



5G Extreme Requirements: Operators' views on fundamental trade-offs

by NGMN Alliance

Version:	1.0
Date:	28-November-2017
Document Type:	Final Deliverable (approved)
Confidentiality Class:	P - Public
Authorised Recipients: (for CR documents only)	

Project:	5G Extreme Requirements Task Force
Editor / Submitter:	Ilaria Thibault, Vodafone
Contributors:	Hakan Batikan, Turkcell; Anass Benjebbour, NTT DoCoMo; Feifei Lou, China Mobile; Wang Fei, China Mobile; Sebastien Jeux, Orange; David Lister, Vodafone; Kevin Smith, Vodafone; Alan Stidwell, Orange; Gustav Wikstrom, Ericsson;
Supporters:	Javan Erfanian, Bell Mobility; Naseem Khan, Verizon;
Approved by / Date:	NGMN Board, 21 November 2017

© 2017 Next Generation Mobile Networks Ltd. All rights reserved. No part of this document may be reproduced or transmitted in any form or by any means without prior written permission from NGMN Ltd.

The information contained in this document represents the current view held by NGMN Ltd. on the issues discussed as of the date of publication. This document is provided "as is" with no warranties whatsoever including any warranty of merchantability, non-infringement, or fitness for any particular purpose. All liability (including liability for infringement of any property rights) relating to the use of information in this document is disclaimed. No license, express or implied, to any intellectual property rights are granted herein. This document is distributed for informational purposes only and is subject to change without notice. Readers should not design products based on this document.

Commercial Address:

ngmn Ltd.,

Großer Hasenpfad 30 • 60598 Frankfurt • Germany

Phone +49 69/9 07 49 98-04 • Fax +49 69/9 07 49 98-41

Registered Office:

ngmn Ltd.,

Reading Bridge House • George Street • Reading •
Berkshire RG1 8LS • UK

Company registered in England and Wales n. 5932387,
VAT Number: GB 918713901

Abstract:

The aim of this work is to highlight what implications and trade-offs related to the delivery of new 5G services are relevant for mobile network operators. Some of these new services, in fact, require extremely low latency and high reliability of the communication link, which have very little in common with the targets that the telecommunications industry has worked towards until today. Mobile cellular networks have in fact traditionally been designed and optimised for the delivery of good voice and data services for mobile broadband customers, so the new 5G requirements now call for a re-think on how the future network will have to be designed and optimised in order to enable the new services.

The purpose of this document is to outline relevant trade-offs that need to be taken into account when delivering 5G services. In particular, the interplay among coverage, packet size, data rate, latency, and reliability, is analysed.

Document History

Date	Version	Author	Changes
10/08/2017	V 0.0	Ilaria Thibault, Vodafone	
24/08/2017	V0.1	Ilaria Thibault, Vodafone	
14/09/2017	V0.1	Alan Stidwell, Orange; Sebastien Jeux, Orange;	Comments and proposed changes added
22/09/2017	V0.2	Ilaria Thibault, Vodafone	Comments by Orange addressed, and more work to shape Sections 2, 3, and 4 has been done.
26/09/2017	V0.2	Alan Stidwell, Sebastien Jeux, Orange	Comments and clarifications
27/09/2017	V0.2	Wang Fei, China Mobile	Comments and proposed changes
29/09/2017	V0.2	Salih Ergüt, Turkcell	Comments and proposed changes
09/10/2017	V0.3	Ilaria Thibault, Vodafone	Harmonisation of contributions
10/10/2017	V0.4	Ilaria Thibault, Vodafone	Added DoCoMo as supporter, and addressed comments by DoCoMo on clarifying the assumptions for Eq. (3) and on the fading gain in the Annex.
21/11/2017	V0.4	Feifei Lou, China Mobile	Comments to the text
28/11/2017	V0.5	Ilaria Thibault, Vodafone	Addressed comments by China Mobile
28/11/2017	V1.0	Klaus Moschner, NGMN	Final clean-up for publication



Table of Contents

1	Introduction and motivation.....	4
2	Definitions and assumptions.....	5
3	Fundamental Trade-off analysis.....	7
4	Conclusion	12
5	Annex	12
6	Bibliography	16

1 INTRODUCTION AND MOTIVATION

New business opportunities for operators in a wide range of vertical industries (e.g., smart manufacturing, logistics, transportation, health, smart cities, agriculture, gaming, etc...) translate into new and sometimes challenging sets of targets that 5th-Generation mobile cellular networks need to meet to be able to successfully deliver the desired services. These targets include an evolution of traditional mobile broadband, which has been the main driver for network development until today, as well as requirements that are completely new to the cellular industry and that mainly address Internet-of-Things type of use cases.

