Wireless Connectivity: An Intuitive and Fundamental Guide

Chapter 11: Using Two, More, or a Massive Number of Antennas

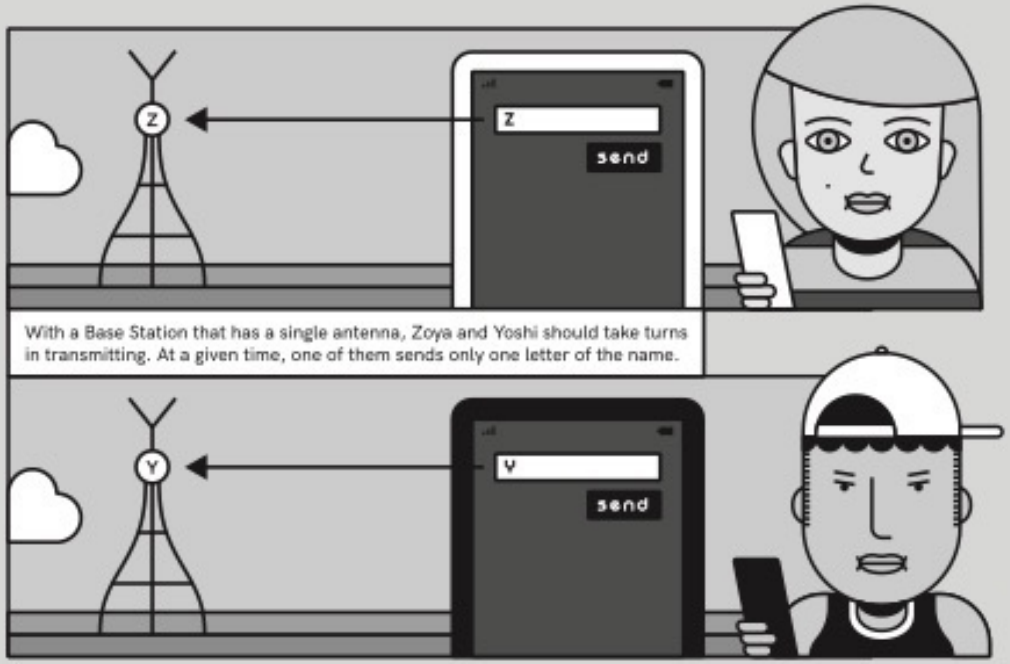
Petar Popovski
Connectivity Section
Department of Electronic Systems
petarp@es.aau.dk

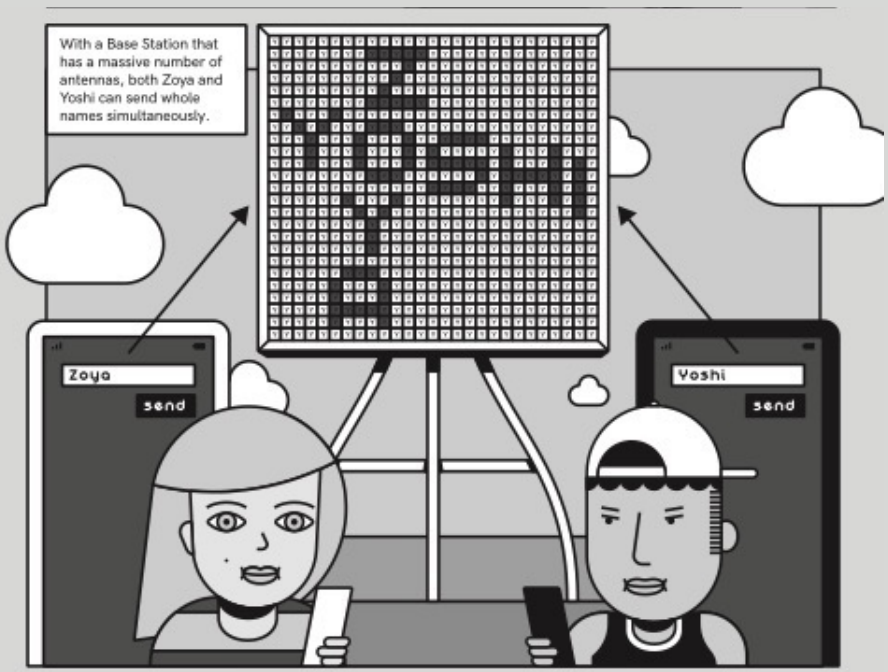
Contributions to the slides:
Israel Leyva-Mayorga
Radoslaw Kotaba
Abolfazl Amiri
Alexandru-Sabin Bana
Robin J. WIlliams

