The Physical Layer of 5G and Beyond

EURECOM, January 24, 2025

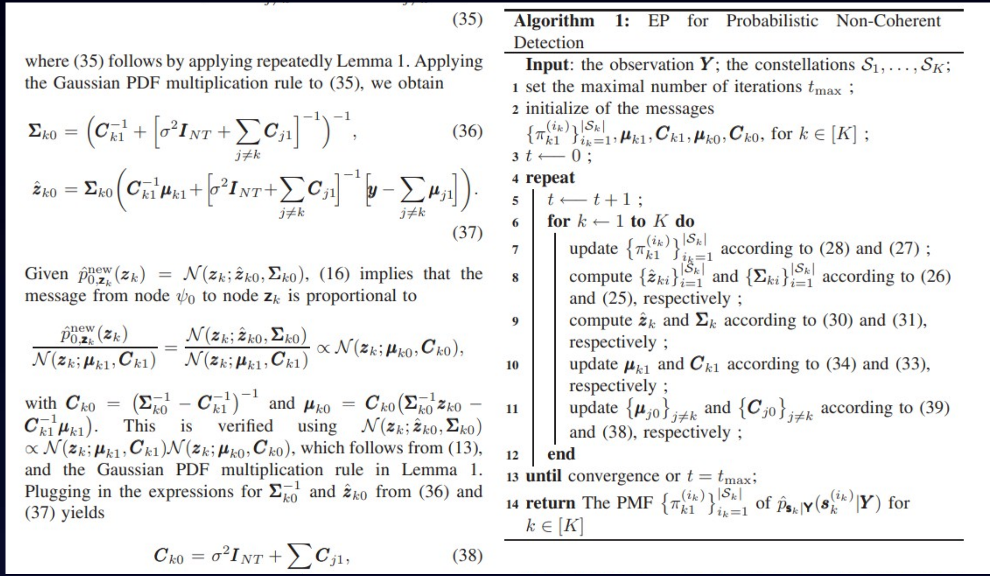
Maxime Guillaud

maxime.guillaud@inria.fr

https://maximeguillaud.github.io/

A Day in the Life of a Researcher

- Read the scientific literature
- Write equations and scientific papers
- Interact with colleagues
- Powerpoint
- Invent and patent things
- Supervise PhD students
- Interact with academic researchers
- Present in scientific conferences
- Teach graduate courses
- Code



(35)

where (35) follows by applying repeatedly Lemma 1. Applying the Gaussian PDF multiplication rule to (35), we obtain

$$\boldsymbol{\Sigma}_{k0} = \left(\boldsymbol{C}_{k1}^{-1} + \left[\sigma^2 \boldsymbol{I}_{NT} + \sum_{j\neq k} \boldsymbol{C}_{j1}\right]^{-1}\right)^{-1}, \quad (36)$$

$$\hat{\boldsymbol{z}}_{k0} = \boldsymbol{\Sigma}_{k0}\left(\boldsymbol{C}_{k1}^{-1}\boldsymbol{\mu}_{k1} + \left[\sigma^2 \boldsymbol{I}_{NT} + \sum_{j\neq k} \boldsymbol{C}_{j1}\right]^{-1}\left[\boldsymbol{y} - \sum_{j\neq k}\boldsymbol{\mu}_{j1}\right]\right). \quad (37)$$

Given $\hat{p}^{\text{new}}_{0,\boldsymbol{z}_k}(\boldsymbol{z}_k) = \mathcal{N}(\boldsymbol{z}_k; \hat{\boldsymbol{z}}_{k0}, \boldsymbol{\Sigma}_{k0})$, (16) implies that the message from node ψ_0 to node $\mathbf{z}_k$ is proportional to

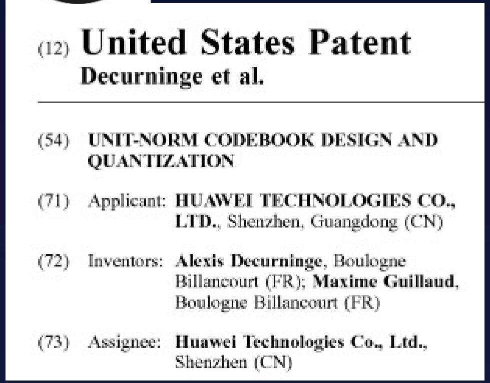
$$\frac{\hat{p}^{\text{new}}_{0,\boldsymbol{z}_k}(\boldsymbol{z}_k)}{\mathcal{N}(\boldsymbol{z}_k; \boldsymbol{\mu}_{k1}, \boldsymbol{C}_{k1})} = \frac{\mathcal{N}(\boldsymbol{z}_k; \hat{\boldsymbol{z}}_{k0}, \boldsymbol{\Sigma}_{k0})}{\mathcal{N}(\boldsymbol{z}_k; \boldsymbol{\mu}_{k1}, \boldsymbol{C}_{k1})} \propto \mathcal{N}(\boldsymbol{z}_k; \boldsymbol{\mu}_{k0}, \boldsymbol{C}_{k0}),$$

with $\boldsymbol{C}_{k0} = (\boldsymbol{\Sigma}_{k0}^{-1} - \boldsymbol{C}_{k1}^{-1})^{-1}$ and $\boldsymbol{\mu}_{k0} = \boldsymbol{C}_{k0}(\boldsymbol{\Sigma}_{k0}^{-1}\hat{\boldsymbol{z}}_{k0} - \boldsymbol{C}_{k1}^{-1}\boldsymbol{\mu}_{k1})$. This is verified using $\mathcal{N}(\boldsymbol{z}_k; \hat{\boldsymbol{z}}_{k0}, \boldsymbol{\Sigma}_{k0}) \propto \mathcal{N}(\boldsymbol{z}_k; \boldsymbol{\mu}_{k1}, \boldsymbol{C}_{k1})\mathcal{N}(\boldsymbol{z}_k; \boldsymbol{\mu}_{k0}, \boldsymbol{C}_{k0})$, which follows from (13), and the Gaussian PDF multiplication rule in Lemma 1. Plugging in the expressions for $\boldsymbol{\Sigma}_{k0}^{-1}$ and $\hat{\boldsymbol{z}}_{k0}$ from (36) and (37) yields

$$\boldsymbol{C}_{k0} = \sigma^2 \boldsymbol{I}_{NT} + \sum \boldsymbol{C}_{j1}, \quad (38)$$

Algorithm 1: EP for Probabilistic Non-Coherent Detection

Input: the observation $\boldsymbol{Y}$; the constellations $\mathcal{S}_1, \ldots, \mathcal{S}_K$;
set the maximal number of iterations $t_{\max}$;
initialize of the messages $\{\pi_{k1}^{(i_k)}\}_{i_k=1}^{|\mathcal{S}_k|}, \boldsymbol{\mu}_{k1}, \boldsymbol{C}_{k1}, \boldsymbol{\mu}_{k0}, \boldsymbol{C}_{k0}$, for $k \in [K]$;
$t \longleftarrow 0$;
repeat
$t \longleftarrow t+1$;
for $k \leftarrow 1$ **to** K **do**
update $\{\pi_{k1}^{(i_k)}\}_{i_k=1}^{|\mathcal{S}_k|}$ according to (28) and (27) ;
compute $\{\hat{\boldsymbol{z}}_{ki}\}_{i=1}^{|\mathcal{S}_k|}$ and $\{\boldsymbol{\Sigma}_{ki}\}_{i=1}^{|\mathcal{S}_k|}$ according to (26) and (25), respectively ;
compute $\hat{\boldsymbol{z}}_k$ and $\boldsymbol{\Sigma}_k$ according to (30) and (31), respectively ;
update $\boldsymbol{\mu}_{k1}$ and $\boldsymbol{C}_{k1}$ according to (34) and (33), respectively ;
update $\{\boldsymbol{\mu}_{j0}\}_{j\neq k}$ and $\{\boldsymbol{C}_{j0}\}_{j\neq k}$ according to (39) and (38), respectively ;
end
until convergence or $t = t_{\max}$;
return The PMF $\{\pi_{k1}^{(i_k)}\}_{i_k=1}^{|\mathcal{S}_k|}$ of $\hat{p}_{\mathbf{s}_k|\mathbf{Y}}(\boldsymbol{s}_k^{(i_k)}|\boldsymbol{Y})$ for $k \in [K]$



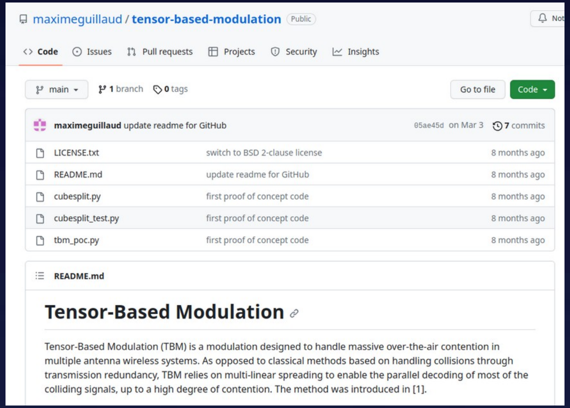
(12) **United States Patent**
Decurninge et al.

(54) UNIT-NORM CODEBOOK DESIGN AND QUANTIZATION

(71) Applicant: **HUAWEI TECHNOLOGIES CO., LTD.**, Shenzhen, Guangdong (CN)

(72) Inventors: **Alexis Decurninge**, Boulogne Billancourt (FR); **Maxime Guillaud**, Boulogne Billancourt (FR)

(73) Assignee: **Huawei Technologies Co., Ltd.**, Shenzhen (CN)



maximeguillaud / **tensor-based-modulation** Public

Code | Issues | Pull requests | Projects | Security | Insights

main · 1 branch · 0 tags · Go to file · Code

maximeguillaud update readme for GitHub · 05ae45d on Mar 3 · 7 commits

LICENSE.txt	switch to BSD 2-clause license	8 months ago
README.md	update readme for GitHub	8 months ago
cubesplit.py	first proof of concept code	8 months ago
cubesplit_test.py	first proof of concept code	8 months ago
tbm_poc.py	first proof of concept code	8 months ago

README.md

Tensor-Based Modulation

Tensor-Based Modulation (TBM) is a modulation designed to handle massive over-the-air contention in multiple antenna wireless systems. As opposed to classical methods based on handling collisions through transmission redundancy, TBM relies on multi-linear spreading to enable the parallel decoding of most of the colliding signals, up to a high degree of contention. The method was introduced in [1].

Outline

- The 5G Standard
  - Objectives, Service Classes
- Spectrum and frequency bands
- Massive MIMO
  - Antenna geometries and channel models
  - CSI acquisition
  - Multi-user aspects
  - Hardware impairments, Full duplex
- mmWave
- Other novel aspects in 5G
  - Waveforms
  - UCNC, CUPS, Cloud/Fog RAN
- Towards 6G
- Applications of Machine Learning

Standardization of 5G

Traffic Classes in 5G

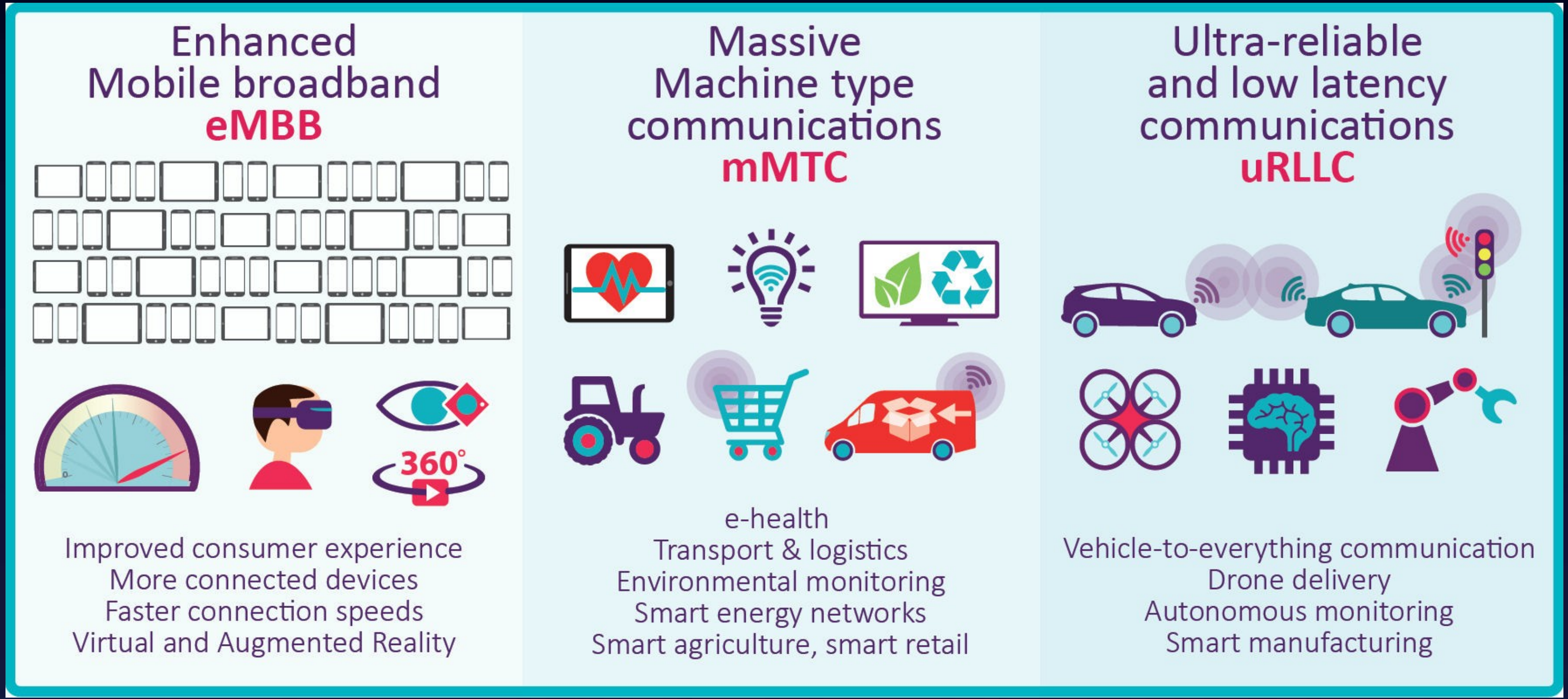
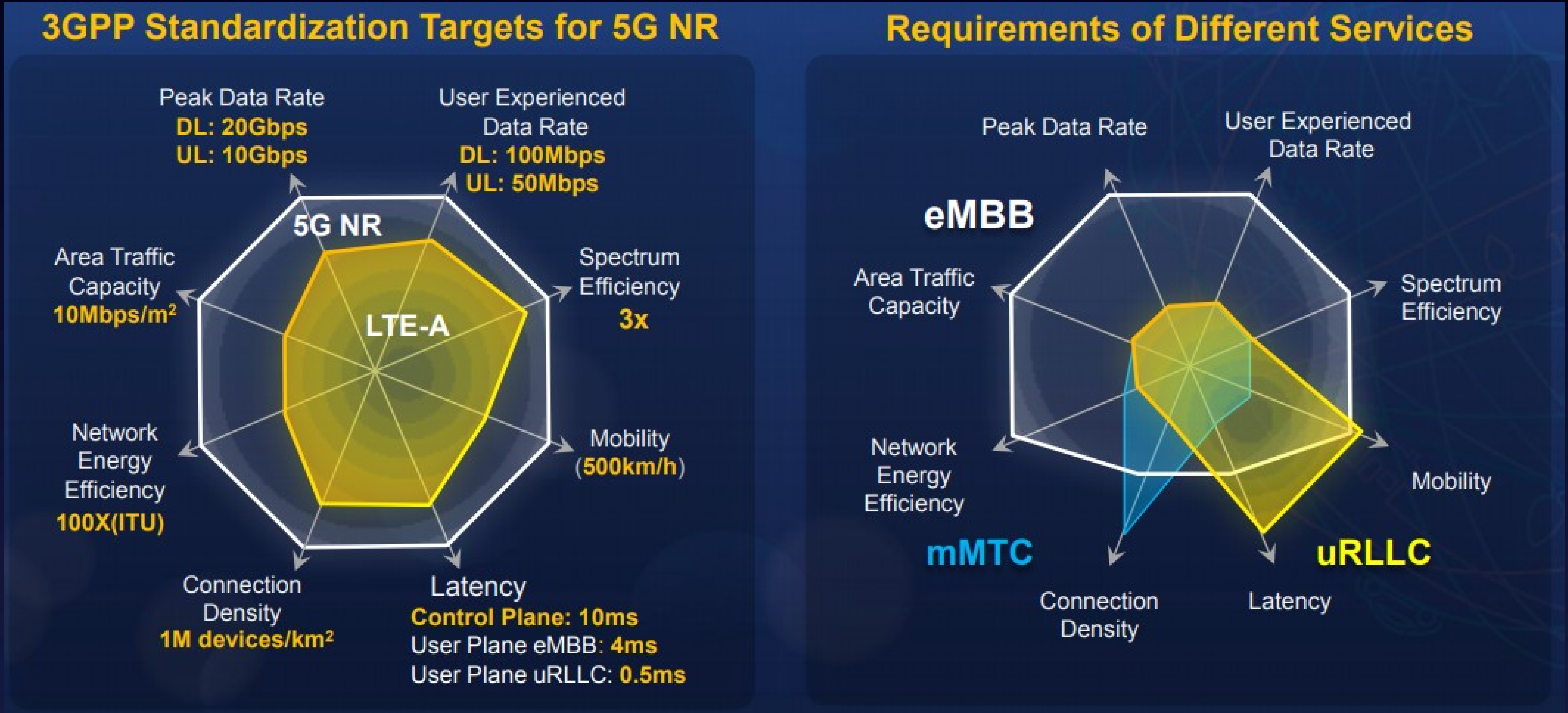


Illustration: OFCOM (UK)

5G Performance Targets

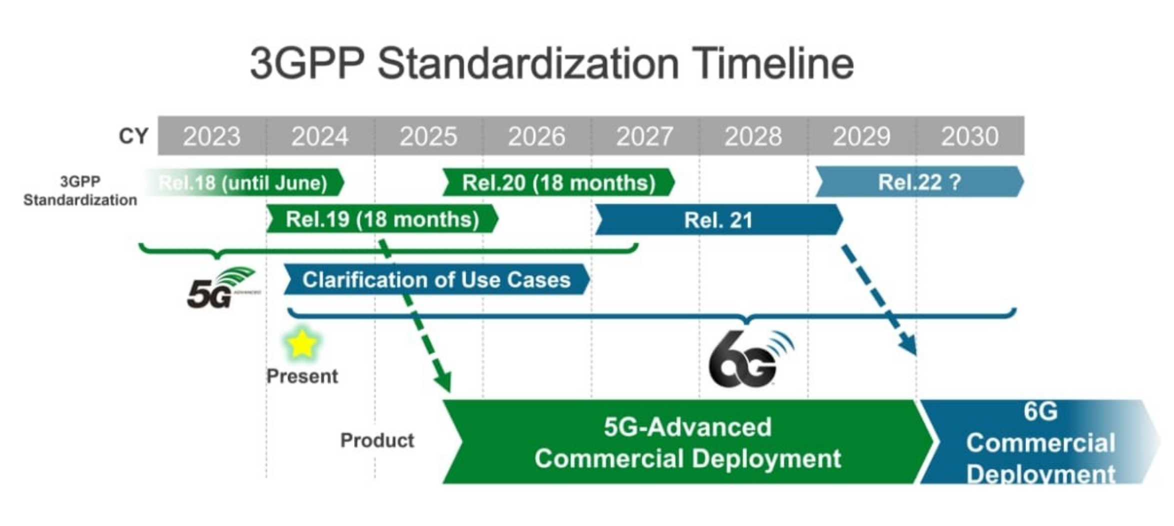


Source: Hamid Reza KARIMI (Huawei), “Bringing 5G to reality.”

