

Intelligent Massive NOMA towards 6G

Tutorials of PIMRC2020, London, UK

Dr. Yuanwei Liu, Prof. Zhiguo Ding and Prof. Lajos Hanzo

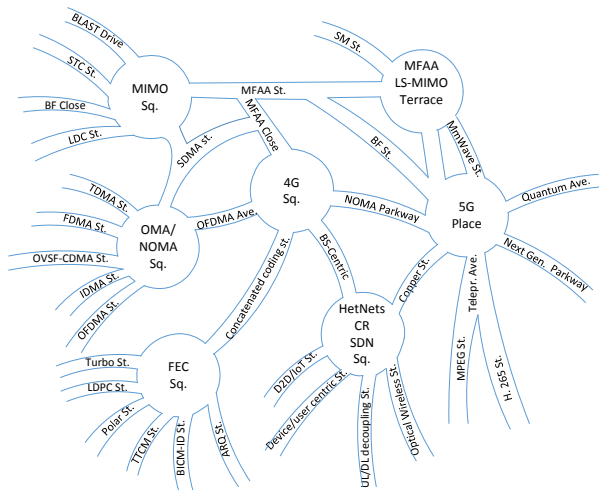
Queen Mary University of London, UK
The University of Manchester, UK
University of Southampton, UK

yuanwei.liu@qmul.ac.uk zhiguo.ding@manchester.ac.uk hanzo@soton.ac.uk

Aug. 31st, 2020

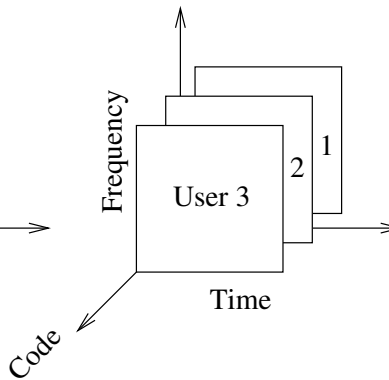
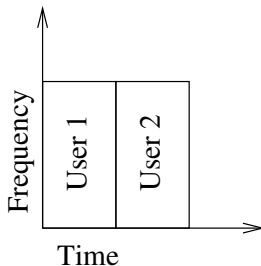
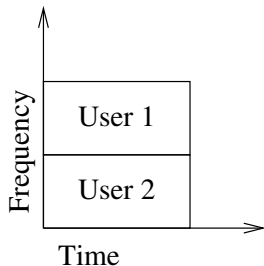
1 Wireless Standardization History: OMA vs NOMA

Brief History of Wireless Standardization

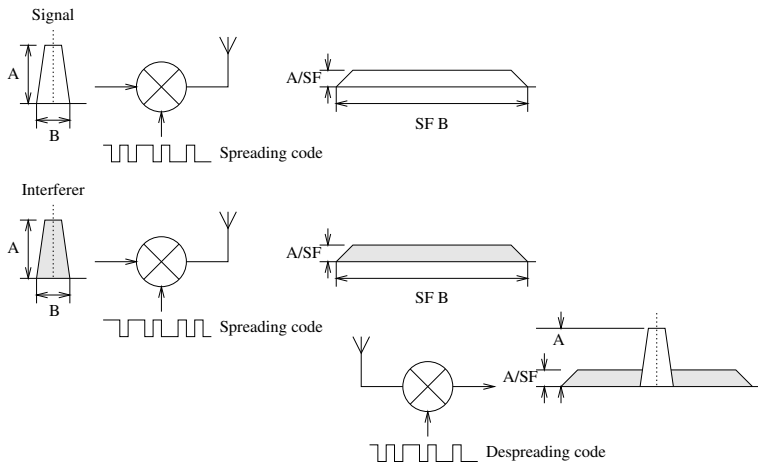


[1] Y. Liu, Z. Qin, M. ElKashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-Orthogonal Multiple Access for 5G", *Proceedings of the IEEE*; Dec 2017.

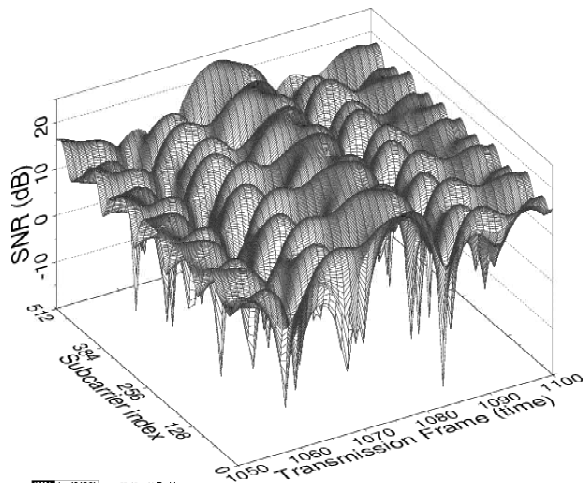
Orthogonal multiple access: FDMA, TDMA and CDMA



Intentional DS-SS-CDMA Spreading

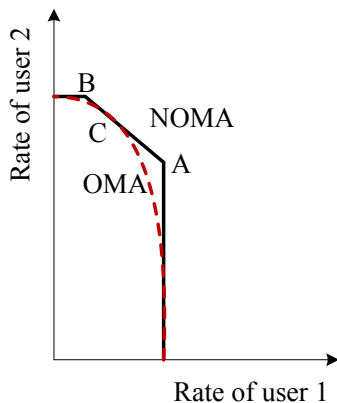


Unintentional Spreading in the FD

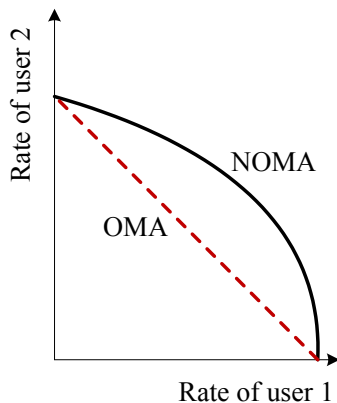


GM 1 Jun 15 16:21 snr_800_1000_16dB_2dlscape

Capacity of OMA vs. NOMA in AWGN channel: (a) Uplink; (b) Downlink.



(a)



(b)

Diverse NOMA contributions

- R. Zhang and L. Hanzo, "A unified treatment of superposition coding aided communications: Theory and practice," *IEEE Commun. Surveys Tutorials*, vol. 13, no. 3, pp. 503–520, Mar. 2011.
- P. Botsinis, D. Alanis, Z. Babar, H. Nguyen, D. Chandra, S. X. Ng, and L. Hanzo, "Quantum-aided multi-user transmission in non-orthogonal multiple access systems," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2016.
- A. Wolfgang, S. Chen, and L. Hanzo, "Parallel interference cancellation based turbo space-time equalization in the SDMA uplink," *IEEE TWC*, vol. 6, no. 2, pp. 609–616, Feb. 2007.
- L. Wang, L. Xu, S. Chen, and L. Hanzo, "Three-stage irregular convolutional coded iterative center-shifting K-best sphere detection for soft-decision SDMA-OFDM," *IEEE TVT*, vol. 58, no. 4, pp. 2103–2109, May 2009.
- S. Chen, L. Hanzo, and A. Livingstone, "MBER space-time decision feedback equalization assisted multiuser detection for multiple antenna aided SDMA systems," *IEEE TSP*, vol. 54, no. 8, pp. 3090–3098, Aug. 2006.
- L. Hanzo, S. Chen, J. Zhang, and X. Mu, "Evolutionary algorithm assisted joint channel estimation and turbo multi-user detection/decoding for OFDM/SDMA," *IEEE TVT*, vol. 63, no. 3, pp. 1204–1222, Mar. 2014.
- S. Chen, A. Wolfgang, C. J. Harris, and L. Hanzo, "Symmetric RBF classifier for nonlinear detection in multiple-antenna-aided systems," *IEEE TNN*, vol. 19, no. 5, pp. 737–745, May 2008.

Diverse NOMA contributions

- S. Chen, A. Livingstone, H. Q. Du, and L. Hanzo, "Adaptive minimum symbol error rate beamforming assisted detection for quadrature amplitude modulation," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1140–1145, Apr. 2008.
- J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Turbo multi-user detection for OFDM/SDMA systems relying on differential evolution aided iterative channel estimation," *IEEE Trans. Commun.*, vol. 60, no. 6, pp. 1621–1633, Jun. 2012.
- J. Zhang, S. Chen, X. Mu, and L. Hanzo, "Joint channel estimation and multi-user detection for SDMA/OFDM based on dual repeated weighted boosting search," *IEEE Trans. Veh. Technol.*, vol. 60, no. 7, pp. 3265–3275, Jun. 2011.
- C.-Y. Wei, J. Akhtman, S.-X. Ng, and L. Hanzo, "Iterative near-maximum-likelihood detection in rank-deficient downlink SDMA systems," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 653–657, Jan. 2008.
A. Wolfgang, J. Akhtman, S. Chen, and L. Hanzo, "Iterative MIMO detection for rank-deficient systems," *IEEE Signal Process. Lett.*, vol. 13, no. 11, pp. 699–702, Nov. 2006.
- L. Xu, S. Chen, and L. Hanzo, "EXIT chart analysis aided turbo MUD designs for the rank-deficient multiple antenna assisted OFDM uplink," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2039–2044, Jun. 2008.

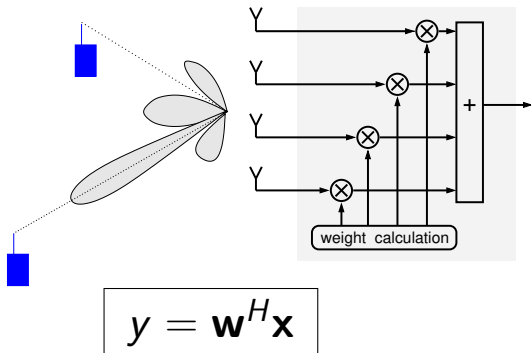
Diverse NOMA contributions

- A. Wolfgang, J. Akhtman, S. Chen, and L. Hanzo, "Reduced-complexity near-maximum-likelihood detection for decision feedback assisted space-time equalization," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2407–2411, Jul. 2007.
- J. Akhtman, A. Wolfgang, S. Chen, and L. Hanzo, "An optimized-hierarchy-aided approximate Log-MAP detector for MIMO systems," *IEEE TWC*, vol. 6, no. 5, pp. 1900–1909, May 2007.

