DIPLOMARBEIT

Gilbert-type
Vehicle-to-Vehicle Link Models
for Capacity Analysis

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Diplom-Ingenieurs

unter der Leitung von
Univ.Prof. Dipl.-Ing. Dr.techn. Christoph F. Mecklenbräuker
Univ.Ass. Dipl.-Ing. Dr.techn. Veronika Shivaldova
Institute of Telecommunications

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik und Informationstechnik

von
Thomas Blazek
Obkirchergasse 38/4/8
1190 Wien, Österreich

Wien, im Oktober 2015

Abstract

In this thesis, we develop and investigate several performance models for Vehicle-to-Vehicle (V2V) communication systems. As starting point, we use a time-variant version of the Gilbert model, which allows low-complexity modelling of burst-error packet channels. We first analyze the general properties of the model with the emphasis on the choice of time resolution. Then we derive a closed-form representation for the channel capacity of such model and prove that it is equal to the Packet-Delivery-Ratio (PDR).

We further establish a relationship between the model parameters and physical quantities. For this, we use real-world measurements performed in 2011 measurement campaign on Austrian highways. A general approach is presented to build the dependence between packet-error trace and a respective physical quantity used as modelling basis. We then apply this approach to our measurements, using both the Signal-to-Noise-Ratio (SNR) and the distance as basis for the modelling. To improve modelling, we introduce channel scenarios. These scenarios account for channel influences not captured by the dependencies between packet-error and SNR or packet-error and PDR.

The scenario-dependent model shows good agreement with the measurements based on the SNR and on the distance. Both models are able to reproduce the underlying measurements while remaining of low complexity. Finally, we extend these investigations based on the video documentation and Global Positioning System (GPS) traces recorded alongside the measurements to analyze the influences causing a change of scenario.

The resulting models allow to acccurately reproduce the real-world measurements that can later be used in the context of a simulation. Furthermore, the developed modelling methods can easily be applied to other measurements.

Zusammenfassung

In dieser Arbeit entwickeln und untersuchen wir mehrere Performance-Modelle für Auto-zu-Auto Kommunikationssysteme. Als Ausgangspunkt wird eine zeitvariante Version des Gilbertmodells gewählt, die das Modellieren von Burst-Fehler Paketkanälen mit niedriger Komplexität erlaubt. Zunächst analysieren wir die grundlegenden Eigenschaften dieses Modells mit einem besonderen Augenmerk auf die Wahl der Zeitauflösung. Außerdem führen wir eine geschlossene Form für die Kanalkapazität dieses Modells ein, und zeigen, dass diese äquivalent zum Packet-Delivery-Ratio (PDR) ist.

Dann stellen wir einen Zusammenhang zwischen den Modellparametern und physikalischen Größen her. Zu diesem Zweck verwenden wir Messungen die in einer Messkampagne 2011 auf österreichischen Autobahnen gesammelt wurden. Wir zeigen einen grundlegenden Zugang, der erlaubt, Abhängigkeiten zwischen Paketfehler-Messungen und entsprechenden Größen, die als Modellierungsbasis verwendet werden, zu bestimmen. Diesen Zugang wenden wir dann auf die verwendeten Messungen an, wobei wir sowohl Signal-to-Noise-Ratio (SNR)- und Abstandsmessungen verwenden. Um die Modellqualität zu verfeinern, führen wir das Konzept der Kanalszenarios ein. Diese Szenarios erlauben, Kanaleinflüsse zu unterscheiden, die nicht in die Abhängigkeiten zwischen den Paketfehlern und SNR oder Distanz eingehen.

Das Kanalszenario-abhängige Modell erreicht SNR- und distanzbasiert gute Übereinstimmung mit den Messungen. Diese Modelle können sowohl basierend auf auf dem SNR als auch der Distanz trotz niedriger Komplexität die Messresultate reproduzieren. Schlussendlich untersuchen wir die Einflüsse die einen Szenariowechsel verursachen anhand von Video- und Global Positioning System (GPS)-Aufzeichnungen die parallel zu den Messungen gemacht wurden.

Die resultierende Modelle erlaubt die Messungen zu reproduzieren und innerhalb einer Simulation zu verwenden. Außerdem kann die entwickelte Modellierungs-Methode leicht auf andere Messungen angewandt werden.

Acknowledgements

As the completion of this thesis not only stands for itself, but marks the end of my diploma studies, I want to take the opportunity to thank everyone who helped or supported me over the past seven years. I do however want to point out a few people in particular.

First and formost, I would like to thank Christoph Mecklenbräuker for the possibility to write this thesis and Veronika Shivaldova for her excellent consulting and input. I couldn't have asked for a better surrounding or topic to write my thesis on.

Second, I would like to thank the Fachschaft Elektrotechnik for tirelessly supporting the students, and providing both help and diversion whenever needed.

Third, I want to thank my family for always supporting me without question during these past years.

And finally, last but not least, to you Doris, thank you, I love you!

Contents

List of Figures

1

Introduction

1.1 Overview and Motivation

Nowadays cars constitute the primary mode of transportation throughout the world. Passenger car traffic accounts for $83.3\,\%$ of travelled kilometers in the EU (European Union) in 2012 [1]. As of the year 2013, the total number of cars worldwide has surpassed 1.2 billion [2], and continues to grow further. Additionally, there exist large markets that have only recently started to fully address private transportation. In China, the number of cars manufactured for private use has shot up from 3 million to 20 million per year in the span of the last 10 years [3].

This ubiquity of cars also means that vehicular traffic has a strong social, economical and ecological impact on the world. Road safety is a primary topic for the World Health Organization (WHO), as it predicts road traffic injuries to be the fifth leading cause of death by 2030 [4] unless action is taken. On the economic side, fuel efficiency and traffic flow are important factors. Especially in cities where the number of cars is still increasing, traffic congestion becomes an increasing problem. Finally, reducing the ecological footprint of the car is also an important part of the sustainability aspect of vehicular traffic.

Innovations in all parts of technology have been continuously introduced into vehicles in order to improve one of those aspects. Examples of such innovations are Anti-lock Breaking System (ABS) and Electronic Stability Control (ESC). For both, studies have shown that these innovations achieved their goal of lowering the risk of accidents ([5] and [6] respectively). Whenever a new technology is introduced, the question arises if it can be beneficial to driving.

One such promising new technology is the concept referred to as Internet of Things (IoT). The IoT is a broad concept, based on the idea to allow *things* to communicate with each other over a self organized network. The idea is that this lets the *things* achieve otherwise impossible feats. Applied to cars, it means that the cars would be able to communicate either with each other, or with the infrastructure unit that may stand alongside the road the cars are on. Both of these scenarios have their respective use cases. A Roadside Unit (RSU) may inform the cars of traffic or weather conditions, imposing a lowered speed limit. In case of dense traffic, a lower speed limit could mitigate jamming, while on a wet street this can of course improve safety. On the other hand, communication between cars would mean that one car can transmit information about abrupt braking maneuvers, triggering the others to do the same. This scenario would avoid accidents during such maneuvers, but of course it requires a large portion of cars to be part of such a network in order to be effective.

For this reason, vehicular connectivity is of high interest. The European Research Cluster on the Internet of Things (IERC) named these Intelligent Transport Systems (ITS) as second driver of the IoT behind smart homes [7], and measures have been taken by many countries to harmonize ITS. In order to allow interoperability, spectrum around $5.9\,$GHz was reserved for use in ITS systems [8]. In 2010, the IEEE 802.11p standard was finalized [9], which defined the respective Medium Access Control (MAC) and Physical Layer (PHY) layer functionalities.

Although the foundation is laid, manufacturers still need incentives to equip their cars with the technology. To this end, Austria, Germany and the Netherlands jointly created the Cooperative ITS Corridor [10]. This is a highway going from Rotterdam to Vienna via Frankfurt and Munich that is equipped with RSUs alongside the road at selected positions. These RSUs will provide useful information to cars which can only be received if the necessary communication equipment is installed. This creates a reason to implement the technology in the car. These Vehicle-to-Infrastructure (V2I) applications are an important first step, as manufacturers have guaranteed service on equipped roads, and do not have to rely on other manufacturers implementing the same technology.

Conversely, one of the first application incentives for Vehicle-to-Vehicle (V2V) is the concept of platooning trucks [11]. These platoons are composed of multiple trucks driving in close proximity while only the lead truck is steering, the others are following guided by an ITS system. These road trains are expected to decrease fuel usage.

Although the first implementation incentives exist, the topic is approached with a high amount of caution, since small mistakes in the implementation can ensue grave repercussions and endanger lives. For this reason, the introduction of such a technology also has to expect to be met with scepticism by the general public. This will understandably slow the innovation pace and requires the systems to be highly reliable.

To achieve the required high reliability, extensive measurements are essential. It is however impossible to measure every aspect of such a complex system. Instead, abstractions have to be made that reduce complexity to a feasible amount, while still producing results that can be relied upon. Therefore, the measurements are usually taken with respect to a small number of system parameters. Then, these measurements are used to derive a model that is able to reproduce the aspects that the measurements captured. It is vital for a good model that it realistically represents the aspects for which it was designed.

Since IEEE 802.11p standardizes PHY and MAC layer, [12] argues that it is crucial to have performance models that realistically capture the packet-error behaviour, yet is simple enough to be effectively implemented in network simulators. The authors focused their modelling goals on the V2I aspect, leaving V2V communications untouched. Focusing on V2I is natural, as it is the technology closer to deployment. We do however require the same models also for the V2V channel. While the fundamental modelling approach is general enough to be applied to the V2V case, the fact that now both transmission ends are moving introduces considerable changes. There is, for instance, no longer a simple correspondence between passed time and transmit distance. Indeed we cannot predict any behaviour of a physical quantity based on time alone, or vice versa. Moreover, no end has fixed coordinates. Since the RSU has a constant location, the distance between car and RSU and the Global Positioning System (GPS) coordinates of the car are directly coordinated via the shape of the highway. Our goal is to approach the additional problems that V2V modelling poses and to derive a performance model analogous to [12] for the V2V case.

In this thesis we develop such a performance model for V2V transmission based on real world measurements. The goal of this model is to replicate the packet-error behaviour of the underlying measurements. To achieve this, it has to link physical parameters to packet-errors. Such a model can then be implemented in network simulators, facilitating replication. This model should follow three rules.

1. The number of resulting model parameters should be small.

2. Since we are dealing with V2V measurements, we will see a large amount of time-variance. This time-variance should be reduced as far as possible.

3. The resulting model should still be able to realistically replicate the measurements it is based on.

The used measurements are taken from a 2011 measurement campaign executed near Vienna.

Analogous to [12], we will use the Gilbert Model [13] as foundation for our performance model. The ultimate goal is to make the model depend on a single physical parameter, such as the Signal-to-Noise-Ratio (SNR) or the distance. For this thesis, the key performance value

we are interested in is the capacity of a given channel. As will be shown, this capacity is equal to the Packet Delivery Ratio (PDR) at any given moment.

1.2 State of the Art

In contrast to channel models, performance models do not try to provide a description of the propagation of electromagnetic waves. Instead, they present abstraction to a higher layer of description. This abstraction often allows to be computationally efficient and therefore be implemented in larger simulations. One possibility is to generate Received Signal Strength Indicator (RSSI) traces based on the distance. [14, 15] derive an RSSI estimate based on an analytical channel model. [16] used empirical measurements to fit this relation based on a dual-slope pathloss description. [17] estimated the effects of buildings and included a shadow-fading model in their approach. This is then combined with 2.5D maps such as OpenStreetMap to calculate pathloss values. Measurements have shown that not only buildings, but also vehicles are able to produce shadow fading. This has been addressed for example in [18–20].

Network simulators such as ns-3 however require a further step of abstraction. They typically expect the model to provide a packet-error trace. This trace again can be based on a model, or on empirical measurements. [21] and [22] both provide packet-error traces based on RSSI values. While the former implements a hard threshold, the latter uses a probabilistic mapping.

In contrast to the RSSI models, few packet-error models exist that are based on real-life measurements. [23] introduced a model that maps distance to packet error rate (PER). This approach however is limited, as many influences on the PER are independent of the distance, as will be shown in this thesis. [24] introduces a four-state markov chain, but lacks the link to a physical quantity such as the distance. [12] has introduced an approach to model the performance based on the distance, but used assumptions only valid for the V2I situation.

1.3 Structure of Thesis

In this section, we briefly discuss the structure of the remaining thesis.

Chapter 1 gives a short introduction into the used measurements. These measurements will be used in the following chapters as basis for deriving a performance model.

Chapter 2 introduces the modified Gilbert model, and presents its key properties. It further analyzes the influences of time-resolution, derives a closed-form representation of the capacity and shows how the model is linked to our measurements.

Chapter 3 approaches the topic of performance estimation based on a physical quantity. First, an algorithm is devised to find a time-invariant mapping from a physical parameter to a performance estimate. This is demonstrated using SNR and distance measurements. In a second step the model is enhanced with low-complexity time-variance to arrive at a more realistic performance model. This is done by introducing two scenarios, and at any given moment, one of the two is valid. We then describe these scenarios seperately.

Chapter 4 takes the results from Chapter 3 and analyzes the results from the estimation. It then proceeds to formulate explicit forms for scenario-dependent performance estimation based on SNR and distance.

Chapter 5 presents the conclusions drawn from the thesis and outlook on possible tasks.

1.4 Measurement Campaign

The measurements used in this thesis are part of the `ROADSAFE2011` measurement campaign [25]. The V2V measurements were made between August 30th and September 2nd on the highway A4 east of Vienna. Two cars equipped with Cooperative Vehicle-Infrastructure Systems (CVIS) nodes [26] were used in the V2V measurements. Both Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS) measurements were recorded.

The next section will briefly illustrate the measurement setup used in the campaign. Please note however that for the remainder of this thesis, the transmit and receiver chains will be regarded as part of the channel. We will therefore take the measurement results as output of a "black box" and will not delve into the dependence on transmit parameters such as the modulation. Instead, after arriving at results, we will try to argue why these results are relevant in general, even when they are obtained from a specific setup.

1.4.1 Measurement Setup

Fig. 1.1 shows the roof-mounted antennas of the two cars used for the measurements. The CVIS nodes transmitted packets at a data rate of 6 Mbit/s at the center frequency of 5.88 GHz. The packet length was 500 Bytes. QPSK modulation with code rate $1/2$ was used. The transmit power was set at 10 dBm. The cars were driving on average at a speed of 80 km/h

(22.22 m/s). Additionally to the packet transmission the RSSI, noise level and GPS coordinates of the cars were recorded, allowing to calculate SNR and distance estimates.
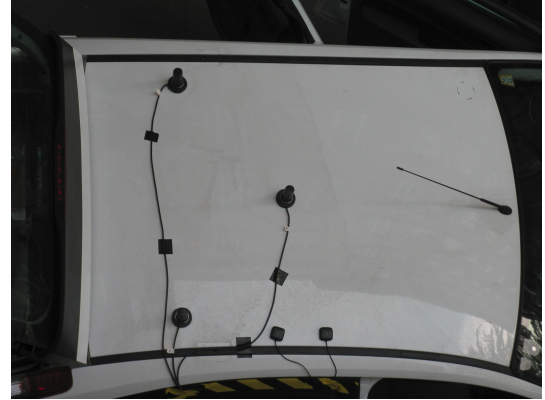
(a) Car1   (b) Car2

Figure 1.1: Antenna mounting

In order to measure a NLOS channel, the truck seen in Fig. 1.2 drove between the two measurement cars.

Figure 1.2: Truck for LOS obstruction

1.4.2 Measurement Results

A total of 33 seperate measurements were used in this thesis. These encompass both NLOS and LOS scenarios. Distances between the measurement cars from almost 0 m during overtaking maneuvers up to 250 m were covered, with the SNR ranging from 0 up to 50 dB.

Appendix A contains a comprehensive set of parameters of all measurements used in this thesis. Moreover, it contains plots of the key performance indicators, such as, PDR, SNR and distance. For these plots, the parameters were always averaged over 1 s intervals.

For illustration purposes, in Fig. 1.3 and Fig. 1.4 we show the key performance indicators for measurement 7 and 8. These measurements will be used for visualisation of the results in the remainder of the thesis. The respective measurement parameters are summarized in Table A.1. Especially when looking at the plots from measurement 8, we see that the PDR is correlated to the SNR. It is also correlated to the distance, but this dependence is less pronounced.

These measurements were chosen for visualization purposes, because they are very illustrative in their behaviour, as will be seen in later sections. At the same time, their behaviour is in line with the other measurements, making sure the chosen examples actually speak for the measurements. Any calculated performance indicator will however always be calculated with regards to all measurements.
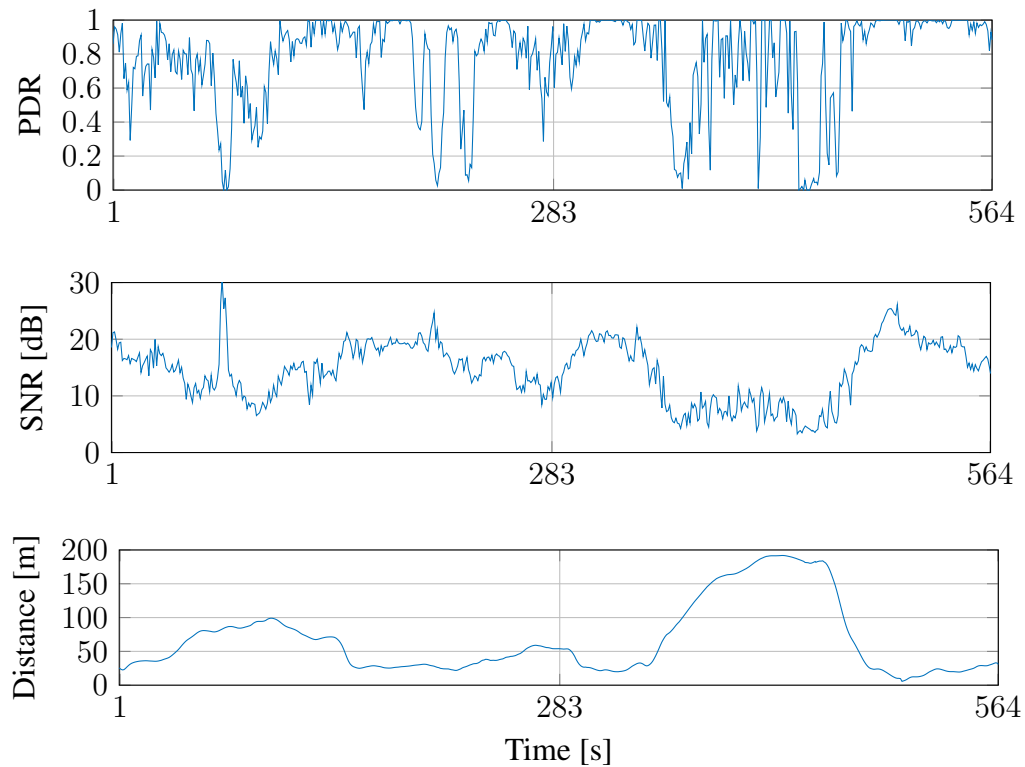
Figure 1.3: Key performance indicators of measurement 7
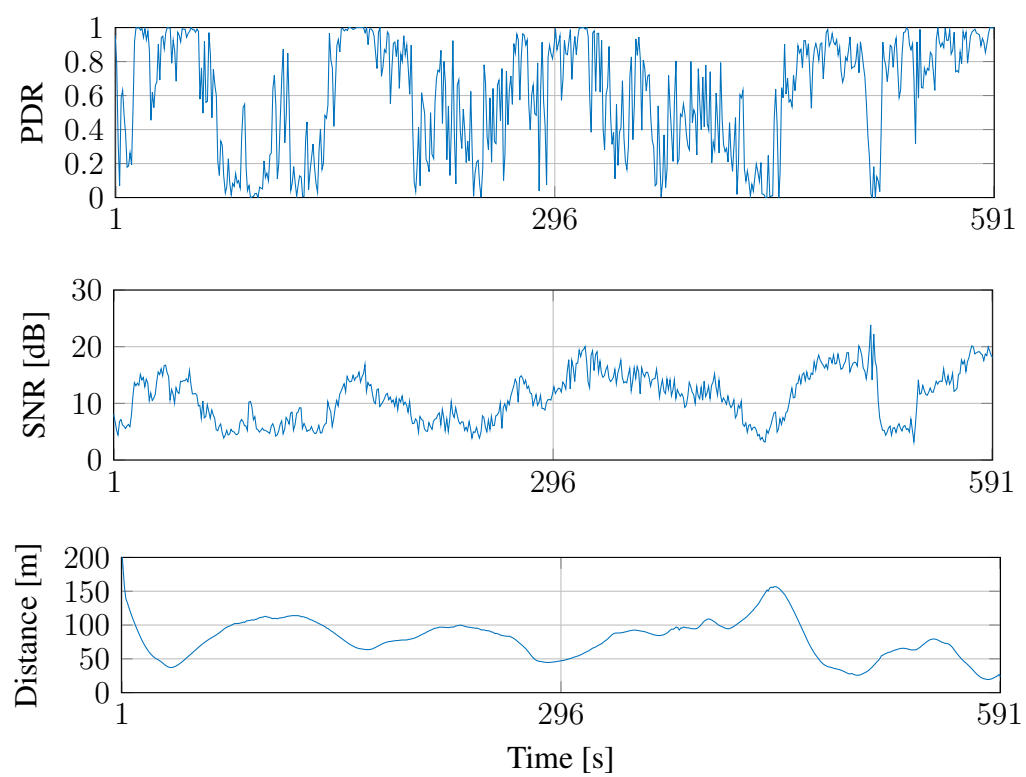
Figure 1.4: Key performance indicators of measurement 8

2

Gilbert Type Performance Modelling

The aim of this thesis is to provide a low complexity link layer model that is able to replicate the measured performance closely. As a starting point, we use the so called Gilbert model. Gilbert introduced his model for burst errors in 1960 [13]. Since then, the model, or variants of it such as the Gilbert-Elliot model [27] have been widely applied to emulate burst error behaviour on link and network layer (see for example [28–31]). The model is based on a two state Markov chain, with errors only occuring in one of the two states. This allows to model alternating error free and error prone time regions, while only three parameters are needed.

The original Gilbert model however is purely time-invariant, as the parameters of the model were assumed constant. This assumption is not applicable to vehicular channels, as many strongly time variant physical quantities influence the transmission quality. Therefore, Shivaldova et al. introduced a modified Gilbert model in [12], which is able to deal with the V2I scenario. The basic idea is to split the channel into distance intervals, and derive a seperate description based on the Gilbert model for every interval. Since in the V2V-case, both transmitter and receiver are moving, the distance-time relation, which is almost linear in the V2I case, is much more complex. Therefore, instead of addressing distance intervals, we take our measurements and split them into time intervals. The reason to do this is that the CVIS box sent with an approximately constant data rate. This means that the time division creates intervals with comparable numbers of events per time slot. This furthermore allows us to choose the interval length such that the major time-variant influences are captured in a well defined manner.
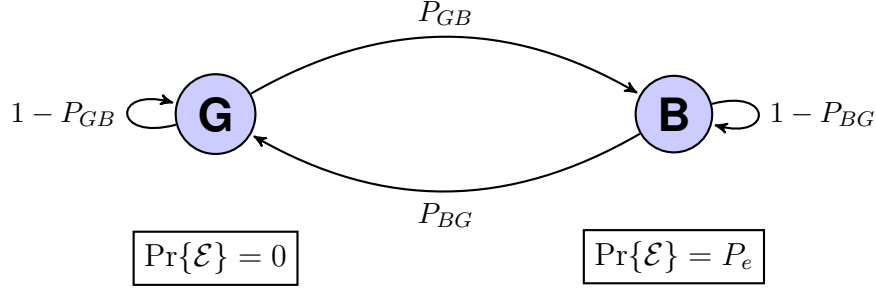
Figure 2.1: Packet level modified Gilbert model.

The following sections will first provide a detailed description of the Gilbert model and how it is applied in our case. Then, we analyze the mathematical properties of the underlying Markov chain. Afterwards, we derive the achievable capacity as a function of the model parameters. Finally, we explain how the model parameters are extracted from our measurements.

Once the model is defined, it can be used to generate new packet streams that, if the model performs well, behave analogous to our measurements.

2.1 The Modified Gilbert Model

The modified Gilbert model introduced in [12] is a packet level model to describe a time-varying burst error packet channel. As seen in Fig. 2.2, the model is based on a two-state Markov chain. This means, that there is a underlying process, a so-called Markov process, that can take one of two states, labelled G for "good" and B for "bad" in our case. The two states are assign transition probabilities P_{GB} from good to bad and P_{BG} from bad to good state. At every time step, a state transition happens according to the transition prbabilities. If the process ends in the bad state, packet-errors occur with a nonzero error probability P_e. Since error-free transmission is possible in both states, we can't infer the underlying state sequence of the Markov process from the received sequence alone, making this model a hidden Markov model [32].

In the original Gilbert model the triple of model parameters (P_{BG}, P_{GB}, P_e) are constant. In the modified version, they are however time sequences that can change with every time interval. Following Shivaldova's definition the model stays in the state during the interval transition.

The original Gilbert model is designed for modelling the bit errors. In this case P_e indicates a bit error probability. However, the model has since been often used to describe packets instead, meaning P_e is a packet error. This is an important difference, as an erroneous packet

can incorporate error detection mechanisms, while a single bit cannot. In our case, the model simulates the whole layer two. That means that the model has packets as input and output, and during transmission every packet is equipped with a Cyclic Redundancy Check (CRC). Based on the CRC erroneous packets are detected, but cannot be corrected.

The next two sections will illustrate further how modelling for those two cases differs.

2.1.1 The Modified Gilbert Model for Packets

All packets are assumed to have the same length. Any packet that fails the CRC check is considered erroneous. Otherwise, the CRC is stripped and the data is processed further. This is equal to labelling any bit with a failed CRC as e, as opposed to 0 or 1. This is called an *erasure*. On the other hand, the assumption is made, that no errors occur in such a way that the CRC is still passed. Therefore, the bit flip probability is 0 in all situations. This channel is called a binary erasure channel (BEC). A visualisation of this channel is seen in Fig. 2.2. Here an erasure occurs with probability P_e in the bad model state.

This approach has benefit that according to our model, we can identify all erasures, meaning no error stays hidden. On the other hand, an erasure bit can not be converted to a binary value again.
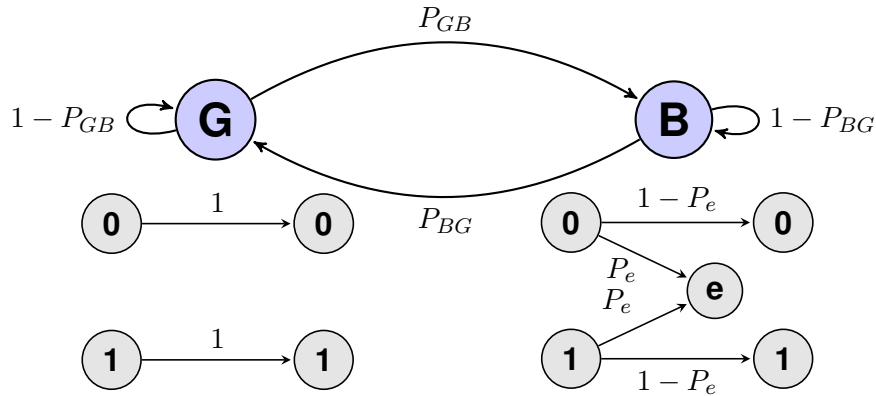
Figure 2.2: Packet level modified Gilbert model, showing the state dependent BECs.

2.1.2 The Modified Gilbert Model for Bits

For the Modified Gilbert Model for bits, the underlying good-bad states are described the same way, and still change at packet instance. But contrary to the packet level description, a failed CRC check does not entail the discarding of the whole packet. This allows for the received bits to still be interpreted. Instead, a failed CRC check will only be used to identify the channel to be in bad state. This means that the bits within the packet must be assigned

bit error probabilities $P_{B,i}$, where the i indicates the bit position within one packet. This channel is depicted in Fig. 2.3.This model only works if there is additional knowledge of the bit error distribution within one packet. This distribution furthermore has to be stationary within channel states, in to be able to make statements.
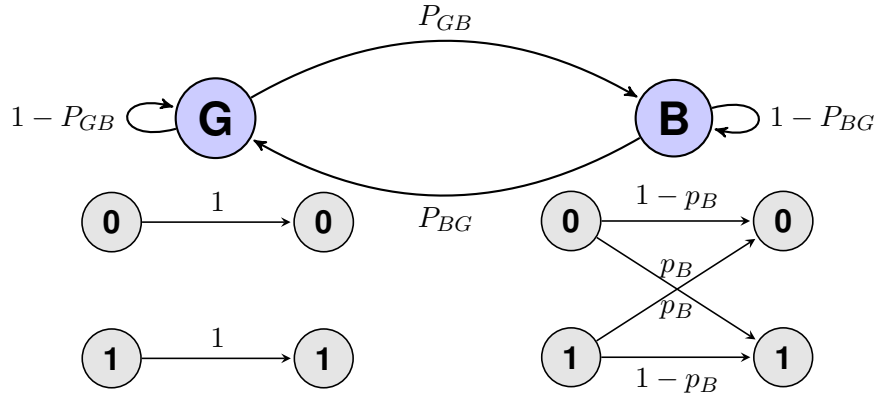
Figure 2.3: Bit level modified Gilbert model, showing the state dependent BSCs.