Standardization: How Does It Work?

3GPP is the standardization forum for cellular networks since 3G

- Regular meetings with technical discussions
- Regular releases between the generations (4G, 5G, 6G…)

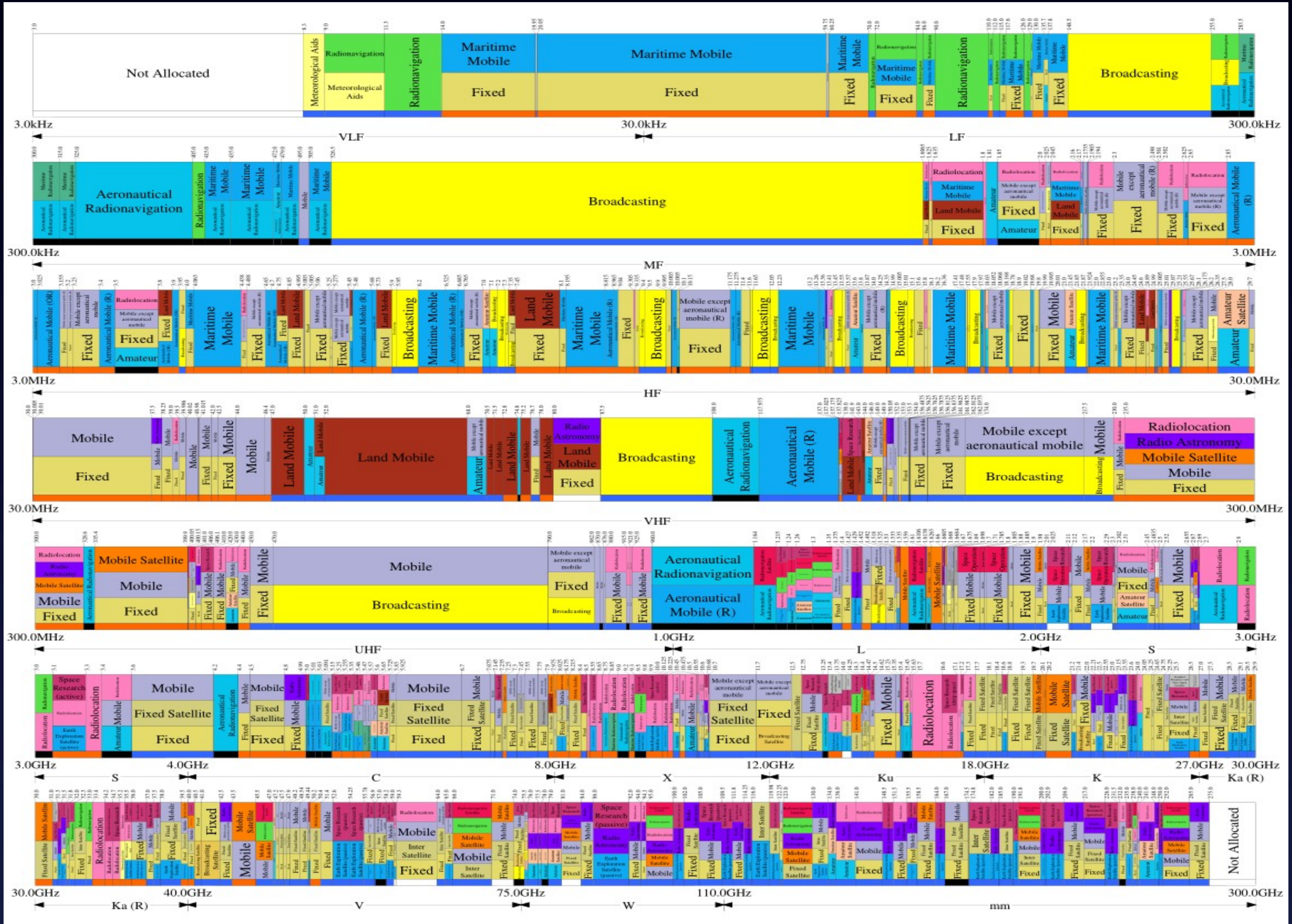


3GPP Member Companies (2017 list)

450connect, 7Layers, Astrid, Aalborg University, AccelerComm, Accuris, Acer, Acorn, adare, ADVA, Aeroflex / Cobham, Affirmed, Airbus, Air-Lynx, Airwave, Alcatel-Lucent, Alibaba, Allot, Altiostar, Amdocs, Analog Devices, Andrew, Anemone, Anritsu, Apple, Applied Communication Sciences, AQSACOM, Arcep, AREA,, ArgoNET, Ascom, Aselsan, ASTRI, Asustek, AT&T, AT4, ATR, Avanti, Azimuth, Azqtel, BAE, BankID Norge, BBC, Beijing Starpoint, Beijing Xiaomi, Beijing Xinwei, Bell, BfV, Bittium, BlackBerry, Bluetest, BMWi, BOCRA, Bolloré, Bouygues, Broadcom, Brocade, BT, Bull, BUPT, Bureau Veritas, C Spire, CableLabs, Cambium, Carnegie, Catapult, CATR, CATT, Cavium, C-DOT, CEA, Cellnex, Celltick, Ceragon, CESG, CETECOM, CEWiT, CGC, Chengdu TD, China Mobile, China Telecom, China Unicom, Chongqing University, CHTTL, Cisco, CITC, CJSC Peter-Service, CMDI, CNES, Cohere, Coherent Logix, Comcast, Comprion, Comptel, Comtech, Continental, Convida, CPqD, CTTC, Czech Technical University, Dai Nippon Printing, Datang, DCMS, DEKRA, DENSO, Deutsche Telekom, Dish, DOCOMO, Dolby, Dongguan OPPO, DSPG Edinburgh, DTS Licensing, East China Institute of Tel., E-Blink, EBU, ECO, EIR, Enensys, Ericsson, ESA, ETELM, ETISALAT, ETRI, ETS-Lindgren Europe, EURECOM, European Commission, European Patent Organisation, Eutelsat, Expway, Fabasoft, Facebook, Fairspectrum, Fastweb, FCC, Fiberhome, FICORA, Finmeccanica, FirstNet, France Brevets, Fraunhofer, Friedrich-Alex-Universität, Fudan University, Fujitsu, Future Cities Catapult, Gemalto, Genband, General Dynamics, Giesecke & Devrient, Gigaset, Gohigh Data, Google, GTRC, Guangdong OPPO, Guardtime, GWT, HangZhou, H3C, Hansung University, Harris, Harting, HCL, Head acoustics, HEPTA 7291, Heron, Hewlett-Packard, Hisense, HiSilicon, Hitachi, HTC, Huawei, Hughes, Husqvarna, Hytera, IAESI, IBM, Idaho National Lab, III, IIT Bombay, Imagination, Indian Institute of Tech, Infineon, Inmarsat, Innovative Technology Lab, Institut Mines-Telecom, Institute Vedecom, Intel, Interdigital, Intertek, IPCom & KG, Irdeto, IRT, Iskratel, Italtel, Itochu, ITRI, Japan Radio, Johns Hopkins University, Juniper, Kapsch, Kathrein, KDDI, Keysight, Knowles, Kodiak, Korea Testing Laboratory, KPN, KRRI, KT, Kyocera, Legrand, Lenovo, Leonardo, LG, Ligado, Lockheed Martin, Marben, McGill University, MCIT, MediaTek, Meizu, Mesaqin, Microsemi, Microsoft, Mitel Mobility, Mitre, Mitsubishi, Mobile Tornado, Morpho Cards, Motorola, MTCC, MTI, MTN, Multi-Tech Systems, MVG, Nadir Hacini, Nanjing, Ericsson Panda, Nanjing Ticom, National Instruments, Taiwan University, nbn, NCSC, NEC, Netas, NetComm, Netscout, Neul, Nextlink, NextNav, NICT, Nkom, Nokia, Nomor, Nordic Semiconductor, NTT, Nubia, NXP, Oberthur, OFCOM, Ogero, Oki, one2many, OnStar, Ooredoo, Openet, Oracle, Orange, OTD, P3, Panasonic, PCTest Engineering Lab, Peking University, Peter-Service, Philips, PIDS, PIWorks, Polaris, Potevio, Prisma, Procera, Proximus, PT, PTS, Public Safety Canada, Qihoo 360, Qorvo, Qualcomm, Quixoticity, RadiSys, Ranzure, RATEL, RED, Redline, Reliance Jio, Robert Bosch, Rogers, Rohde & Schwarz, Sagemcom, Saguna, Samsung, Sanchuan Wisdom, Sandisk, Sandvine, Sanechips, Semtech Neuchatel, Sepura, Sequans, Sernet (Suzhou), SES, SGDSN, SGS Wireless, Shanghai Chen Si, Shanghai Jiao Tong University, Shanghai Tejet Com, Sharp, Shenzhen Coolpad, Shenzhen OPPO, SIA, Siemens, Sierra, Sigfox, Sigos, Sisvel, SK, SkyTL, Skyworks, SoftBank, Softel, Sonus, Sony, Southern LINC, Southwest Jiaotong university, SP Technical Research, SpiderCloud, Spirent, Sporton, Spreadtrum, Sprint, SRG, SRTC, STMicroelectronics, Straight Path, Sumitomo Elec., Suomen virveverkko, Swisscom, SyncTechno, TAIT, Tamum, TCL, TCT, TD Tech, TDF, TechMahindra, Technicolor, Tejas, Telecom Italia, Telefonica, Telekom Deutschland, Telenor, Teleste, TeliaSonera, Telit, Telkom, Telstra, Telus, Teradyne, Texas Instruments, Thales, Netherlands Police, Tianjin Samsung, T-Mobile, TNO, Tongji University, Toshiba, Toyota, Trafikverket, trovicor, TruePosition, Tsofun Algorithm, Turk Telekomunikasyon, TUV Sud, US Department of Commerce, u-blox, UK Broadband, UL VS, Union Inter. Chemins de Fer (UIC), Union, UPV/EHU, US Cellular, UTokyo, Valid, Vasona, Vencore, Veolia, Verizon, Viavi, Virtuosys, vivo, Vodafone, VoiceAge, Volkswagen, Volvo, VT iDirect, VTT, w2bi, Wi-Fi Alliance, Wilus, Wind, Xiaomi, Xidian University, Xilinx, Xura, Yaana, Zhejiang University, Zollkriminalamt, ZTE.

Spectrum and Frequency Bands

Spectrum

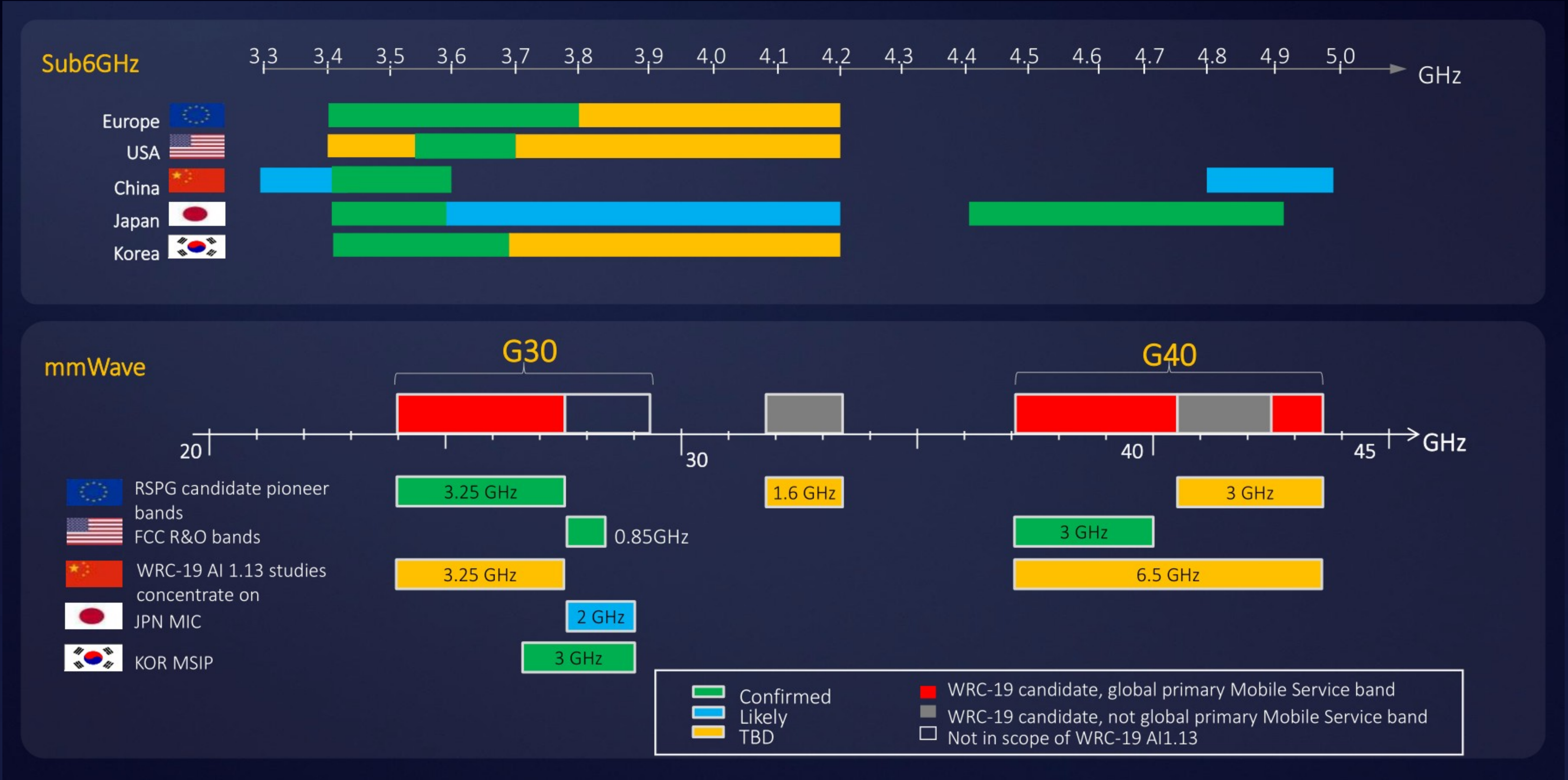


UK Radio Frequency Spectrum Allocation in 2015 - Source: Roke Manor Research

Frequency Bands Considered for 5G

- 700 - 800 MHz band from legacy 2G/3G/4G systems (“re-farming”)
- 3.5 GHz is the primary band for 5G in Europe
- mmWave: 26 GHz

5G Frequency Bands



Path Loss (Attenuation vs. Distance)

In free space:
Friis' transmission equation

$$P_r = G_r \, G_t \left(\frac{c}{4\pi f d}\right)^2 P_t$$

Received power

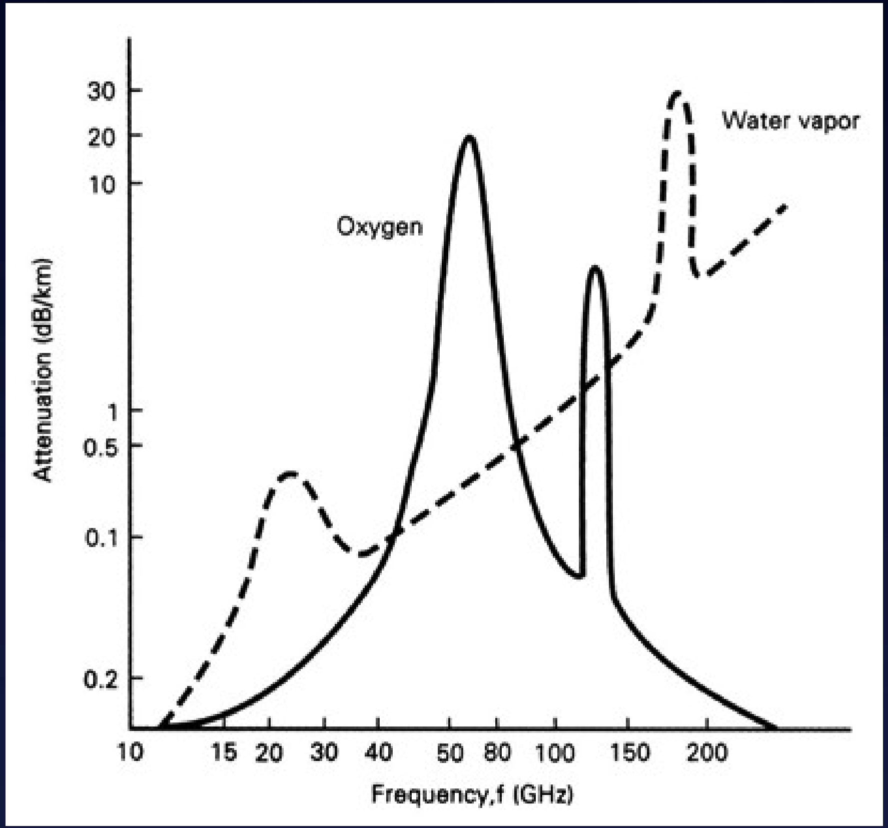
Tx and Rx antenna characteristics

Carrier Frequency

Tx-Rx Distance

Transmitted power

In the atmosphere:
Propagation is affected by gases



Source: US FCC

High Frequency Bands

Why go to higher frequencies?