NOMA Beamforming Example

Uplink/Downlink Beamforming

- Why?
Increase of capacity
- How?
Spatially separated interfering signals are suppressed



MMSE Based Beamforming

- Weights are calculated in order to minimize:

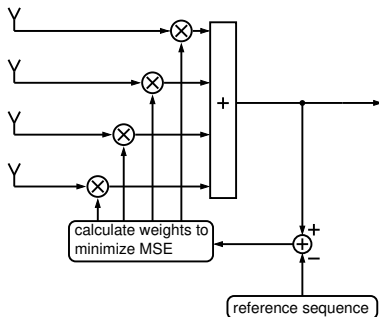
$$\epsilon(t)^2 = (\mathbf{w}^H \mathbf{x}(t) - r(t))^2$$

$\mathbf{w}$: Beamformer weights

$\mathbf{x}(t)$: Channel output

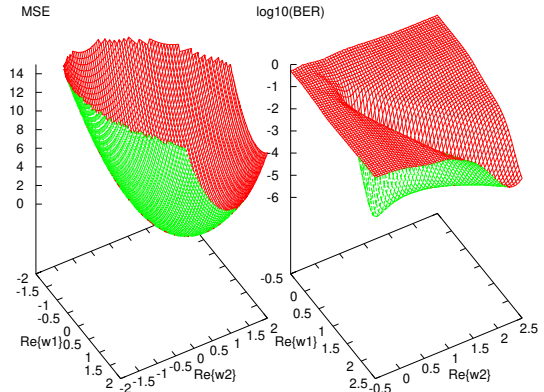
$r(t)$: Reference symbol

- For AWGN channels MMSE weights can be calculated using a closed form expression
- Realizations: LMS, RLS, SMI



MSE and BER Surfaces at the Output of a $[5 \times 2]$ NOMA Beamformer

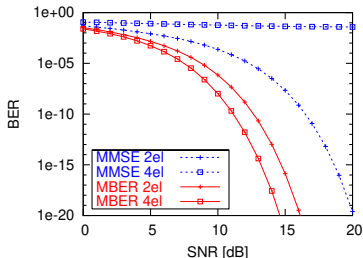
Error surfaces at the receiver's output calculated for five BPSK modulated sources having equal received power and communicating over AWGN channels at SNR=10 dB.



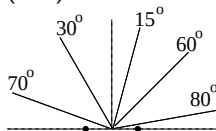
The imaginary part of both weights of the 2-element array was fixed.

MMSE vs MBER NOMA Beamforming

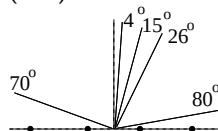
- Test case: BPSK modulated sources having equal received power and communicating over AWGN channels
- MMSE solution calculated analytically
- MBER solution obtained with the aid of conjugate gradient algorithm



Scenario S
(2el.)

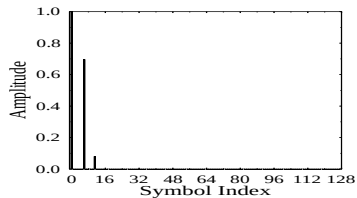
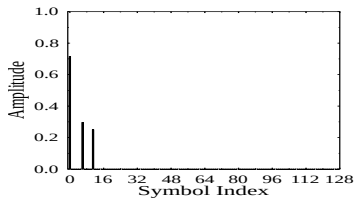
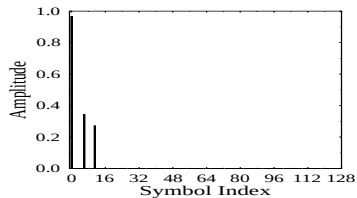
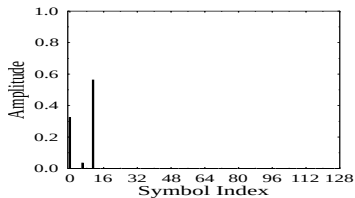


Scenario U
(4el.)



NOMA SDMA Example

Evolution from CDMA-NOMA to SDMA-NOMA



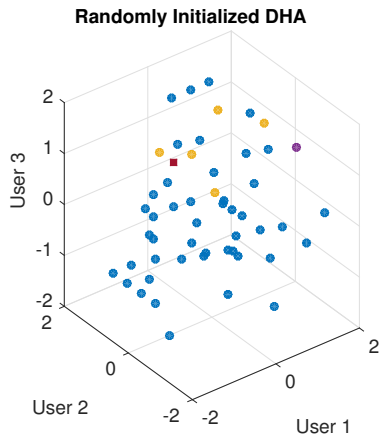
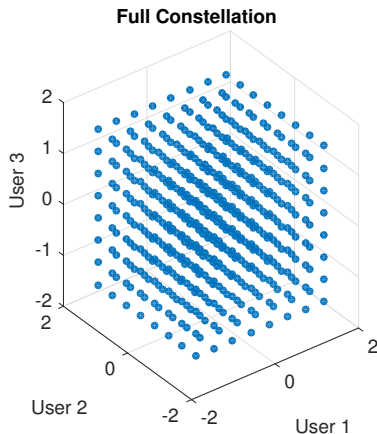
Quantum-Search Aided MUD in NOMA

Multiple Access	SDMA-OFDM
Number of Users	$U = 3$
Number of AEs at the BS	$P = 1$
Normalized User-Load	$U_L = U_q/P = 3$
Modulation	8-PAM $M = 8$
E_b/N_0	0 dB
Channel Code	Turbo Convolutional Code, 8 trellis states, $R = 1/2$
Channel Model	Extended Typical Urban (ETU)
Mobile Velocity	$v = 130$ km/h
Carrier Frequency	$f_c = 2.5$ GHz
Sampling Frequency	$f_s = 15.36$ GHz (77 delay taps)
Doppler Frequency	$f_d = 70$ Hz
Number of Subcarriers	$Q = 1024$
Cyclic Prefix	CP = 128
Interleaver Length	10 240 bits per user
Channel Estimation	Perfect

Quantum-Search Aided MUD in NOMA

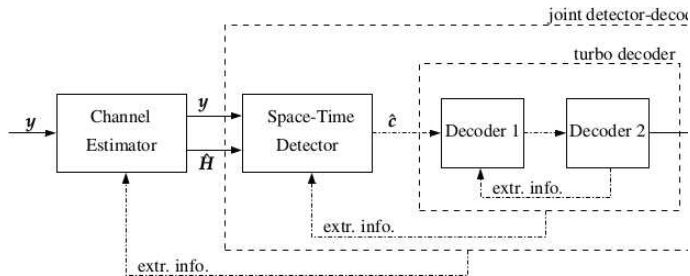
- There are $8^3 = 512$ symbols in the full constellation, while 53 and 46 symbols are obtained by the randomly-initialized and ZF-initialized DHA, respectively.
- The purple circle denotes the random initial input, or the ZF detector's output, which may be used as an initial input. The ZF is as bad as the random one in this rank-deficient scenario.
- By using the DHA, we find symbols better than the previously found symbols, which are denoted by the yellow circles in the 3D figure.
- But we also find symbols that are "worse" than the previously found symbols, as represented by the blue circles in the 3D figure.
- The red square is the optimal symbol which is eventually found.

Dürr-Høyer MUD for CDMA/SDMA NOMA - Userload=2



NOMA CDMA vs SDMA

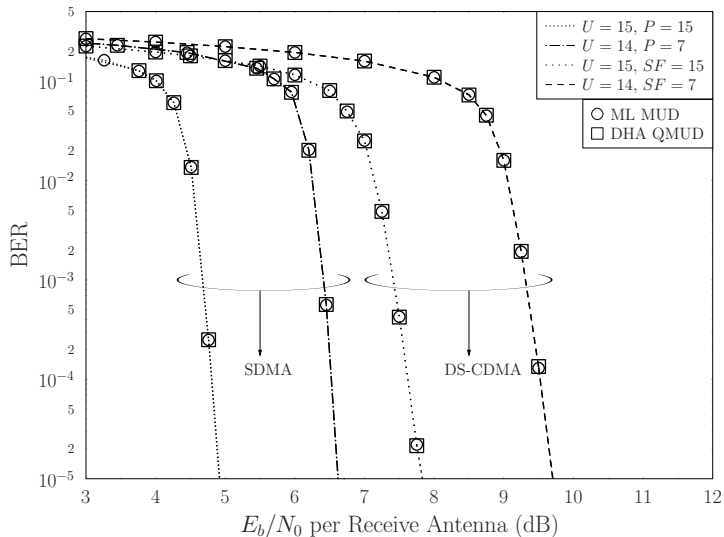
Iterative Joint Channel & Data Estimation Turbo-Receivers for NOMA



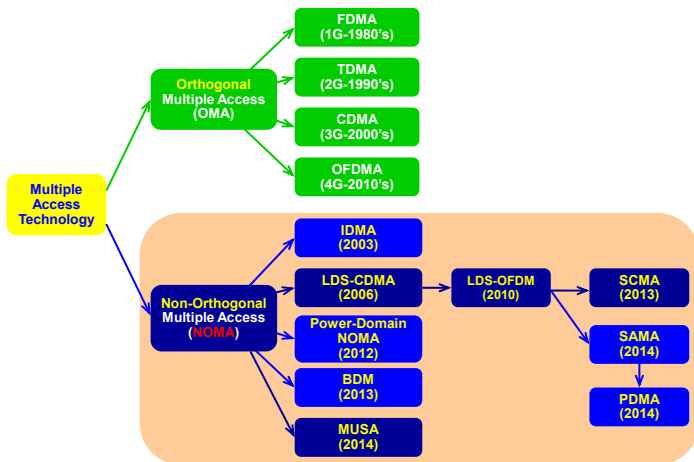
DS-CDMA vs SDMA NOMA Systems

	System 1	System 2	System 3	System 4
Number of Users	$U = 14$	$U = 14$	$U = 15$	$U = 15$
Multiple Access Scheme	DS-CDMA	SDMA	DS-CDMA	SDMA
Number of AEs at the BS	$P = 1$	$P = 7$	$P = 1$	$P = 15$
Spreading Factor	$SF = 7$	N/A	$SF = 15$	N/A
Spreading Codes	m-sequences	N/A	Gold Codes	N/A
Normalized User Load	$U_L = 2$	$U_L = 2$	$U_L = 1$	$U_L = 1$
Bit-based Interleaver Length	42 000	42 000	40 000	40 000
Number of AEs per User	$N_{T_x} = 1$			
Modulation	BPSK $M = 2$			
Channel Code	Turbo Code, $R = 1/2$, 8 Trellis states			
Channel	$I_{inner} = 4$ iterations			
Channel Estimation	Uncorrelated Rayleigh Channel			
	Perfect			

