponownego przemyślenia tradycyjnych paradygmatów FDD / TDD. DUDe, hiper-zagęszczanie, wykorzystanie częstotliwości fal milimetrowych i wysoce kierunkowych anten, może pozwolić na dwustronne podejście do dziedziny przestrzeni. Na przykład, to samo pasmo może być używane dla dwóch różnych urządzeń zlokalizowanych w różnych lokalizacjach, jednego odbierania w DL od pewnego BS, a drugiego transmitującego w UL do innego BS. Przestrzenny skoordynowany mechanizm szeregowania UL / DL skutecznie umożliwiłby uzyskanie pełnego duplexu, bez skomplikowanych mechanizmów eliminacji interferencji w trybie pełnego duplexu. Ponadto, gdy analogowe / cyfrowe mechanizmy interferencyjne są realizowane w celu obsługi pełnego czasowego duplexu, DUDe może być korzystny, ponieważ umożliwia uogólniony odłączony dostęp, który umożliwiłby obsługę DL i niekoniecznie tego samego użytkownika UL w tym samym paśmie częstotliwości.

#### **9.6 Włączanie architektury sieci dostępu radiowego do oddzielenia**

Konfiguracja sieci dostępowej wiąże się z kilkoma problemami, umożliwiającymi korzystanie z odłączonych połączeń, przy zachowaniu zakłóceń i efektywności energetycznej na rozsądnych poziomach. Trzymanie więcej niż jednego połączenia UL jest mniej energooszczędne dla użytkowników, którzy są umieszczeni w pobliżu krawędzi komórki, które są bardziej prawdopodobne,

aby oddzielić; również użycie jednego nośnika do przesyłania wyłącznie sygnałów sterujących doprowadzi do słabej wydajności widmowej. W szczególności, najtrudniejszą częścią opracowywania architektur wspierających DUDe jest to, że zwiększona ilość informacji kontrolnych musi być zasygnalizowana z powrotem do odpowiedniej obsługującej komórki. 3GPP zaproponował kilka architektonicznych alternatyw dla podwójnej łączności DL, a architektura wymagana do obsługi agregacji U-Plane z różnych eNB powinna być bardzo podobna do architektury podwójnej łączności, w oparciu o koncepcję podziału na nośnik. Alternatywnie, te architektury, które umożliwiają całkowite odłączenie UL i DL, powinny wspierać możliwą współpracę między obydwojoma obsługującymi komórkami, nie zagrażając jednocześnie poprawie UL pod względem niezawodności i wydajności. Aby to osiągnąć, konieczne jest zapewnienie dostarczania sygnałów kontrolnych warstwy 1 i 2 przy zachowaniu opóźnienia RAN .

### 9.6.1 Procedury sieciowe

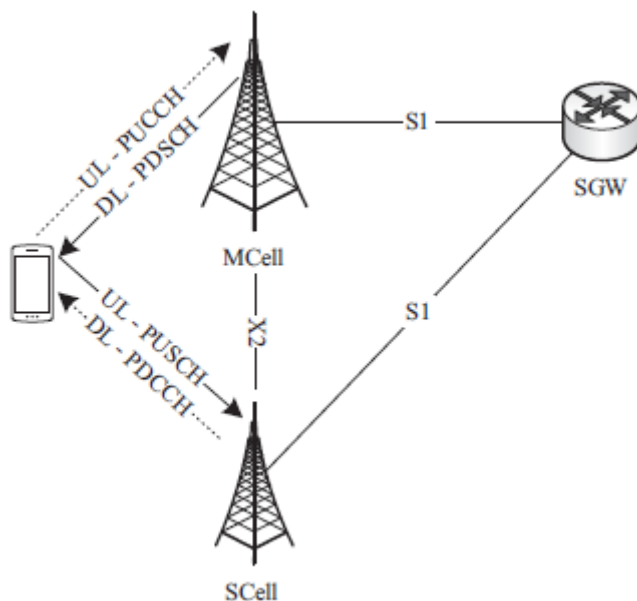
Gdy UE uzyskuje dostęp do systemu i łączy się z eNB, kontrola zasobów radiowych połączenie jest skonfigurowane. RRC jest protokołem, który obsługuje całą sygnalizację płaszczyzny sterowania warstwy 3 pomiędzy UE i EUTRAN. Wśród różnych zadań, RRC odpowiada za konfigurację połączenia, ustalanie i uwalnianie nośnika radiowego, procedury mobilności i konfigurację mechanizmów kontroli mocy. Ilość zasobów, które RRC zużywa dynamicznie zmienia się w zależności od typu ruchu komórkowego, liczby obsługiwanych UE i liczników czasu połączenia. Krótkie wartości licznika czasu generują wysoki koszt sygnalizacji RRC, podczas gdy wartości wysokokanałowe powodują złe korzystanie z trybu bezczynności. Kanały kontrolne warstwy 1 i warstwy 2 są kluczowe dla obsługi transmisji danych płaszczyzny użytkownika. Kanały kontroli fizycznej PDCCH i PUCCH są odpowiedzialne za dostarczanie informacji związanych z aktywną transmisją, jako decyzje planowania lub potwierdzenia przesyłanej informacji, oraz informacje o stanie kanału w celu dokonania dokładnej adaptacji łącza. W szczególności, sygnały sterujące RAN, które mają ogromne znaczenie dla obsługi planowania i innych procedur warstwy MAC, to:

- informacje sterujące przesyłane w PUCCH: DL CQI, raporty o stanie bufora, żądania planowania i raporty o zapasach mocy;
- informacje sterujące w łączu w dół (DCI) przesyłane w PDCCH odpowiedzialne za wskazywanie, między innymi, fizycznych bloków zasobów UL i DL transmisji (UL-SCH i DL-SCH), jak również formularzy adaptacyjnych łącza i mocy transmisji dla UL ;
- komunikaty potwierdzające HARQ w łączu w dół przesyłane w kanale wskaźnika Physical Hybrid-ARQ (PHICH); i
- Komunikaty RRC, które konfiguruje połączenie i zwolnienie UE, jak również
- Pozycja i zasoby PUCCH oraz SRS dla informacji o stanie konfiguracji kanału UL.

Wśród wszystkich tych informacji kontrolnych, najbardziej rygorystyczna pod względem opóźnienia wymaganiem jest HARQ RTT, który wynosi około 8 ms, biorąc pod uwagę informacje przetwarzania po obu stronach i transmisja ramek. HARQ w UL następuje po procesie synchronicznym: z periodycznością 8 podramek wykorzystywany jest ten sam proces identyfikacyjny HARQ i nie są wymieniane jawne informacje o identyfikatorze procesu. Oznacza to, że istnieje ścisła zależność między numerem podramki i identyfikatorem procesu HARQ, co powoduje znaczne opóźnienia, jeśli jedno z potwierdzeń nie może być wysłane w odpowiedniej podramce. Architektura, która umożliwia korzystanie z DUDe, będzie tą, która minimalizuje opóźnienie w procesie synchronizacji potwierdzenia i zapewnia utrzymanie RAN RTT.

### 9.6.2 DUDe z połączeniami pomocniczymi

Zastosowanie wspierających połączeń UL i DL w celu przesyłania sygnałów sterujących jest natychmiastowym rozwiązaniem umożliwiającym wdrożenie DUDe. Terminal jest podłączony do obu komórek i agreguje przepływy danych; w ten sposób cała warstwa sygnalizacyjna 1 / warstwa 2 i RRC może być obsługiwana lokalnie. Prosty sposób na obsługę tego jest użycie CA, gdzie każdy komponent nośny jest skonfigurowany osobno do przenoszenia współdzielonego i kontrolnego kanału. Ta konfiguracja jest pokazana na rysunku 9.12.



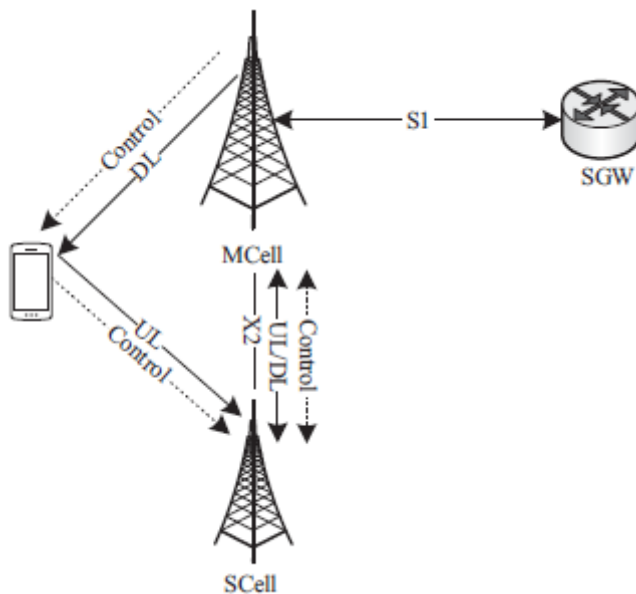
To rozwiązanie pozwala utrzymać opóźnienie RAN na pożądanym poziomie, ponieważ nie wprowadza się sensownego opóźnienia. Jednak potencjalnymi wadami tej konfiguracji są: po pierwsze, posiadanie dwóch równoczesnych połączeń UL może poważnie zagrozić wydajności UL pod względem efektywności energetycznej w UE. Po drugie, ta konfiguracja nie maksymalizuje pojemności w dostępnym widmie, ponieważ nośnik jednego komponentu jest wykorzystywany wyłącznie do obsługi informacji sterujących.

### 9.6.3 Podział na użytkownika płaszczyzny dla DUDe

Najnowsze wydania LTE-A (wydania 12 i 13) rozważają nowe alternatywy dla architektury dla podwójnej łączności, wraz z wprowadzeniem nowej koncepcji podziału nośnika, alternatywy, która ułatwia posiadanie dwóch jednoczesnych transmisji w różnych eNB, na przykład MCell i SCell [ 4]. Z drugiej strony, możliwość UL przyjęcia podziału na okaziciela można argumentować w kategoriach zużycia energii, a dane UL powinny być albo przekazywane bezpośrednio do MCell, albo przesyłane do MCell przez SCell. W związku z tym, w kontekście DUDe, pomoc w połączeniach PUSCH lub PUCCH nie może być realizowana w MCell; a połączenia PDSCH i PDCCH nie mogą być wykonywane w SCell. Brak informacji o sterowaniu sygnalizowanych z powrotem do odpowiedniej komórki obsługującej za pośrednictwem interfejsu użytkownika (tj. Z wykorzystaniem obsługujących połączeń UL i DL), dostarczanie sygnalizacji warstwy 1 i warstwy 2 i RRC opiera się na nieidealnym połączeniu dosyłowym, interfejsie X2, między obydwojema obsługującymi komórkami. Zgodnie z architekturą LTE rozproszonych komórek i wykorzystaniem koncepcji podziału na okaziciela powstają dwa możliwe rozwiązania w celu rozłączenia:

1. Nośnik radiowy jest zarządzany w bramce obsługującej (SGW), a DL przepływa przez S1 z SGW do MCell, a UL przepływa przez S1 do SGW z SCell. Ta opcja zmniejszyła kontrolę przepływu między obsługującymi eNB, ponieważ tylko część sygnałów sterujących, na przykład potwierdzenia HARQ, musi być przekazywana przez X2. Architektura w dużej mierze jest zgodna z tą na rysunku 9.12, gdzie przepływ informacji dla DL i UL przebiega przez interfejs S1.

2. Nośnik radiowy jest zarządzany w MCell, więc konfiguracja jest MAC urządzenia master / slave dla sprzężenia zwrotnego sterowania UL i DL. Oznacza to, że jednostki MAC PDU w czasie rzeczywistym muszą być przekazywane do odpowiedniego eNB poprzez interfejs backhaul X2, z zachowaniem wymogu czasu w obie strony HARQ 8 ms. Konfiguracja master / slave dotyczy UL i DL, a przetwarzanie każdej jednostki MAC PDU odbywa się w odpowiedniej komórce. Rysunek 9.13 pokazuje schemat przepływu informacji dla DL i UL.



Obie konfiguracje pozwalają zmaksymalizować wykorzystanie widma, ponieważ wszystkie nośniki mogą być skonfigurowane do przesyłania informacji U-Plane. Obecne heterogeniczne sieci stanowią duże wyzwanie, aby sprostać RAN RTT ustalonemu w procesie HARQ, ponieważ spodziewane są opóźnienia w zakresie od 5 ms do 30 ms w interfejsie X2. Podobnie jak w przypadku płaszczyzny użytkownika, płaszczyzny kontrolnej, która umożliwia pełny DUDe jest analizowany na podstawie architektury protokołu RRC przedstawionej dla podwójnej łączności. Przedstawiono dwie alternatywy:

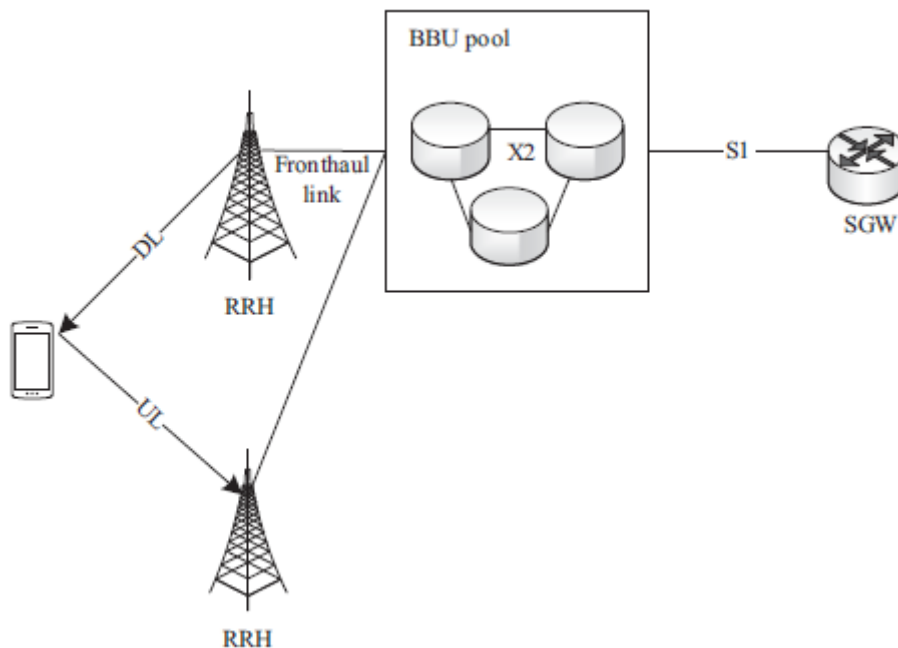
- Na podstawie wcześniejszych negocjacji parametrów pomiędzy SCell i MCell MCell generuje końcowy komunikat RRC. Ta alternatywa jest dobra, ponieważ tylko jedno połączenie RRC jest aktywne, ale wymaga wspomagającego połączenia UL z MCell.
- W tym przypadku komunikat RRC jest budowany przez SCell na podstawie parametrów wejściowych dostarczonych przez MCell. To rozwiązanie zwiększa złożoność po stronie UE, ponieważ wymaga obsługi równoległych procedur RRC

W świetle tego, całkowite oddzielenie UL i DL stanowi ogromny wysiłek w połączeniu RRC.

#### 9.6.4 Rozwiązanie scentralizowane

Poprzednie rozwiązania architektoniczne przedstawiają ograniczenia pod względem wykorzystania widma, wydajności energetycznej i wymagań dotyczących czasu oczekiwania. Jeśli oba eNB, MCell i

SCell, mogą współdzielić tę samą jednostkę BBU, wówczas można zakończyć obsługę DUDe i zminimalizować większość dotychczasowych jej wad. Koncepcja pod wspólnym udostępnianiem BBU jest scentralizowaną architekturą, znaną również jako C-RAN, która przerywa statyczną relację między BBU i RRH, a każdy RRH nie musi być mapowany z jednym konkretnym BBU. W szczególności sygnały radiowe z / do RRH mogą być przetwarzane przez wirtualny eNB, co ułatwia między nimi wzmocnioną współpracę w czasie rzeczywistym. W związku z tym technologia wirtualizacji (wirtualizacja funkcji sieciowych, NFV) zmaksymalizuje elastyczność sieci RAN, umożliwiając różne poziomy skoordynowanych transmisji lub oddzielnych skojarzeń. Rysunek 9.14 pokazuje architekturę C-RAN dla DUDe.



Architektura C-RAN musi wykorzystywać nowy interfejs fronthaul, który umożliwia komunikację między RRH i BBU, co może potencjalnie zwiększyć opóźnienie RAN. Można rozpoznać różne opcje podziału funkcjonalnego protokołu, które zmieniają wymagania dotyczące opóźnień i pojemności w interfejsie fronthaul. Korzyścią jest zlokalizowanie części funkcji przetwarzania sygnałów w pobliżu RRH, wymagania dotyczące przepustowości i opóźnienia są doprowadzane do poziomu, który może być spełniony przez opłacalne sieci transportowe (tj. Ciemne włókna, technologie bezprzewodowe lub inne) oraz w tym samym czasie umożliwiać stosowanie ulepszonych procedur zarządzania zasobami radiowymi w ramach współpracy. Realistyczne opóźnienia fronthaul mogą mieć zakres od 250 ms do 4 ms

Kiedy RRH obsługujący DL i RRH obsługujący UL mają tę samą BBU, utrzymywane jest jedno połączenie RRC. Aby przyjąć tę sytuację, BBU ma jeden pojedynczy identyfikator komórki fizycznej (PCI).

#### 9.6.5 Na drodze do architektur opartych na 5G

W następnych kilku latach nastąpi gwałtowny wzrost liczby badań i rozwoju

obszary 5G. ITU i 3GPP rozszerzają swoją działalność o wymagania i standaryzację 5G, która ma zostać wprowadzona na rynek na początku 2020 roku. Ponadto wzrasta zgodność co do głównych wymagań 5G dotyczących pojemności i opóźnień, a także kluczowych cech technicznych 5G, w tym zagęszczenia

stacji bazowych, masywnego MIMO, częstotliwości fal milimetrowych i prawdopodobnie "bezkomórkowych". "architektura. Czy w tym widoku 5G nowy standard 5G musi zawierać dodatkowe funkcje, aby natywnie obsługiwać DUDe? Ważnym pytaniem, na które należy odpowiedzieć, jest to, czy prosta ewolucja obecnego projektu architektury 3GPP byłaby w stanie skutecznie wspierać DUDe w przyszłych heterogenicznych wdrożeniach 5G. W poprzednich podrozdziałach podjęto dyskusję, w jaki sposób architektura LTE-A obsługuje już implementację DUDe, gdy różne stacje BS są połączone za pośrednictwem światłowodu z tą samą jednostką pasma podstawowego. Omówiono także, w jaki sposób wsparcie DUDe w 4G może być ograniczone do różnych częstotliwości w przypadku agregacji nośnej lub podwójnej łączności. Intuicyjnie, wszelkie przyszłe wydania 5G w 3GPP powinny umożliwiać dwukierunkową łączność o tej samej częstotliwości, która, pomimo wpływu na zarządzanie zasobami i zakłóceniami, nie jest uważana za poważną aktualizację. Innym problemem, który należy rozwiązać, jest zapewnienie właściwego szyfrowania wszystkich danych i kanałów sterujących, zwłaszcza gdy komunikacja między interfejsami X odbywa się za pośrednictwem interfejsu X2. W LTE każda stacja bazowa może obsługiwać dziesiątki tuneli IP Security (IPSec). Jednak zarządzanie bezpieczeństwem za pośrednictwem protokołu IPSec jest tak problematyczne, że operatorzy zwykle wdrażają tylko kilka bramek IPSec w poszczególnych krajach, w których większość bramek jest rozmieszczonych w pobliżu SGW. Oznacza to, że ruch logicznie przebiegający przez X2 jest faktycznie kierowany przez SGW, co powoduje duże opóźnienie, które powoduje, że operacja DUDe jest raczej nieefektywna. Podczas gdy LTE-A ma więcej bramek IPSec rozmieszczonych bliżej krawędzi mobilnej, projekty architektury 5G muszą rozważyć wydajne szyfrowanie ruchu X2 w celu zmniejszenia dodatkowego opóźnienia w 4G. Ponadto niektóre prace integracyjne są potrzebne w przypadku niektórych nowych technik, które okazały się skuteczne w przypadku systemów sprzężonych. Integracja DUDe z odsprężonymi płaszczyznami sterowania / danych i dostępem wspomaganym licencją (LAA) będzie wymagać pewnych zmian w architekturze. Samoorganizujące się sieci (SON) będą również pomocne w zarządzaniu i koordynowaniu zwiększonej swobody wprowadzonej przez DUDe. Biorąc pod uwagę powyższą dyskusję, można stwierdzić, że natywne wsparcie DUDe nie wymaga dużych zmian konstrukcyjnych w 5G z perspektywy architektonicznej

### **9.7 Wyzwania DL / UL i możliwości dla wyższych warstw**

W tej sekcji główny nacisk położono na warstwę transportową, w szczególności na TCP, biorąc pod uwagę, że jest to powszechnie stosowany protokół warstwy transportowej w celu uzyskania niezawodnego transferu danych między końcami w sieciach opartych na protokole IP. Pomimo tego, że TCP został początkowo zaprojektowany dla aplikacji elastycznych, jest on obecnie powszechnie stosowany w różnych popularnych aplikacjach do przesyłania strumieniowego. Warto zauważyć, że Real Media i Windows Media, dwie dominujące aplikacje do przesyłania strumieniowego, oparte są na strumieniu TCP. Tutaj omawiamy wyzwania w wyższej warstwie i TCP w dwóch aspektach: Po pierwsze, wpływ kanału UL na wydajność TCP, a po drugie, wsparcie dostępne w TCP w celu umożliwienia dekompresji DL / UL. W tej sekcji obie strony argumentacji są szeroko opracowane, a na główne pytania odpowiedziano:

1. Jak asymetria DL / UL może wpływać na wydajność TCP?
2. Jakie są efekty alokacji wspólnych zasobów DL / UL na TCP?
3. Jakie są efekty DUDe na TCP?
4. Jak DUDe można włączyć za pomocą dostępnych mechanizmów w TCP?

#### **9.7.1 Wpływ UL na działanie wyższych warstw**

Połączenia TCP, jako z natury dwukierunkowe, wymagają od odbiornika potwierdzenia ACK w celu uzyskania niezawodnej komunikacji. Z jednej strony, powszechne założenie w sieciach centralnych danych polega na tym, że DL przenosi większe obciążenie danych, w ten sposób określona jest mniejsza szerokość pasma dla ścieżki UL. Z drugiej strony pomiary kanałów są przeprowadzane w oparciu o kanał DL. Co więcej, w wielu dzisiejszych aplikacjach, takich jak MTC, nadawca może być terminalem mobilnym, który różni się od tradycyjnych aplikacji, a zatem pojemność UL odgrywa ważną rolę. Wpływ asymetrii łącza na wydajność TCP jest szeroko badany w sieciach przewodowych. Ograniczona dostępna przepustowość i przeciążenie na ścieżce zwrotnej narusza zasadę taktowania ACK i może powodować wzrost częstotliwości RTT, co może zmniejszyć przepustowość TCP na ścieżce do przodu. Kilka prac badawczych bada te problemy i zaproponowano szereg rozwiązań. Niektóre z tych propozycji wymagają jednoznacznego wsparcia routerów lub pośrednich, podczas gdy inne są schematami od końca do końca. Na przykład próby kontroli przeciążenia ACK w celu zmniejszenia natężenia wysyłania ruchu ACK przy założeniu, że obniżenie stawki ACK może pomóc w zmniejszeniu samego przeciążenia. W zakresie komunikacji mobilnej i bezprzewodowej wiele pracy poświęcono optymalizacji działania w kanale DL sieci bezprzewodowych, takich jak na przykład doskonalenie mechanizmu świadomości o zatłoczeniu, tak aby minimalizować wpływ losowych strat bezprzewodowych i optymalizować osiągnięty poziom uczciwości end-to-end częstości TCP. Ograniczona ilość badań w rzeczywistości odnosi się do wpływu takiego braku równowagi na wydajność TCP. W związku z tym ograniczona pojemność lub niższa jakość kanału na ścieżce nie wpłynie negatywnie na wydajność od końca do końca. To rozwiązanie może być zaimplementowane przez dodanie jednego bitu w opcjonalnej części nagłówka TCP w celu wyjaśnienia stanu kontroli przeciążenia, w której działa TCP, tj. Unikania przeciążenia lub powolnego uruchamiania.

#### **9.7.1.1 Modyfikacje sterowania przeciążeniem TCP**

Rdzeniem algorytmu sterowania przeciążeniami TCP jest powolny start i przeciążenie faza unikania. W powolnym starcie przepustowość przepływu TCP zależy od rzeczywistej wartości CW i RTT. Z drugiej strony, przepustowość w fazie unikania przeciążenia może być wyrażona przez przepustowość TCP w stanie ustalonym. Wprowadzenie nowej flagi w nagłówku TCP, flaga SS / CA, może zilustrować aktualny stan TCP, tj. SS / CA równe zero reprezentuje bycie w Slow Start, a jedno jest w Unikaniu zatorów.