2.1.3 The Gilbert Source-Channel
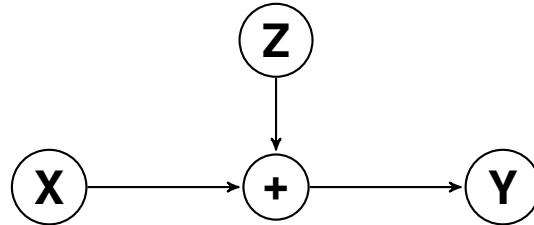
Figure 2.4: Equivalent additive noise model.

In a first step, the Gilbert model is rewritten as an additive Hard-Input Hard-Output (HIHO) channel, as shown in Fig. 2.4. Here, the noise sequence **Z** has to be generated in such a way that the input-output relation follows the Gilbert model. This is possible if we rewrite the Gilbert model as in Fig. 2.5. This is the so-called "source-channel" version of the Gilbert model, so called because it acts as noise source for the transmission. This model has two steps. In the first step, the Markov chain produces one of two values: a 0 if the model was in good state, and a 1 if the model was in bad state. A second step then randomly flips the values, and produces the final noise sequence. A 0 always stays 0, indicating that no errors happen when the model was in good state. In the bad state, the noise sequence either puts out e or 0 depending on P_e. This e stands for either erasure or error, depending on the channel we are looking at. This rewrite makes further analysis more tractable. Since there is no other er-

ror source in our channel, input-output behaviour is perfect but for **Z** from the source-channel model. This in turn means that for entropy analysis, **Z** carries the only unknown properties.
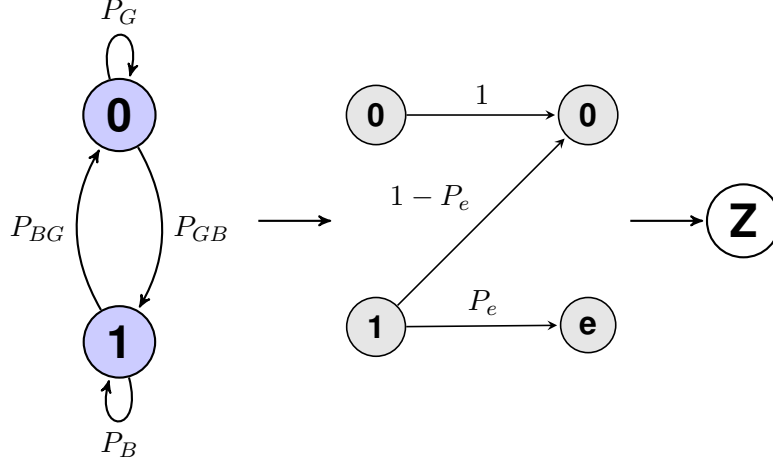
Figure 2.5: The equivalent source-channel noise model.

2.2 Two-State Markov Chain

At this point we want to analyze the properties of the underlying Markov process. The goal of this section is to describe the good-bad state behavior. Furthermore, Gilbert's original analysis dealt with infinitely long sequences, allowing the model to assume steady state. Since the maximum length of the process before switching parameters is bounded, we want to analyze the possible influence of this fact.

Every two-state Markov chain can be described by 2-by-2 state transition matrix $\mathcal{M}$. This state transition matrix describes the possible state transitions and their probabilities of occuring. The probability of being in one of the two states at timestep n is captured by the 2-by-1 vector π_n. Using this notation, we can introduce the eigenvalues and vectors

$$\pi_{n+1} = \mathcal{M}\pi_n \tag{2.1a}$$

$$\pi_{n+1} = \begin{pmatrix} 1 - P_{GB} & P_{BG} \\ P_{GB} & 1 - P_{BG} \end{pmatrix} \pi_n \tag{2.1b}$$

$$\lambda(\mathcal{M})_1 = 1 \tag{2.1c}$$

$$\lambda(\mathcal{M})_2 = 1 - \underbrace{(P_{GB} + P_{BG})}_{=P_{ges}}, \tag{2.1d}$$

where $\lambda(\mathcal{M})$ denotes the eigenvalues of the state transition matrix and π_0 is the initial state. As shown, as long as the system is *indecomposable*, that is as long as $(P_{GB} + P_{BG}) \in (0, 2)$,

there is a unique eigenvalue with magnitude 1 and therefore, after a sufficiently, theoretically infinitely, large amount of time, the following will hold

$$\lim_{n \to \infty} \mathcal{M}^n \pi_0 = \pi_1^* \text{ s.t. } \pi_1^* = \mathcal{M} \pi_1^*. \tag{2.2}$$

This holds for all initial vectors π_0, and therefore the choice of the initial condition is irrelevant. This vector π is given, normalized so all probabilities add to one, as

$$\pi_1^* = \frac{1}{P_{tot}} \begin{pmatrix} P_{BG} \\ P_{GB} \end{pmatrix}. \tag{2.3}$$

The second eigenvector, corresponding to λ_2 is

$$\pi_2^* = \begin{pmatrix} -1 \\ 1 \end{pmatrix}. \tag{2.4}$$

In the next step, we analyze the convergence if we do not allow for infinite iterations. To achieve this, we take an arbitrary vector $\pi_{0,a}$, that is parametrized by the constant a, and decompose it into the eigenvectors

$$\begin{aligned} \pi_{0,a} &= \begin{pmatrix} a \\ 1-a \end{pmatrix} \\ &= \pi_1^* + \left(\frac{1}{P_{tot}} P_{BG} - a \right) \pi_2^* \\ &= \pi_1^* + \frac{1}{P_{tot}} \left((1-a)P_{BG} - aP_{GB} \right) \pi_2^*. \end{aligned} \tag{2.5}$$

With this result, we can now describe the state probablities at the nth state depending on the eigenvalues and the parameter a

$$\pi_{n,a} = \frac{1}{P_{tot}} \begin{pmatrix} P_{BG} - \lambda_2^n[(1-a)P_{BG} - aP_{GB}] \\ P_{GB} + \lambda_2^n[(1-a)P_{BG} - aP_{GB}] \end{pmatrix} \tag{2.6a}$$

$$= \pi_1^* + \begin{pmatrix} -\Delta_n \\ +\Delta_n \end{pmatrix} \tag{2.6b}$$

$$\Delta_n = \underbrace{\lambda_2^n}_{|\cdot| \le 1} \underbrace{\left[\frac{1-a}{P_{tot}} P_{BG} - \frac{a}{P_{tot}} P_{GB} \right]}_{|\cdot| \le 1}. \tag{2.6c}$$

Fig. 2.6 shows the magnitude of the nth power of λ_2 depending on λ_2.
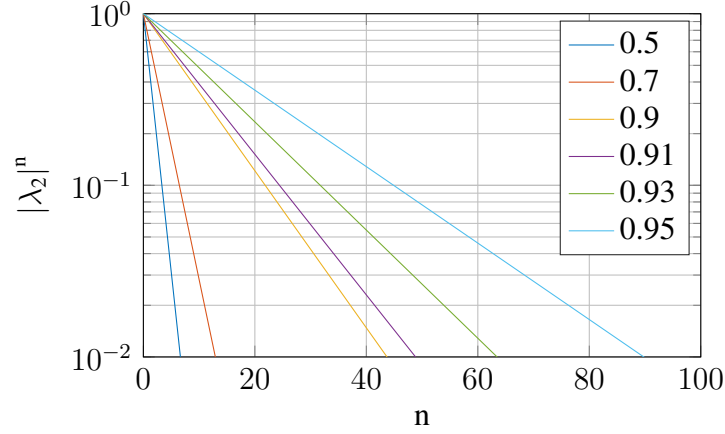
Figure 2.6: Influence of λ_2 after n iterations.

The figure shows that $\lambda_2 = 0.9$ still reaches the 1% in 43 iterations. Taking into account the packet arrival rates of $800\,s^{-1}$, this happens in a twentieth of a second. Even a λ_2 of 0.95 reaches the 1% mark in 90 iterations, or under one tenth. The analysis however is usually pessimistic, as the second factor of Δ_n can in many cases never reach 1. It equals the change of the steady state probabilities before and after a parameter change. Our final results will show λ_2 to be around 0.9 or lower, with an average of 800 events per time interval, meaning that the choice of the initial probabilities has little influence on the state probabilities for the majority of events, as transient properties subside fast.

2.3　Capacity of the Modified Gilbert Model for Packets

In this section, we look for a closed form expression for the capacity of the modified Gilbert model depending on the model parameters. The analogous derivation for the bit model is not relevant for the remainder of the thesis, but was added in Appendix B for the sake of completeness.

The capacity for this problem is introduced by Cover and Thomas [33] as

$$C = \lim_{N \to \infty} \frac{1}{N} \max_{p_X} I(\mathbf{X}^N; \mathbf{Y}^N) \tag{2.7}$$

$$= \lim_{N \to \infty} \frac{1}{N} \max_{p_X} \left(H(\mathbf{Y}^N) - H(\mathbf{Y}^N | \mathbf{X}^N) \right). \tag{2.8}$$

In this notation, boldface signifies an array, and the exponent of an array stands for its dimension. Therefore, we look at asymptotically infinitely long sequences. Furthermore, $I(\mathbf{X}^N; \mathbf{Y}^N)$ is the mutual information of $\mathbf{X}^N$ and $\mathbf{Y}^N$, and $H(\mathbf{Y}^N)$ and $H(\mathbf{Y}^N | \mathbf{X}^N)$ are unconditional and conditional entropies

Using this formula from [33],

$$H(Z_1, Z_2, \ldots, Z_{i-1}) = \sum_{i=1}^{n} H(Z_i | Z_{i-1}, \ldots, Z_1) \tag{2.9}$$

and the fact that for infinitely long history all this conditional entropies converge towards each other, the capacity formula is rewritten

$$C = \max_{p_X} H(Y_i | \ldots, Y_{i-2}, Y_{i-1}) - H(Y_i | \ldots, Y_{i-2}, Y_{i-1}, \mathbf{X}). \tag{2.10}$$

The prefactor N was cancelled with the sum, since all sum terms were equal. Note that the notation $\mathbf{X}$ without superscript stands for the infinitely long sequence. This poses no problem, as the following derivation does not require actual knowledge of this infinitely long sequence.

To analyze the packet capacity, we look at the erasure description of Sec. 2.1.1. When conditioned on $\mathbf{X}$, $\mathbf{Y}$ is perfectly known, except on the occasions where an erasure occurs. Therefore, the entropy is equal to that of the Gilbert source channel conditioned on its past. Additionally, the probabilities of the outcomes of Y is described. Note that the notation $\mathbf{Y}_{i-}$ means the past of Y_i and is used to simplify notation

$$H(Y_i | \mathbf{Y}_{i-}, \mathbf{X}) = H(Z_i | \mathbf{Z}_{i-}) \tag{2.11a}$$

$$\Pr(Y_i = 1 | \mathbf{Y}_{i-} = \mathbf{y}) = (1 - \Pr(e | \mathbf{Z}_{i-} = \mathbf{z})) \Pr(X_i = 1) \tag{2.11b}$$

$$\Pr(Y_i = 0 | \mathbf{Y}_{i-} = \mathbf{y}) = (1 - \Pr(e | \mathbf{Z}_{i-} = \mathbf{z})) \Pr(X_i = 0) \tag{2.11c}$$

$$\Pr(Y_i = e | \mathbf{Y}_{i-} = \mathbf{y}) = \Pr(e | \mathbf{Z}_{i-} = \mathbf{z}) = \Pr(e | \mathbf{Y}_{i-} = \mathbf{y}) \tag{2.11d}$$

In these equations, we used the fact that if we are only interested in erasures, knowledge of $\mathbf{y}$ and $\mathbf{z}$ are equivalent, as both contain full information of when an erasure happened and when not. Therefore, conditioning on a given noise sequence is equivalent for e to conditioning on the respective received sequence.

Finally, we need to evaluate the expression for the conditional entropy $H(Y_i | \mathbf{Y}_{i-})$, by inserting the results from Eq. (2.11)

$$H(Y_i|\mathbf{Y}_{i-}) = \sum_{y \in \mathcal{Y}} \Pr(\mathbf{Y}_{i-} = \mathbf{y}) \Big(\Pr(Y_i = 1|\mathbf{Y}_{i-} = \mathbf{y}) \log_2(\Pr(Y_i = 1|\mathbf{Y}_{i-} = \mathbf{y}))$$

$$+ \Pr(Y_i = 0|\mathbf{Y}_{i-} = \mathbf{y}) \log_2(\Pr(Y_i = 0|\mathbf{Y}_{i-} = \mathbf{y}))$$

$$+ \Pr(Y_i = e|\mathbf{Y}_{i-} = \mathbf{y}) \log_2(\Pr(Y_i = e|\mathbf{Y}_{i-} = \mathbf{y})) \Big). \tag{2.12}$$

By inserting the properties from Eq. (2.11) and regrouping the terms, we arrive at

$$H(Y_i|\mathbf{Y}_{i-}) = \sum_{y \in \mathcal{Y}} \Pr(\mathbf{Y}_{i-} = \mathbf{y}) \Bigg(\underbrace{(1 - \Pr(e|\mathbf{Z}_{i-} = \mathbf{z}) \Pr(X_i = 1) \log_2((1 - \Pr(e|\mathbf{Z}_{i-} = \mathbf{z}))}_{=H(Y_i=1|\mathbf{Y}_{i-}=\mathbf{y},X_i=1)}$$

$$+ \underbrace{(1 - \Pr(e|\mathbf{Z}_{i-} = \mathbf{z}) \Pr(X_i = 0) \log_2((1 - \Pr(e|\mathbf{Z}_{i-} = \mathbf{z}))}_{=H(Y_i=0|\mathbf{Y}_{i-}=\mathbf{y},X_i=0)}$$

$$+ \underbrace{(\Pr(e|\mathbf{Z}_{i-} = z) \log_2((\Pr(e|\mathbf{Z}_{i-} = \mathbf{z}))}_{=H(Y_i=e|\mathbf{Y}_{i-}=\mathbf{y},X_i\in\{0,1\})}$$

$$+ \underbrace{\sum_{x \in \{0,1\}} (1 - \Pr(e|\mathbf{Z}_{i-} = \mathbf{z}) \Pr(X_i = x) \log_2(\Pr(X_i = x))}_{=(1-\Pr(e|\mathbf{Z}_{i-}=\mathbf{z})H(X_i)} \Bigg)$$

$$= H(Y_i|\mathbf{Y}_{i-},\mathbf{X}_{i-}) + \sum_{\mathbf{y} \in \mathcal{Y}} \Pr(\mathbf{Y}_{i-} = \mathbf{y})(1 - \Pr(e|\mathbf{Y}_{i-} = \mathbf{y}))H(X_i). \tag{2.13}$$

The third sum term allows X_i to be anything, since if Y_i is e, the influence of X vanishes. At this point we combine our results and input them in the orignal Eq. (2.10)

$$C = \max_{p_X} H(Y_i|\mathbf{Y}_{i-}) - H(Y_i|\mathbf{Y}_{i-},\mathbf{X}_{i-})$$

$$= \max_{p_X} \sum_{y \in \mathcal{Y}} \Pr(\mathbf{Y}_{i-} = \mathbf{y})(1 - \Pr(e|\mathbf{Y}_{i-} = \mathbf{y}))H(X_i)$$

$$= 1 - \Pr(e). \tag{2.14}$$

In this last step, we see that the conditioning of the erasure probability is summed out, and the maximum entropy for the binary source is 1.

The unconditional erasure probability $\Pr(e)$ is composed of the steady state probability of being in the bad state and the packet error probability in the bad state. This is described depending on the initial state and the nth step as given in Eq. (2.6) as

$$\Pr(e) = \frac{P_{GB}}{P_{ges}}P_e + \Delta_n P_e. \tag{2.15}$$

If it is then assumed that the influence of the initial state is neglectible, we can ignore the Δ term and arrive at the formula for the steady state capacity of the binary erasure channel

$$C_{steady} = 1 - \frac{P_{GB}}{P_{ges}}P_e. \tag{2.16}$$

This result suggests that in the packet channel, the actual past is irrelevant, only the overall error probability matters. This is explained as follows. In a bitflip channel, if the receiver knows that the last 1000 bits were detected wrongly, it can use this knowledge, and flip the next bit right away, assuming to correct it this way.

The erasure works differently. If 1000 consecutive packets are lost due to failed CRC checks, the best thing for the 1001st restransmission is still to try to detect it normally. The erasure allows no guessing about the original symbol, and therefore the best bet is retransmission if the erasure occurs. Therefore, knowing when one will occur does not help to increase the capacity.

2.4 Performance Parameter Estimation

Now we wish to estimate the parameters of the modified Gilbert model from real-world measurements. To this end, we have to define the time interval length that we will use for the remainder of this thesis. Afterwards, we present how the model parameters are extracted from the measurement.

2.4.1 Time Intervals

On the one hand, we aim at characterizing the time-variance of the V2V channel accurately. Therefore, the chosen time interval should be short. On the other hand, the modified Gilbert model is dynamic and stochastic featuring transient behaviour. This motivates choosing larger time intervals to make sure that the transients die out. Finally, we need a sufficient number of transmissions per chosen time interval in order to reduce the variance of the parameter estimates.
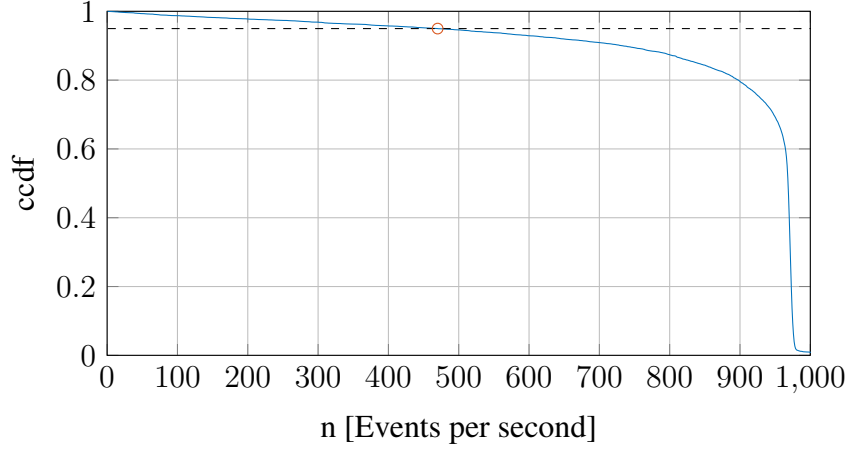
Figure 2.7: CCDF of events per second and 95% line.

First, we need to determine the amount of time it takes for the process to generate a statistically significant amount of events. To this end, Fig. 2.7 shows the Complementary Cumulative Distribution Function (CCDF) of the number of transmission events per second. Additionally, the dashed line indicates the 95 % line. As the figure shows, 95 % of all 1 second intervals saw at least roughly 500 transmission events happening. In 99 % of the cases, there were at least 80 events. This knowledge can then be combined with the results from Fig. 2.6. There we stated, that the influence of the parameter change on the capacity has decreased to below 0.01 in less than 50 events for parameters choices that are relevant to our results.

Together, these facts suggest that 1 s is actually a very good choice for the interval sizes. We showed that in 1 s, the transient properties will usually be finished in less than a tenth of the time. The model also has on average 800 events to compute the parameters. On the other hand, 1 s is still a short interval in which the statistical properties of the V2V channel can be reasonably assumed to stay constant, and represents a resolution that is fine enough to capture even short time influences such as static interferers, as is shown in a later chapter. Following this argument, we will base all further analysis on 1 s time intervals.

2.4.2 Parameter Estimation

Durbin et al. proposed a method for parameter estimation from the output sequence of a hidden Markov model in [34]. The algorithm is based on the Viterbi algorithm, and implemented in MATLAB in the command `hmmviterbi`. We also use this method. For illustration purposes, we ran the algorithm on the individual intervals for measurement 7. The results are seen in Fig. 2.8.

Now that we have those states per second, we can estimate the PDR per second. There are two approaches to do this. The first approach is use the model parameters to generate a
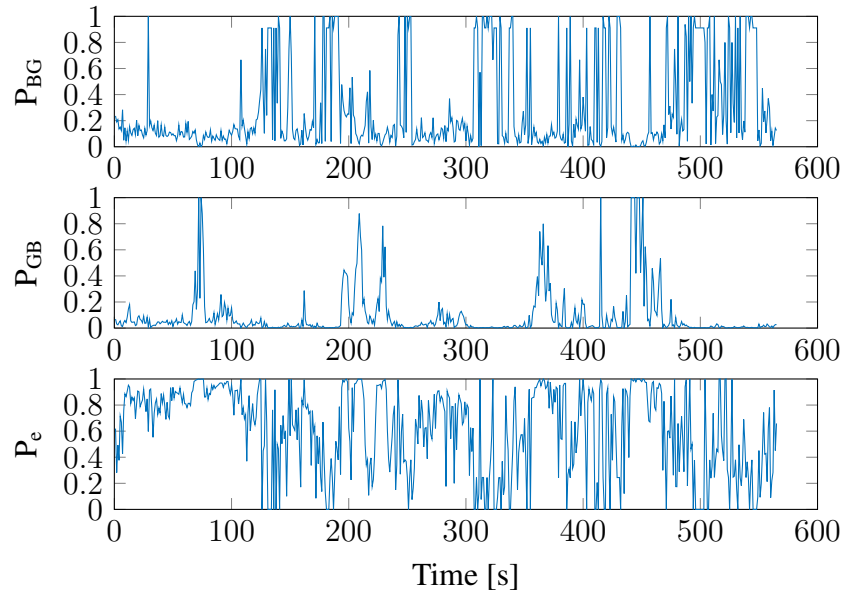
Figure 2.8: State evolution for the per second model calculation.

CRC trace, and calculate the PDR from this trace. The second method is to use the capacity formula that was derived in Eq. (2.16). For this case, we use the steady state version of the formula. The results of both ways are compared to the original PDR in Fig. 2.9.
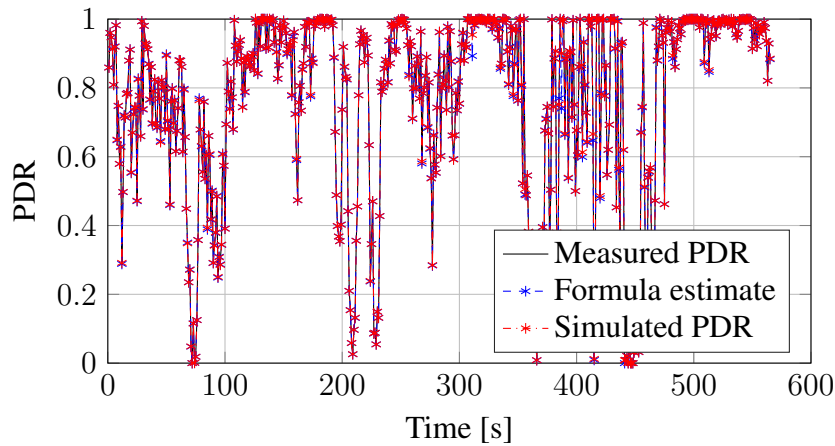
Figure 2.9: Original PDR compared to simulated and calculated PDR estimates.

The figure indicates that both modelling approaches provide the ability to simulate the PDR perfectly. More interestingly, the estimate based on the steady state capacity formula does not experience recognizable deterioration, although it ignores the initial states of the second-long intervals. This is a strong indication that the earlier assertion, that $1\,$s intervals are sufficient to suppress transient influences is true. Indeed, in $97\,\%$ of all times, the absolute difference between the original PDR and the steady-state capacity estimate was below 0.01.

The MSE between the steady-state capacity estimates and the original PDR was $8.3 \cdot 10^{-5}$, showing that the Gilbert model based estimation is in principle perfectly able to capture the PDR properties.

The approach presented here is however merely a toy example to show the capabilities of the model. Using 3 parameters per second is not a practical way to reproduce our measurements. Therefore, the next chapter deals with ways to estimate the model parameters rather than compute them. These estimates will be based on physical quantities such as SNR or distance.

3

Performance Estimation

So far, we presented the modified Gilbert model designed to generate a packet-error stream. The ultimate goal of this is, as stated, to be able to reproduce the packet behaviour of real-world measurements. To this end, we need to find a rule to set the parameters of the model over time. Obviously, the way the Gilbert model was applied in the toy example in the previous section is useless, since we need the packet trace for every second seperately to produce an estimate, which would then be used to hopefully reproduce a similar packet trace.

The real benefit of the Gilbert model only comes in to play if it is possible to identify different time instances that have the same, or at least similar, parameters. Then, those instances can be grouped, and they be assigned a model paramater triple. By doing this, we end up with a finite set of model paramter triples, and every second of every measurement is assigned to one of them.

The next section describes the outlined approach in detail. Afterwards, we present the chosen evaluation metric. Subsequently, the outlined approach is analyzed, both when based on SNR and when based on the distance. Finally, a modification to the approach is introduced, and again evaluated.

3.1 Measurement-based Performance Estimation

In this section, we look for a way to classify time intervals as equal, or at least comparable, based on an underlying physical quantity. A simple way to approach this problem would be to assume equality of two time intervals if some parameters are close enough. An example for this would be to split the measured distance in equidistant intervals, and assume equality within the respective intervals. This was done for V2I measurements in [12]. There it makes sense for two reasons. Firstly, multiple measurements over the same stretch of road were made. Therefore, the channel from within one interval was actually comparable with the same interval from another measurement. Secondly, since the relative speed on the highway will be approximately constant, all of those distance intervals will take roughly the same time to pass through, meaning that all intervals get a fair share of the measurements. This in turn is very relevant as the behaviour of the underlying Gilbert model depends on the ability to approach steady state. To give a very simple example why the equidistant distance seperation does not work in the V2V case, let one interval border lie at $80\,\mathrm{m}$. If the cars happen to be driving at exactly this distance, they will constantly cross the interval border, chopping the packet trace in small chunks of no statistical significance. The first assumption is also violated. We typically only had 3-6 measurements actually passing the same stretch of highway, and no guarantee they happened at the same distance, or in the same LOS conditions.

Therefore, we stick to the approach from the previous chapter, and split the measurement in constant $1\,\mathrm{s}$ invervals. This means an average of 800 packets per interval, which was shown in Sec. 2.4.2 to be sufficient to reach steady state after a possible parameter change. For each of these intervals, we calculated a mean distance in meters and mean SNR in dezibels.

The key steps that are taken to arrive at a performance estimate are depicted in Alg. 1. The algorithm takes 3 inputs. It results in the physical quantity being split into intervals, and each of those intervals is assigned one parameter triple. This can then be used for generating packet streams given only an arbitrary stream of the physical value.

The first input is the trace of any physical ϕ. As mentioned before, the method will be combined with both the SNR and the distance, but not at the same time, and the algorithm works with any parameter, as long as it is correlated with the packet performance. If it is not correlated, the estimation based on such a quantity would be pure guessing. In line 2, we can see that the first step is to group the physical quantity in a set amount of intervals, each of which will be assigned one performance parameter triple. we do this because we have to group our measurements into intervals. Since both SNR and the distance are continous quantities, no two time instances will give the same measurement. However, the next section will illustrate that the model approach is stochastic, and that it needs multiple realizations per value of ϕ to work. The way around this problem is to define intervals. Therefore, we might

as well make the number of intervals a design parameter. This fact also works well together with the stated goal to find a simple, yet realistic performance model. As will be shown in the evaluation, this approach is further validated by the fact that only a small number of intervals is needed to achieve the minimal estimation error for our cases.