Dürr-Høyer CDMA/SDMA NOMA AT Userload=2

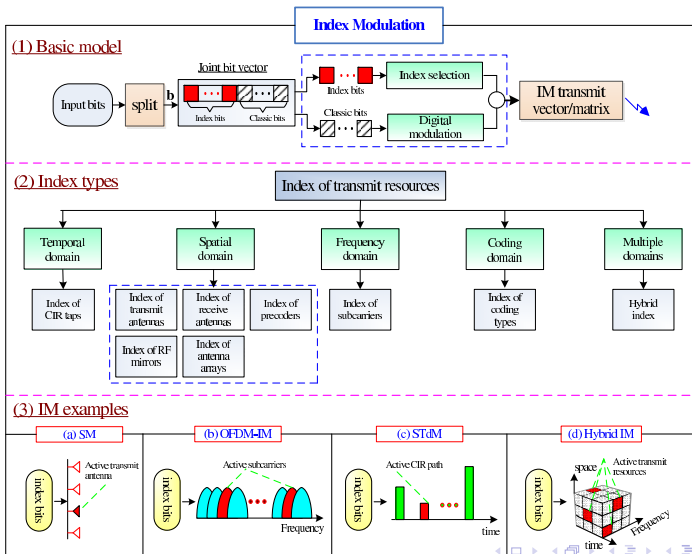


Milestones of multiple access technologies

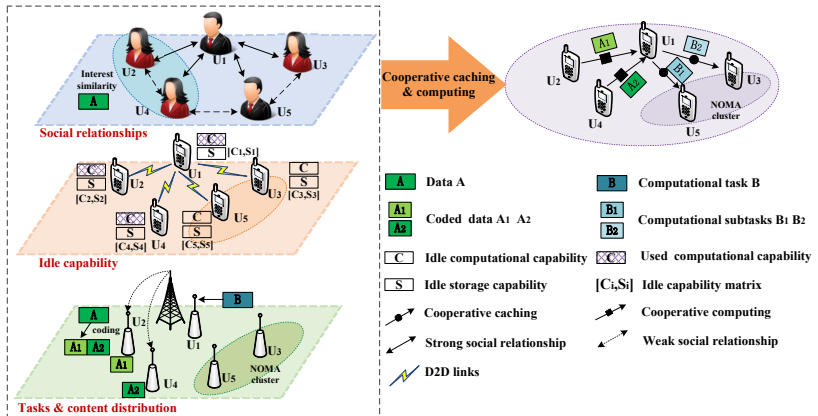


Linglong Dai, Bichai Wang, Zhiguo Ding, Zhaocheng Wang, Sheng Chen and Lajos Hanzo: A Survey of Non-Orthogonal Multiple Access for 5G,

3D Index modulation



Social Networking and Caching Aided Collaborative Computing for the Internet of Things by Ai, Wang, Han, Zhang, Hanzo, IEEE Comms. Mag. 2018

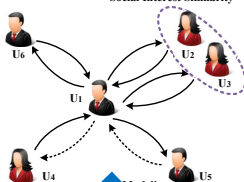


Social relationships

Social Trust Matrix

	U ₁	U ₂	U ₃	U ₄	U ₅	U ₆
U ₁	1.0	0.9	0.8	0.8	0.9	0.3
U ₂	0.8	1.0	0.2	0.3	0.2	0.3
U ₃	0.9	0.2	1.0	0.1	0.0	0.1
U ₄	0.9	0.1	0.2	1.0	0.3	0.2
U ₅	0.9	0.3	0.0	0.4	1.0	0.3
U ₆	0.2	0.2	0.2	0.1	0.2	1.0

Social Interest Similarity



Modeling



Face-to-face social relationship

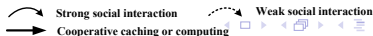
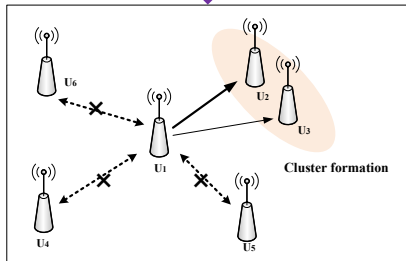


Online social relationship

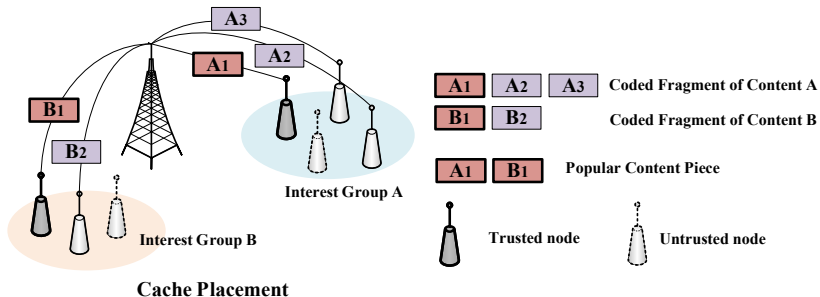
Idle Capabilities Matrix

	U ₁	U ₂	U ₃	U ₄	U ₅	U ₆
Idle computational capability	0	1	0	1	0	0
Idle storage capability	0	1	1	1	1	1
..... capability						

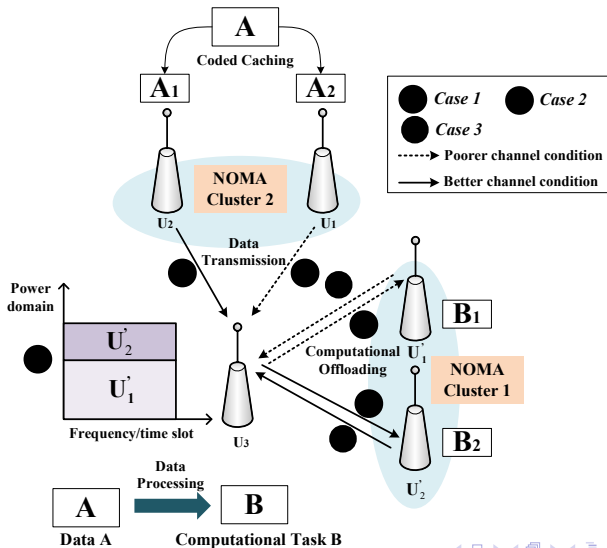
Cooperative Caching and Computing Scheme



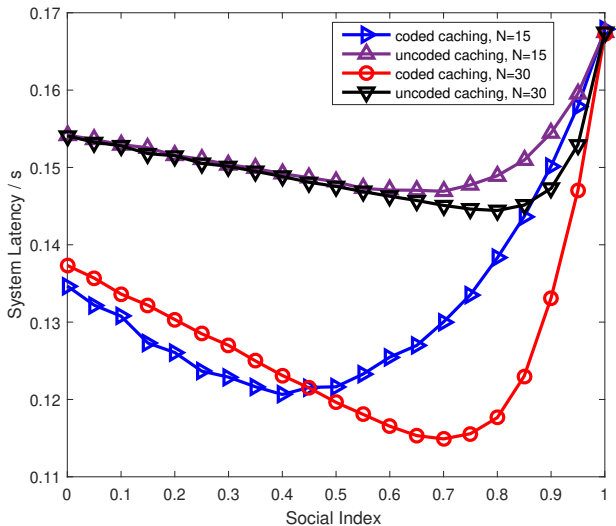
Cache placement based on social relationships



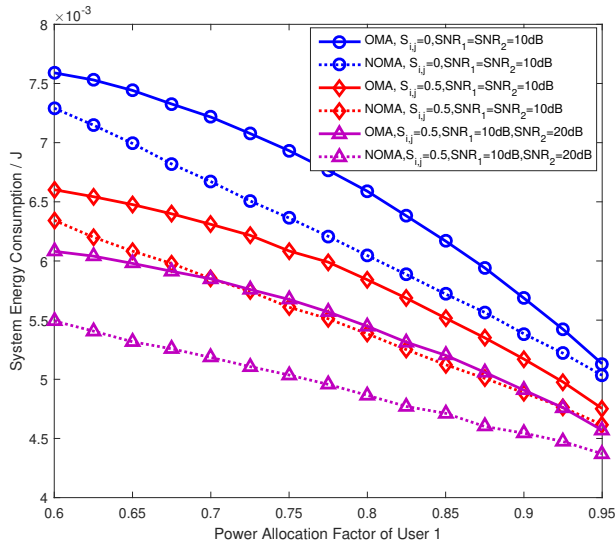
Exploiting coded caching and NOMA in our proposed framework



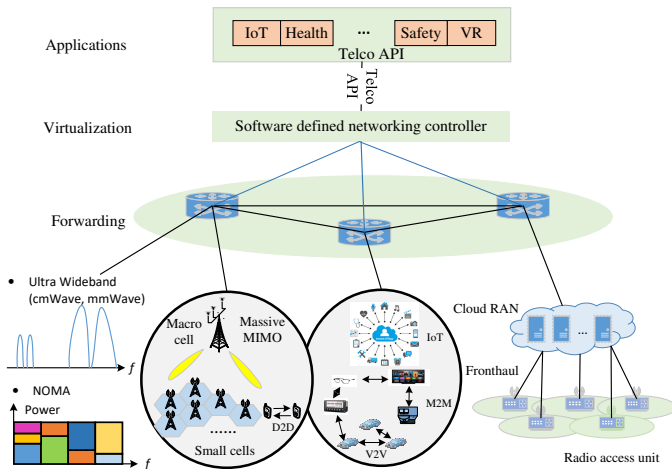
System latency of the proposed framework relying on coded caching versus uncoded caching in a multi-user scenario, where the total number of users is $N = 15$ and 30.



Energy consumption of the proposed framework relying on NOMA exploited and OMA in a multi-user scenario, where the total number of users is $N = 15$.



Future 5G network architecture.



[1] Y. Liu, Z. Qin, M. El-kashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-Orthogonal Multiple Access for 5G", *Proceedings of the IEEE*; Dec 2017.

NOMA - A Paradigm Shift for Multiple Access for 5G and Beyond

Zhiguo Ding

School of Electrical and Electronic Engineering, University of Manchester
A collaboration with Princeton, SWJTU, and FAU

A Tutorial at PIMRC-2020

31 August 2020

Outline

Semi-Grant-Free Transmission

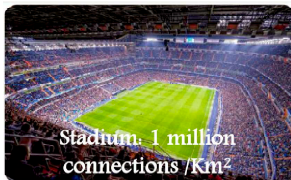
Rethink SIC Designs

Motivations for Semi-Grant-Free Transmission (1/3)

- Why is NOMA important to 5G and beyond:
 - eMBB: Improve the system throughput for downlink and uplink without consuming extra bandwidth
 - URLLC: Users do not have to wait for a long time to be served even if there are not sufficient orthogonal resource blocks
 - mMTC: Encouraging users to share their bandwidth resources instead of solely occupying them
- Extensive studies have been done for the application of NOMA to eMBB, but not to URLLC and mMTC
- Let's focus on grant-free transmission
 - An important feature to be supported in URLLC and mMTC

Z. Ding, R. Schober, P. Fan, and H. V. Poor, "Simple Semi-Grant-Free Transmission Strategies Assisted by Non-Orthogonal Multiple Access", IEEE Trans. Wireless Commun., 2019.

Motivations for Semi-Grant-Free Transmission (2/3)

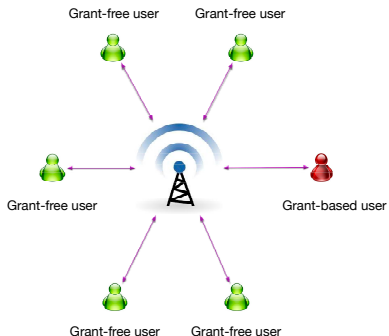


- Use mMTC as an example.
- Characteristics: small data packets, low power consumption, and low cost.
- Requirements: 1 million devices/km and 100 billion connections in total.
- The use of grant-based protocols means that lengthy handshaking is needed for each user
 - Result in large latency and poor spectral efficiency.

