# Methods for Extracting V2V Propagation Models from Imperfect RSSI Field Data

Silvija Kokalj-Filipovic      Larry Greenstein
Bin Cheng      Marco Gruteser
WINLAB, Rutgers University
{*skokalj,ljg,cb3974,gruteser*}*@winlab.rutgers.edu*

*Abstract*—We describe three in-field data collection efforts yielding a large database of RSSI values vs. time or distance from vehicles communicating with each other via DSRC. We show several data processing schemes we have devised to develop opportunistic Vehicle-to-Vehicle (V2V) propagation models from such data. The database is limited in several important ways, not least, the presence of a high noise floor that limits the distance over which good modeling is feasible. Another is the presence of interference from multiple active transmitters. Our methodology makes it possible to obtain, despite these limitations, accurate models of median path loss vs. distance, shadow fading, and fast fading caused by multipath. We aim not to develop a new V2V model, but to show the methods enabling such a model to be obtained from in-field RSSI data, without elaborate measurement design and the associated deployment cost. Finally, models based on field data allow for capturing the multiple effects of an increasing number of simultaneous V2V transceivers under typical extreme traffic scenarios.

## I. Introduction

There has been considerable research to develop and calibrate a Physical (PHY) layer model for Vehicle to Vehicle (V2V) Dedicated Short Range Communications (DSRC) [1]. Such a model can be used to simulate large-scale V2V networks with statistical accuracy, thus allowing studies of V2V enabled driver safety applications. The PHY model consists of characterizing both the channel (propagation) and the receiver. Our focus is on the channel, specifically, modeling the received power as a function of the distance between the transmitter (TX) and the receiver (RX). The particular challenge we address is to develop the channel path loss vs. distance based on the received signal strength (RSSI) data obtained from field experiments, which is an indirect and frequently corrupted measure of the channel quality. While a more accurate picture would come from channel sounding methods, e.g., [2], such an approach is not cost-effective, as it requires expensive and sensitive equipment for each participating vehicle. We note that a database consisting of RSSI samples is, by its very nature, not ideally suited to propagation modeling, for several reasons. One is that RSSI measures total power in the full system bandwidth (no frequency subdivision is done), leading to large noise floors. Another is that RSSI is reported only for packets sufficiently clean to be accurately received. Thus many packets are lost due to noise (especially at larger distances), and also due to interference, in trials with many simultaneously transmitting terminals. However, RSSI sampling is a form of participatory sensing of channel conditions incurring no cost,

as it is a by-product of the primary task, which is in our case testing of V2V safety applications by the Crash Avoidance Metrics Partnership (CAMP) Vehicle Safety Communications 3 (VSC3) Consortium, and the USDOT. If we were to design interference-free RSSI collection campaigns (i.e., with one active transmitter), while covering diverse topologies, vehicles and channel realizations, the deployment costs would be prohibitive. The aim of this paper is not to present a new V2V model, of which there are many, e.g., [3]–[5]. It is to showcase methods that can be used to extract useful models from imperfect data: the RSSI samples collected in typical environments and under dense traffic conditions, with vehicles frequently transmitting simultaneously. The paper describes our experimental approach, the problems with RSSI data, and the methods we have developed for working around these problems.

## II. Problem Description

Our RSSI measurements are made on OFDM transmissions with a 10MHz bandwidth centered near 5.9 GHz, in compliance with V2V DSRC standard IEEE802.11p [6], using Atheros 802.11p chips. A particular contribution differentiating this work from other papers on RSSI-based models [7] is that we tackle RSSI properties due to high density of transmitting vehicles, i.e., in *dense scenarios.*. Apart from the *vehicular density* $\mathcal{V}$, the model should also account for *transmitter density* $\mathcal{T}$. Here, field trials are marked by a certain scale of participating DSRC transceivers, ranging from $\mathcal{T} = 50$ to $400$ radios broadcasting fixed size messages at a constant rate of $\mathcal{R}$ Hz (messages per second), which we refer to as *message density*. Almost every experiment of a certain scale was conducted for both $\mathcal{R} = 5$ and $10$ Hz. Note that higher $\mathcal{R}$ and $\mathcal{T}$ contribute to increasing interference which results in lost packets both due to collisions and lower SINR. Also, increasing $\mathcal{T}$ changes the median propagation model due to additional scatter. These are the factors that need to be decoupled from the losses due to channel propagation in sparse scenarios (signal attenuation due to large distances or fading effects). In the interest of space, we assume that $\mathcal{R}$ is constant.