Data: Physical parameter ϕ and CRC information, Number of Intervals k
Result: ϕ-depentent performance estimation model

1 **begin**
2   Split ϕ in k intervals $\tilde{\Phi}_k$
3   **for** *Every interval $\tilde{\Phi}_k$* **do**
4   Get all time instances where ϕ lies in that inverval
5   Combine CRC-traces of all those instances
6   Run Viterbi algorithm over combined trace to get a new (P_{BG}, P_{GB}, P_e) for the interval
7   **end**
8 **end**

Algorithm 1: Illustration of Measurement based performance estimation.

Once the physical quantity is divided into intervals, it is rather straightforward to split the measurements accordingly, and use the Viterbi algorithm as explained in Sec. 2.4.2 to come up with performance estimate for the interval. However, coming up with the optimal way to divide the physical quantity into intervals is not trivial, and will be explained in the next sections.

3.1.1   Dependence between Physical Quantities and Performance Model

First, the dependence between the physical quantity ϕ and the performance model has to be described. Depending on the approach, ϕ might stand for distance or SNR. Furthermore, as we stated we have to split the domain of ϕ in N intervals $\phi_1, \ldots, \phi_N$. Since neither the PDR-distance dependence, nor the PDR-SNR dependence appears to be deterministic, we will model it stochastically.

All three parameters (P_{BG}, P_{GB}, P_e) are probabilities and therefore can only be in the domain $[0, 1]$. For each interval ϕ_n, we can now compute the conditional pdf of the parameters. This conditional pdf is calculated empirically by the `ksdensity` tool provided by MATLAB.

Since we are however primarily interested in the throughput analyis, it turns out that modelling the influence of ϕ on the steady state capacity $C = 1 - \frac{P_{GB}}{P_{ges}} P_e$ is sufficient. For this, we fitted the empirical pdf $p_{C|\phi}(C|\phi \in \phi_n)$.

3.1.2 Finding the Optimal Intervals

As first step to the interval division, we divide the measurements into fine-grained intervals. These intervals are supposed to be far too many and too fine grained for our purposes. The reason behind this is that we need initial pdfs $p_{C|\phi}(C|\phi \in \phi_n)$. Only if we have these, we can then proceed to reduce the amount of intervals in an optimal way. Therefore, we chose the interval size as $1\,\mathrm{dB}$ for SNR intervals and $10\,\mathrm{m}$ for distance. These intervals are sufficiently small such that they don't influence further processing. They are furthermore large enough that each interval see contains a sufficient amount of measurements, ensuring stochastic relevance. This results in roughly 50 intervals for SNR and 30 for distance. Now it is possible to precompute all conditional pdfs $p_{C|\phi}(C|\phi \in \phi_n)$ for all n. The goal now is to find new intervals $\tilde{\phi}_n$ which are far fewer in number, but retain the maximum amount of information when compared to the original intervals. The optimal solution to finding those intervals is given by the information bottleneck method [35]. This states that the new intervals should be chosen such that the expected Kullback-Leibler divergence $D_{KL}(P||Q)$ [36] is minimized. The Kullback-Leibler divergence is a measure for the information lost when P is approximated by Q. Our optimal set of intervals then becomes

$$\tilde{\Phi}_{opt}^{(K)} = \arg\min_{\tilde{\Phi}^{(K)}} \sum_{n=1}^{N} D_{KL}(p_{C|\phi_n}(C|\phi \in \phi_n)||p_{C|\tilde{\phi}_k}(C|\phi \in \tilde{\phi}_k) \Pr(\phi \in \phi_n). \tag{3.1}$$

As will be shown later, we never needed more than 6 intervals. This combined with the limited valid region for our quantities allowed to look for the optimal interval division by exhaustive search. To find the optimal intervals, we evaluated all possible interval combinations for a certain number of intervals, and chose the one with the minimum Kullback-Leibler divergence. This approach through exhaustive search is possible due to the nature of our problem.

3.1.3 Estimation

In this section, we applied Alg. 1 for interval counts k of 2-6 both based on the SNR and the distance. Table 3.1 shows where the optimal interval borders lie after applying the information bottleneck. The distance dependent result is very remarkable, since when adding another interval, the previous set of borders remains intact and there is merely an additional border added. This is in contrast to the SNR case, where all interval borders shift from one step to the next.

Table 3.2 depicts the capacity estimates, shown in the same order as the intervals of the previous table. The estimates were calculated by taking the (P_{BG}, P_{GB}, P_e) estimates and

k	Interval borders [dB]
2	15
3	8, 15
4	6, 9, 15
5	6, 8, 11, 16
6	6, 8, 12, 20, 24

(a) SNR

k	Interval borders [m]
2	80
3	20, 80
4	20, 80, 250
5	20, 30, 80, 250
6	20, 30, 40, 80, 250

(b) distance

Table 3.1: Results of the optimal interval choices.

k	Capacity
2	0.65, 0.92
3	0.4, 0.78, 0.92
4	0.23, 0.57, 0.79, 0.92
5	0.23, 0.51, 0.72, 0.83, 0.93
6	0.23, 0.51, 0.73, 0.87, 0.94, 0.96

(a) SNR

k	Capacity
2	0.77, 0.7
3	0.94, 0.72, 0.7
4	0.94, 0.72, 0.71, 0.67
5	0.94, 0.76, 0.71, 0.71, 0.67
6	0.94, 0.76, 0.69, 0.71, 0.71, 0.67

(b) distance

Table 3.2: Capacity estimates corresponding to the intervals.

plugging them into the steady state capacity formula from Eq. (2.16). The capacity estimates increase monotonically with increasing SNR and decrease with increasing distance, which is to be expected. The distance estimate does not provide a low capacity estimate, as the lowest estimate is 0.67. The reason for this will become apparent in Sec. 3.4, where we show that the influence of the distance is overshadowed by stronger influences such as the NLOS-LOS distinction. Estimating purely based on the distance does not account for these influences and results in averaging effects for every distance segment. Therefore, the distance-based estimate is close to the overall average PDR for almost all interval choices.

3.2 Evaluation of Performance Estimation

In this section, we applied the SNR- and distance-dependent estimation schemes from the last section to our measurements, and evaluated their performance. To this end, we first present the performance metric used. Afterwards, we present the results obtained from the estimation schemes.

3.2.1 Estimation Performance

The primary goal of this thesis is analyzing the properties of the capacity. Therefore, the metric employed is based on how well our estimation model is able to predict the achieved PDR. For this, we use the mean Mean Squared Error (MSE) of the estimated PDR over all included measurements. The PDR is, as always, calculated on a per second basis as

$$MSE = \frac{1}{T_{total}} \sum_{t \in \mathcal{T}} \left(\hat{PDR}[t] - PDR[t] \right)^2. \tag{3.2}$$

In this equation, $\mathcal{T}$ stands for the set of all time instances in all measurements, and T_{total} is the total recorded time in seconds.

$\hat{PDR}$	MSE
1	0.17
0.5	0.14
0.72	0.09

Table 3.3: Average MSE for constant PDR guesses..

Table 3.3 depicts the results when a constant is used as PDR estimate, with three choice for the constant. The first estimate is the "perfect transmission" assumption that is currently implemented in many network simulators. The second estimate, 0.5 is the minimax estimate [37]. These two do not use the data at all, but are merely guesses based on some argument. Finally, the third estimate is the mean PDR over all measurements, which is the best constant estimate in the MSE sense. From this we can see that we can already improve over the assumption that all packets are received by picking a better constant PDR estimate, however the large steps also suggest that there is still room for improvement.
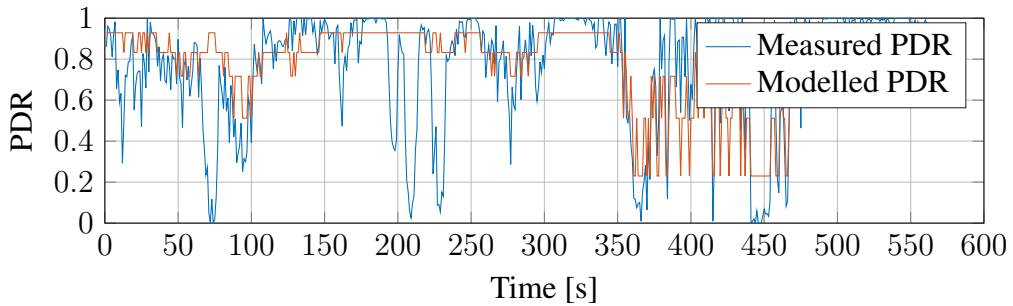
3.2.2 SNR-based Estimation

Figure 3.1: Comparison of measured and SNR-based model PDR for 5 intervals for measurement 7.

Here, we use the measured SNR as physical quantity ϕ. The SNR was highly correlated to the PDR, as is shown in Sec. 1.4.1. Therefore, the estimation based on the SNR is expected to work reasonably well. The originally measured SNR lies in the region between 0 and $50\,\mathrm{dB}$.

Fig. 3.1 compares the measured and modelled PDR for measurement 7. We see qualitatively that the estimation is able to follow the measurement in many instances, but not in all. The following sections contain a thorough investigation of this behaviour and analyze the influence of the number of intervals, the behaviour of the capacity estimates, and finally takes a look at the time behaviour of the MSE.
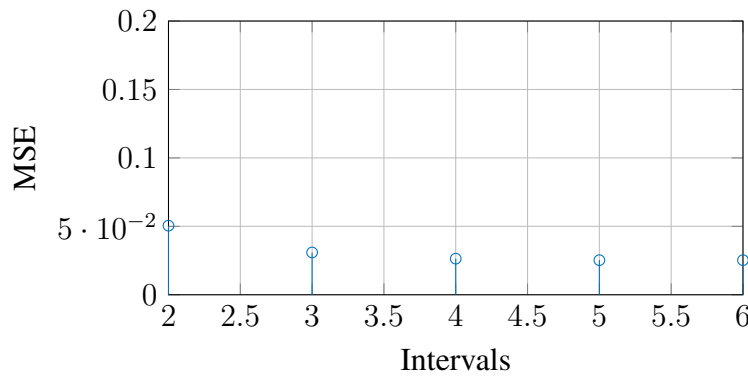
SNR Intervals

Figure 3.2: Average MSE compared to the number of intervals.

Fig. 3.2 shows the overall average MSE of the SNR-based estimation method for 2-6 intervals. Compared to the best constant PDR estimate, which provided an MSE of 0.09, we already see that the approach was able to reduce that by more than half.

The figure shows that using more than 5 intervals does not decrease the MSE further, and even at 3 intervals, the result is almost as good as it gets. This figure shows why we claimed in Sec. 3.1.2 that we do not need to investigate more than 6 intervals. Obviously, already at relatively low interval numbers, the main contributor to the MSE is not captured by SNR dependencies and increasing the interval count does not help anymore.

In the following analysis, we will use the estimation model using 5 intervals.

Capacity Estimates

Fig. 3.3 shows the steady state capacity estimates that are assigned to every dB step of the SNR for a 5-interval model. The predicted capacity grows monotonically with the SNR, which is to be expected. It must be noted, that even at low SNR values, the throughput prediction does not decrease to 0, and that the maximum estimate is reached at roughly $15\,\mathrm{dB}$.
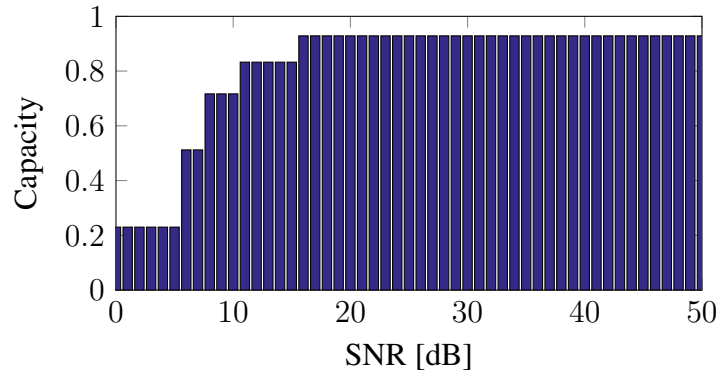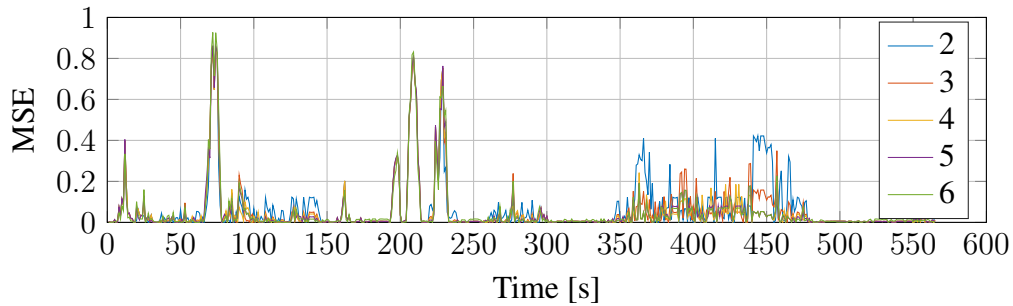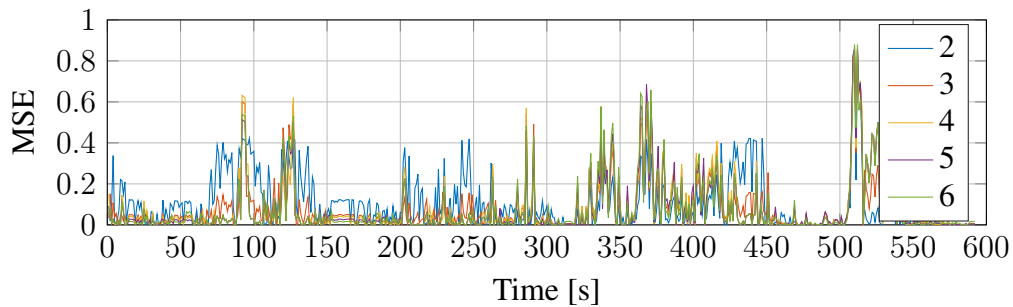
Figure 3.3: Capacity estimate over SNR for 5 intervals.

Furthermore, the quickest changes happen at around $10\,\mathrm{dB}$, which is why the intervals are the narrowest there.

MSE

(a) Measurement 7

(b) Measurement 8

Figure 3.4: MSE over time for two measurements for $2 - 6$ intervals.

Fig. 3.4 shows the behaviour of the MSE for measurements 7 and 8. The two measurements were chosen as illustrations because both of them pass the whole stretch of highway. While measurement 7 was recorded going from northwest to southeast, Measurement 8 went in the opposite direction. Measurement 8 had more reception errors, which lead to being

more difficult to estimate properly. These example measurements show some traits that are found in all measurements. The first thing that stands out are the sharp peaks in the MSE. The MSE is in fact very low for a majority of the time, but at very defined, short periods, the estimation fails completely. When looking for reasons, we see that at those times, the PDR would drop to near 0, while the SNR stayed at a high, or relatively high value. An in-depth analysis of this phenomenon and possible reasons behind it are found in Chapter 4.

The other effect we can see is the influence of the interval counts. Figure 3.2 showed that 3 intervals are already a good estimate, and indeed the time analysis shows that substantial MSE improvements happen at the step from 2 to 3 intervals in many places. However, further increase of the number of intervals doesn't lead to a significant performance improvement.

Finally, we see especially in those places, where the estimation failed, that the interval count hardly matters, suggesting that the model is simply not able to capture those moments.

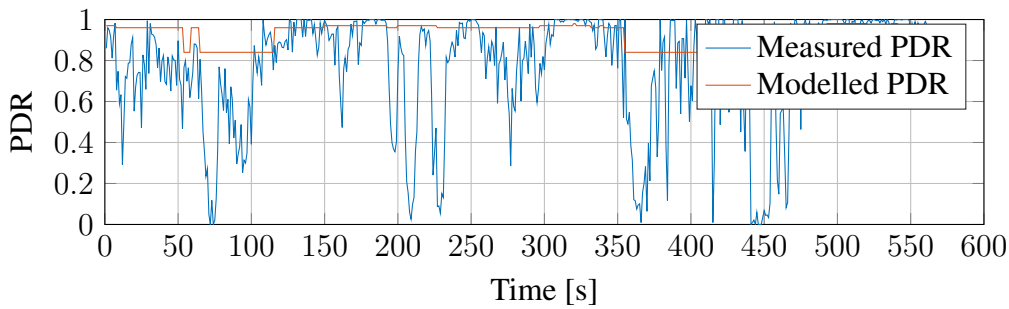### 3.2.3 Distance-based Estimation

Figure 3.5: Comparison of measured and distance-based model PDR for 5 intervals for measurement 7.

Unfortunately, the GPS sensors were not as reliable as the SNR measurements. There were multiple skips and timeshifts, resulting in wrong distance information. In order to keep this influence to a minimum, we restricted the distance based estimation to those time intervals where the distance was reported to be below $300\,\mathrm{m}$. Furthermore, since the valid distances span $300\,\mathrm{m}$ as opposed to $50\,\mathrm{dB}$ in the SNR case, we used $10\,\mathrm{m}$ intervals for the initialization.

In this section we will add the results for SNR based estimation for comparison. Note however, that these SNR results will be computed only for the instances where the distance was valid. They will therefore differ slightly from the graphs in the previous section.

Fig. 3.5 shows the model PDR to illustrate the estimation process. In this case, we see that this estimation fails to follow the measurement, and the reasons behind this will be analyzed in the subsequent sections.
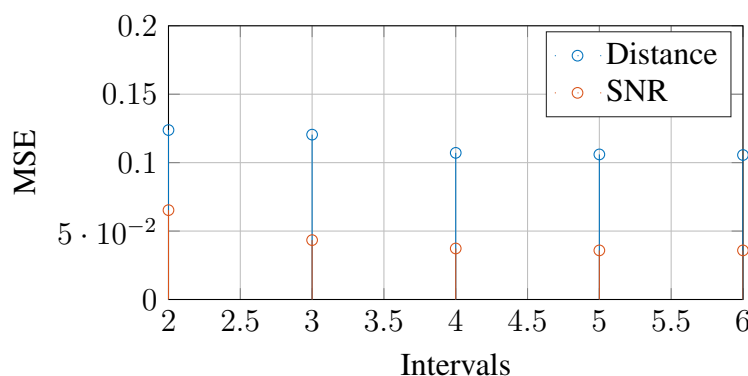
Distance Intervals

Figure 3.6: Average MSE compared to the number of intervals.

Fig. 3.6 shows the overall average MSE for the distance based estimation method. For comparison, we added the same curve for the SNR-based estimation. As is shown, the distance based estimation performs generally worse than the SNR-based one. This is explained with the fact, that distance and PDR are much less correlated than SNR and PDR. For example the difference between LOS and NLOS will obviously have a large impact on the PDR. Blocking the LOS will also influence the SNR, while the distance is oblivious to this factor.

This does, however, not account for the fact how bad exactly the estimation behaves. To recount, estimating the PDR with the constant mean over all measurements yields a MSE of 0.09, which outperforms this distance dependent measurement. The next sections will analyze further how this comes to be, but it is already clear that this approach can not be regarded as a solution to the problem, at least not on its own.
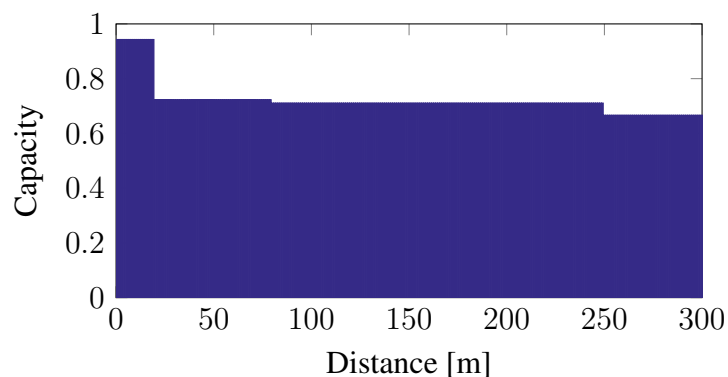
Capacity Estimates

Figure 3.7: Capacity estimate over distance for 5 intervals.

Fig. 3.7 shows the mapping of distance to the capacity estimates for 5 intervals. This figure, together with Fig. 3.6 indicate that a high number of intervals does not improve the estimation. In fact, for the capacity estimate, there are only two distinct regions. Those close to $0\,\mathrm{m}$ with an almost perfect capacity, and the others with a capacity close to 0.7. This result is also interesting insofar as, apart from the region below $20\,\mathrm{m}$, the estimate is very close to the overall mean PDR. This further underlines the fact that this purely distance dependent estimator is not equipped for the task, as it estimates almost everything by something close to the global mean.

MSE
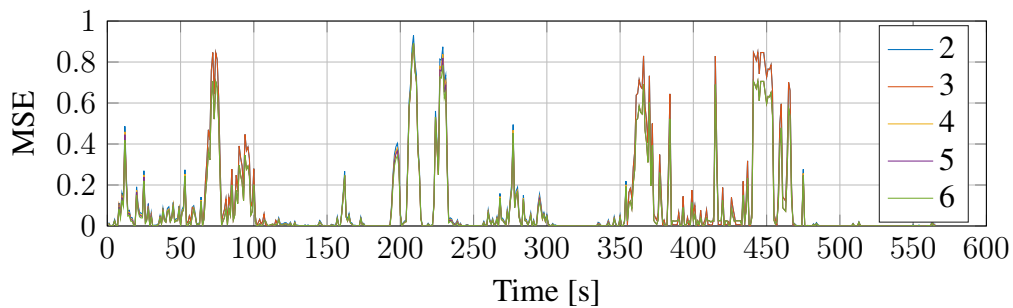
(a) Measurement 7
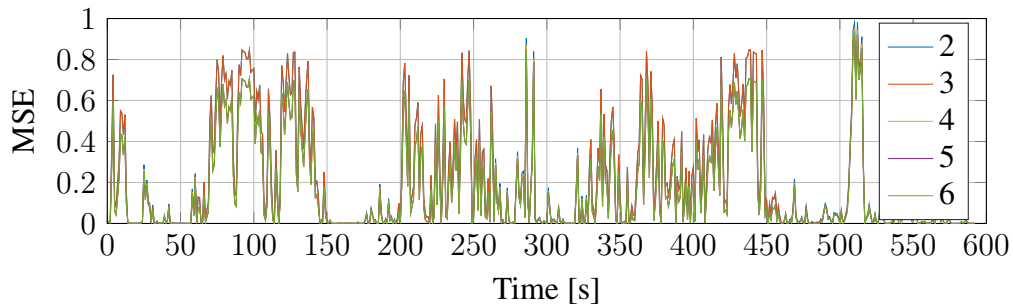
(b) Measurement 8

Figure 3.8: MSE over time for 2 measurements for $2-6$ intervals.

In Fig. 3.8 we see again examples of the estimation MSE for the same 2 measurements as for the SNR case. We see the same effect again as for the SNR, namely the times where the estimation is relatively good, and times where it is bad. This time however, there are more regions that are very bad, and increasing the interval size ostensibly changes almost nothing.

3.2.4 Summary

This mapping of a physical parameter to a performance estimation works as a first step. In the SNR estimation, the times where it does not work are few, and always only for short periods. Therefore, we can state that this estimation constitutes a "usual" case. At this point, we assume that for some reason, the transmission is disturbed in those special cases. The distance dependent case fails at achieving respectable performance altogether. However, in the detailed analysis, we see that the distance dependent is no hopeless case. It also achieves low MSE in many places.

What was left out from the process up till now was of course time-variance of the parameter mapping. The mapping that was proposed here is a purely time-invariant one. Insofar, it is less surprising that the resulting MSE is strongly time variant. Perhaps more surprisingly, the MSE behaves very binary. It is either very good, or very bad. Following this concept, we propose a hidden variable, that can assume two types. The variable is supposed to represent two possible scenarios, "regular" and "irregular". This approach is described in the next section.

3.3 Scenario-Dependent Performance Estimation

We assume that underlying the measurements, one of two scenarios is valid. The channel behaves either as expected, the way our original mapping captured it. Alternatively, it behaves fundamentally different. This approach tries to implement scenario dependence into the estimation in the simplest way possible - a binary variable λ. At any given time interval, it can take one of two values. Formally, this extends the conditional pdf we used to analyze the channel $p_{C|\phi}(C|\phi)$ to $p_{C|\phi,\lambda}(C|\phi,\lambda)$. The next sections will explain how this variable is identified in this case, what influence this extension has on the estimation algorithm, and finally how to calculate a performance estimation model that takes this variable into account.

3.3.1 Overview of Channel Scenarios

If we look back at the previous results, the performance model either works well, or not at all. This implies that the SNR/distance do not contain the full information about the channel. The variable λ is supposed to produce this missing information. Therefore, we label the two values λ can assume as channel scenarios. Our expectation is that the estimate of the first scenario will be equal to the scenario-less estimation. If this is the case, this scenario is called "regular", which describes the usual behaviour. In this case, the second scenario would describe the unusual behaviour.

Reasons for such unsual behaviour can vary. Possibilities include roadside structures that influence the transmission or interferers. In the distance case, the distinction between LOS and NLOS has to be captured. It is, however, also possible that the PDR drops are due to variable performance of the hardware, as this is also part of the described transmission channel.

For this chapter, we differentiate between the channels simply by thresholding the MSE of the original model. It has proven that 0.1 is a sensible choice for the threshold. Everything below this threshold will be regarded as "regular", everything else will be put in scenario 2. In the subsequent estimation, we will take the knowledge of when what scenario is in force as given. The following chapter will then analyze the effects behind the scenarios.

3.3.2 Channel Scenario Assignment

This section describes the assignment of the channel scenarios and the calculation of the scenario-dependent model, which is illustrated in Alg. 2. As is shown, the algorithm requires the physical quantity and the CRC information. The physical quantity is averaged within the 1 s intervals.

The algorithm runs the same routine 2 times. In the first run, it works exactly as the scenario-independent version Alg. 1. In the second run, the results from the first run are thresholded. All following steps are then run seperately for the two steps.

Table 3.4 shows the algorithm output for the SNR-dependent estimation. Scenario 2 never sees more than 2 intervals, which is a result of the information bottleneck algorithm. When compared to Table 3.1 and Table 3.2, we see that scenario 1 includes higher dB-intervals, and higher peak capacity estimates, whereas scenario 2 provides very low capacity estimates. This is the behaviour we hoped to see, as the algorithm is now able to provide the low-PDR estimates that generated a high MSE beforehand. On the other hand, the algorithm now is also capable of providing better estimates for "perfect" conditions, since the PDR-drop effects do not reduce the average PDR anymore in scenario 1.

k	Borders S1 [dB]	Borders S2 [dB]
2	37	30
3	8, 22	29
4	7, 11, 22	29
5	6, 8, 12, 22	29
6	6, 8, 11, 18, 22	29

(a) Interval borders

k	Capacity S1	Capacity S2
2	0.79, 1	0.36, 0.041
3	0.39, 0.86, 0.99	0.36, 0.11
4	0.31, 0.72, 0.9, 0.99	0.36, 0.11
5	0.21, 0.53, 0.77, 0.91, 0.99	0.36, 0.11
6	0.21, 0.53, 0.76, 0.88, 0.95, 0.99	0.36, 0.11

(b) Capacity estimates

Table 3.4: Algorithm output for the SNR case split into the two scenarios (S1 & S2).

    Data: Physical parameter ϕ and CRC information, Number of Intervals k
    Result: ϕ-depentent performance estimation model

1 **begin**
2   **for** $i = 1$ **to** 2 **do**
3   **if** $i = 1$ **then**
4   Split ϕ in k intervals $\tilde{\Phi}_k$
5   **else**
6   Threshold MSE curves from previous loop
7   Assign time instances to scenario 1 or scenario 2 based on thresholding
8   Split ϕ in k intervals $\tilde{\Phi}_k$ for both scenarios seperately
9   **end**
10   **for** *Every interval $\tilde{\phi}_k$ of every scenario* **do**
11   Get all time instances where ϕ lies in that inverval and that belong to that scenario
12   Combine CRC-traces of all those instances
13   Run Viterbi algorithm over combined trace to get a new (P_{BG}, P_{GB}, P_e) for the interval+scenario
14   **end**
15   **end**
16 **end**

Algorithm 2: Scenario dependent performance parameter estimation.

