This document is a preprint of: F. Burmeister, N. Schwarzenberg, A. Trassl et al., "Data-Driven Channel Modeling for Industrial URLLC-Motivated PHY Investigations," in Proceedings of IEEE Wireless Communications and Networking Conference (WCNC 2022), Austin, USA, Apr 2022.

Data-Driven Channel Modeling for Industrial URLLC-Motivated PHY Investigations

Friedrich Burmeister, Nick Schwarzenberg, Andreas Traßl, Richard Jacob, and Gerhard Fettweis
Vodafone Chair Mobile Communications Systems, Technische Universität Dresden, Germany
{friedrich.burmeister, nick.schwarzenberg, andreas.traßl, richard.jacob, gerhard.fettweis}@tu-dresden.de

***Abstract*—Ultra-Reliable Low-Latency Communications (URLLC) is a prerequisite for advancing industrial automation. To verify whether communications systems meet these stringent requirements, physical layer simulations are a powerful tool. Such simulations rely on a large number of channel realizations to obtain statistically significant results. Using exclusively real-world measured channels is challenging, e.g., due to the measurement time needed. In this work, we propose how to derive a channel model that mimics not only the mean but equally the temporal behavior of data from an industrial channel measurement campaign. The approach considers time variation on a large scale, modeled through a Markov process, as well as fading of individual channel components, achieved through Doppler-filtered random processes based on empirical distributions. The model is validated in link-level simulations by comparing the synthesized channels to the original measured data by means of performance metrics relevant to URLLC, including mean and maximum outage durations. Our validations show that the model performs well and especially the outage durations, caused by consecutive packet losses, match the real channel characteristics very well. In the future, we can use the model to investigate how extensive measurements need to be in order to make statements about the performance of communications systems.**

***Index Terms*—Channel Modeling, Radio Channels, Time Variation, Industrial Communications.**

I. Introduction

Ultra-Reliable Low-Latency Communications (URLLC) is considered to have a key role for automation applications coming up with future industrial processes [1]. Above all, a high reliability of the communications system is essential in order to maximize the availability of industrial applications, and thus to avoid downtimes. From an URLLC application point of view, high reliability can be achieved by extremely low Packet Error Rates (PERs) [2]. However, if the application can tolerate occasional packet errors or if one takes the mission duration of those applications into account [3], high reliability in the sense of the application is not necessarily bound to the PER. That is why new dependability metrics are studied for assessing the reliability of URLLC applications, which consider consecutive packet losses and thus the temporal correlation of fading processes and outages [4].

In general, link-level simulations are a first step to investigate the reliability of a communications system, e.g., in [5]. However, the credibility of simulation results strongly depends on the considered radio channel model. Using measured channel data from representative environments in the form of playback simulations would be beneficial for the value of reliability studies. Unfortunately, the pure measurement time required for this would be enormous. Assuming the most relaxed requirements of a closed-loop control application from [6], one needs a PER of 10^{-8} with a transmission interval of $2\,\mathrm{ms}$. To theoretically achieve this accuracy, at least 10^8 packets have to be simulated. With one packet every second millisecond, the channel has to be recorded consecutively for more than 2 days. To nevertheless make comprehensive simulations feasible, channel models have to be derived from the available measurement data. In this context, the question arises what kind of modeling approach is meaningful.

To investigate the temporal occurrence of packet errors by simulations, a channel model needs to reproduce the behavior of the available measurement data, including correlations in time at the scale of communications packets as well as on the large scale. However, to the best of our knowledge, campaigns measuring the channel with mobility for several seconds with a temporal resolution of milliseconds are scarce, especially for industrial indoor environments (see literature overview in our previous work [7]). This fact is also reflected in the number of channel models focusing on modeling temporal variations based on industrial channel measurement data. The authors in [8] present a model for time-varying indoor office channels. Time variation is achieved by using a Markov process to model the appearance and disappearance of taps in the channel impulse response (CIR). However, parameters are derived for Line of Sight (LOS) and Non-Line-of-Sight (NLOS) areas separately with no transition between them. Similar to [8], the existence of multi-path components in the CIR is modeled with a Markov process in [9], but the authors in the latter emphasize the importance of considering transitions between LOS and NLOS. In [10], an industrial channel model, combining a Path Loss (PL) model and a model for small-scale fading, has been parameterized based on measurement data for LOS and NLOS separately. However, the fading parameters of the taps in the CIR do not change over time and changes from LOS to NLOS are not considered either.