In this context, a wide range of use cases with related business opportunities and required network capabilities was identified by NGMN in [1] and [2]. This work then became valuable input for 3GPP when it kicked off its own studies on new services for the next generation of mobile communications. 3GPP categorised all the different use cases into three main categories: massive Internet of Things [3], Critical Communications [4], and enhanced Mobile Broadband [5]. These studies then formed the basis for a single specification [6].

Massive-Internet-of-Things type use cases require the network to support very large numbers of connections for machine-type traffic; Critical Communications call for very low latency and highly reliable wireless access links in order for the network to deliver advanced functionalities for controlling objects; and enhanced Mobile Broadband use cases include data rich and immersive applications that rely on augmented and virtual reality features.

NGMN has recognised the need to gain deeper understanding in what impact these services will have on the future network architecture, both for the radio access and for the entire end-to-end network. Therefore, a task force on 5G Extreme Requirements was kicked off in May 2017. The new requirements are referred to as “extreme” since they go far beyond the boundaries of the traditional targets that have been the main driver for network design until today. This task force has the objective of answering the following questions:

- 1) To which extent can the 5G extreme services be delivered on existing deployments?
- 2) What modifications, if any, are required in the radio access network and/or in the core network to deliver the 5G extreme services?
- 3) How sensitive are the deployment models to the requirements? By relaxing the targets, does the deployment change considerably?

In order to answer the questions above, the 5G Extreme Requirements Task Force is structured into two main phases, which are mapped to a time line in Figure 1:

- **Phase 1: Operators’ view on fundamental trade-offs:**
This is a high-level study that provides preliminary insight for Question 1. The fundamental trade-offs among latency, reliability, packet size, data rate, and service coverage area are analysed. More detailed and rigorous analysis is the scope for Phase 2.
- **Phase 2: Network deployment for extreme services:**
The objective is to identify which network deployment models are best suited to address critical sets of requirements. This phase aims at answering in detail Questions 1, 2 and 3 and is broken down into two sub-phases that address radio access and end-to-end aspects respectively, as described below.
 - o **Phase 2.1: Radio Access Network deployment models:**
For a given set of services, associated with requirements on latency, reliability, throughput, and coverage availability, the required radio access network deployment is derived.
 - o **Phase 2.2: End-to-end considerations:**
This phase extends the scope of Phase 2.1 by identifying what affects latency and reliability in an end-to-end deployment and which changes and new features are required from an end-to-end network perspective to meet the targets associated to the services identified in Phase 2.1.

This report outlines the outcome of Phase 1 and identifies key challenges related to network deployment aspects that are relevant to operators and are addressed more in detail in Phase 2 of this task force.

a positive margin can be met with the baseline deployment with a surplus of resources: this means that by lowering, to the extent allowed by the given margin, bandwidth allocation, link budget, or terminal complexity, the target can still be met. A negative margin indicates that a target cannot be reached with the baseline deployment, and requires additional resources such as bandwidth, antennas, site density, power, or a combination of these.

- **Data rate:** The number of bits that can be transferred in a second on the wireless channel. This includes application layer data plus overhead. Application layer data rate will be used as a metric to understand the efficiency of the transmission, in other words, the ratio of the number of useful bits and the total number of bits that can be transferred on the channel within given latency and reliability constraints will be shown.

Table 1 Baseline deployment as defined in [7].

	Uplink
Duplexing mode	FDD
Deployment	Urban Macro
PER	0.1
Cell-Edge Spectral Efficiency [bps/Hz]	0.08
Bandwidth [MHz]	20
Diversity Order	1

3 FUNDAMENTAL TRADE-OFF ANALYSIS

This section presents simple numerical examples to visualise the interplay among latency, reliability, packet size, application layer data rate, and service coverage area. The focus is on the uplink, as its range is limited by the device capabilities, but discussion is provided for the downlink as well. The Annex provides details on the methodology used to derive the results presented below.

