# List of Abbreviations

| | |
|---|---|
| **BLER** | Block Error Rate |
| **BS** | Base Station |
| **CSI** | Channel State Information |
| **DL** | Downlink |
| **EXP/PF** | Exponential/Proportional Fair |
| **FNBW** | First Null Beam Width |
| **GoB** | Grid of Beams |
| **GoP** | Group of Pictures |
| **HMD** | Head Mounted Display |
| **HOL** | Head Of Line |
| **LoS** | Line-of-Sight |
| **MI-ESM** | Mutual Information Effective SINR Mapping |
| **M-LWDF** | Maximum-Largest Weighted Delay First |
| **MR** | Maximum Ratio |
| **MRC** | Maximum Ratio Combining |
| **OLLA** | Outer Loop Link Adaptation |
| **PF** | Proportional Fair |
| **PRB** | Physical Resource Block |
| **RS** | Reference Signal |
| **RX** | Receiver |
| **SINR** | Signal-to-Interference-plus-Noise Ratio |
| **TB** | Transport Block |
| **TBS** | Transport Block Size |
| **TDD** | Time Division Duplex |
| **TTI** | Transmission Time Interval |
| **TX** | Transmitter |
| **UE** | User Equipment |
| **UL** | Uplink |
| **URA** | Uniform Rectangular Array |

# List of Symbols

## Latin alphabet

not perfectly alphabetically ordered *yet*

| | |
|---|---|
| $a_\phi$ | azimuth lower limit for GoB |
| $a_\theta$ | elevation lower limit for GoB |
| $b_\phi$ | azimuth upper limit for GoB |
| $b_\theta$ | elevation lower limit for GoB |
| $B$ | bandwidth |
| $BLER_0$ | target BLER |
| $C_t$ | coordinates of centre of the table |
| $C_m$ | coordinates of centre of mass of the room |
| $d_f$ | distance to the front of the user for camera placement |
| $d_F$ | Fraunhofer distance |
| $d_s$ | distance to the side of the user for camera placement |
| $d_{out}$ | distance outwards from the head centre for HMD antenna offset |
| $d_{up}$ | distance upwards from the head centre for HMD antenna offset |
| $D$ | largest dimension of radiator to estimate effective area |
| $E_b$ | energy per bit |
| $\boldsymbol{H}_{bul}$ | channel matrix between BS $b$ and UE $u$ in layer $l$ |
| $I$ | interference |
| $j$ | imaginary unit $(j = \sqrt{-1})$ |
| $k_B$ | Boltzmann constant |
| $L_{max}$ | maximum latency for the radio link |
| $N_0$ | noise power spectral density |
| $N_r$ | number of receive antennas |
| $N_t$ | number of transmit antennas |
| $N_{users}$ | number users or participants |
| $N_{phy}$ | number of physical users |
| $N_{vir}$ | number of virtual users |
| $N_{CSI}$ | number of beams with CSI-RS |
| $N_{bs}$ | number of BSs |
| $N_{ue}$ | number of UEs |
| $N_{cam}$ | number of cameras |
| $N_x$ | number of antennas along the x-dimension |
| $N_y$ | number of antennas along the y-dimension |
| $N_{slots}^{DL}$ | number of DL slots in a transmission period |
| $N_{slots}^{UL}$ | number of UL slots in a transmission period |

| | |
|---|---|
| $N_{slots}^{TDD}$ | number slots in a transmission period |
| $N_{slots}^{CSI}$ | number slots between CSI updates |
| $N_{slots}^{SCH}$ | number slots between scheduling updates |
| $N_{ant}^{UE}$ | number of antennas on the UE side |
| $N_{ant}^{BS}$ | number of antennas on the BS side |
| $NF_{BS}$ | BS noise figure |
| $NF_{UE}$ | UE noise figure |
| $NF_r$ | noise figure at the receiver |
| $N_{PRB,bu}$ | number of PRBs allocated for link between BS $b$ and UE $u$ |
| $N_{PRB,bul}$ | number of PRBs allocated for link between BS $b$ and UE $u$ in layer $l$ |
| $N_{symb}^{PRB}$ | number of symbols per PRB |
| $N_{bits}^{symb}$ | number of bits per symbol |
| $N_{infobits}^{symb}$ | number of information bits per symbol |
| $N_{bits}^{slot}$ | number of bits per slot |
| $N_{bits}^{PRB}$ | number of bits per PRB |
| $N_{bits,bul}$ | number of bits to be sent between BS $b$ and UE $u$ in layer $l$ |
| $p$ | scheduling priority |
| $P_N$ | noise power |
| $P_u$ | position of user $u$ |
| $P_r$ | received power |
| $P_s$ | signal power |
| $P_t$ | transmit power |
| $P_{t,max}^{UE}$ | maximum transmit power per UE |
| $P_{t,max}^{BS}$ | maximum transmit power per BS |
| $P_{r,bu}^{UE}$ | received power by the UE for a link between BS $b$ and UE $u$ |
| $P_{t,bu}^{BS}$ | transmit power at the BS for a link between BS $b$ and UE $u$ |
| $P_{t,bu}^{UE}$ | transmit power at the UE for a link between BS $b$ and UE $u$ |
| $P_{t,bul}$ | transmit power for a link between BS $b$ and UE $u$ in layer $l$ |
| $P_{t,bul}^{UE}$ | transmit power at the UE for a link between BS $b$ and UE $u$ in layer $l$ |
| $P_{t,bul}^{BS}$ | transmit power at the BS for a link between BS $b$ and UE $u$ in layer $l$ |
| $P_{IaCI}$ | interference power from intra-cell interference |
| $P_{IeCI}$ | interference power from inter-cell interference |
| $P_{ILI}$ | interference power from inter-layer interference |
| $Q_m$ | modulation order |
| $r_\phi$ | azimuth resolution for GoB |
| $r_\theta$ | elevation resolution for GoB |
| $r_t$ | radius of table |
| $r_u$ | radius of user circumference of disposition |
| $r_{DL/UL}$ | ratio between DL and UL application throughputs |
| $r_{P/I}$ | ratio between P-frame and I-frame sizes |
| $R_b$ | instantaneous bit rate |
| $R_c$ | code rate |
| $R_{packet}$ | packet arrival rate |
| $R$ | instantaneous throughput |
| $\overline{R}$ | average throughput |
| $\overline{R}_{DL}$ | average application throughput in the DL |
| $\overline{R}_{UL}$ | average application throughput in the UL |
| $R_F$ | frame rate |

| | |
|---|---|
| $s$ | speed of user head position change |
| $s_{TDD}$ | TDD split |
| $SINR_{eff}$ | effective SINR, i.e. aggregated over all scheduled PRBs |
| $SINR_i$ | SINR of the i-th PRB |
| $S_I$ | size of I-frame |
| $S_P$ | size of P-frame |
| $S_{packet}$ | size of a packet |
| $S_{GoP}$ | size of a GoP |
| $S_{TB}$ | size of a TB |
| $S_{TB,max}$ | maximum size of a TB |
| $t_w$ | exponential smoothing window size for proportional fair ratio |
| $T$ | noise temperature |
| $T_{sim}$ | simulation duration |
| $T_{slot,\mu}$ | slot duration for numerology $\mu$ |
| $T_{rot}$ | interval between consecutive orientations when head rotating |
| $\boldsymbol{w}$ | vector of beamforming weights |
| $\boldsymbol{w}_{\phi,\theta}$ | beamforming weights from a GoB with direction $(\phi, \theta)$ |
| $\boldsymbol{w}_{i,j}$ | beamforming weights vector with grid indices $(i, j)$, from a GoB |
| $\boldsymbol{w}_{bu}^{BS}$ | beamforming weights between BS panel $b$ and UE $u$, at the BS |
| $\boldsymbol{w}_{bu}^{UE}$ | beamforming weights between BS panel $b$ and UE $u$, at the UE |
| $\boldsymbol{w}_{t,bul}$ | transmit weights vector between BS panel $b$ and UE $u$, in layer $l$ |
| $\boldsymbol{w}_{r,bul}$ | receive weights vector between BS panel $b$ and UE $u$, in layer $l$ |