To fill this gap, we present a data-driven channel model that synthesizes new channel realizations reproducing not only the statistical behavior of the measurement data but also its correlation in time and frequency. Therefore, measured data is actively incorporated in the form of measured distributions. The model reproduces the evolution of large-scale properties, e.g., the Excess Delay (ED), with a state-based

Markov chain to allow for smooth changes of large-scale conditions. On a small scale, the time variation of individual channel realizations is modeled by synthetic fading based on measured fading distributions. The validation of the model relies on a comparison of link-level simulation results between the measured data and synthetic channel data generated by our model. By evaluating the PERs as well as mean and maximum number of consecutive packet errors, we show that the proposed modeling mechanisms are sufficient to mimic the simulation results of the measurement data.

II. Measurement Data

The underlying data used in this work originates from an industrial channel measurement campaign, described in detail in [7]. The measurement environment provided an industry-typical and thus rich scattering with strong multi-path propagation, leading to frequency-selective radio channels. In addition, the receiver was attached to a moving Automated Guided Vehicle (AGV). The resulting relative movement between transmitter and receiver during the measurements caused the measured channel to vary over time. The AGV had a constant velocity of $1\,\mathrm{m/s}$, following a reproducible trajectory on a rectangular round. The measurement campaign consisted of 10 scenarios each differing in terms of obstacle positions. In this work, we focus on the machines scenario shown in [7]. The trajectory of about $20\,\mathrm{s}$ of driving time covered varying transmission distances as well as varying LOS conditions. The bandwidth of the utilized channel sounder was set to $f_S = 100\,\mathrm{MHz}$ at a carrier frequency of $f_C = 3.75\,\mathrm{GHz}$. Due to the deterministic velocity v and f_C, the expected coherence time T_{Coh} of the channel can be estimated using [11]:

$$T_{\mathrm{Coh}} \approx \frac{9c}{16\pi v f_C} = 14.3\,\mathrm{ms} \quad . \tag{1}$$

With one measured CIR per millisecond, the time between consecutive measurement points was clearly below the coherence time. The used bandwidth combined with a high temporal measurement rate allowed us to capture both the frequency and time selectivity of the channel.

III. Modeling

A. Assumptions

Based on the measurements, we determine assumptions concerning a single CIR. First, we assume that there is a start sample in the CIR that indicates the first arriving channel component. Second, samples in the CIR having an amplitude less than a certain fraction of the maximum amplitude of the CIR are assumed to be irrelevant. Thus, there is an end sample and the CIR has a finite length of N samples between start and end point. The density of physical scatterers in the considered environment is high with respect to the sampling frequency and hence to the resolution of resolvable physical paths. We additionally assume the scatterers to be uncorrelated, i.e., in this work, all N samples between start and end point are considered as independent multi-path taps. This assumption seems severely simplifying and contrasts

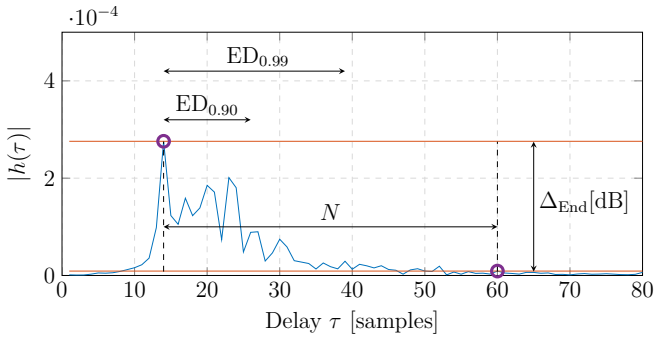


Fig. 1: A typical NLOS CIR from an industrial measurement campaign indicates multi-path propagation. The start and end of the CIR, as well as two exemplary EDs, are depicted.

with [8], [10], where samples of the CIR are grouped into clusters. However, because of the high density of scatterers, the amplitudes of the samples are assumed to originate from the sum of several multi-path components and hence to be subject to fading [11]. That is, according to the Tapped Delay Line model in [11], the CIR

$$h(t,\tau) = \sum_{n=0}^{N(t)-1} c_n(t) e^{j\phi_n(t)} \cdot \delta(\tau - \tau_n) \tag{2}$$

can be interpreted as a time-varying Finite Impulse Response (FIR) filter including $N(t)$ channel taps with its time-varying amplitudes $c_n(t)$ and phases $\phi_n(t)$ and its corresponding delays τ_n. Thereby, $\delta(\cdot)$ corresponds to the Dirac delta function. By assuming that samples in the CIR are equal to taps, the delays τ_n are fixed and equal to n/f_{S}. The number of taps $N(t)$ varies over time due to changing large-scale conditions.