AALBORG UNIVERSITY
DENMARK

Modules

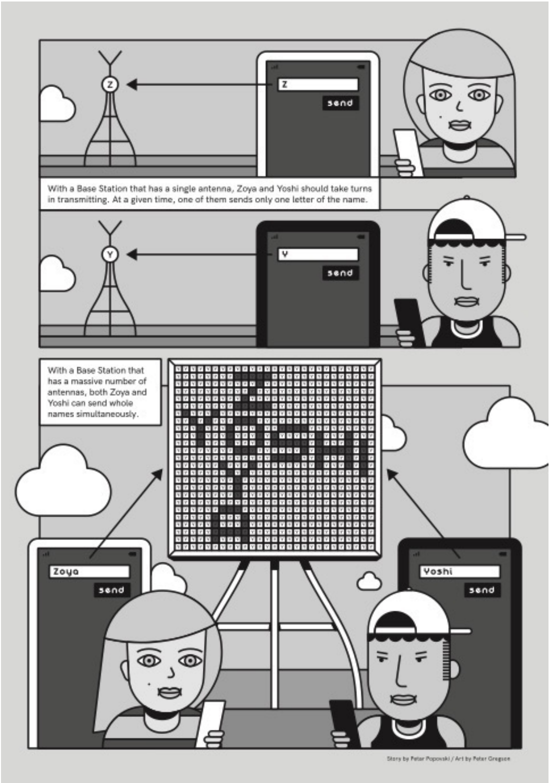
1. An easy introduction to the shared wireless medium
2. Random Access: How to Talk in Crowded Dark Room
3. Access Beyond the Collision Model
4. The Networking Cake: Layering and Slicing
5. Packets Under the Looking Glass: Symbols and Noise
6. A Mathematical View on a Communication Channel
7. Coding for Reliable Communication
8. Information-Theoretic View on Wireless Channel Capacity
9. Time and Frequency in Wireless Communications
10. Space in Wireless Communications
11. **Using Two, More, or a Massive Number of Antennas**
12. Wireless Beyond a Link: Connections and Networks





Story by Petar Popovski / Art by Peter Gregson

Conquering space with multiple antennas



- Multiple antennas bring a spatial dimension on top of time, frequency or code
- Users can be efficiently multiplied in space, multiple antennas can make their signal separable

What will be learned in this chapter

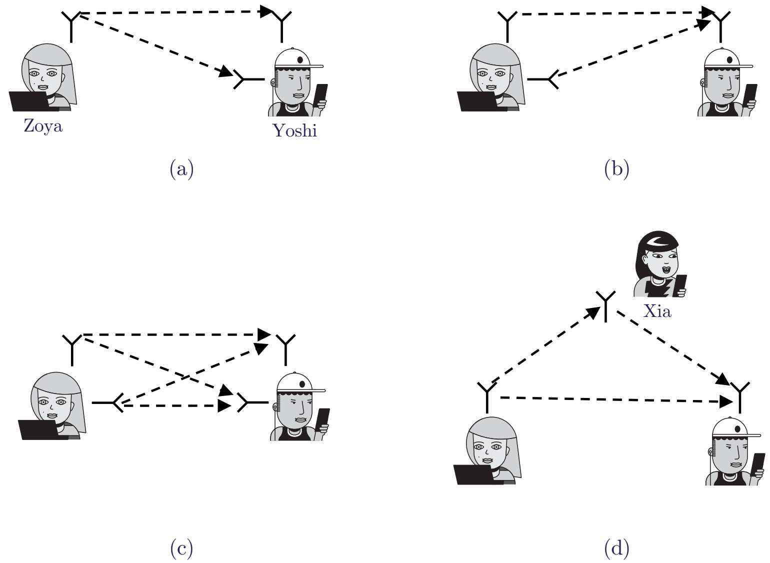
- Rationale behind using multiple antennas
- What changes when two receive or transmit antennas are available
- Spatial multiplexing and multiple access
- What is this beamforming I keep hearing about?
- Increasing the number of antennas from multiple to massive

Communication with multiple antennas

In our quest for expanding the capability of the communication system we have explored the use of more frequencies, coding, longer time, etc.