- There is available spectrum
- Technological barriers disappearing

Cons:

- High penetration loss (walls, foliage)
- High attenuation: low range

Pros:

- Unoccupied spectrum!
- High attenuation: less interference

Frequency Bands in 5G

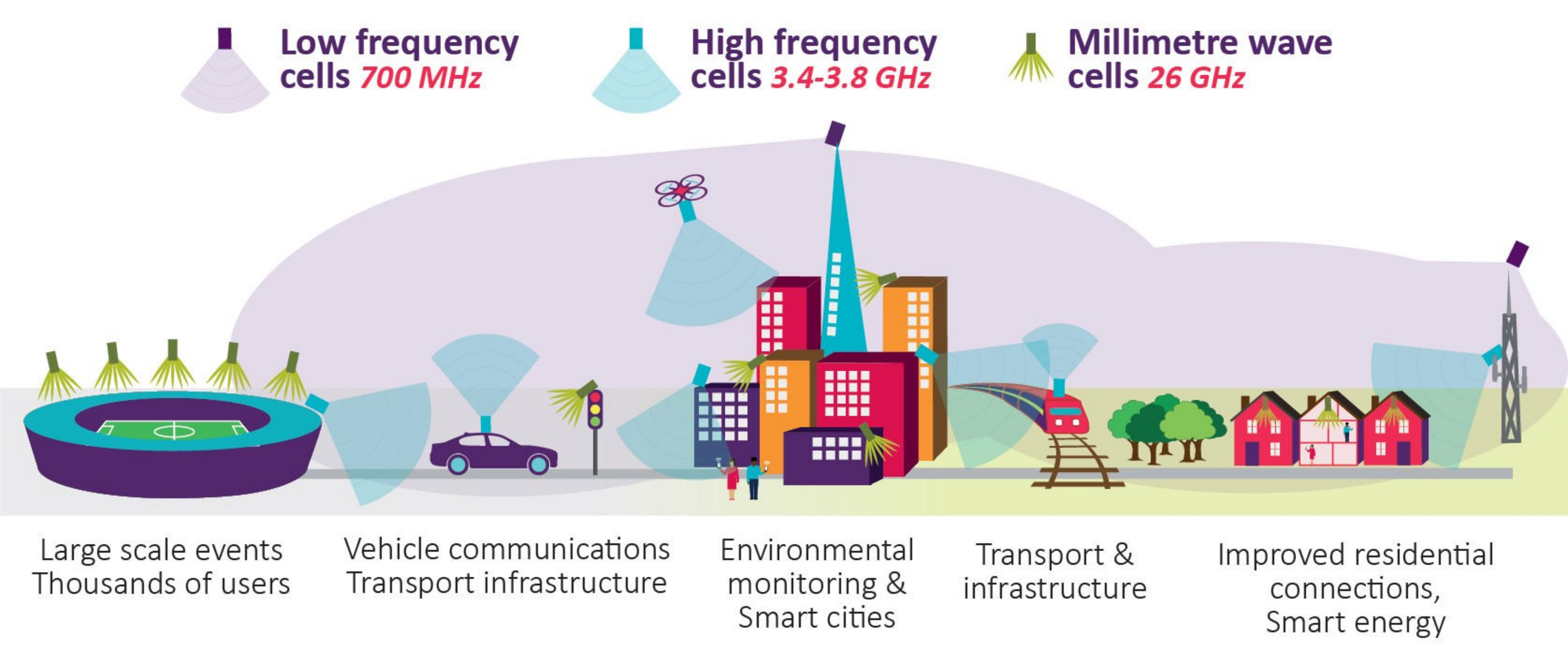


Illustration: OFCOM (UK)

French Spectrum Auctions for 5G (2020)

	Orange	SFR	Bouygues	Free Mobile
Nombre de blocs de 50 MHz	1	1	1	1
Dépense pour le bloc de 50 MHz	350 millions d'euros	350 millions d'euros	350 millions d'euros	350 millions d'euros
Nombre de blocs de 10 MHz	4	3	2	2
Dépense pour les blocs de 10 MHz	504 millions d'euros	378 millions d'euros	252 millions d'euros	252 millions d'euros
Nombre total de fréquences	90 MHz	80 MHz	70 MHz	70 MHz
Dépense totale lors des enchères	854 millions d'euros	728 millions d'euros	602 millions	602 millions

Total: €2.8 Bn

The licenses will run until 2035

Source: Numerama, https://www.numerama.com/tech/651580-si-vous-navez-rien-compris-aux-encheres-5g-qui-demarrent-en-france.html

Italy Spectrum Auction for 5G (2018)

Band	Winning Bidder	Price (€m)	Spectrum Won
700MHz	Vodafone	683.2	20MHz
	TIM	680.2	20MHz
	Iliad	676.5	20MHz
3.7GHz	TIM	1694.0	80MHz
	Vodafone	1685.0	80MHz
	Wind	483.9	20MHz
	Iliad	483.9	20MHz
26GHz	TIM	33.0	200MHz
	Iliad	32.9	200MHz
	Vodafone	32.6	200MHz
	Wind	32.6	200MHz
	Fastweb	32.6	200MHz

Grand Total: €6.5 Bn

Source: Pete Bell, TeleGeography.
https://blog.telegeography.com/italian-5g-auction-causes-concern

The licenses will run until 2037.

A Generic Communications System

Physical layer

Information Source → Source Encoder → bits → Channel Encoder → analog signals → Degrading channel → analog signals → Channel Decoder → bits → Source Decoder → Reconstructed Information

- Sound
- Image
- Video
- HTML
- …

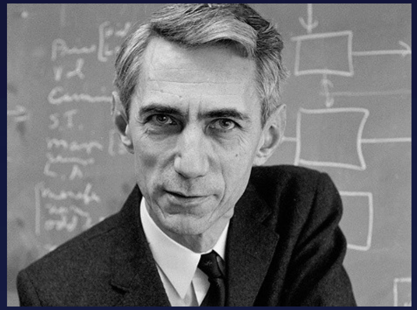
- MP3
- JPEG
- MP4
- Zip
- …

- EM wave propagation
- Analog storage
- Noise
- Noise & interference

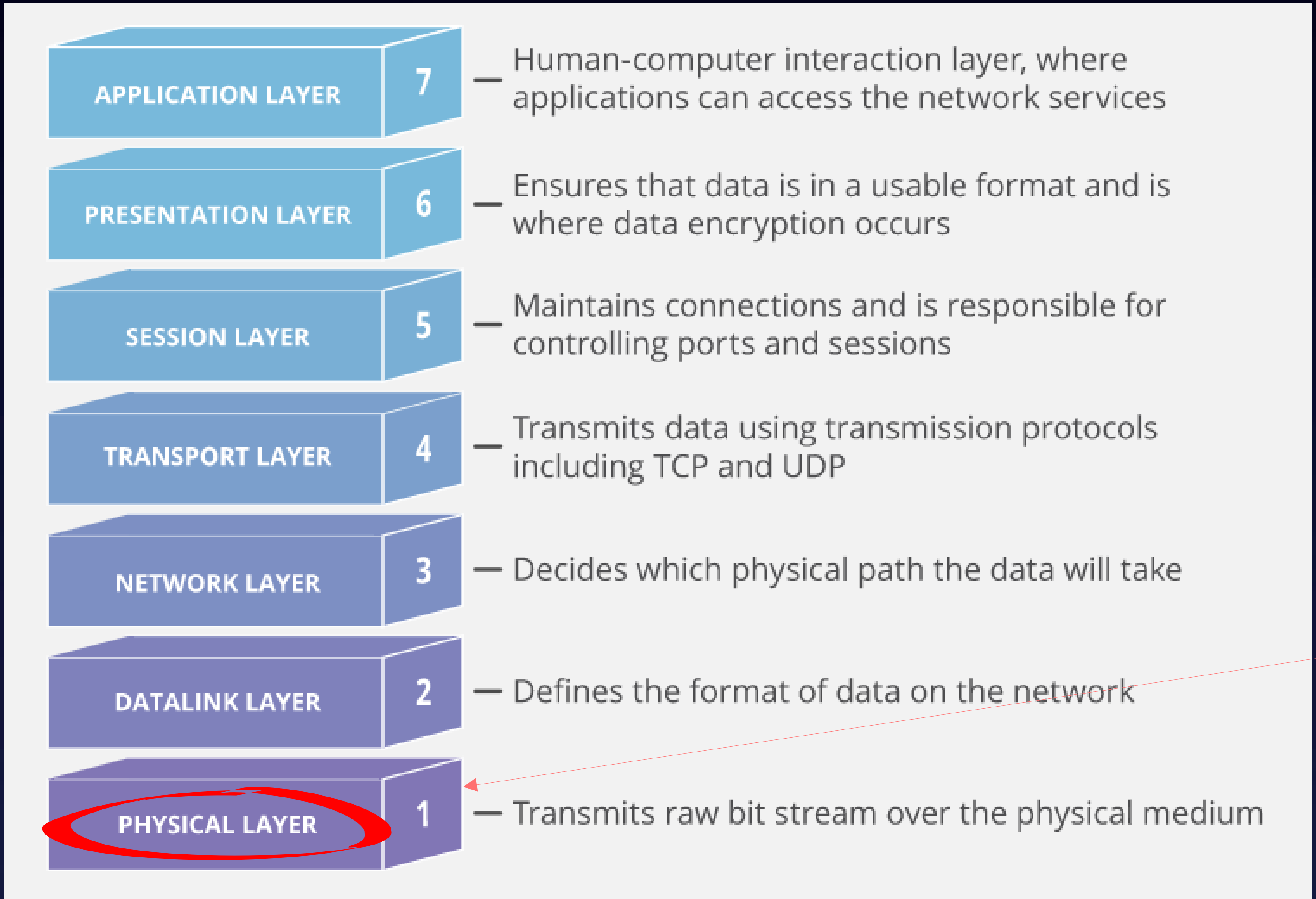
Shannon's Separation Theorem (1948):

Separation between souce and channel coding is optimal

(for long messages, stationary sources)



“Open Systems Interconnection” Layers

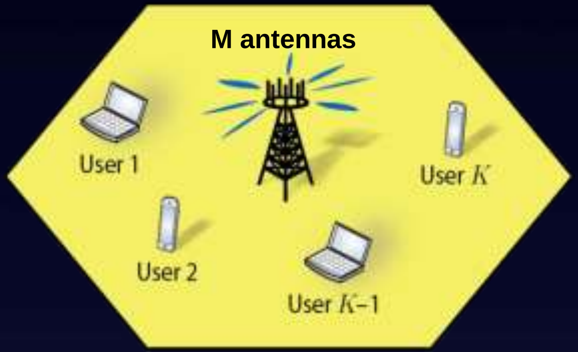


What comes next!

Source: Cloudflare

Massive MIMO

What is Massive MIMO?



- In theory, extension of MIMO (Multiple-Input Multiple Output)
- With some key differences:
  - More BTS antennas (M) than users (K): $\frac{M}{K} \gg 1$
  - Simplified multi-user processing, link adaptation, scheduling
  - Revisited Channel State Information (CSI) acquisition
- **Consider the downlink of a multi-user channel.** BTS has M antennas. Each user has 1 antenna.
- Channel of user i: $\boldsymbol{h}_i$ is a M-dimensional vector
- Consider jointly the (downlink) channels to all users: $\mathbf{H} = [\mathbf{h_1}\ , \dots , \mathbf{h}_K]^{\ T}$.
- Downlink transmission: $\begin{bmatrix} y_1 \\ \vdots \\ y_K \end{bmatrix} = \mathbf{H}\,\mathbf{x}$ (+ noise). $\mathbf{x}$ is the M-dimensional transmitted signal

Massive MIMO:
How did it all start?

3590 IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, VOL. 9, NO. 11, NOVEMBER 2010

Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas

Thomas L. Marzetta

Abstract—A cellular base station serves a multiplicity of single-antenna terminals over the same time-frequency interval. Time-division duplex operation combined with reverse-link pilots enables the base station to estimate the reciprocal forward- and reverse-link channels. The conjugate-transpose of the channel estimates are used as a linear precoder and combiner respectively on the forward and reverse links. Propagation, unknown to both

point-to-point system, but are retained in the multi-user system provided the angular separation of the terminals exceeds the Rayleigh resolution of the array.

Channel-state information (CSI) plays a key role in a multi-user MIMO system. Forward-link data transmission requires that the base station know the forward channel, and reverse-

- **Intuition:** DL linear precoding consisting in superposition coding with simple matched precoder

$$\mathbf{x} = \frac{1}{M}\sum_{i=1\ldots K}\mathbf{h}_i s_i \quad \text{where } s_i \text{ is the data symbols for user i}$$

- At user j: $y_j = \mathbf{h}_j^T \frac{1}{M}\sum_{i=1\ldots K}\mathbf{h}_i s_i = \underbrace{\frac{1}{M}\mathbf{h}_j^T\mathbf{h}_j s_j}_{\text{Signal of interest}} + \underbrace{\frac{1}{M}\sum_{i\neq j}\mathbf{h}_j^T\mathbf{h}_i s_i}_{\text{Interference}} \quad (+\text{ noise})$

- With iid unit-variance fading: $\lim_{M\to\infty}\frac{1}{M}\sum_{i\neq j}\mathbf{h}_j^T\mathbf{h}_i = 0$ while $\lim_{M\to\infty}\frac{1}{M}\mathbf{h}_j^T\mathbf{h}_j = 1$

SINR $\xrightarrow[M\to\infty]{} \infty$

(without Tx power increase)

No fading
on the effective channel: channel hardening

CSI in Massive MIMO Downlink Transmission

- Downlink transmission: $\begin{bmatrix} y_1 \\ \vdots \\ y_K \end{bmatrix} = \mathrm{H}\,\mathrm{x}$ (+ noise). x is the M-dimensional transmitted signal
- MRC Precoding: matched precoder: $\mathrm{x} = \frac{1}{\mathrm{M}} \mathrm{H}^{\mathrm{T}}\,\mathrm{s}$ where $\mathrm{s} = \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix}$ are user symbols

$$\lim_{\mathbf{M}\to\infty} \frac{\mathbf{1}}{\mathbf{M}} \mathbf{H}\mathbf{H}^{\mathbf{T}} = \mathbf{I}_{\mathbf{K}} \quad \Rightarrow \quad \begin{bmatrix} y_1 \\ \vdots \\ y_K \end{bmatrix} = \begin{bmatrix} s_1 \\ \vdots \\ s_K \end{bmatrix}$$

Downlink Channel

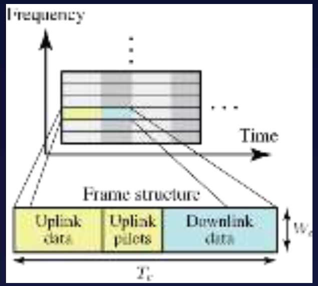
Precoder: must be estimated

Efficient CSI Estimation Method Needed!