Motivations for Semi-Grant-Free Transmission (3/3)

- Importance of grant-free transmission
 - A user is encouraged to transmit whenever it has data to send, without getting the grant from its base station.
 - The lengthy handshaking process is avoided and signalling overhead can be reduced
- Challenges to realize grant-free transmission
 - Existing solutions based on massive MIMO or NOMA are special cases of random access, where the base station is not involved in multiple access (scheduling).
 - Therefore, the number of users admitted to the same channel cannot be not controlled (excessive users → failure of MUD).
 - Blind user activity identification and blind channel estimation are needed.
 - Limited channels are reserved for grant-free transmission.

Semi-Grant-Free Transmission (1/3)



- A semi-grant-free scenario is considered.
- Example of a grant-based user: a sensor for healthcare monitoring or critical care.
- Example of a grant-free user: a sensor for power meters or environmental monitoring
- An important scenario to consider
 - All the channels available in the network, even if they have been reserved by grant-based users, can be used for grant-free transmission
 - Massive connectivity can be supported spectrally efficiently.

Semi-Grant-Free Transmission (2/3)

Two semi-grant-free transmission strategies can be developed:

- Open-loop contention control
 - The base station broadcasts a threshold to the users and sets the criterion for the users to be qualified for transmission
 - A strategy similar to random beamforming
 - No need to know the users' channels at the base station
 - Compared to grant-based transmission, less signalling overhead is consumed by SGF
 - All the users which satisfy the criterion are granted to transmit without going through the individual handshaking process.
 - Compared to grant-free transmission, the number of the users admitted to the same channel in SGF can be carefully controlled
 - Avoid the failure of MUD due to excessive users.

Semi-Grant-Free Transmission (3/3)

- Distributed contention control
 - What is distributed contention control?
 - Once the contention time window starts, each user chooses a backoff value τ_m
 - τ_m is a strictly decreasing (or increasing) function of the user's channel gain
 - A user transmits a beacon to the base station after τ_m expires, under the condition that τ_m is smaller than the time window.
 - A fixed number of users with the desirable channel conditions are granted to transmit.
 - The number of the users admitted by open-loop control is still random
 - Avoid the failure of MUD due to excessive users
 - The performance of the grant-based user can be strictly guaranteed, at a price of more signal overhead, compared to open-loop control.

Outline

Semi-Grant-Free Transmission

Rethink SIC Designs

Channel State Information (CSI)-Based SIC (1/2)

- An uplink communication scenario with $(M + 1)$ users (U_m)
- Two-user power-domain NOMA is the most famous example
 - $U_n, 1 \leq n \leq M$, is paired with U_0
 - Two users transmit simultaneously
 - Assume $|h_n|^2 \geq |h_0|^2$, where h_m denotes U_m 's channel.
 - The base station decodes U_n 's signal first, remove it and then decode U_m 's signal.
 - Data rates of NOMA are

$$R_n^N = \log \left(1 + \frac{P|h_n|^2}{1 + P|h_0|^2} \right), \quad (1)$$

and

$$R_0^N = \log (1 + P|h_0|^2), \quad (2)$$

- Data rates of OMA are $R_i^O = \frac{1}{2} \log (1 + P|h_i|^2), i \in \{0, n\}$

Channel State Information (CSI)-Based SIC (2/2)

- Advantages

- Consider the extreme case $h_0 \rightarrow 0$.
- The sum rate of NOMA can be approximated as follows:

$$R_{sum}^N = R_0^N + R_n^N \xrightarrow{h_0 \rightarrow 0} \log(1 + P|h_n|^2) \xrightarrow{P \rightarrow \infty} \log P, \quad (3)$$

which is almost two times the sum rate of OMA,
 $R_{sum}^O \triangleq R_0^O + R_n^O \rightarrow \frac{1}{2} \log P$. ($P = \frac{1}{2}P^O$ is assumed.)

- Disadvantages

- A user whose signal is decoded first suffers from severe interference $\Rightarrow$ QoS deteriorates $\Rightarrow$ difficult to be generalized
- Users' channels need to be sufficiently different in order to yield a reasonable performance gain over OMA

Quality of Service (QoS)-Based SIC (1/3)

- Cognitive radio inspired (CR) NOMA is a good example
 - U_0 is a delay-sensitive user with a low target data rate (R_0)
 - In OMA, U_0 is allowed to occupy a dedicated resource block
 - In CR-NOMA, $U_0 \rightarrow$ Primary user, $U_n \rightarrow$ Secondary user,
 - A secondary user, U_n , $1 \leq n \leq M$, is scheduled if

$$\log \left(1 + \frac{P|h_0|^2}{1 + P|h_n|^2} \right) \geq R_0. \quad (4)$$

- If (??) is feasible, the first stage of SIC is guaranteed to be successful, and U_n 's achievable data rate is given by

$$R_n^{CR} = \log (1 + P|h_n|^2) . \quad (5)$$

- Rationale behind the used SIC order
 - The user whose signal is decoded first suffers interference $\Rightarrow$ the user's achievable data rate will be small. This is OK.
 - U_n has no interference. U_n 's data rate constitutes the performance gain of CR-NOMA over OMA $\Rightarrow$ unbounded.

Quality of Service (QoS)-Based SIC (2/3)

- Example 1:
 - Assume that there are M delay-sensitive users to be served at low data rates, and one delay-tolarent user.
 - In OMA, $(M + 1)$ time slots are needed to serve these users.
 - With NOMA, all users are served in a single time slot $\Rightarrow$ spectral efficiency improved $(M + 1)$ times
- Example 2:
 - $h_0 = h_n = h$, where CSI-based SIC fails
 - With QoS-based SIC, the sum rate gain of NOMA over OMA is

$$\Delta_{sum} = \mathbf{1}_0 \log(1 + P|h|^2), \quad (6)$$

where $\mathbf{1}_0$ is an indicator function, i.e., $\mathbf{1}_0 = 1$ if

$\log\left(1 + \frac{P|h|^2}{1+P|h|^2}\right) \geq R_0$, otherwise $\mathbf{1}_0 = 0$.

- In Rayleigh fading, we have $P(\mathbf{1}_0 = 1) \rightarrow 1$, for $P \rightarrow \infty$.
Therefore, $\Delta_{sum} \rightarrow \infty$, for $P \rightarrow \infty$

Quality of Service (QoS)-Based SIC (3/3)

The implications of the channel conditions

- It is possible that U_0 's channel conditions are weaker than U_n 's, i.e., $|h_0|^2 < |h_n|^2$.
- This leads to the common question whether this situation results in a decoding failure.
 - Correct decoding depends on whether the data rate supported by the channel is larger than the target data rate.
 - As long as this condition holds, the use of error correction coding can ensure correct decoding, even if the signal strength is weaker than the interference strength.
 - Error correction coding injects redundant information, which reduces the information data rate.
 - But if U_0 's target data rate is small, a significant amount of redundant information can be added.
 - For example, for the case of $R_0 = 0.1$ bits/s/Hz, a repetition code with a code rate of $\frac{1}{10}$ is affordable, where one bit is repeated 10 times.

Dilemma Faced By These Strategies (1/2)

- Same system setup
 - $U_0 \Rightarrow$ Primary user
 - One of the M secondary users is admitted to U_0 's channel.
 - Assume the secondary users are ordered as $|h_1|^2 \leq \dots \leq |h_M|^2$
- CSI-based SIC
 - Schedule the strongest user (U_M) to pair with U_0
 - Decode U_M 's signal $\Rightarrow$ Decode U_0 's signal
 - U_M needs to use the following data rate for its transmission

$$R_M^{CSI} = \log \left(1 + \frac{P|h_M|^2}{1 + P|h_0|^2} \right). \quad (7)$$

- Pros: multi-user diversity exploited; U_0 's QoS guaranteed
- Cons: exhibits error floors in outage probability

$$P \left(\log \left(1 + \frac{P|h_M|^2}{1 + P|h_0|^2} \right) < R_s \right) \xrightarrow{P \rightarrow \infty} P \left(\log \left(1 + \frac{|h_M|^2}{|h_0|^2} \right) < R_s \right)$$

Dilemma Faced By These Strategies (2/2)

- QoS-based SIC

- Decode U_0 ' signal $\Rightarrow$ Decode the secondary user's signal
- Schedule a weak user (U_1) yielding the following data rate:

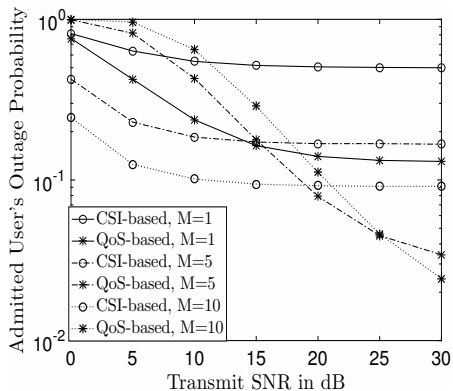
$$R_1^{QoS} = \log(1 + |h_1|^2 P), \quad (8)$$

if $\log\left(1 + \frac{P|h_0|^2}{1 + P|h_1|^2}\right) > R_0$, otherwise $R_1^{QoS} = 0$.

- Pros: R_1^{QoS} is interference free;
- Cons: loss of multi-user diversity; exhibits error floors in outage probability

$$\begin{aligned} & P\left(\log\left(1 + \frac{P|h_0|^2}{1 + P|h_1|^2}\right) < R_0\right) \\ & + P\left(\log\left(1 + \frac{P|h_0|^2}{1 + P|h_1|^2}\right) > R_0, \log(1 + P|h_1|^2) < R_s\right). \end{aligned} \quad (9)$$

Dilemma Faced By These Strategies - Simulation



- QoS-based SIC performs better in the high SNR region.
- Increasing M helps to improve the performance of CSI-based SIC.
- It seems that removing error floors is a mission impossible.

Outage performance achieved by NOMA transmission with the two types of SIC. Independent and identically distributed (i.i.d.) Rayleigh fading is assumed for the users' channel gains. $R_0 = 0.2$ bits/s/Hz, and $R_s = 1$ bits/s/Hz.