What about more antennas?

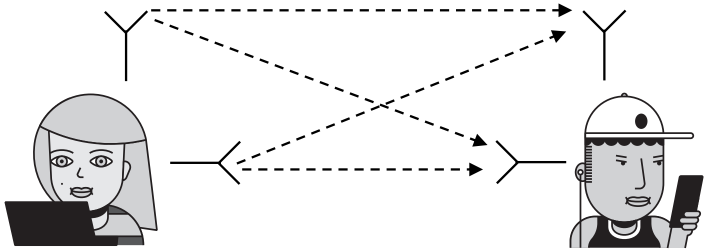
- Depending on the setup and our goal, multiple antennas can provide certain benefits
  - Reliability
  - Throughout
  - Range extension



Assumptions in the multi-antenna model

Before we move further certain clarifications are necessary

- Clearly, the distance between antennas on the same device is much smaller than the transmitter-receiver distance
  - However, it shouldn't be too small as it leads to **correlation**
- We consider some form of OFDM transmission
  - Independent frequency subchannels



Passive vs. active antennas

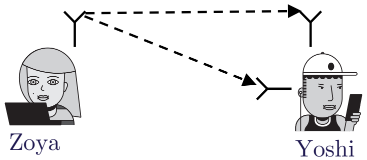
For passive antennas the received signal is simply:

$$y = hz + n$$

In the active case we can have:

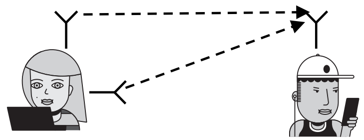
$$y_1 = h_1 z + n_1$$

$$y_2 = h_2 z + n_2$$



(a)

$$y = h_1 z_1 + h_2 z_2 + n$$



(b)

Receiving with multiple antennas

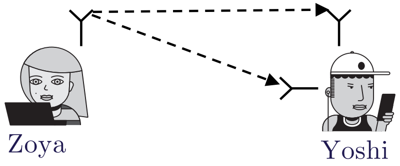
Multiple antennas placed at the receiver allow to collect more energy

Scenario known as **single input multiple output** (SIMO)

Using MRC, receiver combines

$$y_1 = h_1 z + n_1$$
$$y_2 = h_2 z + n_2$$

into $y = h_1^* y_1 + h_2^* y_2$



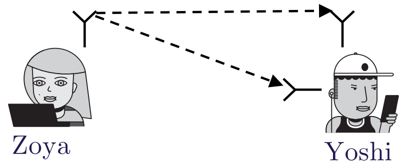
(a)

The resulting SNR is $\gamma_{MRC} = \gamma_1 + \gamma_2$, where $\gamma_i = \frac{|h_i|^2 P}{\sigma^2}$

Receiving with multiple antennas

Other, suboptimal methods exist

- Equal gain combining $\gamma_{EGC} = \frac{(\sqrt{\gamma_1}+\sqrt{\gamma_2})^2}{2}$
- Selection combining $\gamma_{SC} = \max_{1,2} \gamma_i$



(a)

Clearly, each method ensures that the SNR is at least higher than the SNR of the weakest antenna

- Reliability increases since $P(\log_2(1+\gamma_{MRC}) < R) \leq P\left(\log_2\left(1+\min_{1,2}\gamma_i\right) < R\right)$
- Alternatively, Zoya might be able to use higher rate: $R_{adapt} = \log_2(1+\gamma_{MRC})$

Transmitting with multiple antennas

Situation is less obvious when it is the transmitter who has multiple antennas (MISO)