Three distinct field trials are presented here to demonstrate our methods. One is referred to as the Zero Reference (ZR) trial, which included only one TX at a time, mounted on a vehicle that traveled back and forth on a straight road 1400
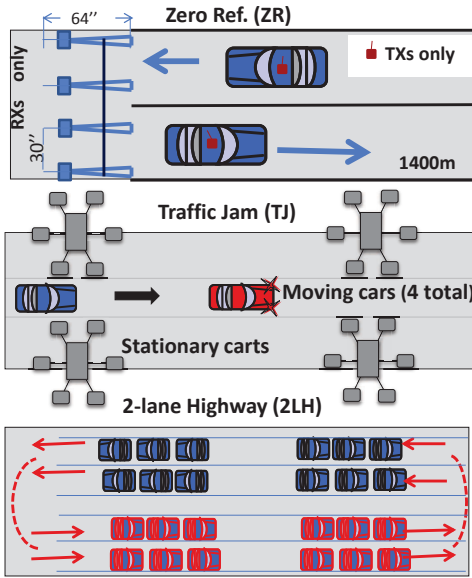
Fig. 1. Scenarios for the three trials. Top: Zero Reference (ZR); Center: Traffic Jam (TJ); Bottom: Two-Lane Highway (2LH).



Fig. 2. Path loss vs. log-distance for ZR data, with $P_t = 18dBm$ (top) and $26dBm$ (bottom); for single TX higher power should be used to minimize losses over the distance range of interest



Fig. 3. Effects of interference on TJ RSSI – Top: RSSI vs. log-d, where interference accounts for the large spread at low distances. Bottom: Interference-caused fading-like fluctuations on a static link between adjacent OBEs.

m long (see Fig. 1, top). At one end of the road there were 4 pods with antennas spread 30 inches apart from each other. They logged RSSI values for packets sent from the TX and tagged by the distance from the TX. The ZR trial is a sparse scenario (low $\mathcal{V}$), and the one with no interference (i.e. low $\mathcal{T}$), enabling us to ignore the influence of interference and scatter. While the absence of other vehicles changes the propagation environment from that of a dense scenario (i.e., less scatter), this scenario helped us assess the impact of receiver noise-floor losses on path loss modeling using RSSI data (e..g., Fig. 2). The 2nd trial, another series of measurements on flat surface with line-of-site and open sky, helped us to study the effects of interference. This trial consisted of four cars moving in a single lane past 66 carts placed along a roadway (see Fig. 1, center). Each cart contains several OBEs (On-Board Equipments, i.e, DSRC transceivers); collectively, 400 cart-mounted OBEs emulated the interference from a typically dense spatial distribution of vehicles (high $\mathcal{T}$). The carts were stationary and distributed along 1400 m of roadway. As the vehicles were passing along a dense deployment of 400 stationary TXs, akin to what might occur in certain real-world traffic jam scenarios, we refer to it as Traffic Jam (TJ) trial. The absence of metallic surfaces–barring that of the four moving vehicles–resulted in reduced scatter, hence, a low $\mathcal{V}$, just as in the ZR trial. In both cases, the scatter plots of RSSI vs. distance $d$ take on the appearance of 2-ray propagation, as presented in Fig. 2 and Fig. 3-top. Note, however, that the typical 2-ray shape is harder to observe in Fig. 3-top, as the interference-induced RSSI fading smears the scatter-plot. The third trial modeled many vehicles moving under free-flow conditions along a 2-lane highway (2LH) in both directions, and turning repeatedly at the end of the lane to assume the opposite bearing (see Fig. 1, bottom). Here, the scatter appeared to be quite strong, leading to RSSI-distance scatter plots that are cloud-like, as in Fig. 4. The data from
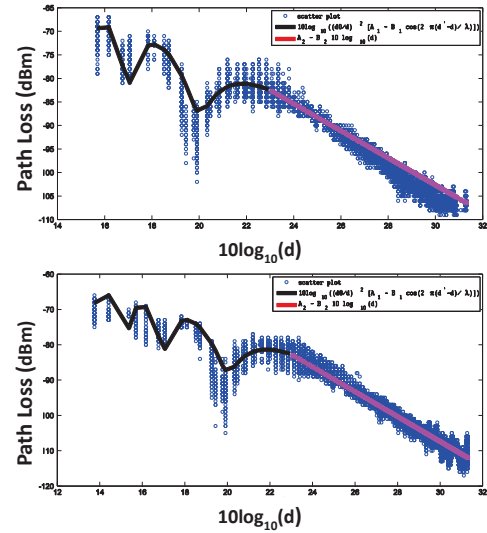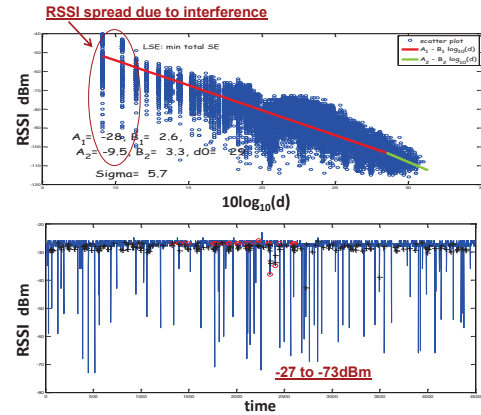
this scenario helped us to model the effects of scatter on path loss, to develop methods for modeling fast fading (FF), and to further explore the mitigation of noise floor effects.