Hybrid SIC (1/2)

- Same system setup
 - $U_0 \Rightarrow$ Primary user
 - One of the M secondary users is admitted to U_0 's channel.
- First set a threshold for the secondary users' channels

$$\tau = \max \left\{ 0, \frac{|h_0|^2}{2^{R_0} - 1} - \frac{1}{P} \right\}. \quad (10)$$

- By using the threshold, the M secondary users can be divided into two groups:
 - $\mathcal{S}_1$ contains the strong users, i.e., $|h_n|^2 > \tau$ and can support CSI-based SIC only $\Rightarrow R_n^1 = \log \left(1 + \frac{P|h_n|^2}{1+P|h_0|^2} \right)$, for $n \in \mathcal{S}_1$.
 - $\mathcal{S}_2$ contains the weak users, i.e., $|h_n|^2 < \tau$, and can support either of the two SIC decoding orders $\Rightarrow \max \left\{ \log \left(1 + \frac{P|h_n|^2}{1+P|h_0|^2} \right), \log(1 + P|h_n|^2) \right\}$, for $n \in \mathcal{S}_2$.

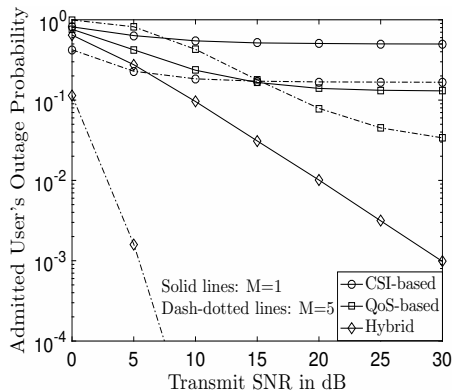
Hybrid SIC (2/2)

- The considered SIC scheme can be viewed as a hybrid version of CSI- and QoS-based SIC, since both decoding orders can be used.
- Error floors can be avoided by using this hybrid SIC
- We used to believe that swapping SIC decoding orders is trivial and does not yield a significant performance gain.
- Dynamically switching the SIC decoding order can achieve a surprising performance improvement that cannot be realized by the two conventional schemes.
- Distributed contention control can be applied to ensure that user scheduling is accomplished in a distributed manner.

Z. Ding, R. Schober, and H. V. Poor, "A New QoS-Guarantee Strategy for NOMA Assisted Semi-Grant-Free Transmission", IEEE TWC (submitted).

Z. Ding, R. Schober, and H. V. Poor, "Unveiling the Importance of SIC in NOMA Systems: Part I & II ", IEEE Communications Letters (invited paper)..

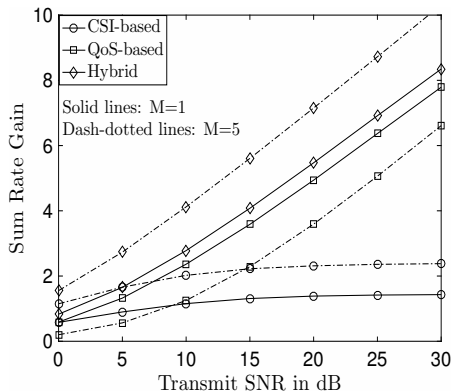
Hybrid SIC - Simulation (1/2)



The performance achieved by NOMA transmission with the three types of SIC. i.i.d. Rayleigh fading is assumed for the users' channel gains. $R_0 = 0.2$ bits/s/Hz, and $R_s = 1$ bits/s/Hz.

- Outage probabilities are used as criteria.
- Both CSI and QoS-based SIC suffer error floors.
- A mission impossible is accomplished by hybrid SIC.

Hybrid SIC - Simulation (2/2)



The performance achieved by NOMA transmission with the three types of SIC. i.i.d. Rayleigh fading is assumed for the users' channel gains. $R_0 = 0.2$ bits/s/Hz, and $R_s = 1$ bits/s/Hz.

- Ergodic rates are used as criteria.
- Hybrid SIC realizes the best performance.
- Increasing M helps to improve the performance of CSI-based SIC.
- QoS-based SIC outperforms CSI-based SIC.

Outline

Semi-Grant-Free Transmission

Rethink SIC Designs

Applications of Hybrid SIC (1/2)

The following findings are important

- Using more than one SIC orders is not trivial.
- A simple strategy to switch between different SIC designs yields a surprising performance improvement that cannot be realized by the two conventional designs.

In the second part of our invited paper at CL, we used NOMA assisted mobile edge computing (NOMA-MEC) as an example

- Conventional NOMA-MEC decides SIC according to the users' offloading deadlines
- With such a fixed SIC design, we previously learned that the gain of NOMA over OMA is related to the deadlines
- Hybrid SIC has been shown to be applicable to NOMA-MEC
- Better gain is realized, and the gain is related to Tx powers.
- So these findings are useful not only for performance analysis but also

Applications of Hybrid SIC (2/2)

- Fundamentals of hybrid SIC
- Green communications
- User clustering and resource allocation
- Multiple-input multiple-output (MIMO) and intelligent reflecting surface (IRS) assisted NOMA
- Emerging applications of NOMA

Thank you for your attention!

Email: zhiguo.ding@manchester.ac.uk

Intelligent Massive NOMA towards 6G

Tutorials of PIMRC2020, London, UK

Dr. Yuanwei Liu, Prof. Zhiguo Ding and Prof. Lajos Hanzo

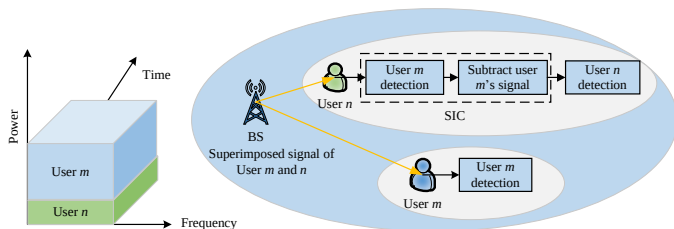
Queen Mary University of London, UK
The University of Manchester, UK
University of Southampton, UK

yuanwei.liu@qmul.ac.uk zhiguo.ding@manchester.ac.uk hanzo@soton.ac.uk

Aug. 31st, 2020

- 1 Power-Domain NOMA Basics
- 2 Signal Processing Advances for NOMA: A Machine Learning Approach
- 3 Emerging Applications for NOMA
 - Emerging Applications for NOMA: Exploiting NOMA in UAV Networks
 - Emerging Applications for NOMA: Interplay Between RIS/IRS and NOMA Networks

Power-Domain NOMA Basics



- 1 Supports multiple access within a given resource block (time/frequency/code), using **different power levels** for distinguishing/separating them [1].
- 2 Apply successive interference cancellation (SIC) at the receiver for separating the NOMA users [2].
- 3 If their power is similar, PIC is a better alternative.

[1] Y. Liu, Z. Qin, M. El Kashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Non-Orthogonal Multiple Access for 5G", *Proceedings of the IEEE*, Dec 2017. ([Web of Science Hot paper](#))

[2] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. El Kashlan, Chih-Lin I, and H. V. Poor (2017), "Application of

Non-orthogonal Multiple Access in LTE and 5G Networks", *IEEE Communication Magazine*; ([Web of Science Hot](#)

Signal Processing Advances for NOMA: A Machine Learning Approach

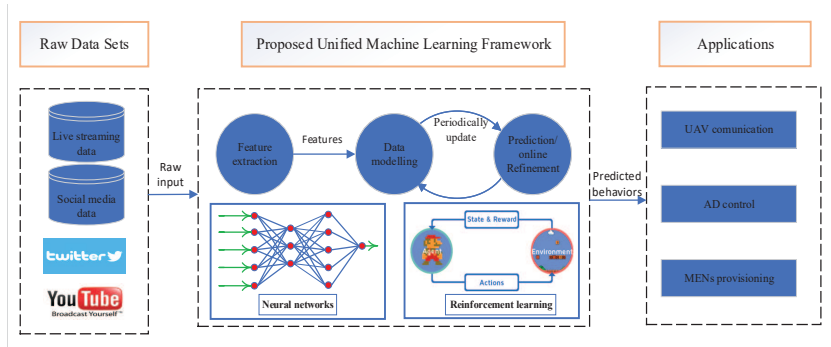


Fig.: Artificial intelligent algorithms for wireless communications.

[1] Y. Liu, S. Bi, Z. Shi, and L. Hanzo, "When Machine Learning Meets Big Data: A Wireless Communication Perspective", *IEEE Vehicular Communication Magazine*, *accept to appear*, <https://arxiv.org/abs/1901.08329>.

System Model and Problem Formulation: Cache-Aided NOMA MEC

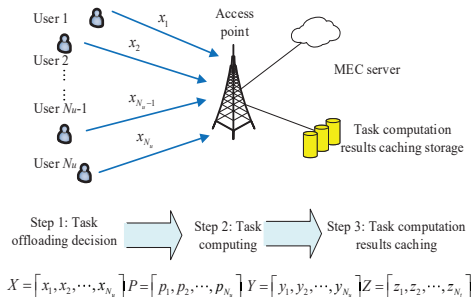


Fig.: An illustration of a multi-user cache-aided MEC.

- Multiple users are served by one MEC server.
- The computation tasks are capable of being computed locally at the mobile devices or in the MEC server.
- The computation results are selectively cached in the storage of the MEC server.

[1] Z. Yang, **Y. Liu**, Y. Chen, N. Al-Dhahir, "Cache-Aided NOMA Mobile Edge Computing: A Reinforcement Learning Approach", *IEEE Transactions on Wireless Communications*, *accept*, <https://arxiv.org/abs/1906.08812>.

System Model: Communication Model

The user with higher channel gain is decoded first, the signal-to-interference-plus-noise ratio (SINR) for user i at time t can be given by

$$\mathbf{R}_i(\mathbf{t}) = B \log_2 \left(1 + \frac{\rho_i(t) |h_i(t)|^2}{\sum_{l=i+1}^{N_{up}} \rho_l(t) |h_l(t)|^2 + \sigma^2} \right), \quad (1)$$

Accordingly, the offloading time for task j with input size π_j at time t is

$$\mathbf{T}_{i,j}^{\text{offload}}(\mathbf{t}) = \frac{\pi_j}{\mathbf{R}_i(\mathbf{t})}. \quad (2)$$

Meanwhile, the transmit energy consumption of offloading at time t is given by

$$\mathbf{E}_{i,j}^{\text{offload}}(\mathbf{t}) = \rho_i \frac{\pi_j}{\mathbf{R}_i(\mathbf{t})}. \quad (3)$$

System Model: Computation Model

- **Local Computing:** The computing time $T_{i,j}^{loc}$ and energy consumption $E_{i,j}^{loc}$ for task j with computational requirement ω_j are

$$T_{i,j}^{loc} = \frac{\omega_j}{\omega_i^{loc}}. \quad (4)$$

$$E_{i,j}^{loc} = P_i^{loc} \frac{\omega_j}{\omega_i^{loc}}. \quad (5)$$

- ω_i^{loc} : the local computing capability,
- P_i^{loc} : the energy consumption per second.

- **MEC Computing:** The computing time $T_{i,j}^{mec}(\mathbf{t})$ and energy consumption $E_{i,j}^{mec}(\mathbf{t})$ are

$$T_{i,j}^{mec}(\mathbf{t}) = \frac{\omega_j}{y_i(\mathbf{t}) C_{MEC}}. \quad (6)$$

$$E_{i,j}^{mec}(\mathbf{t}) = P^{mec} \frac{\omega_j}{y_i(\mathbf{t}) C_{MEC}}. \quad (7)$$

- y_i : the proportion of the computing resources allocated from the MEC server,
- P^{mec} : the energy consumption per second at MEC server.