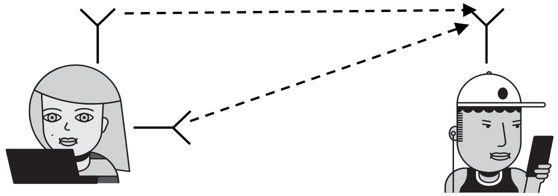
If the channels h_1, h_2 are known to Zoya, rather than transmitting symbol b she can use **precoding** to achieve transmit MRC

$$z_1 = \frac{h_1^*}{\sqrt{|h_1|^2 + |h_2|^2}} b$$

$$z_2 = \frac{h_2^*}{\sqrt{|h_1|^2 + |h_2|^2}} b$$

Resulting in the SNR γ_{MRC} as before

- Another approach involves **antenna selection** where only the stronger of the channels is used
  - Still suboptimal compared to MRC !



Transmitting with multiple antennas

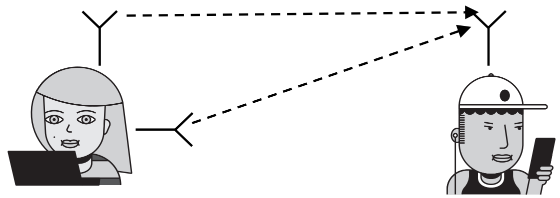
Full knowledge of channel coefficients before transmission is typically a very strong assumption

Is there anything to be gained in MISO with unknown h_1, h_2 then?

- The best Zoya can do is split power in half which yields $\gamma = \frac{|h_1+h_2|^2}{\sigma^2}\frac{P}{2}$
- However, that is only when **a single symbol** transmission is considered

Alamouti space-time block codes

- Let $z_i(j)$ denote the symbol sent from antenna i during j-th symbol transmission
- Zoya sends $z_1(1) = b_1$ and $z_2(1) = b_2$
  then $z_1(2) = -b_2^*$ and $z_2(2) = b_1^*$
- Yoshi receives $y(1) = h_1 b_1 + h_2 b_2 + n(1)$
  then $y(2) = -h_1 b_2^* + h_2 b_1^* + n(2)$



Transmitting with multiple antennas

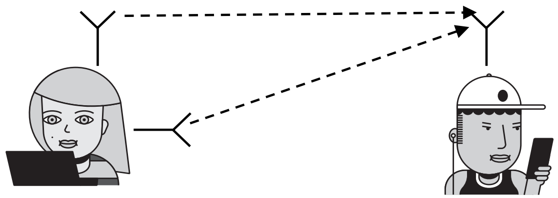
By appropriately weighting the received signals,
Yoshi can retrieve the two **uninterfered** symbols

$$r_1 = h_1^* y(1) + h_2 y^*(2) = |h_1|^2 b_1 + |h_2|^2 b_1 + h_1^* n(1) + h_2^* n(2)$$
$$r_2 = h_2^* y(1) - h_1 y^*(2) = |h_1|^2 b_2 + |h_2|^2 b_2 + h_2^* n(1) - h_1 n(2)$$

The SNR per symbol is $\gamma_{ST} = \frac{|h_1|^2 + |h_2|^2}{\sigma^2} \frac{P}{2}$

Note, that this is always beneficial, unlike the single symbol as it can happen e.g. $|h_1 + h_2|^2 \leq |h_1|^2$

Transmit diversity can be used to improve reliability even when the channels are unknown



Multiple transmit and receive antennas

A natural extension of the previous two schemes is **multiple input multiple output** (MIMO)

In the most general case the input/output relationship is as follows:

$$y_1 = h_{11}z_1 + h_{12}z_2 + n_1$$
$$y_2 = h_{21}z_1 + h_{22}z_2 + n_2$$

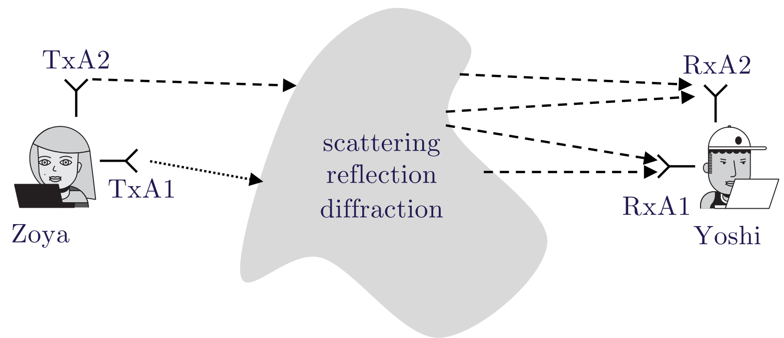
Multiple transmit and receive antennas

The key feature of MIMO is that
it allows to create parallel (independent) spatial channels