**Greek alphabet**

| | |
|---|---|
| $\alpha_P$ | power compensation factor for UL power control |
| $\alpha_u$ | angle from the centre of the table to user $u$ |
| $\beta_x$ | upper limit on uniform distribution for rotation around x axis |
| $\beta_y$ | upper limit on uniform distribution for rotation around y axis |
| $\beta_z$ | upper limit on uniform distribution for rotation around z axis |
| $\gamma$ | burstiness parameter for application traffic |
| $\gamma_{OLLA}$ | step size for OLLA parameter ($\Delta_{OLLA}$) update |
| $\Delta_{OLLA}$ | outer loop link adaptation step |
| $\eta_{OH}$ | efficiency due to overhead |
| $\eta_{slot}$ | efficiency in bit rate from slot format |
| $\lambda$ | wavelength |
| $o$ | overlap parameter for application traffic |
| $\sigma_x$ | standard deviation of normal distribution for position coordinate x |
| $\sigma_y$ | standard deviation of normal distribution for position coordinate y |
| $\sigma_z$ | standard deviation of normal distribution for position coordinate z |
| $\tau_{CSI}$ | CSIs delay, in number of TTIs |
| $\tau_{ACK}$ | delay before acknowledgement, in number of TTIs |

**Sets:**

| | |
|---|---|
| $\mathbb{N}$ | natural numbers |
| $\mathbb{N}_0$ | natural numbers including zero |
| $\mathcal{B}$ | base station panels |
| $\mathcal{U}_b$ | users served by base station $b$ |
| $\mathcal{L}_{bu}$ | layers scheduled between base station $b$ and user equipment $u$ |
| $\mathcal{F}$ | frequencies to simulate |

**Other nomenclature**

| | |
|---|---|
| $\boldsymbol{A}$ | matrix |
| $\boldsymbol{a}$ | column vector |
| $\lvert\boldsymbol{a}\rvert$ | euclidean norm of vector $\boldsymbol{a}$ |
| $\boldsymbol{A}^{\mathsf{T}}$ | transpose of $\boldsymbol{A}$ |
| $\boldsymbol{A}^{\mathsf{H}}$ | Hermitian of $\boldsymbol{A}$, also know as the transpose conjugate of $\boldsymbol{A}$ |
| $\lceil a \rceil$ | ceil $a$, i.e. round up $a$ to the nearest integer |
| $\lfloor a \rfloor$ | floor $a$, i.e. round down $a$ to the nearest integer |
| $\hat{a}$ | estimate of $a$ |

# 1

# Methodology

## Contents

## 1.1 Radio Access Network

In this section, we detail the functions executed by the network equipment to enable data transmission. The network equipment needs to acquire CSI and manage resources accordingly to cope with the incoming application traffic and fulfil service requirements. Firstly, we go over important considerations and assumptions, namely regarding multi-layer transmissions, concentrating on the Downlink (DL), among other miscellaneous but relevant matters. Then we list the steps required to simulate a Transmission Time Interval (TTI) and all processing associated with making the right choices when transmitting and receiving. We summarise these steps with a flowchart and proceed to detail each one.

Firstly, we opt for a Grid of Beams (GoB)-based beamforming approach. With the growing number of antennas at the receivers, full channel knowledge is practically unobtainable, and we need to resort to more overhead-efficient approaches.

Secondly, we address the considerations regarding multi-layer transmission. To reiterate, the difference between single-layer and multi-layer operation is the number of independent streams transmitted per User Equipment (UE). And to transmit independent streams or layers, there must be some orthogonality mechanism that renders such layers independent. The orthogonality domain we are concern with is orthogonality in space. However, by opting GoB-based beamforming although we save in overhead, we loose considerably in transmission flexibility. Free-format beamforming would allow us to send independent layers in the same direction, only focusing different antennas at the reception. But using a GoB we do not have enough beams to do that, instead we have to resort to completely different propagation paths, with paths beyond the first not being the Line-of-Sight (LoS). This would not yield insignificant improvements.

Another option would be to resort to polarisation orthogonality. Instead of using all antenna elements to perform a transmission, we may use the antennas oriented in a given direction to send one layer and the elements oriented perpendicularly to send another. Note that the same beam in the GoB can be used for the different polarisations when they are to be sent over the same path. However, the moments in time where inter-polarisation interference is small, e.g. less than 20 dB, are rare. In other words, often antennas with a given orientation at the receiver get signal from both polarisations at the transmitter. Thus, it would require considerably more complicated interference estimations algorithms to do multi-layer transmissions polarisation-based. This is why we opt for single-layer transmission using all antennas both in the Transmitter (TX) and in the Receiver (RX). Nevertheless, the vast majority of modelling in this section is agnostic to the number of layers.

Thirdly, although in Section **??** we modelled the location of all UEs in the system and application traffic for UL and DL, for conciseness in this section we describe the model for DL transmission procedure and thus we do not consider cameras. In the DL, he number of UEs $N_{ue}$ equal to the number of physical users $N_{phy}$. Moreover, we consider a single-BS with a one or more antenna panels.

In essence, we need to list all procedures that can happen in a TTI. Some may not happen every TTI and we need to state in what circumstances they do happen. See in Figure 1.1 a flowchart of the main steps required to simulate a (DL) TTI.

