Non Orthogonal Multiple Access for 5G and beyond

DIET- Sapienza University of Rome

mai.le.it@ieee.org

November 23, 2018
Introduction

A reference mathematical model of Code-domain NOMA
Theoretical Analysis of Low-density NOMA with perfect CSI
Theoretical Analysis of Low-density NOMA without perfect CSI
Results and discussions
Discussion

Outline

1. Introduction
   - 5G Era
   - Concept of NOMA
   - Classification of NOMA
   - CDM-NOMA in 5G-NR
   - Low-density spreading CDM-NOMA
   - Novelties and Contributions

2. A reference mathematical model of Code-domain NOMA

3. Theoretical Analysis of Low-density NOMA with perfect CSI
   - Optimum decoding
   - Linear decoding

4. Theoretical Analysis of Low-density NOMA without perfect CSI
   - Channel model: Rayleigh block-fading assumptions
   - Capacity lower bound
     - Pilot-based channel model assumptions
     - Pilot-based system model
     - Optimization for pilot-based low-dense NOMA

5. Results and discussions
   - The impact of number of coherence symbols $n_b$
   - The impact of number of users $K$
   - The impact of system load $\beta$

6. Discussion
5G era

**Figure**: Evolution from 1G to 5G

1

---

1Sources:https://www.viracure.com/blog/from-1g-to-5g/
Two key-requirements of 5G are:

- High spectral efficiency
- Massive connectivity

Some potential candidates have been proposed for 5G such as:

- Massive MIMO
- Millimeter wave communications
- Ultra dense network
- Non-orthogonal multiple access (NOMA): (3GPP-LTE-A) standard, the next general digital TV standard (ATSC 3.0), and the 5G New Radio (NR) standard

---

Concept visualization of NOMA

Orthogonal Multiple Access (OMA)

Number of users K

Number of resource elements N
(time, frequency, code slots)
Concept visualization of NOMA

Number of users $K$ increases?

Number of resource elements $N$ (time, frequency, code slots)

How other remain users???
Introduction
A reference mathematical model of Code-domain NOMA
Theoretical Analysis of Low-density NOMA with perfect CSI
Theoretical Analysis of Low-density NOMA without perfect CSI
Results and discussions
Discussion

5G Era
Concept of NOMA
Classification of NOMA
CDM-NOMA in 5G-NR
Low-density spreading CDM-NOMA
Novelties and Contributions

Concept visualization of NOMA

Non-Orthogonal Multiple Access (NOMA)

Number of users $K$ increases?

Number of resource elements $N$ (time, frequency, code slots)

Non Orthogonal Multiple Access for 5G and beyond
Overview of NOMA

- To control collisions in NOMA, it is feasible to share the same signal dimension among users and exploit either power (PDM-NOMA) or code (CDM-NOMA), or space (SDMA) domains.

- **CDM-NOMA** can be inferred from conventional CDMA via spreading matrix, which can be classified as:
  - single-carrier vs. multi-carrier NOMA (based on adopted waveform)
  - dense vs. low-dense NOMA (based on sparsity of spreading matrix)

---

CDM-NOMA in 5G-New Radio (5G-NR)

3GPP RAN NR Standardization

Now 2018

2016
2017
2019
2020

LTE Adv. Pro
5G Phase 1
5G Phase 2

5G RAN NR

Release 14
Release 15
Release 16

First 5G NR Network Deployments

CDM-NOMA is a study item for 5G-NR

Figure: Timeline for 5G-NR

Sources: Rohde & Schwarz MNT May 2018

Non Orthogonal Multiple Access for 5G and beyond
CDM-NOMA in 5G-New Radio (5G-NR)

