Towards Practical Massive Random Access

Maxime Guillaud, Inria / CITI Lab, Lyon

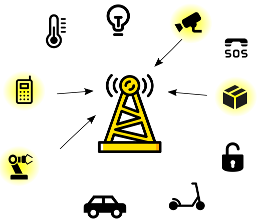
Workshop on Information Theory for Future Networks,
March 4, 2026

Joint work with Charly Poulliat, Sweta Suresh, Claire Goursaud,
Alexis Decurninge, Ingmar Land

This work was supported by ANR under the France 2030 program,
grant NF-PERSEUS (ANR-22-PEFT-0004), and grant WARM-M2M (ANR-24-CE25-2514).

Target Scenario: Massive Random Access

- Multipoint-to-Point, K users
- **Random user activation** (K_a active users), unknown to the receiver
- Regime of interest:
  - Small payload
  - Massively many users ($K \gg K_a$)



- High level of contention: TIN is not desirable due to low SINR, need parallel interference cancelation

First Formulation of the RA Problem

"Proceedings of the NATO Advanced Study Institute on New Concepts in Multi-User Communication, Norwich, UK, August 4-16, 1980"

105

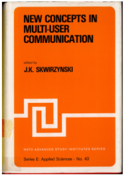
PANEL DISCUSSION

on

'MULTI-USER INFORMATION THEORY'

Friday 15th August 1980 at 16.00 hrs.

Chairman:	Professor J.K. Wolf	University of Massachusetts, USA.
Panel Members:	Dr. B.G. Dorsch	DFVLR, Germany.
	Mr. C. Heegard	Stanford University, USA.
	Dr. P.P. Nuspl	Communications Research Centre, Canada.
	Dr. R.J.F. Fang	COMSAT Laboratories, USA.
	Professor E. van der Meulen	Katholieke Universiteit, Leuven, The Netherlands.
	Mr. D.G.W. Ingram	University of Cambridge, UK.
	Professor H.F.T. Ericson	University of Linkoping, Sweden.
	Professor T. Cover	Stanford University, USA.
Contributors:	Professor A. Ephremides	University of Maryland, USA.
	Dr. P. Fire	US Department of the Navy, UK.
	Professor H. Kobayashi	IBM Research Center, USA.
	Mr. J.J. O'Reilly	University of Essex, UK.
	Professor M. Schwartz	University of Columbia, USA.
	Mr. C.P. Welti	COMSAT Laboratories, USA.



MULTIPLE ACCESS CHANNEL

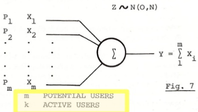


Fig. 7

Wolf's problem: Now let me turn to a problem presented by Professor Wolf. (See Figure 7). In a multiple access channel with, say, 100 potential users, only 5 of which are going to use the channel at any given time, how do you design a code for each of these 100 users when you know that the union of the codewords is going to choke the space of all available codewords. How can a receiver know which codeword was sent? Can one achieve the channel capacity for a multiple access channel with codes designed for only 5 users? If not, how well can you do?

At least in the Gaussian channel we an answer the question. First I repeat the problem more precisely. The multiple access channel has n potential users, only k of which will be active at any given time. Each user has power P and there is additive Gaussian noise Z. In the special case k = m, all the users are active, and the multiple access channel capacity region holds yielding a capacity per user of R = C(kP)k, where C(x) = (1/2)log(1 + x). Notice that the total rate of information is kR = C(kP), and this grows to infinity as log(k). Thus it is pleasing that the total number of bits received at the receiver is not bounded above as the number of users grows to infinity.

The question we now address is what happens if the number of potential users m is much greater than k. We then need many codebooks, say m = 10^6. But the answer is the same as before! The per user rate does not depend on m.

The idea of the proof is that you allocate a little bit of time at the beginning for each of the active users to give his name. That won't affect the rate very much because there are a finite number of potential users. Now use a random codebook for each of the m users. Append to the codewords in the codebook for the kth user the name of the kth user. Receiver Y sees the additive Gaussian noise plus a codeword from each of the k codebooks for the active users of the m potential users. The receiver should not perform a nearest neighbour search for the closest subset of k codewords from the union of the codebooks. This would be a disaster. Rather, he first decodes the names of the active users. He then limits his attention to the k codebooks corresponding to the accurately known k users. Consequently, the answer to Wolf's question is yes: if you have an underutilized set of potential users, you can do as well as if the other threatening potential users were not there at all.