## III. ISSUES WITH RSSI DATA

### A. Major Challenges

Here, we qualitatively characterize the issues in channel modeling with RSSI data and discuss solution approaches. We elaborate upon the latter in Section 4.

**Issue 1: Noise saturation of RSSI data at larger distances** Noise saturation occurs when the RX power is so small that noise dominates (low SNR), and most packets are lost. A clear example is given in Fig. 4, a scatter plot of recorded RSSI values and path distance in the 2LH trial. The trend with distance is downward, as expected, and a dense, fairly uniform cloud of points like this is well-fitted by a straight-line segment, or perhaps two segments with different slopes in different distance ranges. However, as seen, there is an RSSI level below which few if any points are recorded. It
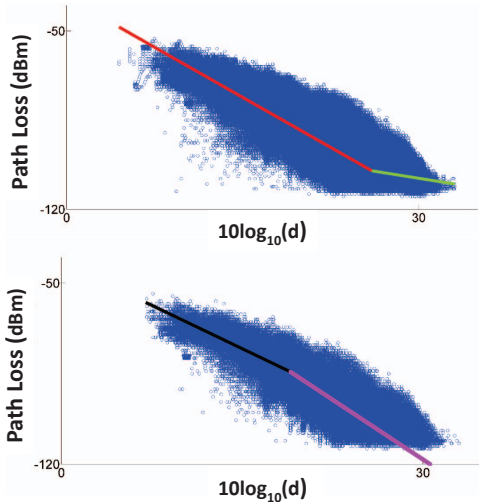
Fig. 4. Scatter plot of path loss vs. log-distance for 2LH data, showing 2-segment fits for the median path loss. Top: The second segment is made shallow by the noise floor. Bottom: For the same scatter plot, the noise floor effect is mitigated by additional processing.
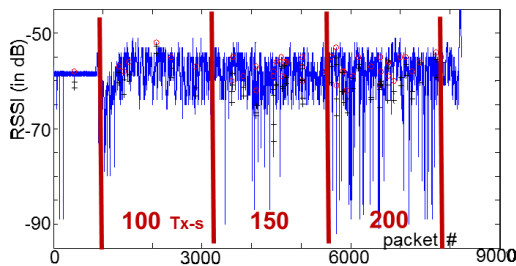


Fig. 5. Time plots for the 2LH trial, showing the effect of different numbers of active TXs. The appearance of what seem like deep fades intensifies as the number increases from 100 to 150 to 200.

corresponds to the RX noise power, -96 dBm in this case, and it creates a flat bottom edge of the scatter plot. Using least square estimates (LSE) to create a 2-segment fit, we obtain the fit in Fig. 4, top. The shallow slope of the 2nd segment is an artifact of the noise floor which, if removed or worked around in the data processing, produces the result of Fig. 4, bottom. The steeper slope of the higher-distance segment is more consistent with realistic propagation [8], [9]. We shall elaborate on the methods used to achieve this improvement. **Issue 2: False fading due to interference** Fading (large and small scale) captures how the instantaneous signal level fluctuates over time, frequency, and space due to mobility and scattering. We present evidence that interference is included in the RSSI measurements, which takes the appearance of sudden fluctuations in the signal level. For large number of TXs, the fluctuations resemble FF due to the effect of summing random phases of the interferers. Under the same conditions except the number of active TXs, these fluctuations are different. In a 2LH trial run with 200 vehicles on the ground, the test first included 100 active TXs; then another 50 were added, and in the final third of the test all 200 vehicles were transmitting. The RSSI vs. time plot for a particular TX-RX pair in this test is presented in Fig. 5, where the three