This preliminary analysis is use-case agnostic, and highlights what can be achieved in terms of delivering a piece of information, i.e., a packet, within given latency and reliability constraints within the baseline coverage described in Table 1. The packet size range and the values of latency have been chosen to identify trends and limits, and not, at this stage, to represent specific use cases. A value of achievable application layer data rate is associated to many of the scenarios on the graphs shown below. As mentioned in Section 2, it is important to note that this data rate does not include overhead and it only takes into account the useful amount of information that can be exchanged with a given set of resources. The extent to which the transfer of information is efficient, in terms of how many useful bits can be transferred in a packet of a given size, is also shown.

In the example in Figure 2, the “x” axis represents different packet sizes, ranging from 10 up to 400 Bytes and the “y” axis represents uplink coverage in dB, as defined in Section 2. Different curves correspond to different requirements on latency: 10, 1, and 0.1 ms respectively. PER, bandwidth allocation, and diversity order are fixed to the baseline values (Table 1). All the requirements that can be met with the baseline deployment are highlighted in yellow on the graph. As can be seen, as the low latency target becomes extreme, i.e., down to 0.1 ms, the baseline deployment is no longer adequate, regardless of the packet size. With a 1ms latency target, small packets up to 200 Bytes (according to the assumptions, 160 Bytes of application layer data plus 40 Bytes of TCP/IPv4 overhead) can be delivered on the baseline coverage, whereas all packet sizes in the considered range can be supported with a 10 ms latency target.

Different values of application layer data rate are mapped to different points of the graph in Figure 2. If low latency can only be supported for very small packets, the application layer information needs to be segmented into independent small chunks, each of which would need to be associated with its own TCP/IPv4 overhead. This may not always be an efficient way of transmitting information since in some cases, the amount of overhead associated with the TCP / IPv4 stack would use most or all the resources available in the given transmission time interval (TTI), and hence no application layer information can be sent within the TTI. The area highlighted in red on the graph emphasises this point: if a connection can support low latency for small packets only, no application layer data can be conveyed in the TTI under the assumptions considered in this analysis (i.e., the TCP/IP stack introduces 40 Bytes of overhead).

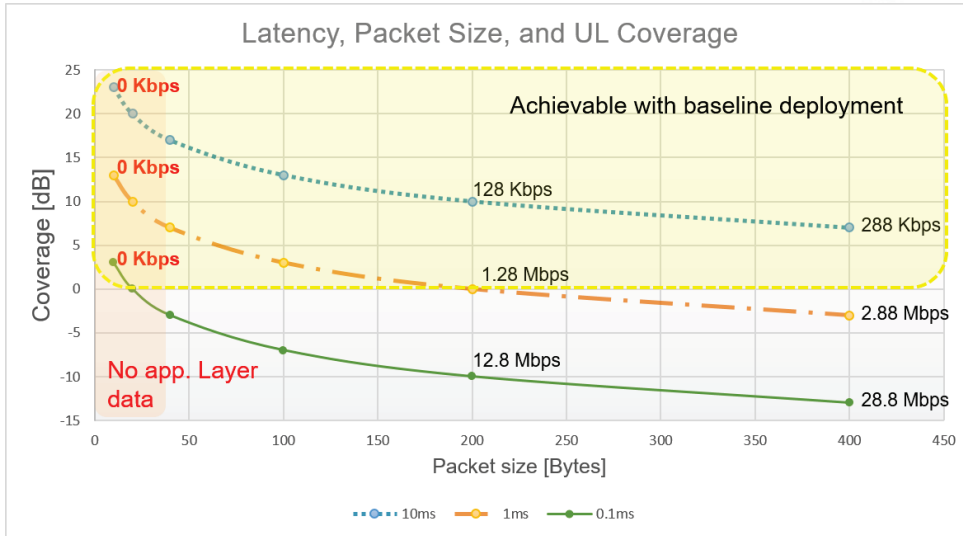


Figure 2 Interplay among latency, packet size, and uplink coverage. PER, bandwidth, and diversity order are fixed to baseline values (Table 1). The data rate values represent application layer data rate that is achievable at the corresponding operating points.