3GPP RAN NR Standardization

2016 2017 Now 2018 2019 2020

LTE Adv. Pro 5G Phase 1 5G Phase 2
Rel-14 Release 15 Release 16

First 5G NR Network Deployments

TS 38.211 NR radio framework

CDM-NOMA is a study item for 5G-NR

Figure: Timeline for 5G-NR

Sources: Rohde & Schwarz MNT May 2018
LDS-NOMA: The idea is to replace dense-spreading sequences of DS-CDMA by sparse counterparts, such that only a part of REs is filled with nonzero entries, while others are zero.
Novelties and Contributions

The major contributions and novelties of this work are as follows:

- A unified mathematical framework for code-domain NOMA.
- The spectral efficiency for Low-density NOMA in the fading channel with linear detection is derived in closed-form, showing that sparse signaling outperforms dense signaling when the network is overloaded ($K > N$).
- The spectral efficiency with optimum detection is derived in closed-form, by finding the limiting spectral distribution of a matrix ensemble that jointly describes spreading and fading: this is a mathematical result of independent interest.
- The results provide an insight into the design of signaling in dense networks, as envisioned in the uplink IoT scenario, sparse signaling can achieve a rate several times larger than that achievable via dense signaling.
A reference mathematical model of Code-domain NOMA

The mathematical model for Code-domain NOMA is:
\[ y = SAb + n, \]
where:
- \( y \in \mathbb{C}^N \) is the received signal with \( N \) signal dimensions (chips)
- \( b = [b_1...b_K]^T \in \mathbb{C}^K \) is the vector of symbols transmitted by \( K \)-users with the power constraint \( \mathbb{E}[s_K] \leq \varepsilon \)
- \( S = [s_1...s_K] \in \mathbb{R}^{N \times K} \) is the spreading matrix, where its \( k_{th} \) column is the spreading sequence \( s_K \) of the \( k_{th} \) user and \( ||s_K|| = 1 \)
- \( A = diag[A_1...A_K] \in \mathbb{C}^{K \times K} \) represents the channel
- \( n \) is a circularly symmetric Gaussian vector with zero mean and covariance \( N_0 I \)
Theoretical analysis for Low-density NOMA with perfect CSI

Assumptions:

- Given $N_s$ the number of nonzero elements in each spreading sequence.
  - $N_s = N$: dense NOMA (DS-CDMA)
  - $N_s < N$: low-dense NOMA (LDS-CDMA).
- In massive communications, the behavior of the system should be considered in the asymptotic limit region, where $N \rightarrow \infty$ and $K \rightarrow \infty$, while the ratio $K/N = \beta$ (system load) remains finite.\cite{VS99}.
  - $\beta < 1$: underloaded system
  - $\beta > 1$: overloaded system
- Optimum and linear decoding are considered in all cases.
Introduction

A reference mathematical model of Code-domain NOMA

Theoretical Analysis of Low-density NOMA with perfect CSI

Theoretical Analysis of Low-density NOMA without perfect CSI

Results and discussions

Discussion

Achievable rates of low-dense NOMA with optimum
decoding for $N_s = 1$

For AWGN channels [FdB15]:

$$C_N^{\text{opt}}(\gamma) = \frac{1}{N} \log_2 \det[\mathbf{I} + \gamma \mathbf{S} \mathbf{S}^H] \rightarrow C^{\text{opt}}(\beta, \gamma) = \sum_{k \geq 0} \frac{\beta^k e^{-\beta}}{k!} \log_2(1 + k \gamma).$$  

(1)

For flat-fading channels (major results):

$$C_N^{\text{opt}}(\gamma) = \frac{1}{N} \sum_{n=1}^{N} \log_2 \left(1 + \gamma \lambda_n(\mathbf{S} \mathbf{H} \mathbf{H}^H \mathbf{S}^H)\right),$$

$$P \rightarrow C^{\text{opt}}(\beta, \gamma) = e^{-\beta} \int_0^\infty \sum_{k=0}^{\infty} \frac{\beta^k x^{k-1} e^{-x}}{k! \Gamma(k)} \log_2(1 + \gamma x) dx.$$  

