Realistic Performance Model for Vehicle-to-Infrastructure Communications

Veronika Shivaldova*†, Andreas Winkelbauer*, Christoph F. Mecklenbräuker*†

*Institute of Telecommunications, Vienna University of Technology
†Christian Doppler Laboratory for Wireless Technologies for Sustainable Mobility, Vienna University of Technology
Gusshausstrasse 25/389, 1040 Vienna, Austria
email: {veronika.shivaldova, andreas.winkelbauer, cfm}@nt.tuwien.ac.at

***Abstract*—This paper introduces extensions of a previously proposed range-dependent modified Gilbert model for generation of realistic error patterns. With the proposed extensions our model can be used to generate signal-to-noise ratio trends corresponding to the error patterns for infrastructure-to-vehicle communications. The model parameters are estimated based on realistic highway measurements using off-the-shelf IEEE 802.11p devices. Finally, a scalar quantization-based framework for reduction of the model complexity is introduced and evaluated.**

I. Introduction

A significant component in the development stage of all communication systems, including vehicular communications, are simulators. The majority of simulators used in the vehicular communication's domain are build to analyze either physical or network layer, but rarely all layers together. The physical layer simulators are mainly developed to evaluate the influence of the signal processing algorithms and the transceiver design on the system performance with (often very complex) propagation channels close to realistic. The physical layer simulators however, are rarely build to investigate realistic, in terms of network density, scenarios, e.g., to model interference caused by several collocated transmissions. This aspects are rather addressed by the network layer simulators, in which physical layer is often abstracted and extremely simplified [1]. However, results of real-world measurements show that deterministic radio propagation models, typically assuming exponential path loss and omnidirectional signal propagation, should be avoided, since they do not capture such significant realistic effects as small scale fading and shadowing. These effects considerably influence transmission range and packet error rates of vehicular communication systems, especially on intersections [2], [3]. Therefore, realistic vehicular channel models, which are capable of capturing effects like small scale fading and shadowing, are crucial for improving the quality of upper layer simulators.

In order to provide accurate representation of physical layer and yet keep computational effort within manageable dimensions, stochastic models describing the wireless channel characteristics from a macroscopic point of view can be used. In this context Markov chains are known to be a powerful and commonly used tool for modeling the error statistics of communication channels. They consists of a finite number of states and corresponding state transition probabilities. The samples produced by the Markov chain depend deterministically on the state. If, however, the samples depend on the states probabilistically, the states are hidden since they cannot be directly observed from the samples. Such models are called hidden Markov models (HMMs). The availability of efficient algorithms for the estimation of HMM parameters from experimental data makes their application in the field of vehicular communications simple and very beneficial. Given a set of suitable HMMs, performance analysis can be carried out on a computer which is cheap compared to highly expensive field tests. Moreover applying HMMs to model and reproduce real-world performance studies would create a valuable cross-layer interface between network layer and physical layer simulators, substantially improving the usefulness of existing simulators such as NS-3 and QualNet.

In this paper we present a novel, computationally inexpensive approach to model packet errors and associated signal-to-noise ratio (SNR) for infrastructure-to-vehicle (V2I) communications. The proposed model incorporates realistic physical layer characteristics, such as standard-compliant transmitter and receiver equipment, vehicle velocity, as well as authentic highway environment and traffic conditions. We suggest an nearly lossless algorithm for reduction of the model complexity. Given a limited set of parameters provided in this contribution, realistic packet error and SNR traces can easily be produced and used as an alternative to unreliable physical layer abstractions of existing simulation tools.

II. Modeling Approach

For modeling the SNR and the packet error patterns of V2I link we suggest to use a simple first-order HMM with two states, the good and the bad. In this model, originally introduced by Gilbert [4], the good state is always error-free, i.e., $P_E = 0$, while in the bad state an error can occur with an emission probability $P_E > 0$. Therefore, the produced samples depend on the states probabilistically and the sequence of states cannot be directly observed from the sequence of samples.