Increasing antenna diversity with respect to the baseline allows for an extension of the uplink coverage area within which extremely low latency services can be delivered. The three curves in Figure 3 represent different values of uplink diversity order, i.e., 1 (baseline), 2, and 8, respectively, and the target latency is 0.1ms for all curves. The PER and bandwidth are fixed to the baseline values (Table 1). As can be seen from the graph, the antenna diversity order needs to be increased 8-fold with respect to the baseline so that a terminal can actually send application layer data within an extremely low latency budget and this would require a considerable increase in terminal complexity as the terminal would have to have at least 8 antennas. More precisely, with diversity order of 8, a packet of up to 150 Bytes (i.e., 110 Bytes of application layer data plus 40 Bytes of overhead) can be delivered in the uplink direction on the considered urban macro grid in a 0.1 ms TTI. This corresponds to an application layer data rate of 8.8 Mbps.

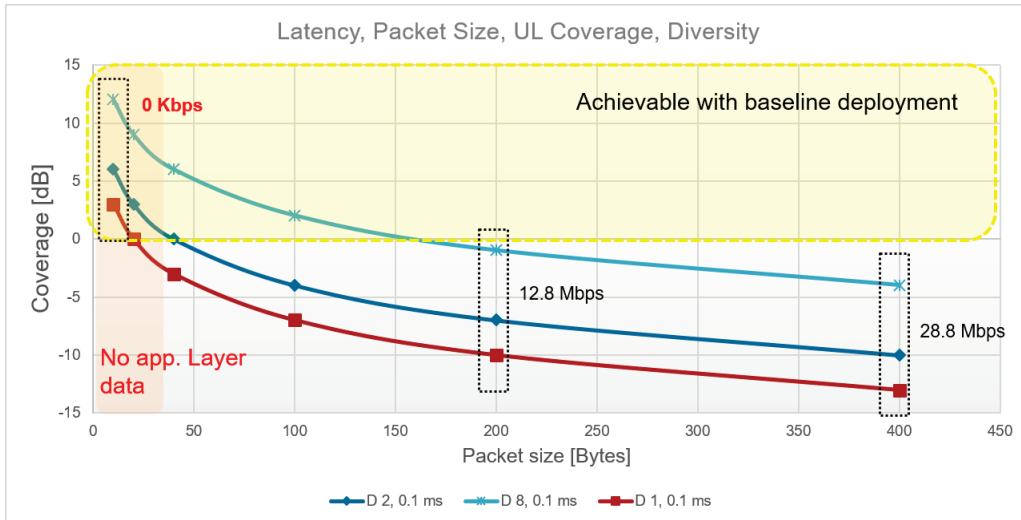


Figure 3 Interplay among latency, packet size, uplink coverage and uplink diversity order (D). PER and bandwidth are fixed to baseline values (Table 1). The application layer data rate that is achievable at the corresponding operating points is provided.

Figure 4 provides a glimpse of what happens when both latency and reliability requirements become extreme, and in particular a PER of $10E-5$ is considered as target reliability, together with 0.1 and 0.5 ms requirements for the latency. Each curve is associated with different requirements on packet size and latency. This scenario can hardly be handled by the baseline deployment and the “x” axis represents different diversity orders. Across the baseline coverage area, if the diversity order is 4, 10-Byte packets can be delivered within a 0.5 ms TTI and with diversity order of 8, 10-Byte packets can be delivered within a 0.1 ms TTI. As mentioned previously, in this context this means sending no application layer data due to the amount of overhead associated with TCP/IPv4 (40 Bytes). A 100-Byte packet can be delivered in a 0.5 ms TTI on the baseline coverage area with diversity order 8.