- Assume that signal from each TxAi arrives through two paths to each RxAi
- Additionally, assume that TxA1 adds constructively at RxA1 and destructively at RxA2, while the opposite is true for signal TxA2
- Yoshi observes

$$y_1 = h_{11} z_1 + n_1$$
$$y_2 = h_{22} z_2 + n_2$$

- The capacity $C = \log_2(1+\gamma_1) + \log_2(1+\gamma_2)$ is achieved when Zoya sends two independent streams on each channel



Spatial multiplexing

The example earlier was rather artificial

Let us slightly relax the assumptions so that $h_{12}, h_{21} \neq 0$, instead

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ h_1 & -h_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$$

In this case, z_1 and z_2 can be decoded by combining $y_1 + y_2$ and $y_1 - y_2$ respectively leading to $\gamma_i = \frac{2|h_i|^2 P_i}{\sigma_2}$ for $i = 1,2$

- This would not possible if $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 \\ h_1 & h_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}$ instead
- In particular, if $n_1 = n_2 = 0$ the two received signals are identical
- Intuitively the quality of the channels (understood as capability to carry information) depends on their similarity or **correlation**

Spatial multiplexing

Let us write more generally now

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} = \boldsymbol{Hz} + \boldsymbol{n}$$

Where matrix $\boldsymbol{H}$ can be represented through SVD as $\boldsymbol{H} = \boldsymbol{U\Lambda V}^H$, where $\boldsymbol{U}, \boldsymbol{V}$ are unitary and $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2)$

- It is assumed that Zoya and Yoshi can learn $\boldsymbol{H}$ first and then use it to calculate $\boldsymbol{U}, \boldsymbol{\Lambda}, \boldsymbol{V}$

Upon receiving $\boldsymbol{y}$, Yoshi transforms it by multiplying with $\boldsymbol{U}^H$

$$\widetilde{\boldsymbol{y}} = \boldsymbol{U}^H\boldsymbol{y} = \boldsymbol{\Lambda V}^H\boldsymbol{z} + \boldsymbol{U}^H\boldsymbol{n} = \boldsymbol{\Lambda}\tilde{\boldsymbol{z}} + \widetilde{\boldsymbol{n}}$$

Spatial multiplexing

Clearly, since $\boldsymbol{\Lambda}$ is diagonal this becomes **similar** to the communication setup from the example at the very beginning

Note that, Zoya also needs to do her part, namely transmit $\tilde{\boldsymbol{z}} = \boldsymbol{V}^H \boldsymbol{z}$ instead

- Moreover, knowing $\boldsymbol{\Lambda}$ she should also perform waterfilling to determine P_1, P_2

The capacity is $C_{2\times 2} = \log_2\left(1 + \frac{P_1 \lambda_1^2}{\sigma^2}\right) + \log_2\left(1 + \frac{P_2 \lambda_2^2}{\sigma^2}\right)$

- It is depending on how well-conditioned $\boldsymbol{H}$ is $\frac{\lambda_2}{\lambda_1} = [0,1]$

In general, the maximum number of spatial channels is $M = \min\{M_T, M_R\}$

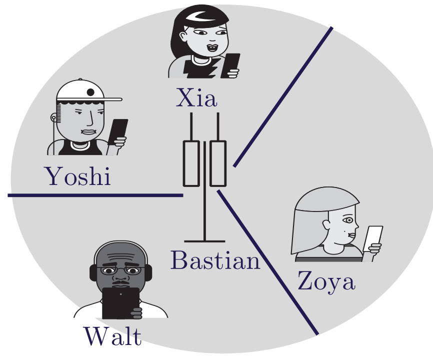
Spatial division of multiple users

Antenna directivity introduces coverage areas, which enables **spatial division multiple access** (SDMA)