Assuming that in time instance n the model is in the bad state, a biased coin is tossed to produce the n^{th} observed error pattern digit, which could be either 1 (with probability

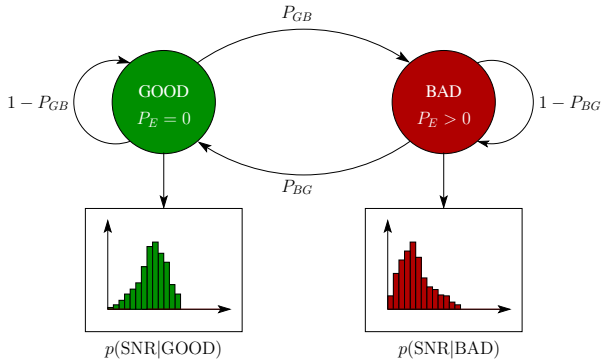

Fig. 1: Schematic illustration of the extended Gilbert model.

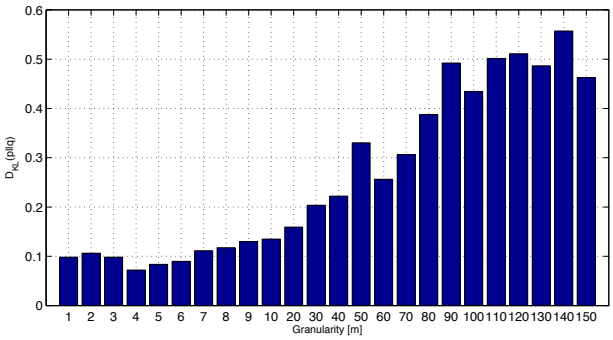

Fig. 2: KL divergence between measured and modeled PDR distributions for granularities range $[1, \ldots, 150]$ m.

P_E) representing a packet error event or 0 (with probability $1-P_E$) representing a successful packet transmission. After a digit in the bad state is produced, a state transition governed by the transition probability P_{BG} takes place and the model either changes from the bad to the good state or remains in the bad state. After the state transition the coin is tossed again to produce the $(n+1)^{th}$ error pattern digit according to the error emission probability of the current state. Thus, given the current state, the corresponding error pattern digit is independent of all previous digits.

The Gilbert model originally designed to generate bursty error patterns can be further extended to produce associated SNR traces. Therefor, we need to know the probability of observing a given SNR value in each of the states, which are hidden initially. However, once the transition probabilities P_{BG}, P_{GB} are estimated from measured error patterns, the Viterbi algorithm can be used to recursively determine the most likely sequence of states. by doing so, each measured SNR trace can be directly linked to the associated sequence of states. With this extension, at any time instance n our model produces two emissions, as shown in Fig. 1, i.e., one error pattern digit driven by the probability P_E and one SNR value, produced according to the conditional probability distribution of SNR given the state $p(SNR|GOOD)$, $p(SNR|BAD)$.

Since the performance of V2I communication link strongly depends on the distance between the transmitter and the receiver, a simple two-state model clearly won't be able to accurately represent the realistic propagation effects. Therefore, for estimating the parameters of our model each measurement is divided into N parts corresponding to N disjoint distance intervals of the same length. This interval length will be referred to as granularity hereafter. The transition probabilities (P_{BG}, P_{GB}) and the emission probabilities (P_E, $p(SNR|GOOD)$, $p(SNR|BAD)$) are estimated for each interval separately using the Baum-Welch algorithm [5]. Once the model parameters for all N intervals are estimated, we can combine them to form a range-dependent modified Gilbert model [6]. This model retains all properties of the original Gilbert model, except for the fact that the model parameters change as soon as the vehicle leaves the current interval. The initial model state is randomly chosen, while for all subsequent intervals the initial state is equal to the final state in the previous interval.