A similar behaviour is seen in Table 3.5 for the distance-dependent estimates. The distance interval borders for both scenarios lie close to the original intervals. However, the capacity now is essentially split into a high-capacity scenario and a low capacity scenario. The distance estimates do not default to the global mean anymore, which is a good sign. It has to be noted, that starting at 4 intervals, the furthest capacity estimate goes up again. This is an artifact due to the GPS problems. Some measurements had the GPS time shifted, which lead to a shift of roughly $300\,\mathrm{m}$ of the distance estimate. Therefore, if an interval is put at above $250\,\mathrm{m}$, these measurements dominate that actually happened at about $100\,\mathrm{m}$.

k	Borders S1 [m]	Borders S2 [m]
2	90	80
3	20, 80	20, 80
4	20, 80, 260	20, 80, 250
5	20, 30, 80, 260	20, 30, 80, 250
6	20, 30, 70, 80, 260	20, 30, 40, 80, 250

(a) Interval borders

k	Capacity S1	Capacity S2
2	0.91, 0.83	0.44, 0.26
3	0.96, 0.88, 0.83	0.43, 0.44, 0.26
4	0.96, 0.88, 0.8, 0.88	0.43, 0.44, 0.24, 0.37
5	0.96, 0.89, 0.88, 0.8, 0.88	0.43, 0.49, 0.43, 0.24, 0.37
6	0.96, 0.89, 0.87, 0.89, 0.8, 0.88	0.43, 0.49, 0.48, 0.42, 0.24, 0.37

(b) Capacity estimates

Table 3.5: Algorithm output for the distance case (S1 & S2) .

3.4 Evaluation of Scenario-Dependent Performance Estimation

This section shows the same evaluations as for the original model. Again, first the results for
SNR are shown, then for distance based modelling. For comparison, the original SNR based
results are added.
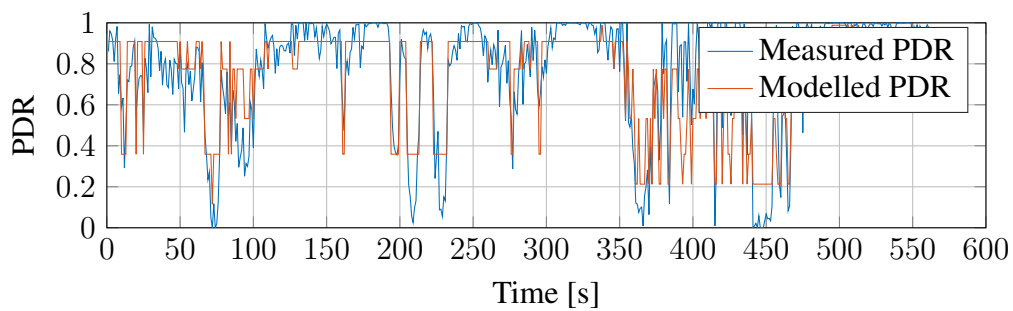
3.4.1 SNR-based

Figure 3.9: Comparison of measured and modelled PDR for 5 intervals for measurement 7.

Fig. 3.9 again provides the PDR comparison for one measurement to illustrate the per-
formance. When compared to Fig. 3.1, we immediately see that the estimation in is able to
follow the PDR drops where the original estimation was not able to. The following sections
will present an analysis of the achieved improvement.
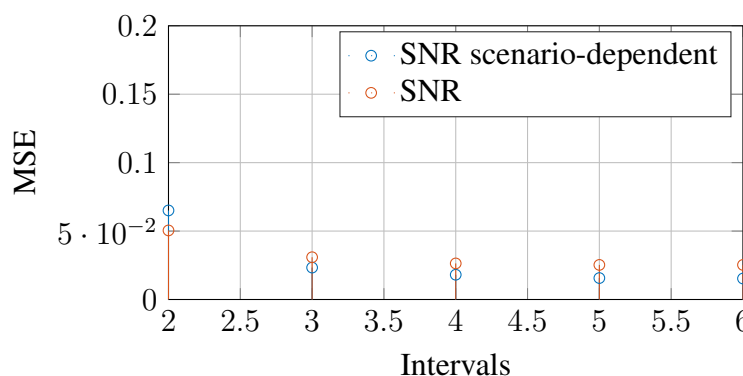
SNR Intervals

Figure 3.10: Average MSE compared to the number of intervals.

Fig. 3.10 shows the influence of the added second channel scenario. While performance decreases for 2 intervals, the MSE is lowered noticably for 3 or more intervals.
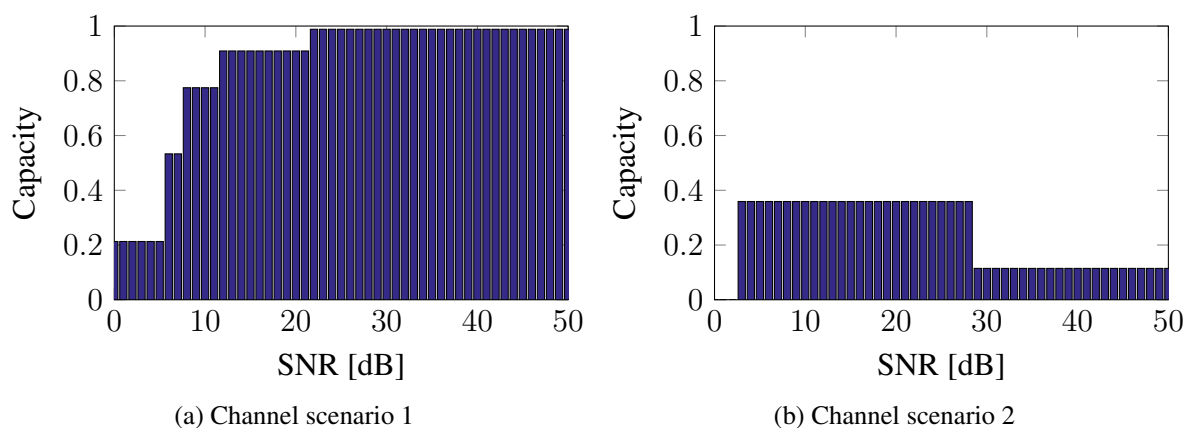
Capacity Estimates

(a) Channel scenario 1   (b) Channel scenario 2

Figure 3.11: Capacity estimate over SNR for 5 intervals for both channel scenarios.

Fig. 3.11 shows the capacity estimates for 5 intervals for both channel scenarios. This illustrates nicely how the regular scenario behaves just like the original version. The second scenario however behaves differently. When comparing the two graphs, there are two things worth noting. In the low SNR regime, the second channel scenario has a slightly higher capacity estimate. In all other regions however, the second scenario has a low capacity estimate, and even more for the very high SNR region.

This result suggests that the aforementioned distinction in "regular" and "inverse" scenario is actually what happens. The second channel scenario adds a higher capacity alterna-

tive to the low-SNR regime, indicating better transmission than would be expected, and a low capacity estimate for the high SNR regime, accounting for bad transmission conditions.
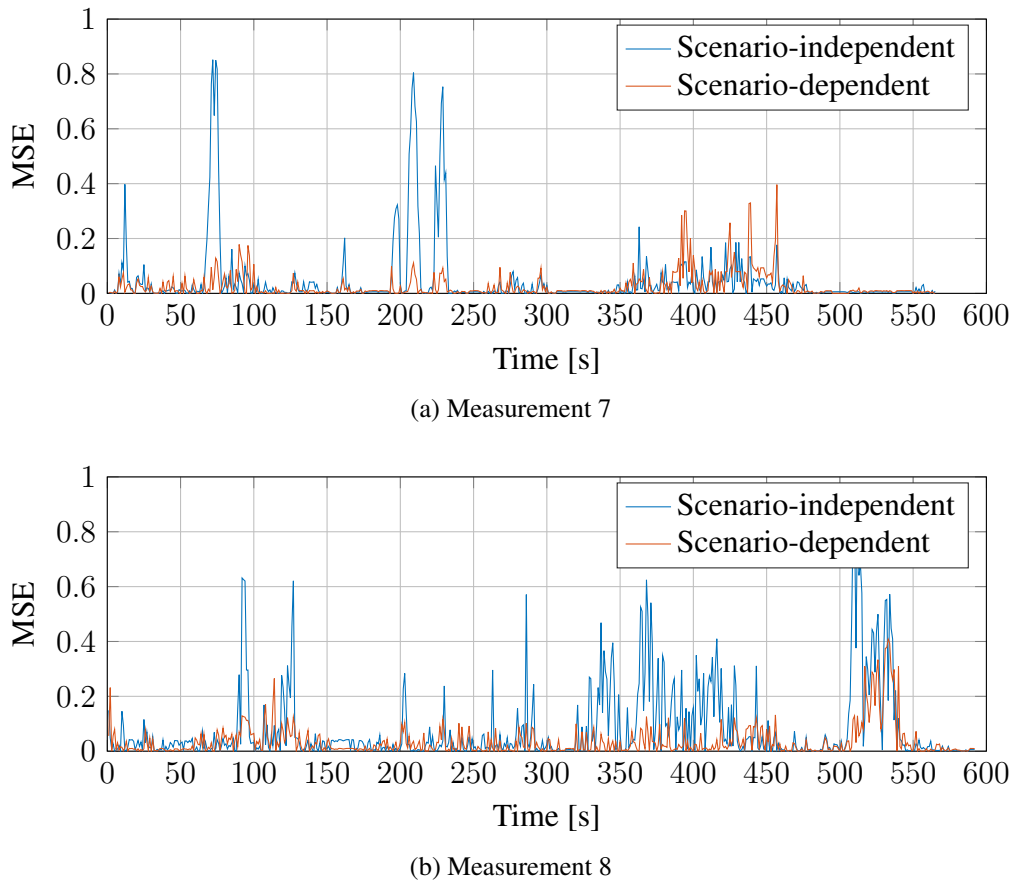
MSE

(a) Measurement 7

(b) Measurement 8

Figure 3.12: MSE over time for 2 measurements for SNR based and scenario-dependent estimations.

Fig. 3.12 depicts the MSE over time for a sample measurement for the pure SNR estimation and the scenario-dependent estimation. The scenario-dependent estimation combats effectively allmost all high peaks from the original estimation scheme. Although there are regions where it actually performs worse, and regions where it does not manage to provide improvements, The figure illustrates how the scenario-dependent estimation is able to significantly reduce the overall MSE.

3.4.2 Distance-based

Fig. 3.13 shows the modelled PDR of the two-scenario distance-based estimation. We can qualitatively see that in contrast to the scenario-independent, distance-based estimation, the
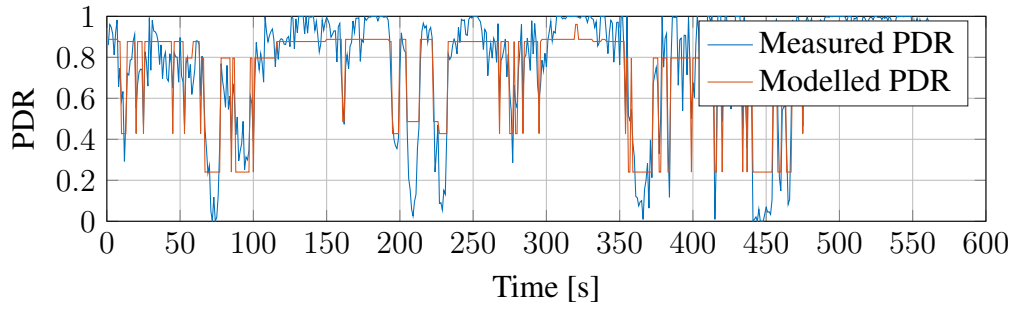
Figure 3.13: Comparison of measured and modelled PDR for 5 intervals for measurement 7.

two-scenario approach enabled the distance-dependent estimation to follow the measurement, and will investigate the estimation quantitatively in the following sections.
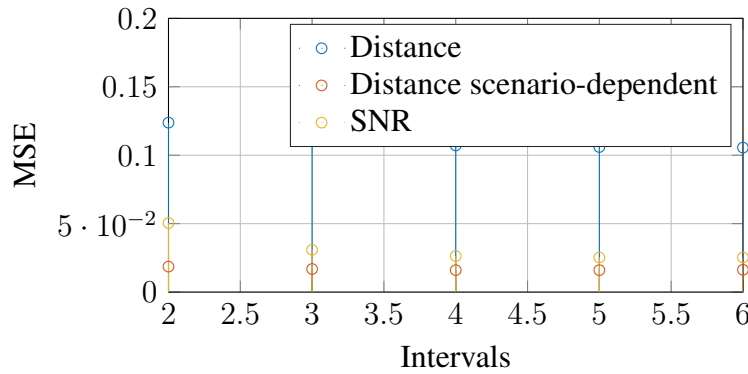
Distance Intervals

Figure 3.14: Average MSE compared to the number of intervals.

Fig. 3.14 shows the performance of the scenario-dependent estimation with respect to the interval count. The figure shows that the number of intervals used is of very little importance. There is a slight decrease from 2 to 3 intervals, but there is no reason to use more than 3. This is in line with the results from Table 3.5, where the capacity estimates have a very close distance behaviour irrespective of the number of intervals.

Capacity Estimates

In Fig. 3.15 we see the capacity estimates that are assigned to the distances in the two channel scenarios. What is interesting, is that the two scenarios behave equally, with a drop in capacity slightly below $100\,\mathrm{m}$. The interpretation of the second scenario as "inverted" channel does not hold for the distance based estimate. Instead, we see only a relatively weak dependence of the capacity on the distance within a range of $300\,\mathrm{m}$. The two scenarios seem to be more in line with a "good" and a "bad" transmission channel, analogous to LOS and NLOS conditions.

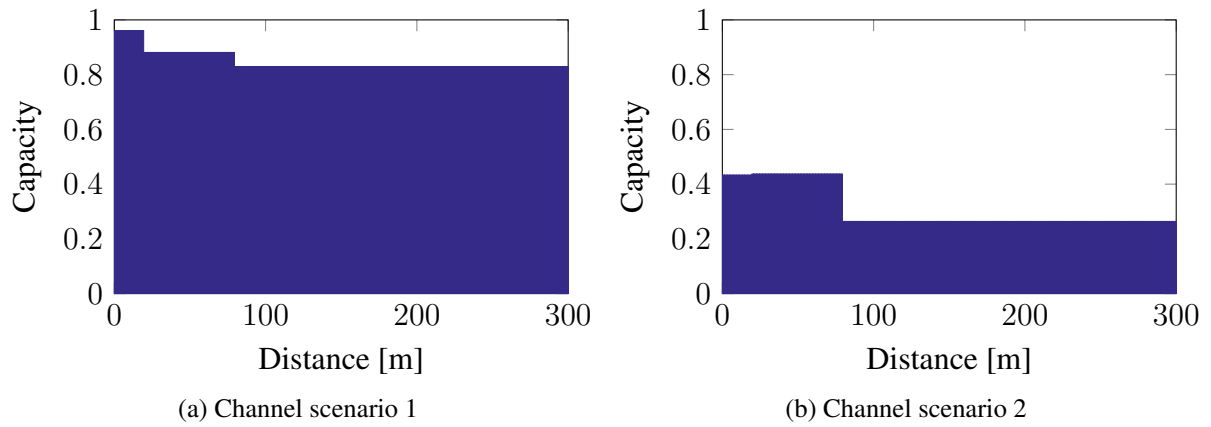(a) Channel scenario 1

(b) Channel scenario 2

Figure 3.15: Capacity estimate over distance for 3 intervals for both channel scenarios.

MSE

Fig. 3.16 shows the MSE over time of the purely distance based estimation and a scenario-dependent estimation. For this measurement, we see how the two channel scenario model achieves its goal to remove the MSE spikes almost perfectly.
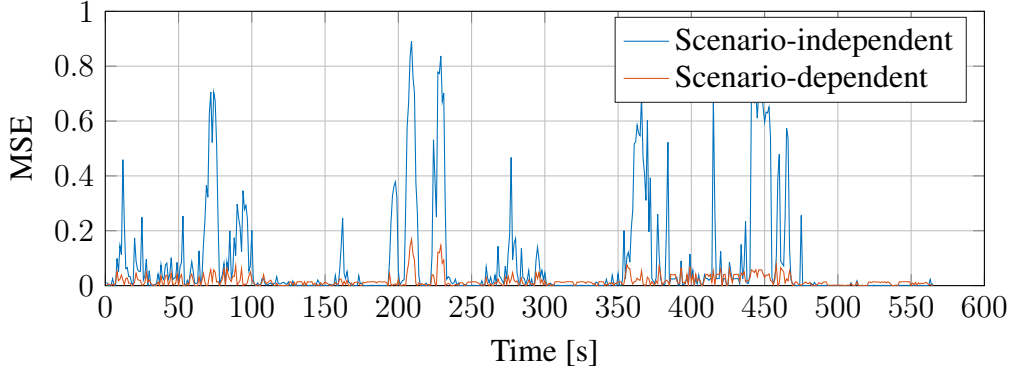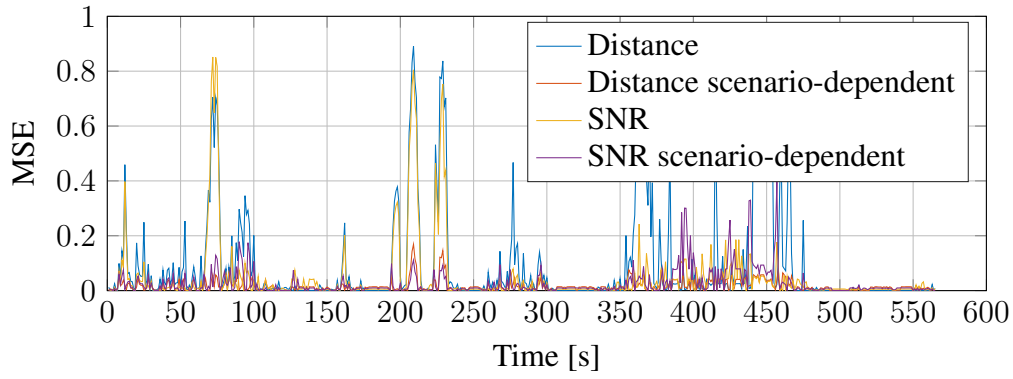
(a) Measurement 7
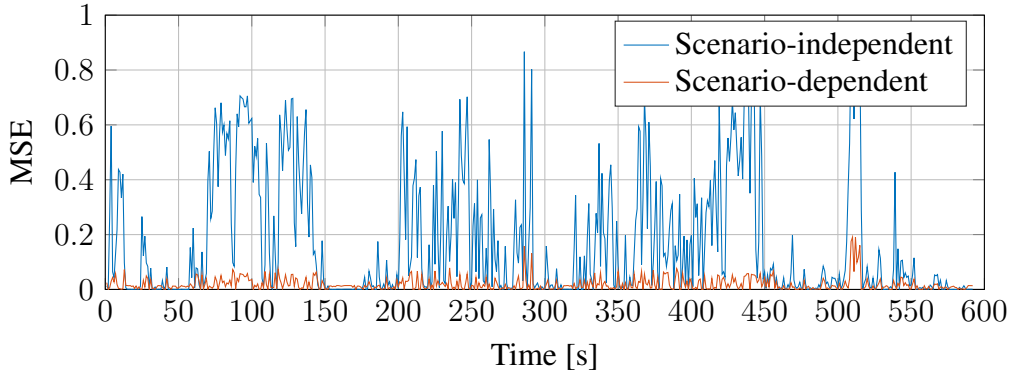
(b) Measurement 8

Figure 3.16: MSE over time for 2 measurements for pure distance based and scenario-dependent estimations.

3.5 Summary

This chapter has provided two estimation schemes, each of which is based on either SNR or distance. Figure 3.17 shows the time evolution of the MSE for all 4 analyzed schemes. We see that if we want to avoid taking the channel scenario into account, the SNR-based scheme is the way to go. There are too many disturbances in the distance-dependent estimation, and as we have seen, the distance dependent estimation resorts practically to estimating the PDR by a constant global mean.

Adding the scenario concept to the estimation scheme let us drastically improve the results. Once we added scenarios, the two schemes achieved almost the same minimum MSE.

Allthough we arrived at an estimation scheme that manages to estimate with a high accuracy, we introduced a new problem. The estimation has now a hidden, time dependent variable. For the estimation up until now, we took perfect knowledge of this variable for granted, but we have not proposed a model for this variable. In the SNR case, as we have seen, we might stick to the channel scenario 1 estimates, and discard the second scenario

(a) Measurement 7
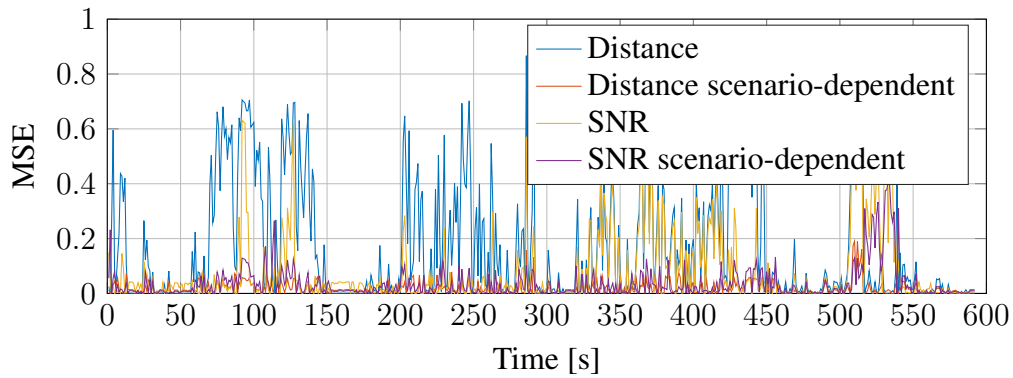
(b) Measurement 8

Figure 3.17: MSE over time for all presented performance estimation schemes.

as random errors. This approach, however would be a oversimplification, and would lead to overestimation of the channel performance. In the distance dependent case, this effect is even more pronounced. Therefore, the next chapter deals with an in-depth analysis of this time-variant model and tries to pinpoint the reasons that enforce the second state.

4

Model Analysis

In the previous chapter, the scenario-dependent performance estimation was presented and evaluated. The evaluation was executed both based on the SNR and the distance. For the evaluation, we assumed perfect knowledge of the channel scenario, which has up to now no meaning assigned.

The goal of this chapter is to reformulate the performance estimation in such a way that, given a distance or SNR trace, the model is able to predict the performance. In order to do this, we have to investigate these channel scenarios. As already concluded in the previous chapter, scenario 1 behaves very similar to the invariant estimation, and can therefore be interpreted as "normal" working condition. It is therefore important to identify when and why the second scenario is assumed. If we are able to reason why the second scenario is in force, we can include this in our predictions, which leads to a model that has no unknown parameters in them, and in this way achieves our ultimate goal.

To this end, this chapter will first investigate the ocurrences of the second scenario. After we analyzed the second scenario for the two estimation schemes, we use this information to formulate the final versions of the SNR-dependent and the distance-dependent estimation models.

4.1   Analysis of Channel Scenario 2

Same as in the previous chapter, we define that the channel has assumed the second channel scenario when the MSE of the estimated PDR exceeds 0.1. For the identification of possible causes of the second scenario, we used multiple methods. On one hand, we consulted the videos that were recorded alongside the measurements. These reveal special constructions, such as bridges, gantries, walls or forests, traffic density and allow differentiation of the LOS-NLOS scenarios.

On the other hand, the GPS data that was also recorded is used to locate these instances on satellite maps provided by Google maps. The reason to do this is to compare different measurements. The measurements provide multiple data points for most locations of the highway stretch. By comparing these different measurements, we hope to find locations where channel scenario 2 was in force consistently across measurements. This could then mean that some structure at that point influenced the packet transmission.

We analyze the SNR-dependent estimation first. This is done because the SNR-dependent estimaion assumed scenario 2 far less frequently, making it easier to identify possible causes. we then proceed to the distance-dependent estimation and compare it to the SNR-dependent one.

4.1.1   Channel Scenario 2 in SNR-Dependent Estimations

Fig. 4.1a and Fig. 4.1b show the aggregated instances where the SNR-dependent estimation assumed scenario 2 due to estimation errors going West and East respectively. The GPS locations for these map plots were taken from the car carrying the CVIS box. In these plots, different colors stand for different measurement runs. The two directions are analyzed independent from one another because the two highway lanes are seperated enough so that different influences may be seen. Note that not all measurements passed all locations.

The map has marked 5 clusters. These clusters are locations that produced estimation errors across multiple measurements. Cluster 1 and 2 are easily distiunguished since they are nicely seperated by error-free stretches, and can be observed in both directions. Cluster 3, 4 and 5 are not as distinct on both maps. They are placed along a stretch of road that produced more errors overall. At the north side of the road, there is a sound protection wall, which heavily influenced one measurement going westwards. This measurement is seen in purple. On the other hand, going eastwards, a new lane joined the traffic at cluster 3 which introduced traffic and frequent overtaking maneuvers. These effects distort the measurements to a certain degree and lead to clusters that are not as well defined. Moreover, cluster 5 is only apparent in measurements going eastwards. There are more places where errors occured, especially
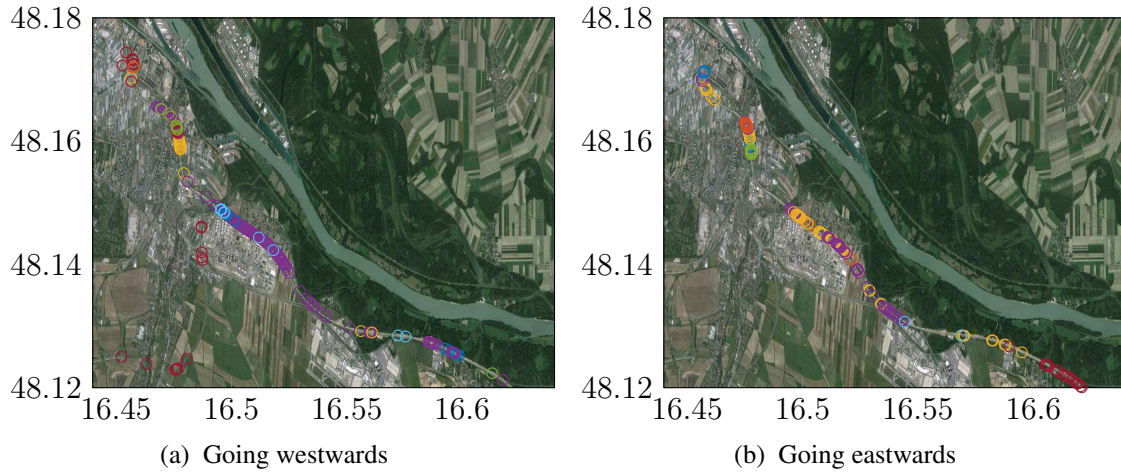
(a) Going westwards   (b) Going eastwards

Figure 4.1: Geographic indicator of high MSE in the SNR case.

further down to the southeast, but none of those places produced errors in more than two measurements.

As mentioned before, we assumed that at locations where multiple measurements deviated from the estimation there might be a structure that influenced the transmission quality. To illustrate this argument, Fig. 4.3 shows the geographic evolution and the PDR curve for one measurement going east. Added to the map plot are all gantries and highway overpasses in the southeast direction.

Figure 4.2: Video camera capture showing the target car and gantries.

In this plot we see that almost all errors happened around the marked clusters. Furthermore, all those errors happened within the vicinity of either a gantry or an overpass. These errors are indicated in the PDR plot as a sharp drop of the PDR, which is very high otherwise. At these instances, the SNR remained around $15\,$dB. The only exception to this is the region around 400 seconds in Fig. 4.3b. There, the cars had a NLOS connection, and the

SNR dropped accordingly to below $10\,\mathrm{dB}$. Errors in this region stem from the fact that the PDR would sometimes rise to almost 1 for brief periods.

(a) Measurement 7 going NW to SE
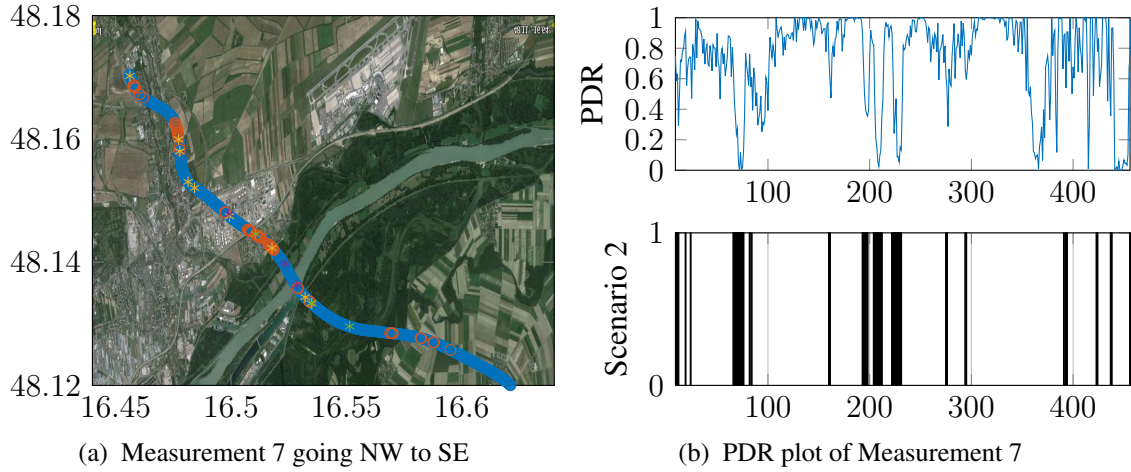(b) PDR plot of Measurement 7

Figure 4.3: Analysis of Measurement 7's estimation errors.