#### **9.7.1.2 Sterowanie korkiem TCP i przydział zasobów DL / UL**

Podczas gdy mechanizmy alokacji zasobów zwykle uwzględniają tylko DL, często przydatne jest rozważenie wspólnego problemu przydziału DL / UL. Takie techniki przydzielania mogą zapewnić dynamiczną granicę między pojemnością łącza DL a przepustowością UL, w celu wydajniejszej alokacji całkowitej ilości zasobów. Ta proporcja może być definiowana dynamicznie w oparciu o różne ograniczenia systemowe, aby zagwarantować wymaganą QoS. Aby w pełni zilustrować wspólną alokację zasobów, tutaj szczegółowo opisano problem alokacji zasobów OFDMA, który jest również ograniczony wymaganiami dwukierunkowego połączenia TCP, tak, że zapewnione jest dostarczanie pakietów TCP ACK dla określonej przydzielonej przepustowości do DL. . Problem alokacji zasobów DL per se może mieć na celu maksymalizację sumy uzależnionej od zużycia energii, minimalizację całkowitego zużycia energii przy jednoczesnym spełnieniu wymagań minimalnej stawki, rozważenie pewnego stopnia QoS lub uczciwości. Tutaj omawiany jest wspólny przydział zasobów DL / UL, aby zapewnić maksymalną przepustowość w DL, a także zagwarantować dostarczanie pakietów DL z odpowiednią szybkością transmisji danych na UL. Cel schematu alokacji DL przedstawiono w dwóch wariantach. Pierwszy zdefiniowany problem, tj. (P1), ma na celu maksymalizację szybkości sumowania, ograniczonej do proporcjonalnej szybkości przesyłania danych w każdym pojedynczym strumieniu TCP. Współczynnik proporcjonalności jest ważony teoretyczną przepustowością TCP, która jest przepustowością, którą może osiągnąć przepływ TCP, w zależności od końcowego RTT i PER

odpowiedniego przepływu - tę przepustowość można uznać za rzeczywistą przepustowość ścieżka od końca do końca. W drugim problemie, tj. (P2), badana jest różnica między przydzielonym łączem bezprzewodowym a teoretyczną przepustowością TCP. (P2) ma na celu zmaksymalizowanie stawki sumy przy jednoczesnej minimalizacji luki między przydzieloną szybkością transmisji danych a każdym przepływem i teoretyczną przepustowością, którą można uzyskać dzięki przepływowi TCP.

### 9.7.1.3 Formułowanie przydziału zasobów OFDMA - (P1)

Sformułowany problem optymalizacji (P1) ma na celu zmaksymalizowanie stopy sumy wskaźnika DL, podczas gdy uczciwość TCP jest zapewniona przez nałożenie zestawu nieliniowych ograniczeń na problem. Innymi słowy, proporcjonalna stawka DL wśród użytkowników w odniesieniu do teoretycznej przepustowości TCP podlega ograniczeniom. W UL, zakładając odbiornik TCP potwierdza każdy pojedynczy pakiet, minimalna wymagana szybkość transmisji danych byłaby proporcją szybkości transmisji danych DL dla każdego określonego przepływu TCP w zależności od rozmiaru pakietu ACK, który można zwiększyć na przykład przy użyciu opcji SACK,

$$R_{u_i} \geq \rho R_{d_i}, \text{ i.e., } 0 < \rho < 1.$$

$$(P1) : \text{Maximize } \sum_{i=1}^n \sum_{j=1}^m c_j a_{ij} w \log_2 \left( 1 + \frac{P_{ij} G_{ij}}{\sigma^2 c_3} \right),$$

$$\text{subject to : } \sum_{i=1}^n a_{ij} \leq 1, \quad \forall j \in \{1, \dots, m\} \quad (9.1)$$

$$\sum_{i=1}^n \sum_{j=1}^m c_j a_{ij} p_{ij} \leq P_T, \quad (9.2)$$

$$\sum_{j=1}^m (1 - c_j) a_{ij} p_{ij} \leq P_t, \quad \forall i \in \{1, \dots, n\} \quad (9.3)$$

$$\frac{R_{d_i}}{B_i} = \frac{R_{d_1}}{B_1}, \quad \forall i \in \{2, \dots, n\} \quad (9.4)$$

$$R_{u_i} \geq \rho R_{d_i}, \quad \forall i \in \{1, \dots, n\} \quad (9.5)$$

$$\sum_{j=1}^m c_j \leq m_d, \quad m_d \in \{1, \dots, m\}, \quad (9.6)$$

$$p_{ij} \geq 0, \quad \forall i \in \{1, \dots, n\}, j \in \{1, \dots, m\} \quad (9.7)$$

$$a_{ij} \in \{0, 1\}, \quad \forall i \in \{1, \dots, n\}, j \in \{1, \dots, m\} \quad (9.8)$$

$$c_j \in \{0, 1\}. \quad \forall j \in \{1, \dots, m\} \quad (9.9)$$

W (P1), pomimo klasycznych podejść w maksymalizacji przepustowości, podnośne  $a_{ij}$  i moc nad każdą podnośną ( $p_{ij}$ ) są przydzielane w taki sposób, aby pewne metryki wydajności TCP są gwarantowane. Te wskaźniki wydajności są dostarczane za pośrednictwem ograniczeń opisanych poniżej. Ograniczenia (9.1) zapewniają, że każdy podnośnik jest przypisany tylko jednemu użytkownikowi. Zakładamy, że  $c_j$  reprezentuje przydział podnośnej  $j$  do DL ( $c_j = 1$ ) lub UL ( $c_j = 0$ ). W ten sposób ograniczenia (9.2) i (9.3) ograniczają całkowitą dostępną moc w stacji bazowej,  $P_T$  i przy każdym użytkowniku mobilnym,  $P_t$ . Co więcej, ograniczenie (9.4) zapewnia sprawiedliwość wśród przepływów TCP z zachowaniem proporcjonalnej szybkości w odniesieniu do przepustowości TCP dla każdego użytkownika. Ograniczenie (9.5) zapewnia wymaganą szybkość transmisji danych dla UL, aby zagwarantować



dostarczenie przydzielonych zasobów DL, a na koniec ograniczenie (9.6) ogranicza liczbę podnośnych DL do md. Zauważ, że optymalną wartość md można znaleźć rozwiązując problem (P1) iteracyjnie dla różnych wartości md.

#### 9.7.1.4 Sformułowanie przydziału zasobów OFDMA - (P2)

Podobny do (P1), (P2) ma na celu maksymalizację stopy sumy DL, ale chwilowy przydział szybkości jest ograniczany przez teoretyczną przepustowość TCP. Różnica między przydzieloną stopą DL dla ego użytkownika a teoretyczną przepustowością TCP przepływu i jest oznaczona przez  $D_i$ :

$$D_i = |\alpha B_i - R_{d_i}|. \quad (9.10)$$

W (9.10),  $\alpha$  reprezentuje narzut nagłówka TCP / IP. Problem przydziału zasobów w tym przypadku próbuje zminimalizować  $D$  przy maksymalizacji stopy sumy. Dlatego (P2) można zdefiniować jako problem optymalizacji wielu celów. Tam są różne podejścia do sformułowania takiego problemu o wielu celach; używamy dobrze zbadanego podejście, które łączy wiele celów w jeden cel funkcja, której rozwiązanie jest optymalne dla Pareto:

$$(P2) : \text{Maximize } \sum_{i=1}^n \sum_{j=1}^m c_j a_{ij} w \log_2 \left( 1 + \frac{p_{ij} G_{ij}}{\sigma^2 c_3} \right) - \mu \sum_{i=1}^n D_i.$$

$$\text{subject to: } \sum_{i=1}^n a_{ij} \leq 1, \quad \forall j \in \{1, \dots, m\} \quad (9.11)$$

$$\sum_{i=1}^n \sum_{j=1}^m c_j a_{ij} p_{ij} \leq P_T, \quad (9.12)$$

$$\sum_{j=1}^m (1 - c_j) a_{ij} p_{ij} \leq P_i, \quad \forall i \in \{1, \dots, n\} \quad (9.13)$$

$$R_{u_i} \geq \rho R_{d_i}, \quad \forall i \in \{1, \dots, n\} \quad (9.14)$$

$$\sum_{j=1}^m c_j \leq m_d, \quad m_d \in \{1, \dots, m\}, \quad (9.15)$$

$$p_{ij} \geq 0, \quad \forall i \in \{1, \dots, n\}, j \in \{1, \dots, m\} \quad (9.16)$$

$$a_{ij} \in \{0, 1\}, \quad \forall i \in \{1, \dots, n\}, j \in \{1, \dots, m\} \quad (9.17)$$

$$c_j \in \{0, 1\}. \quad \forall j \in \{1, \dots, m\} \quad (9.18)$$

Ograniczenia (9.11) - (9.18) są takie same jak (9.1) - (9.3) i (9.5) - (9.9). Jak wspomniano powyżej, problem (P2) ma optymalne rozwiązanie Pareto; dlatego rozwiązanie nie jest unikalne i zależy od wartości  $\mu$ , która równoważy oba cele. W powyższym problemie, zwiększenie wartości  $\mu$  przesunie salda alokacji w kierunku przepustowości TCP, przy jednoczesnym zmniejszeniu wartości  $\mu$  przesunie równowagę w kierunku problemu maksymalizacji szybkości danych.

#### 9.7.1.5 Wyszukiwanie rozwiązań dla (P1) i (P2)

Dobrze stosowaną metodą rozwiązywania problemów z optymalizacją wielu zmiennych jest odseparowanie problemu, tj. przydzielenie  $a_{ij}$  i  $p_{ij}$  osobno. W przydzielaniu nośnej, zakłada się, że moc jest równo rozłożona na wszystkich podnośnych, dlatego rozwiązanie jest nieoptymalne. Następnie do

pewnej alokacji podnośnej można przeformułować problem optymalizacyjny w stosunku do zmiennej ciągłej  $p_{ij}$ . Tak więc, stosując metodę napełniania wodą, moc będzie rozkładana optymalnie. Zasada działania algorytmu DL polega na alokacji podnośnej o najwyższym dostępnym kanale dostępnym dla każdego użytkownika. Ponadto, w pierwszej rundzie przydzielania, użytkownicy o najwyższej teoretycznej przepustowości TCP (największej wartości  $B_i$ ) są pierwszymi, którzy wybierają podnośną. Następnie, w każdej rundzie alokacji, przy rozwiązywaniu problemu (P1) użytkownicy o najniższej proporcjonalnej szybkości mają pierwszeństwo wyboru najlepszej dostępnej podnośnej. Ten etap wykonuje inny problem (P2), tj. Użytkownik z najmniejszą funkcją celu wybiera następną podnośną. Przydzielanie podsięci w UL w obu problemach ma na celu spełnienie wymagań minimalnej stawki UL. Początkowa wartość  $m_d$  utrzymuje proporcję  $\rho$  między liczbą podnośnych DL i UL -  $m_d = m \cdot 1/1 + \rho$ . Następnie, w kilku powtórzeniach, zostanie znaleziona największa wartość  $m_d$ , która spełnia ograniczenia (9.4); wyrażnie ta wartość maksymalizuje funkcję celu. Powyższa procedura dla problemu (P1) jest wyszczególniona w algorytmie 1, w którym  $\Omega$  jest zbiorem przydzielonych podnośnych dla użytkownika  $i$  w DL, a  $\Psi_i$  jest zestawem alokowanych podnośnych dla tego użytkownika w UL. Algorytm 2 opisuje problem rozwiązywania przydziału podnośnych (P2), w którym etapy (a), (b), (d) i (e) są podobne do algorytmu 1, i tylko etap (c) jest przekształcany. Problemem alokacji mocy z predefiniowanym przydziałem podnośnych jest w oparciu o przeformułowanie (P1) na problem maksymalizacji w odniesieniu do zmiennej ciągłej  $p_{ij}$  (podobnie dla (P2)).

#### Algorytm 1 Algorytm alokacji podnośnej dla problemu optymalizacji (P1)

(a) Inicjalizacja

1.  $m_d = 1 / (1 + \rho \cdot m)$
2. Ustawiamy  $R_{di} = 0$  i  $\Omega_i = \emptyset$  dla  $i = 1$  do  $n$  i  $C_d = \{1 \dots m_d\}$
3. Ustawiamy  $R_{ui} = 0$  i  $\Psi_i = \emptyset$  dla  $i = 1$  do  $n$  i  $C_u = \{m_d + 1 \dots m_d\}$
4. sortujemy indeks użytkownika w porządku malejącym  $B_i$ .

(b) dla  $i = 1$  do  $n$

1. Znajdź podnośne  $k$  spełniające  $|G_{ik}| > |G_{ij}|$  dla wszystkich  $j \in C_d$ .
2. Niech  $\Omega_i = \Omega_i \cup \{k\}$  a  $C_d = C_d - \{k\}$
3. Aktualizujemy  $R_{di}$ .

(c) podczas gdy  $C_d \neq \emptyset$

1. Znajdź użytkownika  $l$  spełniającego  $R_{dl}/B_l < R_{di}/B_i$  dla wszystkich  $i \in \{1, \dots, n\}$
2. Dla użytkownika  $l$ , znajdź podnośną  $k$  spełniającą  $|G_{lk}| > |G_{lj}|$  dla wszystkich  $j \in C_d$
3. Niech  $\Omega_l = \Omega_l \cup \{k\}$  a  $C_d = C_d - \{k\}$
4. Aktualizujemy  $R_{dl}$ .

(d) dla  $i =$  do  $n$

1. Znajdź podnośną  $k$  spełniającą  $|G_{ik}| > |G_{ij}|$  dla wszystkich  $j \in C_u$
2. Niech  $\Psi_i = \Psi_i \cup \{k\}$  a  $C_u = C_u - \{k\}$
3. Aktualizujemy  $R_{ui}$

(e) podczas gdy  $C_u \neq \emptyset$

1. Znajdź użytkownika  $l$  spełniającego  $(R_{ui} - \rho R_{di}) < (R_{ui} - \rho R_{di})$  dla wszystkich  $1 \in \{1, \dots, n\}$
2. Dla użytkownika  $l$ , znajdź podnośną  $k$  spełniającą  $|G_{ik}| > |G_{ij}|$  dla wszystkich  $j \in C_u$
3.  $\Psi_l = \Psi_l \cup \{k\}$  a  $C_u = C_u - \{k\}$
4. Aktualizujemy  $R_{ui}$

(f) Jeśli  $\sum_{i=1}^n R_{ui} \geq \rho \cdot \sum_{i=1}^n R_{di}$ .

1. Znajdź największe  $m_d \in \{m_{d1}, \dots, m\}$  takie, że (9.5) jest spełnione
2. W przeciwnym razie znajdź największy  $m_d \in \{1, \dots, m_{d1}\}$ , takie, że (9.5) jest spełnione

#### Algorytm 2 Algorytm alokacji podnośnej dla problemu optymalizacji (P2)

(c) podczas gdy  $C_d \neq \emptyset$

1. Znajdź użytkownika  $l$  spełniającego  $R_{dl} - \mu D_l < R_{dl} - \mu D_l$  dla wszystkich  $i \in \{1, \dots, n\}$
2. Dla użytkownika  $l$ , znajdź podnośną  $k$  spełniającą  $|G_{ik}| > |G_{ij}|$  dla wszystkich  $j \in C_u$
3. Niech  $\Omega_l = \Omega_l \cup \{k\}$  a  $C_d = C_d - \{k\}$
4. Aktualizujemy  $R_d$

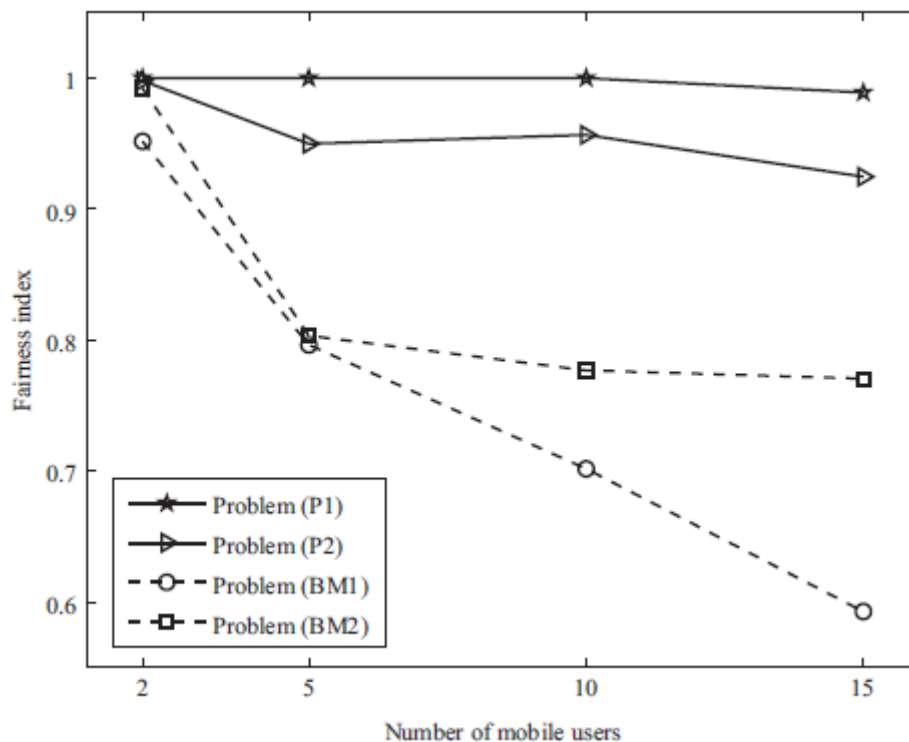
W algorytmach 1 i 2  $\Omega_{i1}$  i  $\Omega_{i2}$  wzajemnie się wykluczają, jeśli  $i_1 \neq i_2$ ; to samo założenie jest również ważne dla  $\Psi_i$ . Problem (P2) można również przepisać jako (P2') nad zmienną ciągłą  $p_{ij}$ .

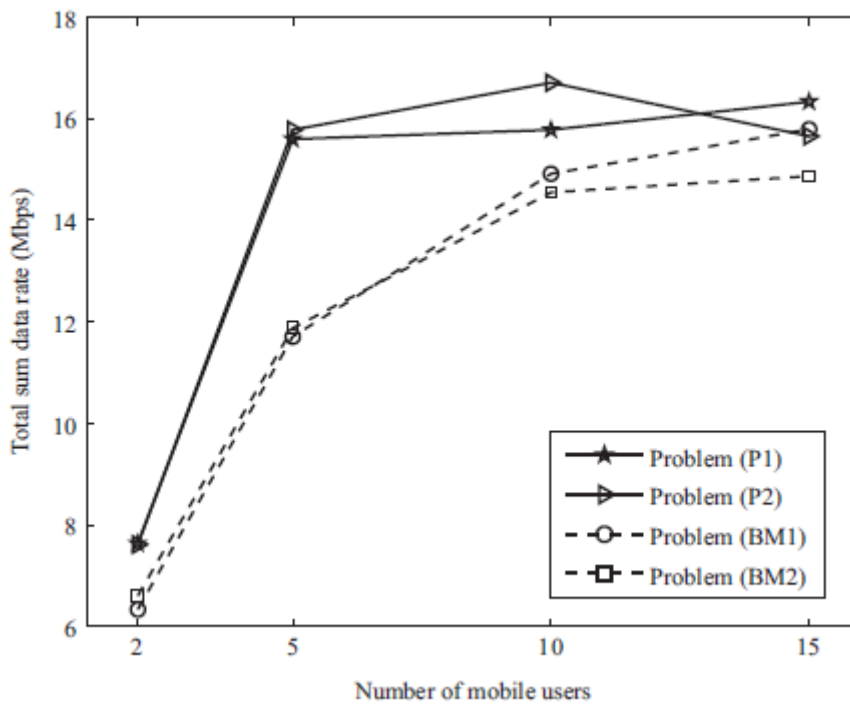
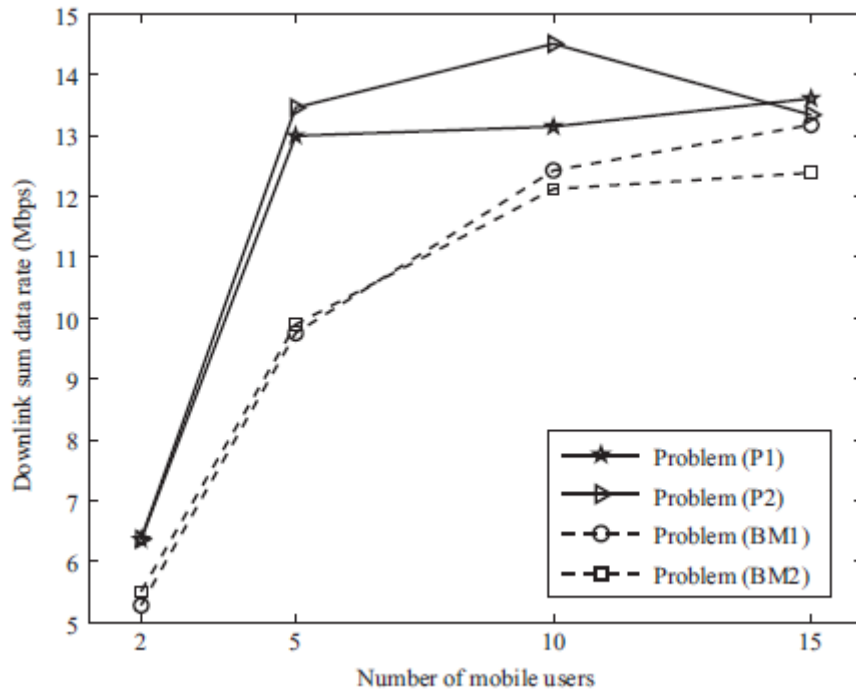
#### **9.7.1.6 Zmniejszenie efektu UL na wydajność TCP**

Oceny numeryczne przeprowadzane są za pomocą parametrów wymienionych w tabeli 9.1, w celu wykazania zmniejszonego wpływu nierównowagi UL na ogólne KPI systemu. Benchmarkami są problem z maksymalizacją stopy sum, oznaczony przez (BM1):

Pasmo	5 MHz
Całkowita liczba podnośnych OFDMA	52
Maksymalna moc na stacji bazowej	43 dBm
Maksymalna moc na urządzeniu mobilnym	23 dBm
Moc szumu cieplnego (hałas Johnson-Nyquist)	107 dBm
Docelowy BER	10E4
Średnia wartość SNR kanału bezprzewodowego	15 dB
Model kanału bezprzewodowego	pieszy model ITU
Maksymalny rozmiar segmentu TCP	1460 bajtów
Typ przepływu TCP	Długo trwający z SS / CA = 1
Liczba użytkowników mobilnych	1-15 użytkowników
Kompleksowy RTT	w [10 ms, 200 ms]

a także maksymalizacja szybkości sumowania z ograniczeniem o równej stopie, oznaczonym przez (BM2).  $m_d$  jest inicjowane z 32 podnośnymi dla DL, podczas gdy  $\rho = 0.2$ . Następnie w kroku (f) algorytmu 1 optymalna wartość  $m_d$  jest obliczana w kilku iteracjach. Dwa problemy z benchmarkami (BM1) i (BM2) działają przy  $m_d = 32$ ; w związku z tym porównanie wyników w tym przypadku głównie pokazuje, w jaki sposób osiąga się wydajność od końca do końca, ustawiając granicę między DL i UL dostosowującą się do aktualnego obciążenia systemu. Dwa badane wskaźniki KPI to przepustowość DL TCP, która również odzwierciedla przepustowość łącza UL i sprawiedliwość wśród przepływów TCP. Uczciwość jest określana ilościowo w oparciu o wskaźnik uczciwości Jaina, który jest dobrze wykorzystanym wskaźnikiem w sieciach komputerowych i komunikacyjnych. Prezentowane liczby na rysunkach 9.15- 9.17 są wartościami średnimi pobranymi ze 150 symulacji Monte Carlo. Z wykresu 9.15 można zaobserwować, że osiągnięty wskaźnik uczciwości z podziałem na zasoby (P1) wzrasta średnio o około 30% w porównaniu z wynikami schematu przydziału (BM1). Ponadto, na Rysunku 9.15 pokazano, że wraz ze wzrostem liczby użytkowników mobilnych konkurujących za pośrednictwem łącza bezprzewodowego, dystrybucja zasobów w uczciwy sposób jest trudniejsza. Największy wzrost wskaźnika rzetelności występuje w przypadku 15 użytkowników, gdzie (P1) jest o 70% lepszy od benchmarków. Co więcej, na rysunkach 9.16 i 9.17 pokazano, że wspólny przydział DL i UL poprawia całkowitą łączną przepustowość, a także zagregowaną przepustowość DL o około 15%.





### 9.7.2 DL i UL, połączone i niezwiązane

Włączenie DUE znacznie przyczynia się do sukcesu przydzielania zasobów DL / UL TCP-ware, a wspólne przydzielanie zasobów poprawia wydajność komunikacji UL i DL. Po pierwsze, wspólny przydział zasobów umożliwi UL posiadanie wystarczającej pojemności do przenoszenia TCP ACK dla DL i na odwrót. Dlatego przydzielone zasoby w sieci RAN będą efektywnie wykorzystywane przez połączenie typu koniec-koniec. Po drugie, niezwiązane z wielkością produkcji DL / UL zapewni dodatkową elastyczność w zakresie wspólnego przydzielania zasobów. W związku z tym rozdzielone

DL / UL, podczas gdy zasoby dla obu są przydzielane w koordynacji ze sobą, a w świadomości z wyższych warstw mogą zapewniać znaczącą dodatkową efektywność komunikacji między punktami końcowymi, zarówno pod względem opóźnienia, jak i szybkości transmisji danych.