Clearly, the choice of the granularity, i.e., the length of the interval in which the model parameters do not change, is essential for the accuracy of our range-dependent modified Gilbert model. To find the optimal granularity, we first convert each error pattern into a packet delivery ratio (PDR).The PDR is defined as the number of error-free packets divided by the number of detection events in a time interval $T = \Delta d/v$. To calculate the PDR as a function of the distance, we compute a moving average of the corresponding error pattern where we set Δd to 10 m and v is the velocity of the test vehicle, which we obtain from GPS data. We note that the PDR values are real numbers between 0 and 1, whereas the error pattern consists only of zeros and ones. Next, we estimate the probability distribution of the measured and modeled PDR values using histograms with L bins, yielding L probabilities p_i and q_i, $i = 1, 2, \ldots, L$, respectively. Finally, we compute the distance (from an information theoretical perspective) between these probability distributions in terms of the Kullback-Leiber (KL) divergence. The KL divergence is calculated as follows:

$$D_{KL}(p||q) = \sum_i p_i \log_2 \frac{p_i}{q_i}$$

Loosely speaking, the KL divergence $D_{KL}(p||q)$ measures the information loss when using the approximate distribution q rather than the true distribution p. We note that $D_{KL}(p||q)$ is always nonnegative and $D_{KL}(p||q) = 0$ if and only if $p = q$. The KL divergence between the measured and modeled PDR distributions with granularities from 1 m to 150 m is shown in Fig. 2. As expected, a better approximation of the measurement can be achieved by estimating the model parameters with smaller granularities.

III. Complexity Reduction

In previous section, we have shown that the model accuracy can be significantly improved by using smaller granularities for estimation of the model parameters. This however, leads to considerable increase in the number of intervals used for

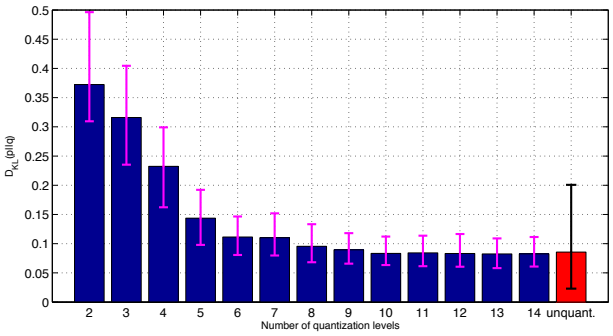

Fig. 3: Average KL divergence between measured and modeled with granularity 1 m PDR distributions. The red and the blue bars show performance of the model with unquantized and quantized parameters, respectively. The error-bars show 95 % confidence intervals of the KL divergence.

parameter estimation and eventually results in a large number of model parameters and thus additional computational complexity. In order to ensure high accuracy of the model while keeping the number of model parameters low, we suggest to use clustering. For this purpose all estimated probabilities are divided into nonintersecting subsets, called clusters, based on some similarity measure. As the similarity measure for clustering, we suggest to use the mean square error (MSE). The partitioning into clusters is then performed such, that an arbitrary probability $P \in \mathcal{C}^k$ is closer (in MSE sense) to any probability of the same cluster $\mathcal{C}^k$ than to any probability from the other clusters $\mathcal{C}^{\sim k}$.

To cluster the probabilities of our model we use the Lloyd's algorithm (also known as K-means algorithm) [7]. This iterative algorithm is applied to each probability set $\{P^1_{GB}, P^2_{GB}, \ldots, P^N_{GB}\}$, $\{P^1_{BG}, P^2_{BG}, \ldots, P^N_{BG}\}$, $\{P^1_{E}, P^2_{E}, \ldots, P^N_{E}\}$ separately, to create K clusters with K respective representatives for each probability set. The clustering procedure consists of the following steps:

1) **Initialization**: The iteration count i is set to 0 and the initial K random numbers $\{\mathbf{c}^1_0, \mathbf{c}^2_0, \ldots, \mathbf{c}^K_0\}$ are selected as representatives, or centroids.
2) **Assignment:** The MSE $d(P^n, \mathbf{c}^k_i)$ is calculated between each probability in the set $\{P^1, P^2, \ldots, P^N\}$ and each centroid $\{\mathbf{c}^1_0, \mathbf{c}^2_0, \ldots, \mathbf{c}^K_0\}$. After which the probabilities are assigned to the cluster $\mathcal{C}^k_{i+1}$ with the smallest MSE. In the other words
$$P^n \in \mathcal{C}^k_{i+1} \text{ if and only if } \mathbf{c}^k_i = \arg\min_j d(P^n, \mathbf{c}^j_i)$$
3) **Update:** The centroids $\mathbf{c}^K_{i+1}$ are found as an average over all probabilities in the cluster $\mathcal{C}^k_{i+1}$:
$$\mathbf{c}^k_{i+1} = \frac{1}{\gamma^k_{i+1}} \sum_{P^n \in \mathcal{C}^k_{i+1}} P^n,$$
where γ^k_{i+1} is the number of probabilities that belong to the k^{th} cluster at the step $i+1$.

The assignment and the update steps are repeated as long as new centroids can be found. The resulting centroids are then used instead of the initial probabilities.

In contrast to the transition probabilities and the emission probability for error pattern, the emission probabilities for SNR ($p(SNR|GOOD)$, $p(SNR|BAD)$) are not scalars and therefore, cannot be clustered by the Lloyd's algorithm. For this reason we use the LBG algorithm [8], which in fact is very similar to the Lloyd's algorithm with the minor difference that all involved quantities are vector-valued. The new centroid-vectors $\{\underline{\mathbf{c}}^1_{i+1}, \underline{\mathbf{c}}^2_{i+1}, \ldots, \underline{\mathbf{c}}^K_{i+1}\}$ are given by the mean (in each vector component) of all probability vectors in the same cluster.

Although both the Lloyd's and the LBG algorithms are guaranteed to converge against a local minimum, the resulting centroids are not guaranteed to yield the global minimum, i.e., there might exist an other set of centroids leading to even smaller overall MSE. Hence, the performance of the used clustering algorithms strictly depends on the initial choice of the centroids. For this reason, in order to choose K representatives for each set of the model parameters, we randomly choose 1000 initial centroids and perform the clustering for each of them. The set of centroids yielding the minimum MSE is then chosen as the optimal representative. Since the centroids for the initialization procedure are chosen randomly, it is necessary to make sure that each initial centroid has minimum MSE to at least two probabilities from the set $\{P^1, P^2, \ldots, P^N\}$. If this condition does not hold, the initial centroids are discarded and the new centroids are randomly chosen. Clearly, the clustering is only performed with the valid initial centroids. We further note, that the randomly chosen elements of the initial centroid-vectors $\{\underline{\mathbf{c}}^1_0, \underline{\mathbf{c}}^2_0, \ldots, \underline{\mathbf{c}}^K_0\}$ are normalized, such that the sum of all elements in a vector is 1.

Fig. 3 shows the dependence of the model accuracy on the number of quantization levels K. Here the bars show the KL divergence calculated between the measured and the modeled PDR values, where the model parameters were estimated with the granularity of 1 m, to yield the highest accuracy. The red bar represents an average KL divergence between the measurement and the model realizations, with unquantized model parameters, i.e., where all N model parameters are used for the data generation (cf. 1^{st} bar of Fig. 2). The blue bars show the difference (in terms of KL divergence) between the measurement and the realizations of the model, parameters of which were replaced by $K < N$ representatives. As expected, the accuracy of the model is the higher, the more values we choose as the representatives of the model parameters. However, starting with $K = 10$ quantization levels, the performance of the model with quantized and unquantized parameters is nearly equal and further increase in the number of quantization levels K yield only negligible improvements. Therefore, we conclude that parameters of our model estimated with the granularity of 1 m can be sufficiently well represented by a new set consisting of only 10 probabilities. Thereby, the complexity of the model is reduced by the factor of 100 without a significant degradation of the model accuracy.