To determine the start and end point, a threshold compared to the maximum power of the CIR is used. The sample that is considered as start point is the first sample with a power higher than $\Delta_{\mathrm{Start}}\,[\mathrm{dB}]$ below the maximum power. Similarly, the end point is found by looking for the last sample with a power higher than $\Delta_{\mathrm{End}}\,[\mathrm{dB}]$ below the maximum power. For our measurement system, we found that $\Delta_{\mathrm{Start}} = 10\,\mathrm{dB}$ and $\Delta_{\mathrm{End}} = 30\,\mathrm{dB}$ work reasonably. While these parameters are relative and therefore applicable to other physical measurement scenarios, they still depend on the measurement hardware and may need to be tuned for other systems. An exemplary CIR from the measurements including the determined start and end points as well as Δ_{End} is shown in Fig. 1.

Further, we infer assumptions that concern consecutive CIRs. Even if single taps in the CIR are subject to fading, the time between consecutively measured CIRs is below T_{Coh}. That means, large-scale characteristics such as the mean receive power as well as LOS conditions are assumed to be constant for a certain number of consecutive CIRs. Those characteristics change on a time period larger than T_{Coh}. Accordingly, we assume that consecutive CIRs can be arranged into segments of length $T_{\mathrm{Seg}} > T_{\mathrm{Coh}}$ as done in [9] (see Fig. 2). Segments are characterized by large-scale parameters that are assumed to be constant within one segment.

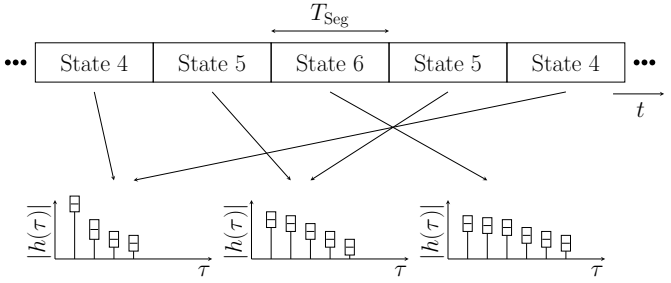


Fig. 2: Consecutive CIRs are divided into segments of length T_{Seg}. Amplitude distributions are represented by box plots in the CIRs.

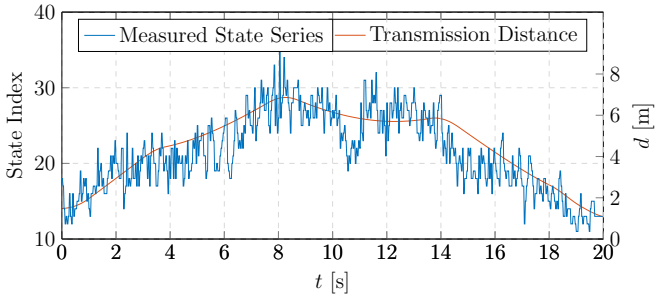


Fig. 3: Applying segmentation and ED-based state labeling to an exemplary measurement round shows the resulting state sequence over time. The distance between transmitter and receiver is plotted on top and shows how state and large-scale conditions are related.

B. Large-scale Modeling

The goal of the large-scale modeling is to classify segments and extract probabilities for their occurrence in order to synthesize new segment time series. The Excess Delay (ED) seems to be a suitable parameter to distinguish segments. On the one hand, it changes with the large-scale conditions [12]. On the other hand, the ED is a metric that indicates the number of relevant taps in a CIR. This is meaningful for the synthesis of CIRs to reduce the number of taps to be modeled to the most relevant ones. In particular, we define $\mathrm{ED}_X \leq N$ as the number of necessary samples to exceed an arbitrary fraction X of the CIR energy. This can be expressed by

$$\sum_{l=0}^{\mathrm{ED}_X-1} |h(\tau-\tau_l)|^2 \overset{!}{\geq} X \sum_{l=0}^{N-1} |h(\tau-\tau_l)|^2 \quad . \tag{3}$$