### **9.7.2 DL i UL, połączone i niezwiązane**

Włączenie DUDe znacznie przyczynia się do sukcesu przydzielania zasobów DL / UL TCP-ware, a wspólne przydzielanie zasobów poprawia wydajność komunikacji UL i DL. Po pierwsze, wspólny przydział zasobów umożliwi UL posiadanie wystarczającej pojemności do przenoszenia TCP ACK dla DL i na odwrót. Dlatego przydzielone zasoby w sieci RAN będą efektywnie wykorzystywane przez połączenie typu koniec-koniec. Po drugie, niezwiązane z wielkością produkcji DL / UL zapewni dodatkową elastyczność w zakresie wspólnego przydzielania zasobów. W związku z tym rozdzielone DL / UL, podczas gdy zasoby dla tych dwóch są przydzielane w koordynacji ze sobą, a w świadomości z wyższych warstw mogą zapewniać znaczącą dodatkową skuteczność komunikacji między punktami końcowymi, zarówno pod względem opóźnienia, jak i szybkości transmisji danych.

#### **9.7.2.1 Włączanie DUDe przez wiele strumieni TCP**

Jednym z wyzwań stojących przed architekturą DUDe jest możliwość posiadania wielu przepływów TCP, które mogą obsługiwać oddzielne połączenia DL i UL. Jednym z głównych istniejących wysiłków w Internecie jest MultiPath TCP (MPTCP) . Protokół MPTCP umożliwia TCP prezentowanie standardowego interfejsu TCP dla aplikacji, podczas gdy kilka interfejsów IP jest używanych jednocześnie. Innymi słowy, komunikacja danych przepływu jest rozłożona na wiele pod-strumieni. Połączenia MPTCP rozpoczynają się regularnie, a jeśli są dostępne dodatkowe ścieżki, na tych ścieżkach tworzone są dodatkowe sesje TCP, określane jako podprzepływy, i są łączone z istniejącą sesją, która nadal pojawia się jako pojedyncze połączenie z aplikacjami na obu końcach. W związku z tym MPTCP umożliwi przełączanie DL / UL sprzężonego i niezwiązanego z siecią dla różnych strumieni ruchu. Podczas gdy MPTCP może być świetnym narzędziem dla architektury DUDe, wyzwaniem jest tutaj dostępność dwóch różnych adresów IP. Innymi słowy, MPTCP może być używany tylko do obsługi DL i UL, jeśli UE jest wieloadresowy. Przykładem innych działań związanych z uruchamianiem wielu i równoległych przepływów TCP jest MuITCP , który zachowuje się tak, jakby był zbiorem wielu wirtualnych połączeń TCP. Taka inicjatywa w domenie internetowej jest często podejmowana w celu zwiększenia szybkości transmisji danych i lepszego wykorzystania przepustowości od końca do końca.

## **9.8 Wnioski**

Od początku istnienia sieci komórkowych połączenie UL było zawsze powiązane z tą samą stacją bazową, która została wybrana do odbioru DL. W tej części postulat ten został ponownie przeanalizowany i wprowadzono funkcję rozłączania zstępującego / łącza uplink (DUDe). DUDe to nowy architektoniczny paradygmat, w którym UL i DL nie są ograniczone do kojarzenia z tą samą stacją bazową i są skutecznie traktowane jako oddzielne sieci. Ta koncepcja staje się bardzo istotna przy oczekiwanych zagęszczaniu w przyszłych sieciach komórkowych, gdzie każdy terminal będzie miał wiele punktów dostępowych w pobliżu i umożliwiając terminalowi wybranie najlepszego połączenia UL i DL, spowoduje znacznie bardziej wydajne działanie sieci komórkowych . W tym rozdziale zidentyfikowano główne wyzwania stojące przed heterogenicznymi sieciami oraz sposób, w jaki DUDe może być rozwiązaniem niektórych z tych wyzwań. Wzrost wydajności DUDe wykazano również pod względem przepustowości UL, zasięgu, równoważenia obciążenia i niezawodności. Podkreślono współdziałanie DUDe z różnymi nowymi trendami w sieciach komórkowych. W końcu, dokładnie przeanalizowano realizację DUDe w zakresie architektury, w której zilustrowano różne opcje implementacji DUDe w LTE, z możliwością zobaczenia, co jest potrzebne w sieciach 5G, aby DUDe był natywnie wspierany, a także efekty DUDe na wyższych warstwach.

# Wirtualizacja sieci bezprzewodowych

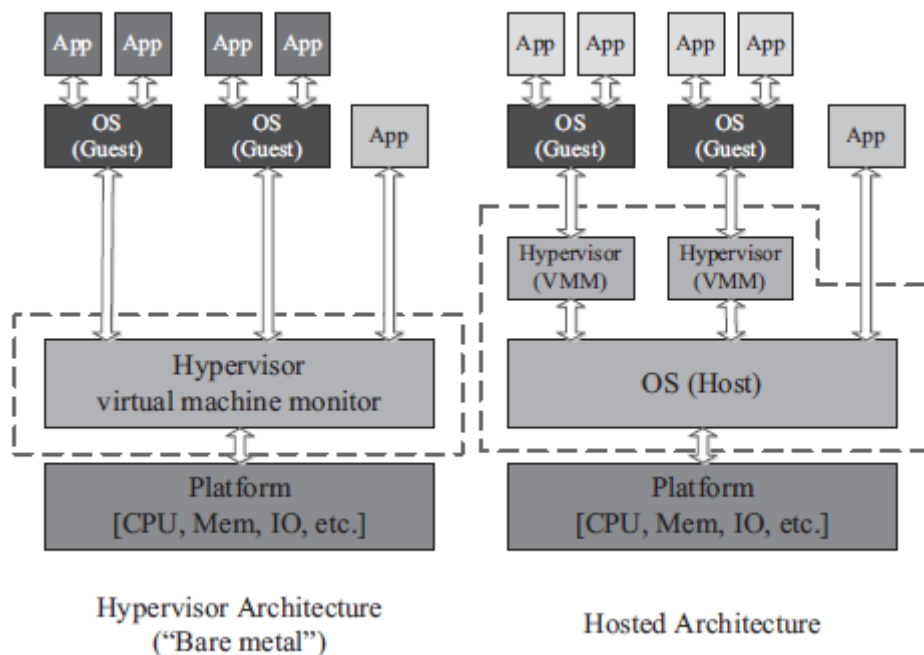
## 10.1 Wprowadzenie

Dzięki imponującemu rozwojowi inteligentnych terminali i kart danych, mobilnego Internetu w ciągu ostatnich kilku lat nastąpił gwałtowny wzrost usług. Jednak ludzie nie widzą, że operatorzy telefonii komórkowej zawsze ponoszą wiele wysiłku, aby odpowiedzieć na stale rosnący ruch. W rzeczywistości, podczas gdy ruch ma wzrost wykładniczy, tempo wzrostu dochodów i zysków większości operatorów jest nie więcej niż 10%, a nawet ujemne. Powodem jest to, że szybko rozwijający się mobilny Internet wywiera ogromną presję na sieci komórkowe. Aby poradzić sobie z codziennym rosnącym ruchem, operatorzy muszą płacić więcej zasobów za wdrożenie, przebudowę lub modernizację sieci, co prowadzi do ogromnych nakładów inwestycyjnych i kosztów operacyjnych (OPEX). Łatwo można sobie wyobrazić, że jeśli w nadchodzących latach mobilny Internet będzie się nadal zwiększał, sieci komórkowe będą coraz bardziej naciskać. Operatorzy mobilni stoją przed niespotykanymi dotąd wyzwaniami. W jaki sposób operatorzy skutecznie redukują OPEX, zwiększają wykorzystanie sieci, a nawet przyspieszają aktywację rozwoju usług, aby sprostać wyzwaniom związanym z erą mobilnego Internetu? Aby rozwiązać ten problem, zaczniemy od wiedzy na temat tradycyjnego sprzętu telekomunikacyjnego, jego formy, trybu rozwoju i debugowania, konserwacji sieci i aktualizacji. Sprzęt sieci telekomunikacyjnej ma dwie główne cechy, mianowicie "zastrzeżoną platformę" i "dedykowane chipsety". "Zastrzeżona platforma" oznacza, że platforma, na której opracowywane jest urządzenie, jest zastrzeżona, różni się od dostawcy do dostawcy. "Dedykowane chipsety" oznaczają, że chipsety, takie jak cyfrowe przetwarzanie sygnału (DSP), programowalna matryca bramkowa (FPGA) i układ scalony specyficzny dla aplikacji (ASIC) są głównie przystosowane do rozwoju sprzętu. W wyniku tego zastrzeżone platformy powodują słabe problemy z kompatybilnością i kompatybilnością, podczas gdy dedykowane użycie chipsetu powoduje problemy ze słabą programowalnością, słabą przenośnością i powolnym rozwojem dodatkowym. Ponadto branża ta wymaga inżynierów o wysokich umiejętnościach, którzy łatwo można znaleźć. W ten sposób ograniczona jest konkurencyjność i innowacyjność usług w całej branży. Poza tym sprzęt zazwyczaj wymaga długiego cyklu programowania, złożonego testowania i weryfikacji. Poza tym, sieć telekomunikacyjna zużywa ogromną ilość energii przy niskim wykorzystaniu. Można przewidywać, że stale rosnące wymagania mobilnego Internetu zwiększą wymagania sieci bezprzewodowych następnej generacji pod względem możliwości rozbudowy sieci, innowacji i terminowego wprowadzania usług, wydajności energetycznej itd., które są tym, czym są tradycyjne tryby rozwoju sprzętu. niezdolny do spełnienia. Następnie zwróćmy naszą uwagę na branżę IT. Krajobraz jest inny. Branża IT słynie z otwartości i miękkości. Różne standardowe interfejsy znacznie ułatwiają współpracę, dzięki czemu programiści mogą skoncentrować się na rozwoju i optymalizacji niektórych komponentów bez konieczności uwzględniania problemów z połączeniami. Dobrymi przykładami są karty wideo, procesor, wspomnienia i tak dalej. Ponadto platforma ogólnego przeznaczenia (GPP), np. Architektura x86 jest powszechnie stosowana, co ułatwia tworzenie, utrzymywanie i przenoszenie oprogramowania. To, w pewnym sensie, zachęca do innowacji usług i przyspiesza czas wprowadzenia produktu na rynek. Rozwój branży IT, w szczególności dobrobytu centrów danych, jest w dużej mierze zasługą kluczowej technologii, a mianowicie technologii wirtualizacji.

## 10.2 Podstawy wirtualizacji

Jeśli chodzi o pochodzenie technologii wirtualizacji, po raz pierwszy użył tego terminu dr Christopher Strachey z Oxford University. W książce "Time Sharing in Large Fast Computers" wspomniano o koncepcji wirtualizacji w latach sześćdziesiątych. Następnie w 1964 roku IBM zainicjował projekt, który nazwano systemem CP / CMS i ewoluował do Virtual Machine Facility / 370 . Od tej pory do odróżnienia

od maszyny fizycznej używano pojęcia maszyny wirtualnej (VM). W obszarze technologii wirtualizacji IBM i środowisko badawcze stwierdziły podstawowe pojęcia związane z wirtualizacją, a Goldberg opublikował prace [3], które dotyczyły koncepcji i korzyści tej technologii. Jednak od kilku lat zainteresowanie spadło, a nic związanego z technologią wirtualizacji nie miało miejsca aż do 1999 roku, kiedy firma VMware Inc. zaprezentowała swój produkt VMware Virtual Platform dla architektury x86-32. Ale czym jest w skrócie wirtualizacja? Jedna opinia jest taka, że wirtualizacja jest aktem ABSTRAKCJI fizycznych granic technologii. Aby osiągnąć fizyczną abstrakcję, istnieje kilka sposobów. Jednym z nich jest to, że aplikacje lub serwery nie potrzebują już dedykowanego sprzętu fizycznego, aby działać jako niezależne jednostki i że maszyna wirtualna może wejść w grę. W przypadku maszyn wirtualnych tradycyjny sprzęt jest emulowany i prezentowany systemowi operacyjnemu tak, jakby rzeczywiście istniał. Poza tym wiele maszyn wirtualnych może działać równolegle w różnych systemach operacyjnych na tym samym sprzęcie. Inna opinia preferuje prostą analogię wirtualizacji do obrazu-obrazu niektórych telewizorów i dekodatorów, ponieważ wyświetla mały wirtualny obraz telewizyjny na drugim obrazie telewizyjnym, umożliwiając jednocześnie odtwarzanie obu programów [6]. W oparciu o zwirtualizowaną platformę, która jest połączeniem wirtualizacji komputerów, sieci i pamięci masowej, tradycyjne aplikacje lub wiele zestawów oprogramowania może być wykonywanych w tym samym czasie. W ścieżce ewolucji technologii wirtualizacji istnieją dwa typy monitorów maszyn wirtualizacji (VMM). Jeden uruchamia się bezpośrednio na sprzęcie z takimi samymi funkcjami, jak w tradycyjnych systemach operacyjnych (OS) w celu kontrolowania lub zarządzania sprzętem. Nazywa się to architekturą "Hypervisors". Drugi działa na istniejącym systemie operacyjnym i zarządza systemem operacyjnym. W tym scenariuszu system operacyjny poniżej VMM nazywa się Hostem, a ten jest nazywany systemem gościa. Tryb nazywa się architekturą "Hosted". Rysunek 10.1 pokazuje przykład tych dwóch typów VMM i hipernadzorców.



W przypadku wirtualizacji na poziomie systemu istnieją trzy podejścia: pełna wirtualizacja, parawirtualizacja i wirtualizacja systemu operacyjnego. Możemy po prostu wprowadzić różnice między tymi trzema metodami:



- Pełna wirtualizacja ma jeden dedykowany VMM lub hiperwizor do zarządzania bazowym sprzętem. Dzięki asystentowi sprzętowemu procesory Intel i AMD mogą obsługiwać takie wirtualizacje. Może obsługiwać różne systemy operacyjne gościa.
- Parawirtualizacja różni się nieco od pełnej wirtualizacji. Ma lepszą wydajność niż pełna wirtualizacja. Ale wymaga również modyfikacji systemu gościa.
- Wirtualizacja systemu operacyjnego wymaga jednej warstwy do wirtualizacji i ma najlepsze skalowalności wydajności. Ale obsługuje tylko jeden system operacyjny. W przypadku powyższych trzech podejść nie ma potrzeby wprowadzania żadnych zmian w aplikacjach. Różnica polega na złożoności, możliwościach i liczbie obsługiwanych systemów operacyjnych.

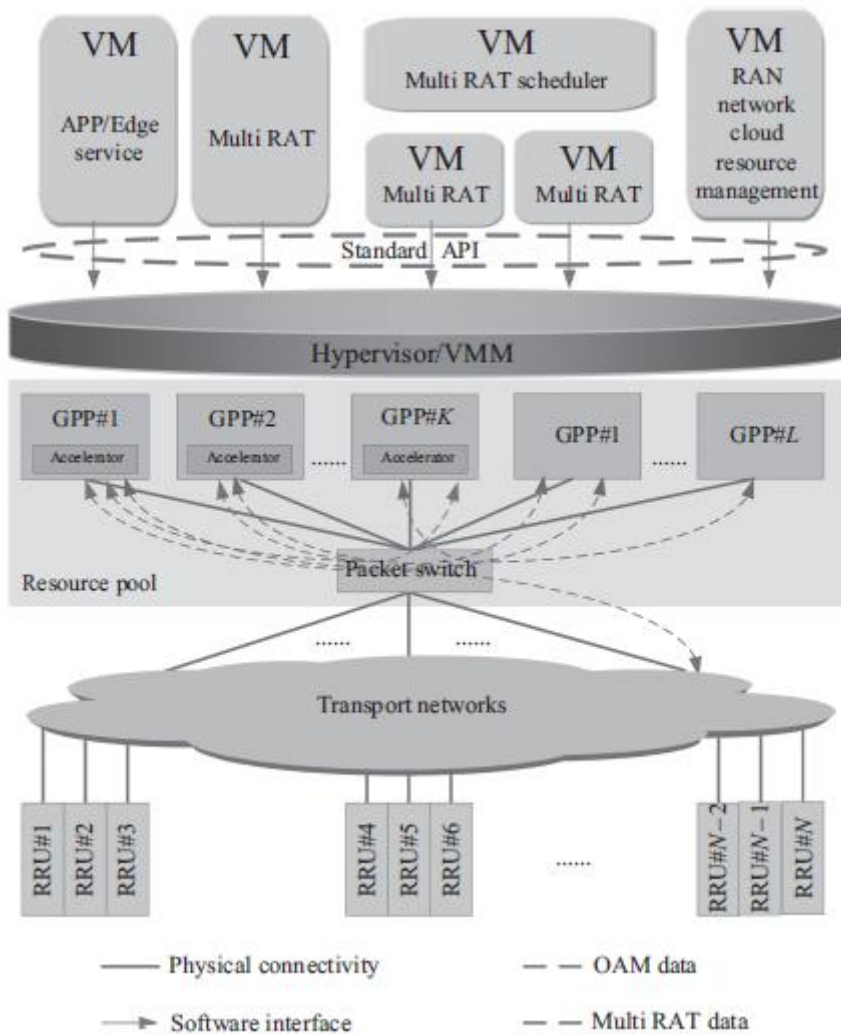
### **10.3 Wirtualizacja Cloud RAN i jej wyzwania**

Z jednej strony operatorzy stają przed wyzwaniami narzucanymi przez coraz bardziej ruch mobilny. Z drugiej strony jest świadkiem, że technologia wirtualizacji wykroczyła poza branżę IT i rozszerzyła swój wpływ na kilka innych branż. Ludzie zaczynają myśleć o możliwości wprowadzenia wirtualizacji do branży telekomunikacyjnej. W 2012 r. Utworzono Industry Specification Group (ISG) o nazwie Network Functions Virtualization (NFV) pod patronatem European Telecommunications Standards Institute (ETSI). Podstawową ideą NFV jest "konsolidacja wielu typów urządzeń sieciowych na standardowe, o wysokim natężeniu serwery, przełączniki i pamięć masową, które mogą znajdować się w centrach danych, węzłach sieci oraz w pomieszczeniach użytkownika końcowego ". Ludzie oczekują, że wirtualizacja może przynieść operatorom takie korzyści, jak szybszy proces rozwoju, uaktualnienie oprogramowania podobne do oprogramowania, mniej skomplikowane zarządzanie, szybszy czas wejścia na rynek i lepszą moc wydajność i tak dalej. Sieci bezprzewodowe składają się głównie z sieci rdzeniowej i sieci dostępu radiowego (RAN). Jest oczywiste, że RAN jest znacznie bardziej skomplikowany niż sieć rdzeniowa z punktu widzenia obliczeń i wymagań w czasie rzeczywistym. W związku z tym istnieje szeroki konsensus w branży, że realizacja wirtualizacji RAN jest znacznie trudniejsza niż w przypadku sieci rdzeniowej. W tej części skupimy się głównie na wirtualizacji RAN. W szczególności skupimy się na Cloud RAN (C-RAN), która jest niezbędną architekturą w stosunku do 5G.

#### **10.3.1 Podstawowy C-RAN**

Koncepcja C-RAN została po raz pierwszy zaproponowana przez China Mobile w białej księdze w 2009 roku. "C" tutaj oznacza "Scentralizowany, współpracujący, Cloud i Clean". Podstawowa idea C-RAN rozpoczyna się od centralizacji, która polega na agregowaniu różnych jednostek pasma podstawowego (BBU), które w tradycyjnym wdrożeniu są geograficznie rozdzielone na tę samą lokalizację. Po scentralizowaniu możliwe byłoby, aby różne jednostki BBU komunikowały się ze sobą w bardziej punktualny sposób, łącząc je z szybkimi sieciami przełączników, umożliwiając w ten sposób implementację algorytmów współpracy w celu poprawy wydajności systemu. Prosta centralizacja nie obsługuje udostępniania zasobów między różnymi fizycznymi jednostkami BBU. Dzięki wirtualizacji to się wkrótce zmieni. Ostatecznym celem C-RAN jest realizacja tak zwanej cechy " zasocjowania zasobów ", która polega na połączeniu zasobów przetwarzania pasma podstawowego, aby zasoby mogły być zarządzane i dynamicznie przydzielane na żądanie. Wirtualizacja jest kluczem do realizacji procesu chmury zasobów. Wirtualizacja sieci RAN ma na celu architekturę sieci RAN z ewoluującym standardem IT technologii wirtualizacji w celu konsolidacji tradycyjnych typów urządzeń sieciowych na standardowych serwerach, przełącznikach i pamięciach o dużej pojemności, które będą znajdować się w witrynach bezprzewodowych lub biurach centralnych. Polega ona na implementacji części funkcji RAN w oprogramowaniu, które może działać na szeregu standardowych urządzeń serwerowych i które mogą być przenoszone do lub lokalizowane w różnych lokalizacjach w sieci zgodnie z wymaganiami,

bez konieczności instalowania nowych sprzęt. Rysunek 10.2 ilustruje zwirtualizowany system C-RAN oparty na platformie Commercialoffs-shelf (COTS).



Dzięki wirtualizacji standardowe serwery IT są używane jako ogólna platforma z obliczeniami i pamięcią masową jako wspólne zasoby. Jak pokazano na Rysunku 10.2, na górze serwerów działają różne aplikacje w postaci maszyn wirtualnych (VM). Niezbędne aplikacje w C-RAN to te, które realizują różne technologie dostępu radiowego (RAT), w tym 2G, 3G, 4G i przyszłe 5G. Dodatkowe aplikacje użytkownika, takie jak sieć dostarczania treści, mogą być również wdrażane na tej samej platformie.

### 10.3.2 Wyzwania związane z wirtualizacją C-RAN

Komunikacja bezprzewodowa, w szczególności przetwarzanie pasma podstawowego w RAN, różni się od centrów danych IT tym, że komunikacja bezprzewodowa ma niezwykle surowe wymagania dotyczące przetwarzania w czasie rzeczywistym. Na przykład, w przypadku systemów LTE wymagane jest, aby sygnał ACK / NACK był wytwarzany i wysyłany z powrotem do UE / eNB w ciągu 3 ms po odebraniu ramki. Tradycyjna technologia wirtualizacji centrów danych z procesorem ogólnego zastosowania (GPP) nie może spełnić tego wymogu. Dlatego zastosowanie wirtualizacji w stacjach bazowych wymaga starannego projektowania i specjalnej optymalizacji kluczowych bloków funkcjonalnych. Aby wdrożyć wirtualne stacje bazowe w czasie rzeczywistym w scentralizowanej puli pasm podstawowych, należy rozwiązać następujące problemy:

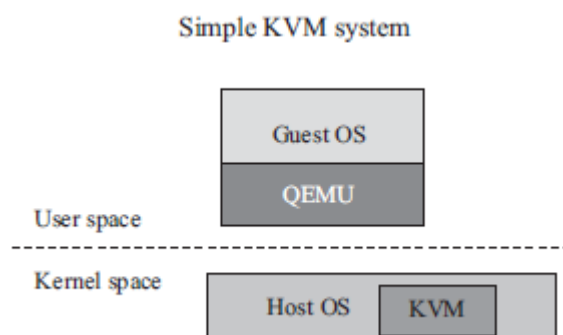
- Procesor ogólnego przeznaczenia i zaawansowany algorytm przetwarzania do przetwarzania sygnałów w czasie rzeczywistym.
- Topologia międzysystemowa BBU o wysokiej przepustowości, niskim opóźnieniu i niskich kosztach wśród fizycznych zasobów przetwarzania w puli pasma podstawowego. Obejmuje on połączenie między układami w BBU, pomiędzy BBU w stelażu fizycznym, a także pomiędzy wieloma stelażami.
- Zdolność tradycyjnej platformy COTS do spełnienia wymagań RAN w czasie rzeczywistym. Obejmuje to optymalizację różnych kluczowych modułów, w tym systemu operacyjnego, hipernadzorca, operacji we / wy, sieci przełączników, systemów zarządzania i tak dalej. Szczegółowe omówienie rozwiązań znajdziecie w następnym rozdziale.
- Wyzwaniem jest również badanie ziarnistości wirtualizacji. Ze względu na różnice w protokołach komunikacyjnych wymagania dotyczące zasobów przetwarzania dla GSM, TD-SCDMA i TD-LTE różnią się znacznie pod względem możliwości obliczeniowych i wydajności w czasie rzeczywistym. W praktyce należy rozważyć różne różnice w wirtualizacji. Na przykład w praktycznych wdrożeniach maszyna wirtualna może obsługiwać tylko jednego przewoźnika w ramach BBU lub nawet mniejszych jednostek. Istnieje inne podejście warstwowe, w którym maszyna wirtualna obsługuje tylko jedną warstwę, taką jak L1, L2 lub L3, lub nawet pewną część warstwy nośnej.

#### 10.4 Najważniejsze technologie i rozwiązania wirtualizacji RAN

Aby osiągnąć wirtualizację RAN, należy zająć się wieloma wyzwaniami technologicznymi. Z punktu widzenia sprzętu i oprogramowania platforma wirtualizacji powinna mieć określone ulepszenia i modyfikacje. Badania i doświadczenie wyraźnie pokazały, że standardowe rozwiązania informatyczne klasy korporacyjnej i korporacyjnej nie zapewniają determinizmu, wydajności i niezawodności wymaganych do skutecznej implementacji RAN - wymagane jest specjalnie zbudowane, zoptymalizowane rozwiązanie. Następujące kluczowe technologie są ważne dla osiągnięcia skutecznej wirtualizacji RAN.