Problem Formulation

The sum energy consumption is

$$\mathbf{E}_i(t, x_i(t), y_i(t), f_i(t), \rho_i(t)) = \left((1 - y_i(t)) \left(x_i(t) \mathbf{E}_{i,j}^{\text{loc}} + (1 - x_i(t)) \left(\mathbf{E}_{i,j}^{\text{offload}}(\mathbf{t}) + \mathbf{E}_{i,j}^{\text{mec}}(\mathbf{t}) \right) \right) \right). \quad (8)$$

The optimization problem is

$$(\mathbf{P1}) \min_{X, Y, F, P} \sum_{t=1}^T \sum_{i=1}^{N_u} \mathbf{E}_i(t, x_i(t), y_i(t), f_i(t), \rho_i(t)), \quad (9a)$$

$$\text{s.t. } C_1 : x_i(t) \in \{0, 1\}, \forall i \in \mathcal{N}, \quad (9b)$$

$$C_2 : y_i(t) \in \{0, 1\}, \forall i \in \mathcal{N}, \quad (9c)$$

$$C_3 : \sum_{i \in N_u} f_i(t) \leq F, \quad (9d)$$

$$C_4 : \sum_{i \in N_u} \rho_i(t) \leq P_i, \quad (9e)$$

$$C_5 : \mathbf{T}_{i,j}^{\text{offload}}(\mathbf{t}) + \mathbf{T}_{i,j}^{\text{mec}}(\mathbf{t}) \leq T, \forall t \in [1, T]. \quad (9f)$$

Feature Extraction: Neural Networks (NN) for Prior Information Prediction

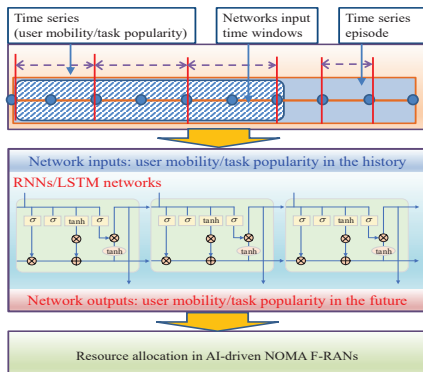


Fig.: Recurrent neural networks for user mobility/content popularity prediction.

- The expert fitting characteristic of NN enables the ability of predicting time series.
- We invoke recurrent neural networks (RNNs)/long-short term memory (LSTM) networks for their skills of processing time-sequential data using internal memory units.

Resource allocation: From the Formulated Problem to Reinforcement Learning Model

A Markov decision process (MDP) model is a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{R} \rangle$.

- 1 **Objective:** maximize the sum reward

$$V_{\pi}(s) = \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_0 = s \right] \text{ of a trajectory}$$

$$s_0 \xrightarrow{a_1|r_1} s_1 \xrightarrow{a_2|r_2} s_2 \cdots \xrightarrow{a_n|r_n} s_n.$$

- 2 **State space ($\mathcal{S}$):**

$$s(t) = [x(t), y(t), f(t), \rho(t)] \in \mathcal{S} = \mathcal{X} \times \mathcal{Y} \times \mathcal{F} \times \rho.$$

- 3 **Action space ($\mathcal{A}$):**

$$a(t) = [\Delta x(t), \Delta y(t), \Delta f(t), \Delta \rho(t)] \in \mathcal{A}.$$

- 4 **Reward function (r):** the sum energy consumption of taking

$$\text{an action on a state } r_t = - \sum_{i=1}^{N_u} \mathbf{E}_i(t, a(t) | s(t)).$$

How to define State and Action Space? From Maze to the Proposed Framework

- **State Space:** High dimensional matrix related to the parameters in the objective function.
- **Action Space:** Moving granularity in each element of the state space.

	Maze game	UAV trajectory	Cache-Aided NOMA MEC
State Space	$s(t) = [x(t), y(t)]$ $\in \mathcal{S} = \mathcal{X} \times \mathcal{Y}$	$s(t) = [x(t), y(t), z(t)]$ $\in \mathcal{S} = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$	$s(t) = [x(t), y(t), f(t), \rho(t)]$ $\in \mathcal{S} = \mathcal{X} \times \mathcal{Y} \times \mathcal{F} \times \mathcal{G}$
Action Space	$a(t) = [\Delta x(t), \Delta y(t)]$	$a(t) = [\Delta x(t), \Delta y(t), \Delta z(t)]$	$a(t) = [\Delta x(t), \Delta y(t), \Delta f(t), \Delta \rho(t)]$
	2 Dimensional	3 Dimensional	4 Dimensional

Fig.: Setting of state and action space.

Reinforcement Learning Model

The goal of reinforcement learning is to find an optimal policy that maximize the long-term sum rewards:

$$\pi^* = \arg \max_{\pi} \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t r_t \mid \pi \right]. \quad (10)$$

- Policy π : a function from state to action that specifies what action to take in each state.

The Q-value function is adopted to measure the performance of the policy.

$$Q^*(s, a) = \max_{\pi} \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t r_t \mid s = s_0, a = a_0, \pi \right]. \quad (11)$$

The optimal Q-value function satisfies the Bellman Equation

$$Q^*(s, a) = \mathbb{E}_{s' \sim \mathcal{E}} \left[r + \gamma \max_{a'} Q^*(s', a') \mid s, a \right]. \quad (12)$$

How does the Intelligent Agent Learn?

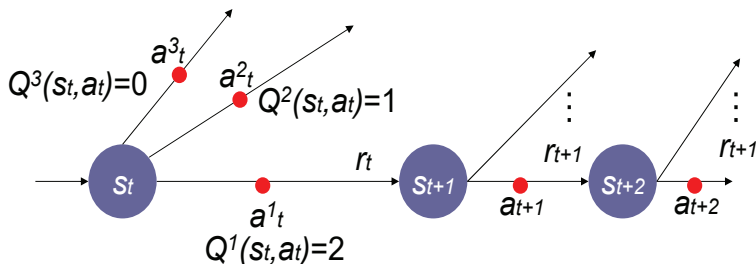
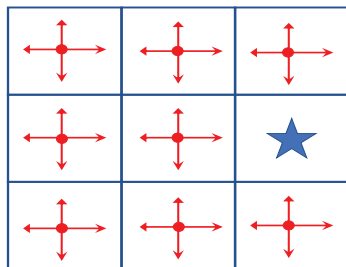


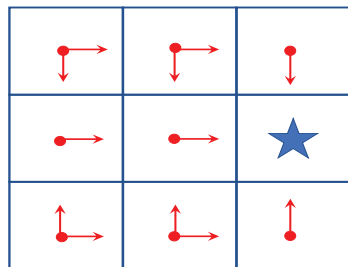
Fig.: Q-learning flow.

- The agent takes action a^1_t , because the corresponding Q value $Q^1(s_t, a_t)$ is max.

The Learning Results: A Maze Case Example



Random policy before learning



Optimal policy after training

Fig.: Q-learning expected result (star represents the treasure).

- After learning, we obtain the optimal action for each state.

Resource Allocation: The Proposed Reinforcement Learning for Cache-Aided NOMA MEC

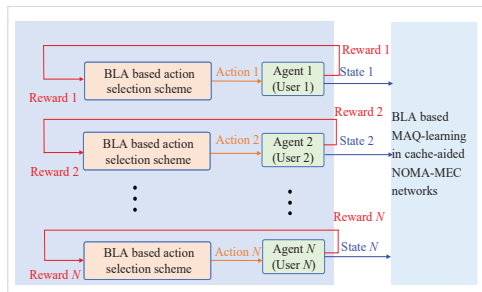


Fig.: Bayesian learning automata based multi-agent Q-learning for resource allocation.

- Each mobile user is set as a intelligent agent.
- BLA is capable of obtaining optimal action for two action case.
- The multiple intelligent agents operate in a selfish manner.

Numerical Results: Feature Extraction (RNNs for User Mobility Prediction)

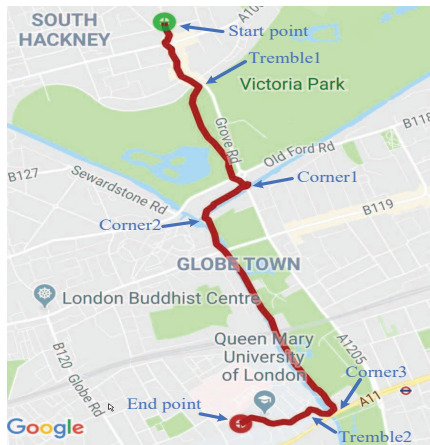


Fig.: User trajectory in Google Maps.

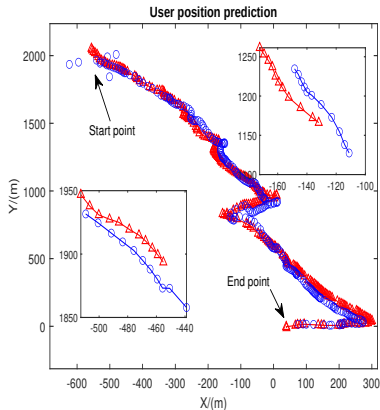


Fig.: Comparison between real user trajectory and predicted trajectory.

Numerical Results: Resource Allocation (the proposed Reinforcement Learning Algorithm)

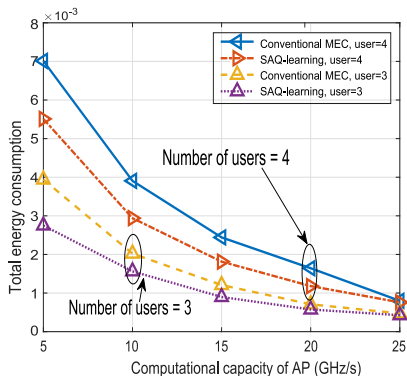


Fig.: Total energy consumption vs. the computation capacity of the AP.