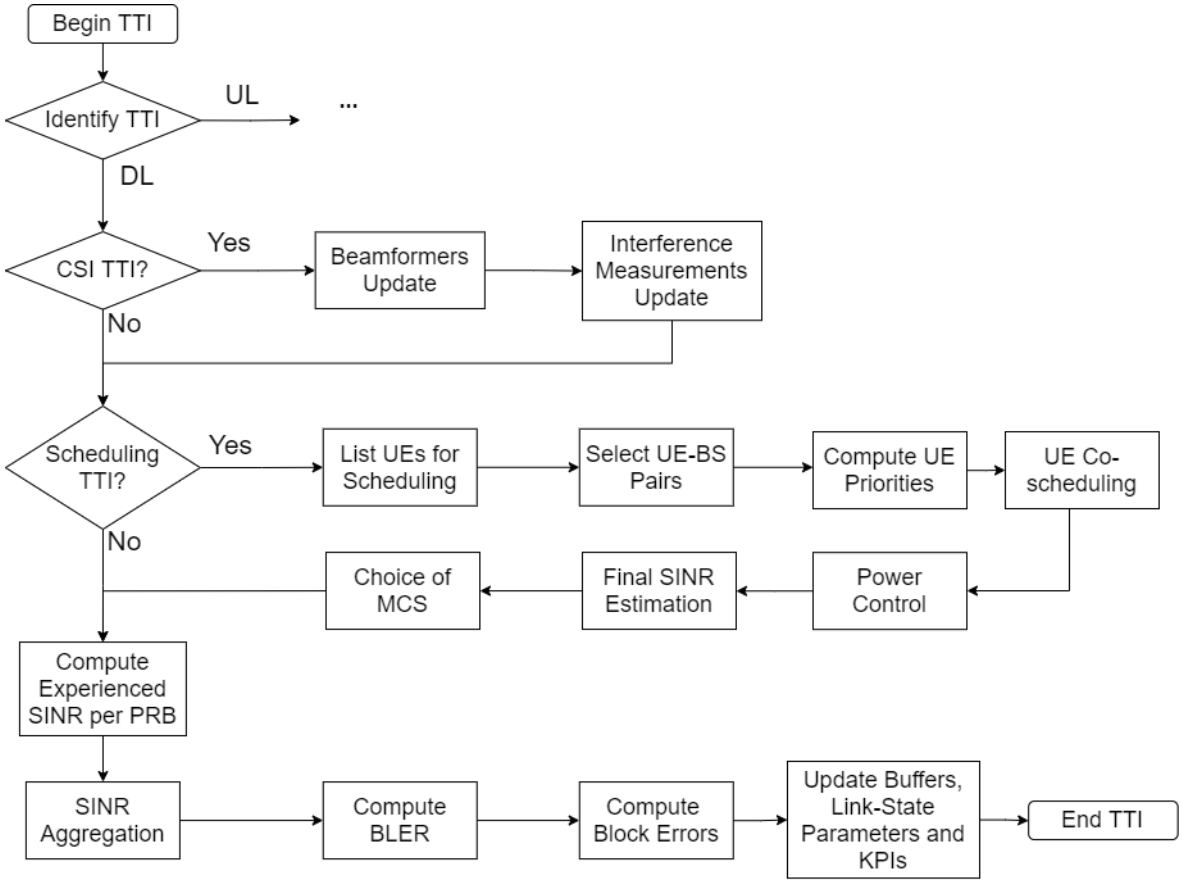


**Figure 1.1:** Flowchart for of simulation steps for each TTI.

Several verifications are made to decide whether some procedures should take place. The first is to identify the nature of the current TTI - UL and DL TTIs have different steps. The second is checking whether CSI should be updated. Thirdly, it is to verify whether the current user scheduling information for that TTI is to be updated. Only after those verifications and respective procedures, the transmissions scheduled for the present TTI are processed.

We start by assessing the nature of the TTI. It depends on the slot-structure and TDD split.

## TDD Split and Slot Format

We recognise two options. The first is to use self-contained slots, at a cost of about $2/14 \approx 14\%$ lower bit rate since 2 out of 14 symbols are used for guard and control, but having the benefit of feedback about block errors in the same TTI, thus allowing triggering retransmissions of the lost information the next TTI. This way the likelihood of packet dropping due to transgressions of time constrains is reduced since latency is reduced, leading to more opportunities to transmit the data on time. The second is simpler and more throughput-efficient, at the cost of latency performance. It consists on using slots that only have DL/Uplink (UL) symbols, respectively, and we ignore the guard time in the transition slot.

Therefore our definition of UL/DL split, or Time Division Duplex (TDD) split depends on the option. Let us define $s_{TDD}$ as the ratio between UL and DL slots. We represent this ratio as $N_{slots}^{DL} : N_{slots}^{UL}$, e.g. 4:1, meaning that for each UL slot there are 4 DL slots. The slot structure is completely defined by the number of slots in a transmission period $N_{slots}^{TDD}$.

In order to optionally change between both options, we introduce a transport block acknowledgement delay $\tau_{ACK}$ (in TTIs) and a slot efficiency $\eta_{slot}$. The acknowledgement delay is the number of TTIs before the transmitter receives the acknowledgement, thus $\tau_{ACK} + 1$ is the number of TTIs until the erroneous transport block can be transmitted again. With self-contained slots, $\tau_{ACK} = 0$. Without self-contained slots it depends on the $s_{TDD}$.

The slot efficiency $\eta_{slot} = 0.86$ in the example of self-contained slots where 14% of symbols are not used for data, and $\eta_{slot} = 0$ in the DL/UL heavy slots. It is applied to the instantaneous throughput $R$ as $R_{modified} = R\eta_{slot}$

As mentioned, we solely present, and posteriorly evaluate, modelling for DL TTIs. Thus after making the distinction between TTIs, the next step is to update the CSI information based on our beamforming strategy. Therefore, let us first state how the GoB is created.

## Grid of Beams

To create a GoB we need to know which directions to steer the beam. The beam-steering directions are all possible combinations of values in the azimuthal and elevation angular domains, relative to the antenna boresight (direction perpendicular to the plane the antenna array is inserted). And to create a beam grid in one such domain, one simple way is to use the resolution and the values of the extremes. We define in Equation (1.1) an interpolation function to perform the operation of creating

a set of values from $a$ to $b$, given $b$ strictly greater than $a$, with intervals of resolution $r$.

$$F_I(a, b, r) = \{a + i \times r \; \forall \; i \in \mathbb{N}_0 : i \times r \leq b - a\} \tag{1.1}$$

This way, we define in the azimuthal angular domain as $\mathcal{A}_\phi = F_I(a_\phi, b_\phi, r_\phi)$ and the elevation angular domain as $\mathcal{A}_\theta = F_I(a_\theta, b_\theta, r_\theta)$. For instance, if the antenna is positioned in the centre of the room, on the ceiling, pointing downwards, then the most logical approach is a symmetric approach because in that position the coverage of the room would be uniform since we consider our room with equal length and width(for room and user behaviour modelling, see Section **??**). More concretely, the GoB should cover all positions the UEs may potentially be. Thus, given the position and movement of the users in relation to the size of the room described in the example of Section **??**, choosing the lower limits to $a_\phi = a_\theta = -60°$ and the upper limits to $b_\phi = b_\theta = 60°$ covers all possible UE positions.

The resolutions should depend on the array size. To create a pseudo-non-interfering GoB, where the maximum of the main lobe of one beam points at the a minimum of an adjacent beam, the resolution should be roughly half the First Null Beam Width (FNBW). It is 'pseudo-non-interfering' because the FNBW varies with the direction at which the beam is steered, which causes the maximums to not align perfectly with the nulls. This effect is unnoticeable in adjacent beams, and gets more noticeable the more far apart beams are from each other. So, this method is a simplistic yet effective approach to minimise the interference between beams, but it does not eliminate this interference.

Thus, the possible directions are defined as a cartesian product between the azimuthal and elevation domains, shown in Equation (1.2).

$$\mathcal{D} = \mathcal{A}_\phi \times \mathcal{A}_\theta = \{(\phi, \theta) : \; \phi \in \mathcal{A}_\phi, \; \theta \in \mathcal{A}_\theta\} \tag{1.2}$$

Having the directions, we need the precoder that will construct a beam pointing in that direction. In Equation (1.3) we define the $M$ by $N$ beamforming matrix $\boldsymbol{W}_{\phi,\theta}$ that contains the relative amplitudes and phases that are applied to the signal of each antenna element of an $M$ by $N$ planar array, obtaining as a result a beam directed to $\phi$ degrees on the horizontal plane and $\theta$ degrees on the vertical plane. Note that such planes depend on the orientation of the array and the angles $\phi$ and $\theta$ are null in the interception of both planes, corresponding to the direction orthogonal to the array plane (see Appendix **??** for a complete derivation).

$$\boldsymbol{W}_{\phi,\theta} = \begin{bmatrix} 1 & u_2 & \dots & u_2^{(N-1)} \\ u_1 & u_1 u_1 & \dots & u_1 u_2^{(N-1)} \\ \vdots & \vdots & \ddots & \vdots \\ u_1^{(M-1)} & u_1^{(M-1)} u_2 & \dots & u_1^{(M-1)} u_2^{(N-1)} \end{bmatrix}, \text{ with } \begin{cases} u_1 = e^{-j\pi \sin(\phi)\sin(\theta)} \\ u_2 = e^{-j\pi \cos(\phi)\sin(\theta)} \end{cases} \tag{1.3}$$

Subsequently, to obtain every precoder in the GoB we need to build a precoding matrix for each direction in $\mathcal{D}$. Let us define in Equation (1.4) the set $\mathcal{W}$ containing all precoders $\boldsymbol{W}_{\phi,\theta}$ in the GoB, formed for an $M$ by $N$ Uniform Rectangular Array (URA).

$$\mathcal{W}^{\text{GoB}} = \{\boldsymbol{W}_{\phi,\theta} : (\phi,\theta) \in \mathcal{D}\} \tag{1.4}$$

As a last step, we vectorise the matrix $\boldsymbol{W}_{\phi,\theta}$ into a vector $\boldsymbol{w}_{\phi,\theta}$, to facilitate its usage. This process is exactly the same as stacking the columns of $\boldsymbol{W}_{\phi,\theta}$.