##### 10.4.1 Rozszerzenie na hostowym systemie Linux

Najpierw wprowadzimy Wirtualną Maszynę (KVM) opartą na jądrze. KVM jest przeznaczony dla architektur x86 i x86\_64, aby korzystać z natywnego wsparcia dla wirtualizacji sprzętowej, podczas gdy KVM obsługuje również inne architektury procesorów, takie jak ARM i Power. KVM opiera się na emulacji Quick Emulator (QEMU), aby zapewnić kompleksowe rozwiązanie wirtualizacji. Proste użycie dla KVM jest pokazane na rysunku 10.3.



Wirtualizacja RAN jest bardziej złożonym scenariuszem użycia niż niektórymi aplikacjami sieci rdzeniowej, ponieważ istnieje kilka maszyn wirtualnych lub systemu-gościa, a wydajność jest bardzo wrażliwa na rygorystyczne wymagania (trudne w czasie rzeczywistym, wymagające przetwarzania co 500 ms bez jittera). Podobnie, wydajność bez determinizmu jest bezużyteczna w tym kontekście.

Kompletne zwirtualizowane rozwiązanie musi zapewniać ten sam poziom wydajności i gwarantowaną przewidywalność dla każdego obciążenia związanego ze zwirtualizowaną aplikacją RAN. Należy zająć się następującymi ograniczeniami:

- Obsługa KVM w czasie rzeczywistym i QEMU.
- Lepsza izolacja gości dzięki technikom minimalizacji interferencji procesorów.
- Udoskonalona konfiguracja systemu, partycjonowanie i izolacja maszyn wirtualnych i środowisk z wykorzystaniem elastycznych funkcji skryptowania.
- Dobrze zdefiniowane konteksty bezpieczeństwa dla maszyn wirtualnych i środowisk.
- Obsługa pełnego NO\_HZ (jądro łaskocze). Zestaw poprawek do jądra, aby poprawić izolację określonego procesora ze źródeł zakłóceń systemu, takich jak Inter Processor Interrupts (IPI), tykanie zegara, odzyskiwanie pamięci i inne. Modyfikacje globalnych maszyn stanów (RCU) i usług (pomiar czasu) są wykonywane bez okresowego dostępu do rdzenia.

Ponadto wciąż istnieje wiele innych kluczowych parametrów systemu, które są nazywane opcjami w systemie Linux i mogą wpływać na wydajność w czasie rzeczywistym. W przypadku innego sprzętu i oprogramowania konieczne będzie wykonanie ustawień systemu. Opcje są przekazywane do jądra przez program rozruchowy grub i pojawiają się w pliku / proc / cmdline, gdy cel zakończy proces uruchamiania.<sup>1</sup> W przypadku jądra gościa są one przekazywane do binarnego Qemu-Kvm podczas uruchamiania czas. Poniżej wymieniono najważniejsze opcje, które mogą mieć wpływ na wydajność systemu w czasie rzeczywistym:

### **1. zegar źródłowy**

Służy do przesłonięcia domyślnego źródła zegara, a jiffie to źródło zegara cofania. W architekturze x86\_64 dostępne są opcje hpet i tsc. Podczas gdy QEMU może emulować źródło zegara HPET dla gościa, OVP używa tsc jako źródła zegara, aby uniknąć narzutu emulacji.

### **2. beczynny**

Służy do definiowania zachowania podczas uśpienia procesora, gdy jest on beczynny. Poprawne parametry to pool, mwait, stop i nomwait. Podczas korzystania z odpytywania, beczynne zadanie powoduje przebudzenie CPU przez okresowe pobieranie, uniemożliwiając przejście do trybu oszczędzania energii. Zmniejsza to opóźnienia związane z przełączaniem kontekstów kosztem większego zużycia energii. Podczas używania MWAIT, beczynne zadanie używa specjalnych instrukcji MONITOR / MWAIT, które osiągają prawie taką samą wydajność opcji odpytywania, ale przy mniejszym zużyciu energii. Wszystkie nowsze procesory Intel'a obsługują te instrukcje. Opcja nomwait pozwala CPU spać w stanach oszczędzania energii C, ale nie stosuje instrukcji MONITOR / MWAIT. Zwróć uwagę, że od czasu Linuksa 3.0 pojawiło się ostrzeżenie o czasie rozruchu po wywołaniu idle ¼mwait, co oznacza, że ta opcja zostanie usunięta w przyszłych wydaniach. Tak właśnie jest, ponieważ jądro 3.9. Jądra skonfigurowane za pomocą CONFIG\_ACPI=n i CONFIG\_INTEL\_IDLE=n po uruchomieniu na sprzęcie obsługującym MWAIT będą po prostu używać HLT. Jeśli MWAIT jest pożądany w tych systemach, te dwie opcje konfiguracji powinny być włączone.

### **3. Niepewność**

Służy do ustawienia domyślnej maski powinowactwa IRQ, czyli zestawu procesorów, które powinny przetwarzać przerwania. Poprawne parametry to lista, zakres lub kombinacja identyfikatorów procesora. Na przykład, jeśli na 16 rdzennych użytkownikach komputerów chcemy zapobiec

przerywaniu przejść do procesorów 1, 2 i 3. Mogliby użyć wartości zakłóceń =0,4-15 na linii startowej hosta. Opcja `irqaffinity` jest szczególnie wygodna w połączeniu z opcją `isolcpus`, aby w pełni wyizolować jeden lub więcej procesorów z planera jądra systemu Linux i podsystemów zarządzania przerwaniami.

#### **4. `isolcpus`**

Ta opcja służy do określenia jednego lub więcej procesorów w celu odizolowania od ogólnych algorytmów równoważenia i planowania SMP. Użytkownicy mogą przenosić proces na izolowany procesor lub z niego, korzystając z wywołań systemowych powinowactwa CPU lub poleceń zestawu zadań i `cpuset`. Numer procesora zaczyna się od 0, a maksymalna to liczba procesorów w systemie -1. Ta opcja jest również używana w wierszu poleceń jądra QEMU-KVM, aby wyizolować wirtualne procesory, które mają być używane do określonych zadań. Na przykład dla dwóch instancji gościa wirtualnego `isolcpus=1` pozostawiłoby zerowe wirtualne CPU dla zadań porządkowych, podczas gdy procesor wirtualny 1 będzie przeznaczony tylko dla tych zadań jawnie ustawionych do działania na nim. Ogólnie, uzyskanie pełnej izolacji procesora nie jest całkowicie możliwe. W szczególności niektóre zadania jądra, takie jak miękkie przetwarzanie IRQ przez wątek `ksoftirqd`, będą prawdopodobnie zawsze obecne. Lista `ksoftirqd` dozwolona dla procesora w jądrze jest zasadniczo tylko do odczytu.

#### **5. `rcu-nocbs`**

Służy do zapobiegania wykonywaniu procedur zwrotnych RCU w docelowych procesorach. Poprawne parametry to lista, zakres lub kombinacja identyfikatorów procesora. Połączenia zwrotne RCU to funkcje, które wykonują porządki po upływie okresu prolongaty RCU. W przypadku niektórych obciążeń liczba wywołań zwrotnych może być dość duża. Na przykład, aby zażądać, aby procesory 1, 2, 3 i 6 nie były używane do wykonywania wywołań zwrotnych RCU, użytkownicy mogą używać `rcu-nocbs=1-3,6` na linii startowej hosta. Nie jest możliwe przypisanie każdego procesora w systemie do listy bez wywołań zwrotnych; co najmniej jeden procesor, CPU 0, musi pozostać w trybie tradycyjnym lub przetwarzanie okresu próbnego RCU nie będzie działać poprawnie. Zazwyczaj zakres `rcu-nocbs` będzie pasował do parametrów `isolcpus`, aby dodatkowo poprawić stan izolacji docelowych procesorów.

#### **6. `rcu_nocb_poll`**

Służy do wymuszania trybu odpytywania na wątkach RCU, które obsługują wykonywanie wywołań zwrotnych RCU. Ta opcja nie ma parametrów. Kiedy procesor działa bez wywołań zwrotnych RCU, będzie osobny proces obciążony obsługą wywołania zwrotnego. Gdy proces ten się obudzi, pobierze listę wyjątkowych wywołań zwrotnych dla przypisanego procesora, poczeka, aż upłynie okres prolongaty RCU, a następnie przejdzie przez wszystkie wywołania zwrotne. Opcja rozruchu `rcu_nocb_poll` nakazuje temu procesowi odpytywanie od czasu do czasu za nową pracę. Tryb odpytywania przyczynia się do zmniejszenia obciążenia odciążonych procesorów, które wciąż wykonują wywołania zwrotne RCU, oszczędzając im starań, aby regularnie budzić procesy czyszczenia.

#### **7. `tsc`**

Służy ona do przyjmowania założeń dotyczących dostępności i zachowania licznika znaczników czasu. Użycie `tsc=perfect` pozwala uniknąć niektórych sprawdzeń wykonawczych wykonywanych na niezawodności TSC. Jest używany w hostach i gościach.

### **10.4.2 Poprawa komunikacji VM**

Komunikacja między VM jest bardzo ważna dla platformy wirtualizacji. Jedną z wyróżniających właściwości wirtualizacji jest możliwość izolowania maszyn wirtualnych współlokatorów na tej samej

fizycznej platformie. Te wirtualne maszyny często muszą się komunikować i wymieniać znaczną ilość danych. Wiele aplikacji RAN ma różne tryby działania w obrębie węzła, które muszą wymieniać wiadomości i dane podczas trwania połączenia głosowego lub transmisji danych. Ważne jest, aby dostawcy RAN mogli zminimalizować zmiany w oprogramowaniu, aby to samo oprogramowanie mogło działać na dedykowanym sprzęcie lub w maszynach wirtualnych, prowadząc do wymogu wysokiej wydajności przesyłania komunikatów między urządzeniami VM. Wydajność będzie wąskim gardłem, jeśli przepustowość lub opóźnienie nie spełni wymagań wirtualizacji RAN. Ponieważ RAN nie jest taki sam jak dotychczasowy system informatyczny, wymagania są niezwykle surowe.

Istnieje wiele rodzajów metod komunikacji między VM, w tym

- Technika współdzielonej pamięci
- Technika oparta na sieci
- \* Przełącznik wirtualny hosta
- \* Przełącznik asystenta sprzętowego

Tutaj skupimy się na komunikacji sieciowej. Główną przyczyną jest to, że korzystanie z komunikacji sieciowej wymaga bardzo niewielu modyfikacji maszyny wirtualnej.

Standardowy KVM i QEMU ma obsługiwane niektóre funkcje, takie jak

- Urządzenia Virtio. VirtIO to specyfikacja parowo-wirtualnych operacji we / wy na maszynie wirtualnej. Tradycyjnie hiperwizor emuluje prawdziwe urządzenia, takie jak interfejs Ethernet lub kontroler dysków, aby zapewnić maszynę wirtualną we / wy. Ta emulacja jest często nieefektywna.
- Vhost. Sterowniki vhost w systemie Linux zapewniają emulację urządzenia virtio w jądrze. Normalnie proces przestrzeni użytkownika QEMU emuluje dostęp I / O z gościa. Vhost umieszcza kod emulacji virtio w jądrze, usuwając przestrzeń użytkownika QEMU poza obraz. Dzięki temu kod emulacji urządzenia może bezpośrednio wywoływać podsystemy jądra zamiast wykonywania wywołań systemowych z przestrzeni użytkownika.

Obecnie dostępnych jest kilka wirtualnych przełączników. Podstawowym, domyślnym wyborem jest stabilny, ale słabo działający Open vSwitch, który jest open source. Istnieją również niektóre komercyjne przełączniki wirtualne, takie jak wydajny przełącznik Wind River Accelerated vSwitch, zintegrowany z platformą infrastruktury NFV Titanium Server.

Aby zapewnić wysoką przepustowość między sieciami VM, przełączniki wirtualne mogą mieć następujące funkcje:

- Obsługa pakietów przestrzeni użytkownika
- Dedykowany rdzeń procesora przypisany

Aby zapewnić sieć VM o małym opóźnieniu, przełączniki wirtualne mogą mieć następujące funkcje:

- Sterownik trybu pobierania
- Brak projektu systemu zamków

#### **10.4.3 Ulepszanie w czasie rzeczywistym**

Aby osiągnąć wysoką wydajność i przewidywalność, dystrybucja systemu operacyjnego musi zostać zbudowana i przetestowana przy użyciu określonego zestawu parametrów konfiguracyjnych i tuningowych:

- Preempt-RT powinien obsługiwać jądro hosta i gościa: Zestaw łat jądra dla jądra Linux w celu poprawy determinizmu i zachowania w czasie rzeczywistym poprzez zmiany semantyczne blokad, modyfikacje algorytmów i zmiany sterowników. Wszystkie poziomy pierwszeństwa są obsługiwane do wartości PREEMPT\_RT\_FULL w zależności od przypadku użycia.
- Sterowniki i funkcje w czasie rzeczywistym: adaptacje typowych narzędzi i operacji do pracy w semantyce blokującej preempt-rt.
- Pełny NO\_HZ: zestaw poprawek jądra, które poprawiają izolację określonego procesora ze źródeł zakłóceń systemu, takich jak Inter Interruptor (IPI), tykanie zegara, odzyskiwanie pamięci i inne. Modyfikacje globalnych maszyn stanów (RCU) i usług (pomiar czasu) są wykonywane bez okresowego dostępu do rdzenia.
- Ulepszona obsługa KVM i KVM w czasie rzeczywistym: Oprócz podstawowej funkcjonalności jądra rteci, obsługa gości i hosta KVM została zoptymalizowana pod kątem opóźnień przerwania i izolacji procesu. Pozwala to maszynom wirtualnym na działanie z wydajnością zbliżoną do wydajności macierzystego BSP, z dodatkowym zabezpieczeniem, izolacją, abstrakcją, emulacją i przyspieszeniem przy jednoczesnym wykorzystaniu obsługi funkcji KVM, takich jak migracja na żywo

#### **10.4.4 Migracja na żywo**

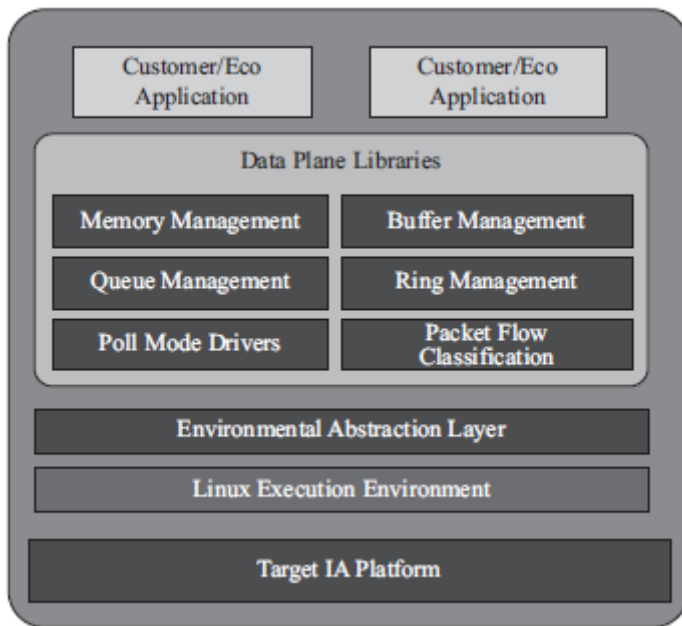
Migracja na żywo odnosi się do procesu przenoszenia uruchomionej maszyny wirtualnej lub aplikacji między różnymi maszynami fizycznymi lub między różnymi procesorami na tej samej maszynie fizycznej bez zatrzymywania maszyny wirtualnej. Pamięć, pamięć i łączność sieciowa maszyny wirtualnej są przenoszone z oryginalnego gościa maszyna do miejsca przeznaczenia. Migracja może być używana w następujących przypadkach użycia:

- Równoważenie obciążenia - goście mogą zostać przeniesieni do hostów z mniejszym użyciem, gdy host zostaje przeciążony.
- Sprzętowe przełączanie awaryjne - gdy urządzenia sprzętowe na hoście zaczynają się zawieszać, goście mogą zostać bezpiecznie przeniesieni, aby host mógł zostać wyłączony i naprawiony.
- Oszczędność energii - goście mogą być redystrybuowani do innych hostów i systemów hostów, aby oszczędzać energię i obniżyć koszty w krótkich okresach użytkowania.
- Migracja geograficzna - goście mogą zostać przeniesieni do innej lokalizacji w celu zmniejszenia opóźnień lub w krytycznych okolicznościach.

Aby uzyskać migrację na żywo, stan systemu-gościa na hoście źródłowym musi być replikowane na hoście docelowym. Wymaga to migracji stanu procesora, pamięci zawartość, pamięć lokalna i stan sieci. Na polu wirtualizacji RAN istotna jest przerwa w migracji na żywo. Migracja na żywo nie powinna wpływać na jakość użytkownika końcowego, a połączenia i dane nie powinny być odrzucane. Aby uzyskać najlepszą wydajność migracji na żywo, można zastosować specjalną technikę migracji pamięci. Coraz ważniejszym czynnikiem, który należy wziąć pod uwagę przy wyborze platformy wirtualizacji do wdrożenia vRAN, jest obsługa migracji na żywo aplikacji opartych na Intel Data Plane Development Kit (DPDK). Ponieważ coraz więcej aplikacji jest budowanych za pomocą DPDK, migracja na żywo dla tych aplikacji zostanie zablokowana, chyba że rozwiązanie wirtualizacji wyraźnie obsługuje tę funkcję.

#### 10.4.5 Przyspieszenie sieci

W celu przyspieszenia sieci skupimy się na zestawie do programowania płaszczyzn danych (DPDK). DPDK to zestaw bibliotek i sterowników do szybkiego przetwarzania pakietów. Został zaprojektowany do pracy na dowolnym procesorze. Pierwszym obsługiwanym procesorem był Intel x86, który został rozszerzony na IBM Power 8, EZchip TILE-Gx i ARM. Działa głównie w przestrzeni użytkownika Linuksa. Port FreeBSD jest dostępny dla podzbioru funkcji DPDK. DPDK jest projektem licencjonowanym na licencji Open Source BSD



Głównym celem DPDK jest zapewnienie prostej, kompletnej struktury do szybkiego przetwarzania pakietów w zastosowaniach do płaszczyzn danych. DPDK implementuje model "run-to-completion" do przetwarzania pakietów, w którym wszystkie zasoby muszą zostać przydzielone przed wywołaniem aplikacji Płaszczyzny danych, działając jako jednostki wykonawcze na logicznych rdzeniach przetwarzania. Model nie obsługuje programu planującego, a do wszystkich urządzeń można uzyskać dostęp za pomocą odpytywania. Głównym powodem niewykorzystywania przerwań jest ogólny nakład pracy związany z przetwarzaniem przerwań. Oprócz modelu "run-to-completion" można również zastosować model potokowy, przekazując pakiety lub wiadomości między rdzeniami za pośrednictwem pierścieni. Pozwala to na wykonywanie etapów i pozwala na bardziej efektywne wykorzystanie kodu na rdzeniach. Przełącznik Intel DPDK vSwitch jest rozwidleniem wielowarstwowego wirtualnego przełącznika Open vSwitch. Reimplementuje moduł przesyłania dalej jądra (płaszczyznę danych), budując logikę przełączania w górnej części biblioteki Intel DPDK. Znacząco zwiększa to przepustowość przełączania pakietów i przybliża Open vSwitch do ogólnej wydajności komercyjnych rozwiązań vSwitching. Niestety ogólna implementacja jest wciąż niedojrzała, cierpiąca z powodu ograniczeń użyteczności i niezawodności. Chociaż nie jest jeszcze gotowy do użytku komercyjnego, wygląda obiecująco. Silnik spedycyjny zawiera tablice dużych tabel Intel DPDK, a demon kontrolny Open vSwitch jest modyfikowany w celu połączenia z tymi tabelami. Moduł przesyłania działa w przestrzeni użytkownika Linux z prawami licencyjnymi BSD. Intel DPDK vSwitch implementuje podzbiór funkcji przełączania Open vSwitch.

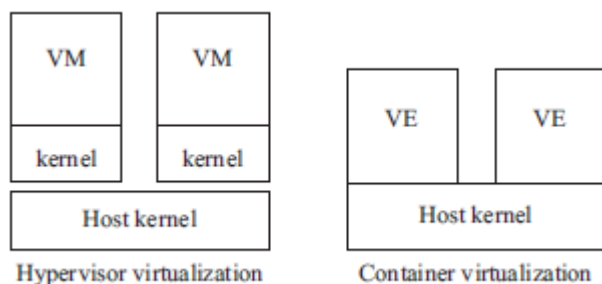
#### 10.4.6 Pojemnik



Kontenery są wynikiem wirtualizacji i separacji systemu operacyjnego. Zapewniają lekkie wirtualne środowisko, które grupuje i izoluje zestaw procesów i zasoby takie jak pamięć, procesor, dysk itp. z hosta i innych kontenerów. Podstawowe technologie kontenerowe obejmują:

- cgroups
- przestrzenie nazw
- chroot
- Linux Security Module (LSM), obowiązkowa kontrola dostępu (MAC)
- Możliwości / SECCOMP
- Tradycyjna wirtualizacja
- Runtime / management

Zrozumienie różnic między hiperwizorem a wirtualizacją kontenerów może lepiej przygotować użytkowników do projektowania i rozwijania wirtualizacji opartych na systemie końcowym. Dwa wirtualne węzły są przedstawione na rysunku 10.5



„jedna uruchomiona wirtualizacja hiperwizora z dwoma maszynami wirtualnymi i inna wirtualizacja kontenera z dwoma wirtualnymi środowiskami (VE). Wspólne, współdzielone jądro Linux zapewnia:

- Narzędzia do zarządzania przestrzenią użytkownika dla funkcji zabezpieczania systemu Linux.
- Izolacja zasobów sprzętowych (procesor, pamięć, blokowe operacje we / wy, sieć itd.) Bez konieczności korzystania z pełnych maszyn wirtualnych.
- Izolacja widoku aplikacji dla zasobów O / S (drzewa procesów, sieć, identyfikatory użytkowników i zamontowane systemy plików).

Technologia pojemników jest lekka niż wirtualizacja poziomu VM. Ponieważ żaden hiperwizor nie jest używany, wszystkie kontenery mają wspólne jądro hosta.

#### 10.4.7 CAT / CMT

Cache Allocation Technology (CAT) to nowa funkcja umożliwiająca hiperwizor systemu operacyjnego lub VMM, aby kontrolować przydział współużytkowanej pamięci podręcznej ostatniego poziomu procesora. Po skonfigurowaniu CAT procesor umożliwia dostęp do części pamięci podręcznej zgodnie z ustaloną klasą usług (COS). Procesor przestrzega zasad COS, gdy uruchamia wątek aplikacji lub proces aplikacji. Można to osiągnąć, wykonując następujące czynności:

- Ustal, czy CPU obsługuje funkcję CAT.

- Skonfiguruj COS, aby określić ilość dostępnych zasobów (przestrzeń pamięci podręcznej). Ta konfiguracja jest na poziomie procesora i jest wspólna dla wszystkich procesorów logicznych.
- Powiązać każdy procesor logiczny z dostępnym COS.
- Uruchom aplikację na procesorze logicznym, który używa żądanego COS.

Cache Monitoring Technology (CMT) to nowa funkcja, która pozwala systemowi operacyjnemu lub hypervisor lub monitorowi maszyny wirtualnej (VMM) na określenie użycia pamięci podręcznej dla aplikacji działających na platformie. CMT może być używany do wykonywania następujących czynności:

- Aby wykryć, czy platforma obsługuje tę funkcję monitorowania;
- W przypadku systemu operacyjnego lub maszyny wirtualnej, aby przypisać identyfikator dla każdej aplikacji lub maszyn wirtualnych, które są zaplanowane do uruchomienia na rdzeniu. Ten identyfikator jest nazywany identyfikatorem monitorowania zasobów (RMID);
- Aby monitorować obciążenie pamięci podręcznej na podstawie RMID, oraz
- W przypadku systemu operacyjnego lub maszyny wirtualnej odczytać zajętość pamięci podręcznej ostatniego poziomu dla danego RMID w dowolnym momencie.