(2)

where $\{\lambda_n(\mathbf{S} \mathbf{H} \mathbf{H}^H \mathbf{S}^H)\}$ is the set of eigenvalues of the matrix.
Achievable rates of low-dense NOMA with optimum decoding for $N_s = 1$

![Graph: Optimum capacity $C^{opt}(\text{bits/s/Hz})$ vs. load factor $\beta = K/N_s$](image)
Achievable rates of low-dense NOMA with linear decoding for $N_s = 1$

As demonstrated in [FdB15], the distinguished difference of LDS-CDMA to DS-CDMA is that all linear receiver (ZF, SUMF, MMSE) result in the same mutual information. Achievable rates of low-dense NOMA for AWGN channels [FdB15]:

$$R_{TH}^{ZF} = R_{TH}^{SUMF} = R_{TH}^{MMSE} = \beta \sum_{k \geq 0} \frac{\beta^k e^{-\beta}}{k!} \log_2 \left( 1 + \frac{\gamma}{k \gamma + 1} \right)$$  \hspace{1cm} (3)

For flat-fading channels, the closed-form expression is as followed (major results):

$$R_{LDS}^{SUMF}(\beta, \gamma) = \beta \log_2 e \int_0^1 e^{-t} \left( \frac{\beta + \frac{1}{(1-t)\gamma}}{1 - t} \right) dt.$$  \hspace{1cm} (4)
Achievable rates of Low-dense NOMA with linear decoding for $N_s = 1$ without fading
Achievable rates of low-dense NOMA with linear decoding for $N_s = 1$ with fading
Channel model: Rayleigh block-fading assumptions

Continuous vs. block-fading model is linked via the duality property

\[ n_b = \frac{1}{2f_m T_s}, \]

- \( n_b \) is the number of coherent symbols within a block in which the channel is considered stationary (\( T_s \) is the symbol period for continuous-fading model, respectively).
  - For 3GPP LTE and IEEE 802.16 WiMAX: \( n_b \) may range from unity to several hundreds.
  - For 5G-NR: \( n_b \) may range from unity to thousands.
- \( f_m = \nu f_c / c \) is the maximum Doppler frequency:
  - \( f_c \) being the carrier frequency,
  - \( \nu \) is the velocity of interest and \( c \) is the speed of light.

---

$n_b$ in 5G-NR scenario

The calibration of $n_b$ for New Radio (5G-NR) is based on:

- $f_c = 1–100$ [GHz], with the deployment of macro sites at lower frequencies, and micro and pico sites at higher frequencies $^7$.

- Supported mobility speeds $\nu$ are up to 500 km/h, while the vehicular velocities of interest are about 120 km/h $^7$.

- Numerology$^8$, i.e. the subcarrier spacing $\Delta_f$, is the most distinguished feature in frame structure of 5G-NR compared to LTE.
  - In LTE: $\Delta_f = 15$ [kHz],
  - In 5G-NR: $\Delta_f = 2^\mu \times 15$ [kHz] with $\mu = 0–4$, corresponding to the symbol period $T_s = 4\mu s$ to 66.7$\mu s$.

$\rightarrow$ The range of $n_b$ in the 5G-NR context is from one to thousands!


$^8$ 3rd Generation Partnership Project (3GPP), TS 38.211. NR; Physical channels and modulation, 2017
Capacity lower bound: Pilot-based channel model

- A pilot-based scheme includes a pilot phase and a data transmission phase, being implemented after a MMSE channel estimation obtained by the pilot phase.

- Among the total $n_b$ symbols of a fading block:
  - $n_p$ symbols, called pilot symbols, are allocated for learning the channel,
  - remaining $(n_b - n_p)$ symbols are dedicated for data transmission.