CSI Acquisition Strategies

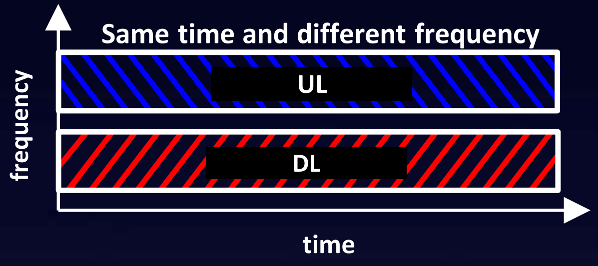
Time Division Duplex (TDD)

- UL channel estimation **at the BTS** based on pilot sequences sent by the UEs



- DL channel obtained by electromagnetic **reciprocity** (same as UL channel)

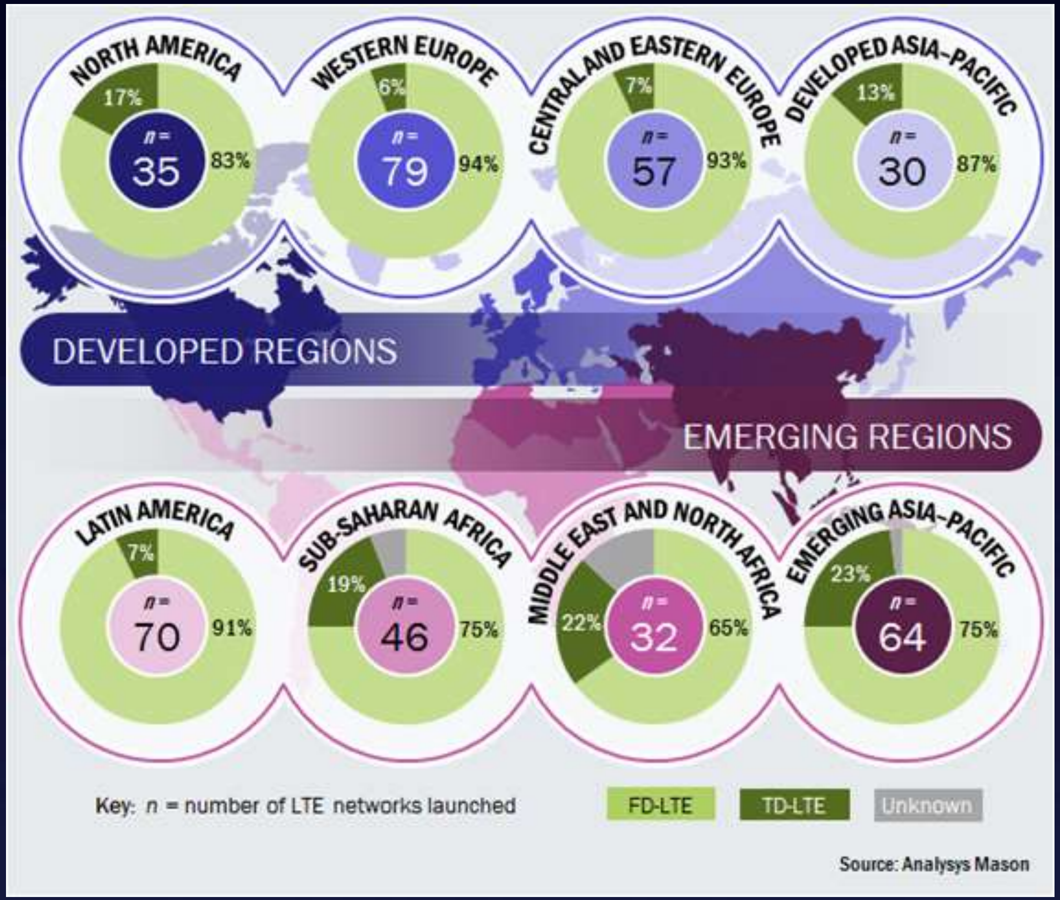
Frequency Division Duplex (FDD)



- DL channel estimation **at the UE** based on pilot sequences
- Feedback (UE -> BTS) of estimated CSI (the data encoding the quantized CSI is transmitted on the uplink channel)

Why FDD Matters

- Depending on the region, **65 to 94% of 4G spectrum is FDD**
- Difficult to change due to regulatory and/or technical (coexistence with other systems) constraints



Uplink Multi-User Channel Estimation

- **Uplink Channel estimation:** users transmit pilot sequences simultaneously:

  $p_i(t)$ is the (known) pilot symbol for user i at time t,

$$\mathbf{y}(\mathrm{t}) = \sum_{i=1\ldots K} \mathbf{h}_i \mathrm{p}_i(t) = \mathbf{H}^{(ul)}\,\mathbf{p}(t)$$

- Length-L training phase in matrix form: $\mathbf{Y} = [\mathbf{y}(1), \ldots, \mathbf{y}(L)]$, $\mathbf{H}^{(ul)} = [\mathbf{h}_1, \ldots, \mathbf{h}_K]$,

  $\mathbf{P} = [\mathbf{p}(1), \ldots, \mathbf{p}(L)]$

- CSI acquisition for all K users:

$$\mathbf{Y} = \mathbf{H}^{(ul)}\,\mathbf{P}\ (+\text{noise})$$

- Intuition: **linear estimation problem** (observation $\mathbf{Y}$ is a linear combination of the estimee $\mathbf{H}^{(ul)}$).
- Trivial solution if $\mathbf{P}$ is invertible: $\mathbf{H}^{(ul)} = \mathbf{Y}\mathbf{P}^{-1}$

Uplink Multi-User Channel Estimation

- CSI acquisition for all K users treated jointly:

$$\underbrace{\mathbf{Y}}_{M\times L} = \underbrace{\mathbf{H}^{(ul)}}_{M\times K} \underbrace{\mathbf{P}}_{K\times L} \quad (+\text{noise}\ldots)$$

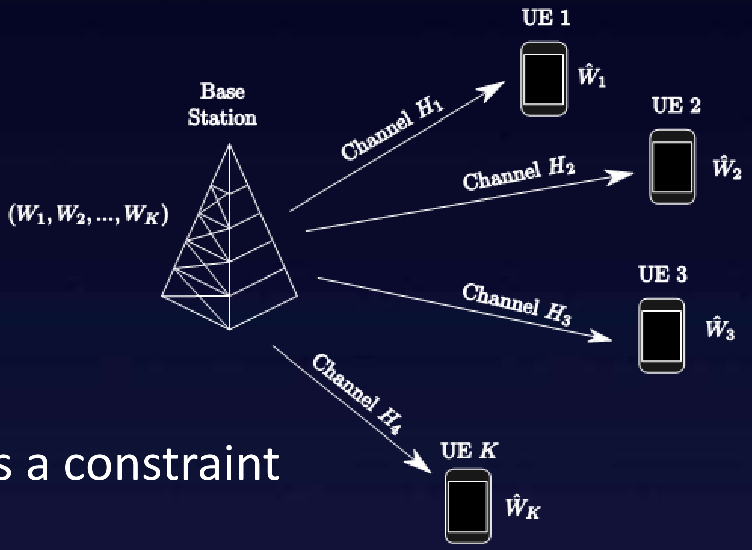
- Pilot design:
  - $\mathbf{P} = I_K$: round-robin CSI estimation across the users
  - $\mathbf{P}\mathbf{P}^H = I_K$: orthogonal pilots across the users (requires $L \geq K$)
  - Non-orthogonal pilots: not a problem as long as rank($\mathbf{P}$)=K
  - rank($\mathbf{P}$) $<$K: H can not be identified (under-determined linear system)
  - Pilot reuse across cells: $\mathbf{P} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_1 \end{bmatrix}$ (pilot contamination)

The properties of the pilot matrix P govern CSI estimation

Precoding and Multiple-Access Strategies

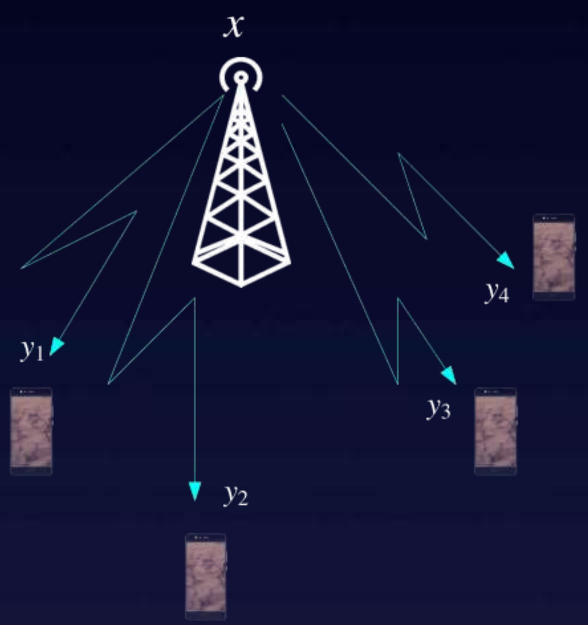
Multi-user Precoding in Massive MIMO (Downlink)

- **Linear precoding:** Zero-Forcing, Regularized ZF, MMSE…
- **Non-linear** approaches are also an option
- Large number of antennas and users, computational complexity is a constraint



Favour Relatively Simple Precoding Approaches

Linear Precoding



$$\mathbf{y} = \mathbf{H}\,\mathbf{x} + \mathbf{n}$$

Linear precoding: $\mathbf{x} = \mathbf{G}\,\mathbf{s}$ where

- $\mathbf{G} = \alpha\,\mathbf{H}^{\dagger}(\mathbf{H}\mathbf{H}^{\dagger})^{-1}$ (zero-forcing)
- $\mathbf{G} = \alpha\,\mathbf{H}^{\dagger}(\mathbf{H}\mathbf{H}^{\dagger} + \frac{\mathbf{1}}{\mathbf{SNR}}\mathbf{I})^{-1}$ (MMSE)

Scalar transmit power normalization

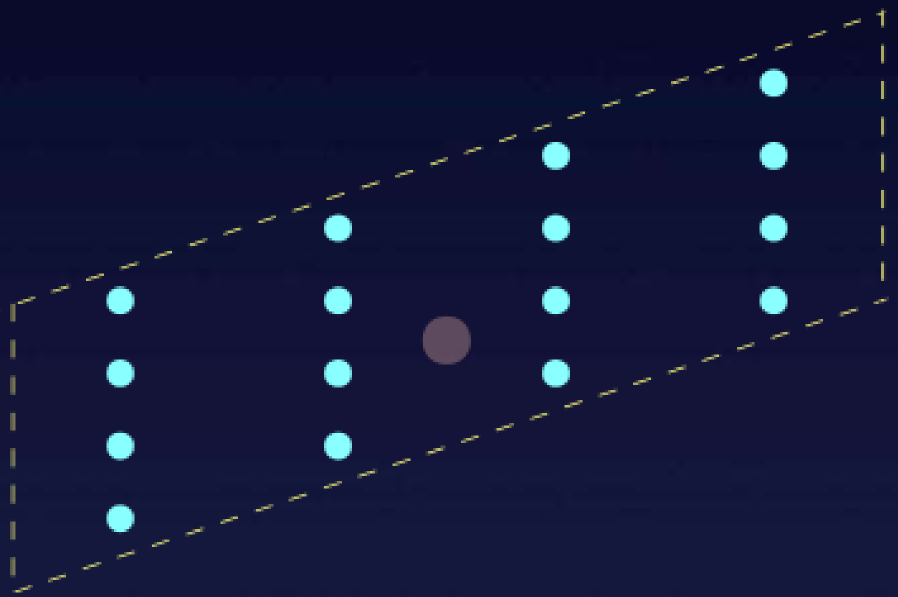
Classical Linear Precoding

Original constellation

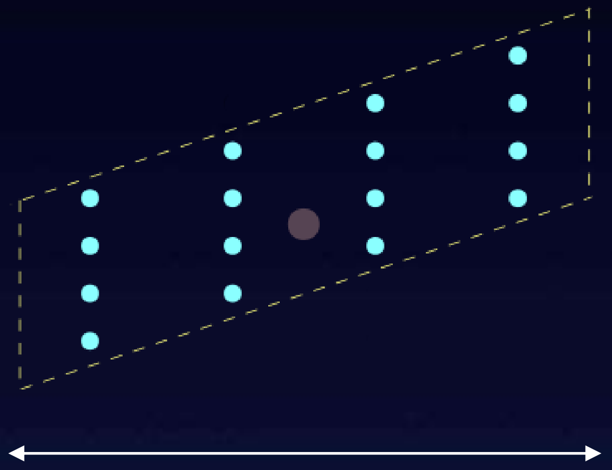
symbols s

Tx Precoded signal

$\mathbf{x} = \mathrm{G}\, s$



Linear Precoding Does Not Scale



- **Large Tx power required for one dimension.**
- **The power normalization coefficient α gets small, it would "kill" the other dimensions.**
- **This phenomenon gets works in large dimension**

Simple analysis of why this approach fails in large dimension (from "A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiuser Communication," Hochwald, Peel, Swindlehurst)

Assume

- equal number of Tx and Rx antennas (K antennas at the Tx, K single-antennas Rx)
- Rayleigh-fading unit variance iid channels
- Unit-variance complex Gaussian iid symbols s
- Simple zero-forcing where $\mathrm{x} = \mathrm{H}^{-1}\,\mathrm{s}$

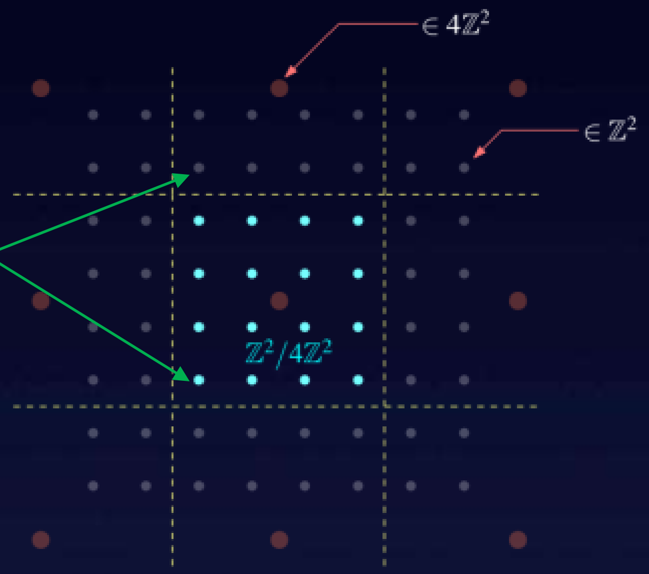
Then:

- $\mathrm{E}_{\mathrm{s,H}}[||\mathrm{x}||^2] = +\infty$ for Gaussian i.i.d. s and H
- The sum-capacity for large K is $\lim_{K\to\infty} C_{sum} = SNR \cdot \log e$
- Sum-capacity does not increase with K!

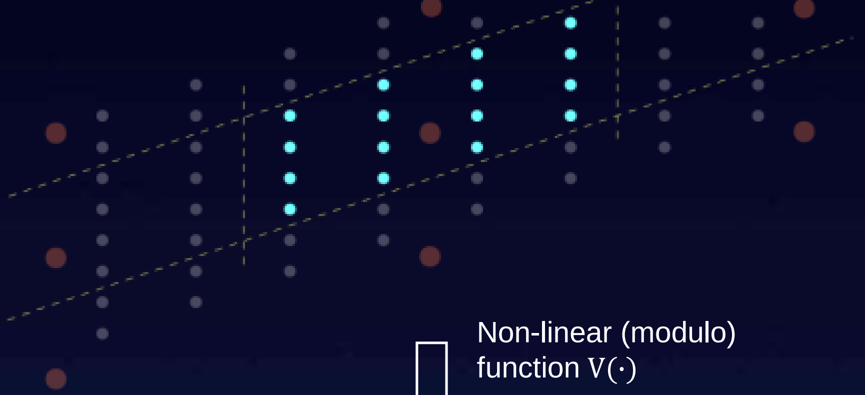
Non-linear Precoding: Vector Perturbation

Original constellation of $\mathbf{s}$

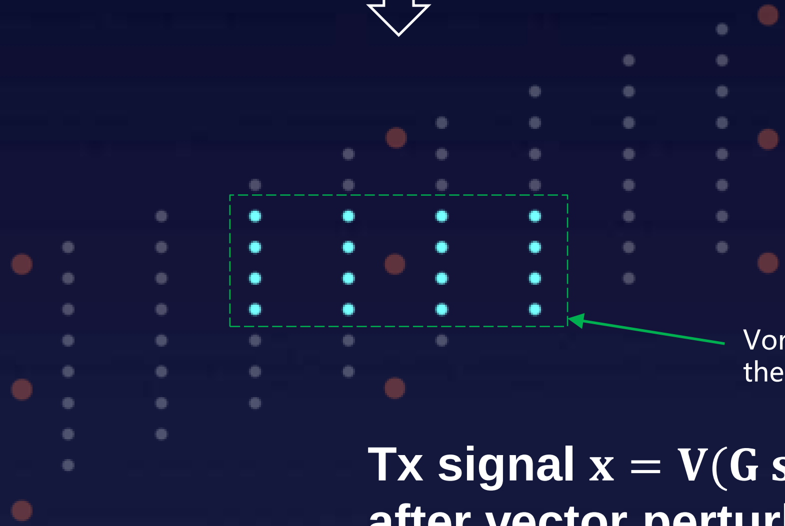
Points belonging to the same coset represent the same information



Linearly precoded signals $\mathrm{G}\,\mathbf{s}$



Non-linear (modulo) function $\mathrm{V}(\cdot)$

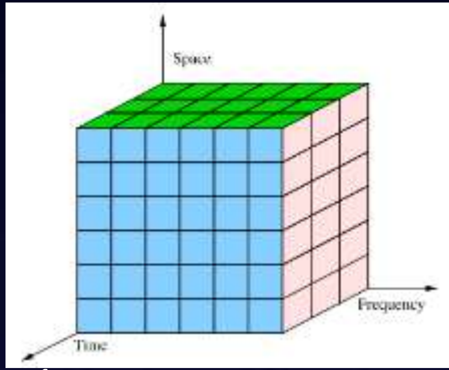


Tx signal $\mathbf{x} = \mathbf{V}(\mathbf{G}\,\mathbf{s})$ after vector perturbation

The receiver interprets the constellation modulo a coarse (red) lattice

Multiple Access (Uplink)

- Many users, many BTS antennas, large bandwidth… calls for an **efficient multiple-access scheme**