#### **10.4.8 Obsługa OpenStack**

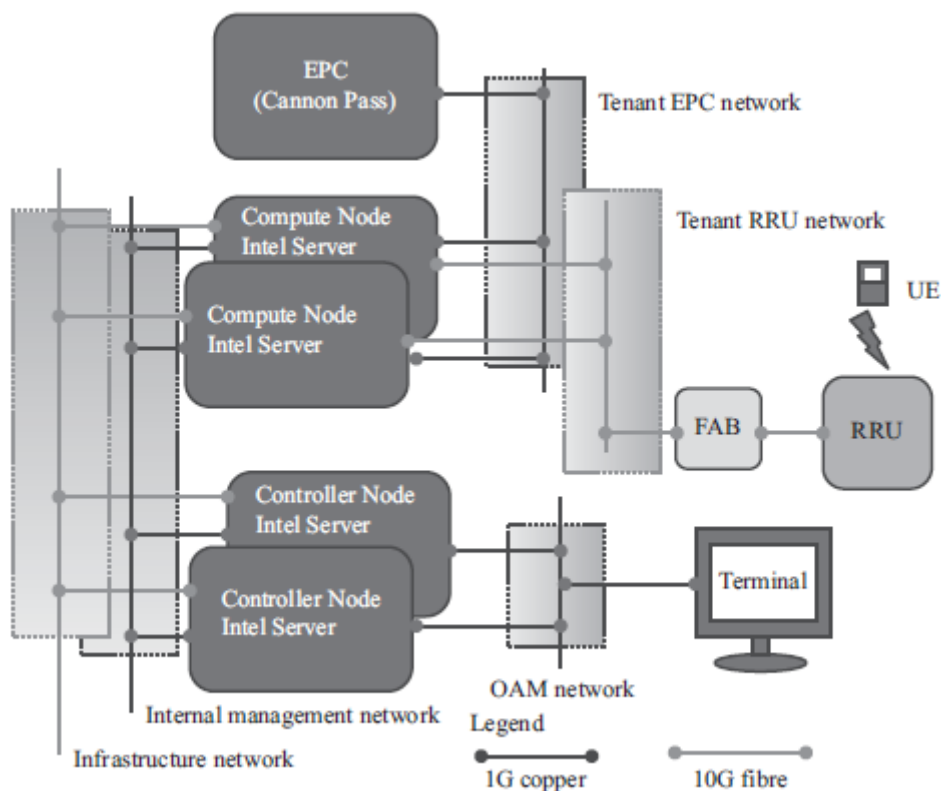
OpenStack [11] to darmowa i otwarta platforma oprogramowania do przetwarzania w chmurze, najczęściej wdrażane jako infrastruktura jako usługa (IaaS). Platforma oprogramowania składa się ze wzajemnie powiązanych komponentów kontrolujących puli sprzętowe przetwarzania, zasoby pamięci i sieci w całym centrum danych. OpenStack zawiera wiele podstawowych usług, takich jak Nova, Swift, Neutron itp., które mają kluczowe znaczenie dla wirtualizacji RAN.

#### **10.5 Proof-of-Concept rozwoju zwirtualizowanego RAN**

W celu oceny wykonalności wirtualizacji RAN utworzono prototyp z różnymi funkcjami optymalizacyjnymi zaimplementowanymi na platformie wirtualizacji, która została opisana w tej sekcji.

##### **10.5.1 Zwirtualizowany system demonstracyjny RAN**

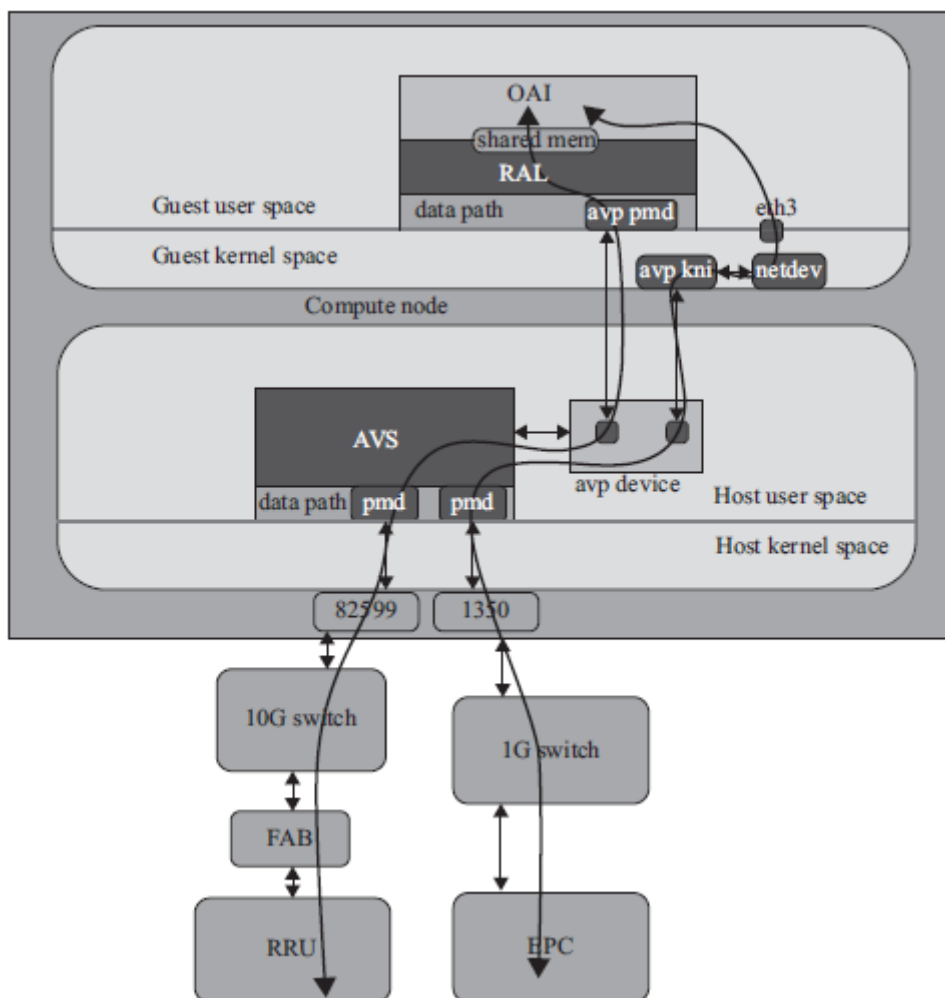
Rysunek 10.6 przedstawia topologię konfiguracji Proof-of-Concept (PoC) C-RAN.



Dwa węzły kontrolera działają w usługach Titanium Server w celu zarządzania infrastrukturą chmury. Dwa węzły obliczeniowe działają na serwerze Titanium, aby zapewnić komputer usługi i host maszyn wirtualnych. System wykorzystuje serwer Wind River Titanium jako infrastrukturę NFV z następującymi funkcjami:

- Automagiczne odzyskiwanie maszyny wirtualnej po awarii węzła obliczeniowego hosta (awaria węzła wykrywanie w sekundach zamiast w minutach typowych dla standardowej klasy IT)
- Automagiczne odzyskiwanie maszyny wirtualnej w przypadku awarii maszyny wirtualnej (wykrywanie awarii maszyny wirtualnej 60x szybciej niż w przypadku standardowego poziomu IT)
- Migracja na żywo maszyn wirtualnych za pomocą procesora Intel DPDK (nie dostępne w OpenStack klasy IT)
- Redundancja węzła sterownika i automagiczne przełączanie awaryjne (nie dostępne w wersji opartej na technologii IT OpenStack)
- Monitorowanie VM związane z kontrolą stanu aplikacji w obrębie gościa (zapewnia precyzyjne wykrywanie błędów aplikacji)
- Grupy zabezpieczeń VM (zapewniające tworzenie maszyn wirtualnych z tej samej grupy na różnych węzłach obliczeniowych)

Przepływ danych jest pokazany na rysunku 10.7.



W węzle obliczeniowym, Titanium Server Accelerated vSwitch (AVS) odpowiada za cały ruch Ethernet. Wysoko wydajny vSwitch przestrzeni użytkownika oparty na DPDK umożliwia wysoką wydajność komunikacji VM-VM bez potrzeby korzystania z jądra Linux, a także wysokowydajne przetwarzanie pakietów z karty sieciowej do aplikacji w maszynach wirtualnych. Ethernet z obsługą Intel DPDK, SRIOV oraz 1G, 10G i 40G zapewnia ultraszybkie przetwarzanie pakietów. Kluczowe moduły na rysunku obejmują:

- Accelerated Virtual Port (AVP) to przyspieszony interfejs sieci Ethernet typu TiS dla zwiększenia przepustowości.
- avp pmd to urządzenie, które jest sterownikiem trybu sond kompatybilnym z Intel DPDK.
- avp\_kni jest urządzeniem zgodnym ze standardowym urządzeniem sieciowym jądra Linux.

Każde wirtualne urządzenie Ethernet w maszynie wirtualnej będzie miało odpowiednie urządzenie w systemie AVS (takie jak "urządzenia avp"). Cały ruch jest dwukierunkowy. Weźmy jako przykład przykład dla uplink. Sygnał ze zdalnej jednostki radiowej jest przekształcany w pakiety Ethernet. Następnie przechodzi w przełącznik 10G. Gdy węzeł obliczeniowy 82599 NIC odbierze pakiety, zostanie on odpytany przez AVS. AVS podejmie decyzję w oparciu o MAC / IP, a następnie przekaże ją do dedykowanej wirtualnej karty sieciowej maszyny wirtualnej. Po zakończeniu procesu VM pakiety zostaną przekazane do sieci szkieletowej. Pakiety przechodzą do określonej wirtualnej karty sieciowej, a następnie wprowadzają ponownie AVS. AVS podejmie decyzję w oparciu o MAC / IP, a następnie

przekazuje pakiet do I350, kolejnej karty sieciowej. W tym eksperymentalnym środowisku serwer HP DL380 jest używany jako serwer, a inne elementy wyposażenia są wymienione poniżej:

- Serwer: HP DL380 Gen8 z 2 gniazdami, 8 rdzeni na gniazdo,
- Procesor: Intel (R) Xeon (R), E-2690 v2, na 3,00 GHz
- Karta sieciowa (NIC): Intel 82599
- Pamięć: 64 GB (węzeł obliczeniowy)

Oprogramowanie uruchomione w systemie jest wymienione poniżej:

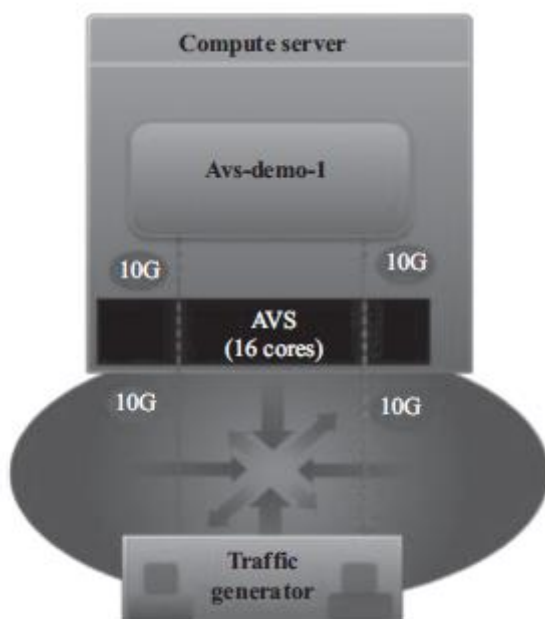
- WindRiver Linux OVP 6 (Open Virtualization Platform)
- Ulepszony kilogram OpenStack
- WindRiver AVS (Advance Vswitch)

### 10.5.2 Wyniki testu

Testy są przeprowadzane w dwóch scenariuszach za pomocą testera Ethernet. W każdym scenariuszu opóźnienie, jitter i przepustowość są odpowiednio mierzone.

#### I. scenariusz vloop

W scenariuszu pokazanym na rysunku 10.8 pakiety są wysyłane z testera Ethernet do systemu operacyjnego.

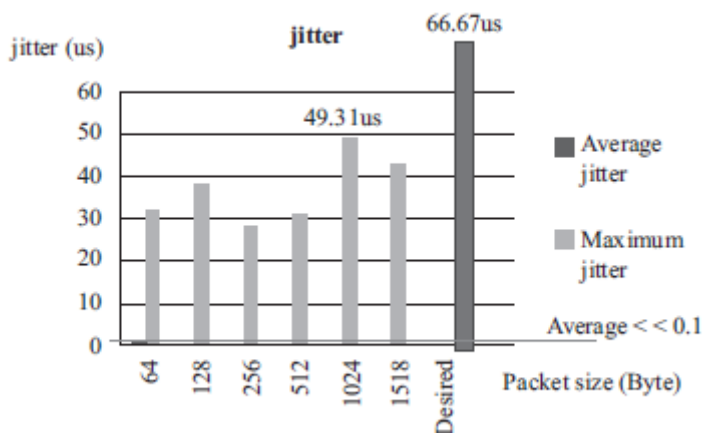
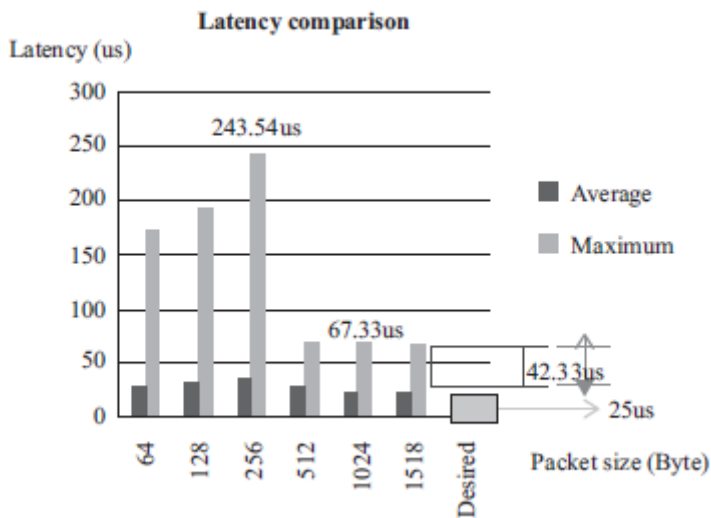


Następnie system hosta przekazuje pakiety do AVS. Gdy pakiety docierają do Avs-demo1, który jest maszyną wirtualną, proces przekazywania L2 odbywa się w ramach demo1. W końcu, po wysłaniu pakietów z vNIC demo1, wrócą oni do testera Ethernet i wykonają analizę pakietów. Przepustowość jest mierzona dla różnych rozmiarów klatki w środowisku wspomnianym powyżej, jak pokazano w Tabeli 10.1.

Rozmiar ramki (bajt) 64 128 256 512 1024 1518

Wydajność (%) 33,03 58,516 99,298 99,298 99,298 99,298

Zgodnie z tabelą, gdy rozmiar ramki staje się większy, przepustowość wzrasta. Powodem jest to, że w przypadku tej samej ilości danych potrzeba więcej pakietów dla mniejszego rozmiaru klatki. System będzie dokładać więcej starań, aby zdemontować pakiety. W rezultacie przepustowość jest niska, gdy rozmiar jest mały. W przypadku różnych rozmiarów ramek opóźnienie pętli pokazano na rysunkach 10.9 i 10.10.

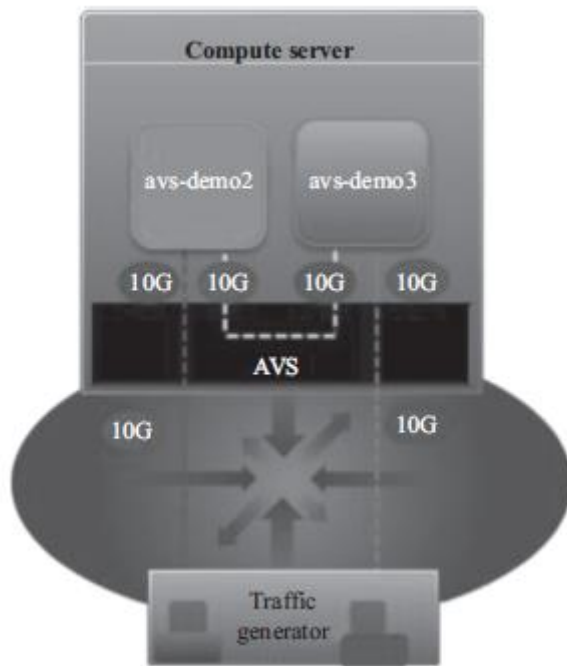


Wykres drgań względem wielkości ramy można zobaczyć na rysunku 10.10. Z rysunków 10.9 i 10.10 można zauważyć, że:

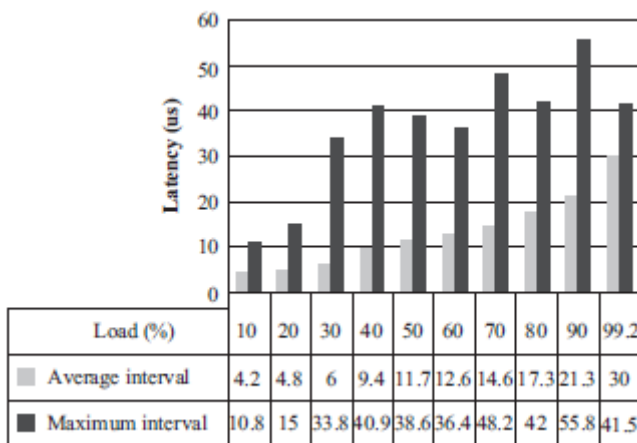
1. Przy zerowym poziomie utraty pakietów, gdy rozmiar pakietu jest nie mniejszy niż 256 bajtów, przepustowość może osiągnąć prawie pełną przepustowość 10 Gb / s.
2. Gdy rozmiar pakietu jest nie mniejszy niż 512 bajtów, opóźnienie transmisji pakietu pozostaje prawie stałe, z maksymalnym opóźnieniem wynoszącym 67.33us.
3. Wszystkie testowane drgania są mniejsze niż 66.67us, co spełnia wymóg komunikacji bezprzewodowej.

## II. Scenariusz modelu M.

Test ma na celu oszacowanie opóźnienia komunikacji VM, która ma kluczowe znaczenie dla migracji na żywo



W tym teście rozmiar ramki jest ustalony na 1024 bajty. Dane są generowane z generatora ruchu i wysyłane do demo2. Po przekazaniu pakietu L2-forward jest przekazywany do demo3. W odróżnieniu od scenariusza vloop, w tym scenariuszu, po dostarczeniu pakietów testowych do avs-demo2, docelowy adres MAC jest ustawiony jako inna maszyna wirtualna (avs-demo3). Następnie pakiety przechodzą z avs-demo3 z powrotem do systemu operacyjnego. Na koniec pakiety zostaną przechwycone przez centrum testowe Ethernet. Na podstawie opóźnienia scenariusza vloop i bieżącego scenariusza różnica między nimi jest mierzona jako opóźnienie komunikacji VM. Zgodnie z rys. 10.12 opóźnienie między maszynami wirtualnymi w dowolnym momencie jest mniejsze niż 60 sekund.



Opóźnienie to jest lepsze niż wydajność obecnych komercyjnych serwerów i spełnia wymagania dynamicznej migracji

#### 10.6 Wniosek

W drodze do 5G wirtualizacja jest postrzegana jako jedna z kluczowych technologii, aby zmaksymalizować elastyczność, skalowalność i wydajność sieci, aby dostosować ją do różnych aplikacji i scenariuszy. Pomimo prostoty idei wdrożenia wirtualizacji w branży telekomunikacyjnej wdrożenie jest trudniejsze w praktyce. Komunikacja bezprzewodowa różni się od centrów danych IT tym, że komunikacja bezprzewodowa ma bardzo surowe wymagania dotyczące przetwarzania w czasie rzeczywistym. W tym rozdziale wirtualizacja w sieciach dostępu radiowego, w szczególności C-RAN, jest uważnie analizowana za pomocą analizy wyzwań i potencjalnych rozwiązań. Biorąc pod uwagę, że funkcje RAN, takie jak przetwarzanie pasma podstawowego, wymagają intensywnych obliczeń i przetwarzania w czasie rzeczywistym, wirtualizacja RAN wymaga następnie wspólnej optymalizacji różnych modułów systemowych, w tym systemów operacyjnych, hiperwizora itp. Wymaga to również ulepszenia i optymalizacji istniejącego tradycyjnego oprogramowania. Technologie informatyczne, w tym przyspieszenie sieci, migracja na żywo, komunikacja VM, wsparcie dla OpenStack i tak dalej. Opracowano prototyp zoptymalizowanej platformy do wirtualizacji o zwiększonej wydajności w czasie rzeczywistym. Testy wykazały, że zoptymalizowany obieg w obie strony L2 Opóźnienie przetwarzania platformy można zmniejszyć średnio do około 30, podczas gdy w tradycyjnym systemie Linux zwykle wynosi około 300-500. W międzyczasie opóźnienie komunikacji VM można zoptymalizować do około 10, w porównaniu do kilkuset mikrosekund dla tradycyjnego systemu Linux. Zwiększenie wydajności platformy w czasie rzeczywistym jest tylko pierwszym krokiem. W przyszłości należy zaangażować znacznie więcej zadań, takich jak zarządzanie w czasie rzeczywistym, szczegółowość wirtualizacji, przyspieszenie sprzętowe, projektowanie szybkich sieci przełączników o małej zwłóce itp. W celu realizacji całkowicie zwirtualizowanych systemów C-RAN



# Licencjonowany dostęp współdzielony (LSA) i trójpoziomowe modele podziału spektrum: regulacje, perspektywy biznesowe i technologiczne

## **Abstrakt**

Współdzielenie widma jest ważnym czynnikiem umożliwiającym przyszłym mobilnym sieciom szerokopasmowym zaspokojenie rosnących potrzeb użytkowników w zakresie przesyłania danych. Biorąc pod uwagę dzielenie się widmami od regulacji, perspektywy biznesowe i technologiczne są kluczem do pomyślnego opracowania modeli udostępniania, które można wdrożyć w praktycznych systemach. W niniejszym rozdziale przeanalizowano najnowsze koncepcje dzielenia się widmem z perspektywy regulacyjnej, biznesowej i technologicznej. Prezentujemy europejską koncepcję dzielenia się widmem LSA (License and Shared Access) obejmującą jej aspekty prawne, biznesowe i techniczne, ze szczególnym uwzględnieniem przykładu LSA, na przykład dzielenia się LTE i programami zasiedziały i usługami specjalnymi (PMSE) w 2.3-2.4. Pasma GHz. Ewolucja LSA, obejmująca bardziej dynamiczne współdzielenie w kierunku amerykańskiego trójpoziomowego modelu współdzielenia widma dla Citizen Broadband Radio Service (CBRS) opartego na systemie Spectrum Access System (SAS), przedstawiona jest pod kątem ekosystemu, korzyści biznesowych i technicznych podejść. Perspektywy na przyszłość przewidywane są na rozwój modeli udostępniania dla mobilnego internetu szerokopasmowego.

## **11.1 Wprowadzenie**

Bezprzewodowy wzrost ruchu danych zmotywował przemysł i środowisko akademickie do opracowania nowych modeli współdzielenia widma do przyjęcia przez regulatorów widma. Dzielenie widma odnosi się do sytuacji, w której dwa lub więcej systemów radiowych działa w tym samym paśmie częstotliwości. Począwszy od poznawczych badań radiowych ponad 10 lat temu, techniczne podejścia do dzielenia się widmem stały się dojrzałe i można je zastosować do konkretnych scenariuszy współdzielenia między rzeczywistymi systemami bezprzewodowymi w określonych pasmach częstotliwości. Opracowanie modeli współdzielenia widma wymaga dogłębnego rozważenia relacji rozporządzenia - biznes - technologia, ponieważ te trzy wymiary są bardzo wysoko powiązane i ich odrębne wymagania muszą zostać spełnione. Po szeroko zakrojonych badaniach nad udostępnianiem widma w tzw. Białych przestrzeniach telewizyjnych, często opartych na nielicencjonowanym dostępie z wykorzystaniem podejścia do bazy danych geolokalizacyjnych, zainteresowanie domeną regulującą spektrum rozszerzyło się na licencjonowane metody udostępniania, które mogą chronić obecnych użytkowników przy jednoczesnym wprowadzaniu dodatkowych licencjonowanych użytkowników na rynku. wspólna podstawa. Licencjonowane współdzielenie może skutkować środowiskiem kontrolowanej interferencji, w którym zarówno operator zasiedziały, jak i systemy wchodzące na rynek działające w tym samym paśmie mogą korzystać z gwarancji jakości usług (QoS). Po wprowadzeniu w 2011 r. Koncepcja Licencjonowanego Wspólnego Dostępu (LSA) zyskała znaczną uwagę w przepisach europejskich jako sposób na wprowadzenie dodatkowych licencjonowanych użytkowników na zespoły, które są obecnie wyłącznym użytkownikiem, a jednocześnie chronią dotychczasowy użytek. Jako hybryda dwóch podejść do dzielenia się, a mianowicie typu nielicencjonowanego i licencjonowanego, Federalna Komisja ds. Komunikacji (FCC) w USA zaproponowała trójstopniowy model udostępniania dla Citizen Broadband Radio Services (CBRS), wprowadzając dwie dodatkowe warstwy współdzielenia, podczas gdy ochrona obecnych operatorów. Model składa się zarówno z licencjonowanej warstwy

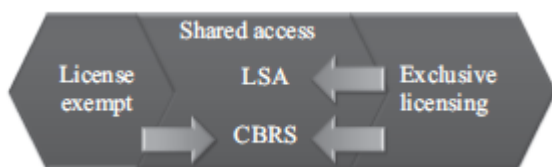
współużytkowania, jak i oportunistycznej warstwy współużytkowania z różnymi poziomami praw dostępu i gwarancjami ochrony przed zakłóceniami. Model ten jest opracowywany dla pasma 3,5 GHz jako pierwszy przypadek zastosowania, ale mógłby zostać przyjęty szerzej, gdyby udowodniono jego sukces. Ta część zajmie się koncepcjami dzielenia LSA i CBRS z perspektywy regulacji, biznesu i technologii. Koncepcje dzielenia widma LSA i CBRS są prezentowane na podstawie najnowocześniejszych osiągnięć z tej perspektywy i przedstawiono perspektywę na przyszłość.

## 11.2 Dzielenie widma

Aby zrozumieć rozwój nowych modeli współużytkowania spektrum, ta sekcja najpierw dzieli współdzielenie widma na szerszy obraz metod dostępu do widma i wprowadza perspektywy regulacji, biznesu i technologii. Następnie LSA i wprowadzono koncepcje CBRS.