Professor Wolf - No more questions; we are going to the banquet. I would like to thank all the speakers and the audience for a most interesting discussion.

Multiple Access vs. Random Access

Multiple Access: divide-and-conquer approach

- Resource **grants**
- **Closed-loop** synchronization (OFDM symbol, TA)
- Power control, rate selection
- **Coordinated** assignment of orthogonal pilots
- CSI estimation
- MU-MIMO equalization
- Carrier frequency **offset compensation**
- FEC coding for the AWGN channel

Random Access: sporadic, uncoordinated transmission

All of the above functionality has to be implemented at the time scale of a single packet arrival

Gallager's Perspective: Impact on Coding

IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. IT-31, NO. 2, MARCH 1985

A Perspective on Multiaccess Channels

Invited Paper

ROBERT G. GALLAGER, FELLOW, IEEE

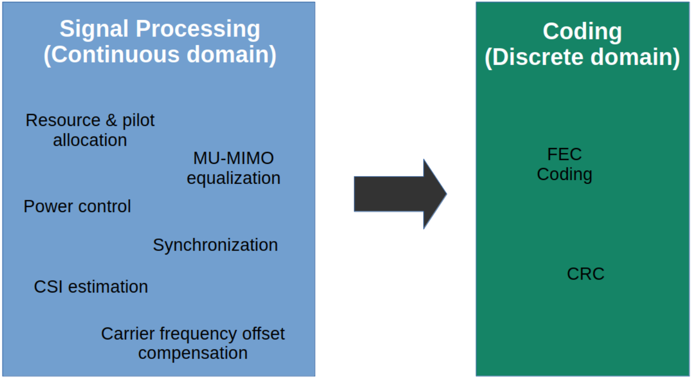
sonable for multiaccess channels. For a point to point channel, one normally assumes an infinite reservoir of data to be transmitted. The reason for this is that it is a minor practical detail to inform the receiver when there is no data to send; furthermore, there is no other use for the channel, so potential lack of data might as well be left out of the model. For multiaccess channels, on the other hand, most transmitters have nothing to send most of the time, and only a few are busy. The problem is then to share the channel between the busy users, and this is often the central technical problem in multiaccess communication.

From a more practical point of view, the single user limit theorems of information theory are interesting both because they put an upper limit on what is achievable and because the limit is usually not too far from what is practically achievable. For a multiaccess channel, however, the long time intervals required for the source arrivals to appear smoothed out are typically far greater than the tolerable delays. Conversely, the time interval required for coding to be effective (i.e., the time for the noise to be smoothed out) is typically smaller than the tolerable delay. What is needed then is an information theoretic model that somehow precludes the possibility of imposing long delays on source messages.

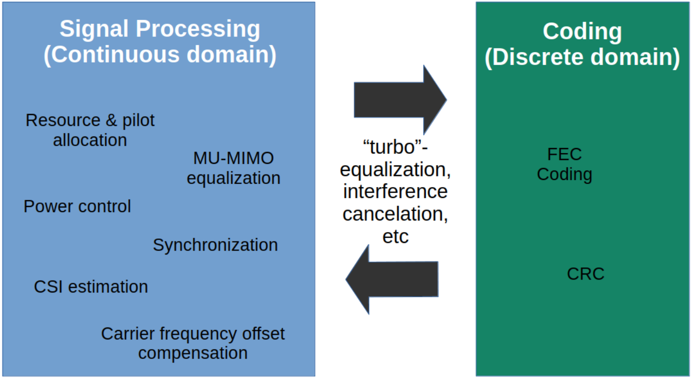
One approach to this, which is used in the collision resolution field, is to assume an infinite number of sources, or equivalently, that a new transmitter is created for each new arriving message and then destroyed when the message is successfully transmitted. The received sequence or waveform would then be some function of noise and whatever was being transmitted by the active transmitters. It seems that to develop understanding in this area, it is necessary first to develop some understanding of coding (as opposed to coding theorems) in a multiaccess environment. This understanding should involve decoding in the presence of several messages being transmitted simultaneously, since otherwise the problem simply reduces to conflict resolution with coding added for reliable transmission in the absence of conflicts.