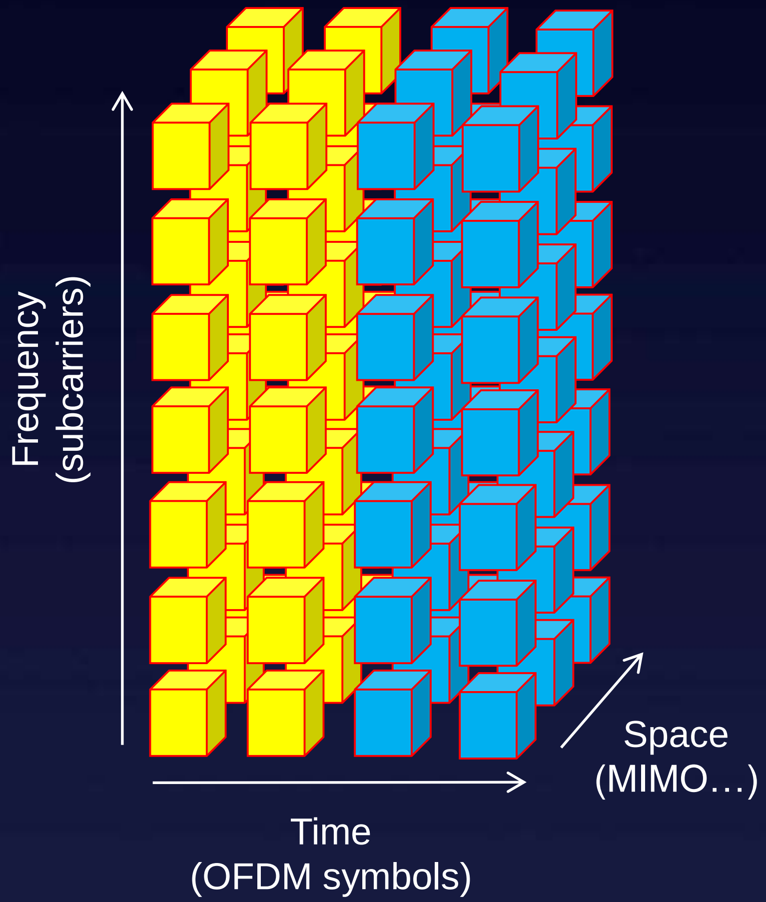
- Classically (up to 4G): orthogonal resources (in time/frequency/space) allocated by the BTS
  - Allows for simple receiver architecture (no multi-user decoding)

- Recent trend towards **overloaded (non-orthogonal) multiple-access:** more than 1 user per resource element
  - **More degrees of freedom in the resource allocation (e.g. several low-rate users sharing a resource)**

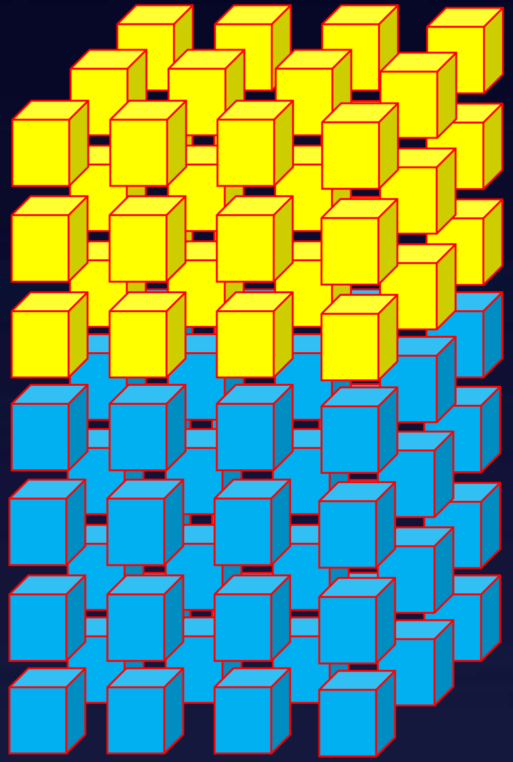
- **Grant-free** access schemes: in low-rate applications (sensors, IoT…) the overhead of signaling to request an exclusive channel grant and CSI acquisition is impractical
  - **New grant-free access schemes allowing true random access**

Orthogonal Multiple Access Schemes

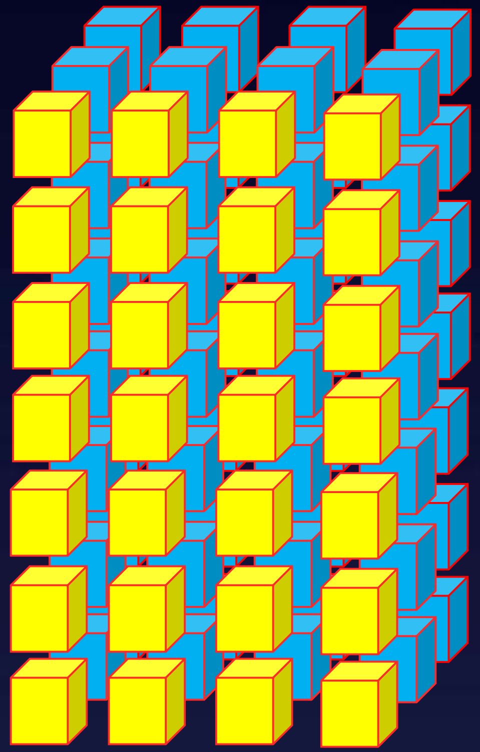
Time Division Multiple Access - TDMA



Orthogonal Frequency Division Multiple Access -OFDMA

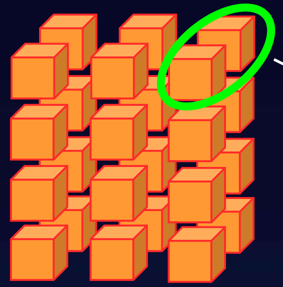


Space-Division Multiple Access - SDMA



Orthogonal Multiple Access with SDMA

- The signal received from two users are mixed in the spatial domain:

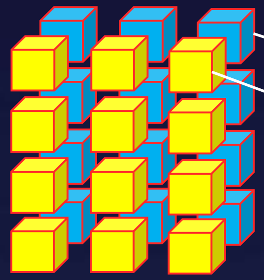


Symbols transmitted by users 1 and 2 on
a given time-frequency resource

$$\underline{y} = \underline{h}_1 s_1 + \underline{h}_2 s_2 = \begin{bmatrix} \underline{h}_1 & \underline{h}_2 \end{bmatrix} \cdot \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

(vector) channels of users 1
and 2 to the BTS

- An equalizer matched to the MIMO channel: (without noise)



$$\begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} = \begin{bmatrix} \underline{h}_1 & \underline{h}_2 \end{bmatrix}^{-1} \underline{y}$$

Only works if the channel matrix is (pseudo-) invertible: more antennas than users

Non-Orthogonal Multiple Access

- Consider a single resource element (r.e.)
- On which multiple signals are transmitted:

$$\longrightarrow \quad y = h_1 s_1 + h_2 s_2 = [h_1 \ \ h_2] \cdot \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

(scalar) channels of
users 1 and 2

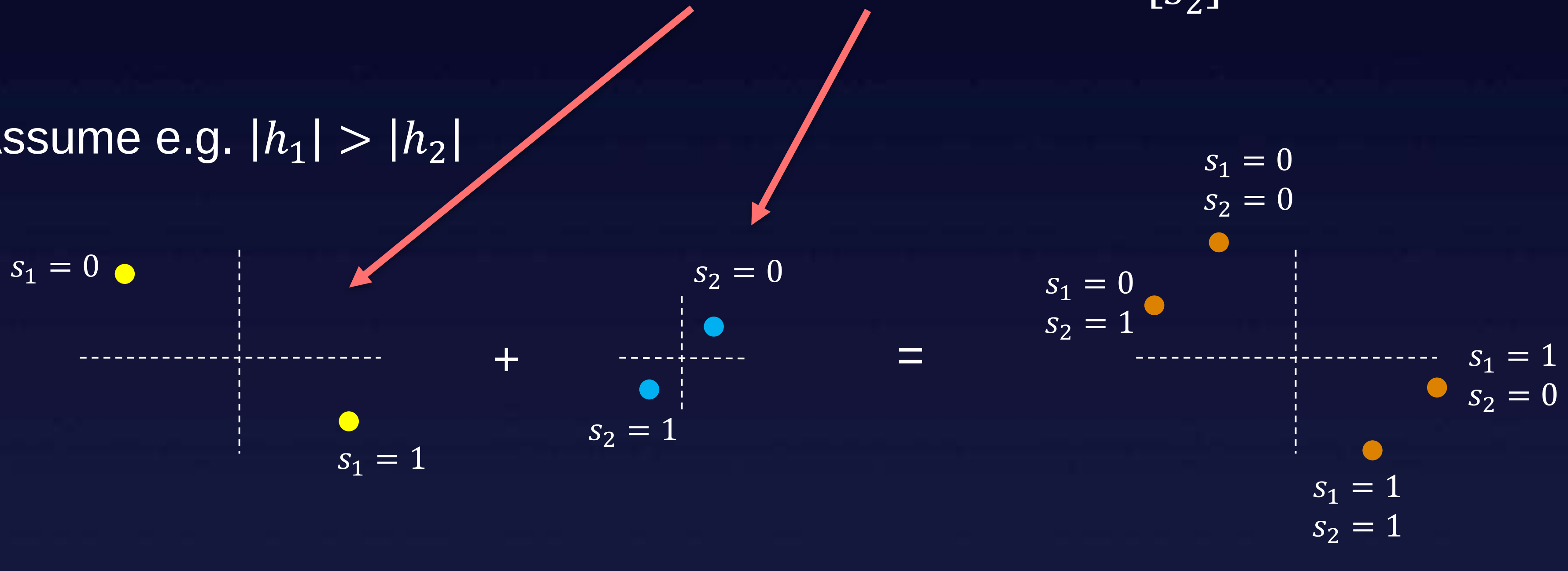
- Can we recover s_1 and s_2?

Non-Orthogonal Multiple Access (NOMA)

- Assume BPSK constellations for s_1 and s_2:

$$y = h_1 s_1 + h_2 s_2 = [h_1 \;\; h_2] \cdot \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}$$

- Assume e.g. $|h_1| > |h_2|$



Non-Orthogonal Multiple Access (NOMA) Properties

- In general, NOMA consists in having K users transmitting simultaneously over **N<K** (time, frequency, spatial) resource elements: allows **overload**

- **Flexibility**: we are not bound to allocating integer multiples of the numbers of resource elements to each user:

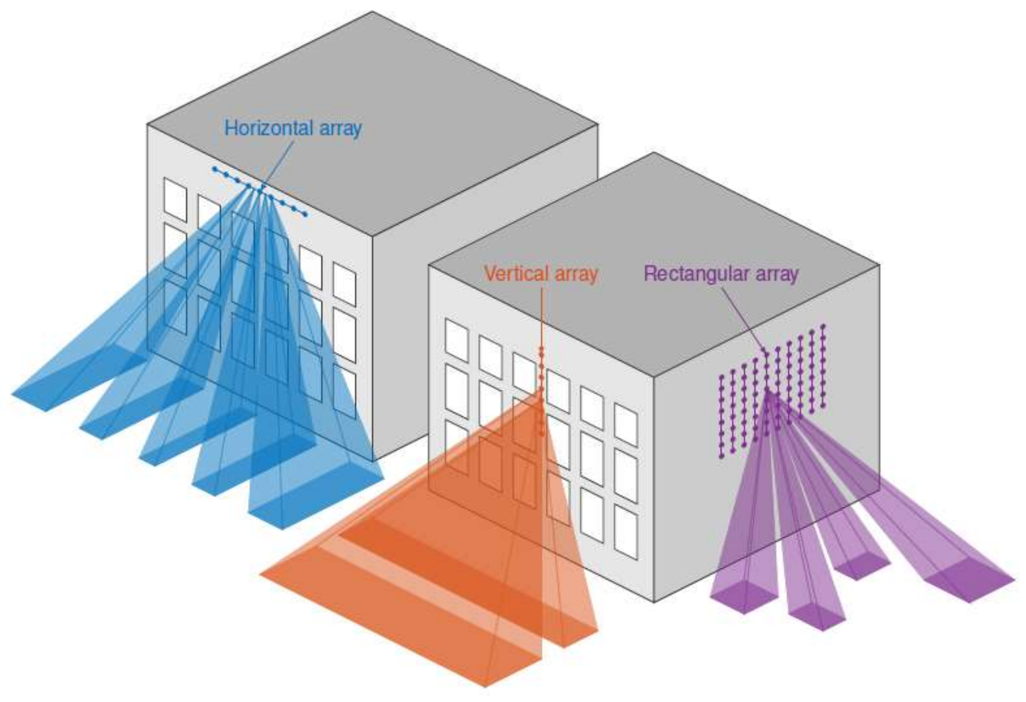
  N resource elements shared among K users → $\frac{N}{K}$ r.e. per user

- **More complex decoding**: linear equalization does not work, need to to multi-user decoding using more complex message-passing algorithms

Massive MIMO Channel and Hardware Specificities

Massive MIMO Antenna Array Geometries

- Centralized M-MIMO arrays
  - Uniform linear (horizontal/vertical)
  - Rectangular (3D)
  - Cylindrical



Large Aperture Arrays

- **Sparse channel models**

  Signal strength from one user is not uniform over the array

- **Spherical wave fronts** (instead of planar)

  The far-field assumption does not hold

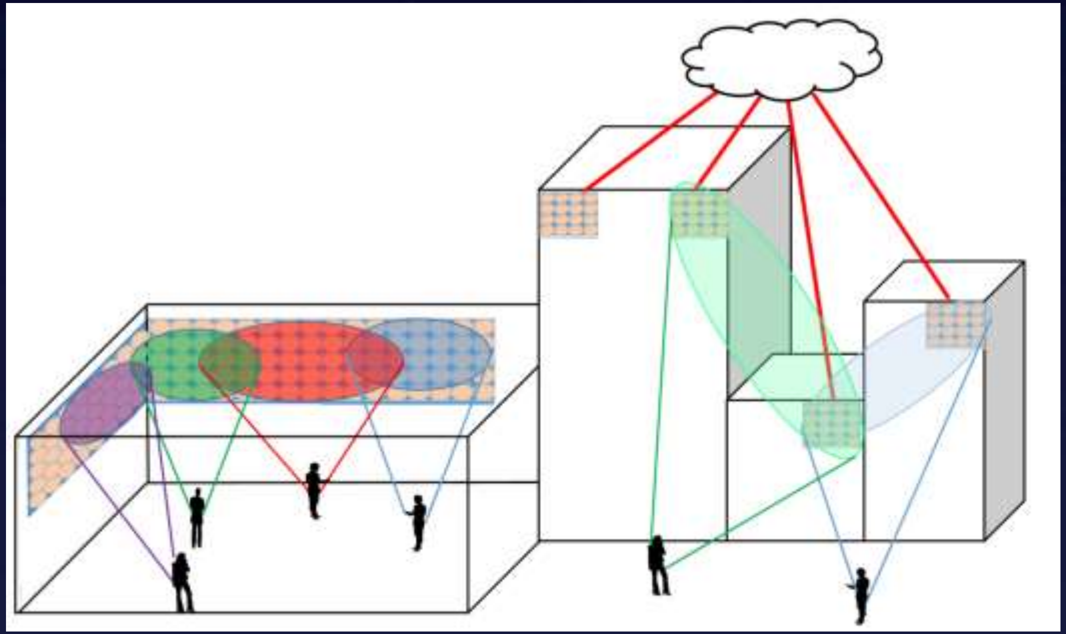


Illustration from “Non-Stationarities in Extra-Large Scale Massive MIMO,” De Carvalho, Ali, Amiri, Angjelichinoski, Heath, 2019.

Channel Model Evolutions

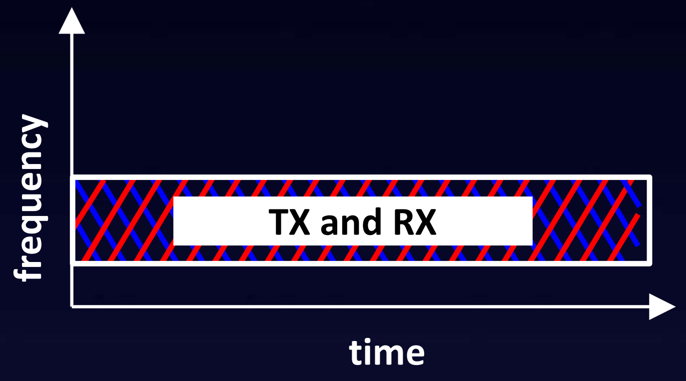
New features in channel models:

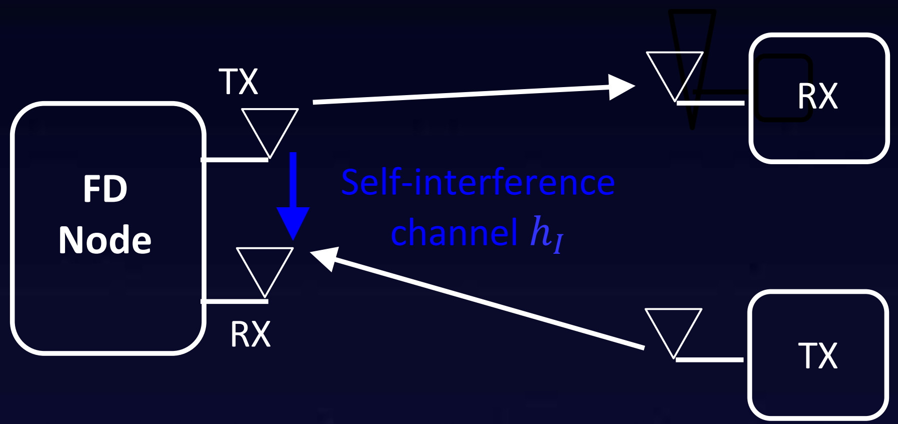
- Spatial dimension (joint modeling of the covariance of several users!)
- New (large rectangular/linear) antenna array geometries
- Spherical wave fronts (instead of planar waves)

New (higher!) frequencies:

- higher pathloss, shorter reach
- line-of-sight becoming more crucial
- more sensitive to Doppler effect

Self-Interference & Full Duplex





- **Self-interference (I) >> Signal of interest (S).**
  - I/S >> 80 dB (85 to 110dB SI cancellation can be achieved)

- **Solution in theory:**
  - The interference is known.
  - This is a strong interference regime.
  - Estimate channel h_I.
  - Cancel: Subtract from the received signal.