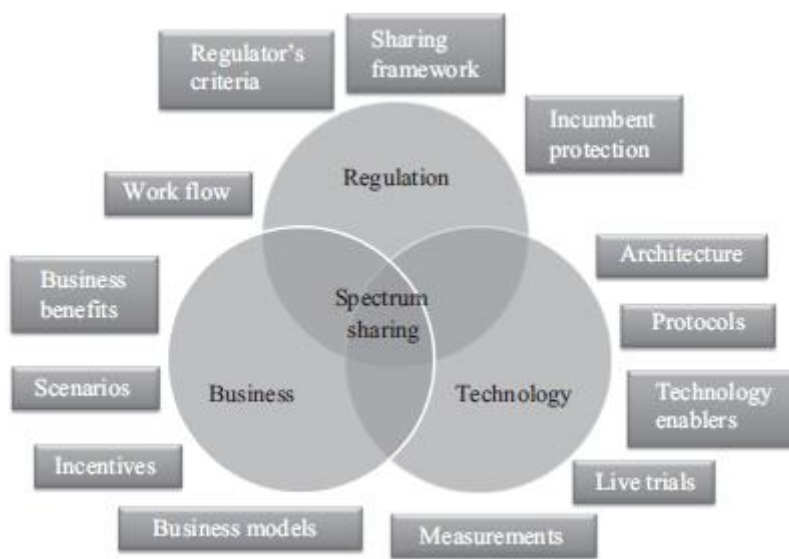
### 11.2.1 Modele dostępu do widma i perspektywy udostępniania

Tradycyjne modele dostępu do widma są w większości oparte na licencjonowaniu wyłącznym i operacjach zwolnionych z licencji. Operacje zwolnione z obowiązku udzielenia licencji umożliwiają łatwy dostęp do określonych pasm widma do urządzeń, które działają zgodnie z wcześniej określonymi warunkami eksploatacji, takimi jak ograniczenia mocy i cyklu roboczego. Ponieważ ilość urządzeń dopuszczonych do pasma nie może być kontrolowana, ten typ operacji nie może zagwarantować pewnego poziomu QoS. Do tej pory dostęp do widma dla operatorów sieci komórkowych (MNO) opierał się wyłącznie na wyłącznych licencjach o długim okresie ważności licencji i szerokim zasięgu. Jednak rosnące wyzwania związane z wyszukiwaniem wolnych zasobów (lub nawet mogą być realnie uwolnione) od obecnego użycia zainspirowały bardziej elastyczne sposoby wyznaczania zasobów widma dla operatorów sieci ruchomej. Współdzielenie widma pozwala dwóm lub więcej systemom radiowym działać na tym samym paśmie częstotliwości pod pewnymi regułami i warunkami, które zapewniają wykonalne środowisko operacyjne dla systemów. Rysunek 11.1 ilustruje dwa modele dostępu do wspólnego widma, omawiane obecnie w regulacji widma, jeden w Europie i jeden w USA



Ogólne modele dostępu regulacyjnego nie są ograniczone wyłącznie do zastosowań mobilnych, ale mogą być stosowane między dowolnymi dwoma lub więcej systemami radiowymi. Jednak w zakresie tego artykułu są one omawiane jako narzędzia umożliwiające mobilny dostęp do nowych pasm widmowych na zasadzie współdzielenia. Koncepcja europejskiego spektrum częstotliwości LSA przypomina ekskluzywne licencjonowanie z rozróżnieniem, że pasmo widmowe nie jest ponownie przydzielane wyłącznie do użytku mobilnego, ale istniejące wykorzystanie w paśmie jest utrzymywane w długim okresie. MNO, jako Licencjobiorca LSA, może wykorzystywać te części pasma widma, które nie są wykorzystywane przez dotychczasowego użytkownika widma w określonej lokalizacji, w określonych okresach czasu, zgodnie z ustalonymi zasadami i warunkami. Istniejący użytkownik utrzymuje wyższe prawa użytkowania i może odzyskać pasmo widmowe lub jego część w określonym obszarze, a sieć komórkowa musi wdrożyć narzędzia do zarządzania, które umożliwiają reagowanie na zmieniające się wykorzystanie widma w istniejącym systemie. CBRS, rozpatrywane w USA, jest bardziej złożoną koncepcją udostępniania, która umożliwia dodatkowe wykorzystanie pasma z istniejącym wykorzystaniem IU na obu licencjonowanych (PAL) oraz bez licencji (General Authorized Access (GAA))

)) podstawa. CBRS wprowadza dodatkowych licencjonowanych użytkowników w oparciu o PAL, które przypominają koncepcję LSA. Dwie główne różnice między tymi dwoma są takie, że centralna częstotliwość operacji nie jest ustalona w PAL, a zasób widma jest dostępny tylko wtedy, gdy jest używany, zatem zasób PAL może być wykorzystywany oportunistycznie przez licencjonowanego (oportunistycznego) General Authorized Urządzenia Access (GAA), gdy nie są używane przez posiadacza PAL. Koncepcja CBRS wprowadza również oportunistyczne warstwy GAA, których główną różnicą w porównaniu do wykorzystania bezzwrotnej licencji jest to, że użytkownicy GAA powinni być zarejestrowani jako użytkownicy CBRS. Dzielenie widma to coś więcej niż kwestia techniczna, co rozważano we wczesnych dniach badań kognitywnego systemu radiowego. Należy ją uważnie rozważyć również pod kątem regulacji i perspektyw biznesowych. Rysunek 11.2 ilustruje biznesowe, regulacyjne i technologiczne wymiary podziału widma oraz tematy, które są interesujące z tych perspektyw.

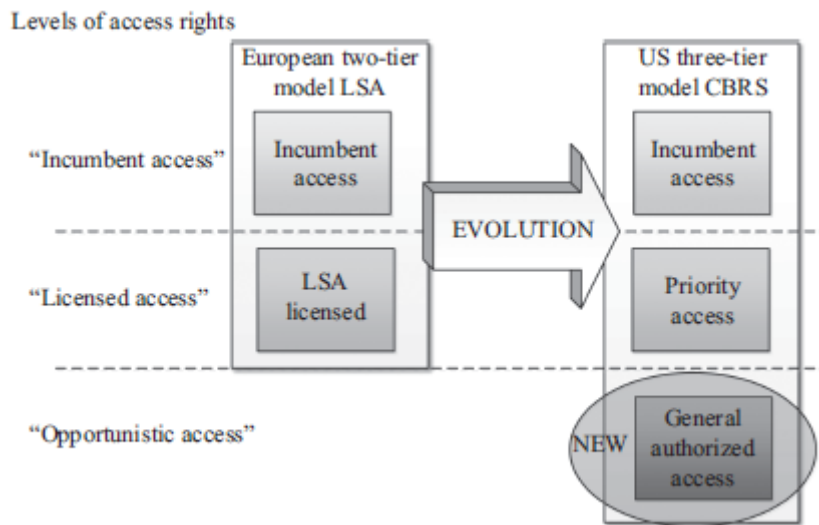


Z punktu widzenia regulacji, krajowe organy regulacyjne (KOR) mają zestaw kryteriów, według których model dzielenia się jest wykonalny, patrz [8], aby uzyskać szczegółowe informacje. Na przykład ochrona operatorów zasiedziałych przed szkodliwą ingerencją, uczciwością i pewnością prawną to ważne kwestie z punktu widzenia regulatora, które muszą być spełnione przez techniczne wdrożenia w celu wdrożenia systemu udostępniania. Z biznesowego punktu widzenia, dzielenie widma musi zapewniać korzyści biznesowe zaangażowanym interesariuszom, a także modele biznesowe i zachęty dla tych, którzy dzielą wspólne zasoby. Wreszcie, z punktu widzenia technologii i dostosowania się do biznesowego punktu widzenia, ważne jest opracowanie wykonalnej architektury i protokołów udostępniania, które wdrażają ramy udostępniania w zharmonizowany sposób, który umożliwia ekonomię skali. Próby na żywo i pomiary wydajności są ważne, aby pokazać możliwości dzielenia się pomysłami dla organów regulacyjnych w celu przyjęcia nowych modeli. W tym rozdziale wykorzystamy trzy perspektywy regulacji, biznesu i technologii oraz omówimy najnowsze koncepcje dzielenia się z tych perspektyw.

### 11.2.2 Pojęcia dzielenia widma

Obecnie najbardziej obiecującymi modelami udostępniania są obecnie europejski model LSA i amerykański trójkondygnacyjny model świadczenia usług CBRS. Koncepcja LSA wprowadza dodatkowych licencjonowanych użytkowników w zespołach, które obecnie mają wyłączne użytkowanie zasiedziałe. Koncepcja CBRS wprowadza dwie warstwy dodatkowych użytkowników z różnymi prawami dostępu, a jednocześnie chroni operatorów zasiedziałych. Rysunek 11.3 ilustruje

poziomy różnych poziomów praw dostępu w koncepcjach LSA i CBRS na wysokim poziomie. Poniżej przedstawiono te dwie koncepcje.



### 11.2.2.1 Licencjonowany dostęp współdzielony

Koncepcja Licencjonowanego Wspólnego Dostępu (LSA) to model współdzielenia widma zaproponowany przez Komisję Europejską (KE) w celu wprowadzenia dowolnego systemu radiowego do pasma częstotliwości z istniejącym wykorzystaniem. Dzielenie się odbywa się na podstawie licencji LSA wydanej przez krajowy organ regulacyjny oraz uzgodnionych ram podziału między zaangażowanymi stronami zainteresowanymi. LSA opiera się na dobrowolności, a zasiedziali użytkownicy mogą decydować, które pasma częstotliwości mają być wyznaczone do współdzielenia w jakich obszarach geograficznych. W ten sposób zespoły o stałym lub częstym użytkowaniu mogą pozostać poza układem LSA i można oczekiwać bardziej stabilnych warunków udostępniania dla licencjobiorcy LSA. Licencja LSA wydana przez organ regulacyjny gwarantuje koncesjonariuszowi LSA pewną operacyjność na czas trwania, obszar i warunki pasm widma niewykorzystywane przez operatora zasiedziałego, zgodnie z definicją zawartą w licencji i ramach udostępniania. Dotychczasowi użytkownicy widma mają prawa do użytkowania na wyższym poziomie i mogą odzyskać pasmo lub jego części w dowolnym miejscu w dowolnym momencie. Mechanizmy zarządzania i techniki, które licencjobiorca LSA musi wdrożyć, aby odpowiedzieć na zmieniającą się dostępność pasma LSA omówiono później.

### 11.2.2.2 CBRS

W kwietniu 2015 r. FCC przyjęła trzypoziomowy model podziału częstotliwości o nazwie Citizens Broadband Radio Service (CBRS) [5], zaproponowany przez Radę Doradców ds. Nauki i Technologii (PCAST) w 2012 r. [4]. CBRS wykorzystuje "szybką ścieżkę" o częstotliwości 3,550-3,700 MHz i wprowadza tam dwa typy współdzielonych warstw przy jednoczesnej ochronie operatorów zasiedziałych. Model składa się zarówno z licencjonowanej warstwy współużytkowania, jak i oportunistycznej warstwy współużytkowania z różnymi poziomami praw dostępu. Reguły są zoptymalizowane pod kątem wykorzystania w małych celach, ale mogą również obejmować punkt-punkt i punkt-wielopunkt, szczególnie w przypadku usług wiejskich. Zarządzanie ochroną interferencyjną jest realizowane za pośrednictwem systemu Spectrum Access Systems (SAS), który koordynuje wykorzystanie widma przez uczestników w celu ochrony operatorów zasiedziałych. Oprócz widma podejście bazy danych, która została przyjęta do koncepcji LSA The CBRS wprowadza pojęcie środowiska Sensing Capability (ESC) do monitorowania aktywności zasiedziałego, która wykrywa

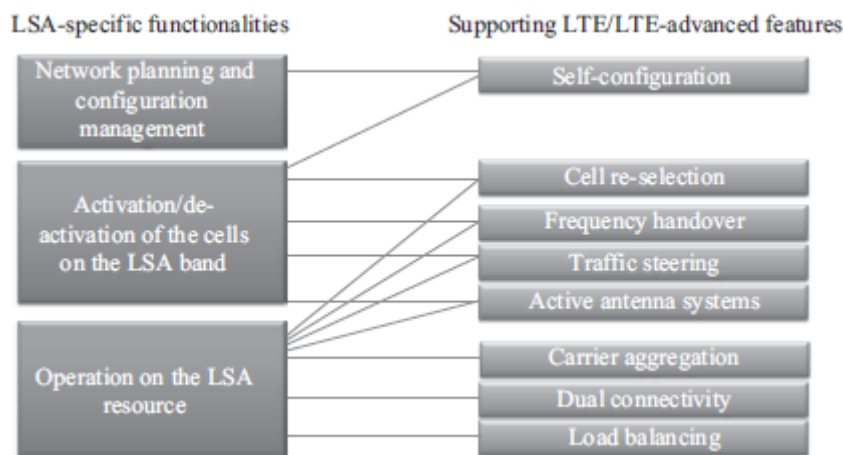
pojawienie się konkretnych operatorów, takich jak radar marynarki na obszarach przybrzeżnych i śródlądowych w pobliżu baz wojskowych. Zgodnie z dyskusjami, w przypadku wykrycia istniejącej aktywności, ESC przekazuje te informacje SAS, który powinien ponownie skonfigurować lokalne punkty dostępu w celu uniknięcia szkodliwej ingerencji w zasiedziałe radary.

### 11.3 Podejścia techniczne

Koncepcje dzielenia LSA i CBRS są najpierw omawiane z perspektywy technologicznej, dostarczając ich kluczowych elementów technicznych.




#### 11.3.1 Aspekty techniczne LSA

Jedną z głównych zalet LSA jest to, że przewiduje się wdrożenie w oparciu o dwie dodatkowe jednostki funkcjonalne nad istniejącą architekturą sieci komórkowej. Usługa LSA nie wymaga modyfikacji istniejących wewnętrznych procedur sieci komórkowej ani funkcji zarządzania zasobami radiowymi. Po stronie sprzętu użytkownika (UE) nie ma wpływu poza wdrażaniem obsługi dowolnego nowego pasma częstotliwości. W związku z tym dodatkowe spektrum LSA można wykorzystać bez długiego procesu regulacyjnego i standaryzacji. Dwie wymagane dodatkowe jednostki funkcjonalne to repozytorium LSA i kontroler LSA. Repozytorium LSA przechowuje i aktualizuje informacje na temat dostępności i wykorzystania pasma widma LSA wraz z obowiązującymi zasadami i warunkami (moc nadawania itp.) W celu ich wykorzystania. Repozytorium LSA koordynuje również wymianę informacji między obecnymi użytkownikami a kontrolerami LSA. Kontroler LSA zapewnia ochronę i swobodną pracę operatora zasiedziałego i sieci mobilnej, obliczając obszary ochrony na podstawie informacji otrzymanych z repozytorium LSA i informacji o układzie sieci komórkowej, wykorzystanych mocach transmisji itp. Obecny LTE Standard LTEAdvanced obejmuje kilka funkcji, które umożliwiają i wspierają korzystanie z widma LSA o zmiennej dostępności przez sieć komórkową. Te cechy podsumowano na rysunku 11.4.



Komórka / sektor w paśmie LSA może wymagać dezaktywacji z powodu żądań ewakuacji od obecnego użytkownika, a następnie ponownego skonfigurowania po ponownym włączeniu. Dzięki automatycznej konfiguracji komórka LSA może zostać ponownie uruchomiona przy minimalnej ręcznej interwencji. W przypadku aktywacji komórki w paśmie LSA, procedura ponownej selekcji komórek pozwala UE wybrać komórkę w paśmie LSA na podstawie jej pomiarów siły sygnału. W przypadku dezaktywacji komórki w paśmie LSA, terminale są w stanie automatycznie uruchomić procedurę ponownej selekcji komórek. Ponadto narzędzia takie jak kierowanie ruchem, przekazywanie częstotliwości i równoważenie obciążenia pomagają kierować użytkowników do najlepszej obsługującej komórki na podstawie wcześniej określonych kryteriów. Aktywne systemy antenowe

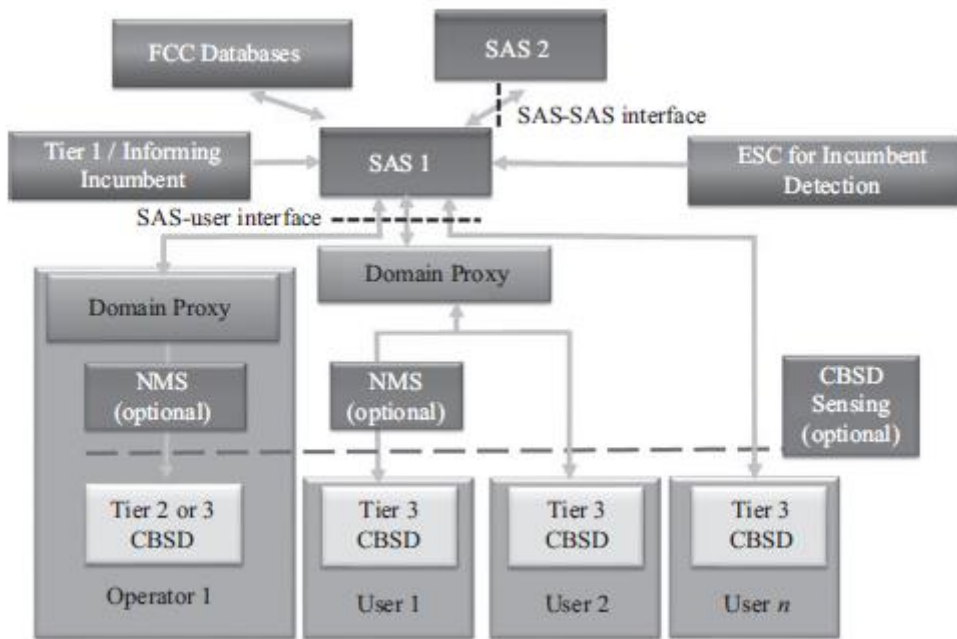
(AAS) mogą być wykorzystywane do zmniejszania rozmiarów stref wykluczenia i zakłóceń między użytkownikiem zasiedziałym a siecią mobilną poprzez optymalizację zasięgu i przepustowości sieci. Korzystając z AAS, rozmiar komórek może być automatycznie dostosowany i dostosowany do wydajności poprzez zmianę parametrów anteny, takich jak pochYLENIE, azymut lub kształt wiązki. Agregacja nośna (CA) umożliwia wykorzystanie zasobów radiowych na nośnikach wieloskładnikowych w celu zapewnienia szerszej efektywnej przepustowości dla użytkownika końcowego. Korzystając z urzędu certyfikacji, operator w paśmie LSA może być używany razem z operatorem w paśmie licencjonowanym, aby zapewnić dodatkową przepustowość swoim użytkownikom, bez ryzyka przerwania połączenia. Podwójna łączność umożliwia połączenie UE do zarówno istniejącej komórki FDD LTE, jak i komórki LSA TD-LTE w tym samym czasie. Pierwsza na świecie próba LSA dotycząca wymiany między LTE a operatorami zasiedziałymi w paśmie 2,3 GHz została pokazana w Finlandii w 2013 r. Rysunek 11.5 ilustruje ewolucję fińskiego procesu LSA.

2013	2014	2015
<ul style="list-style-type: none"> <li>• World's first live LSA field trial</li> <li>• All LSA components available</li> <li>• Evacuation of LSA band</li> <li>• QoS measurements</li> </ul>	<ul style="list-style-type: none"> <li>• TDD/FDD handover</li> <li>• Visualization</li> <li>• Multiple incumbent types</li> <li>• Small cells</li> </ul>	<ul style="list-style-type: none"> <li>• Tracking of mobile incumbent</li> <li>• LSA Controller integrated to network management</li> </ul>
		

W 2015 r. Fińskie testy LSA pokazały, w jaki sposób aplikacja mobilna może automatycznie śledzić obecnego użytkownika, a sieć komórkowa jest w stanie automatycznie dostosować się do tego. Dodatkowo wykazano nowy algorytm oszacowania interferencji, który może uwzględniać zagregowaną interferencję ze wszystkich stacji bazowych. Ta próba została zaprezentowana w celu udoskonalenia sterowania mocą w celu koordynacji zakłóceń między siecią LTE a obecnym systemem PMSE.

### 11.3.2 Aspekty techniczne CBRS

Głównym elementem technicznym CBRS jest SAS. Oprócz omówionych funkcji SAS, koncepcja CBRS składa się z połączonych z SAS baz danych spektrum, ESC, urządzeń CBRS (CBSD), urządzeń użytkowników końcowych (EUD) oraz opcjonalnie serwerów proxy domeny i systemu zarządzania siecią (NMS), jak pokazano na rysunku 11.6.



CBSD są stałymi lub przenośnymi stacjami bazowymi lub punktami dostępowymi lub sieciami takich stacji i mogą działać jedynie pod nadzorem i zarządzaniem scentralizowanego SAS wybranego SAS, który może być wielokrotny. Komisja zatwierdziła urządzenia CBRS (CBSD) i musi zarejestrować w SAS informacje wymagane przez przepisy, np. Identyfikator operatora, identyfikację urządzenia i parametry oraz informacje geolokacyjne. W przypadku CBSD jest zarządzana sieć, jak w typowym przypadku MNO, CBSD obejmuje funkcjonalność serwera proxy domeny i zarządzania siecią. Serwer proxy może być czystym dwukierunkowym procesorem przetwarzania i routingu informacji lub bardziej inteligentną funkcją mediacji, np. łączenie małych komórek centrum handlowego lub hali sportowej z wirtualną jednostką BTS, która obejmuje centrum handlowe lub arenę sportową. Ta ostatnia opcja umożliwi elastyczną samokontrolę i optymalizację interferencji w takich sieciowych urządzeniach końcowych użytkowników (EUD), np. Telefony nie są uważane za CBSD. SAS przypisuje widmo i dynamicznie określa i wymusza maksymalne poziomy mocy. do certyfikowanych CBSD w określonym miejscu i czasie geograficznym, kontroluje środowisko interferencyjne i wymusza kryteria ochrony i strefy wyłączenia w celu ochrony użytkowników o wyższym priorytecie, a także zajmuje się rejestracją, uwierzytelnianiem i identyfikacją informacji użytkownika oraz wykonuje inne funkcje określone w Zasady FCC. Ponieważ użytkownicy IA mają podstawowe prawa do widma w każdym czasie i we wszystkich obszarach w porównaniu z PA i GAA, wszystkie CBSD i EUD muszą być zdolne do dwukierunkowej komunikacji w całym zakresie 3,5 GHz i zaprzestania działania lub zmiany częstotliwości w kierunku SAS w celu ochrony IA. Obowiązkowe jest, aby wszystkie urządzenia CBSD chroniły użytkowników IA w paśmie. W oparciu o naturę i krytyczne wymagania federalnego operatora zasiedziałego, FCC przyjęła zasady wymagające od ESC wykrycia zasiedziałej aktywności radarowej na obszarach przybrzeżnych iw pobliżu śródlądowych baz wojskowych w pasie 3,5 GHz i sąsiadującym z nim. Po wykryciu działania IA, ESC przekazuje te informacje do SAS, które w razie potrzeby mogłyby nakazać użytkownikom warstwy komercyjnej opuszczenie zasobu widma w ciągu 60 s w częstotliwości, miejscu lub czasie, który w pobliżu radaru federalnego IA stwarza ryzyko szkodliwego działania. ingerencja. Federalna ochrona IA zostanie wprowadzona w dwóch etapach: Po pierwsze, duży obszar kraju poza strefami wyłączeń statycznych będzie dostępny po tym, jak SAS zostanie zatwierdzony i komercyjnie dostępny przez FCC. W drugiej fazie strefy wyłączone zostaną przekształcone w strefy ochronne za pośrednictwem systemu ESC, dzięki czemu reszta kraju, w tym główne obszary przybrzeżne, stanie się dostępna. ESC składa się z jednej lub więcej komercyjnie obsługiwanych sieci

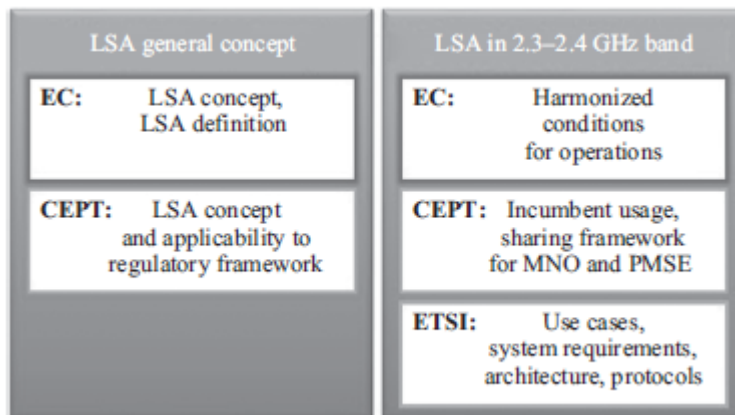
opartych na czujnikach lub opartych na infrastrukturze czujników CBSD, które byłyby wykorzystywane do wykrywania sygnałów z federalnych systemów radarowych w pobliżu stref zamkniętych. Potencjalni operatorzy ESC muszą zatwierdzić swoje systemy w tym samym procesie co administratorzy SAS i SAS. SAS uzyskałby informacje FCC, np. O zarejestrowanych lub licencjonowanych użytkownikach komercyjnych, obszarach strefy wykluczenia wymagających ESC z bazy danych FCC. Architektura funkcjonalna ma opcję informowania aktualnego przypadku, że federalna IA chce poinformować SAS przed planami wykorzystania widma w pewnym obszarze, np. Związany z planowanym wykorzystaniem widma w szkoleniu.