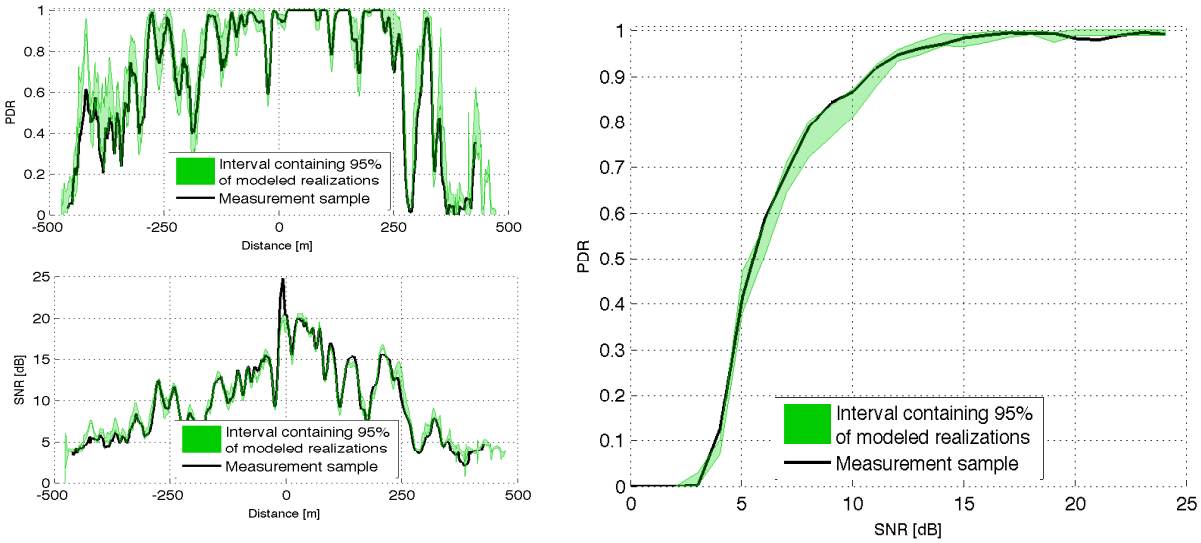

Fig. 4: PDR and SNR performance comparison. Black solid line represents an average measured performance, while the green bands show the the intervals containing $95\dot{\%}$ of all model realizations. Parameters of the corresponding model were estimated with granularity $1\,\mathrm{m}$ and subsequently replaced by 10 representatives.

All KL divergence values shown in Fig. 3 show an average difference between the measurement and 1000 realization of the model. The magenta and black error-bars show the $95\,\%$ confidence intervals of KL divergence, for the model with quantized and unquantized parameters, respectively. We clearly see, that the confidence intervals shrink with the increasing number of quantization levels. In fact the confidence intervals for the model with quantized parameters are even tighter than for the unquantized model. Therefore, we conclude that the model parameters estimated with the granularity of $1\,\mathrm{m}$ can be optimally represented by 10 quantization levels, without any significant performance losses[1]. The optimal set of model parameters is given in Tab. I.

IV. Results

To demonstrate the high precision and accuracy of our model, in this section we compare the measured PDR and SNR performance with the performance resulting from 1000 realizations of our model with a granularity of $1\,\mathrm{m}$ and number of quantization levels $K = 10$. Plots on the left hand side of Fig. 4 shows the corresponding results in terms of PDR and SNR curves. The curves are plotted versus the distance d, where the origin of the abscissa ($d = 0\,\mathrm{m}$) corresponds to the position of the transmitter. Negative values on the abscissa correspond to locations where the vehicle was approaching the RSU and positive distances represent the vehicle locations after passing the transmitter. The solid lines show the average PDR and SNR of the measured traces. The green bands represent the intervals containing $95\dot{\%}$ of all model realizations.

[1]Carefully note, that the optimal number of quantization levels depends on the granularity, i.e., for larger granularities the model parameters can be sufficiently well represented by $K < 10$ quantization levels

Here both the PDR and the SNR as functions of the distance, are computed as a moving average of the corresponding error pattern and single SNR values, respectively. Each PDR is calculated as the number of error-free packets divided by the number of detection events occurred within the distance $\Delta d_{PDR} = 10\,\mathrm{m}$. The average SNR is calculated as a mean of the single SNR values within the interval $\Delta d_{SNR} = 1\,\mathrm{m}$. The averaging window size $\Delta d_{PDR} > \Delta d_{SNR}$, since for reliable PDR calculation the number of binary error pattern digits should be large, while averaging over the large number of SNR values would lead to undesired performance smooting effect.