Figure 1.2 illustrates the result of a cut at zero degrees elevation on beams of two grids. The two grids are built for square antenna arrays, with 16 and 1024 elements, respectively, left and right sides of the figure, hence the noticeably different directivity. Using the 3GPP-defined elements in [1], the maximum directivities are 20 dBi and 38 dBi, respectively, for the 16-element array and for the 1024-element array. Furthermore, since the resolutions were purposely set to match half of the FNBW, the grid on the left spans 120° of angular domain, from -60° to 60°, with steps of 30°, while the grid on the right does so with a resolution of 4°. In total, this equates to 25 distinct beams of the small array and 961 beams in the larger array.

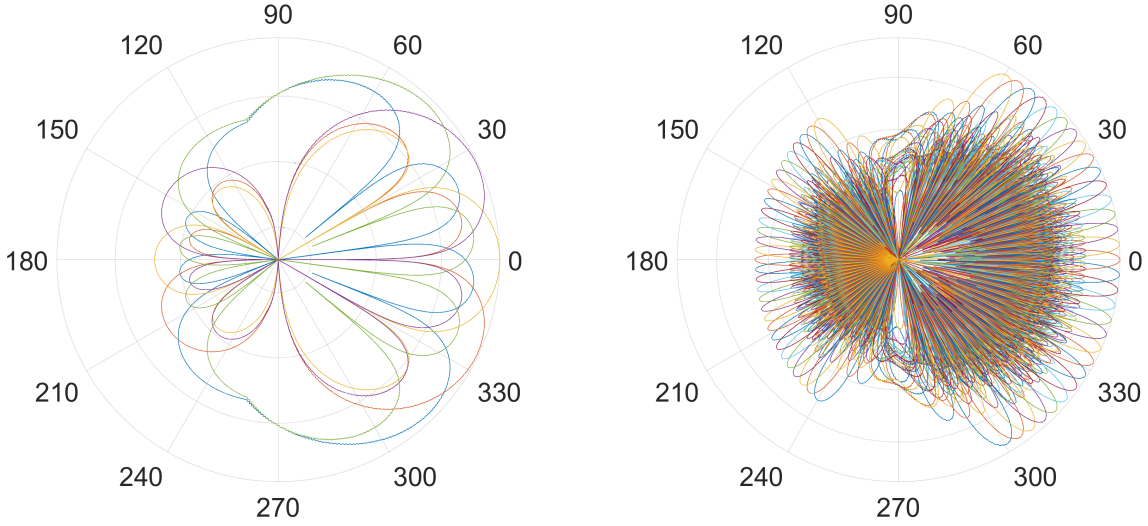


**Figure 1.2:** GoB azimuth cuts for 4 by 4 (left) and a 32 by 32 (right) antenna element array.

### 1.1.1 Channel State Information

CSI updates happen every $N_{slots}^{CSI}$ slots or TTIs. Not all TTIs have CSI updates because the channel does not change enough to be worth updating that frequently, and due to prohibitive overheads since reference signals are sent in place of data.

The overheads associated with different Channel State Information (CSI) feedback schemes is not modelled. The overhead would depend on the type and quality of said measurements, thus we simply define a CSI-slot efficiency $\eta_{CSI}$ meant to reduce the bit rate of CSI slots.

CSI is required for operation of two important mechanisms. First, to direct and receive signals optimally, in accordance with the paths where attenuation is lower. As such, it is used to update matching beamformers at the transmitter and receiver, or beam pairs. Second, to assess received power and interference, which are crucial to estimate channel quality, which is then used to, e.g. determine which MCS to use.

Let us address the beam pair establishment first. Our formulation holds beam correspondence [2], this means the beam computed for the transmitting are used for receiving as well. Therefore we refer to weights vectors as beamformers, instead of the direction-specific nomenclatures like precoder or combiner.

#### Beam pairs Update

To update the best beam pairs between UEs and BS panels, the BS should transmit $N_{CSI}$ CSI-RSs precoded in GoB beams and the UE reports how well it received each RS. However, this would require a mechanism for the BS to identify, based on previous channel measurements, which beams are more likely to best serve the UE. As such, instead we check all beams in the GoB to assess which best suit the channel. Furthermore, we keep received power information about $N_{CSI}$ of them, which is useful for future SINR estimations.

The best beam pairs are chosen to maximise the channel gain achieved from performing a transmission with a given GoB beam, with a best effort reception using MRC. Therefore, for a link between UE $u$ and BS panel $b$, the beamformer on the BS side $\boldsymbol{w}_{bu}^{BS}$ is always a $N_{ant}^{BS} \times 1$ beam-steering vector from the GoB, i.e. $\boldsymbol{w}_{bu}^{BS} \in \mathcal{W}_{b}^{GoB}$. The UE-side beamformer $\boldsymbol{w}_{bu}^{UE}$ is always the Maximum Ratio (MR) beamformer that fits the BS beamformer used over the $N_{ant}^{UE} \times N_{ant}^{BS}$ channel $\boldsymbol{H}_{bu}$. As such, the received signal in each of the $N_{ant}^{UE}$ UE antennas is $\boldsymbol{H}_{bu} \cdot \boldsymbol{w}_{bu}^{BS}$. Here, $N_{ant}$ refers to the number of single-polarised antenna elements. The computation of the UE-side beamformer is given in Equation 1.5, from using Equation **??**.

$$\boldsymbol{w}_{bu}^{UE} = \frac{\left(\boldsymbol{H}_{bu} \cdot \boldsymbol{w}_{bu}^{BS}\right)^{H}}{|\boldsymbol{H}_{bu} \cdot \boldsymbol{w}_{bu}^{BS}|} \tag{1.5}$$

When $\boldsymbol{w}^{UE}$ is a MR beamformer, the channel gain under transmit and receive beamforming is a real number. Therefore, to choose the $\boldsymbol{w}$ that achieves the highest gain, we simply have to choose $\boldsymbol{w}$ that results in highest norm of its internal product with the channel. This shortcut is represented in Expression (1.6). In essence, this means that because we are computing the UE-side beamformer already taking into account the transmit-side beamformer, to maximise the norm of the received signal it is sufficient to choose the appropriately the transmit-side beamformer.

$$\boldsymbol{w}_{bu}^{BS} = \underset{\boldsymbol{w} \in \mathcal{W}_{b}^{GoB}}{\operatorname{argmax}} \left|\boldsymbol{w}_{bu}^{UE} \cdot \boldsymbol{H}_{bu} \cdot \boldsymbol{w}\right| = \underset{\boldsymbol{w} \in \mathcal{W}_{b}^{GoB}}{\operatorname{argmax}} |\boldsymbol{H}_{bu} \cdot \boldsymbol{w}| \tag{1.6}$$