Application to Massive MIMO
is technologically complex

Massive MIMO in 5G

Blue Danube

Ericsson radio stripes

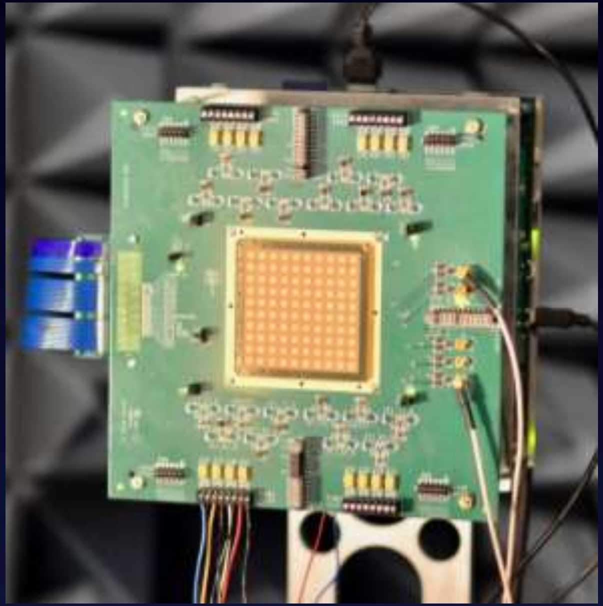
Millimeter Wave (mmWave)

Communicating over mmWave Bands

Technological constraints due to the use of higher frequencies:

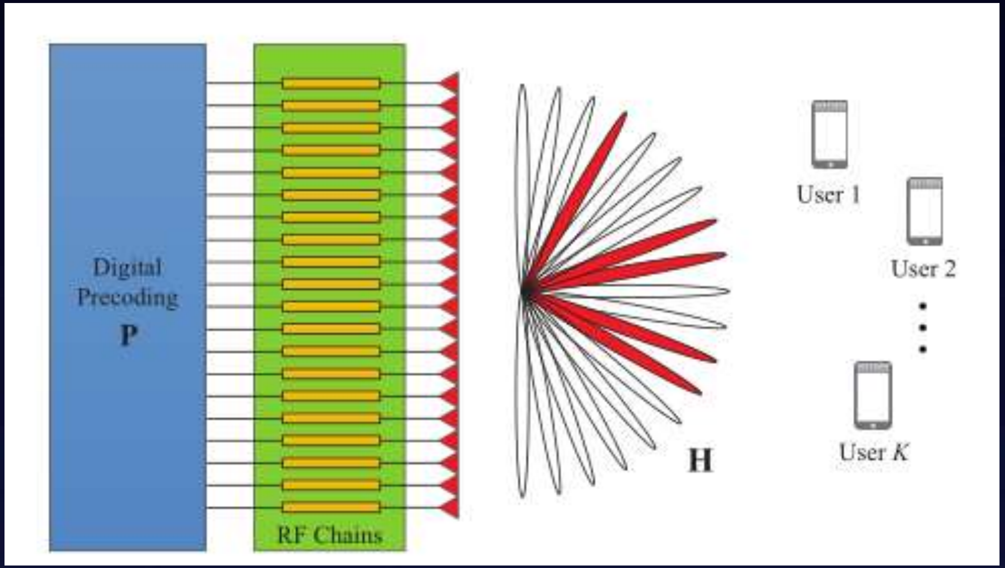
- **Smaller antenna elements** (proportionate to the wavelength)
- **Lower power** transmitted per element

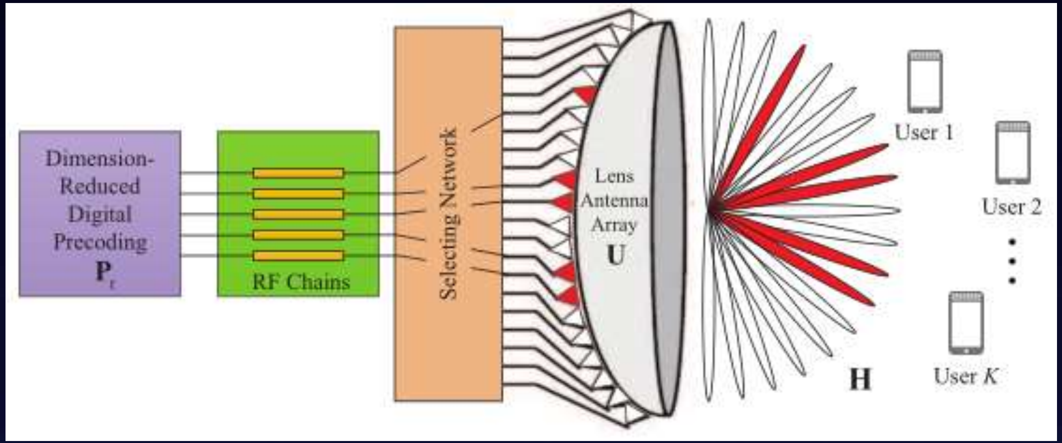
mmWave requires to coherently combine signals from many elements (array gain) to achieve sufficient range



28GHz 64 dual-polarized elements phased array.
Total dimensions: 2.8” x 2.8”
Source: IBM-Ericsson research Zurich,
https://www.flickr.com/photos/ibm_research_zurich

mmWave RF Architectures (I)





Fully digital architecture:

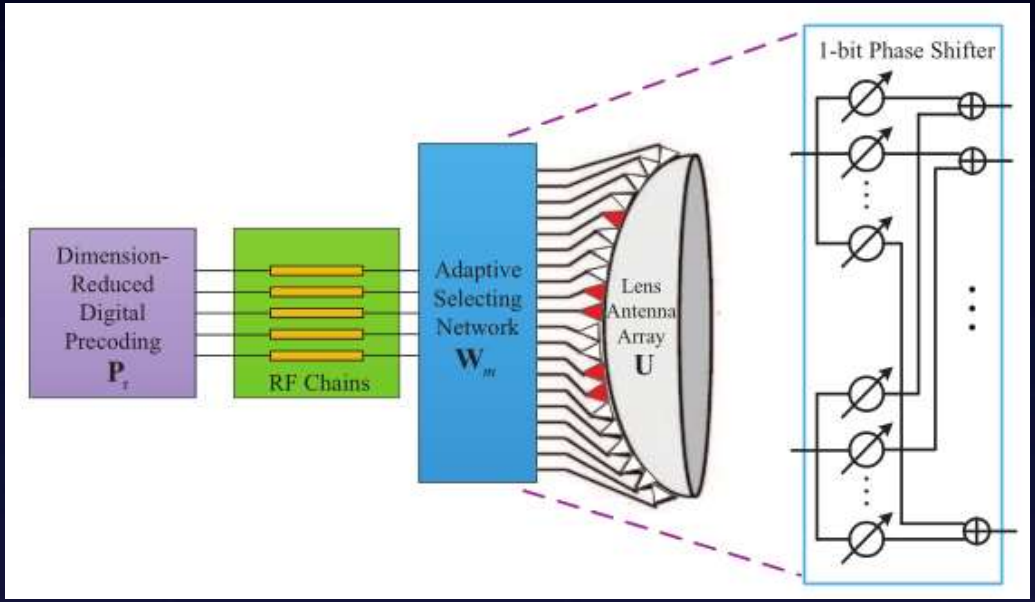
- Similar in spirit to Massive MIMO
- Allows for multi-user multiplexing
- **Impractical to have so many RF chains**

Lens antenna:

- Fixed beams defined by the geometry of the elements and the RF lens
- Needs a simple RF selecting network

Illustrations: “Beamspace Channel Estimation for Millimeter-Wave Massive MIMO Systems with Lens Antenna Array,” Dai, Gao, Han, I, Wang, ICCC 2016.

mmWave RF Architecture (II)



Hybrid (analog + digital approaches):

- Analog selecting network can only assume a subset of the possible precoders, using **phase-shifters**
- A digital beamformer of reduced dimension

mmWave: Analog and Digital Processing

Complex processing required for CSI acquisition

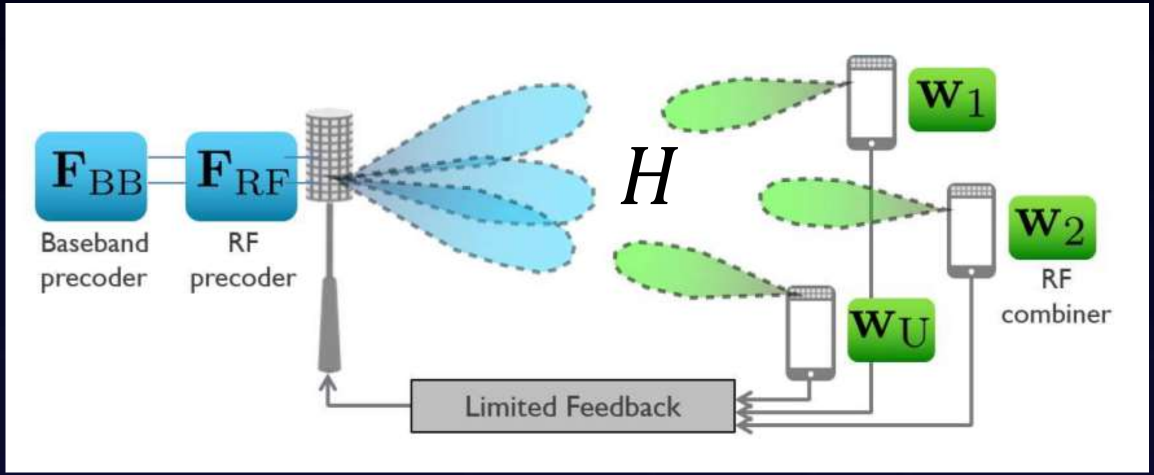


Illustration: "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," Heath, Gonzalez-Prelcic, Rangan, Roh, Sayeed.

- Channel H considering all Tx and Rx elements is a high-dimensional matrix (possibly of low rank)
- We can only measure $F_{RF} \cdot H \cdot W_i$, i.e. the channel can be measured only in the direction(s) pointed by the analog precoders. Need to send pilots again if we want to change F_{RF} or W_i
- Beams are typically very narrow (3 to 15 degrees beamwidth in 2 dimensions) – many directions to try to get a full picture of the channel state

New Concepts

(considered, and to some extent implemented)

in 5G

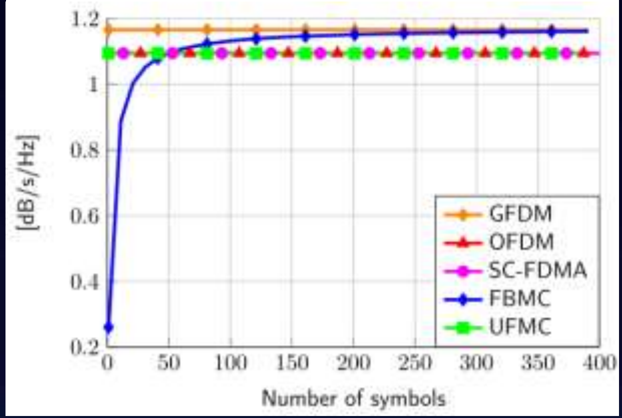
Waveforms

Waveform	Description	Pros	Cons
Cyclic-Prefix Orthogonal Frequency Division Multiplexing (CP-OFDM)	Classical OFDM with cyclic prefix	Low complexity (FFT)	Poor frequency localization, poor peak-to-average power ratio
Zero Padding OFDM (ZP-OFDM)	CP is replaced by zeros	Better resilience to "notches" in channel frequency response	Complex equalization
Filterbank Multi-carrier (FBMC)	Each subcarrier filtered individually	Spectrally efficient	Very narrowband digital filters induce design problems
Universal filtered multi-carrier (UFMC)	Filters applied to groups of subcarriers, no cyclic prefix		Interference from multipath, high complexity
Generalized Frequency Division Multiplexing (GFDM)	2D filtering (time+frequency)	Good PAPR and frequency localization	Sacrifice carrier orthogonality
Filtered OFDM	Subbands can have different waveform and numerologies (symbol rate, FFT size…)	Flexibility, spectral efficiency	Complexity

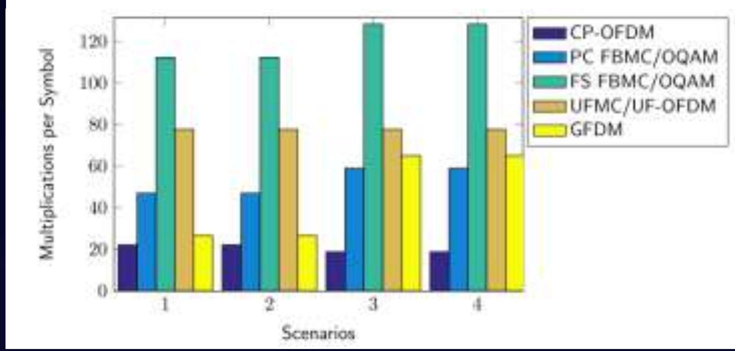
Waveforms: A Difficult Trade-off

Source: **"The 5G candidate waveform race: a comparison of complexity and performance,"** Gerzaguet et al. EURASIP Journal on Wireless Communications and Networking (2017) 2017:13

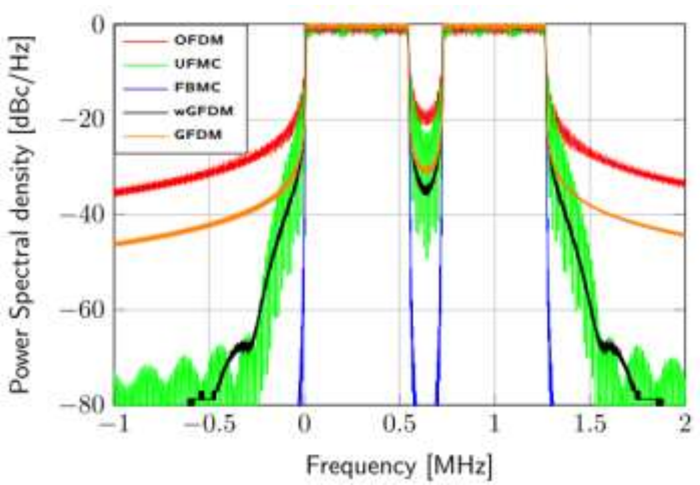
Spectral efficiency



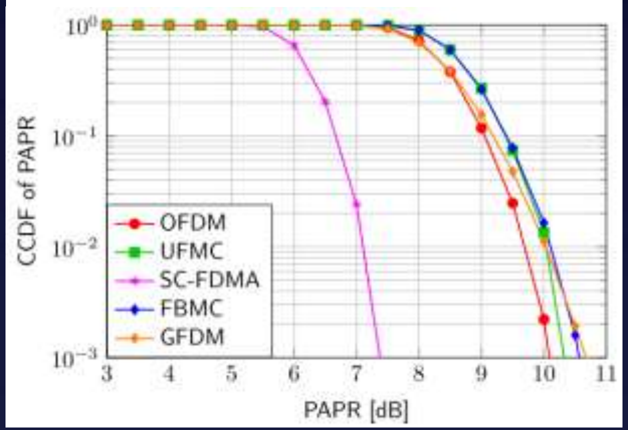
Complexity



Frequency Localization



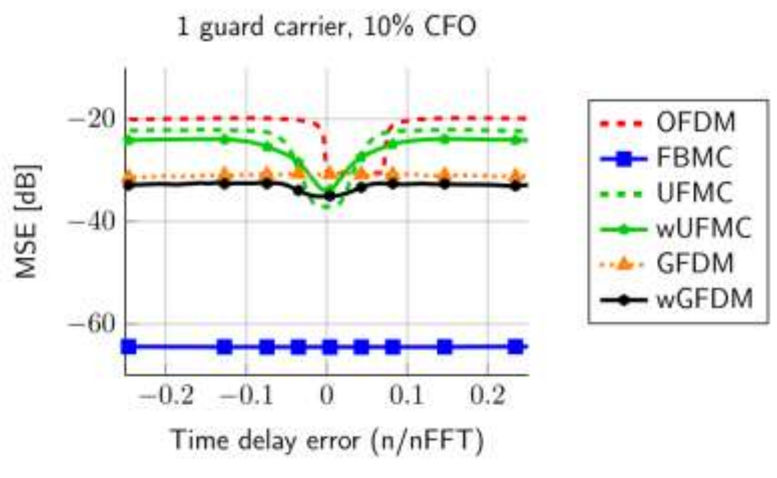
Peak-to-Average Power Ratio



And more questions:

- Oscillator phase noise
- Doppler
- …

Robustness to Carrier frequency offset & multi-user synchronization



Decision:

CP-OFDM will be used for both uplink and downlink!