Black solid line on the right hand side of Fig. 4 shows the average dependence of the PDR on the SNR derived from the measurements. The green band again represents the interval containing $95\dot{\%}$ of all model realizations. To compute this dependence, we averaged all PDR values obtained for the given SNR value.

In all plots of Fig. 4 the intervals containing $95\dot{\%}$ of model realization are very tight. Therefore we conclude, that all realizations of our range-dependent modified Gilbert model with granularity of $1\,\mathrm{m}$ and 10 quantization levels can reproduce the measurement result with high precision.

V. Conclusion

In this contribution a computationally low-cost range-dependent modified Gilbert model that allows to accurately reproduce the packet error statistics and the associated SNR performance of the real-world wireless V2I links was introduced. The model parameters have been estimated using data acquired during an extensive IEEE 802.11p V2I field-testing campaign. The influence of the model granularity was

TABLE I: Parameters of the range-dependent modified Gilbert model that were estimated with granularity 1 m and subsequently replaced by 10 representatives.

	$\mathbf{P_{GB}}$	$\mathbf{P_{BG}}$	$\mathbf{P_E}$	$\mathbf{p(SNR|GOOD) = \{\rho_{0dB}, \rho_{1dB}, \cdots \rho_{29dB}\}}$	$\mathbf{p(SNR|BAD) = \{\rho_{0dB}, \rho_{1dB}, \cdots \rho_{19dB}\}}$
Level 1	0.007	0.006	0.001	0 0 0 0 0.004 0.021 0.052 0.081 0.122 0.188 0.231 0.156 0.083 0.032 0.011 0.005 0.007 0.005 0 0 0 0 0 0 0 0 0 0 0	0 0.001 0.006 0.024 0.123 0.242 0.394 0.157 0.044 0.008 0 0.001 0 0 0 0 0 0 0
Level 2	0.035	0.044	0.198	0 0 0 0 0 0.002 0.007 0.017 0.031 0.041 0.047 0.055 0.090 0.171 0.250 0.201 0.068 0.013 0.005 0.002 0.002 0 0 0 0 0 0 0 0	0 0.003 0.009 0.021 0.030 0.063 0.039 0.071 0.113 0.156 0.190 0.131 0.068 0.054 0.027 0.012 0.007 0.004 0.001
Level 3	0.082	0.088	0.402	0 0 0 0.001 0.008 0.033 0.088 0.186 0.275 0.252 0.101 0.024 0.007 0.006 0.005 0.002 0.004 0.005 0.003 0 0 0 0 0 0 0 0 0 0	0.001 0.009 0.061 0.227 0.344 0.258 0.077 0.017 0.003 0.001 0 0 0 0 0 0 0 0 0
Level 4	0.145	0.144	0.539	0 0 0 0 0.001 0.003 0.008 0.012 0.015 0.020 0.030 0.041 0.059 0.095 0.129 0.152 0.196 0.136 0.063 0.021 0.011 0.004 0.003 0.002 0 0 0 0 0	0 0 0 0.002 0.010 0.056 0.149 0.225 0.407 0.086 0.035 0.019 0 0 0.01 0 0 0 0
Level 5	0.243	0.230	0.625	0 0 0 0.003 0.018 0.085 0.230 0.312 0.236 0.079 0.017 0.006 0.002 0.004 0.001 0.002 0.001 0.002 0.002 0 0 0 0 0 0 0 0 0 0	0 0.032 0.151 0.322 0.347 0.107 0.025 0.005 0.001 0 0 0 0 0 0.001 0.007 0.002 0 0
Level 6	0.373	0.323	0.706	0 0 0 0 0.003 0.011 0.027 0.045 0.061 0.078 0.107 0.143 0.181 0.163 0.096 0.043 0.021 0.011 0.006 0.003 0.001 0 0 0 0 0 0 0 0	0 0.002 0.021 0.042 0.063 0.122 0.208 0.203 0.150 0.101 0.042 0.021 0.016 0.002 0.003 0.002 0.001 0 0
Level 7	0.524	0.472	0.792	0 0.001 0.013 0.077 0.266 0.410 0.204 0.029 0.001 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.004 0.027 0.195 0.495 0.143 0.084 0.025 0.016 0.003 0 0 0 0 0 0.003 0.004 0.001 0 0
Level 8	0.703	0.657	0.868	0 0 0 0 0 0 0 0 0 0 0.001 0.002 0.003 0.006 0.007 0.011 0.019 0.024 0.049 0.083 0.131 0.175 0.152 0.114 0.080 0.044 0.018 0.016 0.018 0.011	0.001 0.009 0.056 0.162 0.233 0.242 0.185 0.083 0.021 0.004 0.001 0 0 0 0.001 0 0 0 0
Level 9	0.826	0.910	0.942	0 0 0.002 0.007 0.074 0.240 0.450 0.192 0.033 0.002 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.001 0.004 0.025 0.096 0.230 0.389 0.191 0.044 0.012 0.002 0 0.007 0 0 0 0 0 0 0
Level 10	1	1	1	0 0 0 0 0 0.001 0.002 0.002 0.005 0.014 0.028 0.044 0.035 0.032 0.048 0.060 0.090 0.159 0.212 0.168 0.050 0.019 0.012 0.008 0.006 0.001 0 0.001 0	0 0 0.011 0.067 0.016 0.025 0.150 0.533 0.133 0.054 0.011 0 0 0 0 0 0 0 0