In the map of this figure we see that although the errors cluster around the gantries, there are many gantries that do not incur errors. However, there are structures that provoke errors consistently. Figure 4.4 shows all 6 measurements that passed cluster 2 in southeast direction. For this plot, we calculated the mean position of the two cars, and chose the first gantry in cluster 2 as center point, indicated by the red dashed line. This gantry is seen in Fig. 4.2 on the opposite lane on the left side. All measurements passing cluster 2 experienced a sharp PDR drop at the same time, apart from the third measurement, which experienced it a little later. These drops were all accompanied by a rising or constant SNR. A very similar behaviour is found at cluster 2 going in direction northwest.

The plot uses the moment where the mean position passed the gantry as 0 point, and shows PDR and SNR curves from 20 seconds before until 20 seconds after the gantry. As can be seen, all measurements experienced a dip in PDR close to the gantry. While none of them experienced a decrease in SNR, some even experienced a sharp increase for the duration of the PDR dip. For this gantry, due to the coupling of PDR drop and SNR increase, which can be seen in both directions, in all measurements, always on the same side of the gantry, we conclude that there was an active interferer mounted on the gantry, pointing northwest.
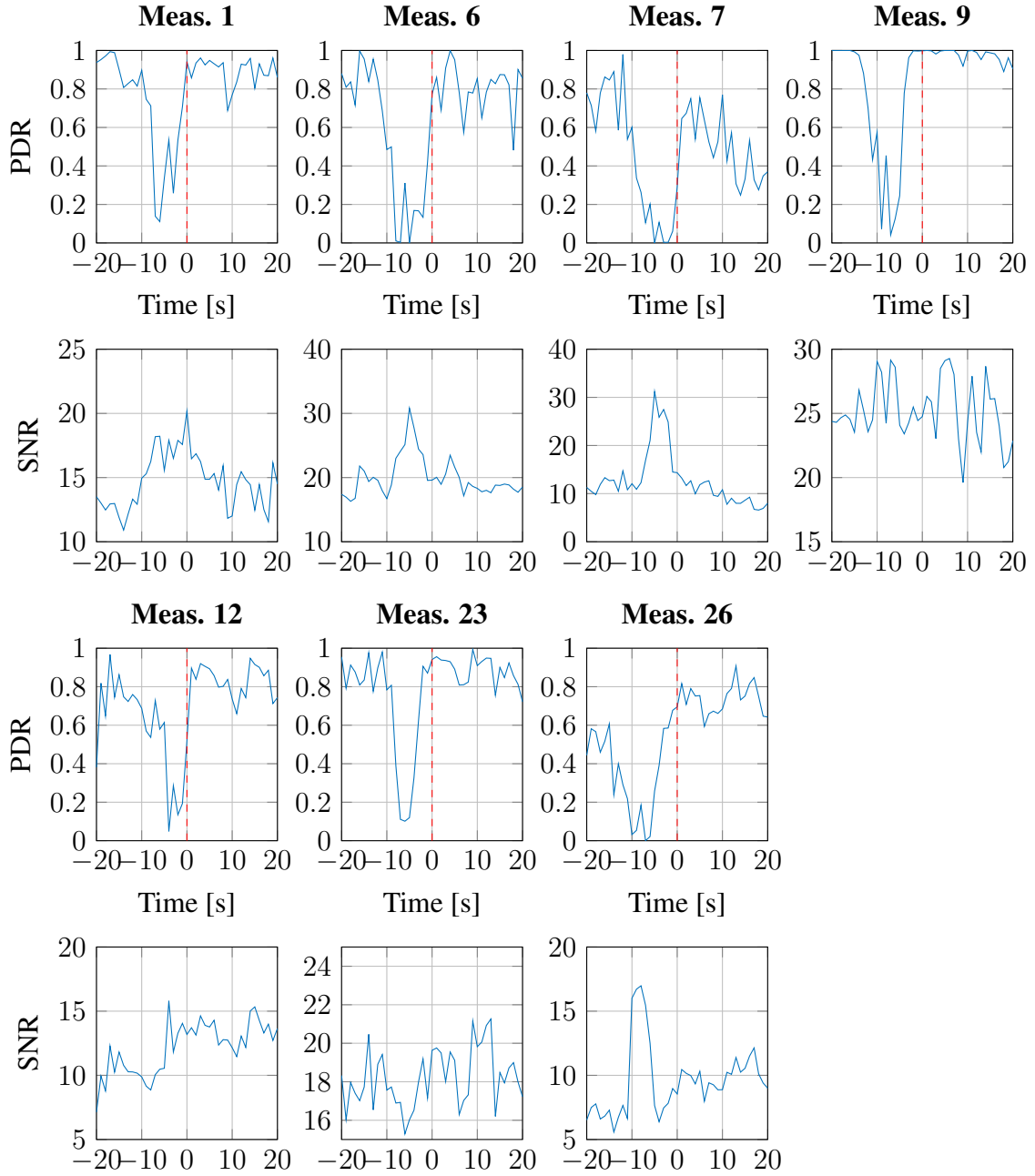
Figure 4.4: PDR and SNR curves of all measurements passing cluster 2 going eastwards.

Other gantries along the road also provoked PDR drops on multiple occasions, and a sound protection wall that was erected also had an influence on the measurements. A complete analysis of the clusters can be seen in Appendix C.

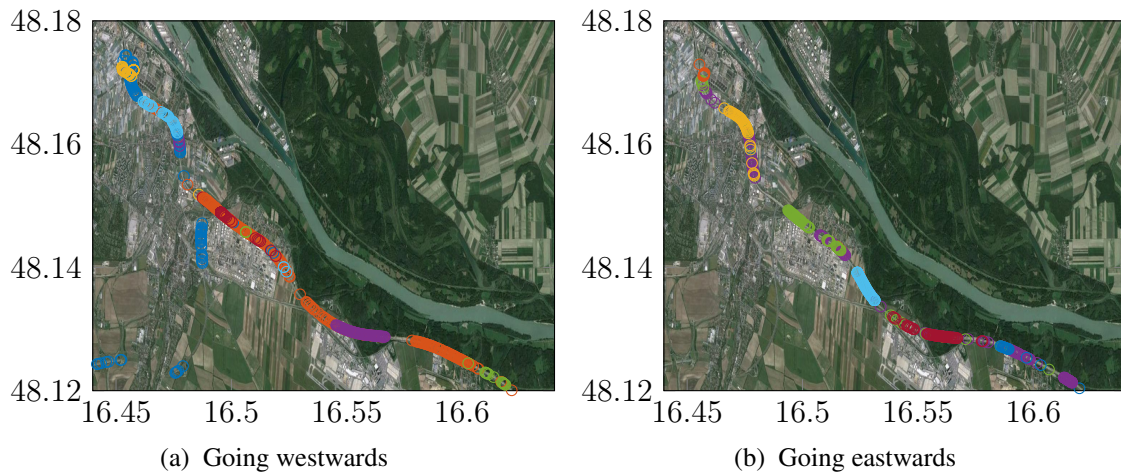## 4.1.2 Channel Scenario 2 in Distance-Dependent Estimation

(a) Going westwards   (b) Going eastwards

Figure 4.5: Geographic indicator of high MSE in the distance case.

Fig. 4.5a and Fig. 4.5b shows the analogous maps from the previous section for distance based estimation. These maps however do not tell much, as high MSE occured almost everywhere, at some point. Instead, we compared the single measurements. Figure 4.6 shows side-by-side the measured SNR, distance, and when the respective estimators assumed scenario 2 for Measurement 9. Areas where the SNR-dependent estimation assumed scenario 2 also feature scenario 2 in the distance dependent estimation. This is to be expected if our interpretation as external structural influences holds, as these of course do not influence the distance, and therefore the distance is not fit to capture these PDR drops. Compared to the SNR curves however, the distance dependent estimation features three regions where the distance dependent estimation was in scenario 2 while the SNR dependent wasn't. The first can be seen as a second region at roughly $100\,\text{s}$. The second is the stretched-out region shortly before $400\,\text{s}$, and the final region is after $400\,\text{s}$.

When compared to the video, we see that in the first of this regions, a small car was in between our two test vehicles. This can actually be seen in Fig. 4.2 as the small white car before the target. At the moment the picture was captured, the cars were still in the curve, so the LOS was not obstructed, but, shortly after it would be. This is seen in the PDR as a drop to about 0.5. In the second region, the two cars overtook a truck, while being at a distance of about $100\,\text{m}$. During this overtaking, the truck obstructed the LOS for a short time, which is

seen as a drop of the PDR to almost 0. Finally, in the third region a van placed itself between the two cars, leading to a true NLOS scenario for a short period, which can again be seen as a PDR drop to almost 0.

Analyzing the other measurements in the same way, we came to similar results. The distance-dependent measurements assume channel scenario 2 almost always if the SNR dependent counterparts did for a noticable time. In addition to those times, the distance dependent estimations assumed scenario 2 for an extended amount of time if the LOS path was at least partly obstructed.

(a) SNR based
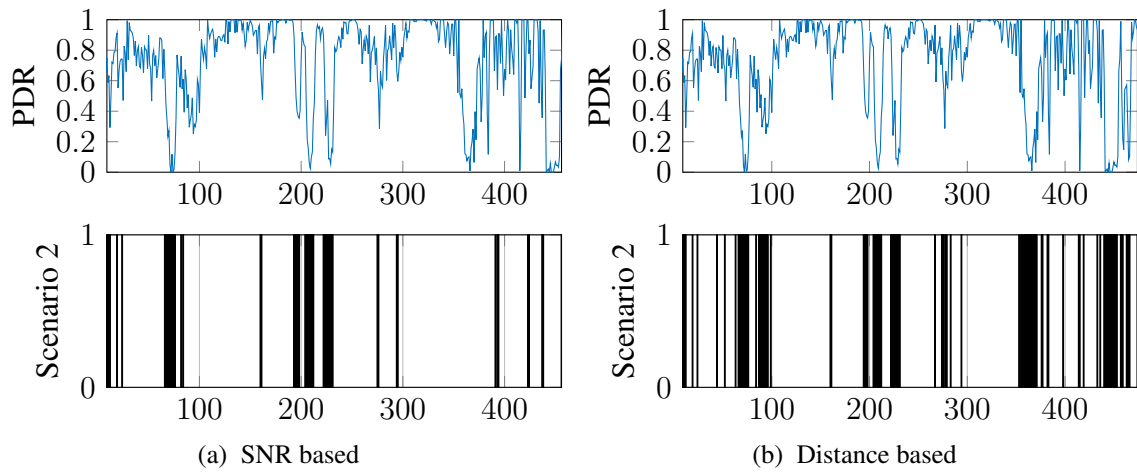
(b) Distance based

Figure 4.6: Comparison of SNR and distance based measurements, and when the estimation assumed channel scenario 2.

4.2   Time Analysis of Channel Scenario 2

Estimator	Total time [s]	S1 [s]	S1 [%]	S2 [s]	S2 [%]
SNR dependent	7965	7394	92.8	571	7.2
Distance dependent	5742	4617	80.4	1125	19.6

Table 4.1: Comparison of total time used for the estimation and time spent in channel scenario 1 and 2 (S1 & S2).

In this section, we analyze the time an estimation spends in channel scenario 2 once it switches both for the SNR dependent and the distance dependent case. Table 4.1 shows the time spent in scenario 1 and 2 for both estimation schemes. Total time means the total amount that was used for evaluation of the estimation schemes. As mentioned before, the distance dependent scheme used less time instances due to faulty GPS data. It is noteworthy that more

than 92% of the SNR dependent estimations were in scenario 1. This means that scenario 1 holds for a significant amount of the time, while only a small minority was in scenario 2. For the distance dependent case, scenario 1 is still in force $4/5$ of the time, but is not as dominant as in the SNR dependent case.

Fig. 4.7 shows the ECDF of the scenario 2 durations for both estimation strategies. Looking at the SNR case first, we see that across all measurements, there was no time when scenario 2 was in force for more than $20\,s$ at a time. Furthermore, 90% of all scenario 2 bursts were only $7\,s$ or shorter, and over 50% remained in scenario 2 for only $1\,s$.

The picture for the distance-dependent case is similar. The 90% mark only moved from 7 to $10\,s$, however, the maximum length reaches $60\,s$ in this case. This is not surprising, as the measurements happened on limited time, and long bursts could not appear as often as short bursts.

<div style="display:flex">

(a) SNR dependent estimation   (b) Distance dependent estimation

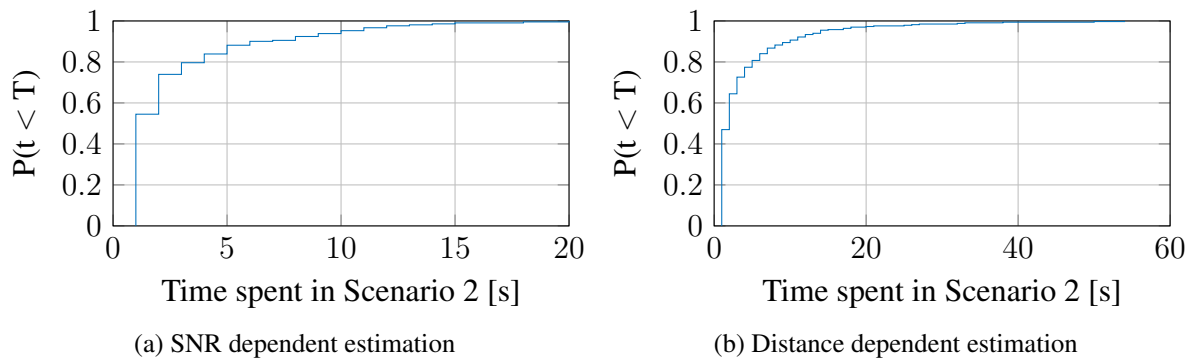</div>

Figure 4.7: ECDF of Scenario 2 burst lengths.

Fig. 4.8 shows the scenario 1 bursts, again for both estimation schemes. In this case it is more noticable, that the SNR case still has 10% of scenario 1 bursts being longer than $100\,s$, while that same mark lies at roughly $20\,s$ for the distance dependent estimation. This is intuitive, as the SNR dependent estimation has generally less reason to enter scenario 2, making for longer scenario 1 stretches.

4.3 The SNR-dependent Modified Gilbert Model

As could be seen, the vast majority of the instances where the SNR-dependent estimation assumed channel scenario 2 can be attributed to roadside structures. Moreover, scenario 2 only appeared less than 8% of the time, and only for a maximum of $20\,s$ at once. Based on these results, we define a regular highway scenario based on the estimates for scenario 1, and a complementary 2nd scenario that accounts for roadside structures that impede the transmission as seen in Table 4.3. Similar to the previous chapter, we chose 5 intervals for the

(a) SNR dependent estimation   (b) Distance dependent estimation
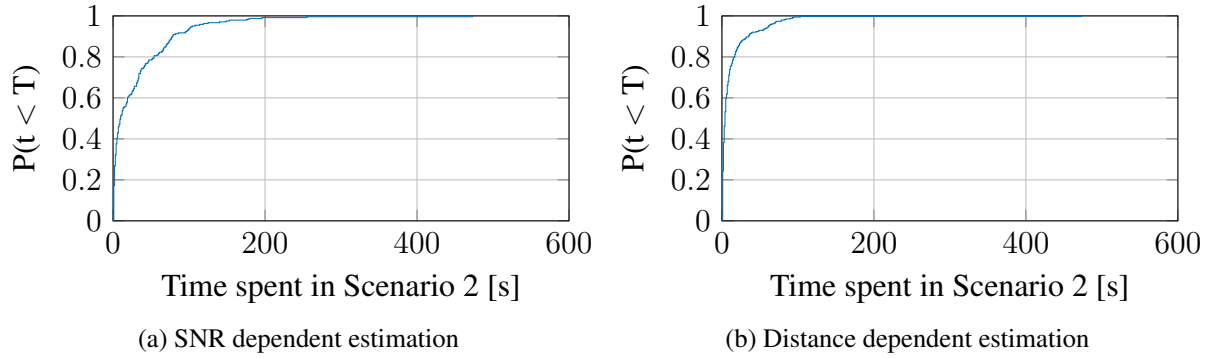
Figure 4.8: ECDF of scenario 1 burst lengths.

SNR range, as more intervals did not decrease the estimation error further (see Fig. 3.10). Furthermore, we use the fact that the gantry related PDR drops spanned roughly $10\,\text{s}$, which combined with the fact that the cars were going roughly $80\,\text{km/h}$, corresponds to a distance of about $200\,\text{m}$.

With these components, we propose the estimation rule, that the "regular" case should be used, unless a roadside structure is within $100\,\text{m}$ of the center between the cars. In this case, the model should assume scenario 2. Within the scenario, performance parameters are to be taken from Table 4.3. This is a pessimistic rule, since not all gantries lead to PDR drops, but it is better to underestimate rather than overestimate the performance.

Scenario 1					
SNR [dB]		P_{BG}	P_{GB}	P_e	C
	$x \quad < 6$	0.044	0.206	0.955	0.213
$6 \leq$	$x \quad < 8$	0.089	0.101	0.877	0.533
$8 \leq$	$x \quad < 12$	0.1	0.04	0.79	0.775
$12 \leq$	$x \quad < 22$	0.104	0.017	0.634	0.909
$22 \leq$	x	0.075	0.002	0.433	0.988
Scenario 2					
SNR [dB]		P_{BG}	P_{GB}	P_e	C
	$x \quad < 29$	0.053	0.113	0.939	0.36
$29 \leq$	x	0.049	0.405	0.994	0.115

Table 4.2: The SNR-dependent modified Gilbert model.

4.4   The Distance-Dependent Modified Gilbert Model

Finally, we present the distance-dependent, modified Gilbert model for the V2V case. It only needs 3 intervals per scenario, as more does not increase the estimation quality. In this model,

the regular scenario corresponds to usual highway conditions, while the scenario 2 is inforce either if a roadside structure impedes transmission, or the LOS is partially or fully obstructed.

We argue that the same rule as for the SNR dependent case with respect to roadside structures should be used, and in addition, scenario 2 should be inforce whenever the LOS is at least partially obstructed.

Scenario 1				
Distance [m]	P_{BG}	P_{GB}	P_e	C
$0 \leq x < 20$	0.086	0.008	0.479	0.961
$20 \leq x < 80$	0.11	0.002	0.703	0.881
$80 \leq x$	0.085	0.026	0.729	0.829
Scenario 2				
Distance [m]	P_{BG}	P_{GB}	P_e	C
$x < 20$	0.082	0.159	0.859	0.432
$20 \leq x < 80$	0.074	0.119	0.915	0.437
$80 \leq x$	0.0487	0.1653	0.954	0.264

Table 4.3: The SNR-dependent modified Gilbert model.

5

Conclusions and Outlook

In this thesis, we developed and analyzed several performance models for the V2V link, based on the vehicle-to-infrastructure (V2I) models in [12]. We introduced one model based on the SNR and another model based on the distance between the cars, and evaluation showed that both models are capable to replicate well the packet-error behaviour of the extensive real-world measurements that were used. The presented approach can be easily applied to different measurements or quantities, allowing for extension and cross-validation of the model. It furthermore requires a small amount of model parameters, allowing easy implementation in simulations. In this way, we were able to provide a measurement abstraction model that facilitates the exchange and comparison of measurements by way of a small set of parameters.

A mathematical analysis of the model showed that the capacity of the described Gilbert model equals the achievable PDR. This is a remarkable result, as it means that the main optimization goal is the PDR itself, since no improvement beyond the PDR is possible without applying cross-layer design.

To overcome the challenges posed by the V2V situation, we introduced additional steps to the original model. We used constant time intervals of $1\,\mathrm{s}$. This choice was shown in the thesis to be large enough for the modified Gilbert model to reproduce the PDR well, yet small enough to see short-time influences like static interferers were still well-resolved. We also quantized the physical quantity using the information bottleneck method before applying the modified Gilbert model. This ensured that the interval borders of the quantity lie optimally both with respect to the distribution of the quantity and the distribution of the PDR. Moreover, it allowed us to make the number of model parameters a design choice.

We have also introduced channel scenarios as means to model channel influences not captured by, depending on the model, SNR or distance. Regarding the SNR-based model,

this approach was able to reveal roadside structures such as gantries as possible sources of transmission disruptions. The model results also allowed to identify a roadside interferer, who was able to easily jam the transmission in its vicinity. The same analysis for the distance-based model showed that within our measurement range, the transmission was very weakly dependent on the transmission distance. This is illustrated by the fact that the MSE saturated at a very low interval count for the model.

The results show that within the measured range of up to $300\,\mathrm{m}$, distance is not the main factor impacting transmission quality between two cars. The deciding factors are rather the traffic and the topology of the road. Other cars obstructing the LOS, even partly, strongly impact the PDR. Roadside structures such as gantries, entries and exits, overpasses and sound protection walls were shown to also influence the transmission quality.

The final results of this analysis are very important, as they allow to more closely identify problem sources we have to expect when deploying V2V communications. While transmit distance was no problem, traffic proved to be a of considerable importance. If many cars in traffic are equipped with V2V units, a large amount of interference is to be expected. Additionally, we have to expect RSUs positioned alongside the road, which will add aditional interference for the V2V link. On the other hand, if only a few cars are equipped, all others will be obstacles obstructing the LOS, again leading to a poor PDR. This, combined with the influence of roadside structures, show that there are still many topics to be tackled. Among those are some that we were not able to answer within the scope of this thesis.

Since the presented models are based in measurements, one goal has to be analyzing different measurements and comparing the resulting parameters. This will then reveal how strongly our parameters were influenced by circumstances specific to our measurements.

Further analysis of interferer influence was shown in this thesis to be relevant. There was only one interferer placed on our measurement route, but if ITS become commonplace, the behaviour we saw is not acceptable and has to be mitigated.

Our focus in this thesis lay on capacity analysis, but another important factor of the Gilbert model is the burst behaviour. This has not been touched in this thesis, but is essential for worst-case analysis.

Finally, Appendix B presented the capacity formula for transmitted bits. In order to capitalize on this result, cross-layer design can be applied to use the mutual information of the bits.

List of Acronyms

ABS Anti-lock Breaking System

ESC Electronic Stability Control

IoT Internet of Things

ECDF Empirical Cumulative Distribution Function

CCDF Complementary Cumulative Distribution Function

HIHO Hard-Input Hard-Output

BEC binary erasure channel

RSSI Received Signal Strength Indicator

GPS Global Positioning System

IERC European Research Cluster on the Internet of Things

RSU Roadside Unit

WHO World Health Organization

PDR Packet Delivery Ratio

CRC Cyclic Redundancy Check

ITS Intelligent Transport Systems

LOS Line-of-Sight

MAC Medium Access Control

MSE Mean Squared Error

NLOS Non-Line-of-Sight

PHY Physical Layer

SNR Signal-to-Noise-Ratio

V2I Vehicle-to-Infrastructure

V2V Vehicle-to-Vehicle

CVIS Cooperative Vehicle-Infrastructure Systems

Appendices

A
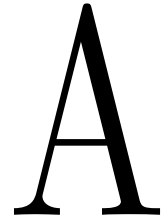
List of Used Measurements

This appendix lists all measurements that were used in this thesis. Figure A.1 lists key features of all used measurements. All measurements were taken in 2011, and the date format used in the table is DD-MM. The traffic condition is differentiated between light and dense, and the channel is either LOS or NLOS, depending on which was predominant. The direction is either southeast (SE) or northwest (NW), indicating the direction in which the cars were driving. The notes section is used to point out measurements where the GPS failed alltogether or provided offset data. Figures A.1-A.33 illustrate the three core parameters used in this thesis, PDR, SNR and distance over time for all measurements seperately. All parameters were averaged in one second intervals.

Measurement	Date	Duration [s]	Traffic	Channel	Direction	Notes
Measurement 1	30-08	137	Dense	LOS	SE	GPS shifted by roughly 300 m.
Measurement 2	30-08	124	Dense	LOS	SE	GPS shifted by roughly 300 m.
Measurement 3	30-08	195	Dense	LOS	SE	GPS shifted by roughly 300 m.
Measurement 4	30-08	133	Dense	LOS	SE	GPS shifted by roughly 300 m.
Measurement 5	30-08	67	Dense	LOS	NW	GPS shifted by roughly 300 m.
Measurement 6	02-09	648	Light	LOS	NW	GPS recording started at 160 s.
Measurement 7	02-09	564	Light	LOS	SE	
Measurement 8	02-09	591	Light	LOS	NW	
Measurement 9	02-09	611	Light	LOS	SE	
Measurement 10	30-08	69	Light	LOS	NW	GPS shifted.
Measurement 11	30-08	66	Dense	LOS	NW	GPS shifted.
Measurement 12	30-08	137	Light	LOS	SE	GPS shifted.
Measurement 13	30-08	72	Light	LOS	SE	GPS shifted.
Measurement 14	30-08	113	Dense	LOS	SE	GPS shifted.
Measurement 15	30-08	95	Dense	LOS	NW	GPS shifted.
Measurement 16	30-08	149	Light	LOS	SE	GPS shifted.
Measurement 17	30-08	178	Dense	LOS	SE	GPS shifted.
Measurement 18	30-08	72	Light	LOS	NW	GPS shifted.
Measurement 19	30-08	137	Light	LOS	NW	GPS shifted.
Measurement 20	30-08	76	Light	LOS	NW	GPS shifted.
Measurement 21	30-08	78	Light	LOS	NW	GPS shifted.
Measurement 22	30-08	39	Light	LOS	SE	GPS shifted.
Measurement 23	01-09	2000	Dense	LOS	NW	GPS had short errors.
Measurement 24	01-09	69	Dense	NLOS	NW	
Measurement 25	01-09	87	Dense	NLOS	NW	
Measurement 26	01-09	77	Light	NLOS	SE	
Measurement 27	01-09	118	Dense	NLOS	SE	
Measurement 28	01-09	67	Dense	NLOS	NW	
Measurement 29	01-09	78	Light	NLOS	NW	GPS recording stops at 20 s
Measurement 30	01-09	68	Light	NLOS	SE	
Measurement 31	01-09	68	Dense	NLOS	NW	
Measurement 32	01-09	72	Light	NLOS	SE	GPS recording starts at 72 s
Measurement 33	01-09	885	Dense	NLOS	NW	GPS broken