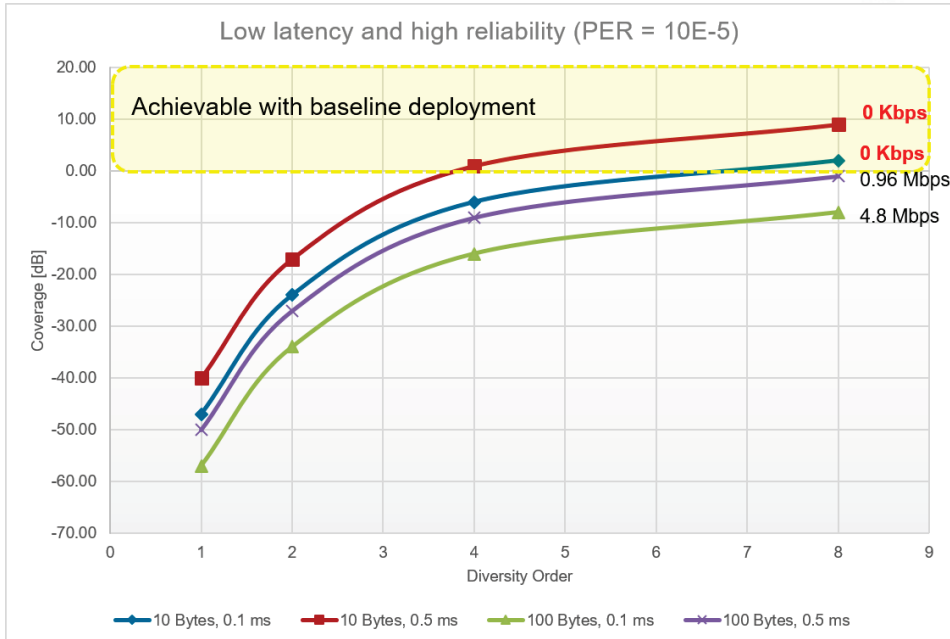


Figure 4 Interplay among latency, packet size, uplink coverage, reliability, and diversity. Bandwidth is set to the baseline value (Table 1). The application layer data rate that is achievable at the corresponding operating points is provided.

In Figure 5, the bandwidth is scaled up to 10-fold with respect to the baseline. If diversity and site deployment are fixed as per the baseline, bandwidth is in fact another parameter that can theoretically be tweaked to address extreme targets on the wireless link. The “x” axis is the bandwidth, the PER target is fixed to 10E-5, and different curves represent different combinations of latency and packet size. This graph shows how, even with a 10-fold increase in bandwidth allocation, it is not possible to simultaneously meet very high reliability and low latency requirements on the coverage area. Hence, bandwidth may play a role but it must be allocated in conjunction with other measures (e.g., more antennas).

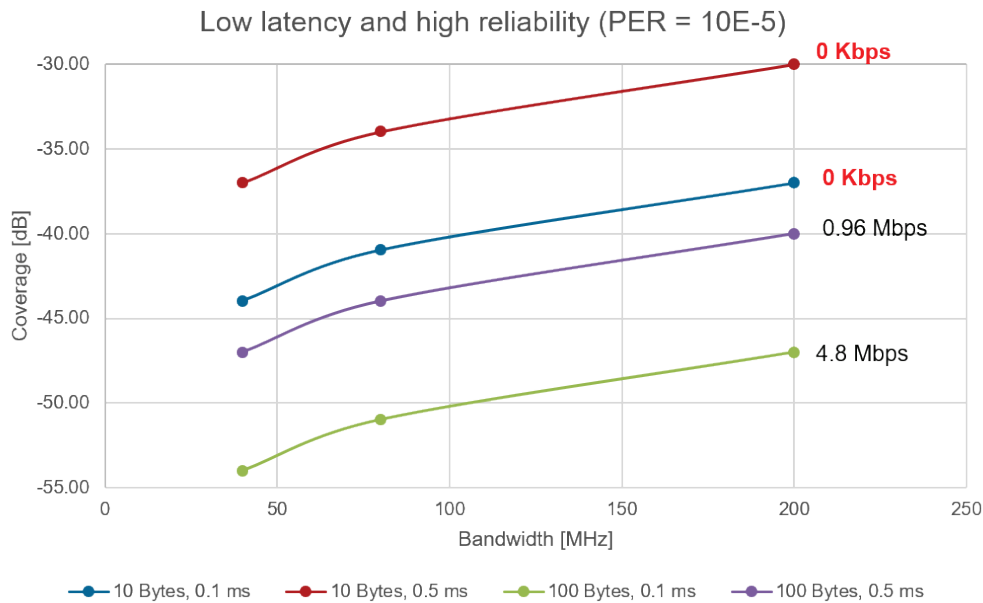


Figure 5 Interplay among latency, packet size, uplink coverage, reliability, and bandwidth. Diversity order is set at the baseline value (Table 1). The application layer data rate that is achievable at the corresponding operating points is provided.