parts of the trial are separated by vertical bars. Clearly, the fluctuation process intensifies with the number of active TXs. The interference is also to blame for unusually high RSSI spread at small distances in a Traffic Jam trial run (Fig. 3, top), and the appearance of deep fades for a static link in a stationary field test (Fig. 3, bottom). A particularly difficult challenge was to decouple interference effects from real fades in modeling small scale fading. For static nodes, it is easier to remove outliers (more in the next section) since we know that there should be no FF, and whatever appears to be a sudden deep fade (e.g., see Fig. 3, bottom) is to be ignored. For mobile links, we used the ZR logs as the no-interference baseline to gauge the contribution of interference to RSSI fluctuations. In [10] we describe a more elaborate technique based on density estimation. **Issue 3: Missing points due to interference** Packets are lost not only when signal levels fall below noise; they can be lost when signal levels are good, due to interference. This is illustrated by Fig. 6, top, which shows, for the Traffic Jam trial, the number of recovered packets (red curve) and the number of packets lost (blue rectangles), both as functions of distance. The transmit (TX) power was 20 dBm. We see that severe losses occur even at distances where the signal should be strong compared to RX noise; note the log scale of the distance causing an illusion of more packets at higher distances. We illustrate that these losses are due to interference by comparing them with Fig. 6, bottom, which shows results for a similar (mobile link) run in the ZR trial. As opposed to Fig. 6, top, there are no losses at the lower distances, a result of having no interference. The other change, related to the onset of packet losses, is discussed next.

### B. Handling the Noise Floor (Censored Data)

One method for mitigating the influence of RX noise is to increase the TX power. In the ZR trial leading to Fig. 6, bottom, the TX power was 26 dB, which is 6 dB above that used in the Traffic Jam trials leading to Fig. 6, top. As seen in the bottom plot, the onset of the noise saturation begins at higher distances and is a lot milder. Another example is given in Fig. 2, where data from the ZR trials are collected for two TX powers: 18 dBm (top) and 26 dBm (bottom). No bottoming out of the scatter plot at large distances is evident on the bottom. The fitting of the straight-line segment is therefore better: the RMS variation about it is 1.6 dB, in contrast to 2.3 dB using a TX power of 18 dBm. The second method is by binning the points over the distance range, and attempting to model the RSSI points within each bin as a Gauss-like distribution, based on the mode and histogram. The expectation was that if we identify several bins wherein RSSI has a clean uni-modal distribution, we may infer the missing points based on assuming symmetry of the distribution. We refer to this binning approach as the mode-fitting method. Because of the lost points due to interference, the bin-based distributions of RSSI are not always uni-modal. This can be mitigated by first separating the interference [10] and then applying mode-fitting. Also, at and beyond some point in distance, the mode is likely to fall below the noise threshold, thus
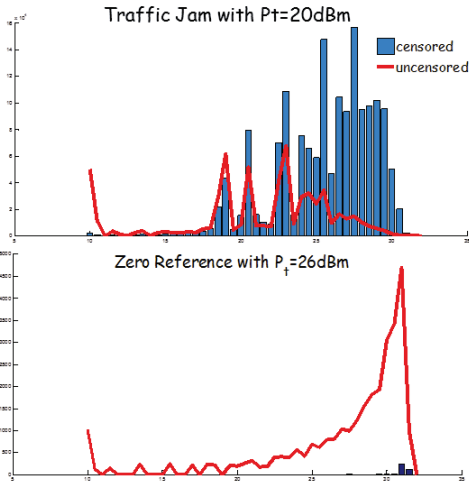
Fig. 6. Number of packets received vs. $10log(d)$ – **Top:** TJ logs with $P_t = 20dBm$, where missing data are due to both noise floor and interference. **Bottom:** ZR logs with $P_t = 26dBm$, with no interference and diminished effect of noise floor due to $P_t$.

avoiding detection. However, if we can find several bins that have appropriate-looking distributions and regard their RSSI modes as likely median values, we can fit a path loss function (e.g. linear or piece-wise linear) through these medians. This method was used to obtain the second segment in Fig. 4, bottom, which is steeper than the first segment, as expected [8], [9].

### C. Handling Interference

For static links, we model the median path loss and shadow fading based on a database obtained by calculating RSSI modes (most frequent RSSI values) for each TX-RX pair. Note that all RSSI readings for one TX-RX pair have the same distance value (due to static OBEs). Since there is sufficient number of pairs for each distance value, we have sufficient statistics to model shadow fading and to obtain a complete path loss model. For mobile links, we address the problem of interference-caused fluctuations by using only the ZR data set to model small-scale fading, or by applying statistical signal-interference separation techniques ( [10]). With multipath thus modeled, we can delineate the fast fading from the slow fading model, as explained in detail in the next section.