When $N_{CSI} > 1$, instead of the best beamformer, we save the $N_{CSI}$ best GoB beamformers. For sake of practicality, let us assume $N_{CSI} = 1$ for now on. Furthermore, beam pairs computed in this way profit from beam-reciprocity, i.e. the beams used for receiving can be used for transmitting as well. And doing this way, the received power is already present (see Equation (1.7)), thus we only need to update the interference now.

$$P_{r,bu}^{UE} = P_{t,bu}^{BS} \left|\boldsymbol{w}_{bu}^{UE} \cdot \boldsymbol{H}_{bu} \cdot \boldsymbol{w}_{bu}^{BS}\right|^{2} \tag{1.7}$$

**Interference Measurements Update**

To measure interference, the Base Station (BS) should schedule an empty UE-specific RS for interference measurements. It should result in measuring the power received by the interfering sources. The main drawback is the outdatedness of the measurement. It takes around 4 TTIs until the information is available since it needs to be sent, received, processed and fed back. Therefore, when the interference measurement is available, it refers to $\tau_{\text{TTI}}$ TTIs back, e.g. 4 TTIs ago.

A major disadvantage of estimating the interference in this manner comes from the fact that the experienced interference is extremely dependent on current scheduling. If the scheduled UEs or beamformers in use change, then it is expected that a major change in the experienced interference takes place, thus possibly rendering the measurement completely invalid. We foresee precise interference estimation algorithms, perhaps driven by learning mechanisms, to be a future direction of work. We discuss this matter further in Section **??**.

### 1.1.2 User Scheduling

Analogous to the CSI update procedure, the scheduling information is only updated every $N_{slots}^{SCH}$ TTIs. In a first stage, we renovate the scheduling information on which UEs are considered for scheduling and which BS panels are used for each UE. In essence, only UEs with non-empty buffers are examined to potentially be part of the scheduled list; and each UE is served by a single BS panel with the best beam pair to that UE. This constitutes the simplest panel selection scheme.

The scheduling process then continues to estimate SINRs and achievable throughputs for each UE, to compute UE priorities, to make user co-scheduling decisions, assign powers for each transmission and derive MCS to be used. The SINR estimation step is presented first.

#### SINR Estimation

The received powers for the best $N_{CSI}$ beams have been reported in the CSI acquisition step, as well as the interference levels computed from experienced interference from $\tau_{CSI}$ TTIs ago. Also, the channel gain can be derived directly knowing the transmit power that was used. Thus, we assume an equal distribution of the maximum transmit power at the BS $P_{t,max}^{BS}$ over the number of scheduled UEs with non-empty buffers. And the only missing piece in the SINR expression is the noise.

We use wideband scheduling, i.e. allocating all available spectrum to every transmission, relying on spatial separation to prevent excessive interference. Therefore, assuming $B$ to be the system bandwidth, using thermal noise we get a noise power $P_N$ given by Equation (1.8), with the Boltzmann constant $k_B = 1.380649 \times 10^{-23}$ J/K, the noise temperature $T$ and an upscaling with the receivers' noise figure $NF_r$. All hardware imperfections are abstracted by considering noise figures in the BS and in the UEs, respectively, $NF_{BS}$ and $NF_{UE}$, in dB.

$$P_N = k_B T B \times 10^{\frac{NF_r}{10}} \tag{1.8}$$

To summarise, the expression used for SINR estimation uses information from $\tau_{CSI}$ TTIs ago on the received power $P_s$ and total interference $I$. See Equation 1.9

$$\hat{SINR}_{eff} = \frac{\hat{P}_s}{\hat{I} + P_N} \tag{1.9}$$

### Instantaneous Throughput

To compute the instantaneous throughput, we need to quantify the value of serving each user. Then we can weigh transmission options regarding fairness, maximum aggregated throughput, or likelihood of fulfilling latency constraints. It is also needed to calculate the estimated and realised bit rates.

The SINR is used to choose which CQI should be reported from the BLER curves represented in Figure 1.3, with equations in Appendix **??**. The point at which each curve intercepts the BLER probability of 10% is marked. The corresponding MCS choice consists on selecting the highest MCS that achieves a lower percentage of block errors than the Block Error Rate (BLER) target $BLER_0$.

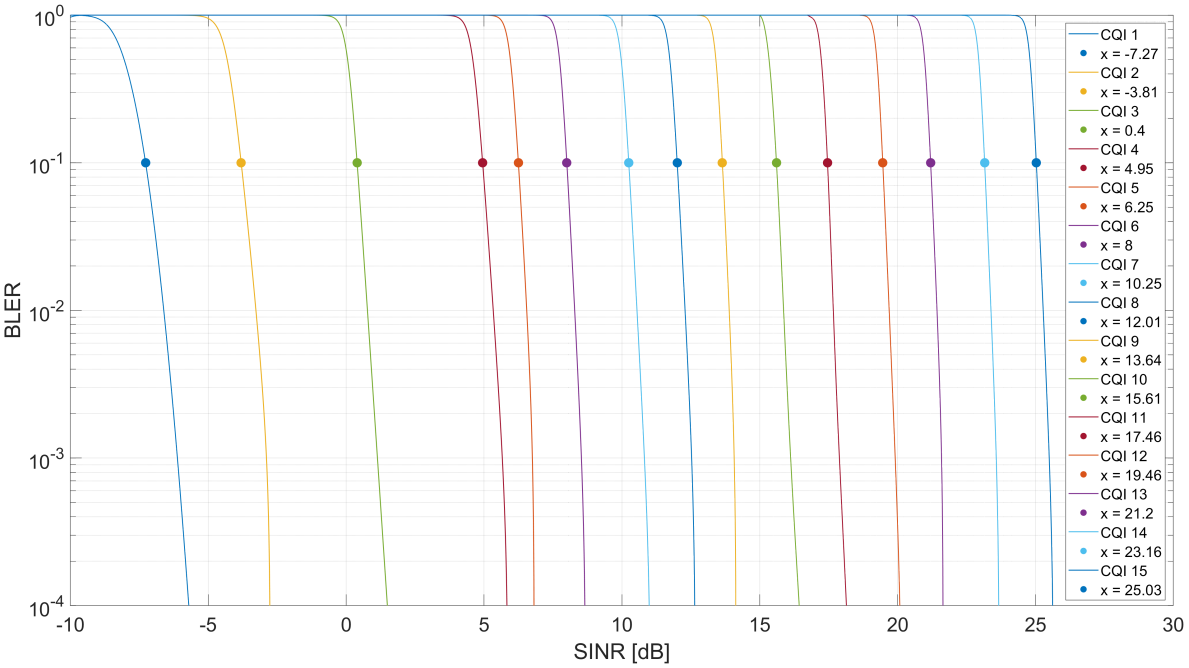


**Figure 1.3:** BLER curves for all MCSs. Simulated with Vienna Link-Level Simulator [3].

The selected MCS is then adjusted with the OLLA parameter as described in Subsection 1.1.2. The resultant MCS tells us the number of bits in a symbol $N_{bits}^{symb}$, which can be computed from the modulation order M, through $\log_2(M)$. To compute the bits per PRB we assume initially that all REs in a PBR are used for data, i.e. $N_{symb}^{PRB} = 168$, we multiply by $N_{bits}^{symb}$ and take into account the code rate $R_c$. In essence, Equation (1.10) computes the bit rate by dividing the number of bits transmitted in a PRB by the duration of that PRB, which corresponds to a slot duration $T_{slot,\mu}$, that depends on the numerology $\mu$.

$$R_b = \frac{N_{bits}^{symb} \times N_{symb}^{PRB} \times R_c}{T_{slot,\mu}} \tag{1.10}$$

To balance the excessively optimistic assumptions, like assuming all symbols are used for data, we adjust to the bit rate, namely due to signalling overheads and self-contained slots, respectively, by multiplying the efficiencies $\eta_{OH}$ and $\eta_{slot}$. Equation (1.11) has the final bit rate efficiency $\eta$. An estimation for the instantaneous throughput per TTI $R$ is $R = R_b \times \eta$.

$$\eta = \eta_{OH} \times \eta_{slot} \tag{1.11}$$

**Compute UE Priorities with Scheduler**

A scheduler task is to compute UE priorities $p$ according to a trade-off of resource sharing fairness, achieving the maximum instantaneous aggregated throughput, or attain the lower average latencies, to name a few. These priorities allow us to select UEs by order of importance according to the weighted trade-off relation we choose.