analyzed and it was shown that the accuracy of the model is significantly improved when using smaller granularity. Smaller granularities however, lead to larger number of model parameters and thus, boost the computational complexity. To reduce the model complexity, we proposed to use simple clustering algorithms. It has been shown, that parameters of the model with granularity of 1 m can be replaced by only 10 representatives, with only negligible accuracy loss. This simplification results in the complexity reduction by the factor of 100.

ACKNOWLEDGMENT

This work was performed with partial support by the Christian Doppler Laboratory for Wireless Technologies for Sustainable Mobility and the ROAD-SAFE project, a scientific cooperation between FTW, TU Wien, ASFINAG Maut Service GmbH, Kapsch TrafficCom AG and Fluidtime GmbH. We acknowledge the Federal Ministry for Transport, Innovation, and Technology of Austria (BMVIT) for granting a test license in the 5.9 GHz band. We further appreciate support of COST Action IC1004 on cooperative radio communications for green smart environments.

REFERENCES

[1] J. Mittag, S. Papanastasiou, H. Hartenstein, and E. Ström, "Enabling Accurate Cross-Layer PHY/MAC/NET Simulation Studies of Vehicular Communication Networks," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1311 –1326, July 2011.

[2] J. Otto, F. Bustamante, and R. Berry, "Down the Block and Around the Corner The Impact of Radio Propagation on Inter-vehicle Wireless Communication," in *29th IEEE International Conference on Distributed Computing Systems, 2009*, June 2009, pp. 605 –614.

[3] E. Giordano, R. Frank, A. Ghosh, G. Pau, and M. Gerla, "Two Ray or not Two Ray this is the price to pay," in *6th IEEE International Conference on Mobile Adhoc and Sensor Systems, 2009*, October 2009, pp. 603 –608.

[4] E. N. Gilbert, "Capacity of a burst-noise Channel," *Bell System Technical Journal*, vol. 39, no. 9, pp. 1253–1265, 1960.

[5] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *The Annals of Mathematical Statistics*, pp. 164–171, 1970.

[6] V. Shivaldova, A. Winkelbauer, and C. Mecklenbräuker, "Vehicular Link Performance: From Real-World Experiments to Reliability Models and Performance Analysis," *IEEE Vehicular Technology Magazine*, vol. 8, no. 4, pp. 35–44, 2013.

[7] S. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[8] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *Communications, IEEE Transactions on*, vol. 28, no. 1, pp. 84–95, 1980.