Thereby, the ED can be understood as measure of significance. Samples in a CIR that are incorrectly considered as taps, e.g., due to an unfavorably tuned Δ_{End} are thus ignored, making the modeling more robust. Figure 1 indicates that 99% of the energy within this CIR is accumulated already after a fraction of the N samples. The determined excess delays $\mathrm{ED}_X(t)$ are averaged segment-wise and rounded to an integer number of samples. This numerical value per segment is defined as state index S, indicating to which state this segment belongs. The state series observed during a measurement round is visualized in Fig. 3. One can see the states depending on the transmission distance (increasing state index with increasing distance) as well as on changing LOS conditions. In the areas with higher state indices (from 6 s to 10 s and from 11 s to 14 s), the LOS was blocked by two metallic obstacles.

Due to segmentation with subsequent labeling of the segments, the state index S representing the large-scale conditions lies in a finite state space. To characterize and model the evolution of states, a Discrete-Time Markov Chain (DTMC) is used. Therefore, the overall state probabilities as well as the state transition probabilities have to be extracted from the measurements. The goal is to use the obtained probabilities to model synthetic large-scale state time series later on.

For modeling time series, we want to dispense with the classical Markov property, meaning that the next state S_{n+1} solely depends on the current state S_n. Instead, a Markov chain with a memory of M steps is considered in order to allow for the synthesis of state sequences that are similar to measured time series. That is, the next state S_{n+1} depends on $M+1$ states in total, i.e., it depends on the state sequence $(S_n, S_{n-1}, ..., S_{n-M})$. To obtain the overall probability for a specific sequence, we count the number of occurrences of this sequence and divide it by the total number of sequences in the measurements. The overall probabilities are necessary to know for the beginning of the synthesis procedure, i.e., to randomly choose a starting sequence. More important are the transition probabilities for the next state S_{n+1} based on a specific sequence. To obtain this transition probability, we count the number of occurred changes into the state based on this specific sequence and divide this by the total occurrence of this particular sequence. Since the state space is finite, all transition probabilities can be represented by a $(M+2)$-dimensional probability tensor T. For $M=1$, the (i,j,k)-th element of T is

$$p_{ijk} = \Pr(S_{n+1} = i | S_n = j, S_{n-1} = k) \tag{4}$$

and gives the probability for entering state i, if the current state is j and the previous state was k. The size M of the Markov chain memory can be interpreted as a hyperparameter. Note that having no correlation between states, i.e., $M=0$, might end in inaccurate time variations the channel model produces. Choosing M very high removes the randomness from synthetic state series.

After obtaining the overall state probabilities and the state transition probabilities from the measurements, we are interested in synthesizing random state series. The starting point is a randomly selected sequence of length $M+1$. Thereby, the current state s_n and the M previous states are given. From T, all probabilities for the next possible states s_{n+1} are known, i.e., we can construct the discrete Probability Density Function (PDF) $f_{S_{n+1}}(s_{n+1})$ for the next occurring state. To randomly draw the next state according to the probability densities in $f_{S_{n+1}}(s_{n+1})$, Inverse Transform Sampling (ITS) is used. This allows to sample randomly from a certain PDF given its Cumulative Distribution Function (CDF). By accumulating the individual probabilities in $f_{S_{n+1}}(s_{n+1})$, one obtains the discrete CDF $F_{S_{n+1}}(s_{n+1})$. The inversion of the CDF and a random sample u from a uniform distribution $U \sim \mathrm{Unif}[0,1]$

allows to draw random values from $f_{S_{n+1}}(s_{n+1})$. That is, the next state of the synthetic time series results from

$$s_{n+1} = F^{-1}_{S_{n+1}}(u) \quad . \tag{5}$$

The new state is appended to the synthetic time series and the procedure repeats until a desired length of the time series is reached. For more details about ITS, please refer to [13]. In this work, we always consider ITS to be used if a random sample has to be drawn from a specific (measured) PDF.