The most common and widely used scheduler is the Proportional Fair (PF), presented in Equation (1.12). PF takes the ratio between the estimated instantaneously attainable throughput $\hat{R}$ and average attained throughput $\overline{R}$ to balance immediate reward and fairness across users, for each TTI $t$. The average $\overline{R}$ is computed using exponential smoothing with a parameter $t_w$, according to Equation (1.13).

$$p(t) = \frac{\hat{R}(t)}{\overline{R(t)}} \tag{1.12}$$

$$\overline{R}(t) = \left(1 - \frac{1}{t_w}\right) \overline{R}(t-1) + \frac{1}{t_w} R(t-1) \tag{1.13}$$

As seen, PF does not consider latencies. Yet, for our case where each user has the same amount of data to receive (and each camera the same amount of data to transmit), the PF also levels latencies by weighting fairness, not leaving any user waiting for long. However, it may not perform as well as latency-aware alternatives.

Two latency-aware alternatives are Exponential/Proportional Fair (EXP/PF) [4] and Maximum-Largest Weighted Delay First (M-LWDF) [5]. The latter is almost as simple as the PF, only weighting the Head Of Line (HOL) latency as well. EXP/PF is more complex and considers a maximum delay and increases priorities exponentially as latencies approach the limit. Both use the PF ratio described in Expression (1.12). M-LWDF outperforms EXP/PF in practically every scenario, besides when the load is very high [5]. Therefore, both M-LWDF and EXP/PF seem worthwhile alternatives, but we choose the PF for this work.

### Co-schedule users

This step lists the users to be scheduled together until the next update to the schedule. The co-scheduling rule for a single-BS-panel operation is to add one UE layer at a time to the list, by order of UE priority (computed in the previous step), if the best beams used for those layers are compatible with the previously added UE layers. And we define as compatible beams when the BS-side beam, belonging to the GoB is at least $\kappa$ beams apart, with $\kappa \in \mathbb{N}_0$. If $\kappa$ is 0, then all layers are accepted. If $\kappa = 1$, then the beams must be different - adjacent beams have a distance of 1, so are still used together. Beams located diagonally adjacent of the GoB are considered to have a distance of 2, hence they may be co-scheduled when $\kappa \leq 2$. Figure 1.4 illustrates the beams that cannot be co-scheduled with certain values of $\kappa$, representing in filled blue circles as incompatible beams with respect to the orange one, and empty circles as compatible beams with the central orange beam. More generally, the beam distance is defined by the sum of absolute differences of the beam indices in the grid. Mathematically, the beamformers $\boldsymbol{w}_{i,j}$ and $\boldsymbol{w}'_{i',j'}$, having $(i, j)$ and $(i', j')$ as the GoB indices, respectively, are compatible if $|i - i'| + |j - j'| \geq \kappa$.

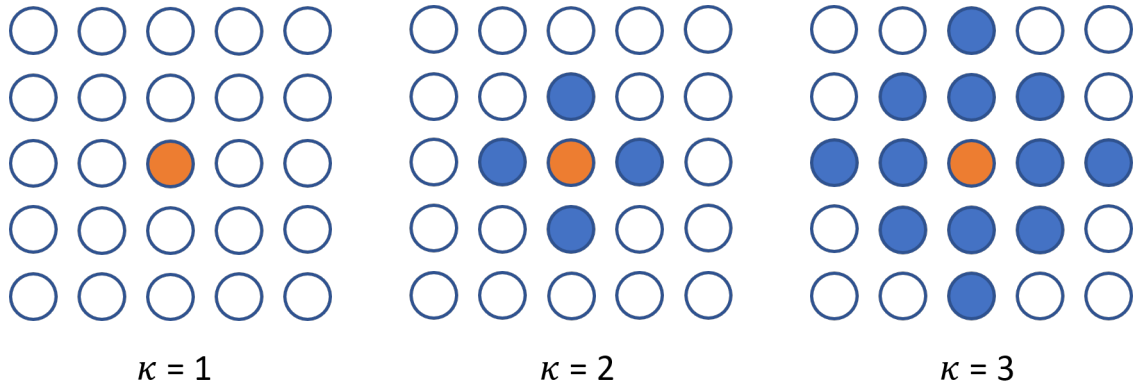


**Figure 1.4:** Beam co-scheduling incompatibility distance.

In this step there is space for more elaborate algorithms that attempt to choose different combinations of the best $N_{CSI}$ beams of each user, in an attempt to maximise the metrics we care about. Of course, if scheduling one more user considerably reduces the quality of the channel to many others, it is likely not worth doing.

### Power Control

Depending on the beamforming strategy, it may be necessary to scale down all precoders due to excessive power per antenna constraints. This, however, does not apply to our case because beam-steering beamformers always have uniform amplitude. Power control in the downlink is as simple as distributing the maximum total transmit power equally amongst the scheduled UEs and assigned PRBs.

### MCS selection

By now the transmit power to each user and the beam pairs are fixed. Therefore, we obtain an estimation of the SINR for each user, as done previously but with the new information on powers and beam pairs. And we obtain a MCS to be used for each transmission, according to the same process used in the throughput estimation.

The choice of which MCS to use for transmission has major impact in performance. Since choosing one MCS above what the channel quality allows can lead to excessive errors, while one lower than ideal will wastes resources. Factors such as the outdatedness of measurements complicate the process of choosing the correct MCS. We present a mechanism that adapts the MCS choice according with block errors. This procedure takes place every time a MCS is used, including in instantaneous throughput calculations.

It is a Outer Loop Link Adaptation (OLLA) mechanism [6] and it is UE-specific. When a MCS is estimated, it is subsequently adjusted with the OLLA parameter. The OLLA parameter $\Delta_{OLLA}$ is initialised at zero and is updated in every TTI the given UE is scheduled. When a Transport Block (TB) is successfully transmitted, the OLLA parameter is updated with a step-size of $\gamma_{OLLA}$ according to Equation (1.14) such that the BLER long-term average converged to the target $BLER_0$. If the block is erroneous, Equation (1.15) is used instead.

$$\Delta_{OLLA} = \Delta_{OLLA} + BLER_0 \times \gamma_{OLLA} \tag{1.14}$$

$$\Delta_{OLLA} = \Delta_{OLLA} - (1 - BLER_0) \times \gamma_{OLLA} \tag{1.15}$$

Observe the subtlety of the asymmetry in update. The term that multiplies the step size $\gamma_{OLLA}$ is much bigger in Equation (1.15) than in (1.14), since $BLER_0$ is usually 0.1 or smaller, depending on the QoS reliability requirements. It is a defensive approach, to take bigger steps towards more conservative MCSs when there are errors because it is always better to have some throughput than none. Contrarily, the progression to increasing the MCS is slower.