In Figure 6, the efficiency of communication on the radio link is shown. The efficiency is calculated as the ratio between the achievable application layer data rate and the total data rate (including overhead) for a given packet size and latency budget. Clearly, if the packet size is small, the efficiency is low, and it then grows as the packet size grows. Since TCP/IP headers take up a significant amount of transmitted data, and workarounds such as Robust Header Compression are expensive and incur additional processing latency, it may be worth considering lighter protocol stacks for the delivery of low latency for small packets on the wide area network.

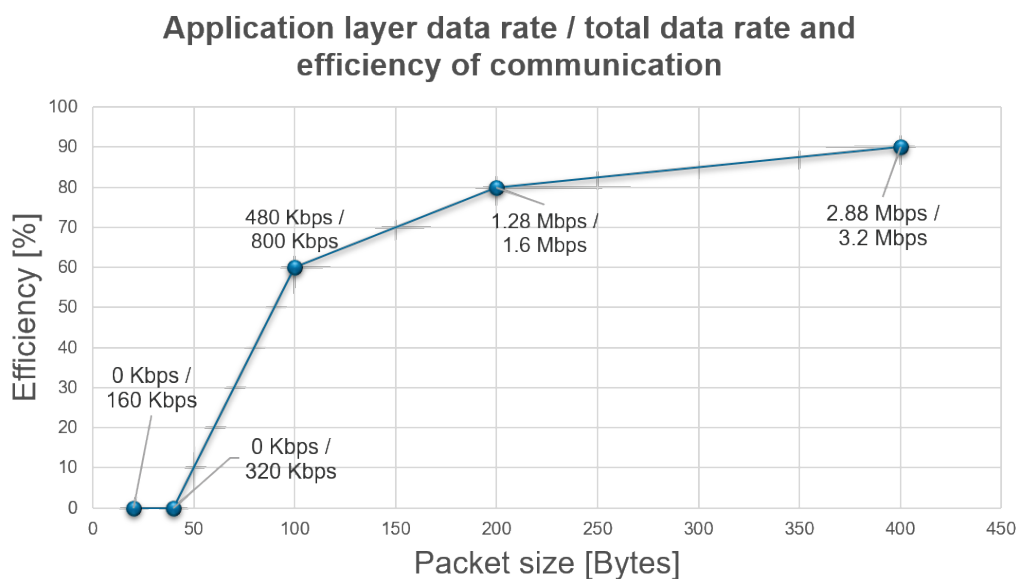


Figure 6 Communication efficiency when the requirement on latency per packet is 1ms

4 CONCLUSION

This work presented a few simple examples to show how delivering 5G extreme services on existing cellular networks may be challenging as new requirements might have a considerable impact on the coverage area of existing deployments that are optimised for the delivery of mobile broadband.

Increasing the number of antennas and bandwidth allocation are measures that can be put into place to boost the coverage availability. In the uplink direction, the terminal is power limited, so the option of increasing transmit power at the device is not viable. Instead, increasing the number of base stations may be another solution for service coverage enhancement. Differently, in the downlink, increasing the base station output power (within limits imposed by regulations) is another option for boosting the link budget and hence improving the downlink coverage area for extreme services.

If the above measures are not viable on the wide area network, one option is to accept that the service can only be delivered on a portion of the coverage area. Alternatively, if there is a specific need for an extreme service in a confined and localised area (e.g., a manufacturing plant), it may be worth deploying an ad hoc (i.e., private) network that is designed and optimised to meet certain requirements.

Since TCP/IP headers take up a significant amount of transmitted data, and workarounds such as Robust Header Compression are expensive and incur additional processing latency, it may be worth considering lighter protocol stacks for the delivery of low latency for small packets.

Clearly, more work needs to be done to extend this analysis, such as considering the downlink direction, TDD as another duplexing scheme, different radio interfaces (e.g., LTE-Advanced and NR), and extending the scope to system-level evaluations to quantify what network deployment models are most appropriate to meet extreme targets. All this will be the scope of Phase 2 of this work.