#### 11.4 Współdzielenie statusu modelu w regulacji i standaryzacji

Następnie koncepcje dzielenia się LSA i CBRS omawiane są z perspektywy regulacyjnej, w tym standaryzacji, oraz statusów dwóch koncepcji w tych forach są recenzowane.

##### 11.4.1 LSA

Działania regulacyjne i standaryzacyjne związane z LSA w Europie były jak dotąd skoncentrowane w paśmie 2,3-2,4 GHz z MNOs będącymi licencjobjorcami LSA i operatorami zasiedziałymi różniącymi się w zależności od sytuacji krajowej. Korzystanie z LSA jest sprawą ogólnokrajową jest regulowany przez krajowy organ regulacyjny, ale wydaje się, że zharmonizowane ramy pomagają w opracowaniu zharmonizowanego podejścia i rynku LSA w Europie. Szczegółowy opis różnych działań i obecnego stanu regulacji i standaryzacji LSA są udokumentowane. Omówienie tego przedstawiono na rysunku 11.7.



KE opracowała najpierw ogólną koncepcję LSA, wprowadzając ją, a następnie opracowując oficjalną definicję LSA. Ponadto WE rozpoczęła badania nad LSA w paśmie 2,3- 2,4 GHz, wydając zarówno upoważnienie do standaryzacji ETSI, jak i mandat regulacyjny do CEPT. W odpowiedzi na mandat regulacyjny, CEPT przedstawiła przegląd różnych obecnych usług w różnych krajach europejskich oraz istniejących przykładów wdrożenia. CEPT opracowała również ramy LSA ze zharmonizowanymi warunkami technicznymi i regulacyjnymi pomiędzy mobilnymi i programowymi i specjalnymi wydarzeniami (PMSE), bezprzewodowymi kamerami, które są najczęściej stosowanym obciążeniem dla zespołu. Ramy regulacyjne LSA w Europie dla pasma 2,3-2,4 GHz są gotowe do przyjęcia na poziomie krajowym, ale jak dotąd nie przeprowadzono żadnych wdrożeń. Jednak prowadzone są już próby, aby zapewnić administrację wykonalności koncepcji. ETSI opracował przypadki użycia, wymagania systemowe, architekturę i funkcje, przepływy informacji, procedury i interfejsy do operacji LSA na potrzeby mobilnego internetu szerokopasmowego w paśmie 2,3 GHz. W oparciu o badania przeprowadzone w ETSI i CEPT, EC zapewni zharmonizowane warunki dla operacji na 2,3-2,4 GHz. Będą one wiążące dla wszystkich państw członkowskich UE.



#### 11.4.2 CBRS

CBRS to krajowa koncepcja zaproponowana przez FCC w USA. Kluczowe komunikaty polityczne raportu PCAST zostały dodatkowo wzmocnione w 2013 r. Memorandum mieszkaniowe stwierdzające, że . . . Musimy udostępnić jeszcze więcej spektrum i stworzyć nowe drogi dla innowacji bezprzewodowych. Jednym ze sposobów jest umożliwienie i zachęcanie do wspólnego dostępu do widma, który jest obecnie przydzielany wyłącznie do użytku federalnego. Tam, gdzie jest to technicznie i ekonomicznie wykonalne, dzielenie się może i powinno być wykorzystywane w celu zwiększenia wydajności wszystkich użytkowników i przyspieszenia komercyjnego dostępu do dodatkowych zakresów częstotliwości, z zastrzeżeniem odpowiedniej ochrony przed zakłóceniami dla użytkowników federalnych, . . . powinniśmy również dążyć do zniesienia ograniczeń w zakresie możliwości przewoźników komercyjnych w negocjowaniu umów o współdzieleniu z agencjami. Aby kontynuować te wysiłki, nadal chroniąc chronione zasiedziały systemy, które mają kluczowe znaczenie dla interesów federalnych i wzrostu gospodarczego, niniejsze memorandum nakazuje agencjom i biuram przyjęcie szeregu dodatkowe działania w celu przyspieszenia wspólnego dostępu do widma. Po intensywnych dyskusjach i konsultacjach z branżą opublikowano FCC Raport i kolejność oraz drugie dalsze zawiadomienie o proponowanych przepisach w celu ustalenia nowe zasady wspólnego korzystania z pasma 3,550-3,650 MHz w kwietniu 2015 r. FCC postrzega otwarcie pasma 3,5 GHz jako "nowy rozdział w historii administracji jednego z najcenniejszych zasobów naszego narodu – elektromagnetyzmu widmo radiowe." Struktura definiuje sąsiedni blok 150 MHz na 3550-3 700 MHz dla mobilnego Internetu szerokopasmowego (MBB), który FCC nazywa Citizens Broadband Radio Service (CBRS). Zakres częstotliwości 3,550-3,650 MHz jest obecnie przydzielany do użytku przez systemy radarowe Departamentu Obrony USA (DoD) i Fixed Satellite Services (FSS) podczas gdy operatorzy widma w zakresie 3650-3 700 MHz to FSS i komercyjne bezprzewodowe usługi szerokopasmowe. FCC określa CBRS jako "zespół innowacyjny", w którym mogą przypisać spektrum do komercyjnych systemów MBBB na zasadzie współdzielenia z zasiedzianymi systemami radarowymi i FSS oraz promować różnorodność heterogenicznych technologii sieciowych, w szczególności małych komórek. Neutralność technologiczna CBRS będzie w szczególności odgrywać rolę w oportunistycznych poziomach autoryzacji dostępu (GAA), otwierając nowe możliwości. Ramy udostępniania w CBRS obejmują trzy poziomy: istniejący dostęp (IA), priorytetowy dostęp (PA) i ogólny autoryzowany dostęp (GAA), jak pokazano na rysunku 11.3. Obowiązkiem wszystkich użytkowników CBRS (użytkowników PA i GAA) będzie ochrona użytkowników IA w paśmie. Użytkownicy PA otrzymają licencję PA (PAL) od FCC do obsługi do 70 MHz segmentu widmowego 3,550-3,650 MHz i są chronione przed szkodliwymi zakłóceniami z operacji GAA. Użytkownicy PA otrzymują krótkoterminowe upoważnienie priorytetowe do działania w wyznaczonych obszarach geograficznych z PAL, takimi jak kanał sparowany 3 rokiem 10 MHz, na pojedynczym torze spisowym, nagradzanym licytowaniem konkurencyjnym. Tylko podczas pierwszego okna aplikacji wnioskodawca może ubiegać się o maksymalnie dwa kolejne trzyletnie warunki dla danego PAL. Licencje będą mogły pomieścić nie więcej niż cztery PAL w jednym przewodzie spisowym jednorazowo. Zapewni to dostępność widma PAL przynajmniej dwóm licencjonowanym użytkownikom w obszarach geograficznych o największym zapotrzebowaniu. PAL są przypisywane określone częstotliwości w ich obszarze usług, a ich przypisanie częstotliwości nie powinno być kontrolowane dynamicznie przez SAS. Jednak częstotliwość środkowa może się różnić w zależności od istniejącej aktywności. Pod koniec okresu obowiązywania PAL automatycznie wygasa i nie może zostać przedłużony. Trzecia warstwa GAA będzie działała zgodnie z ramami licencjonowanych reguł i będzie dozwolona w paśmie 150 MHz bez żadnej ochrony przed zakłóceniami od innych użytkowników CBRS. Ta struktura ma na celu ułatwienie szybkiego wdrażania zgodnych z wymogami małych urządzeń komórkowych przy jednoczesnym zminimalizowaniu administracyjnych kosztów i obciążeń dla społeczeństwa, licencjobiorców i FCC. Użytkownicy GAA mogą używać wyłącznie certyfikowanych

urządzeń FCRS z certyfikatem CBRS i muszą rejestrować się w SAS z informacjami wymaganymi zgodnie z przepisami, np. Identyfikatorem operatora, identyfikacją urządzenia i informacjami geolokacyjnymi. Urządzenia CBRS (CBSD), które są stałymi stacjami (punktami dostępowymi) lub sieciami takich stacji, będą dynamicznie przydzielane widmem przez wybraną przez FCC SAS, która może być wielokrotnością. Sprzęt użytkownika, np. Telefony komórkowe, nie są uważane za urządzenia CBSD. SAS kontroluje środowisko interferencyjne i wymusza strefy wyłączenia, aby chronić użytkowników o wyższym priorytecie, a także zajmuje się rejestracją, uwierzytelnianiem i identyfikacją informacji o użytkowniku. Ponieważ użytkownicy IA mają podstawowe prawa dostępu do widma przez cały czas i we wszystkich obszarach na PA i GAA, wszyscy użytkownicy CBRS muszą być zdolni do działania w całym paśmie 3,5 GHz i zaprzestania działania lub zmiany częstotliwości w kierunku SAS w celu ochrony IA. Automatyczne przydzielanie kanałów przez SAS będzie po prostu wiązało się z instrukcjami dla tych użytkowników, aby używali określonego kanału, w określonym miejscu i czasie, w zakresie 3550-3 700 MHz. W oparciu o naturę i krytyczne wymagania federalnego operatora zasiedziałego, FCC przyjęła zasady, aby wymagać funkcji wykrywania środowiska (ESC) do wykrywania federalnego wykorzystania widma w pasmie 3,5 GHz i sąsiadującym z nim. Opcjonalne GAA bez ochrony przed zakłóceniami od innych użytkowników CBRS ma na celu zapewnienie taniego wejścia do pasma CBRS dla szerokiej gamy użytkowników i usług, podczas gdy operacje systemu PAL muszą czekać na rozpoczęcie procesu aukcyjnego po 600\$ Oficjalne aukcje na mecie na 2016 r. Tymczasem FCC zachęciła grupy wielopodmiotowe do rozważenia różnych kwestii poruszonych w przepisach. Komitet ds. Udostępniania widma WIREless iNNovation Forum (WInnForum) z przedstawicielami mobilnego internetu szerokopasmowego, bezprzewodowego dostępu szerokopasmowego, Internetu i ekosystemów obronnych rozpoczął wstępną pracę normalizacyjną nad interfejsami między systemem MBB a działaniami SAS ukierunkowanymi na umożliwienie współdzielenia CBRS w 2016 r. Punktem wyjścia do ewentualnego przyszłego dostosowania systemów LSA i CBRS WInnForum jest wymiana informacji pomiędzy organizacjami normalizacyjnymi (ETSI, 3GPP) w zakresie pasma 3,5 GHz oraz obsługa systemu zarządzania siecią dla LSA. Rząd USA początkowo zidentyfikował dodatkowe 2 GHz spektrum poniżej 6 GHz należące do DoD i innych użytkowników do przyszłego wspólnego wykorzystania komercyjnego warunkowo, jeśli udostępnianie widma na 3,5 GHz zakończy się sukcesem. To otwiera drogę, aby licencjonowane spektrum udostępnić trzeciemu głównemu nurtu spektrum licencjonowania użytkownikom komercyjnym, uzupełniając tradycyjne wyłączone licencjonowanie i dostęp do nielicencjonowanego widma. FCC ma wizję powtórzenia sukcesu Wi-Fi poprzez obniżenie bariery wejścia QoS bariery wejścia dla nowych uczestników i pionów, np. Przedsiębiorstwa, zakłady użyteczności publicznej, służba zdrowia, bezpieczeństwo publiczne, inteligentne miasta.

### **11.5 Względy biznesowe**

Wraz z transformacją w kierunku dzielenia się, branża komunikacji mobilnej jest świadkiem różnorodnych przejść, które wpływają na sposób prowadzenia biznesu w branży. Zrozumienie środowiska biznesowego wokół udostępniania dla mobilny internet szerokopasmowy jest ważny dla rozwoju realistycznych koncepcji dzielenia się widmem Trendy omówione powyżej w kontekście biznesowym prowadzą do powstania i ekosystemowego "myślenia o gospodarce dzielenia się" w ramach mobilnego internetu szerokopasmowego, gospodarki, która wymaga warstwowego i sieciowego spojrzenia na modele biznesowe. Aby zrozumieć transformację, warstwowy model biznesowy 4C składający się z modeli biznesowych łączności, treści, kontekstu i handlu przedstawia przeplatające się możliwości biznesowe i modele biznesowe. Są różne zainteresowane strony i sposoby prowadzenia biznesu na różnych poziomach. Dostępne usługi mobilnego dostępu szerokopasmowego umożliwiają oferowanie różnych usług związanych z treścią, takich jak usługi audio i wideo. Usługi kontekstowe stają się istotne w obfitości usług związanych z treścią, np. W postaci usług wyszukiwarek,

aby użytkownicy końcowi mogli znaleźć treści, których szukają, lub w formie dostarczania informacji kontekstowych, takich jak informacje o dostępności widma dla operatorów MNO w LSA lub dla użytkowników PA w ramach CBRS. Usługi handlu z kolei opierają się na platformach, w których interesariusze zaangażowani w układy LSA lub CBRS kupują lub sprzedają w sposób ciągły w czasie rzeczywistym dowolną z usług niższej warstwy lub oferują użytkownikom końcowym możliwość zostania prosumentami. Ogólnie rzecz biorąc, w przypadku współdzielenia operatorów tradycyjnych istnieje możliwość zwiększenia prawa do korzystania z widma w dłuższej perspektywie, ponieważ dopuszczają oni dodatkowych użytkowników w paśmie, kontynuując własną działalność. Mogłoby to również umożliwić budowanie dodatkowych źródeł dochodu i umożliwić korzystanie z infrastruktury MNO, a tym samym osiągnąć oszczędności w inwestycjach infrastrukturalnych. Jeśli chodzi o operatorów sieci ruchomych, współużytkowanie może pomóc w uniknięciu lub obniżeniu kosztów widma, otworzyć nowe źródła przychodów z usług, zwiększyć skalowalność zasobów infrastruktury i pomóc w wykorzystaniu profili klientów. Dla innych interesariuszy uczestniczących w wymianie, takich jak wewnętrzni dostawcy usług komunikacyjnych (CSP) lub integratorzy, chmura oparte na sieci jako usługa może otworzyć nowe możliwości biznesowe. Podobnie, dostarczanie informacji kontekstowych, np. Dotyczące dostępności widma, może tworzyć nową klasę usług potrzebnych w mobilnym ekosystemie szerokopasmowym.

### 11.5.1 Aspekty biznesowe LSA

Przeprowadzono szeroko zakrojone prace badawcze nad aspektami biznesowymi koncepcji LSA, polegającej na dzieleniu się między urządzeniami mobilnymi a operatorami zasiedziały. Ekosystem biznesowy wokół LSA koncentruje się głównie na MNO, NRA i operatorach zasiedziały. Obowiązują tu dwa typy operatorów zasiedziały i operatorów sieci ruchomych, urzędnicy państwowi (OG) i prywatni operatorzy (PI), a także dominujący i motywujący operatorzy sieci ruchomej. Wcześniejsze rozważania biznesowe dotyczące LSA przedstawiono na rysunku 11.8.



Przede wszystkim dla krajowych organów regulacyjnych LSA stwarza okazję do promowania innowacji, zagwarantowania sprawiedliwego dostępu do widma, odpowiedzi na zapotrzebowanie na widmo przez operatorów sieci komórkowych oraz utrzymania rozsądnych poziomów cen na rynkach konsumenckich. Osiągalne korzyści operatora zasiedziały od LSA mogą być utożsamiane z uzyskaniem lepszej pozycji w możliwych negocjacjach realokacji widma oraz zwiększeniu obrotu i potencjału zysku, np. Poprzez wykorzystanie infrastruktury operatorów sieci ruchomej. Istotnie, operator zasiedziały może przyjąć strategię obronną lub agresywną dotyczącą udostępniania [24], zwiększyć efektywność kosztową w istniejących przedsiębiorstwach lub wygenerować nowe dochody z nowych możliwości biznesowych. W podobny sposób mogą przyjąć otwartą lub zamkniętą strategię dotyczącą swojego systemu, skupić się na konwergencji technologii i przedsiębiorstw lub na zabezpieczeniu krytycznej infrastruktury. W przypadku operatorów telefonii komórkowej szybki i tani

dostęp do dodatkowego widma w sposób efektywny pod względem kosztów i potencjalnie bez dodatkowych zobowiązań dotyczących zasięgu może przynieść korzyści zarówno ze strony kosztowej, jak i dochodowej modelu biznesowego. Modele biznesowe MNO są prezentowane dominującym i stawiającym wyzwania MNO. Dominujący MNO mogą lepiej wykorzystywać segmentację klientów indywidualnych i korporacyjnych, gdy szukają wzrostu lub unikają dodatkowych kosztów. Po stronie operacyjnej dominujące systemy zarządzania danymi klienta i usługami dla operatorów sieci ruchomych oraz dynamiczne funkcje zarządzania ruchem zyskują na znaczeniu w LSA w porównaniu z tradycyjnymi modelami biznesowymi. Dla zaawansowanych operatorów telefonii komórkowej szansą jest pozyskanie klientów od dominujących operatorów sieci ruchomych, w szczególności z dostosowanymi lub specyficznymi usługami dla wybranych klientów, dzięki lepszemu doświadczeniu klientów lub w obszarach, w których dominujący MNO mają mniejszą koncentrację. Aby wyciągnąć wnioski na temat biznesowych aspektów LSA, główne otwarte pozycje są zachętami dla dotychczasowych podmiotów, ponieważ koncepcja LSA opiera się na dobrowolności. Co więcej, modele biznesowe dla operatorów sieci ruchomej są otwarte, ponieważ najwięksi operatorzy sieci komórkowych preferują ekskluzywne widmo na wspólnym widmie i nie promują udostępniania. Dlatego skalowalność modeli biznesowych opartych na LSA jest bardziej ograniczona biznesowo w porównaniu z CBRS, ponieważ opiera się głównie na warstwie połączenia

### **11.5.2 Aspekty biznesowe CBRS**

Ekosystemem biznesowym wokół CBRS jest znacznie bardziej wszechstronny w porównaniu z podstawowym podejściem LSA. Jednak do chwili obecnej nie ma zbyt wielu opublikowanych materiałów dotyczących rozważań biznesowych dotyczących CBRS. Rysunek 11.9 przedstawia uproszczoną ilustrację ekosystemu biznesowego wokół koncepcji CBRS i kluczowych interesariuszy. Po stronie organizatora CBRS, rozporządzenie (FCC) i organizacje normalizacyjne (WInnForum itp.) Odgrywają kluczową rolę w definiowaniu zasad i warunków oraz zharmonizowanych podejść do spełnienia tych wymagań, a także zdolności SAS i ESC w zapewnieniu dostępności widma. Informacje dla użytkowników CBRS. Po stronie użytkowników pasma CBRS, trzy warstwy praw dostępu do pasma 3,5 GHz są przechowywane przez różnych interesariuszy. Przykładowo, współrzędnościowe kanały międzykanałowe obejmują, na przykład, radar, FSS, usługę lokalizacji radiowej (RLS), bezprzewodową usługę szerokopasmową, i radar oraz FSS jako sąsiednich operatorów kanałów. Jako uczestnicy zapewniający CBRS, są operatorzy warstwy PA i operatorzy warstwy GAA, którzy mogą być również tymi samymi graczami, a także dostawcami sprzętu CBRS. W ramach systemowych zmiana taka jak wprowadzenie CBRS, wszyscy interesariusze w ekosystemie biznesowym odgrywają istotną rolę w przyjęciu nowej koncepcji CBRS i ogólnie dzielenia widma. Ponadto, przy opracowywaniu i analizowaniu ram możliwości, trzy dziedziny regulacji, biznesu i technologii, mające wpływ na koncepcję dzielenia się widmem, powinny przebiegać w parze. Zidentyfikowaliśmy umożliwiające, ograniczające i stawiające wyzwania elementy dotyczące możliwości biznesowych dla operatorów sieci ruchomej w kontekście CBRS, które są omówione poniżej i wymienione w tabeli

#### Elementy ramek biznesowych

- Regulatorzy rozważający wspólne ramy widma w UE i USA
- Wspólna alokacja widma poprawia ogólną efektywność wykorzystania widma
- Brak wyłącznego spektrum wyzwala nowe podejścia do dostępu do widma
- Konsumenci Nawyki konsumpcyjne MBB zmieniają się w kierunku asymetrycznym korzystanie z wielu urządzeń

- Uwolnienie inwestycji w spektrum, infrastrukturę sieci i usługi
- Dodatkowe niższe możliwości kosztowe, aby poradzić sobie z asymetrycznym ruchem i poprawić wydajność
- Może obniżyć bariery wejścia dla zaawansowanych operatorów telefonii komórkowej i nowego rodzaju operatorów
- Lepsze spektrum QoS może zwiększyć gęstą działalność w obszarach miejskich
- Dodatkowa zdolność GAA do rozładunku
- Zharmonizowana baza technologii LTE umożliwia optymalizację zasobów heterogenicznych sieci i oferuje skalę
- Duże możliwości analizy danych i danych z domeną internetową
- Potrzeba globalnych i krajowych regulacji poza Stanami Zjednoczonymi może spowolnić wejście . Harmonizacja jest warunkiem wstępnym do skalowania i pełnego wykorzystania potencjalnych korzyści.
- Konieczna standaryzacja funkcjonalności SAS dla potrzeb technologii IMT
- Ograniczona dostępność widma i przewidywalność ogranicza działalność MNO możliwości
- Niewyraźne lub brakujące rzeczywiste zachęty dla obecnych operatorów federalnych
- Federalne zasiedziały specjalne wymagania w szczególności związane z bezpieczeństwem i potrzebami do wykrywania
- Stopień dzielenia się informacjami krytycznymi biznesowo (MNO) i tajemnicą informacji (urzędnik federalny) i potrzebny system ESC
- Wpływ na model wyłącznego licencjonowania widma i jego dostępność w przyszłości
- Niepewność i ryzyko związane z regulacją w czasie, okresie, licencjach i elastyczność tworzy ryzyko i naraża operatora sieci ruchomej na kontynuowanie inwestycji.
- Atrakcyjny i dynamiczny rynek widma o potencjalnie niższych kosztach transakcji.
- Może zwiększać i zmieniać konkurencję. Nowi operatorzy i inne domeny biznesowe.
- Zwiększona złożoność techniczna i operacyjna (SAS) z powiązaniem kapitałem i koszty operacyjne
- Nowe kompetencje i możliwości potrzebne do zarządzania siecią i optymalizacja
- Terminowa dostępność terminali i potencjalny wpływ na koszty i złożoność

Elementy biznesowe i technologiczne można zidentyfikować jako czynniki umożliwiające współtworzenie wartości. Szybko rosnący popyt i brak ekskluzywnego spektrum w połączeniu z drastycznymi zmianami w nawykach konsumpcyjnych będą zachęcać do przyjęcia nowych, bardziej elastycznych i wydajnych koncepcji zarządzania widmem. Ramy radykalnie rozpraszają inwestycje w widmo, infrastrukturę sieci i usługi, które umożliwiają nowe usługi i modele biznesowe. Co więcej, różne programy podziału spektrum mają duże znaczenie w programach organów nadzorczych, których celem jest obniżenie bariery wejścia na widmo dla nowych rodzajów operatorów, którzy mogliby rozważyć wejście do bezprzewodowego Internetu szerokopasmowego. Wykorzystanie skali i harmonizacji ekosystemu LTE zredukuje dojrzałość technologii związanej z ryzykiem i zapewni