Cloud (or Centralized) Radio Access Network (C-RAN)

RF signals are brought back to a **single pool of processors**, after downmixing and quantization

- Centralized baseband processing enables to better deal with interference
- Economies of scale
- Power efficiency: baseband units (BBUs) can be turned on/off dynamically to adjust to the number of active users

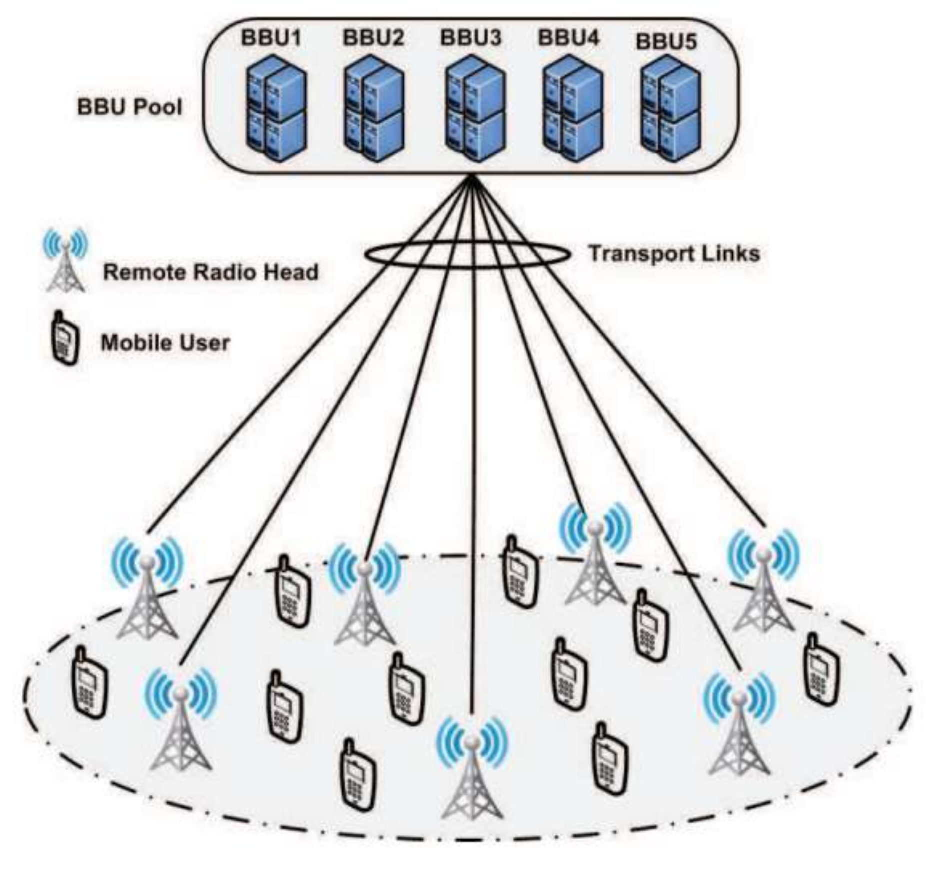


Illustration: "Group Sparse Beamforming for Green Cloud-RAN," Shi, Zhang, Letaief, 2013.

Cloud RAN, Fog-RAN

Recent trends in RAN architecture:

- **Edge Computing**: Optimally distribute the processing between what can be processed locally and what must be processed centrally (interference between users)
- Optimally distribute **data caches** (think videos…) near where it will be consumed

CLOUD

FOG

Centralized

Distributed

Power efficient

Low Fronthaul Requirements

Economies of scale

Spectrally efficient

Low Latency

Cumulus cloud. (Glg / Wikipedia / CC BY-SA 2.0 DE)

CLOUD

View from Blassenstein mountain near Scheibbs, Austria. By Uoaei1. CC-BY-SA 3.0.

FOG

User-Centric, No Cell (UCNC), or *cell-free* networks

Move away from the concept of “each user is connected to one cell”

- On the network side, signals are processed across several *transmission points (TRPs)*
- The mobile can simultaneously communicate with different TRPs, possibly over several frequency bands

Control and User Plane Separation (CUPS)

Use mmWave (when available) for data:
- Cost-effective high-throughput

Use low (sub-6G) frequencies for the control plane:
- Ensure ubiquitous coverage
- Reliable control link to ensure handover and speed up HF beam search

Practical consequences:
- Mobiles will have to support several bands, simultaneously
- Synchronization across bands is difficult

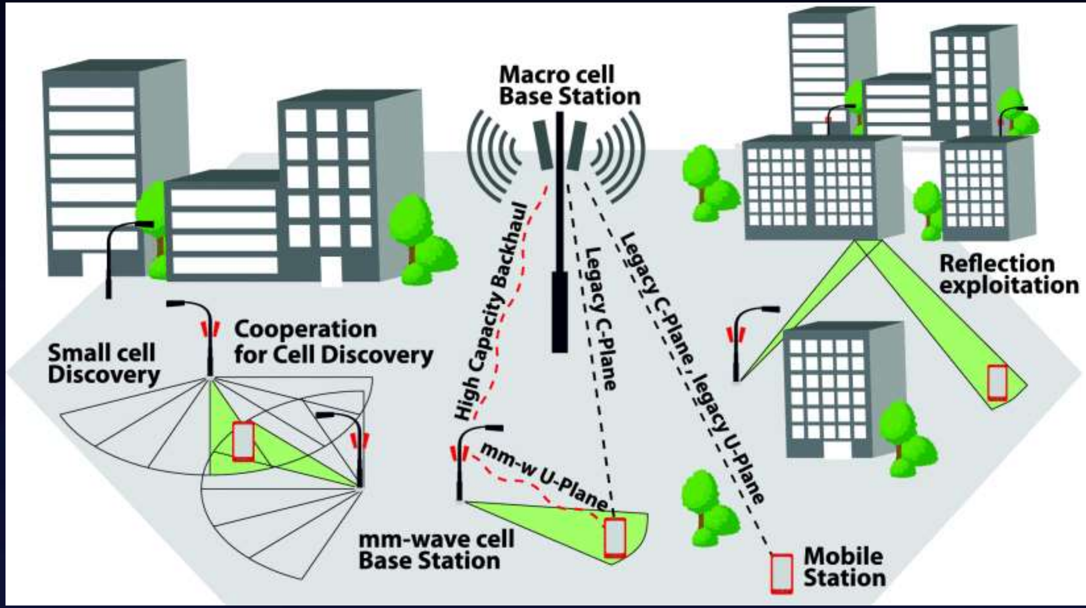
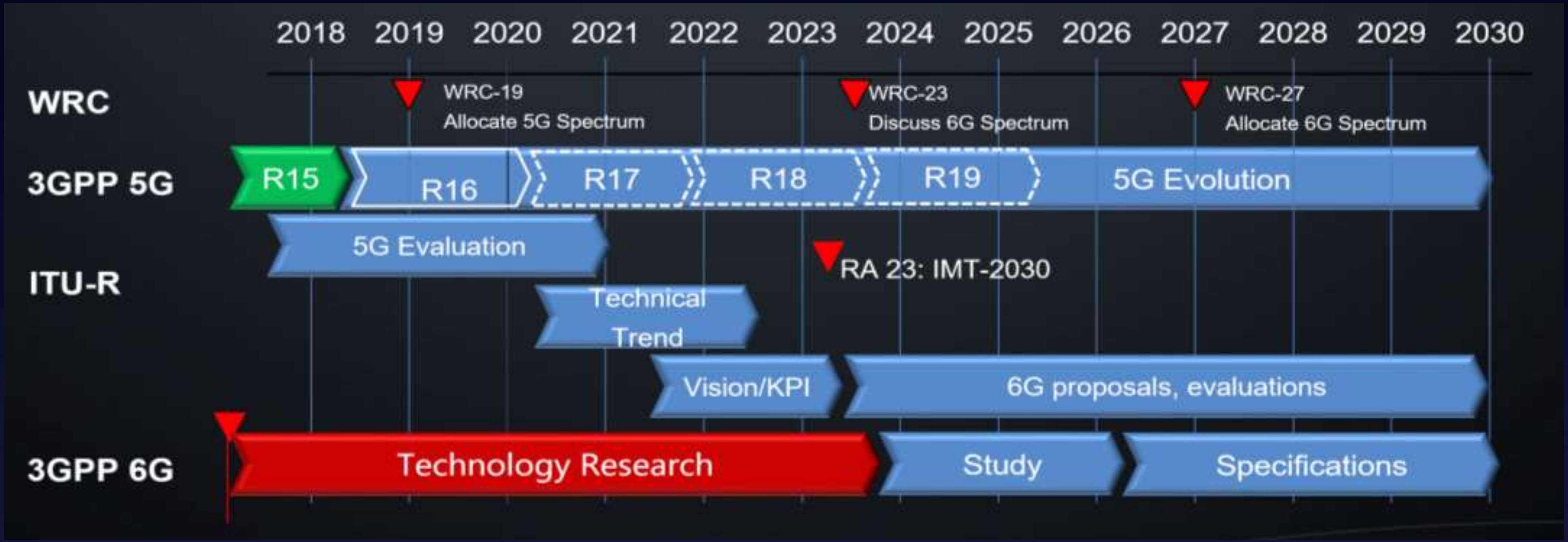


Illustration: Facing the Millimeter-Wave Cell Discovery Challenge in 5G Networks With Context-Awareness, Devoti, Filippini, Capone, IEEE Access 2016.

Beyond 5G Technologies

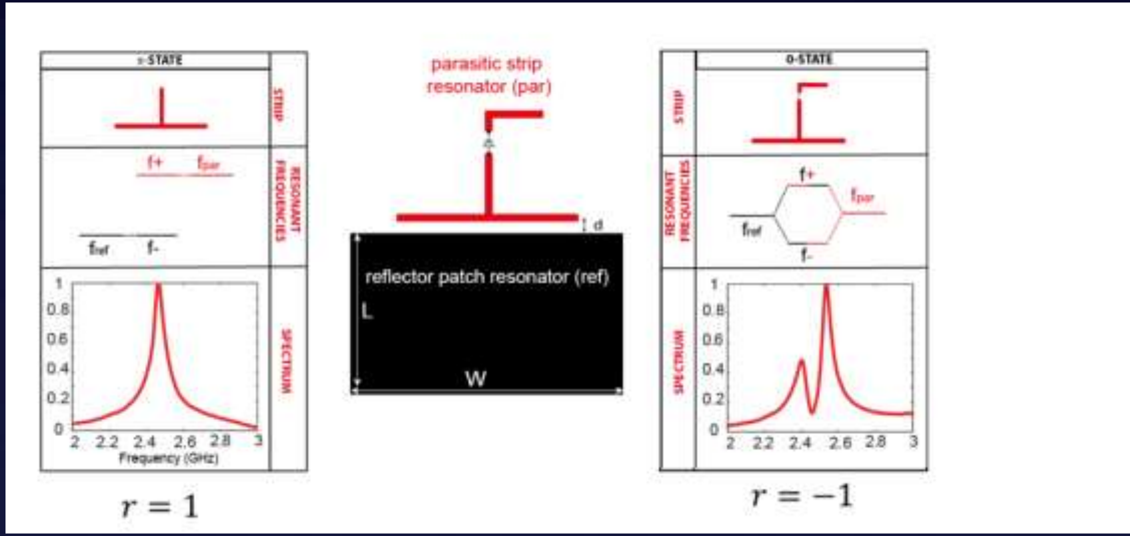
Roadmap to 6G



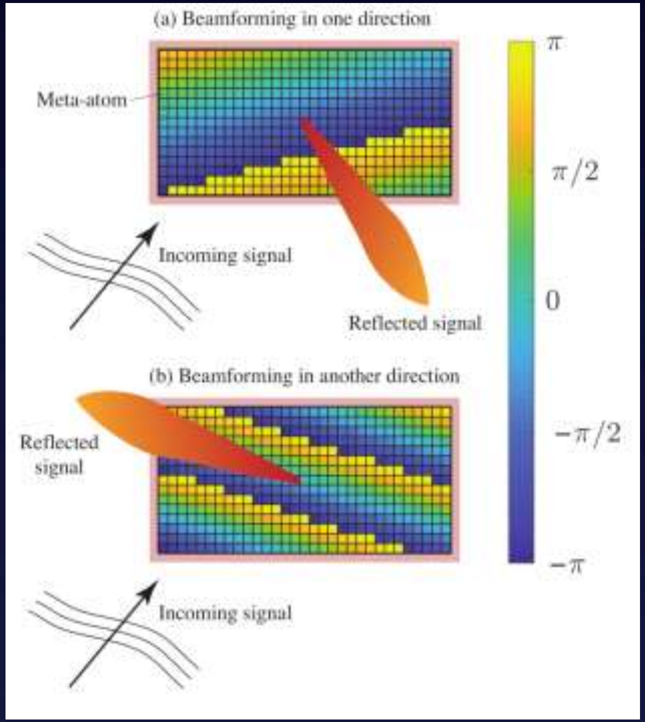
Beyond Massive MIMO: Intelligent Reflecting Surfaces

- Electronically controlled parasitic resonators
- Behave like a set of controllable scatterers (no power is transmitted)

GreenerWave prototype

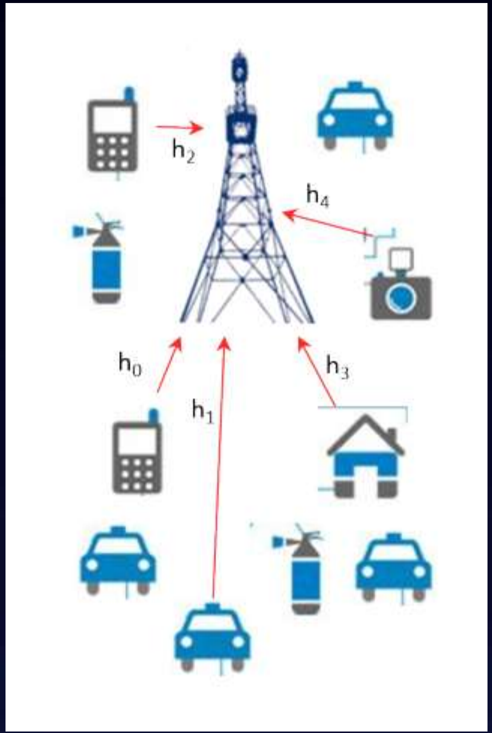


From Kaina, N., Dupré, M., Fink, M. and Lerosey, G. "Hybridized resonances to design tunable binary phase metasurface unit cells". Optics Express 22(16), 18881-18888 (2014).



From "Multiple Antenna Technologies for Beyond 5G," by Jiayi Zhang, Björnson, Matthaiou, Ng, Hong Yang, and Love, 2019

Massive Random Access



- **Random Access** fading channel: $\mathbf{y} = \sum_{i\in\mathcal{A}} \mathbf{s_i}\, h_i + \mathbf{w} \in \mathbb{C}^{\mathrm{T}}$
  - The set $\mathcal{A}$ of active users is unknown
  - Sporadic activity: $|\mathcal{A}| \ll$ K
  - The Rx must jointly estimate $\mathcal{A}$ and the channels
  - Allocating orthogonal pilot sequences is not always feasible

- Possible approaches:

1. Activity detection (estimate $\mathcal{A}$ and $\{h_i\}_{i\in\mathcal{A}}$ based on pilots) using compressed sensing or contention method (slotted ALOHA…)

2. *Unsourced approach (Polyanskiy. "A perspective on massive random access," IEEE International Symposium on Information Theory, 2017)* : let all users use the same codebook (no user-specific pilot sequences). Simpler decoder (problem dimension of order $|\mathcal{A}|$ instead of K)

Massive Random Access

Current PHY designs are based on a divide-and-conquer approach:

- Coordinated assignment of orthogonal pilots
- CSI estimation
- MU-MIMO equalization
- Power control, rate selection
- Synchronization (OFDM symbol, timing advance)
- Carrier frequency offset compensation
- Resource grants
- Coding for the AWGN channel
- Authentication

In eMBB or URLLC, the overhead required to make this happen is amortized over many packets

These only make sense for a connection-oriented PHY.