- Three assumptions rule the pilot-based scheme in this work:
  - $n_b > 2K$ since $n_b$ is much greater than both $\{K, N\}$,
  - perfect channel estimation (via pilot phase) is assumed,
  - each transmission is assumed to be self-contained, that is, both pilot and data phases are done within a block of $n_b$ symbols.
Pilot-based system model

- **Pilot phase:**
  \[ Y_p = SAP + N_p, \]
  where \( P \in \mathbb{C}^{K \times n_p} \) is the matrix of known pilot symbols replacing the transmitted signal \( X \) with the constraint \( PP^* = n_p I \), \( Y_p \) and \( N_p \) are matrices of size \( N \times n_p \).

- **Data phase:** a similar equation to the perfect CSI model can be applied as such:
  \[ Y_d = SAX_d + N_d, \]
  with new dimensions of output and input being \( Y_d \in \mathbb{C}^{N \times (n_b - n_p)} \) and \( X_d \in \mathbb{C}^{K \times (n_b - n_p)} \), respectively.
Optimization for pilot-based low-dense NOMA

- The effect of pilot power allocation:
  - No constraint on pilot power allocation:
    \[ C_{LB}(\beta, \gamma_{eff}) = \left(1 - \frac{K}{n_b}\right) \sum_{k \geq 1} \frac{e^{-\beta} \beta^k}{k!} \int_0^\infty \frac{e^{-\lambda} \lambda^{k-1}}{(k-1)!} \log_2(1 + \gamma_{eff} \lambda) \partial\lambda, \]
    where \( \gamma_{eff} = \frac{n_b \text{SNR}}{n_b - 2K} (\sqrt{\alpha} - \sqrt{\alpha - 1})^2. \)
  - With constraint on pilot power allocation:
    \[ \max_{1 \leq n_p \leq n_b} \left(1 - \frac{n_p}{n_b}\right) C\left(\frac{\text{SNR}^2 n_p/K}{1 + \text{SNR}(1 + n_p/K)}\right). \]

- Optimizing the number of pilot symbols yields \( n_p^{opt} = K, \) given
  - if \( n_p \) is too small, the time dedicated to channel sounding may be insufficient to provide good estimates,
  - if \( n_p \) is too large, the data transmission rates will reduce.
In this section, capacity bounds of low-dense NOMA are analyzed as a function of:

- number of coherence symbols $n_b$
- number of users $K$
- system load $\beta = K/N$

The analysis of lower bound focuses on the case of no pilot power constraint. An example with pilot power constraint is shown for reference in the last section.
The impact of number of coherence symbols $n_b$

**Figure:** Capacity bounds (bits/s/Hz) of low-dense NOMA as a function of $E_b/N_0$ for fixed $\beta = 1$, with upper bound defined by coherent capacity (solid line) and lower bounds by a pilot-based scheme with fixed $n_p = K = 10$ for different values $n_b$. 

Non Orthogonal Multiple Access for 5G and beyond
The impact of number of users $K$

**Figure**: Capacity lower bounds (bits/s/Hz) of low-dense NOMA pilot-based scheme with fixed $\beta = 1$ and $E_b/N_0 = 10$ [dB] as a function of $n_b$ with different values $K = \{10, 50, 100\}$.
The impact of system load $\beta$

**Figure:** Capacity bounds of low-dense NOMA scheme with upper bound being the coherent capacity and lower bound obtained with a pilot-based scheme as a function of $\beta = K/N$ and fixed $E_b/N_0 = 10$ [dB].
Discussions

For perfect CSI, results of CDM-NOMA can be drawn as follows:

- For optimum decoding, achievable rates of dense systems are always higher than low-dense counterparts, despite fading, \( \beta \) and \( N_s \).

- For linear decoding, achievable rates in the low-dense regime are equal for all linear decoders, and are shown higher than those of dense systems for \( \beta \in [1.2, 5] \), and increase with fading.

For no CSI, lower capacity bounds reach the upper bound in 5G-NR context under some conditions (\( n_b \) high, \( K \) low).

In conclusion, by changing the spreading strategy from dense to low-dense, specific theoretical limits hold, showing that, to obtain higher achievable rates for linear decoders while still enjoying the lower receiver complexity, it is advisable to adopt sparse communications.
Main References