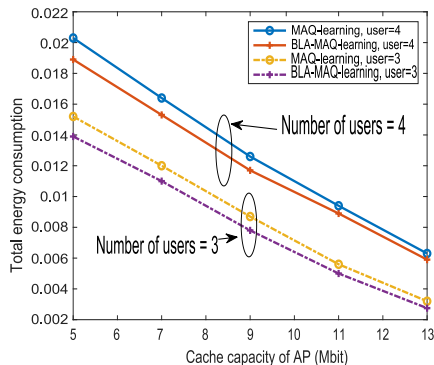
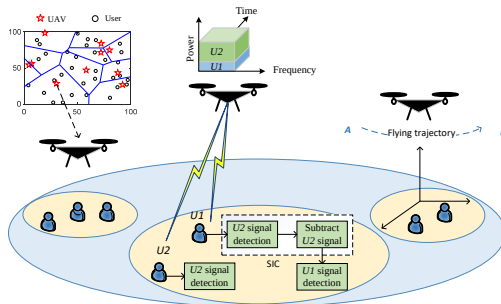


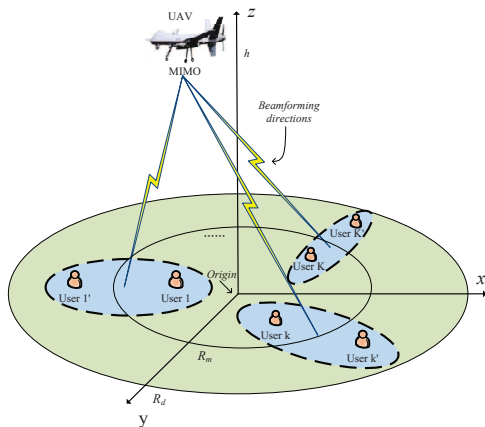
Fig.: Total transmit energy consumption vs. cache capacity of the AP..

Emerging Applications for NOMA: Exploiting NOMA in UAV Networks



[1] Y. Liu, Z. Qin, Y. Cai, Y. Gao, G. Ye Li, and A. Nallanathan, "UAV Communications Based on Non-Orthogonal Multiple Access", *IEEE Wireless Communications*, vol. 26, no. 1, pp. 52-57, Feb. 2019.

Single UAV: MIMO-NOMA UAV Networks



- 1) There are probabilistic line-of-sight links.
- 2) The small-scale fading follows Nakagami fading or Rice fading.
- 3) The height of UAV can be a random variable or any arbitrary value.

[1] T. Hou, Y. Liu, Z. Song, X. Sun, Y. Chen, "Multiple Antenna Aided NOMA in UAV Networks: A Stochastic Geometry Approach", *IEEE Transactions on Communications*, vol. 67, no. 2, pp. 1031-1044, Feb. 2019.

From Single UAV to Multiple UAVs: NOMA enabled UAV Communications

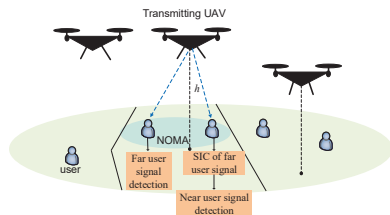


Fig.: An illustration of NOMA UAV in Cellular Networks.

- Massive UAV-BSs are located in the sky.
- Users are located on the ground
- NOMA technique is deployed.
- User-association is complicated.

[1] T. Hou, Y. Liu, Z. Song, X. Sun, Y. Chen, "Exploiting NOMA for Multi-UAV Communications in Large-Scale Networks", *IEEE Transactions on Communications, accept to appear*.

NOMA enabled UAV Communications—User-centric Scenario

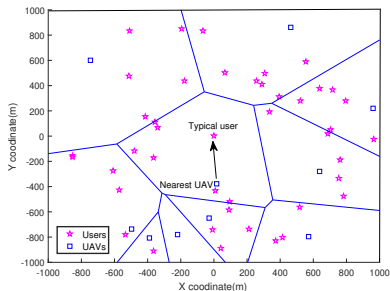


Fig.: The proposed user-centric Scenario, which is a potential solution for emergency communications.

- Ground users and UAVs are distributed according to HPPP.
- All the ground users must be served.
- Association is decided by users according to distance.

NOMA enabled UAV Communications—UAV-centric Scenario

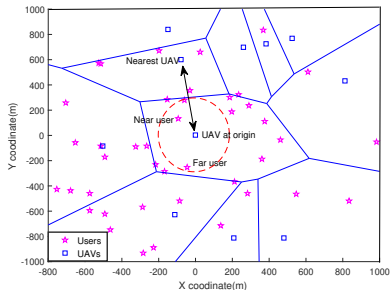


Fig.: The proposed UAV-centric Scenario, which is a potential solution for offloading communications.

- Ground users and UAVs are distributed according to HPPP.
- UAV only provides access services to users located in hot spot areas.
- This is supplementary communications.

NOMA UAV-to-Everything (U2X) Networks

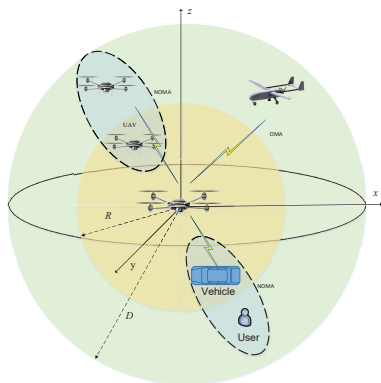
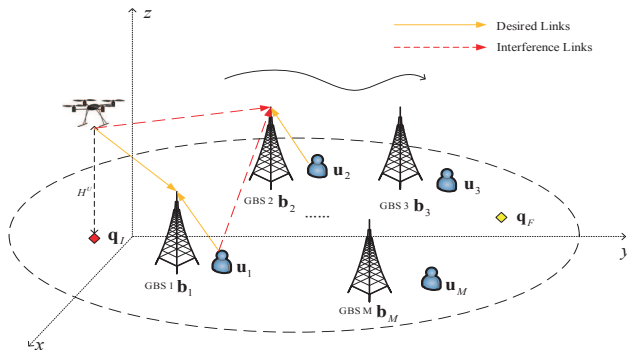


Fig.: The illustration of NOMA enhanced UAV-to-Everything networks.

- Users or receivers are located on the ground or in the sky.
- The coverage space is a sphere.
- NOMA is deployed.

[1] T. Hou, **Y. Liu**, Z. Song, X. Sun, Y. Chen, "Non-Orthogonal Multiple Access in UAV-to-Everything (U2X) Networks", *IEEE Internet of Things*, *accept to appear*, <https://arxiv.org/abs/1907.05571>.

Air-to-Ground NOMA: Trajectory Design and Resource Allocation



- A rotary-wing UAV has a mission of travelling from an predefined initial location $\mathbf{q}_I$ to a final location $\mathbf{q}_F$, while uploading specific information bits to M GBSs.

[1] X. Mu, Y. Liu, L. Guo, and J. Lin, "Non-Orthogonal Multiple Access for Air-to-Ground Communication", *IEEE Transactions on Communications*, accept, <https://arxiv.org/abs/1906.06523>.

System Model

- The instantaneous coordinate of the UAV is denoted as $(x(t), y(t), H^U)$, $0 \leq t \leq T$. The coordinate of GBS m is fixed at (x_m^G, y_m^G, H^G) and its served GUE is located at $(x_m, y_m, 0)$. We denote $\mathbf{q}(t)$, $\mathbf{b}_m$ and $\mathbf{u}_m$ as the horizontal coordinates of the above locations
- The line-of-sight (LOS) channel power gain between the UAV and m th GBS at time instant t is $|h_m^{UAV}(t)|^2 = \frac{\rho_0}{(H^U - H^G)^2 + \|\mathbf{q}(t) - \mathbf{b}_m\|^2}$
- The rayleigh fading channel power gain between GUE j and GBS m is $|h_{j,m}^{UE}|^2 = g_{j,m}^{UE} 10^{-\frac{L_{j,m}^{UE}}{10}}$
- A binary variable $a_m(t) \in \{0, 1\}$ is defined to represent the UAV-GBS association state at time instant t .

Optimization Problem

The considered UAV mission complete time minimization problem:

$$(P1) : \min_{\mathbf{Q}, \mathbf{A}, T} T \quad (13a)$$

$$\text{s.t. } \mathbf{q}(0) = \mathbf{q}_I, \quad (13b)$$

$$\mathbf{q}(T) = \mathbf{q}_F, \quad (13c)$$

$$\|\dot{\mathbf{q}}(t)\| \leq V_{\max}, 0 \leq t \leq T, \quad (13d)$$

$$U_m \geq \tilde{U}_m, m \in \mathcal{M}_{BS}, \quad (13e)$$

$$a_m(t) \|\mathbf{q}(t) - \mathbf{b}_m\|^2 \leq D_m^{NOMA}, \forall m \in \mathcal{M}_{BS}, 0 \leq t \leq T, \quad (13f)$$

$$\|\mathbf{q}(t) - \mathbf{b}_m\|^2 \geq (1 - a_m(t)) D_m^{QoS}, \forall m \in \mathcal{M}_{BS}, 0 \leq t \leq T, \quad (13g)$$

$$\sum_{m=1}^M a_m(t) = 1, 0 \leq t \leq T, \quad (13h)$$

$$a_m(t) \in \{0, 1\}, \forall m \in \mathcal{M}_{BS}. \quad (13i)$$

Optimization Problem

- Constraints (13b)-(13d) are the UAV mobility constraints.
- Constraint (13e) are the required UAV uploading information bits of each GBSs.
- Constraints (13f) and (13g) represent the UAV is required stay in the specific feasible regions when it is associated with different GBSs.
- Constraints (13h) means the UAV need to maintain connectivity during T and associate with at most one GBS at each time instant.

There are two main reasons that make Problem (P1) is challenging to solve.

- First, (P1) is a mixed integer non-convex problem due to the non-convex constraints (13e) and integer constraints (13i). Constraints (13f) and (13g) further make $\mathbf{Q}$ and $\mathbf{A}$ coupled together.
- Second, the UAV trajectory $\mathbf{Q}$ and the UAV-GBS association vectors $\mathbf{A}$ are continuous functions of t , which make (P1) involve infinite number of optimization variables.

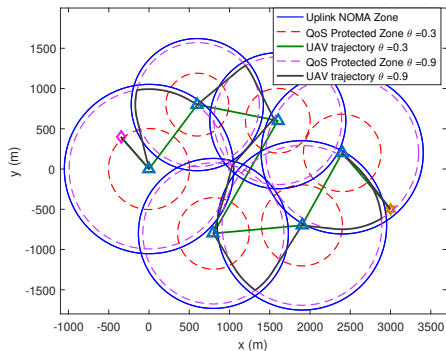
Proposed Solutions: Fly-Hover-Fly Scheme

Theorem 1: Without lose of optimality to (P1), the optimal UAV trajectory can be assumed to be following fly-hover-fly structure: Except hovering at specific locations, the UAV travels at maximum speed $V_{\max}$. Based on **Theorem 1**, the total mission completion time of (P1) can be expressed as

$$\begin{aligned} T(D_{\text{fly}}) &= T_{\text{fly}} + T_{\text{hover}} \\ &= \sum_{m=1}^M \left(\frac{D_{\text{fly},m}}{V_{\max}} + \frac{\tilde{U}_m - U_{\text{fly},m}}{R_{\text{hover},m}} \right) \\ &= \frac{D_{\text{fly}}}{V_{\max}} + \sum_{m=1}^M \frac{\tilde{U}_m - U_{\text{fly},m}}{R_{\text{hover},m}}. \end{aligned} \quad (14)$$

where $D_{\text{fly},m}$ is the total travelling distance when UAV is associated with GBS m , $U_{\text{tr},m}$ is the UAV uploaded information bits to GBS m during travelling through $D_{\text{fly},m}$ and $R_{\text{hover},m}$ is the communication rate when UAV is associated with GBS m and hovers at the corresponding optimal location.