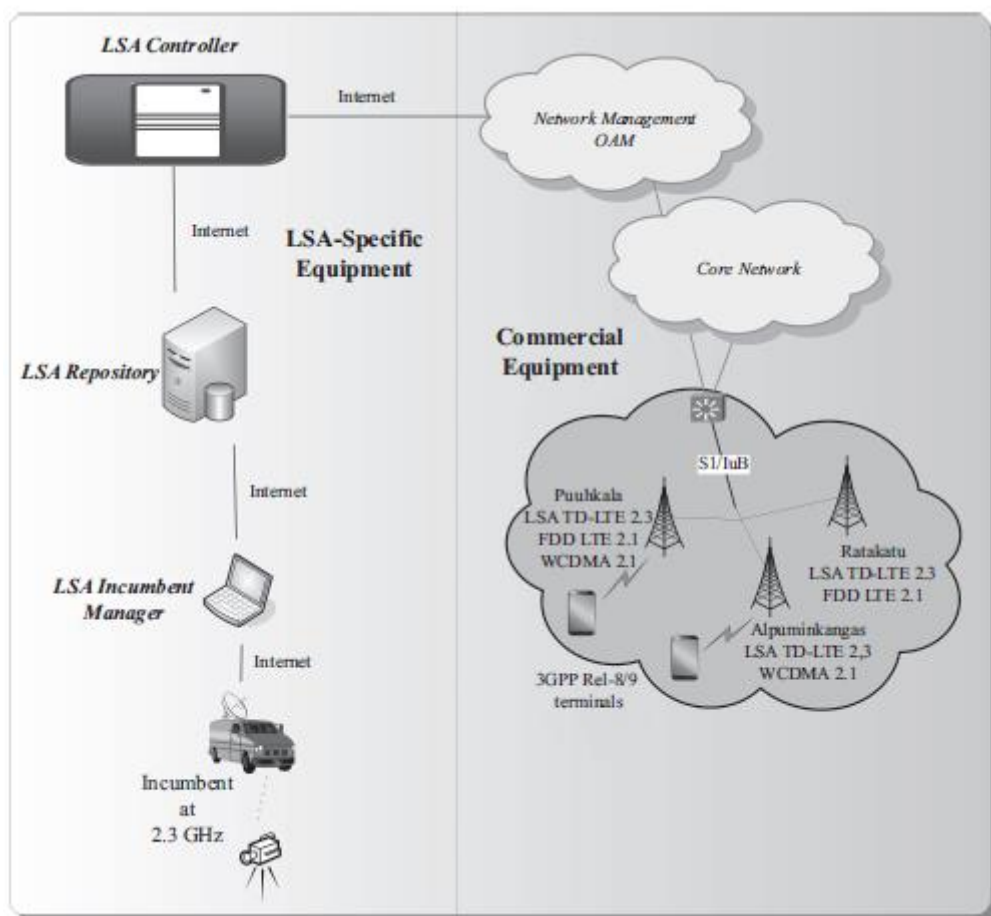
narzędzia do bezproblemowej integracji dodatkowej przepustowości z heterogenicznymi sieciami MNO, np. Za pośrednictwem technologii Carrier Aggregation (CA), LTE Unlicensed (LAA) i samoorganizującej się sieci. Funkcje analizy dużych danych będą odgrywać główną rolę w radzeniu sobie z dynamiką SAS i umożliwieniu niskich kosztów transakcyjnych. W odniesieniu do czynników ograniczających, rozsądne, zrównoważone i zharmonizowane otoczenie regulacyjne może być ograniczeniem, którym należy się zająć, zanim MNO będzie mógł współtworzyć i współtworzyć wartość z nim z partnerami ekosystemowymi. Ograniczona dostępność częstotliwości w zakresie częstotliwości, czasu lub lokalizacji z potencjalnym ograniczeniem i niepewnością może negatywnie wpłynąć na prognozy MNO dotyczące wspólnego użytkowania i wyceny widma. Specyficzną pozycją technologiczną, którą należy wziąć pod uwagę, jest stopień informacji biznesowych (MNO) i misji (DoD), które są niezbędne do udostępniania i wynikającej z tego potrzeby systemu ESC. Oprócz możliwości związanych z operatorami ruchomymi istotne jest rozważenie wzajemnych zachęt dla obecnych posiadaczy widma federalnego w celu dalszego przejścia na CBRS. Ryzyko polityczne i niepewność to główne elementy konkurencyjnych wyzwań w dziedzinie konkurencji. Rozdrobniona struktura rynku krajowego i globalnego pozbawia korzyści skali i zakresu, podnosząc koszty i hamując innowacje w ekosystemie. Ponadto wprowadzenie modeli współdzielenia może mieć wpływ na obecny model MNO w zakresie wyłącznego licencjonowania widma i jego dostępność w przyszłości. Złożoność struktury CBRS, w szczególności SAS, może mieć wpływ na wartość widma i wymagany czas odzyskania inwestycji sieciowych. W dziedzinie kompetencji MNO muszą zwracać uwagę na dynamiczne możliwości potrzebne do wdrażania, zarządzania i optymalizacji wielowarstwowych heterogenicznych sieci w warunkach udostępniania. Tradycyjne wsparcie MNO dla widma 3,5 GHz w ich sieciach ma kluczowe znaczenie dla zachęcenia producentów układów i urządzeń do obsługi całego pasma 3,5 GHz w konkurencyjnych terminalach. Atrakcyjny i dynamiczny rynek widma o potencjalnie niższych kosztach transakcji może się zwiększyć i zmienić konkurencję, np. poprzez wprowadzenie nowych typów operatorów lokalnych i innych domen. Podsumowując, aby zrealizować potencjał biznesowy CBRS, operatorzy MNO mają okazję do jednoczesnego współtworzenia i współtworzenia wartości z podmiotami ekosystemowymi w konkurencyjnym środowisku biznesowym, w którym istnieje współpraca (widmo) i konkurencja (klienci i usługi) równoległe do siebie. MNO są w wyjątkowej sytuacji, aby wykorzystać dodatkowe wielopoziomowe oferty CBRS. Szybszy dostęp do zoptymalizowanego widma zoptymalizowanego pod kątem małych komórek QoS, bez obowiązkowych obowiązków dotyczących zasięgu, pomoże im na czas poradzić sobie z rosnącymi asymetrycznymi potrzebami w zakresie danych. Dodatkowe skalowalne i elastyczne zasoby widma wykorzystujące technologie LTE pozwolą operatorom sieci komórkowych lepiej utrzymać i rozbudować istniejącą bazę klientów, zmieniając popyt i zwyczaje konsumentów, oferując usługi różnicujące i odkrywając nowe pionowe segmenty. W odniesieniu do modeli biznesowych w ramach CBRS, niektóre uwagi dotyczące można przedstawić modele biznesowe MNO i CSP. W przypadku MNO możliwości biznesowe w CBRS przypominają LSA: unikaj kosztów lub uzyskaj dodatkowe możliwości. Jednak klienci MNO mogą teraz być lepiej sprofilowani niż segmentować, jak to ma miejsce w przypadku LSA. Oczekujemy, że rola zarządzania doświadczeniami klientów będzie jeszcze bardziej widoczna w CBRS, a także można oczekiwać, że schematy cenowe i taryfikacyjne operatorów sieci ruchomej będą budowane w celu lepszego profilowania klientów. Ciekawym przypadkiem są nowo powstające wewnętrzne "mikrooperatorzy" lub CSP, którzy mogą w określonych "zamkniętych" lokalizacjach, takich jak centra handlowe lub kampusy, zapewniać doskonałe lokalne usługi. W związku z tym klienci indywidualni mogą korzystać z kontekstowych segmentów klientów za pomocą dostosowanych pakietów usług, takich jak połączona łączność i określona treść. W zamkniętych kontekstach łączność może być oferowana za darmo dla klientów, ponieważ klienci korporacyjni, tacy jak reklamodawcy lub inne strony trzecie, mogą pokryć koszty usługi, a klienci korporacyjni z kolei mogliby być zainteresowani zakupem "hurtowej" łączności z lokalnego CSP. . Można również przewidywać, że operatorzy sieci

ruchomej i dostawcy usług publicznych niekoniecznie zawsze będą ze sobą konkurować, ale będą od siebie zależni. Na przykład, operatorzy sieci ruchomych mogli czasami udostępniać swoje usługi łączności lub usługi billingowe wyłącznie dla dostawców CSP. Omawiając skalowalność modelu biznesowego w artykule CBRS stwierdza się, że koncepcja CBRS ułatwia skalowalność modelu biznesowego nie tylko w modelach biznesowych warstwy łączącej, ale także zapewnia lepsze możliwości modelom biznesowym opartym na treści i kontekście, które pomyślnie się powiodą.

## 11.6 Wyniki

### 11.6.1 Implementacja referencyjna LSA

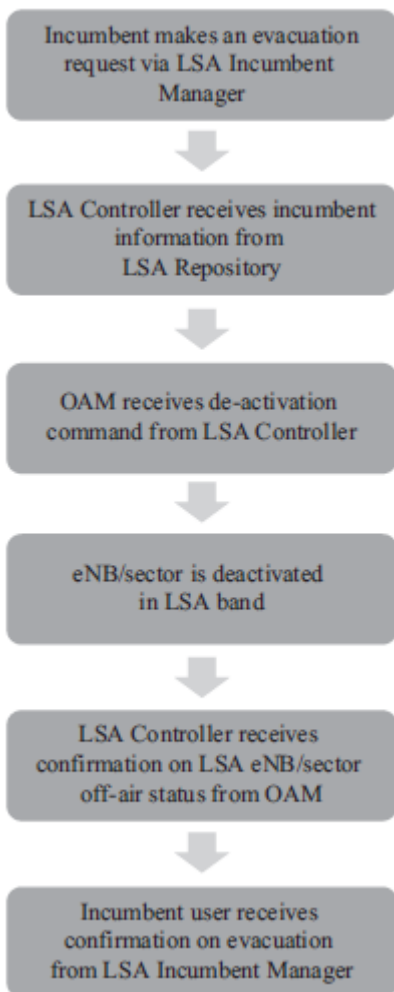
Pierwsza na świecie referencyjna implementacja koncepcji LSA w paśmie 2.3-2.4 GHz została opracowana w Finlandii w 2013 r., Jak pokazano na rysunku 11.10,



w celu dzielenia się usługami LTE i PMSE. Jest to kombinacja elementów, które są dostępne na rynku lub specjalnie zaprojektowane dla LSA pojęcie. Wyposażenie komercyjne w środowisku próbnym LSA obejmuje komórkową sieć LTE, która składa się z dostępow radiowych zgodnych z 3GPP i rzeczywistej sieci rdzeniowej. Od 2013 r. Środowisko testowe LSA zostało rozszerzone o trzy zmienione czasowo LTE (TD-LTE) rozwinięte NodeBs (eNB) z pięcioma sektorami i kilkoma mniejszymi komórkami zlokalizowanymi w Centria w Ylivieska w Finlandii. Te eNB działają w paśmie częstotliwości LSA 2,36-2,40 GHz. Środowisko testowe zawiera także eNB z duplexowaniem z podziałem częstotliwości LTE (FDD-LTE) z czterema sektorami, które zapewniają zasięg na tym samym obszarze z wykorzystaniem pasma 2,1 GHz. Węzły eNB zapewniają wsparcie dla przekazywania TD i FDD oraz równoważenia obciążenia. Wszystkie eNB są połączone z siecią szkieletową LTE i są zarządzane przez komercyjny system operacyjny, administracyjny i zarządzający (OAM). Dostępne są gotowe wielostanowiskowe

urządzenia użytkownika (UE) LTE, które obsługują dual-mode LTE i zapewniają bezproblemowe przekazywanie TD-FDD. Oprócz sprzętu komercyjnego, fińskie środowisko testowe LSA zawiera komponenty specyficzne dla LSA opracowane do testów, w tym LSA Incumbent Manager, Repozytorium LSA i kontroler LSA. LSA Incumbent Manager to specjalne narzędzie opracowane, aby pomóc dostawcy usług PMSE w rezerwowaniu częstotliwości poprzez składanie wniosków o ewakuację w prosty i bezpieczny sposób za pośrednictwem aplikacji na telefon komórkowy lub przeglądarki internetowej. Użytkownik zasiedziałego widma może zarezerwować pasmo poprzez złożenie żądania ewakuacji poprzez ustawienie następujących informacji w przeglądarce internetowej: lokalizacja w interfejsie mapy, typ łącza, zakres częstotliwości nadawania oraz okres czasu rezerwacji. Aplikacja internetowa umożliwi starszemu użytkownikowi zażądanie wielu zabezpieczeń z wyprzedzeniem, zgodnie z planowanym użyciem w przyszłości. Korzystając z aplikacji mobilnej, operator zasiedziały może złożyć wniosek o ewakuację zgodnie z aktualną pozycją. Aplikacja mobilna korzysta z GPS, aby zlokalizować zasiedziałego użytkownika na żądanie. Po uzyskaniu szczegółów rezerwacji LSA Incumbent Manager przesyła informacje zasiedziałego użytkownika do repozytorium LSA. Proces ochrony może również obejmować algorytmy służące do maskowania rzeczywistego wykorzystania obecnego użytkownika, jeśli jest to konieczne. Repozytorium LSA to baza danych zawierająca aktualne informacje na temat pasm widmowych LSA wraz z warunkami korzystania z każdego pasma. Gromadzi, przechowuje i zarządza danymi na temat wykorzystania widma w paśmie LSA dla różnych czasów i lokalizacji geograficznych. Dodatkowo repozytorium LSA zawiera informacje, np. o kontrolerach LSA, licencjach i obecnym łączu i typy sprzętu. Po zmianie aktywności użytkownika zasiedziałego, repozytorium LSA przekazuje go do kontrolera LSA, umożliwiając sieci MNO reagowanie na zmiany. Kontroler LSA zapewnia koncesjonariuszowi LSA środki umożliwiające dostęp do widma i automatyczne kontrolowanie sieci w odniesieniu do zmian w aktywności użytkownika zasiedziałego. W przypadku pojawienia się operatora zasiedziałego w obszarze operacyjnym sieci LSA LTE, kontroler LSA wykorzystuje informacje zasiedziałego użytkownika otrzymane z repozytorium LSA i podane odległości ochrony w celu obliczenia dotkniętych eNB / sektorów w paśmie LSA. Odpowiednio, kontroler LSA wysyła polecenia deaktywacji / aktywacji do OAM. Po usunięciu przez operatora zasiedziałego żądania ewakuacji od obecnego operatora lub z góry ustalonego okresu ewakuacji kontroler LSA wyśle żądanie aktywacji do OAM. Fińskie testy LSA ewoluowały iteracyjnie, aby przedstawić bardziej skomplikowane konfiguracje LSA od czasu jego wprowadzenia w 2013 r. W kilku publicznych testach w latach 2014-2015 przedstawiono ulepszone funkcje obejmujące śledzenie operatora telefonii komórkowej, ulepszone algorytmy ochrony zasiedziałej oraz kontroler LSA zintegrowany jako samoorganizujące się sieci ( SYN) w systemie zarządzania siecią. funkcje różnych komponentów LSA, opisane powyżej, mogą być przedstawione jako różne fazy procesu ewakuacji w paśmie LSA, jak pokazano na rysunku 11.11.





Po pierwsze, operator widma zasiedziatego przesyła żądanie ewakuacji do LSA Incumbent Manager, który z kolei przekazuje te informacje do repozytorium LSA. Repozytorium LSA przechowuje informacje w bazie danych i przekazuje je do odpowiedniego kontrolera LSA. Na podstawie informacji o zasiedziatym użytkowniku otrzymanym z repozytorium LSA, kontroler LSA oblicza, które eNB lub sektory w paśmie LSA zostaną naruszone. Kontroler LSA przekazuje odpowiednie polecenie de aktywacji do OAM. OAM następnie wykonuje plan radiowy deaktywacji dla dotkniętych eNB / sektorów w paśmie LSA, albo blokując eNB / sektory dotknięte LSA, wyłączając w ten sposób ich interfejsy powietrzne, albo stosując płynne wyłączenie, gdy moc eNB / sektorów LSA stopniowo zmniejsza się przed zablokowaniem całkowicie go wyłączyć. W obu przypadkach UE są przekazywane innym sieciom, aby nie zakłócać ich działania. Następnie OAM kończy wykonywanie planu radiowego i rozpoczyna kontrolę stanu eNB / sektora LSA, którego wynik jest wysyłany do kontrolera LSA. Gdy wszystkie potrzebne eNB / sektory LSA osiągną status powietrza, sterownik LSA zostanie zakończona ewakuacja i przekazuje uzupełnione informacje ewakuacyjne do repozytorium LSA. Wreszcie, operator widma zasiedziatego otrzymuje potwierdzenie ewakuacji od obecnego menadżera LSA.

### 11.6.2 Ocena wydajności LSA

Opisana powyżej implementacja referencyjna LSA została przetestowana, a pierwsze pomiary wydajności procesu ewakuacji w paśmie LSA zostały przedstawione w celu oceny zaangażowanych skal czasowych dla różnych faz przepływu procesu przy użyciu oryginalnej konfiguracji LSA.

Pomiary uwzględniają blokowanie eNB /sektory w dwóch oddzielnych przypadkach: normalna ewakuacja, w której jeden eNB / sektor jest zablokowany lub ewakuacja awaryjna, w której zablokowane są wszystkie 3 eNB / 5 sektorów. Po rozważeniu ewakuacji wszystkich sektorów za pomocą pojedynczego polecenia w środowisku próbnym LSA (tj. Ewakuacji awaryjnej), odprawa zespołu trwa prawie w tym samym czasie, jednak wymaga więcej niż 10 sekund, aby potwierdzenie dotarło do operator zasiedziały. W bardziej szczegółowej analizie skali czasu wyniki pomiarów można podzielić na opóźnienia z platformy LSAresearch i opóźnienia pochodzące z komercyjnego OAM LTE. Pierwotne opóźnienie dla jednego eNB / sektora 61 s składa się z 20,6 s z badań i 40,4 s z platform komercyjnych. Pierwotne opóźnienie ewakuacji ewakuacyjnej z 76 s obejmuje 33 s z badań i 43 s z platform komercyjnych. Opóźnienia po stronie sprzętu komercyjnego wynikają z rozpoczęcia uruchamiania planowania planu radiowego do momentu, w którym moc nadawania LTE zniknie z kanału, po czym dostarczanie planu radiowego nadal trwało przez 9-10 sekund. Opóźnienie z komercyjnego sprzętu można zminimalizować za pomocą wcześniej zatwierdzonych planów radiowych, aby skrócić czas wykonania operacji dostarczania. Można również zauważyć, że dostarczenie planu radiowej deinstalacji wielosieciowej zajmuje tylko około 3 więcej sekund niż pojedyncza komenda lokacji, co jest obiecujące dla zarządzania większymi sieciami LSA. Powyższe pomiary pochodziły z oryginalnych konfiguracji LSA w latach 2013-2014. Kontroler LSA z ulepszonymi funkcjami ochrony zasiedziały. W nowej konfiguracji LSA, LSA Controller został wdrożony jako rozwiązanie SON zintegrowane z komercyjnym systemem zarządzania siecią, w tym ulepszone algorytmy sterowania mocą w celu zoptymalizowania stref ochrony dla istniejącej ochrony przy maksymalizacji dostępności pasma LSA dla MNO. Pomiary wydajności ulepszanego układu LSA pokazały, że do ewakuacji pierwszej komórki zajęło 21 s od zgłoszenia żądania ewakuacji aż do wyczyszczenia pasma. Dodatkowe 13 s były potrzebne, dopóki potwierdzenie operacji ewakuacji nie było widoczne dla operatora zasiedziały. Średni czas ewakuacji dla optymalizacji strefy ochronnej wynosił 24 s, a wdzięczna rekonfiguracja mocy 58 s, które są uważane za odpowiednie szczególnie w przypadku użycia PMSE w paśmie 2,3-2,4 GHz. Fińskie środowisko testowe LSA można rozbudować, aby pokazać również trójwymiarowy model udostępniania. Pierwsze publiczne testy CBRS zostały zaprezentowane na konferencji IEEE GLOBECOM 2015, prezentującej różne poziomy praw dostępu i sposób, w jaki sieć LTE może reagować na obecną działalność

### **11.7 Perspektywy na przyszłość i wnioski**

W niniejszej części podsumowano status europejskiej koncepcji LSA w paśmie 2,3-2,4 GHz oraz trzypoziomą koncepcję współdzielenia widma w USA w celu zapewnienia łączności CBRS w paśmie 3,5 GHz z wykorzystaniem SAS z perspektywy regulacji, biznesu i technologii. Dotychczasowe rozwinięcie koncepcji LSA koncentrowało się na wykorzystaniu pasma 2,3-2,4 GHz przez systemy komunikacji mobilnej na zasadzie współdzielonej, zapewniającej przewidywalne warunki QoS dla operatorów zasiedziały, które różnią się w zależności od konfiguracji krajowej. Możliwość udostępnienia LSA w paśmie 2,3- 2,4 GHz do dzielenia się została udowodniona w Finlandii, a ramy regulacyjne są gotowe do przyjęcia. W przypadku krajowego rozmieszczenia potrzebna jest bardziej szczegółowa analiza konkretnych operatorów w danym kraju, w tym wyprowadzenie wystarczających i praktycznych kryteriów ochrony. Kolejnym obszarem zastosowań LSA, który będzie badany w Europie po pasmie 2,3-2,4 GHz, jest pasmo 3,6-3,8 GHz, w którym rozpoczęto badania.

## **Epilog: bezprzewód poza 5G**

### **12.1 Wizja bezprzewodowa powyżej 5G**

Zakładając sukces, 5G wprowadzi nie tylko nowy sposób myślenia w projektowaniu i optymalizacji sieci bezprzewodowych, ale także nową erę, w której "łączość bezprzewodowa jako narzędzie" całkowicie zmieni sposób, w jaki ludzie, rzeczy i (dowolne) urządzenia aplikacyjne komunikować, przechowywać i przetwarzać informacje. Szybkość (szybkość transmisji), objętość danych, odległość, zwinność (zwłoka), bezpieczeństwo i niezawodność stają się "nie problemami", "nieskończonymi" możliwościami, a nieograniczona łączność inicjuje zmianę paradygmatu ekosystemu, zgodnie z którą sieć infrastruktura jako całość jest zawsze dostępna, skalowalna i funkcjonująca "w tle", podczas gdy dostęp bezprzewodowy jest wszechobecny, a technologia agnostyczna (niezależna). Ta nowa funkcja jakości doświadczenia wyraźnie przesuwą środek ciężkości ze strony sieci (infrastruktury) w stronę końca użytkownika (krawędź sieci) oraz od technologii radiowych i sieciowych do aplikacji. Rozprzestrzenianie się wielu nowych aplikacji, obejmujących zróżnicowany krajobraz zastosowań branż pionowych, przełamuje bariery społeczno-technologiczne, jak nigdy dotąd: łączność bezprzewodowa, poprzez udostępnianie zawsze, wszędzie i dla wszystkich, może być traktowana jako pewnik, a nie tylko korzyść technologiczna jest obecnie uważana za prawo społeczne i ludzkie. Co więcej, łączność bezprzewodowa jako narzędzie będzie katalizować rynek, pomagając w realizacji wizji Industry 4.0: cyfryzacji sektora produkcyjnego. Opierając się na zakłóceniach, takich jak przetwarzanie dużych ilości danych, mała moc obliczeniowa, inteligencja analityki dużych zbiorów danych i zaawansowana interakcja człowiek-maszyna, oczekuje się, że łączność bezprzewodowa będzie odgrywać kluczową rolę we wspieraniu tworzenia wzrostu w każdym sektorze gospodarki, społeczeństwie, i kultura. Ale czy sukces 5G będzie oznaczał koniec badań nad siecią bezprzewodową, jakie znamy? Czy imponująca podróż ewolucyjna, która rozpoczęła się ponad 30 lat temu, wraz z wprowadzeniem komunikacji mobilnej i bezprzewodowego dostępu lokalnego, ma wejść w fazę "dojrzwania" / nasycenia? Czy jest dokładnie odwrotnie? Optymistyczny scenariusz przewiduje, że 5G otworzy nowe, zielone kierunki badań terenowych, które tylko pod pryzmatem nowego, bezprzewodowego myślenia mogą uwolnić ich prawdziwy potencjał. W tym kontekście oczekuje się, że sieci wykraczające poza 5G zapewnią niespotykaną dotąd doskonałość w zakresie wydajności, nie tylko poprzez ukierunkowanie na szybkości transmisji danych w systemie Terabitper-sekund (Tbps), ale także z natury wspierając szeroki zakres dynamicznych nowych scenariuszy użytkowania i aplikacji, które łączą te ekstremalne szybkości transmisji danych ze zwinnością, niezawodnością, zerowym czasem reakcji i sztuczną inteligencją. Obecność wirtualna, drukowanie 3D, systemy cyberprzestrzeni dla inteligentnego transportu i przemysłu 4.0 to tylko kilka przykładów kilku bardzo trudnych, przewidywanych przypadków użycia. Próbuując wyartykułować wizję ponad 5G w dzisiejszych kategoriach, na pewno zawierałoby trzy główne atrybuty:

- Łącza bezprzewodowe Tbps, zawsze i wszędzie dostępne;
- Prawdziwa zbieżność między optycznymi (niezawodność) i bezprzewodowymi (elastyczność);
- Aplikacje sztucznej inteligencji dla Cyber-fizycznych systemów (aby odnieść sukces IoT).

Chociaż 5G wydaje się być bardziej niż chętni do objęcia kilku zasad projektowania gier, takich jak wirtualizacja i softwarzacja, w celu zwiększenia skalowalności, elastyczności i wydajnego wykorzystania zasobów, można łatwo zrozumieć, że podstawowe ograniczenia wydajności związane z dostępną przepustowością, transmisją i opóźnienie przetwarzania oraz koszt i zużycie energii wciąż definiują obwiednię możliwości 5G. Aby przełamać te bariery technologiczne w sieciach innych niż 5G, należy wnieść niewiele zbadanych zasobów i technologii (np. Komunikację THz) do walidacji i

eksploatacji, kierując badania ku zmniejszeniu ryzyka koncepcji technologicznych, komponentów, architektur i koncepcji systemów (np. przetwarzanie / optymalizacja, ekstremalne udostępnianie zasobów i commoditization). Potrzebne są innowacyjne wspólne badania, ocena i projektowanie modeli teoretycznych, dostosowane i wspierane przez ekstrakcję i zatwierdzanie parametrów eksperymentalnych z tego powodu.

## **12.2 Oczekiwania i wyzwania dla sieci bezprzewodowych powyżej 5G**

Biorąc pod uwagę trendy w rozwoju sieci bezprzewodowej 5G i przyszłość w perspektywie 2030 r. I później pojawia się kilka podstawowych pytań, otwierających nowe kierunki, rosnące oczekiwania i motywujące ekscytujące badania i innowacje. Poniższe pytania zostały zadane liderom bezprzewodowych badań, znanym na całym świecie z innowacji w zakresie zrozumienia podstaw i wprowadzenia nowych koncepcji i technologii do realizacji.

1. Czy nauka podstawowa / podstawowa ma do odegrania rolę w przyszłości bezprzewodowej? Jakie dyscypliny są potencjalnymi kandydatami?
2. Jakie nieodkryte technologie mogłyby odegrać kluczową rolę w sieci bezprzewodowej poza 5G?
3. Oczekuje się, że zmiany paradygmatu w łańcuchu wartości w przemyśle bezprzewodowym, wywołane przez 5G, będą katalizować przyszłą sieć bezprzewodową? Czy można przewidzieć pojawienie się nowych graczy i nowych rynków niszowych w epoce 5G, które mogłyby radykalnie zmienić krajobraz branży ICT?