- Full buffer assumption does not hold for RA
- Closed-loop "divide and conquer" approaches are excluded

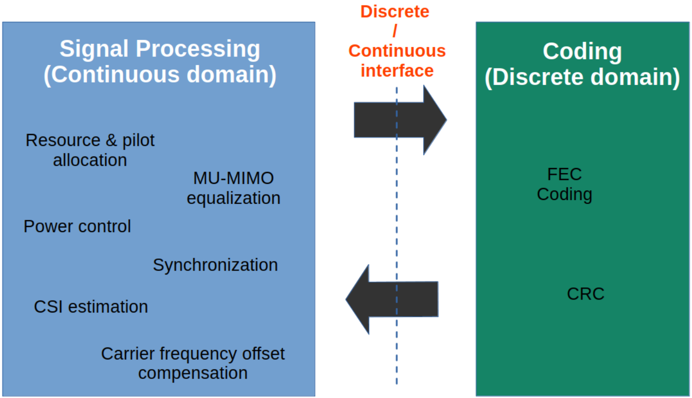
Multiple Access vs. Random Access



Multiple Access vs. Random Access



Multiple Access vs. Random Access



Channel Model: SIMO, Non-Coherent, Quasi-Static Block-fading

- Set $\mathcal{A}$ of active users, unknown to the receiver
- Block-fading with blocklength T
- Block synchronicity across the users
- M receive antennas, channel state $\mathbf{h}_k \in \mathbb{C}^M$ for user k
- User k transmits a sequence $\mathbf{s}_k \in \mathbb{C}^T$
- At the receiver:

$$\mathbf{y} = \sum_{k \in \mathcal{A}} \mathrm{vec}\left(\mathbf{s}_k \mathbf{h}_k^T\right) + \mathbf{w} = \sum_{k \in \mathcal{A}} \mathbf{s}_k \otimes \mathbf{h}_k + \mathbf{w} \in \mathbb{C}^{TM}$$

Multilinear Spreading

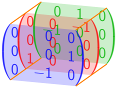
- Assume that the blocklength factorizes as $T = \prod_{i=1}^{d} T_i$
- Let the transmitted sequence $\mathbf{s}_k$ be defined as a **multilinear product of information-bearing vectors**:

$$\mathbf{s}_k = \mathbf{x}_{1,k} \otimes \cdots \otimes \mathbf{x}_{d,k}$$

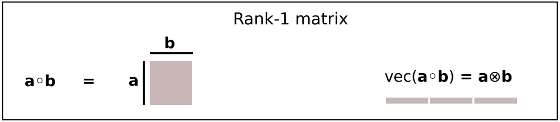
- $\mathbf{x}_{i,k}$ is a vector modulation suitable for **single-user, non-coherent** communications (e.g. 1 reference symbol + $T_i - 1$ QAM-modulated symbols)

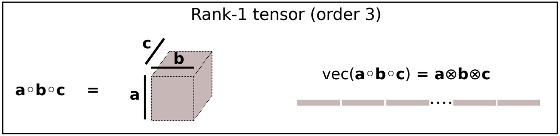
Tensor Algebraic Interpretation

Tensors generalize matrices to $d \geq 2$ dimensions



- Rank 1, outer product $\circ$ and Kronecker product $\otimes$:





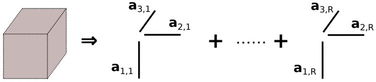
- Generalization to tensors of order $d \geq 2$:

$$\mathrm{vec}(\mathbf{a}_1 \circ \mathbf{a}_2 \circ \cdots \circ \mathbf{a}_d) = \mathbf{a}_1 \otimes \mathbf{a}_2 \otimes \cdots \otimes \mathbf{a}_d$$

- $\mathrm{vec}(\cdot)$ is an isomorphism between the space of $(T_1, \ldots, T_d)$-dimensional tensors and the space of $(\prod_{i=1}^{d} T_i)$-dimensional vectors (with the respective sums)

Tensor CP Decomposition (PARAFAC)

- A tensor of **sub-generic rank** almost surely admits a **unique canonical polyadic (CP) decomposition** into rank-1 components



- Tensors can have "high" rank (min. # of CP components)
  - Example: a (5,5,5,5) generic tensor has rank 37 a.s.
- "Unicity" is defined
  - up to a permutation between the R terms, and
  - up to $d-1$ scalars per rank-1 term