5 ANNEX

This section provides details on the methodology adopted to obtain the results presented in Section 0. The methodology is based on link-level considerations, and the following steps show how each graph has been derived. From spectral efficiency and bandwidth, the supported data rate can be derived as follows:

$$\text{Spectral Efficiency} \left[\frac{\text{bps}}{\text{Hz}} \right] \times \text{Bandwidth}[\text{Hz}] = \text{Data Rate}[\text{bps}]$$

(1)

Given the baseline values shown in Table 1, the following cell-edge data rate can be derived:

$$0.08 \times 20 \times 10^6 + 6 = 1.6 \text{ Mbps}$$

(2)

This is equivalent to sending 200 Bytes in 1 ms, and it represents the cell-edge capability of the baseline deployment. Hence, 1ms is considered as baseline latency, and 200 Bytes is considered as baseline packet size for the purpose of the following calculations.

The coverage points in Figure 2 have been derived as follows:

$$Coverage [dB] = 10 \times \log_{10} \left[\frac{Latency Requirement}{Baseline Latency} \times \frac{Baseline Packet Size}{Packet Size Requirement} \right] \quad (3)$$

Where:

- The Baseline Latency is 1ms as shown above,
- The Baseline Packet size is 200 Bytes as shown above,
- The Latency Requirement is an input,
- The Packet Size Requirement is an input
- Coverage is the link budget margin that derives from either having to send more information in less / equal time (negative margin), or less information in more / equal time (positive margin).

Equation (3) is obtained assuming that, at cell edge, where the Signal-to-Noise Ratio (SNR) is low, the data rate is proportional to the SNR.

Table 3 provides the numbers that have been obtained with Equation 3 with different packet sizes and latency values as inputs, and that have been used to populate the graph in Figure 2. The baseline has been highlighted in yellow.

The coverage points in Figure 3 have been derived as follows:

$$Coverage [dB] = 10 \times \log_{10} \left[\frac{Latency Requirement}{Baseline Latency} \times \frac{Baseline Packet Size}{Packet Size Requirement} \right] + 10 \times \log_{10} \left[\frac{Diversity Order}{Baseline Diversity Order} \right] \quad (4)$$

Where:

- The first term to the right hand side of the equation is equivalent to the right hand side of Equation 3,
- The second term is the gain obtained by increasing the diversity order, e.g., if the diversity order doubles the gain is 3dB.

Table 4 provides the numbers that have been obtained with Equation 4 with different packet sizes, latency requirements, and diversity orders as inputs and that have been used to populate the graph in Figure 3.

The coverage points in Figure 4 have been derived as follows:

$$Coverage [dB] = 10 \times \log_{10} \left[\frac{Latency Requirement}{Baseline Latency} \times \frac{Baseline Packet Size}{Packet Size Requirement} \right] + 10 \times \log_{10} \left[\frac{Diversity Order}{Baseline Diversity Order} \right] - FG \quad (5)$$

Where:

- The first two terms to the right hand side of the equation are equivalent to the right hand side of Equation 4,
- The third term is the fading gain (FG) needed to increase reliability. The values shown in Table 2 are used as FG values for different diversity orders (Source: [9])

Table 5 provides the numbers that have been obtained with Equation 5 with different packet sizes, latency requirements, diversity orders, and PER = 10E-5 as inputs and that have been used to populate the graph in Figure 4.

The coverage points in Figure 5 have been derived as follows:

$$Coverage [dB] = 10 \times \log_{10} \left[\frac{Latency Requirement}{Baseline Latency} \times \frac{Baseline Packet Size}{Packet Size Requirement} \right] + 10 \times \log_{10} \left[\frac{Diversity Order}{Baseline Diversity Order} \right] - FG + 10 \times \log_{10} \left[\frac{Bandwidth}{Baseline Bandwidth} \right] \quad (6)$$

Where:

- The first three terms to the right hand side of the equation are equivalent to the right hand side of Equation 5,
- The fourth term is the gain obtained by increasing bandwidth.

Table 6 provides the numbers that have been obtained with Equation 6 with different packet sizes, latency requirements, PER = 10E-5, and different bandwidth values as inputs and that have been used to populate the graph in Figure 5.