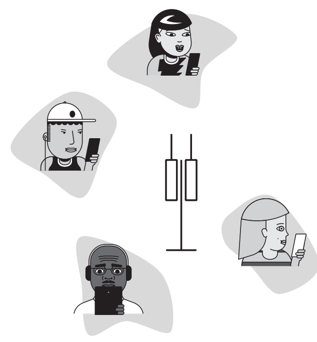
The straightforward geometric beamforming can be further generalized into **digital beamforming**

Bastian utilizes multiple antennas and the principles explained earlier to create spatial channels to each user

This is known as **Zero-Forcing (ZF)** precoding



(a)



(b)

Spatial division of multiple users

Let us start with a simple example of downlink MISO with 5 Tx antennas and two recipients: Zoya and Yoshi

- Their respective channels are $\boldsymbol{h}_Z = [h_{Z1}\ h_{Z2}\ h_{Z3}\ h_{Z4}\ h_{Z5}]$, $\boldsymbol{h}_Y = [h_{Y1}\ h_{Y2}\ h_{Y3}\ h_{Y4}\ h_{Y5}]$
- Bastian sends s_Z to Zoya by mapping it to all antennas with precoding vector $\boldsymbol{v}_Z = [v_{Z1}\ v_{Z2}\ v_{Z3}\ v_{Z4}\ v_{Z5}]^T$ (similarly for s_Y)

$$\boldsymbol{b}_{TX} = \boldsymbol{v}_Z s_Z + \boldsymbol{v}_Y s_Y$$

- What Bastian wants to achieve (through appropriate selection of ZF precoding vectors) is that $\boldsymbol{h}_Z \boldsymbol{v}_Y = \boldsymbol{h}_Y \boldsymbol{v}_Z = 0$
  - As a result $y_Z = \boldsymbol{h}_Z \boldsymbol{v}_Z s_Z + n_Z$ and $y_Y = \boldsymbol{h}_Y \boldsymbol{v}_Y s_Y + n_Y$
  - $\boldsymbol{v}_Z$, $\boldsymbol{v}_Y$ should be chosen such that the individual SNRs are maximized

Importantly, due to channel reciprocity, $\boldsymbol{v}_Z$, $\boldsymbol{v}_Y$ can be reused when Zoya and Yoshi transmit simultaneously in the uplink!

Spatial division of multiple users

Multiple antennas allow to support multi-user communication in yet another way

As this is independent from time/frequency multiplexing,
all these multiplexing methods can be used jointly

The prerequisite is that channels to different users are uncorrelated

Without rich scattering ZF can be suboptimal or downright infeasible

Zero-forcing vs. other types of precoding

ZF precoding is not the only possible way of achieving digital beamforming, nor it is generally the best one

- While it cancels the interference, it can also enhance the noise
- A better performance is achieved through Wiener or MMSE precoding

Common property of ZF, MMSE precoding is that they are **linear** i.e. they are independent of data symbols transmitted

As always, even better performance can be achieved when transmission is tailored to the data

However, this comes at the cost of complexity and in practice linear methods may be preferable

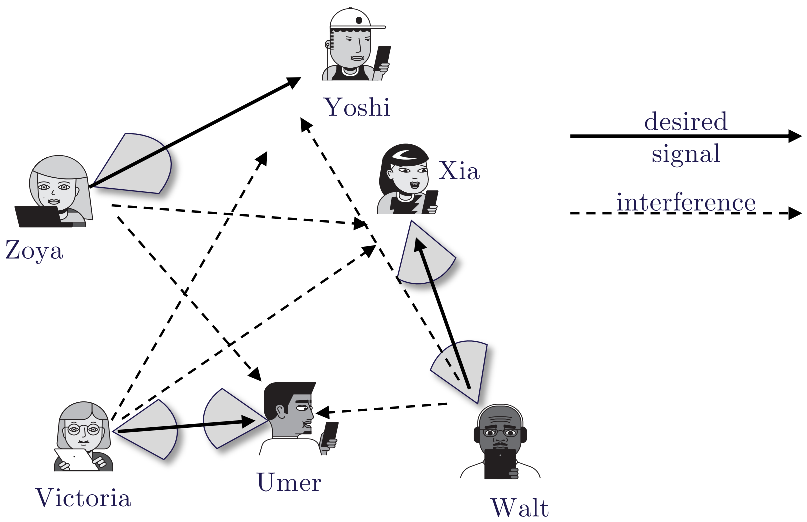
Beamforming and spectrum sharing

The SDMA is not restricted to
a centralized communication system

By carefully selecting beams and nulls,
multiple simultaneous connections could be active