“Tensor-Based” Random Access [1]

- At each transmitter: $\mathbf{s}_k$ is a **rank-1 tensor** of dimension $(T_1, \ldots, T_d)$:

$$\mathbf{s}_k = \mathbf{x}_{1,k} \otimes \cdots \otimes \mathbf{x}_{d,k}$$

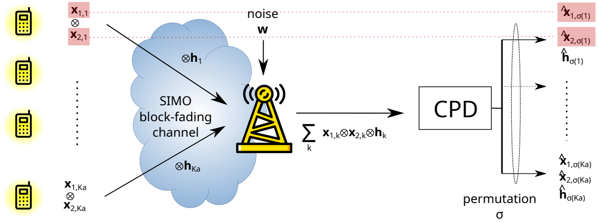
- At the receiver:

$$\mathbf{y} = \sum_{k \in \mathcal{A}} \underbrace{\mathbf{x}_{1,k} \otimes \cdots \otimes \mathbf{x}_{d,k} \otimes \mathbf{h}_k}_{\substack{\text{rank-1 tensor of order } d+1 \\ \text{and dimension} (T_1, \ldots, T_d, M)}} + \mathbf{w}$$

- User separation is a **low-rank tensor decomposition** problem, with mild unique decomposition conditions

[1] A. Decurninge, I. Land, and M. Guillaud, “Tensor-based modulation for unsourced massive random access,” **IEEE Wireless Communications Letters**, vol. 10, no. 3, Mar. 2021.

Noise-Free Performance Characterization



- Noise-free case: $\hat{\mathbf{x}}_{i,\sigma(k)} \propto \mathbf{x}_{i,k}$ as long as the decomposition is uniquely identifiable
- Each rank-1 component (=1 user) is equivalent to the parallel, noise-free transmission of d Grassmannian variables
  - Per-user **complex DoF**: $\sum_{i=1}^{d}(T_i - 1)$

Maximum Achievable K_a (Noise-Free Case) [2]

- For $(T_1, \ldots, T_d, N)$-tensors with general choices of the $\mathbf{x}_{i,k}, \mathbf{h}_k$, the CPD is **almost surely unique** for $K_\mathrm{a} < \overline{R}$ where

$$\overline{R} = \begin{cases} R^1 - 1 & \text{for } T_1 \geq R^1 \\ R^2 - 1 & \text{for } M \geq R^2 \\ R^0 & \text{otherwise} \end{cases}$$

where

$$R^1 = 2 - M + M \prod_{i=2}^{d} T_i - \sum_{i=2}^{d} (T_i - 1),$$

$$R^2 = 1 + T - \sum_{i=1}^{d} (T_i - 1) \quad \text{and}$$

$$R^0 = \left\lceil \frac{M \prod_{i=1}^{d} T_i}{M + \sum_{i=1}^{d} (T_i - 1)} \right\rceil \quad \text{(expected generic rank)}$$

- **The maximum supported K_a increases with # of Rx antennas M**

[2] L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven, "An algorithm for generic and low-rank specific identifiability of complex tensors," **SIAM Journ. Matrix Analysis and Applications**, vol. 35, no. 4, 2014.

TBM DoF Bounds

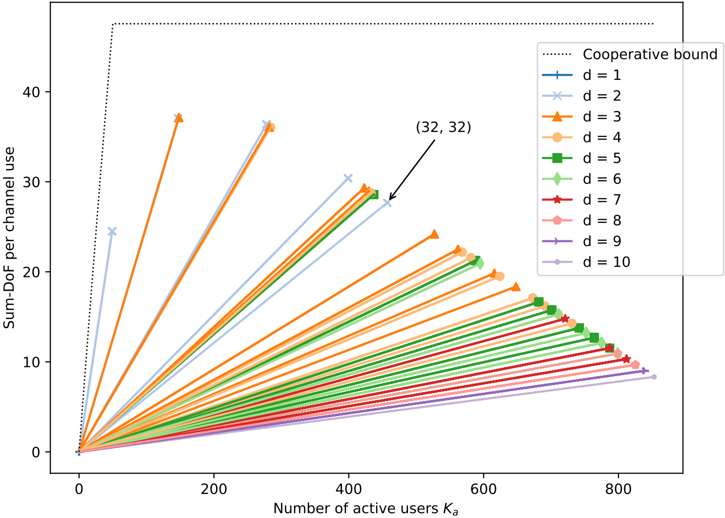
- At most $K_{\mathrm{a}} = \overline{R} - 1$ users can be identifiable, hence the maximum Sum-DoF:

$$\mathrm{D}_{\mathrm{TBM}}(\overline{R}-1) = (\overline{R}-1)\sum_{i=1}^{d}(T_i - 1)$$