Table 2 Fading margin used for different diversity orders (Source: [9])

	Div Order = 1	Div Order = 2	Div Order = 4	Div Order = 8
Fading Margin for 10E-5				
PER [dB]	50	30	15	10

Table 3: Interplay among packet size, latency, and UL coverage. Diversity order, PER, and bandwidth are fixed as per the baseline (shown in Table 1).

Latency [ms]	0.1	0.5	1	2.5	5	10
UL Coverage [dB]						
Packet size [Bytes]						
10	3.01	10.00	13.01	16.99	20.00	23.01
20	0.00	6.99	10.00	13.98	16.99	20.00
40	-3.01	3.98	6.99	10.97	13.98	16.99
100	-6.99	0.00	3.01	6.99	10.00	13.01
200	-10.00	-3.01	0.00	3.98	6.99	10.00
400	-13.01	-6.02	-3.01	0.97	3.98	6.99

Table 4 Interplay among packet size, latency, and UL coverage. Diversity order is set to 2. PER and bandwidth are fixed as per the baseline (shown in Table 1).

Diversity Order		2					
Latency [ms]	0.1	0.5	1	2.5	5	10	
UL Coverage [dB]							
Packet size [Bytes]							
10	6.02	13.01	16.02	20.00	23.01	26.02	
20	3.01	10.00	13.01	16.99	20.00	23.01	
40	0.00	6.99	10.00	13.98	16.99	20.00	
100	-3.98	3.01	6.02	10.00	13.01	16.02	
200	-6.99	0.00	3.01	6.99	10.00	13.01	
400	-10.00	-3.01	0.00	3.98	6.99	10.00	

Table 5: Interplay among latency, packet size, reliability, and UL coverage. PER and bandwidth are set as per the baseline (shown in Table 1).

PER		10E-5					
Diversity Order		2					
Latency [ms]		0.1	0.5	1	2.5	5	10
		UL Coverage [dB]					
Packet size [Bytes]							
10		-23.98	-16.99	-13.98	-10.00	-6.99	-3.98
20		-26.99	-20.00	-16.99	-13.01	-10.00	-6.99
40		-30.00	-23.01	-20.00	-16.02	-13.01	-10.00
100		-33.98	-26.99	-23.98	-20.00	-16.99	-13.98
200		-36.99	-30.00	-26.99	-23.01	-20.00	-16.99
400		-40.00	-33.01	-30.00	-26.02	-23.01	-20.00

Table 6: Interplay among latency, packet size, reliability, bandwidth, and UL coverage.

PER		10E-5					
Bandwidth [MHz]		200					
Latency [ms]		0.1	0.5	1	2.5	5	10
		UL Coverage [dB]					
Packet size [Bytes]							
10		-36.99	-30.00	-26.99	-23.01	-20.00	-16.99
20		-40.00	-33.01	-30.00	-26.02	-23.01	-20.00
40		-43.01	-36.02	-33.01	-29.03	-26.02	-23.01
100		-46.99	-40.00	-36.99	-33.01	-30.00	-26.99
200		-50.00	-43.01	-40.00	-36.02	-33.01	-30.00
400		-53.01	-46.02	-43.01	-39.03	-36.02	-33.01



6 BIBLIOGRAPHY

- [1] NGMN, "NGMN 5G White Paper," 2015.
- [2] NGMN, "NGMN Perspectives on Vertical Industries and Implications for 5G," 2016.
- [3] 3GPP, "FS_SMARTER - Massive Internet of Things," *TR 22.861*.
- [4] 3GPP, "FS_SMARTER - Critical Communications," *TR 22.862*.
- [5] 3GPP, "FS_SMARTER - enhanced Mobile Broadband," *TR 22.863*.
- [6] 3GPP, "Service Requirements for the 5G System," *TS 22.261*.
- [7] 3GPP, "Feasibility study for further advancements for E-UTRA (LTE-Advanced)," *TR 36.912*.
- [8] ITU-R M.2135, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," 2008.
- [9] N. A. Johansson, E. Y.-P. Wang, E. Eriksson and M. Hessler, "Radio Access for Ultra-Reliable and Low-Latency 5G Communications," in *IEEE ICC - Workshop on 5G & Beyond - Enabling Technologies and Applications*, 2015.