The OLLA parameter adjusts the MCS choice by flipping an appropriately biased coin and adding either $\lfloor \Delta_{OLLA} \rfloor$ or $\lceil \Delta_{OLLA} \rceil$ to the MCS index estimated in the previous step. An appropriately biased coin in this situation is a coin that selects to round down the OLLA parameter with a probability of $\lceil \Delta_{OLLA} \rceil - \Delta_{OLLA}$. This makes sense because $\Delta_{OLLA}$ is decreased when a block has errors, thus making more likely that the MCS is reduced when the link has worse quality than expected. When the block

does not have errors, it makes it more likely to increase the MCS estimate, such that a good link condition can be taken advantage of to increase the bit rate. Note that this formulation still works as supposed for negative values, i.e. the OLLA mechanism works for increasing and decreasing the MCS.

### 1.1.3 Transmission Realisation

Here we obtain the outcomes of the realised transmissions. Firstly, we calculate the number of TBs in which the data to be transmitted in a given TTI is segmented. Secondly, the SINRs each UE experiences in each PRB are computed and then we present how these SINRs can be aggregated in something more easily useable to conclude on overall channel quality, an effective SINR. Finally, effective SINRs are probabilistically used to determine the success or failure of the transmitted transport blocks according with the MCS used for transmission and then the link quality adaptation mechanism is updated appropriately, as well as buffers and PF ratios, to be used in upcoming transmissions to assure a balanced operation of the system in line with the result of the transmission in the present TTI. As usual, the steps follow.

#### Transport Block Size Calculation

To obtain the Transport Block Size (TBS), essentially two ways have been modelled. The first is to consider the same number of TBs on every transmission, $N_{TB}$. Therefore the numbers of bits to be transmitted $N_{bits,bul}$ is divided equally over TBs and the size of each TB is the same, see Equation (1.16).

$$S_{TB} = \lceil N_{bits,bul}/N_{TB} \rceil \tag{1.16}$$

The second is to consider a maximum TBS $S_{TB,max}$, obtain $N_{TB}$ from Equation (1.17) and then use Equation (1.16).

$$N_{TB} = \lceil N_{bits,bul}/S_{TB,max} \rceil \tag{1.17}$$

This such manner, $N_{TB}$ TBs are sent and the experienced bit rates depend on how many of them are delivered with no errors. If there are no errors, the bit rate computed in Equation (1.10) is achieved, otherwise only a fraction of that bit rate is achieved, corresponding to the successfully transmitted TBs over total TBs. One of the modelled methods is chosen by fixing either $N_{TB}$ or $S_{TB}$, respectively, for the first and second methods.

Another alternative way is to follow an extensive list of steps described in [7], making the Transport Block Size depend on the number of layers $\#\mathcal{L}_{bu}$ carrying the same

QoS flow, modulation order $M$, code rate $R_c$, number of allocated PRBs $N_{PRB,bul}$ and transmission duration, which we assume to be always $T_{slot}$.

### Compute Realised SINR: A Multi-layer SINR Framework with Beamforming

Although the rest of this chapter assumes simplifications for downlink single-layer transmission, for future purposes we derive a general multi-layer framework that works for uplink as well.

To accurately compute the SINR experienced during a transmission, we need to know the power received from each transmitter, for any scheduled UE, taking into account the different channel responses in each PRB of the assigned bandwidth.

Let $l$ be the layer that links a set of antennas in BS $b$ to a set of antennas in a UE $u$, with $P_{t,l}$ the total transmit power and $P_{r,l}$ is the received power in that layer, after combining the contributions of each receive antenna. Then, let $P_{r,ll'}$ be the power received by layer $l$ receiver using combiner $\boldsymbol{w}_{r,l}$, transmitted by layer $l'$ transmitter using precoder $\boldsymbol{w}_{t,l'}$, with $\boldsymbol{H}_{ll'}$ the channel matrix that connects the receiver and the transmitter. Equation (1.18) shows how these quantities relate.

$$P_{r,ll'} = P_{t,l'} \left| \boldsymbol{w}_{r,l} \cdot \boldsymbol{H}_{ll'} \cdot \boldsymbol{w}_{t,l'} \right|^2 \tag{1.18}$$

The powers are scalars, $\boldsymbol{w}_{r,l}$ is a $1 \times N_r$ vector, $\boldsymbol{w}_{t,l'}$ is $N_t \times 1$ vector and $\boldsymbol{H}_{ll'}$ is a $N_r \times N_t$ matrix, where $N_t$ and $N_r$ are the number of antenna elements at the transmitter and receiver antenna arrays, respectively.

Knowing how to calculate this quantity we can compute the powers of all parts of the SINR expression on a PRB basis: the signal $P_s$, the intra-cell interference $P_{IaCI}$, the inter-cell interference $P_{IeCI}$, the inter-layer interference $P_{ILI}$ and the noise $P_N$. Here, the term cell refers to a panel. See in Equation (1.19) the expression for the SINR of a specific layer $l$ in a given PRB. Subsequently we present equations for each quantity in the SINR expression, along with the rationale behind them.

$$SINR = \frac{P_s}{P_{ILI} + P_{IaCI} + P_{IeCI} + P_N} \tag{1.19}$$

Moreover, and to reiterate, all quantities mentioned in this section are time (TTI) and frequency (PRB) specific. These SINRs need to be posteriorly aggregated in an effective SINR for each transmission in the given TTI. We choose to omit the $i$ index to simplify notation, as we did with the TTI since this chapter is TTI-specific.

The received signal power $P_s$ is in Equation (1.20).

$$P_s = P_{r,ll} \tag{1.20}$$

In case of multi-layer transmission, other layers scheduled to/from the same UE may interfere among themselves. The power of inter-layer interference $P_{ILI}$ takes into account this interference by summing the interferences caused in layer $l$ by every other layer $l'$ scheduled between BS $b$ and UE $u$. See Equation (1.21), where $\mathcal{L}_{bu}$ is the set of layers scheduled between BS $b$ and UE $u$.

$$P_{ILI} = \sum_{\substack{l' \in \mathcal{L}_{bu} \\ l' \neq l}} P_{r,ll'} \tag{1.21}$$

Interference power contributions from the same cell/BS come from every transmission that takes place to other UEs in the same cell/served by the same BS. See Equation (1.22), where $\mathcal{U}_b$ is the set of users served by BS $b$.

$$P_{IaLI} = \sum_{\substack{u' \in \mathcal{U}_b \\ u' \neq u}} \sum_{l' \in \mathcal{L}_{bu'}} P_{r,ll'} \tag{1.22}$$

Interference contributions from outside the cell come from all non-serving BSs, all UEs and in all layers. We see this Equation (1.23), where $\mathcal{B}$ is the set of all BS (or BS panels) in the system.

$$P_{IeCI} = \sum_{\substack{b' \in \mathcal{B} \\ b' \neq b}} \sum_{u' \in \mathcal{U}_{b'}} \sum_{l' \in \mathcal{L}_{b'u'}} P_{r,ll'} \tag{1.23}$$

The noise power $P_N$ is computed according to Expression 1.8, using the bandwidth of a single PRB, which depends on the numerology as evidenced in Table **??**.

This framework is also applicable when several BS are jointly serving one user, or when one user is transmitting to several BS simultaneously, i.e. Distributed-MIMO (D-MIMO). This is true because we simply account for power contributions, abstracting from the content of the spatial streams.