- Cooperative DoF upper-bound (point-to-point $M \times K_{\mathrm{a}}$ non-coherent channel):

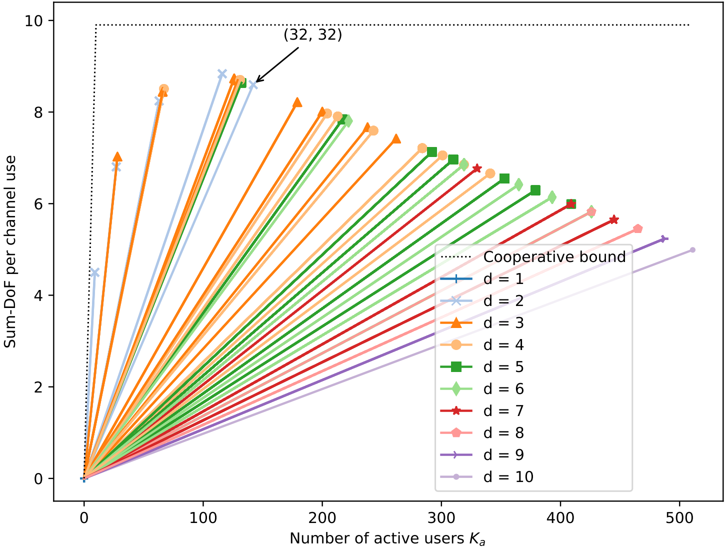
$$\mathrm{D}_{\mathrm{coop}}(K_{\mathrm{a}}) = M^*(T - M^*) \quad \text{where} \quad M^* = \min(K_{\mathrm{a}}, M, \lfloor T/2 \rfloor)$$

Achievable DoF, $T = 1024$, $M = 50$ antennas



Achievable sum-DoF per channel use $(\mathrm{D}_{\mathrm{TBM}}(K_{\mathrm{a}})/T)$ vs. K_{a} for different tensor sizes (d and T_i). Markers denote $K_{\mathrm{a}} = \overline{R} - 1$, while the slope of the lines going through the origin represents the per-user DoF.

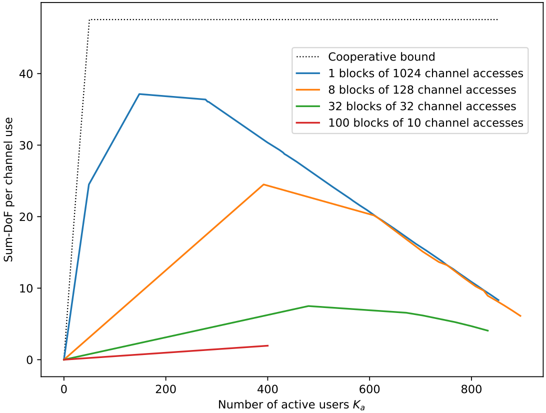
Achievable DoF, $T = 1024$, $M = 10$ antennas



Achievable sum-DoF per channel use $(\mathrm{D_{TBM}}(K_\mathrm{a})/T)$ vs. K_a for different tensor sizes (d and T_i). Markers denote $K_\mathrm{a} = \overline{R} - 1$, while the slope of the lines going through the origin represents the per-user DoF.

Genie-Aided Orthogonalization

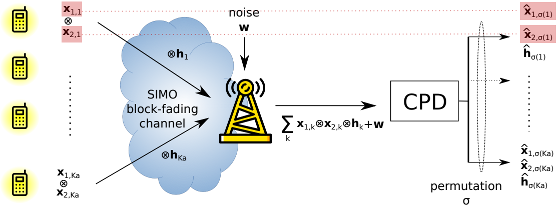
- Active users are distributed into L even-sized groups that use orthogonal resources (sub-blocks of length $\frac{T}{L}$)
- TBM is used **within** each subblock