For sporadic communications, these assumptions need to be questioned and revised

Machine Learning and Big Data for the PHY Layer

ML for Network Parameter Optimization

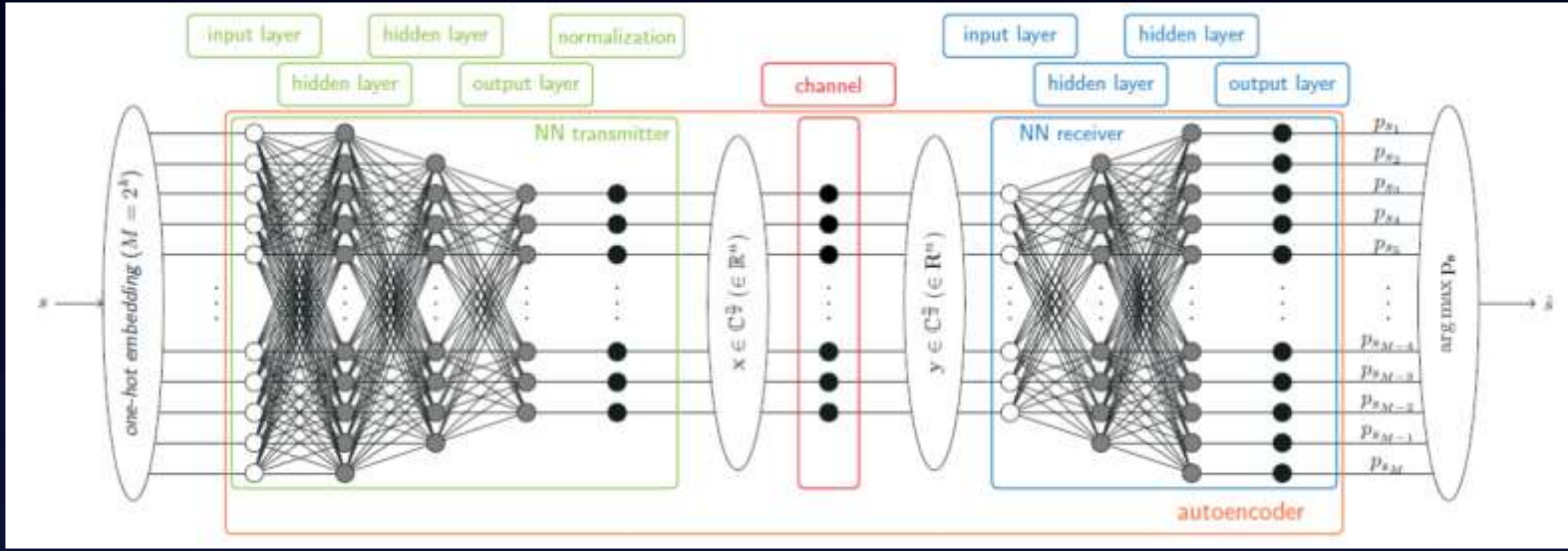
- Cellular networks can be tuned with hundreds/thousands of parameters (resource allocation, antenna downtilt…)

- The interactions between these are difficult to model accurately (depends on interference, geography, user behavior…)

- Use ML and historical data for
  - Performance prediction and optimization
  - Fault detection

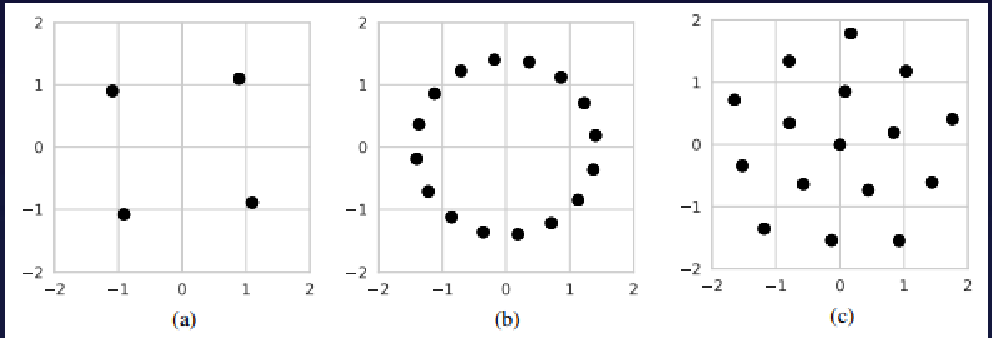
ML for Code and Decoder Design

Design of channel constellations and demapper/decoder through an *autoencoder*:



Note: training is specific to the (current) channel state ⇒ tough real-time implementation constraints!

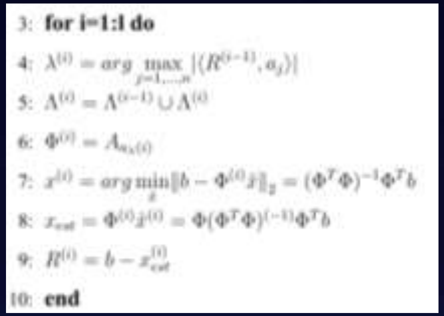
Constellations produced by autoencoders for 2 channel accesses (x-y axes) for (a) 2 bits, (b) 4 bits, (c) 4 bits with average power constraint

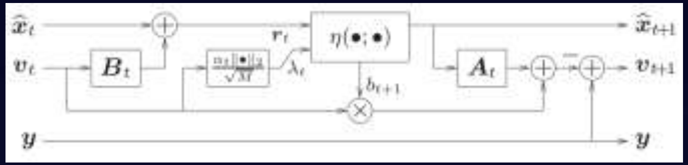


From:“An Introduction to Deep Learning for the Physical Layer,” O’Shea, Hoydis, 2017.

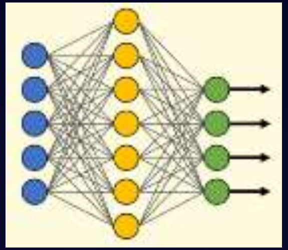
“Learned” Algorithms

- Many classical signal processing algorithms (message-passing, iterative…) can be construed as multiple layers of sums, products and non-linear functions





From “Onsager-Corrected Deep Learning for Sparse Linear Inverse Problems,” Borgerding and Schniter, 2016.

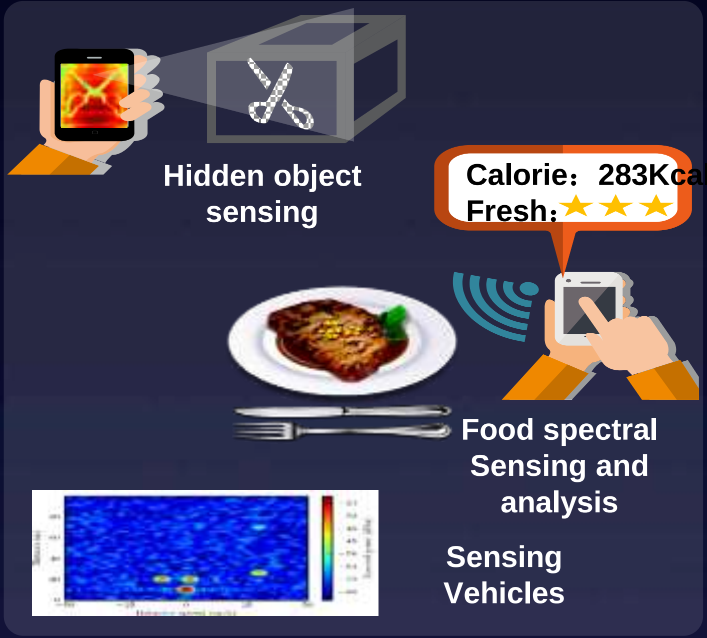


- Apply the **neural network paradigm**:
  1. Existing algorithms provide the NN architecture and weights initialization
  2. A data-driven phase improves the weights

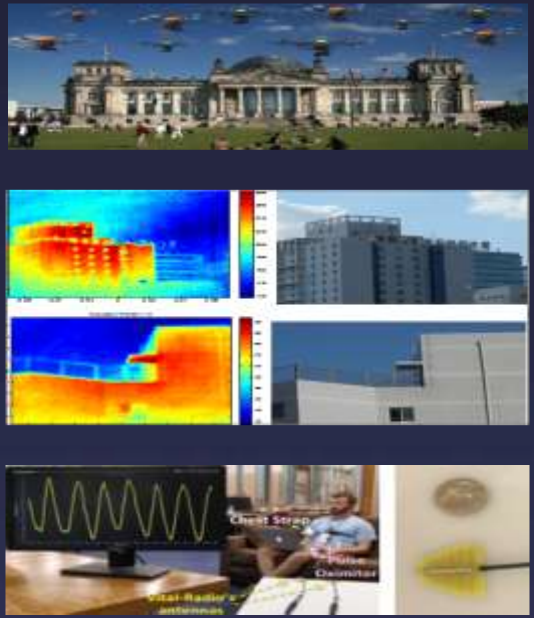
- Get the best of both worlds:
  - Well chosen NN architecture (model based, incorporates decades of expert knowledge!)
  - Data-driven: can overcome the model imperfections if trained with real signals

Convergence of *Communication* and *Sensing*

Space | Chemistry | Biology | Medical



Terminal Sensing

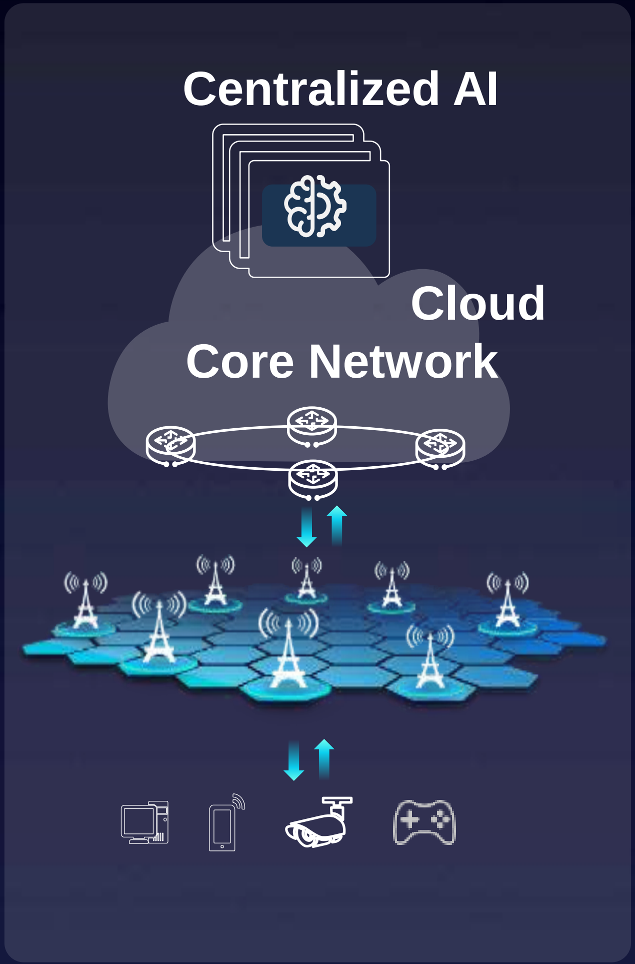


Infrastructure Sensing

Analytics

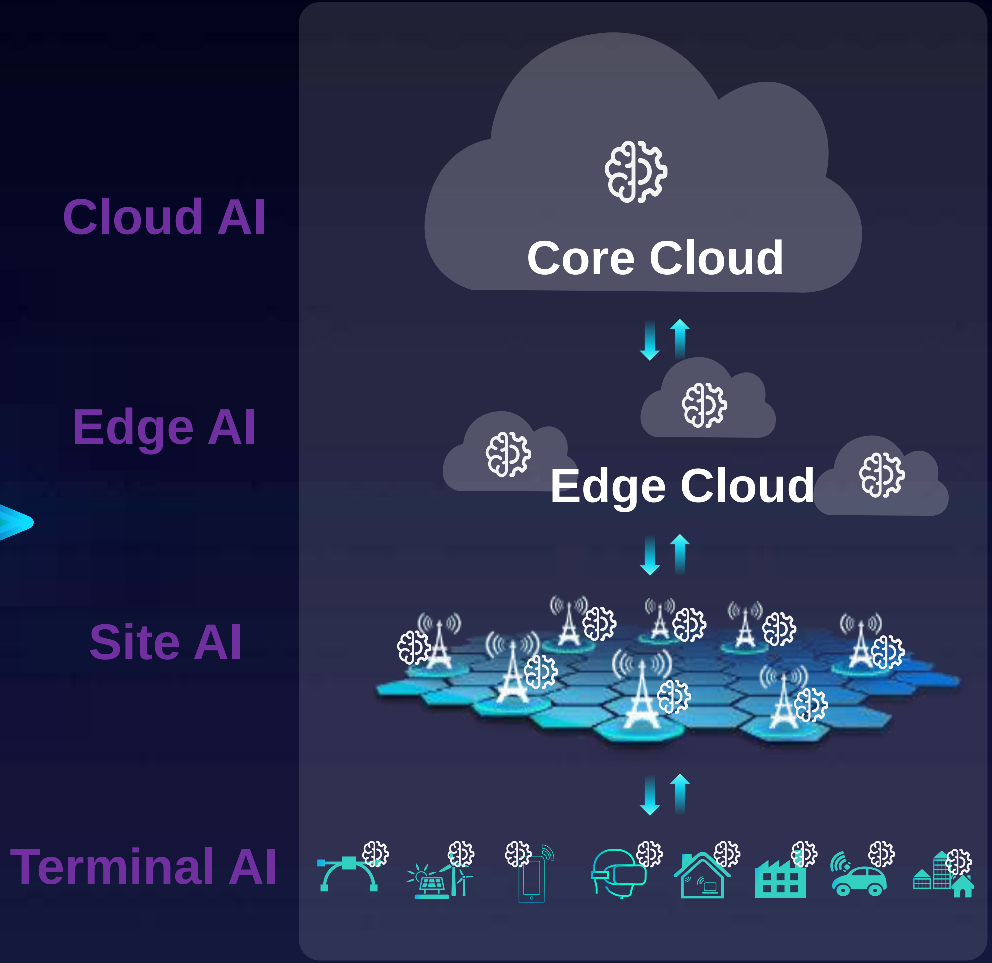
Artificial Intelligence

Information + *Intelligence*: New Paradigms for Networks



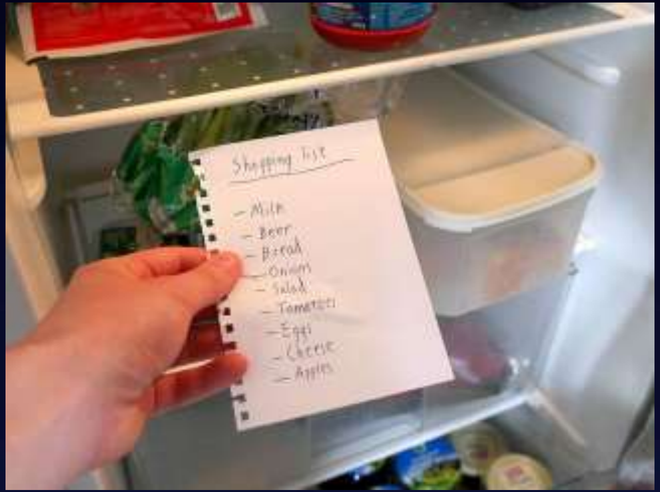
Current Cloud and AI

AI in Cloud



Integrated Cloud, Network & AI

Semantic and Goal-Oriented Communications



Raw data (10MB)

"Shopping list: milk, beer, bread, onions…"

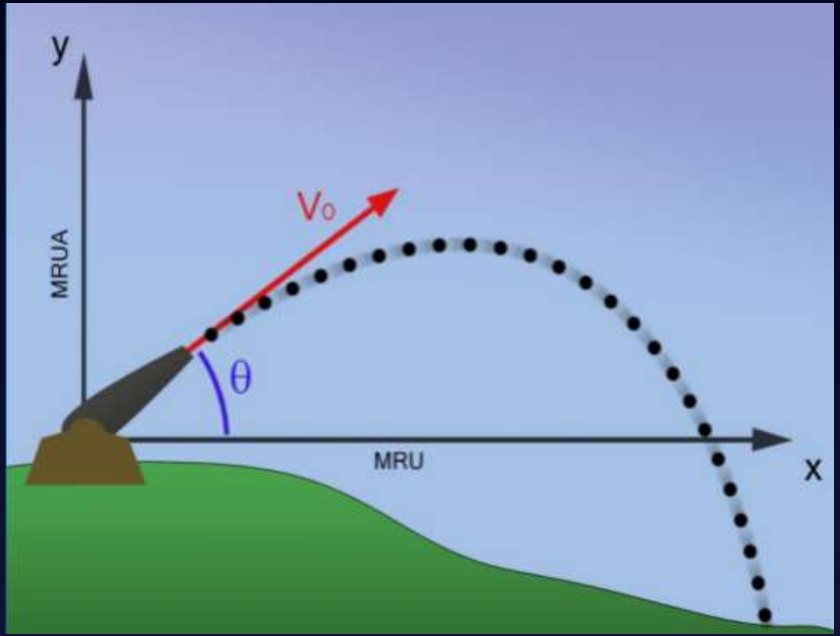
Semantic representation (1kB)

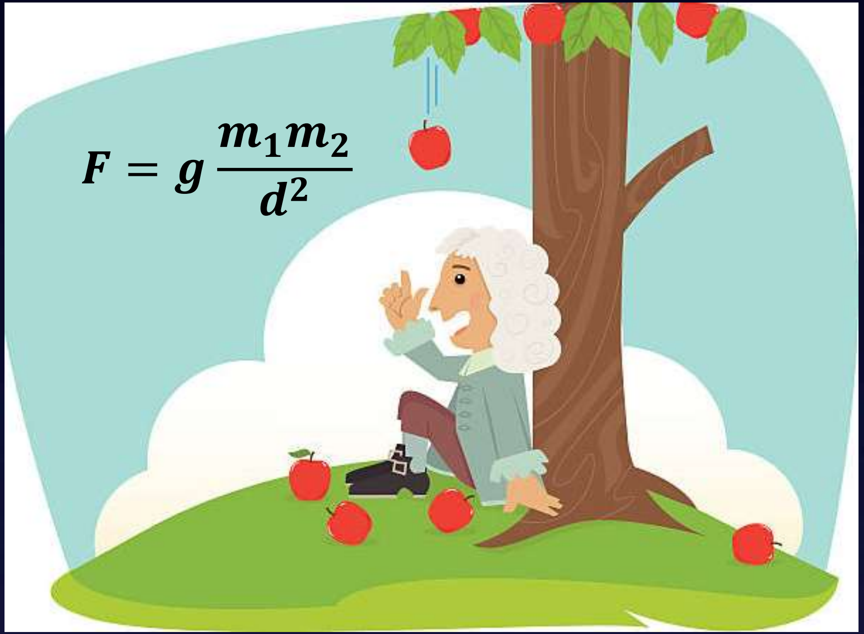
Goal-oriented

From bit pipes to goal-oriented communication

- **Large efficiency improvement**
- **Need to take ultimate objective of communication into account**

Towards true "Intelligence"?





Machine learning: Numerous training samples allow to perform fairly accurate inference

- **Costly training**
- **Little generalization value**

Intelligence: Based on few observations, infer a law of general value

"Machine learning is advanced statistics, and artificial intelligence is advanced PowerPoint"
Prof. Henning Schulzrinne, Columbia U.

Conclusion

- Radio Access Networks are moving from ``bit pipes'' to complex networks with information **processing** capabilities

- At the physical layer, need to go back to the basics (physics…) to find new solutions for 6G

- More emphasis on
  - Latency and Random access
  - In-network processing
  - Semantic / Goal-oriented communications