C. Small-scale Modeling

The state indices given by the large-scale modeling determine the fading parameters of a segment, but not the individual fading realizations. That is why we need a small-scale modeling layer dealing with the generation of CIRs and their small-scale variations. After dividing the measured CIRs into segments and classifying them based on ED_X, we need to collect statistical amplitude distributions. In [14] is shown that recurring characteristics of CIRs occur in industrial environments. Inspired by this, we propose to merge amplitude distributions of all CIRs from segments with the same state index. This is illustrated with the help of Fig. 2. From these distributions, we aim to generate concrete CIRs during synthesis. If time-varying CIRs have to be synthesized, two questions arise. First, how many taps have to be considered in the delay domain? Second, how to model the amplitudes $c_n(t)$ and phases $\phi_n(t)$ per taps and their temporal behavior?

As described in III-B, the ED_X is calculated per CIR. We determine the number of taps to be modeled in a CIR of one state differently from the classification for large-scale modeling by determining the maximum ED_X occurring from all segments of that state. To address the second question, the focus first lies on the amplitudes before the modeling of the phases is discussed later. We propose to split the amplitude $c_n(t)$ of delay index n into a constant mean amplitude $\bar{c}_n$ and a time-varying fading amplitude $\tilde{c}_n(t)$ such that:

$$c_n(t) = \tilde{c}_n(t) \cdot \bar{c}_n \tag{6}$$

The empirical PDFs $f_{\tilde{C}_{s,n}}(\tilde{c}_{s,n})$ and $f_{\bar{C}_{s,n}}(\bar{c}_{s,n})$ for generating the above amplitude components during synthesis must first be obtained from the measurement data. To collect the amplitudes for the empirical PDFs, we start with the first segment which belongs to a certain state s according to the classification introduced in III-B. Now, for a single delay index n, the mean amplitude $\bar{c}_n$ is computed and put into the respective PDF $f_{\bar{C}_{s,n}}(\bar{c}_{s,n})$. For the same delay index n, the amplitudes $c_n(t)$ of this segment are normalized to the previously determined $\bar{c}_n$ such that the fading amplitudes

$$\tilde{c}_n(t) = \frac{c_n(t)}{\bar{c}_n} \tag{7}$$

result. These fading amplitudes are put into the data collection for $f_{\tilde{C}_{s,n}}(\tilde{c}_{s,n})$. This is repeated for all delay indices in this segment. After that, the next segment is evaluated and the amplitudes are put into the respective data collections to obtain $f_{\bar{C}_{s,n}}(\bar{c}_{s,n})$ and $f_{\tilde{C}_{s,n}}(\tilde{c}_{s,n})$ depending on the state of the next segment. This is done for all segments in the data set.

To allow for better synthesis later on, we do not only collect the empirical PDFs of the mean powers for each state, but we also determine the correlation coefficients between the mean powers of all delay indices per state. This gives us one correlation matrix K_s per state.

Now, in order to synthesize CIRs for a certain state, one needs to draw random mean amplitudes and a random fading amplitudes per delay index. Thinking of the proposed large-scale segmentation, we propose to keep the mean amplitude at a given delay constant for one segment, i.e., we first draw the mean amplitudes for all delay indices. We again use ITS and consider the correlations between the mean amplitudes by generating random realizations u_n (comp. Sec. III-B) for all delay indices such that they have the correlations from K_s between each other. After having the mean amplitudes, one has to draw the fading amplitudes per delay index. By drawing consecutive fading amplitudes for the same delay independently from each other, the channel taps are uncorrelated in time and hence do not yet reflect a correlated temporal variation of the channel with finite Doppler frequency.

To incorporate temporal correlation in the drawing of fading amplitudes, we adapt the fading simulation approach proposed in [15] to the distributions $f_{\tilde{C}_{s,n}}(\tilde{c}_{s,n})$. Fig. 4 illustrates this process. Simulated Rayleigh fading with a certain Doppler spectrum and a maximum Doppler frequency f_m serves as starting point. In contrast to [15], we consider the inverse of the measured CDF $F^{-1}_{\tilde{C}_{s,n}}(\tilde{c}_{s,n})$ depending on the state and delay index in this process. The synthesized fading amplitudes

$$\tilde{c}_{s,n}(t) = F^{-1}_{\tilde{C}_{s,n}}(F_R(r(t))) \tag{8}$$

have amplitude distributions corresponding to the measurement data and the temporal correlation of simulated Rayleigh fading. F_R corresponds thereby to the CDF of the Rayleigh distribution and $r(t)$ is the simulated Rayleigh fading amplitude. An exemplary time series of the temporally correlated fading amplitude $\tilde{c}_{s,n}(t)$ is shown in Fig. 5. The correlation becomes apparent in a finitely fast changing amplitude.