- Grouping (even if perfectly even) is detrimental to Sum-DoF

Finite SNR Performance Characterization

- Characterization of the $\mathbf{x}_{i,k} \longmapsto \hat{\mathbf{x}}_{i,\sigma(k)}$ equivalent channel is needed for an information-theoretic performance metric



- The $\mathbf{x}_{i,k}$ are related to the tensor singular vectors [3]

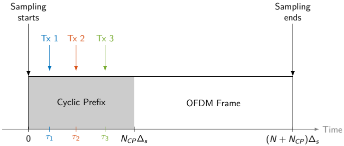
- Some tools based on random matrix theory for perturbation analysis [4] (asymptotic, low-rank)

[3] L.-H. Lim, "Singular values and eigenvalues of tensors: A variational approach," in **IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing**, IEEE, 2005, pp. 129–132.

[4] M. El Amine Seddik, M. Guillaud, and R. Couillet, **Annals of Applied Probability**, vol. 1A, no. 34, pp. 203–248, 2024. DOI: 10.1214/23-AAP1962.

Multilinear Spreading with Channel Impairments

- OFDM, single-path channel with unknown lag τ_k per user [5]



- The channel gain includes a multiplicative phase ramp

$$\boldsymbol{D}(\phi_k) = \text{diag}\left(1, e^{\imath\phi_k}, \cdots, e^{\imath(N-1)\phi_k}\right) \quad \text{with} \quad \phi_k = \frac{2\pi\tau_k}{N\Delta_s}$$

- Under some technical conditions on $T_1, \ldots T_d$ and N,

$$\boldsymbol{D}(\phi_k) = D_1(\phi_k) \otimes \ldots \otimes D_d(\phi_k)$$

$$\Rightarrow \boldsymbol{y} = \sum_{k\in\mathcal{A}} (\boldsymbol{D}_1(\phi_k)\boldsymbol{x}_{1,k}) \otimes \ldots \otimes (\boldsymbol{D}_d(\phi_k)\boldsymbol{x}_{d,k}) \otimes \boldsymbol{h}_k + \boldsymbol{w}$$

- The CPD can still separate the (rank-1) users!

[5] A. Decurninge, P. Ferrand, and M. Guillaud, "Massive random access with tensor-based modulation in the presence of timing offsets," in **Proc. GLOBECOM**, 2022, pp. 1061–1066.

Multilinear Spreading

$$\mathbf{s}_k = \mathbf{x}_{1,k} \otimes \cdots \otimes \mathbf{x}_{d,k}$$

Joint coding (redundancy) and blind user separation
in the modulation domain

- Pros: makes equalization (blind user separation, impairment compensation. . .) easy

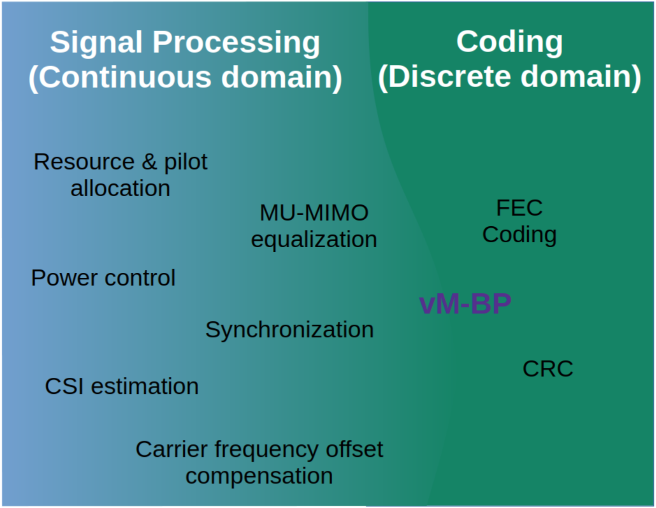
Multilinear Spreading

$$\mathbf{s}_k = \mathbf{x}_{1,k} \otimes \cdots \otimes \mathbf{x}_{d,k}$$

Joint coding (redundancy) and blind user separation
in the modulation domain

- Pros: makes equalization (blind user separation, impairment compensation...) easy
- Cons: what is this "code" actually doing?

Conclusion


Signal Processing
(Continuous domain)
Coding
(Discrete domain)
Resource & pilot allocation
MU-MIMO equalization
FEC Coding
Power control
vM-BP
Synchronization
CRC
CSI estimation
Carrier frequency offset compensation