The use of multiple antennas enables
spectrum sharing

While **digital** beamforming is more powerful,
similar functionality can be achieved with
analog and hybrid **analog-digital** beamforming



Beamforming and spectrum sharing

It is important to understand that all access techniques described so far have their own challenges, advantages and disadvantages

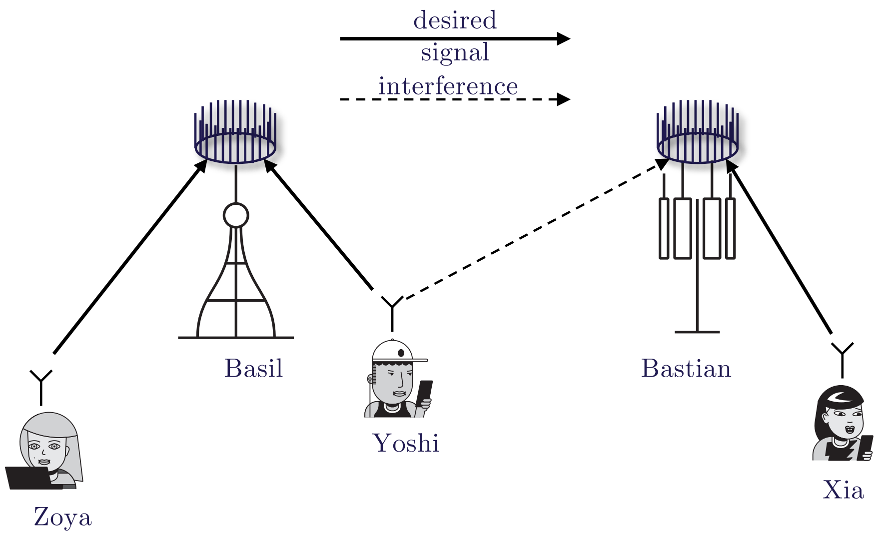
- SDMA requires high quality CSI, and hence introduces **overhead**
  - It also relies on the user separation which can't always be controlled
- In CDMA system designer has more control through the selection of spreading sequences
  - Precise **synchronization** is required
- In FDMA time synchronization and CSI are not required, which makes it the simplest solution
  - However dedicating spectrum portion to each user and the need for guard bands make it the most **wasteful in terms of bandwidth**

What if the number of antennas is massive?

Even a perfect knowledge of channel coefficients
is not very helpful
if the conditions are simply poor

However, if the number of antennas used for beamforming is massive,
the bad-luck cases in SDMA disappear due to statistics of large numbers

This is the rationale behind **massive MIMO**



What if the number of antennas is massive?

Consider uplink scenario with Basil receiving signals from Zoya and Yoshi (their channels assumed to be known)

$$\boldsymbol{b}_{RX} = \boldsymbol{h}_Z^T s_Z + \boldsymbol{h}_Y^T s_Y + n$$

Basil extracts the signal from Zoya as follows:

$$\widehat{b_Z} = \frac{1}{M}\boldsymbol{h}_Z^* \boldsymbol{b}_{RX} = \frac{1}{M}\boldsymbol{h}_Z^* \boldsymbol{h}_Z^T s_Z + \frac{1}{M}\boldsymbol{h}_Z^* \boldsymbol{h}_Y^T s_Y + \frac{1}{M}\boldsymbol{h}_Z^* \boldsymbol{n}$$