The phase of the synthesized fading $\phi(t)$ is taken from the simulated Rayleigh fading process, as proposed in [15]. Finally, the complex fading amplitude is multiplied with the segment-wise constant mean amplitude $\bar{c}_{s,n}$ drawn from $f_{\bar{C}_{s,n}}(\bar{c}_{s,n})$. The resulting synthesized amplitudes at tap index n within a segment of state s lasting from $t = 0 \ldots T_\mathrm{Seg} - 1$ can be described as follows:

$$c_{s,n}(t) = F^{-1}_{\tilde{C}_{s,n}}(F_R(r(t)))\, \bar{c}_{s,n}\, e^{j\phi(t)} \quad . \tag{9}$$

For each segment, this procedure has to be performed for all delay indices to be considered. For state transitions, the mean amplitude per delay changes and also the fading process is interrupted, resulting in sudden changes of the channel between segments. In Sec. IV, we investigate whether this is a limitation of the model.

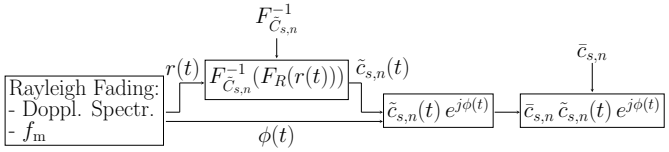


Fig. 4: Flow chart for generating measurement-like fading

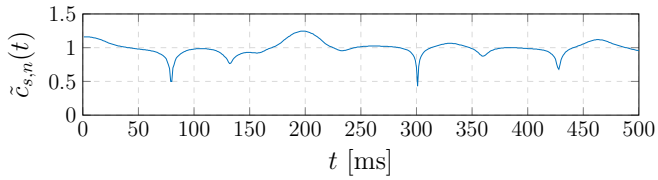


Fig. 5: Exemplary time series of synthesized fading amplitudes

IV. Validation

The primary objective of this model is to mimic the outage behavior that measurement data causes in link-level simulations. Therefore, a playback simulation with the data from Sec. II serves as an evaluation basis for the model. The results of the Physical Layer (PHY) simulations with synthesized channels from our model should then resemble those of the measured data. Before the results are compared, we briefly introduce the simulation framework.

A. Simulation Framework

Since IEEE Wireless Local Area Network (WLAN) is handled as a candidate for industrial communications [2], we use the *MathWorks WLAN Toolbox* to perform simulations with IEEE 802.11ac as PHY. Note that the particular choice of the PHY for validation is of subordinate importance as we are not interested in absolute performance results but in differences between measured and synthesized channels. The chosen PHY is based on Orthogonal Frequency Division Multiplexing (OFDM) and thus inherently reflects effects like the frequency selectivity of the channel. We assume that the validation of the model applies as well to other OFDM-based wideband PHYs such as 5G NR.

Table I gives an overview of the simulation parameters. The standard system bandwidth closest to the measured bandwidth is $B_{Sys} = 80\,\mathrm{MHz}$. The antenna configuration is Single-Input Single-Output (SISO) to match the measurement setup. We consider closed-loop remote control as a typical URLLC application, e.g. the control of an AGV. Therefore, we set the Modulation Coding Scheme (MCS) to 1, which equals a robust Quadrature Phase Shift Keying (QPSK) mapping with a code rate of $1/2$. We assume small control data packets of 300 Bytes to be sent periodically with a transmission interval of $1\,\mathrm{ms}$. The considered measurement scenario was repeated 10 times (200,000 measured CIRs). A playback simulation with 200,000 packets thus serves as a comparison for the synthesis.

TABLE I: Simulation Parameters

System Bandwidth	80 MHz	MCS	1
Antenna Config.	SISO	Payload Size	300 Bytes
Transmission Interval	1 ms	Packets	200 k

B. Validation Results

The proposed channel model includes mechanisms on different time scales, i.e., the DTMC on the large scale and synthetic fading on the small scale. To assess the suitability of each of these mechanisms, we compare simulations using the measured channels with three types of synthesis:

1) Uncorrelated Small-Scale Synthesis: The time series of the large-scale states originates from the measurements and synthetic fading is not temporally correlated, i.e., consecutive CIRs have no correlation (corresponding to $f_\mathrm{m} \to \infty$).