Table A.1: Measurement key parameters.
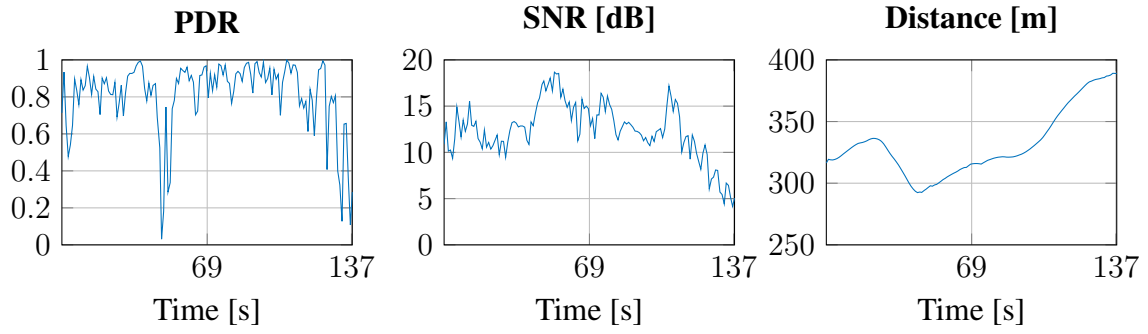
Figure A.1: Measurement 1.
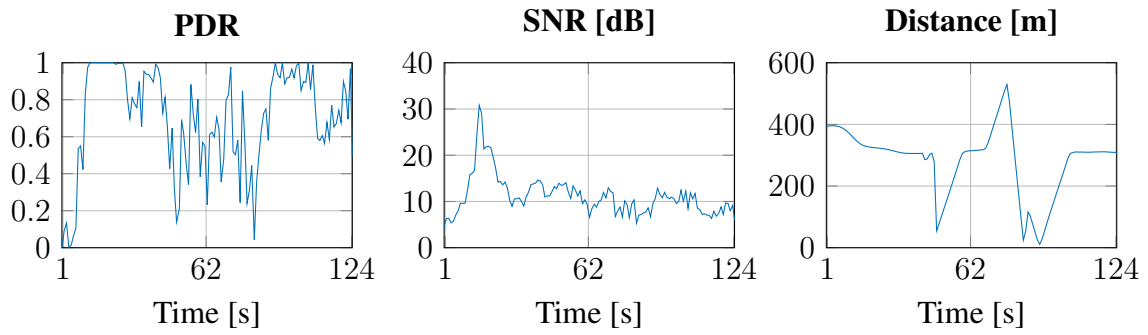
Figure A.2: Measurement 2.
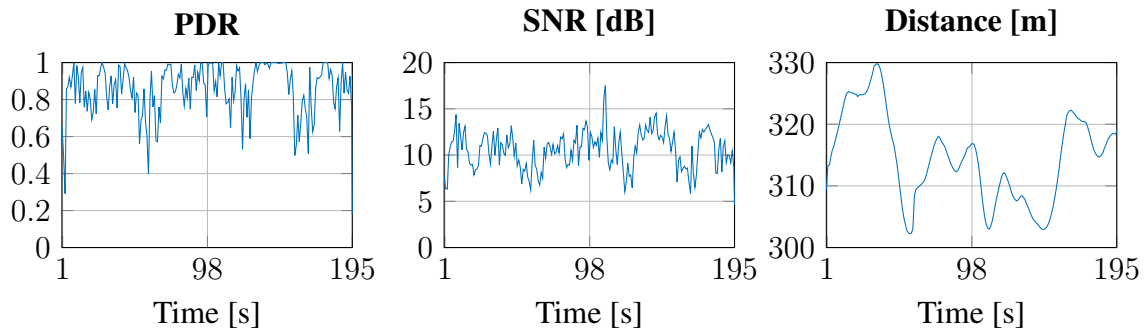
Figure A.3: Measurement 3.
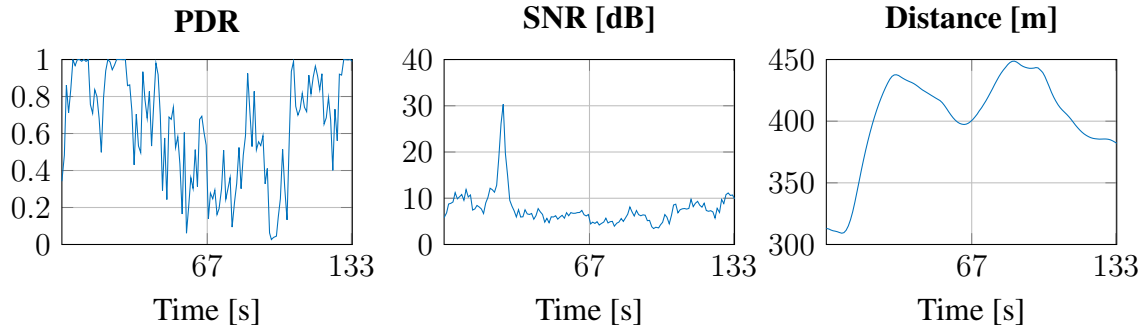
Figure A.4: Measurement 4.
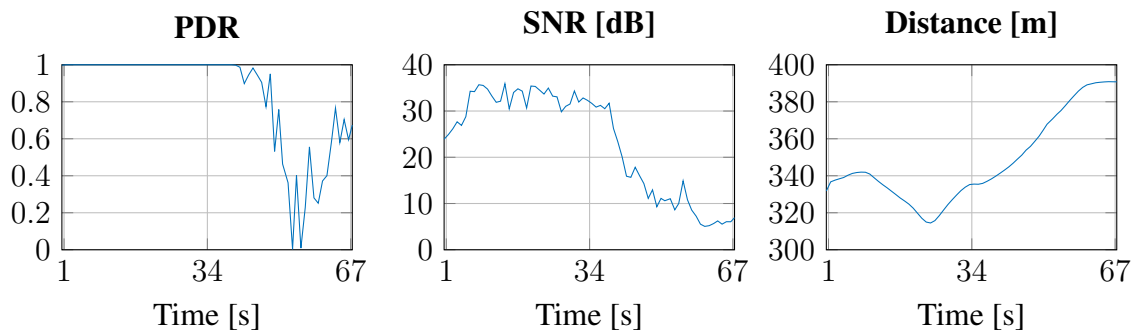
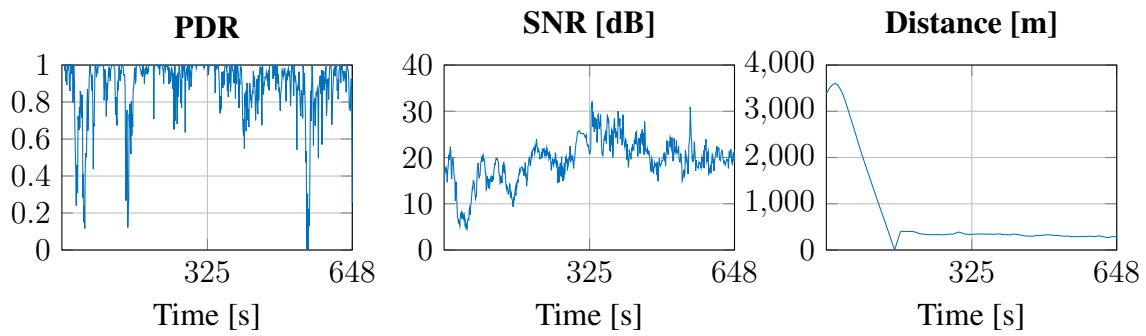Figure A.5: Measurement 5.

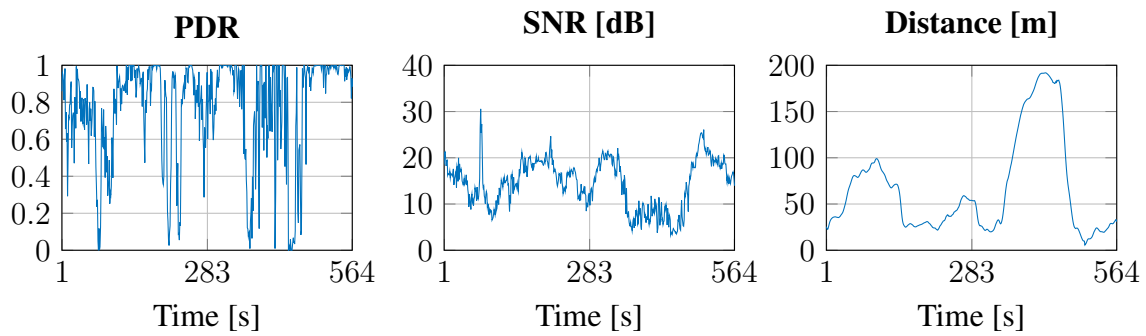Figure A.6: Measurement 6.

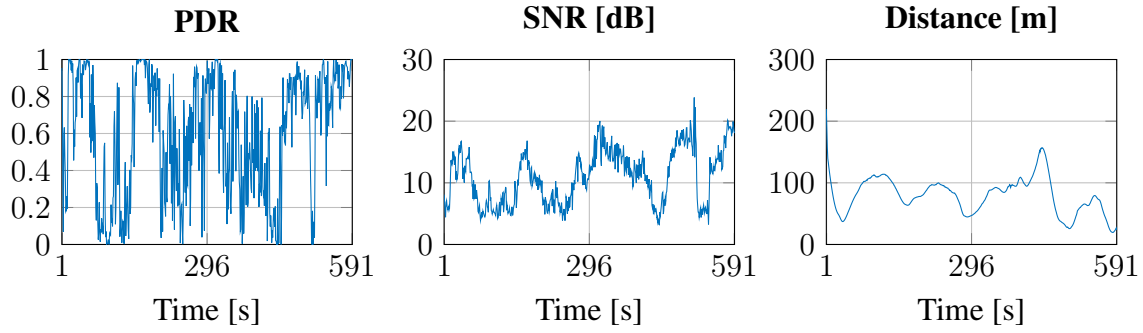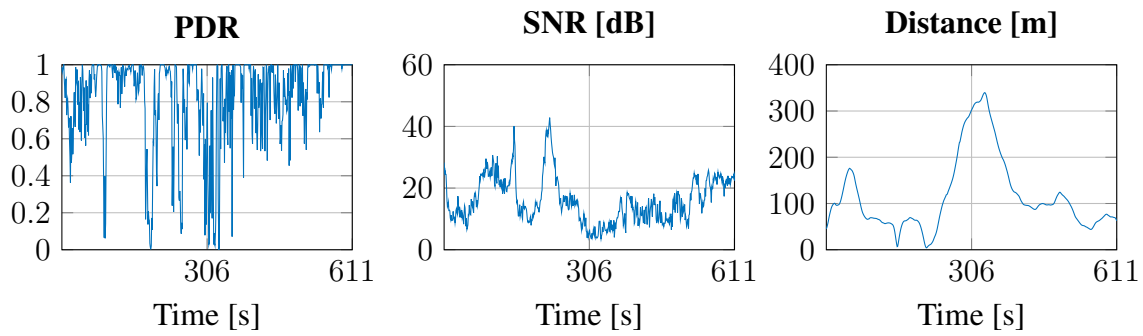Figure A.7: Measurement 7.

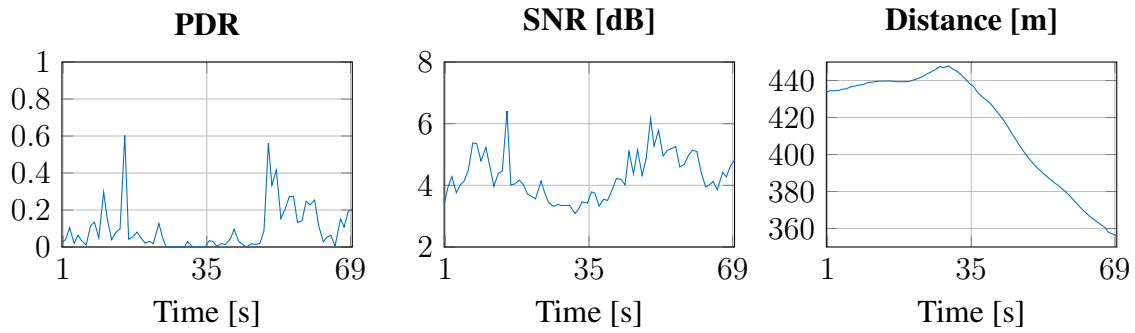Figure A.8: Measurement 8.

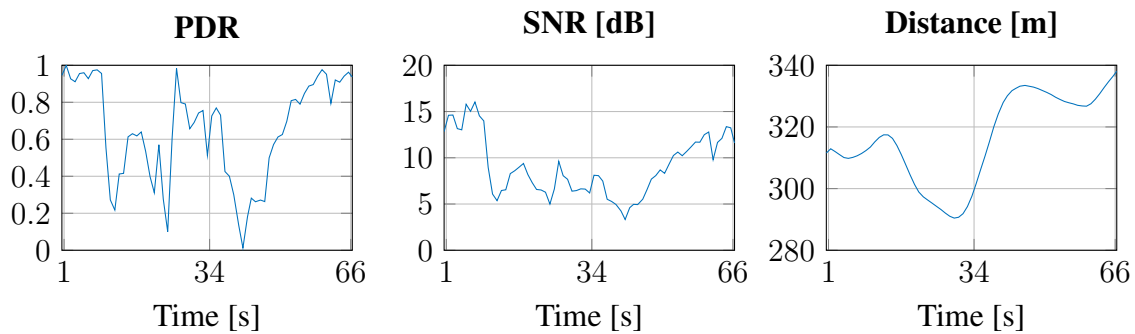Figure A.9: Measurement 9.

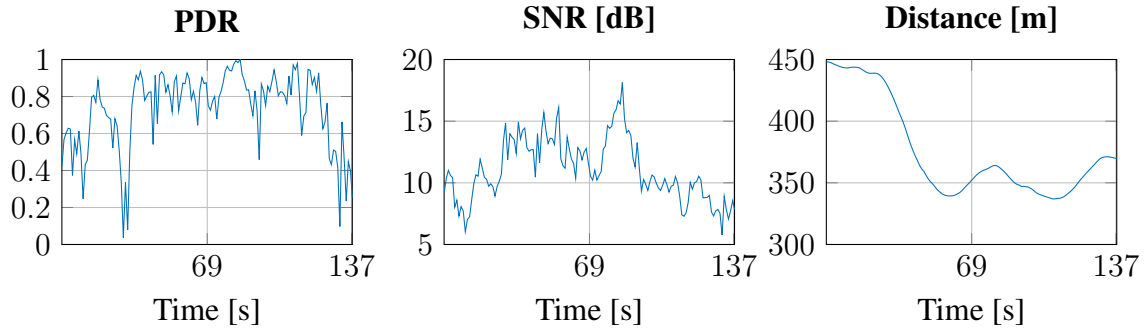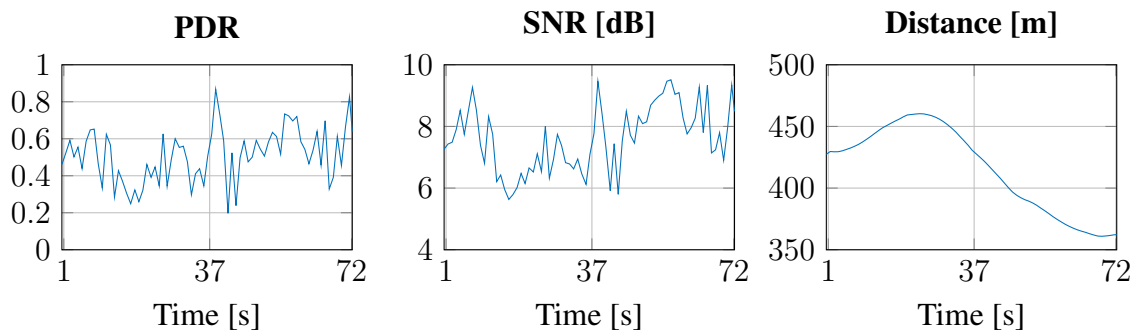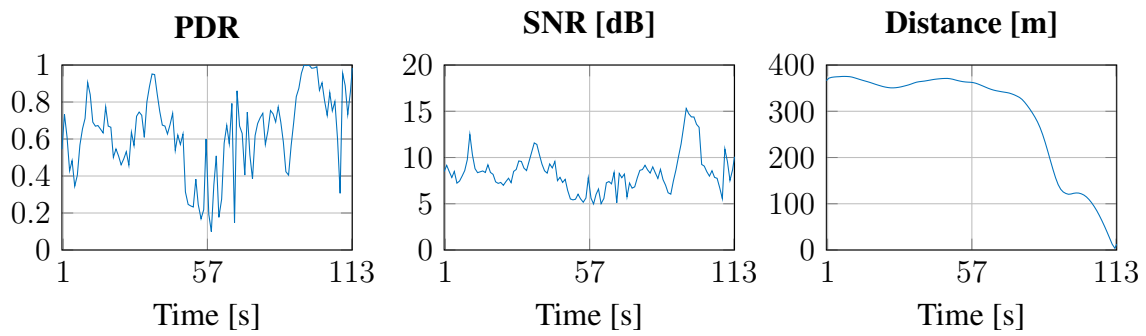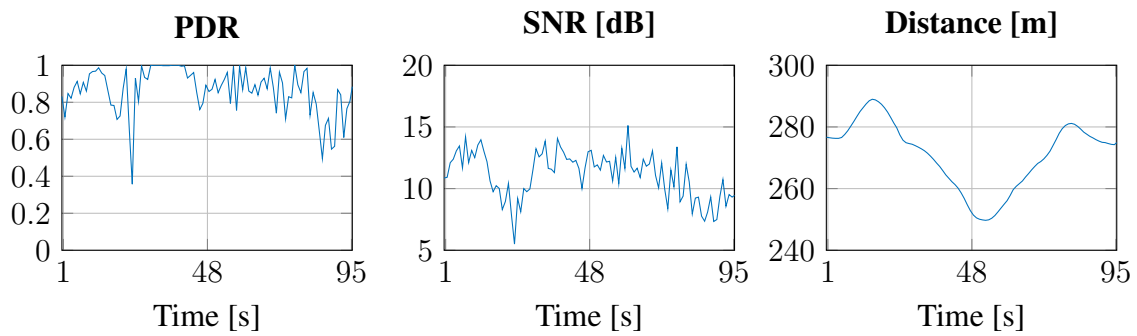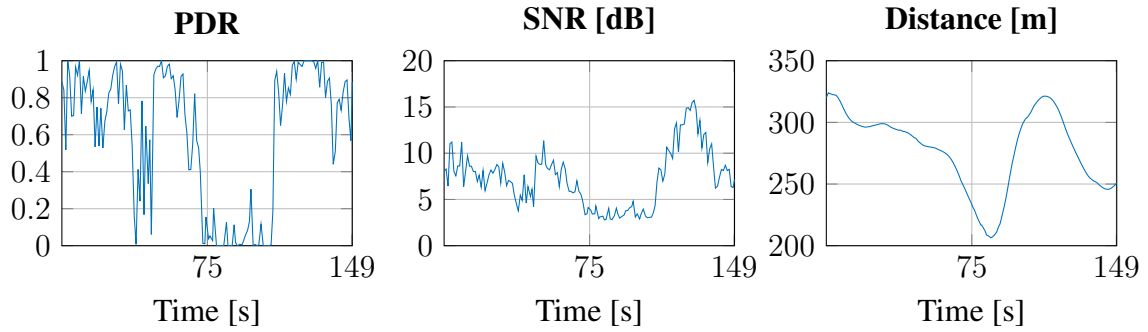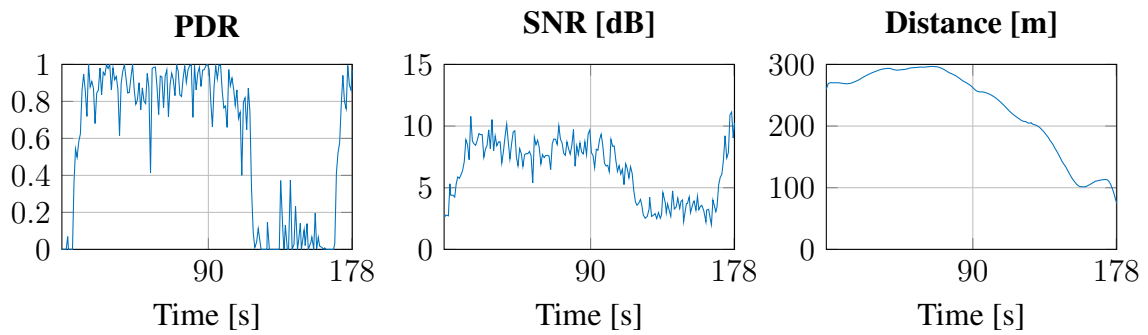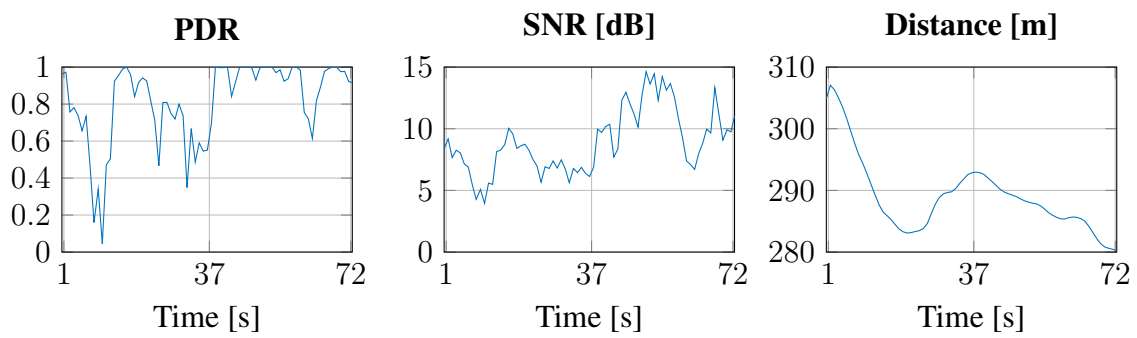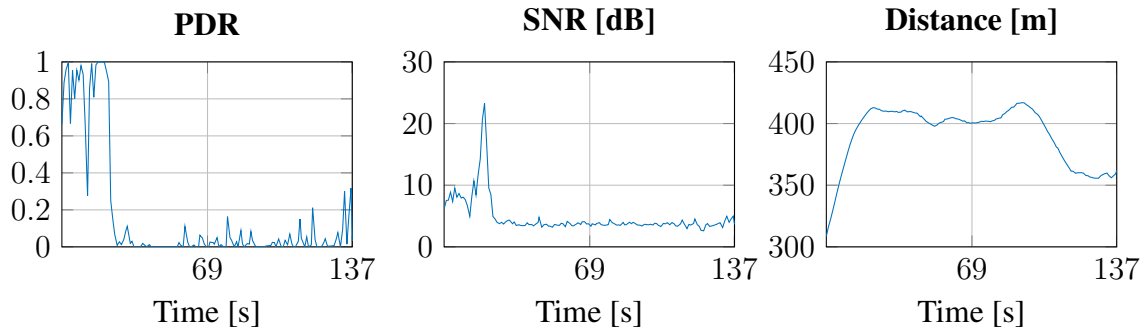Figure A.10: Measurement 10.

Figure A.11: Measurement 11.

PDR **SNR [dB]** **Distance [m]**

Figure A.12: Measurement 12.

PDR **SNR [dB]** **Distance [m]**

Figure A.13: Measurement 13.

PDR **SNR [dB]** **Distance [m]**

Figure A.14: Measurement 14.

PDR **SNR [dB]** **Distance [m]**

Figure A.15: Measurement 15.

Figure A.16: Measurement 16.

Figure A.17: Measurement 17.

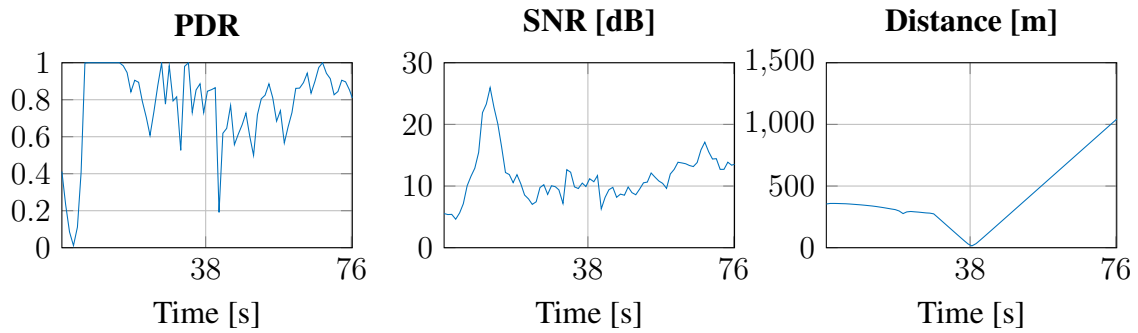Figure A.18: Measurement 18.

Figure A.19: Measurement 19.

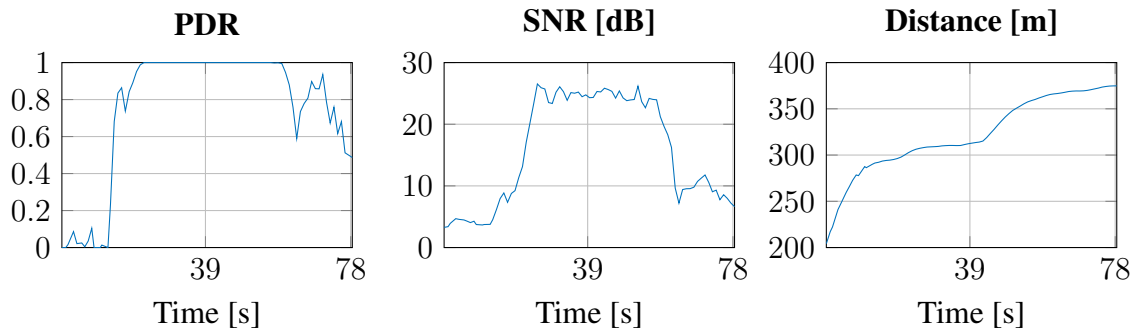Figure A.20: Measurement 20.

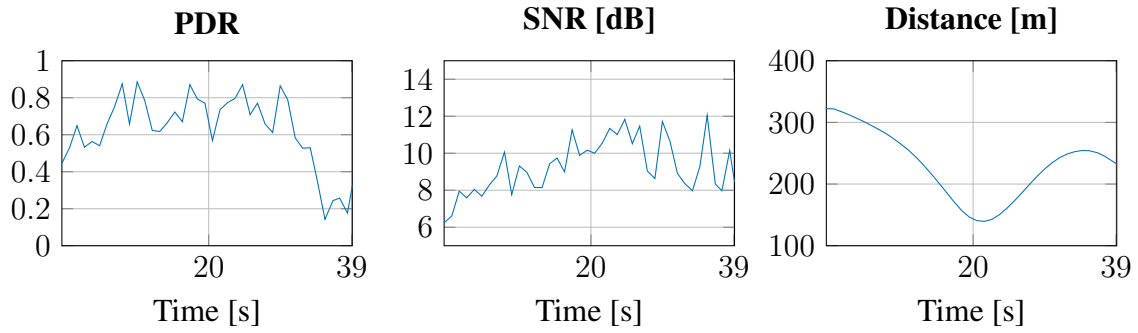Figure A.21: Measurement 21.
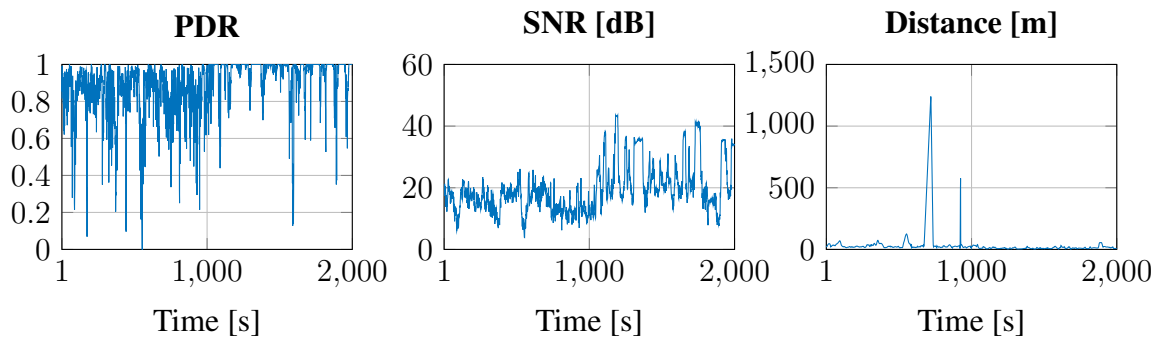
Figure A.22: Measurement 22.
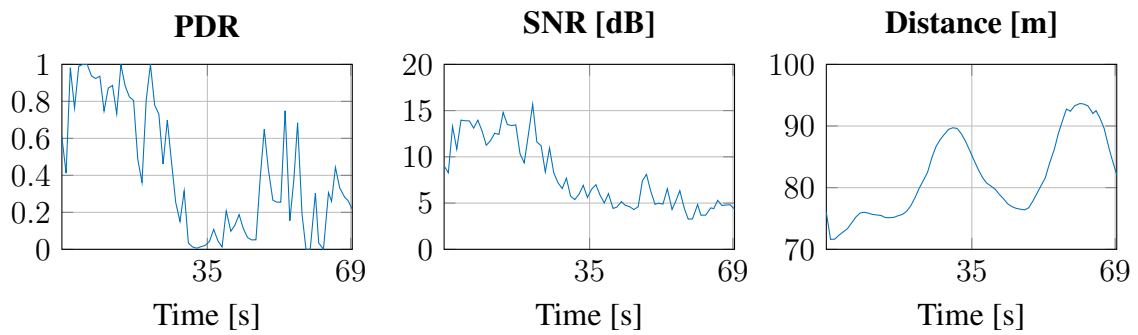

Figure A.23: Measurement 23.
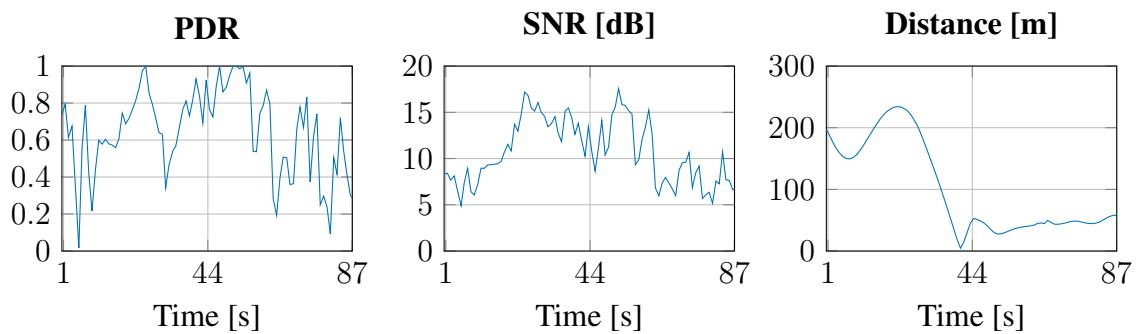

Figure A.24: Measurement 24.
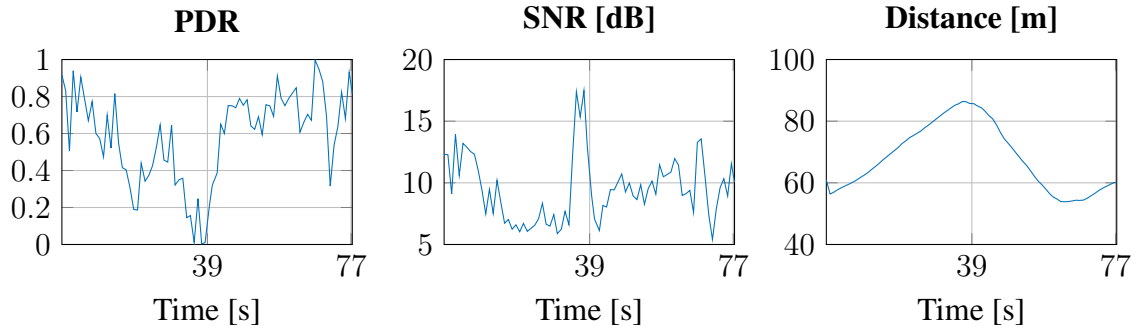

Figure A.25: Measurement 25.