#### Aggregate SINRs: Mutual Information Effective SINR Mapping

MI-ESM is an SINR aggregation technique that allows us to attribute one SINR to a transmission where the quality of the channel varies across the transmission band, namely across PRBs. We choose this SINR mapping strategy because [8–12] show

that it unquestionably achieves very good results without the need of calibration for different MCSs. Equation (1.24) sums how it works.

$$SINR_{eff} = I_k^{-1}\left(\frac{1}{N}\sum_{i=1}^{N} I_k\left(SINR_i\right)\right) \tag{1.24}$$

Above, $I_k$ is the mutual information function that for a given SINR and MCS (with $k$ bits per symbol) gives the bits of information that are conceivably extracted for a transmission with that SINR. For low SINRs, the mutual information is practically zero. As the SINR grows, the quantity of information bits extracted approaches $k$. Appendix **??** goes into further detail on the mutual information function works.

Therefore, Equation (1.24) obtains the mutual information achievable in each PRB, averages it and computes the SINR that would achieve that average information. Thus, the effective SINR is determined as the SINR that would yield this average mutual information if it were applied on all PRBs.

#### Compute Block Errors

Subsequently, with the effective SINR $SINR_{eff}$ and the MCS used for the transmission, we get the resultant $BLER$ from the correspondent MCS curve in Figure 1.3. Then we flip a BLER-biased coin to determine whether each block was received well.

#### Update Link Adaptation, Buffers and Performance Indicators

Firstly, the link adaptation mechanism is updated based on the block errors in accordance with Section 1.1.2.

Then, the information that was successfully transmitted needs to be removed from the buffers. We model an ordered buffer where the information in one transport block has a direct mapping to IP packets. Therefore if that TB gets lost, those packets with information carried in the lost TB stay in the buffer.

This means that block errors may cause packets to arrive out of order. This phenomenon is represented in Figure 1.5 where the size of a TB is set to the same size as a packet for illustration purposes. We see the bits in the transport blocks that did not arrive successfully are kept in the transmission buffer. Thus, if those bits are eventually successfully sent in the future, they would be out of order. Note that this is something common in packet networks. Successfully transmitted TBs get their share of packets removed from the buffers.

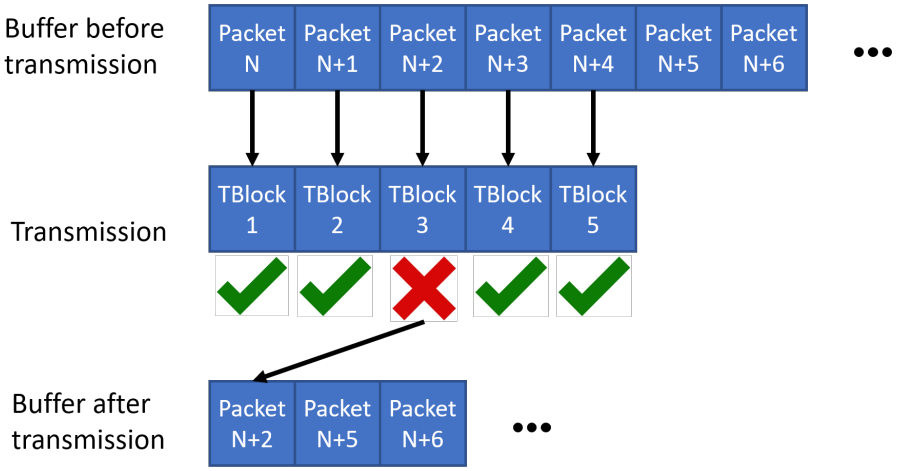


**Figure 1.5:** Buffer state before and after a transmission, with $S_{TB} = S_{packet}$.

These modelling considerations make the system considerably more realistic compared to a pool-of-bits formulation where no packet has a specific latency budget. Since a packet is dropped when the time it has passed since it arrived in the buffer exceeds the latency budget for the radio link, such realistic modelling considerations increase the likelihood that a packet is discarded due to excessive delay, so the latencies supported in a given scenario are higher.

Finally, the experienced throughputs are used to update the PF ratio according with Equation (1.13).

## Conclusion

In this section we followed the required steps to perform data transmission. We started with TTI identification and slot format. Then, we modelled CSI acquisition and presented an intuitive way of creating a GoB. Afterwards, we went through the user scheduling steps, where we addresses SINR estimation, instantaneous throughput calculation from an SINR, user priority calculation, user co-scheduling, power control and finally the selection of MCS for the transmission, adjusted with a link adaptation algorithm.

Finally, we simulate a data transmission, where we introduced calculations on the number and size of TBs. Then we presented a flexible SINR framework and a broadly accepted SINR aggregation algorithm, respectively, to compute the SINR each UE experienced in a given transmission in each PRB, and to aggregate those SINRs in one effective SINR that describes the quality of the transmission. Lastly, we used that SINR to determine the block errors, which are in turn used to adapt the link adaptation parameter, to remove the TB contents from the buffer in case it is successfully transmitted and to update the schedulers PF ratios.

In the next section we simulate single and multi-user scenarios and assess the relations between several parameters described in this section.

# Bibliography


[1] 3GPP, *TS 38.901 - 5G; Study on channel model for frequencies from 0.5 to 100 GHz.*, v16.1.0, Rel. 16, 2020.

[2] Y. R. Li, B. Gao, X. Zhang, and K. Huang, "Beam Management in Millimeter-Wave Communications for 5G and Beyond," *IEEE Access*, 2020.

[3] S. Pratschner, B. Tahir, L. Marijanovic, M. Mussbah, K. Kirev, R. Nissel, S. Schwarz, and M. Rupp, "Versatile mobile communications simulation: the Vienna 5G Link Level Simulator," *EURASIP Journal on Wireless Communications and Networking*, 2018.

[4] Jong-Hun Rhee, J. M. Holtzman, and Dong-Ku Kim, "Scheduling of real/non-real time services: adaptive EXP/PF algorithm," *The 57th IEEE Semiannual Vehicular Technology Conference, Spring*, 2003.

[5] F. Afroz, K. Sandrasegaran, and P. Ghosal, "Performance analysis of PF, M-LWDF and EXP/PF packet scheduling algorithms in 3GPP LTE downlink," *2014 Australasian Telecommunication Networks and Applications Conference (ATNAC)*, 2014.

[6] K. I. Pedersen, G. Monghal, I. Z. Kovacs, T. E. Kolding, A. Pokhariyal, F. Frederiksen, and P. Mogensen, "Frequency Domain Scheduling for OFDMA with Limited and Noisy Channel Feedback," *IEEE 66th Vehicular Technology Conference*, 2007.

[7] 3GPP, *TS 38.214 - 5G NR; Physical layer procedures for data.*, v15.3.0, Rel. 15, 2020.

[8] E. Tuomaala and Haiming Wang, "Effective SINR approach of link to system mapping in OFDM/multi-carrier mobile network," 2005.

[9] A. M. Cipriano, R. Visoz, and T. Salzer, "Calibration Issues of PHY Layer Abstractions for Wireless Broadband Systems," *2008 IEEE 68th Vehicular Technology Conference*, 2008.

[10] Z. Hanzaz and H. D. Schotten, “Performance evaluation of Link to system interface for Long Term Evolution system,” *7th International Wireless Communications and Mobile Computing Conference*, 2011.

[11] X. Li, Q. Fang, and L. Shi, “A effective SINR link to system mapping method for CQI feedback in TD-LTE system,” *IEEE 2nd International Conference on Computing, Control and Industrial Engineering*, 2011.

[12] J. C. Ikuno, “System Level Modeling and Optimization of the LTE Downlink,” *Ph.D. dissertation, TU Wien*, 2013.