2) Correlated Small-Scale Synthesis: The series of states originates from the measurements and synthetic fading is temporally correlated. For all taps in the CIR, baseline Rayleigh fading is produced with Jakes spectrum and $f_\mathrm{m} = 12\,\mathrm{Hz}$.

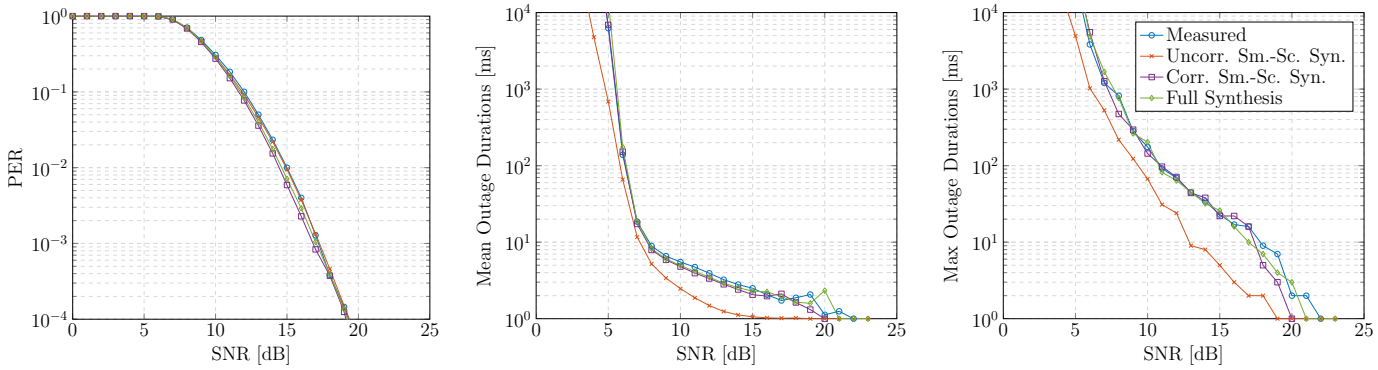
3) Full Synthesis: The time series of the large-scale states originates from the Markov process and synthetic fading is temporally correlated (as for *2)*).

As model-specific parameters, a segment length $T_\mathrm{Seg} = 40\,\mathrm{ms}$, an energy fraction $X = 0.99$ and a memory length of $M = 1$ proved to be suitable parameters for this environment.

As first, classical metric related to PHY investigations, the PER is evaluated. This metric serves as a rough assessment of whether measured and synthesized channels behave similar over a complete measurement round. The resulting PERs are shown in Fig. 6a. The good fit between synthesis and measurement data indicates that the model is meaningful and that parameters are well-chosen. For this metric, it is important that the slopes of the curves coincide for higher SNR values. However, the plot shows the limitations of this metric for assessing the modeling quality since the mechanisms producing temporal correlation seem to have no effect.

To assess the temporal outage behavior that the model causes, we consider mean and maximum outage durations per SNR. We expect a difference between the synthesis with and without temporal correlation which can be seen in Fig. 6b. The synthesis with uncorrelated fading causes outage durations being too short. The reason is the transmission interval of packets being lower than the expected coherence time T_Coh. By allowing the channel to change infinitely fast, bad channel conditions can simply disappear in the subsequent CIR. Thereby, consecutive packet errors become more unlikely. At the considered velocity and carrier frequency, a bad channel condition would affect multiple consecutive packets.

This effect is well represented by both syntheses including temporal correlation. The good match of the *Correlated Small-Scale Synthesis* and the *Full Synthesis* with the measured data indicates that the proposed mechanism is suitable. An industrial URLLC application usually has a tolerable outage duration that must not be exceeded to avoid a shutdown of the application [6]. For this reason, we consider it reasonable to take the maximum occurred outage duration into account, as shown in Fig. 6c. By looking at this metric, the capability of emulating small-scale fading can be equally confirmed.



(a) PER vs. Signal-to-Noise Ratio (SNR) (b) Mean outage duration vs. SNR (c) Maximum outage duration vs. SNR

Fig. 6: The model is validated by comparing the PER as well as outage durations between measured and synthesized channels.

Obviously, outage durations of $10^4\,\mathrm{ms}$ are not relevant for URLLC but presented for the sake of validation completeness.