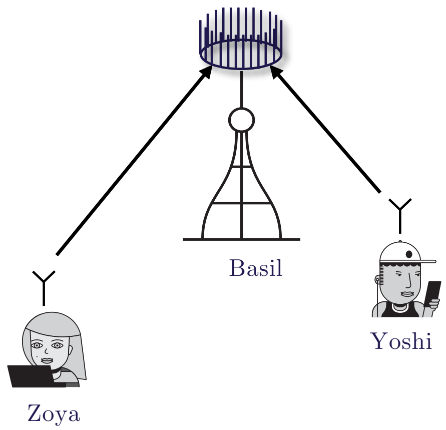
Massiveness of M has two consequences:

- **Channel hardening**:

$$\frac{1}{M}\boldsymbol{h}_Z^* \boldsymbol{h}_Z^T s_Z = \frac{1}{M}\sum_{m=1}^{M} |h_{Zm}|^2 \xrightarrow{M\to\infty} \beta_Z$$

- **Asymptotic orthogonality**:

$$\frac{1}{M}\boldsymbol{h}_Z^* \boldsymbol{h}_Y^T s_Y = \frac{1}{M}\sum_{m=1}^{M} h_{Zm}^* \, h_{Ym} \xrightarrow{M\to\infty} 0$$



What if the number of antennas is massive?

As a result the received signal becomes:

$$\widehat{b_Z} = \beta_Z s_Z + \frac{1}{M}\boldsymbol{h}_Z^* \boldsymbol{n}$$

The SNR becomes independent of particular channel realizations and instead approaches

$$\gamma_Z = \frac{M\beta_Z}{\sigma^2}$$

Similar results can be observed in the downlink

- In this case, Basil needs to construct the transmitted signal as

$$\boldsymbol{b}_{TX} = \rho_Z \boldsymbol{h}_Z^H s_Z + \rho_Y \boldsymbol{h}_Y^H s_Y$$

- Where ρ_Z, ρ_Y ensure transmit power constraints are met

> In the downlink, beamforming allows to concentrate energy at a very precise point in space
>
> In the uplink, beamforming allows to almost perfectly mute all interference

What if the number of antennas is massive?

SDMA is conditional on the availability of CSI which can be challenging in massive MIMO

A feasible tactic is to have terminals transmit limited number of **pilot signals** in the uplink

Basil can simultaneously estimate all M coefficients of $\boldsymbol{h}_Z$

- Can you see why this training would be problematic in the downlink?

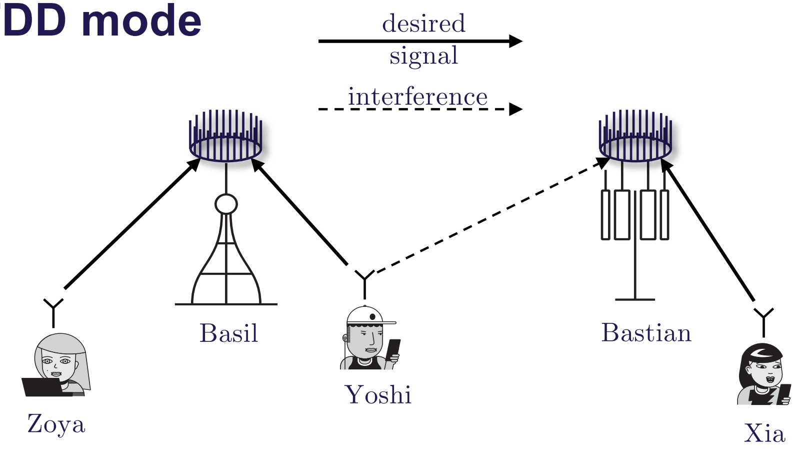
To use $\boldsymbol{h}_Z$ in the downlink, system must operate in **TDD mode**

The number of required pilots scales **linearly** with the number of users

This can get prohibitive, especially recalling that there are multiple cells around

In those instances **pilot contamination** phenomenon can occur

$$\boldsymbol{b}_{RX} = \boldsymbol{h}_X^T s + \boldsymbol{g}_Y^T s + n = (\boldsymbol{h}_X^T + \boldsymbol{g}_Y^T) s + n$$



Outlook and takeaways

- Multipath propagation and geometry introduce new phenomena that can be exploited with multiple antennas
- The potential benefits depend on the scenario
- MIMO introduces new flavor of channels – spatial ones
- Beamforming enables new type of multiple access
- As number of antennas goes massive some new interesting properties emerge