Figure A.26: Measurement 26.
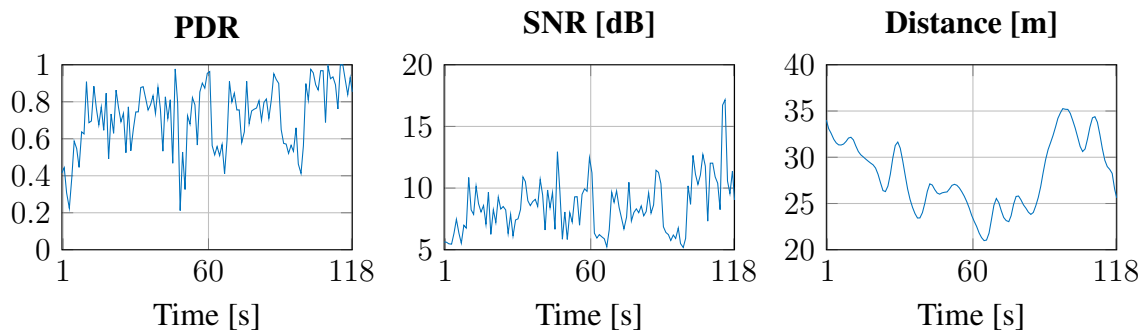
Figure A.27: Measurement 27.
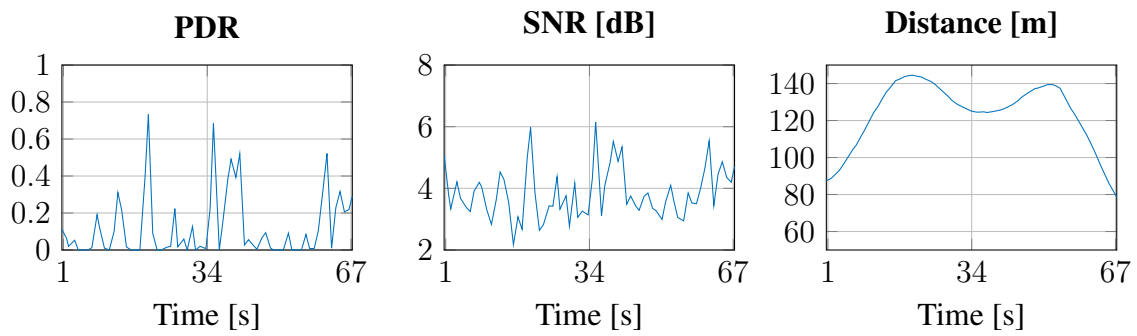
Figure A.28: Measurement 28.
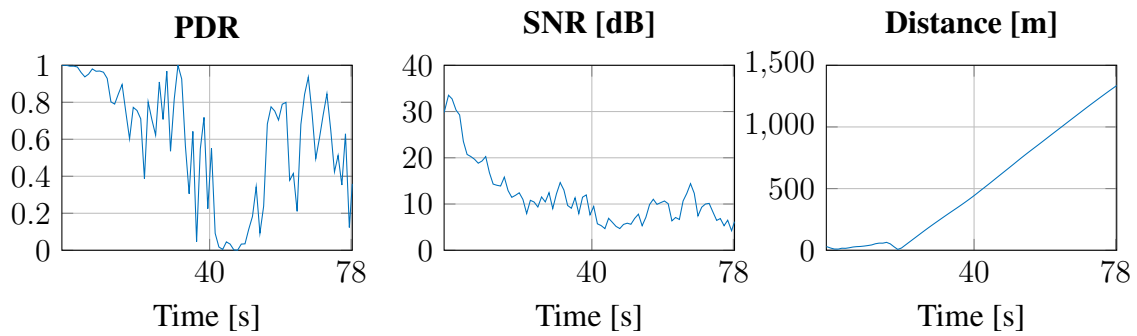
Figure A.29: Measurement 29.

PDR

SNR [dB]

Distance [m]
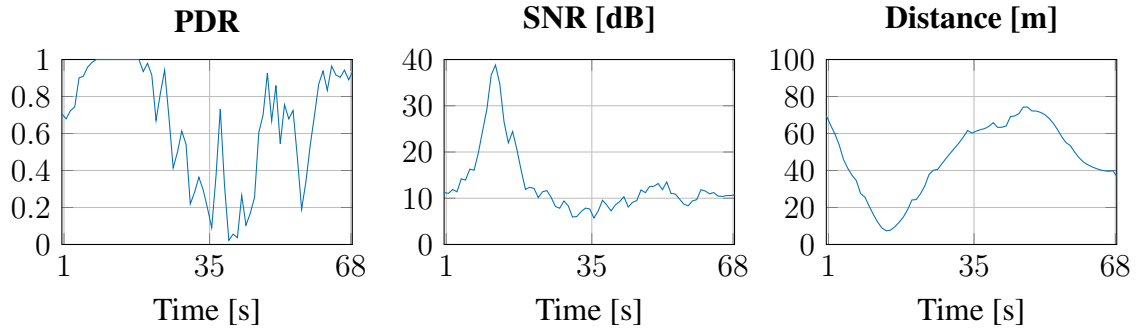
Figure A.30: Measurement 30.

PDR

SNR [dB]

Distance [m]
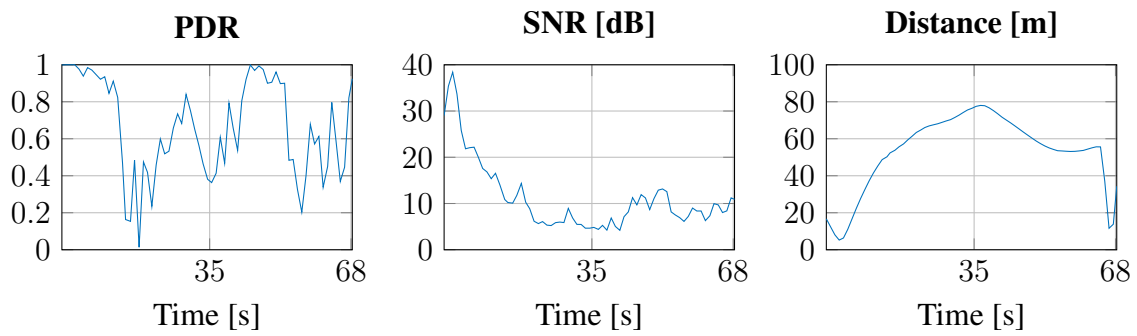
Figure A.31: Measurement 31.

PDR

SNR [dB]

Distance [m]

Figure A.32: Measurement 32.

PDR

SNR [dB]

Distance [m]

Figure A.33: Measurement 33.

B

Capacity of the Time-Variant Bit Model

In this section, the capacity of the modified Gilbert model for bits is analyzed. In a first step, Gilbert's capacity derivation for the time-independent bit model is retraced. Afterwards, it is expanded to concepts of time dependence.

B.1   Capacity of the Gilbert Model

Using the the equivalent additive noise HIHO channel model from Sec. 2.1.3, the mutual information of source $\mathbf{X}$ and sink $\mathbf{Y}$ can be written as in [33]

$$I(\mathbf{X}^N; \mathbf{Y}^N) = H(\mathbf{X}^N) - H(\mathbf{X}^N|\mathbf{Y}^N) \tag{B.1a}$$

$$= H(\mathbf{X}^N) - H(\mathbf{Y}^N \oplus \mathbf{Z}^N|\mathbf{Y}^N) \tag{B.1b}$$
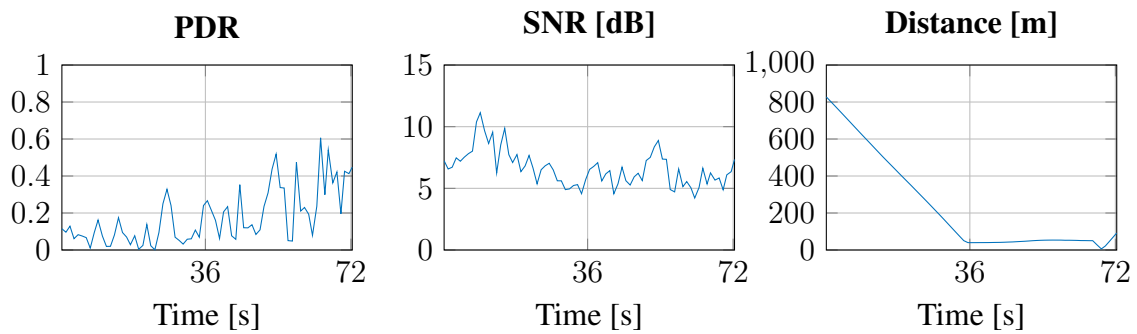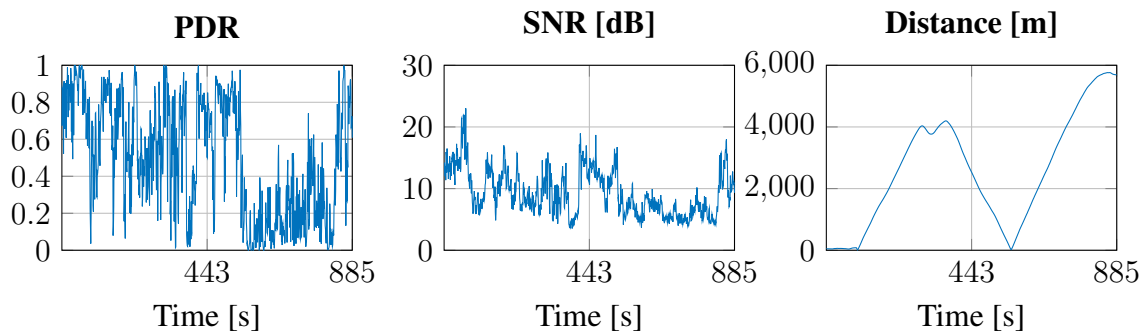
$$= H(\mathbf{X}^N) - H(\mathbf{Z}^N). \tag{B.1c}$$

In (B.1b) the fact was used, that subtraction and addition are equivalent in $GF(2)$. Using the same capacity expression as in Equation (2.7), and the fact that the optimal source has one bit per bit entropy, the capacity boils down to

$$C = 1 - \lim_{N \to \infty} \frac{1}{N} H(\mathbf{Z}^N). \tag{B.2}$$

After combining Eq. (B.2) and (2.9) and acknowledging that in the infinite series, every entry has a inifite number of predecessors and the conditional entropies converge towards each other, the capacity expression becomes [13]

$$C = 1 - \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{n} H(Z_i|Z_{i-1}, \ldots, Z_1) \tag{B.3a}$$

$$= 1 - H(Z_i|Z_{i-1}, Z_{i-2}, \ldots) \tag{B.3b}$$

$$= 1 - \sum_{z_j \in \{0,1\}} \Pr(\ldots, z_{i-2}, z_{i-1}) h(\ldots, z_{i-2}, z_{i-1}), \tag{B.3c}$$

$$h(\ldots, z_{i-2}, z_{i-1}) = \sum_{l=0}^{1} \Pr(z_i = l|z_{i-1}, z_{i-2}, \ldots) \log_2(\Pr(z_i = l|z_{i-1}, z_{i-2}, \ldots)). \tag{B.3d}$$

 Note that the state of the Gilbert model is known to be in bad state at the kth step if $z_k = 1$. In that case, the sequence becomes conditionally independent of previous states. Therefore, the condition in Eq. (B.3d) can be reduced to the most recent 1 followed by all 0s, and becomes

$$h(\mathbf{10}^i) = \sum_{l=0}^{1} \Pr(z_i = l|\mathbf{10}^i) \log_2(\Pr(z_i = l|\mathbf{10}^i)). \tag{B.4}$$

Finally, the capacity formula can be reformulated as done by Gilbert into

$$C = 1 - \sum_{N=0}^{\infty} \Pr(\mathbf{10}^N) h(\mathbf{10}^N). \tag{B.5}$$

Gilbert then introduces $u(K) = \Pr(\mathbf{0}^K|1)$ and $v(K) = P(\mathbf{0}^K 1|1)$ to reach the final form of

$$C = 1 + \Pr(1) \sum_{N=0}^{\infty} v(N) \log_2 v(N). \tag{B.6}$$

B.1.1   Capacity of the Modified Gilbert Model

This section derives a capacity expression for the single bitstreams, assuming decoding attempts are made, even when the CRC fails. As in the core thesis, we discritize constant time intervals. One thing that is decidedly different however is the meaning of capacity. Since the channel is time-variant, there can be multiple meanings.

B.1.2   Ergodic Capacity

Now to go back to Eq. (B.3). As additional prerequisite, we assume that the triple (P_{GB}, P_{BG}, P_B) is given as a sequence. This triple will be denoted the current channel characterization $\mathcal{H}_n$ in place for n iterations. According to the assumptions made in the model, this triple will be constant for a certain amount of iterations, then change. It is furthermore enhanced by the initial state the process was in when changing parameters, resulting in the quadruple channel characterization $\mathcal{H}_n^0 = (P_{GB}, P_{BG}, P_B, \pi_0)$.

In Eq. (B.3a), it is now assumed that a parameter change means that the past becomes uninformative, truncating the memory of the entropy function. Assuming all constant durations $K \in [1, N]$ are allowed, the capacity becomes:

$$
\begin{aligned}
C &= \lim_{N \to \infty} \frac{1}{N} \left[N - \sum_{K=1}^{N} \int_{\mathrm{h}^0 \in \mathcal{H}_K^0} \underbrace{N \Pr(K|\mathrm{h}^0) f(\mathrm{h}^0)}_{n'} H_{\mathrm{h}^0}(Z_i|Z_{i-1}, \ldots, Z_{i-K}) \, \mathrm{dh}^0 \right] \\
&= \sum_{K=1}^{\infty} \int_{\mathrm{h}^0 \in \mathcal{H}_K^0} \Pr(K|\mathrm{h}^0) f(\mathrm{h}^0) \underbrace{(1 - H_{\mathrm{h}}(Z_i|Z_{i-1}, \ldots, Z_{i-K}, \pi_0))}_{C_e(\mathrm{h}, \pi_0, K)} \, \mathrm{dh}^0 \\
&= \sum_{K=1}^{\infty} \int_{\mathrm{h} \in \mathcal{H}_K} \int_{\pi_0 \in \Pi} \Pr(K|\mathrm{h}^0) f(\mathrm{h}|\pi_0) f(\pi_0) \underbrace{(1 - H_{\mathrm{h}}(Z_i|Z_{i-1}, \ldots, Z_{i-K}, \pi_0))}_{C_e(\mathrm{h}, \pi_0, K)} \, \mathrm{dhd}\pi_0.
\end{aligned}
$$

$$\tag{B.7}$$

In this equation, n' is roughly equal to the number of occurences of the parameter-quadrupel in a N-iteration sequence. $C_e(\mathrm{h}, \pi_0, K)$ denotes the expected capacity of a given channel h that has been constant for K iterations, and a given initial state π_0. Now to develop

this further

$$
\begin{aligned}
C_e(\mathrm{h}, \pi_0, K) &= 1 - H_{\mathrm{h}}(Z_i | Z_{i-1}, \ldots, Z_{i-K}, \pi_0)) \\
&= 1 - \sum_{z_j \in \{0,1\}} \Pr(z_{i-K}, \ldots, z_{i-2}, z_{i-1} | \pi_0) h(z_{i-K}, \ldots, z_{i-2}, z_{i-1} | \pi_0) \\
&= 1 - \sum_{N=0}^{K-1} \Pr(\mathbf{1}\mathbf{0}^N | \pi_0) h(\mathbf{1}\mathbf{0}^N | \pi_0) - \Pr(\mathbf{0}^K | \pi_0) h(\mathbf{0}^K | \pi_0).
\end{aligned}
\tag{B.8}
$$

Finally, putting it all together, ends in this

$$
\begin{aligned}
C_e(\mathrm{h}, \pi_0, K) = 1 &+ \sum_{n=0}^{N-1} \Pr(\mathbf{1}_{N-n} | \pi_0) v(n) \log_2 v(n) + \Pr(\mathbf{1}_1 | \pi_0) u(N) \log_2(u(N)) \\
&+ r(N+1 | \pi_0) \log_2(r(N+1 | \pi_0)) - r(N | \pi_0) \log_2(r(N | \pi_0)) \\
&+ s(N | \pi_0) \log_2(s(N | \pi_0))
\end{aligned}
\tag{B.9a}
$$

$$
r(K | \pi_0) = \Pr(\mathbf{0}^K | \pi_0)
\tag{B.9b}
$$

$$
= \Pr(0 | \pi_0) - \Pr(\mathbf{1}_0 | \pi_0) \sum_{k=1}^{K-1} u(k)
\tag{B.9c}
$$

$$
= r(K-1) - u(K-1) \Pr(1 | \pi_0),
\tag{B.9d}
$$

$$
s(K | \pi_0) = r(K | \pi_0) - r(K+1 | \pi_0)
\tag{B.9e}
$$

$$
= u(K) \Pr(\mathbf{1}_0 | \pi_0),
\tag{B.9f}
$$

$$
u(K) = \Pr(\mathbf{0}^K | 1),
\tag{B.9g}
$$

$$
v(K) = u(K) - u(K+1).
\tag{B.9h}
$$

This formula can also be expressed in terms of Δ_i, the difference of the actual state vector from the steady state vector

$$
\begin{aligned}
C_e(\mathrm{h}, \pi_0, K) = {} & 1 + \Pr(1) \sum_{n=0}^{N-1} v(n) \log_2 v(n) + \sum_{n=0}^{N-1} \Delta_i P_B v(n) \log_2 v(n) \\
& + \Pr(1) u(N) \log_2(u(N)) + \Delta_1 P_B \log_2(u(N)) \\
& + \left(r(N+1) - \Delta_0 P_B \sum_{k=0}^{N} u(k) \right) \log_2 \left(r(N+1) - \Delta_0 P_B \sum_{k=0}^{N} u(k) \right) \\
& - \left(r(N) - \Delta_0 P_B \sum_{k=0}^{N-1} u(k) \right) \log_2 \left(r(N) - \Delta_0 P_B \sum_{k=0}^{N-1} u(k) \right) \\
& + \left(s(N) + \Delta_0 P_B u(N) \right) \log_2 \left(s(N) + \Delta_0 P_B u(N) \right).
\end{aligned}
\tag{B.10}
$$

These equations are unwieldy, because they depend on the initial state. However [38] showed, that the statistical modelling quality does actually not decrease, even when only few triple-configurations are allowed. A similar statement can be seen in the given formulas. If the product of the initial difference and the error probability $\Delta_0 P_B$ are sufficiently small, they can be equalled to zero, and the influence of the initial state disappears completely. The underlying assumption for this is then that the channel parameters have not changed dramatically, and the steady-state vectors are roughly the same before and after change. Additionally, low bit error rates also diminish the influence of the initial vector.

For this case, where the steady state vector π^* stays roughly equal during the channel change, the capacity can be simplified into

$$
\begin{aligned}
C_e(\mathrm{h}, \pi_0 = \pi_1^*, K) = {} & 1 + \Pr(1) \sum_{n=0}^{N-1} v(n) \log_2 v(n) + r(N+1) \log_2(r(N+1)) \\
& - r(N) \log_2(r(N)) + s(N) \log_2(u(N)^2 \Pr(1)).
\end{aligned}
\tag{B.11}
$$

B.1.3 Block Transmission Capacity

The actual channel model describes the aggregation of blocks over time in a uniform manner. This means, that the channel is always assumed constant for a fixed amount of time T_c, and then allowed to change. The amount of iteration steps happening in one such timespan is called the block length L in the following. It is characterized by $p_L(l|T_c) = \Pr(L = l|T_c)$. Since the timespan is a design parameter, and once set then fixed, this conditioning will be suppressed.

Recalling the integral form of the overall capacity Eq. (B.7)

$$C = \sum_{K=1}^{\infty} \int_{\mathrm{h}\in\mathcal{H}_K} \int_{\pi_0\in\Pi} \Pr(K|\mathrm{h}^0) f(\mathrm{h}|\pi_0) f(\pi_0) C_{h,K}^0 \, \mathrm{d}\pi_0 \mathrm{d}\mathrm{h},$$

the conditional probability $\Pr(K|\mathrm{h},\pi_0)$ becomes independent due to the model properties. Therefore, the capacity can be reordered as

$$C = \int_{\mathrm{h}\in\mathcal{H}_K} \int_{\pi_0\in\Pi} f(\mathrm{h}|\pi_0) f(\pi_0) \sum_{l\in\mathcal{L}} p(l) \underbrace{\sum_{K=1}^{l} \frac{1}{l} C_e(\mathrm{h},\pi_0,K)}_{=C_B(\mathrm{h},\pi_0,l)} \, \mathrm{d}\pi_0 \mathrm{d}\mathrm{h}. \qquad \text{(B.12)}$$

This equation additionally takes into account, that within one transmission block of length L, every sequence length $K \in [1,L]$ occurs exactly once, as every consecutive bit in the sequence has one more past symbol to take into account. Additionally the block transmission capacity $C_B(\mathrm{h},\pi_0,l)$ is averaged over all possible lengths.

$$C_B(\mathrm{h},\pi_0,l) = 1 + \frac{1}{l} \sum_{N=1}^{l}$$

$$\Pr(1) \sum_{n=0}^{N-1} v(n) \log_2 v(n) + \sum_{n=0}^{N-1} \Delta_i P_B v(n) \log_2 v(n)$$

$$+ \Pr(1) u(N) \log_2(u(N)) + \Delta_1 P_B \log_2(u(N))$$

$$+ \left(r(N+1) - \Delta_0 P_B \sum_{k=0}^{N} u(k)\right) \log_2\left(r(N+1) - \Delta_0 P_B \sum_{k=0}^{N} u(k)\right)$$

$$- \left(r(N) - \Delta_0 P_B \sum_{k=0}^{N-1} u(k)\right) \log_2\left(r(N) - \Delta_0 P_B \sum_{k=0}^{N-1} u(k)\right)$$

$$+ \left(s(N) + \Delta_0 P_B u(N)\right) \log_2\left(s(N) + \Delta_0 P_B u(N)\right). \qquad \text{(B.13)}$$

$$C_B(\mathrm{h}, \pi_0, l) = 1 + \Pr(1)\sum_{n=0}^{l-1}(1 - \frac{n}{l})v(n)\log_2 v(n) + \sum_{n=0}^{l-1}(1 - \frac{n}{l})\Delta_i P_B v(n)\log_2 v(n)$$

$$+ \frac{1}{l}\sum_{N=1}^{l}\Pr(1)u(N)\log_2(u(N)) + \Delta_1 P_B \log_2(u(N))$$

$$+ \frac{1}{l}(r(l+1) - \Delta_0 P_B \sum_{k=0}^{l}u(k))\log_2(r(l+1) - \Delta_0 P_B \sum_{k=0}^{l}u(k))$$

$$- \frac{1}{l}(r(1) - \Delta_0 P_B)\log_2(r(1) - \Delta_0 P_B)$$

$$+ \frac{1}{l}\sum_{N=1}^{l}(s(N) + \Delta_0 P_B u(N))\log_2(s(N) + \Delta_0 P_B u(N)). \tag{B.14}$$

Again, this equation becomes somewhat easier to grasp if we look at the case where the $\Delta_0 P_B$ product is insignificant. Then, the equation becomes

$$C_B(\mathrm{h}, \pi_0 = \pi_1^*, l) = 1 + \Pr(1)\sum_{n=0}^{l-1}(1 - \frac{n}{l})v(n)\log_2 v(n)$$

$$+ \frac{1}{l}\sum_{N=1}^{l}\Pr(1)u(N)\log_2(u(N)) + \frac{1}{l}\sum_{N=1}^{l}(s(N))\log_2(s(N))$$

$$+ \frac{1}{l}(r(l+1)\log_2(r(l+1)) - r(1))\log_2(r(1)). \tag{B.15}$$

C
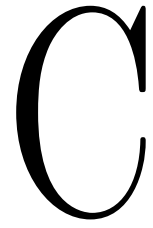
Gantry analysis

C.1 Analysis per Cluster

C.1.1 Cluster 1

This cluster lies around a turnaround point, where the measurement cars switched directions. Since these are entirely different conditions to the other taken measurements, we will not try to analyze these moments.

C.1.2 Cluster 2

Figures C.1 and C.2 show PDR and SNR curves for cluster 2, with the red line signifying the time the center position between the cars passed a chosen gantry. In this case, the behaviour strongly suggests that the gantry had an interferer mounted, as both directions almost universally see the combination of PDR drop and SNR rise.

C.1.3 Cluster 3

Figures C.3 and C.4 show PDR and SNR curves for cluster 3, with the red line signifying the time the center position between the cars passed a chosen gantry. At this cluster, there were multiple gantries in close proximity, and we also see multiple drops, however, they are not as consistent as for cluster 3, and not accompanied by SNR increases.

C.1.4   Cluster 4

Figures C.5 and C.6 show PDR and SNR curves for cluster 4, but this time the red line is an overpass. The eastwards measurements again show PDR drops, however the westwards measurements did not show any apparent different behaviour.

C.1.5   Cluster 5

Here, only the eastwards route showed interesting results, seen in Fig. C.7. We see the same behaviour again, a PDR drop that is not correlated to the SNR behaviour.