While the synthetic fading is responsible for a realistic outage duration within a segment, the concatenation of correlated segments ensures realistic outages exceeding a segment length. The DTMC is capable of achieving this as the similarity between the *Full Synthesis* and the *Measured* curve indicates in Fig. 6b. Taking a more detailed look, the *Measured* curve shows two different slopes. The point of changing slopes falls together with the coherence time of about $10\,\mathrm{ms}$ to $20\,\mathrm{ms}$ on the y-axis. We conclude that the length of segments T_{Seg} must at least exceed the expected coherence time of the channel. If we choose segments to be too small, the modeled fading deviates from the measured data. Segments being too long carry the risk of changing large-scale properties within a segment. Thereby, the accuracy of the empirical distributions would suffer. In the scope of the investigations, different segment lengths were systematically tested and a segment length corresponding to half a wavelength,i.e., ($T_{\mathrm{Seg}} = 40\,\mathrm{ms}$), proved to provide the best match with the measured data.

V. Conclusion

This work was initiated by the question of how a data-driven channel model can be derived such that the temporal outage behavior of measurement data can be reproduced for URLLC-related PHY investigations. By using suitable metrics, e.g., the mean and maximum occurring outage duration, we show that the model works on both the large and the small scale. Even though transitions between large-scale segments are not modeled to happen smoothly, consecutive outages are nevertheless represented accurately. This is thanks to carefully choosing the segment length under consideration of the channel's coherence time. In the future, it will be interesting to study whether the model works similarly on data generated within other environments and with different transmission parameters. By artificially reducing the amount of data from which the distributions are derived, this model can also help to understand how much data is to be recorded in a new environment such that metrics like the reliability can be assessed. From the proposed modeling approach for SISO channels, we can extend our studies to model Multiple-Input Multiple-Output (MIMO) channels.

Acknowledgment

This work was supported by the German Federal Ministry of Education and Research (BMBF) as part of the project "AI4Mobile" under grant 16KIS1177.

References

[1] G. Aceto *et al.*, "A Survey on Information and Communication Technologies for Industry 4.0: State-of-the-Art, Taxonomies, Perspectives, and Challenges," *IEEE Communications Surveys Tutorials*, 2019.

[2] A. Frotzscher *et al.*, "Requirements and current solutions of wireless communication in industrial automation," in *2014 IEEE International Conference on Communications Workshops (ICC)*, 2014.

[3] T. Hößler *et al.*, "Mission Reliability for URLLC in Wireless Networks," *IEEE Communications Letters*, 2018.

[4] I. Muhammad *et al.*, "Mission Effective Capacity - a Novel Dependability Metric: A Study Case of Multi-Connectivity Enabled URLLC for IIoT," *IEEE Transactions on Industrial Informatics*, 2021.

[5] A. Traßl *et al.*, "On Dependability Metrics for Wireless Industrial Communications - Applied to IEEE 802.11ax," in *2019 IEEE 2nd 5G World Forum (5GWF)*, 2019.

[6] M. Gundall *et al.*, "5G as Enabler for Industrie 4.0 Use Cases: Challenges and Concepts," in *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2018.

[7] F. Burmeister *et al.*, "Measuring Time-Varying Industrial Radio Channels for D2D Communications on AGVs," in *IEEE Wireless Communications and Networking Conference*, 2021.

[8] C.-C. Chong *et al.*, "A novel wideband dynamic directional indoor channel model based on a Markov process," *IEEE Transactions on Wireless Communications*, 2005.

[9] S. Nie *et al.*, "A Three-Dimensional Time-Varying Channel Model for 5G Indoor Dual-Mobility Channels," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017.

[10] A. Traßl *et al.*, "Deriving an Empirical Channel Model for Wireless Industrial Indoor Communications," in *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2019.

[11] T. S. Rappaport, *Wireless Communications, Principles and Practice*. Prentice Hall, 2002.

[12] D. W. Matolak *et al.*, "Modeling wireless channel delay spread trends," in *2009 IEEE Radio and Wireless Symposium*.

[13] L. Devroye, *Non-Uniform Random Variate Generation*. Springer, New York, NY, 1986.

[14] M. Kashef *et al.*, "Clustering and Representation of Time-Varying Industrial Wireless Channel Measurements," in *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*.

[15] N. Beaulieu and C. Cheng, "Efficient nakagami-m fading channel simulation," *IEEE Transactions on Vehicular Technology*, 2005.