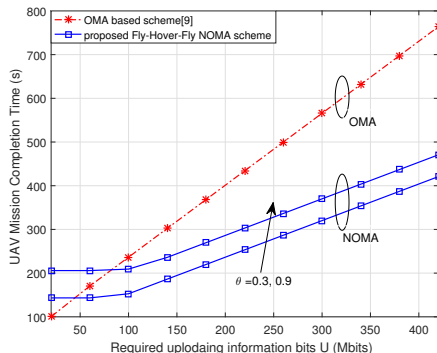
Numerical Results



“ Δ ” are the locations of GBSs.
“ $\diamond$ ” is the UAV initial location.
“ $\star$ ” is the UAV final location.

- Higher QoS requirements contribute larger QoS protected zones.
- When $\theta = 0.3$ bit/s/Hz, the designed UAV trajectory (green line) is only composed of several line segments. It is due to the fact that smaller $\{D_m^{QoS}\}$ impose little constraints on UAV trajectory design..
- When $\theta = 0.9$ bit/s/Hz, the UAV trajectory (black line) is designed to exactly avoid the GUE QoS protected regions to have a shortest travelling distance.

Numerical Results



- The proposed NOMA scheme significantly outperforms OMA scheme when U increase due to the spectrum sharing, which implies the proposed scheme is suitable for rate demanding UAV communication.
- When θ increases, the UAV mission completion time increases for same U . This is due to the increase of θ impose more constraints on the UAV trajectory design and enlarge the minimum UAV travelling distance.

MIMO-Reflecting Intelligent Surfaces (RISs) Networks

■ Motivation

- This is **Next Generation Relay Networks**.
- **Meet the diversified demands** of services and applications of smart communications, e.g., receivers on the died-zones or in the sky by controllable reflections.
- **Provide stronger received power** for mobile devices.
- **Provide interference cancellation services** for mobile devices.

■ Challenges

- **Multiple antennas** are equipped on both BS and Users.
- **How multiple RISs** reflect received signals?
- We have to design **Active beamforming (also called precoding matrix)** at the BS, **passive beamforming at RISs**, **detection vectors at users**.

[1] T. Hou, Y. Liu, Z. Song, X. Sun, Y. Chen and L. Hanzo, "Reconfigurable Intelligent Surface Aided NOMA Networks", *IEEE Journal on Selected Areas in Communications; accept to appear*, <https://arxiv.org/abs/1910.13636>.

Interplay Between IRS and NOMA Networks

Motivations

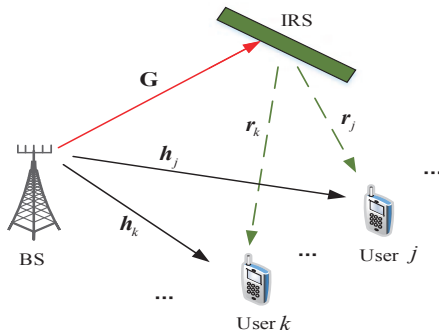
- IRS-aided communication: low cost, programmable wireless environment, spectrum and energy efficiency enhancement.
- NOMA transmission: High spectrum efficiency, fairness and massive connectivity.
- explore the potential performance improvement brought by effective integration NOMA technology with IRS-aided communication.

Challenges

- For multi-antenna NOMA transmission, additional decoding rate conditions need to be satisfied to guarantee successful SIC.
- Both the active and passive beamforming in IRS-NOMA affect the decoding order among users.

[1] X. Mu, Y. Liu, L. Guo, J. Lin and N. Al-Dhahir "Exploiting Intelligent Reflecting Surfaces in Multi-Antenna Aided NOMA Systems", *IEEE Transactions on Wireless Communications*, <https://arxiv.org/abs/1910.13636>.

System Model



- An N -antenna base station serves K single-antenna users through the NOMA protocol with the aid of an IRS with M passive reflecting elements
- $\Theta = \text{diag}(\mathbf{u}) \in \mathbb{C}^{M \times M}$ denotes the diagonal reflection coefficients matrix of the IRS with $\mathbf{u} = [u_1, u_2, \dots, u_M]$ and $u_m = \beta_m e^{j\theta_m}$.

IRS elements assumptions

- **Ideal IRS:** Both the reflection amplitudes and phase shifts can be optimized.

$$\Phi_1 = \{u_m || u_m|^2 \in [0, 1]\} . \quad (15)$$

- **Non-ideal IRS:**

- Continuous phase shifters with the unit modulus constraint.

$$\Phi_2 = \{u_m || u_m|^2 = 1, \theta_m \in [0, 2\pi)\} . \quad (16)$$

- Discrete phase shifters with B resolution bits:

$$\Phi_3 = \{u_m || u_m|^2 = 1, \theta_m \in \mathcal{D}\} , \quad (17)$$

where $\mathcal{D} = \{\frac{n2\pi}{2^B}, n = 0, 1, 2, \dots, 2^B - 1\}$.

Received Signal Model

- The received signal at user k can be expressed as

$$y_k = (\mathbf{h}_k^H + \mathbf{r}_k^H \Theta \mathbf{G}) \sum_{k=1}^K \mathbf{w}_k s_k + n_k, \quad (18)$$

- Based on the NOMA principle, the received **SINR** of user j to decode user k is given by

$$\text{SINR}_{k \rightarrow j} = \frac{|\left(\mathbf{h}_j^H + \mathbf{r}_j^H \Theta \mathbf{G}\right) \mathbf{w}_k|^2}{\sum_{\Omega(i) > \Omega(k)} \left|\left(\mathbf{h}_j^H + \mathbf{r}_j^H \Theta \mathbf{G}\right) \mathbf{w}_i\right|^2 + \sigma^2}. \quad (19)$$

- The corresponding decoding rate is $R_{j \rightarrow k} = \log_2 (1 + \text{SINR}_{j \rightarrow k})$.
- Conditions of success SIC:** $R_{k \rightarrow j} \geq R_{k \rightarrow k}$ for $\Omega(j) > \Omega(k)$.

Optimization Problem

The considered sum rate maximization problem:

$$(P1) : \max_{\Omega, \Theta, \{\mathbf{w}_k\}} \sum_{k=1}^K R_{k \rightarrow k} \quad (20a)$$

$$\text{s.t. } R_{k \rightarrow j} \geq R_{k \rightarrow k}, \Omega(j) > \Omega(k), \quad (20b)$$

$$\left| (\mathbf{h}_k^H + \mathbf{r}_k^H \Theta \mathbf{G}) \mathbf{w}_{\Omega(i)} \right|^2 \leq \left| (\mathbf{h}_k^H + \mathbf{r}_k^H \Theta \mathbf{G}) \mathbf{w}_{\Omega(j)} \right|^2, \forall k, i, j, \Omega(i) > \Omega(j), \quad (20c)$$

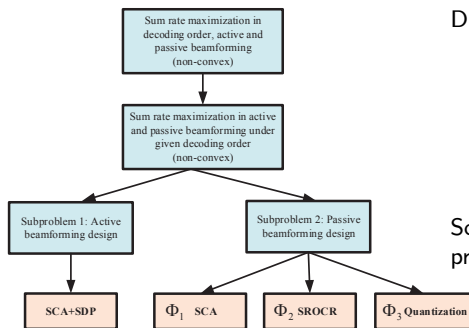
$$\sum_{k=1}^K \|\mathbf{w}_k\|^2 \leq P_T \quad (20d)$$

$$\mathbf{u}_m \in \Phi, \quad (20e)$$

$$\Omega \in \Pi. \quad (20f)$$

- Φ denotes different IRS assumptions.
- Π denotes the set of all possible SIC decoding orders.

Proposed Solutions



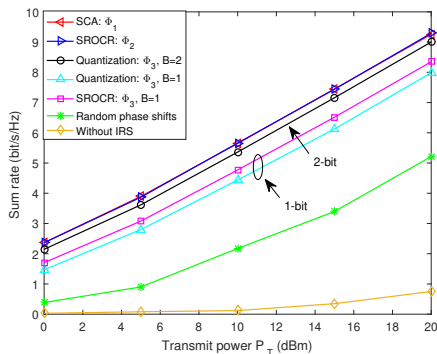
Difficulties:

- Decoding order and beamforming vectors are highly coupled.
- Active and passive beamforming vectors both affect the conditions of success SIC.

Solutions: Divide the complicated problem into some ease of subproblems.

Numerical Results

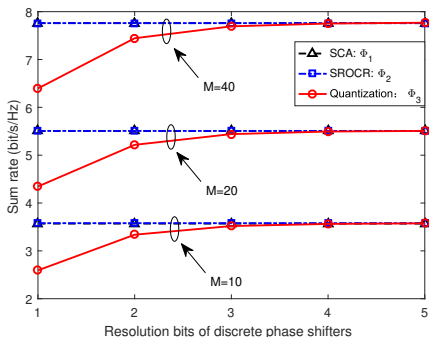
Sum Rate versus Transmit Power



- Significant sum rate gains can be achieved by deploying IRSs with the proposed algorithms.
- The performance gaps between the case of ideal IRS and continuous phase shifters can be ignored.
- The performance degradation caused by finite resolution phase shifters decreases as the bit resolution increases.

Numerical Results

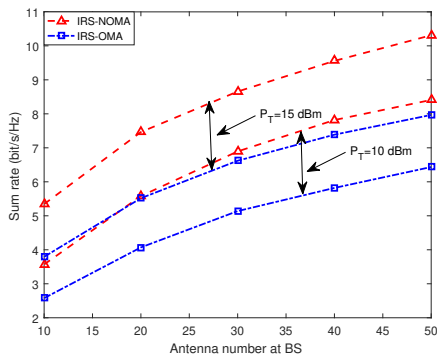
Sum Rate versus the Resolution Bits



- The ideal IRS case achieves the best performance, while the discrete phase shifters case achieves the worst performance.
- “1-bit” and “2-bit” schemes can achieve 80% and 90% performance of the ideal IRS case, respectively.
- The performance loss between the “3-bit” scheme and the ideal IRS is negligible.

Numerical Results

Performance Comparison with OMA



- The IRS-NOMA scheme significantly outperforms the IRS-OMA scheme with more than 5 dB performance gain since all users can be served simultaneously through the NOMA protocol compared with the OMA scheme.

Discussion: Research Opportunities and challenges for NOMA

- 1 Error Propagation in SIC.
- 2 Imperfect SIC and limited channel feedback.
- 3 Synchronization/asynchronization design for NOMA.
- 4 Different variants of NOMA.
- 5 Novel coding and modulation for NOMA.
- 6 Hybrid multiple access
- 7 Efficient resource management for NOMA
- 8 Security provisioning in NOMA
- 9 Different variants of NOMA
- 10 Massive NOMA in IoT Networks

Thank you!

Thank you!