C.2   Conclusion

The measurement-by-measurement analysis has shown, that roadside structures such as gantries and overpasses can indeed semi-consistently influence the channel performance. We also identified an obvious interferer.
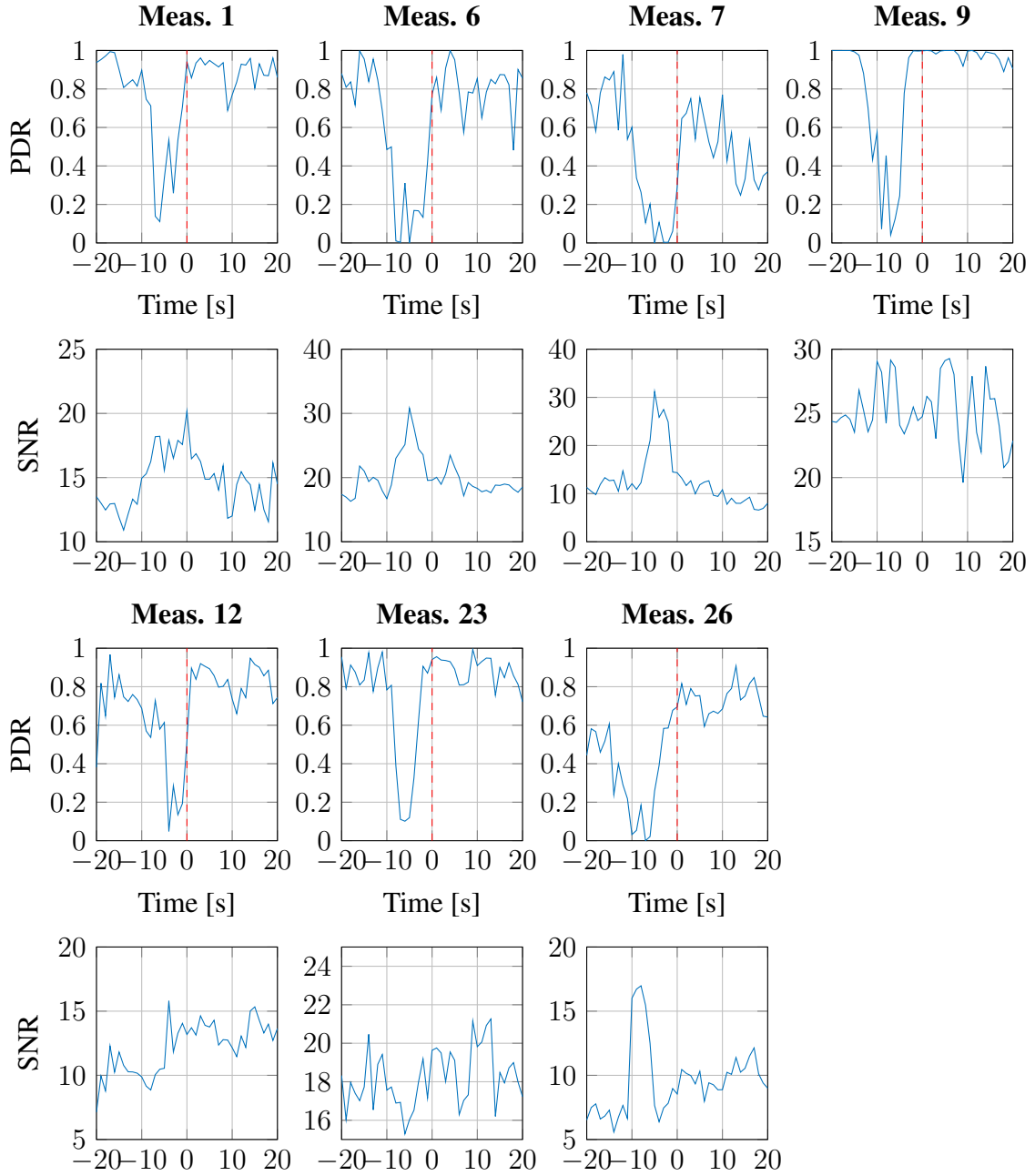
Figure C.1: PDR and SNR curves of all measurements passing cluster 2 going eastwards.
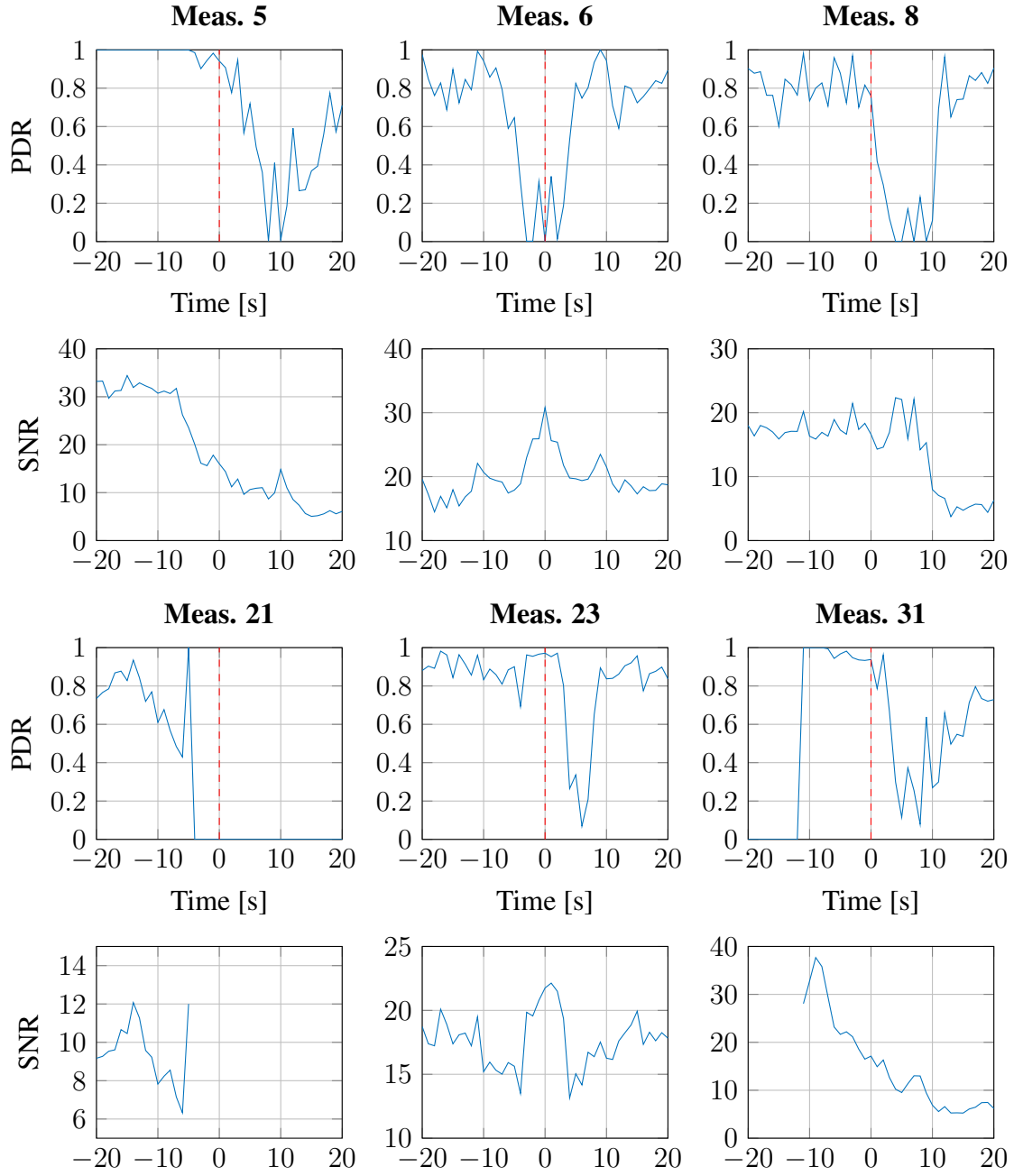
Figure C.2: PDR and SNR curves of all measurements passing cluster 2 going westwards.

Figure C.3: PDR and SNR curves of all measurements passing cluster 3 going eastwards.

Figure C.4: PDR and SNR curves of all measurements passing cluster 3 going westwards.
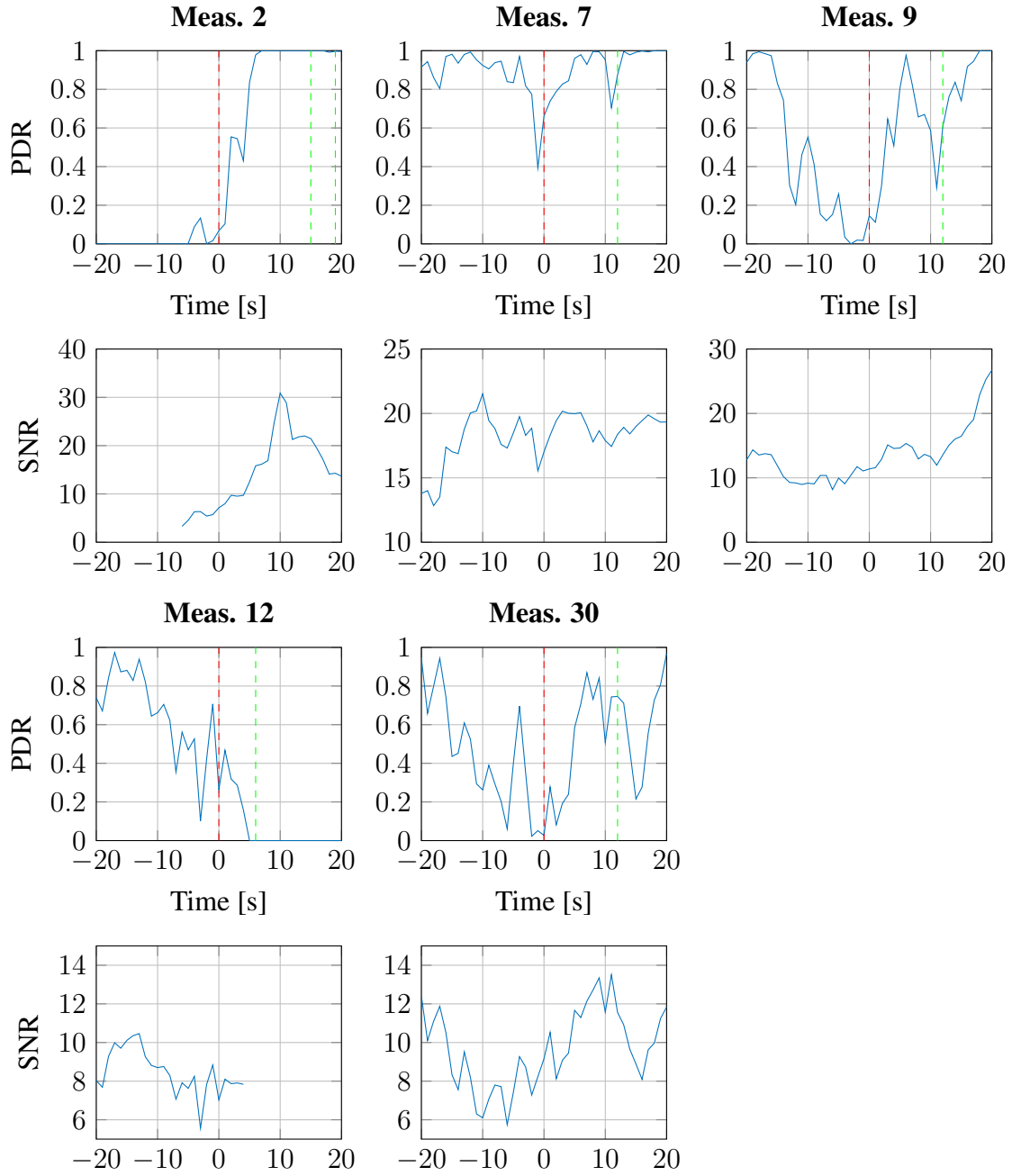

Figure C.5: PDR and SNR curves of all measurements passing cluster 4 going eastwards.
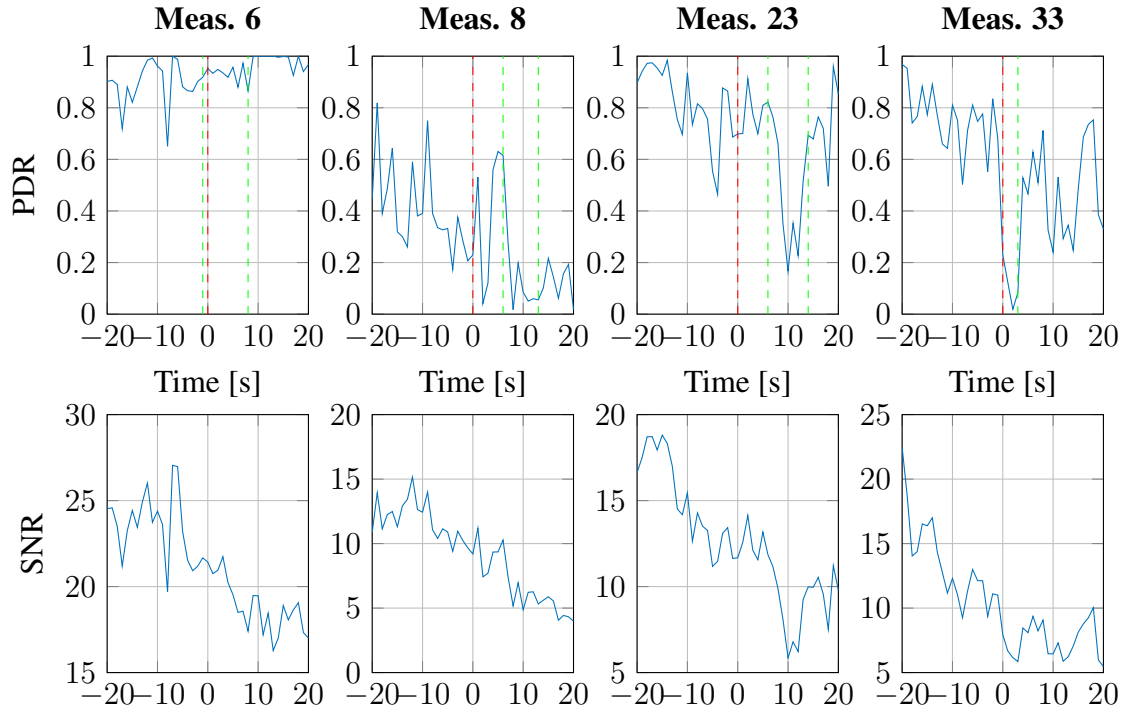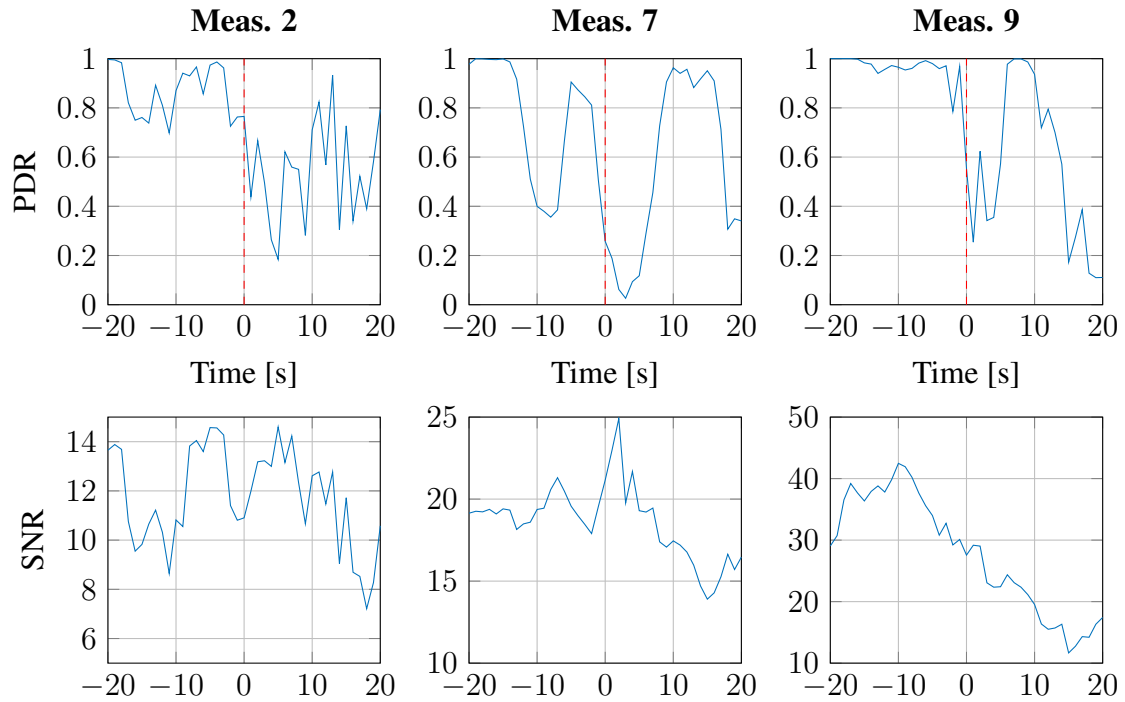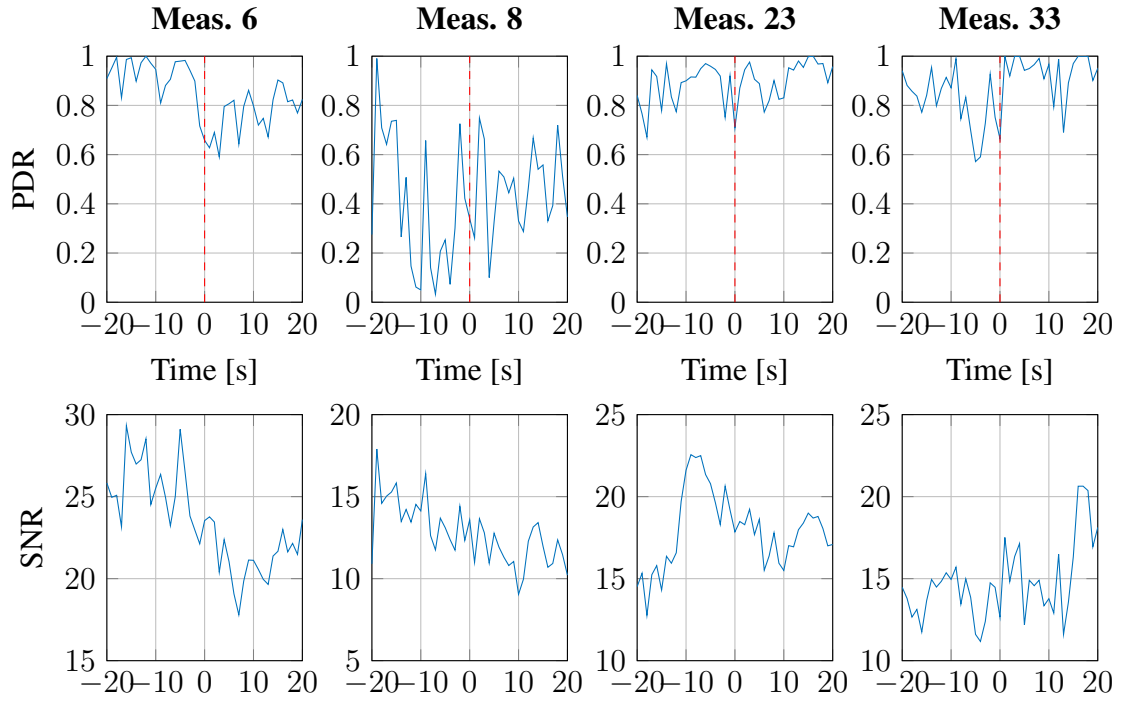
Figure C.6: PDR and SNR curves of all measurements passing cluster 4 going westwards.
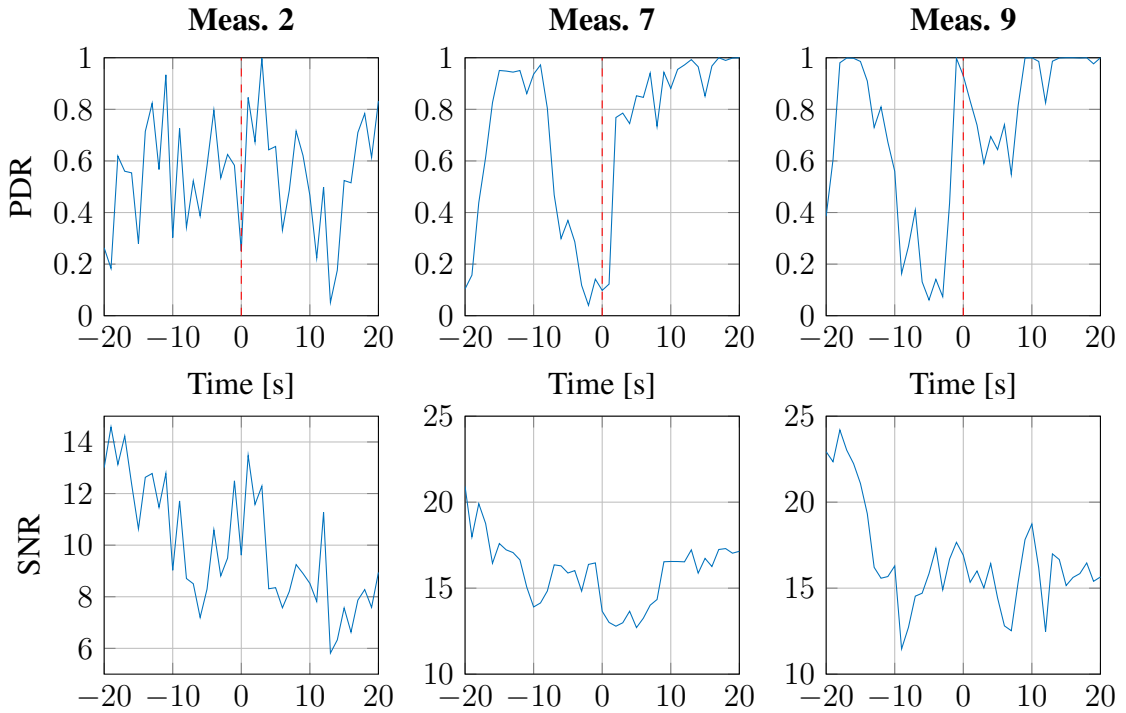

Figure C.7: PDR and SNR curves of all measurements passing cluster 5 going eastwards.

Bibliography

[1] "Passenger transport statistics." Available online: http://ec.europa.eu/eurostat/statistics-explained/index.php/Passenger_transport_statistics.

[2] "Number of passenger cars and commercial vehicles in use worldwide from 2006 to 2013 (in millions)." Available online: http://www.statista.com/statistics/281134/number-of-vehicles-in-use-worldwide/.

[3] "Production of cars in china from 2005 to july 2015." Available online: http://www.statista.com/statistics/281133/car-production-in-china/.

[4] "Global status report on road safety 2013: Supporting a decade of action," tech. rep., World Health Organization, 2013.

[5] J. Broughton and C. Baughan, "The effectiveness of antilock braking systems in reducing accidents in great britain," *Accident Analysis & Prevention*, vol. 34, no. 3, pp. 347–355, 2002.

[6] A. Erke, "Effects of electronic stability control (ESC) on accidents: A review of empirical evidence," *Accident Analysis & Prevention*, vol. 40, no. 1, 2008.

[7] P. F. Drucker, "Internet of things," 2015.

[8] The Commission of the European Communities, "Commission decision on the harmonised use of radio spectrum in the 5875-5905 MHz frequency band for safety-related applications of intelligent transport systems (ITS)." 008/671/EC, Aug. 2008.

[9] IEEE Computer Society, "IEEE standard for information technology– local and metropolitan area networks– specific requirements– part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 6: Wireless access in vehicular environments," July 2010. Available online: http://ieeexplore.ieee.org/servlet/opac?punumber=5514473.

[10] P. Ross, "Thus spoke the autobahn," *IEEE Spectrum*, vol. 52, pp. 52–55, Jan. 2015.

[11] C. Bergenhem, Q. Huang, A. Benmimoun, and T. Robinson, "Challenges of platooning on public motorways," in *17th world congress on intelligent transport systems*, pp. 1–12, 2010.

[12] V. Shivaldova, A. Winkelbauer, and C. Mecklenbräuker, "Vehicular Link Performance: From Real-World Experiments to Reliability Models and Performance Analysis," *Vehicular Technology Magazine, IEEE*, vol. 8, pp. 35–44, Dec 2013.

[13] E. N. Gilbert, "Capacity of a burst-noise channel," *Bell System Technical Journal*, vol. 39, no. 5, pp. 1253–1265, 1960.

[14] E. Giordano, R. Frank, G. Pau, and M. Gerla, "CORNER: a realistic urban propagation model for VANET," in *Proc. 7th International Conference on Wireless On-demand Network Systems and Services (WONS)*, pp. 57–60, Feb. 2010.

[15] C. Sommer and F. Dressler, "Using the right two-ray model? a measurement based evaluation of PHY models in VANETs," in *Proc. 17th ACM International Conference on Mobile Computing and Networking (MobiCom)*, ACM, Sept. 2011.

[16] G. Grau, D. Pusceddu, S. Rea, O. Brickley, M. Koubek, and D. Pesch, "Vehicle-2-vehicle communication channel evaluation using the CVIS platform," in *Proc. 7th International Symposium on Communication Systems Networks and Digital Signal Processing (CSNDSP)*, pp. 449–453, July 2010.

[17] C. Sommer, D. Eckhoff, R. German, and F. Dressler, "A computationally inexpensive empirical model of ieee 802.11p radio shadowing in urban environments," in *Proc. 8th International Conference on Wireless On-Demand Network Systems and Services (WONS)*, pp. 84–90, Jan. 2011.

[18] R. He, A. Molisch, F. Tufvesson, Z. Zhong, B. Ai, and T. Zhang, "Vehicle-to-vehicle propagation models with large vehicle obstructions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, pp. 2237–2248, Oct. 2014.

[19] R. Meireles, M. Boban, P. Steenkiste, O. Tonguz, and J. Barros, "Experimental study on the impact of vehicular obstructions in VANETs," in *Proc. IEEE Vehicular Networking Conference (VNC)*, pp. 338–345, Dec. 2010.

[20] T. Mangel, O. Klemp, and H. Hartenstein, "A validated 5.9 GHz non-line-of-sight path-loss and fading model for inter-vehicle communication," in *Proc. 11th International Conference on ITS Telecommunications (ITST)*, pp. 75–80, Aug. 2011.

[21] M. Boban, T. Vinhoza, M. Ferreira, J. Barros, and O. Tonguz, "Impact of vehicles as obstacles in vehicular ad hoc networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, pp. 15–28, Jan. 2011.

[22] Y. Zang, L. Stibor, G. Orfanos, S. Guo, and H.-J. Reumerman, "An error model for inter-vehicle communications in highway scenarios at 5.9GHz," in *Proc. 2nd International Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks*, pp. 49–56, Oct. 2005.

[23] S. Demmel, A. Lambert, D. Gruyer, G. S. Larue, and A. Rakotonirainy, "IEEE 802.11p empirical performance model from evaluations on test tracks," *Journal of Networks*, vol. 9, no. 6, pp. 1485–1495, 2014.

[24] L.-T. Yang, H. Jiang, C.-C. Guo, Y.-H. Wang, J. Wu, and L.-J. Chen, "A four-state Markov model based on measurements for evaluating the packet-level performance of VANET," in *Proc. 68th IEEE Vehicular Technology Conference (VTC Fall)*, Sept. 2008.

[25] "Project "Robust And Distributed Safety-Improved Traffic Telematics" (ROADSAFE)." Available online: https://portal.ftw.at/projects/roadsafe/.

[26] "Cooperative vehicle-infrastructure systems (CVIS) project." Available online: http://www.cvisproject.org/.

[27] E. Elliott, "Estimates of error rates for codes on burst-noise channels," *Bell System Technical Journal*, vol. 42, pp. 1977–1997, Sept 1963.

[28] G. Haßlinger and O. Hohlfeld, "The Gilbert-Elliott model for packet loss in real time services on the internet," in *Proc. Measuring, Modelling and Evaluation of Computer and Communication Systems , 2008*, pp. 1–15, VDE, 2008.

[29] D. Dhoutaut, A. Régis, and F. Spies, "Impact of radio propagation models in vehicular ad hoc networks simulations," in *Proc. of the 3rd international workshop on Vehicular ad hoc networks*, pp. 40–49, ACM, 2006.

[30] M. Yajnik, S. Moon, J. Kurose, and D. Towsley, "Measurement and modelling of the temporal dependence in packet loss," in *Proc. IEEE INFOCOM*, vol. 1, pp. 345–352 vol.1, Mar 1999.

[31] J. Korhonen and Y. Wang, "Effect of packet size on loss rate and delay in wireless links," in *Wireless Communications and Networking Conference, 2005 IEEE*, vol. 3, pp. 1608–1613 Vol. 3, March 2005.

[32] S. R. Eddy, "Hidden Markov models," *Current opinion in structural biology*, vol. 6, no. 3, pp. 361–365, 1996.

[33] T. M. Cover and J. A. Thomas, "Entropy, relative entropy and mutual information," *Elements of Information Theory*, pp. 12–49, 1991.

[34] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison, *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. Cambridge university press, 1998.

[35] N. Tishby, F. Pereira, and W. Bialek, "The information bottleneck method," in *Proc. 37th Allerton Conference on Communication, Control and Computing*, pp. 368–377, 1999.

[36] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, pp. 79–86, 03 1951.

[37] E. L. Lehmann and G. Casella, *Theory of point estimation*, vol. 31. Springer Science & Business Media, 1998.

[38] V. Shivaldova and C. Mecklenbräuker, "Quantization-based complexity reduction for range-dependent modified gilbert model," in *Proc. 8th IEEE Sensor Array and Multi-channel Signal Processing Workshop (SAM)*, pp. 345–348, June 2014.