## IV. MODELING METHODOLOGY

### A. Generic Path Loss

Conventionally, the path loss (PL) on a wireless link between a transmitter and receiver $d$ meters apart can be written as:

$$PL = G_{med} + 10\log_{10}(F_{sh}) + 10\log_{10}(F_{mp}), \quad (1)$$

where $G_{med}(d)$ is the median path loss, at distance $d$, over the terrain; $10\log_{10}(F_{sh})$ is the dB shadow fading; and $10\log_{10}(F_{mp})$ is the dB multipath fading. In our processing, we arrive at PL by subtracting the dB transmit power from the processed value of RSSI, i.e., the recorded RSSI adjusted for the manufacturer-specific offset between RSSI and dB received power, $10\log(p_r)$. Because shadowing and multipath

terms have different fading rates and statistics, it is desirable to separate them [11]; we show a way to do that in our processing.

### B. Median Path Loss

We have discerned two distinct patterns of RSSI vs. $d$ at distances up to around $d$ = 200 m: **(1)** from the Traffic Jam and Zero Reference trials, we see the effects of 2-ray propagation, wherein the dominant rays consist of a direct (line-of-sight) ray and a ray reflected from the road (see Fig. 2); **(2)** we see a cloud of points, suggestive of a lot of multipath scatter (see Fig. 4). We have thus devised two approaches to modeling $G_{med}$ in this d-range. For higher distances, all trials suggest the use of a straight-line segment, where this segment is contiguous with the one for $G_{med}$ in the lower range. More details on how we derive $G_{med}$ vs. $d$ from our trial data are given in [12], including the method of fitting the 2-ray scenarios.

### C. Slow (Shadow) Fading

There is a large body of evidence, e.g., [13], [14], supporting the notion that the shadow fading term, $10\log(F_{sh})$, is Gaussian. It can thus be modeled as $\sigma u$, where $u$ is a zero-mean, unit-variance Gaussian random variable, and $\sigma$ is the standard deviation *(std)* of the scatter of RSSI points about the median fit. Note, however, that the RSSI data include variations due to both shadowing and multipath. Therefore, the variance computed from data is actually $\sigma^2 = \sigma_{sh}^2 + \sigma_{mp}^2$, i.e., the variances of the two kinds of fading seem to be inextricably coupled. We now show that they can be decoupled by ascertaining the fast fading distribution.

### D. Fast (Multipath) Fading

The RSSI data base is plagued with missing data due to lost packets. Hence, the RSSI time series do not correspond to consecutive time samples, which presents problem for FF modeling. We developed filters to extract portions of the database that offer long sequences of RSSI values with no packet losses. Such a sequence, referred to as RSSI signature, is singled out in Fig. 7 among more than 10000 contiguous RSSI samples (representing both received and lost packets). Note that the RSSI values that have not been received, marked by black crosses, were interpolated only for the sake of visualization. The close-up of the highlighted RSSI signature (in the upper left corner of Fig. 7) exhibits a couple of such interpolated values, but they do not affect the FF results, being few and far apart among more than 100 samples. Note that Fig. 7 represents one of the worst cases of packet losses (2LH), which exhibit periodic bursts due to the network topology and mobility pattern of the vehicles (bottom pane of Fig. 1). The signature sequences, with the RSSI values de-logged to yield sequences of $p_r$, can be processed to estimate the underlying fading pdf. This is helped by the fact that, for the packet rates, wavelength and vehicle speeds involved, adjacent RSSI samples were assumed to be independent. To a first approximation, the correlation time $(T_c)$ is the inverse of the Doppler bandwidth, as the autocorrelation function
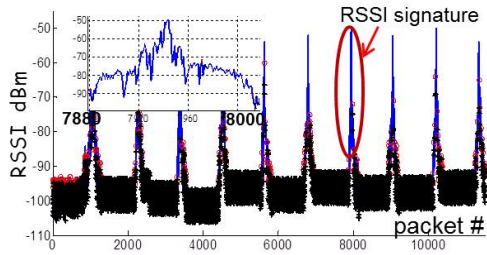
Fig. 7. 2LH logs of RSSI vs. time, where black areas denote unreceived packets. By extracting signature data blocks with few losses (the selected signature is blown up in the inset), processing can be used to estimate FF distributions.
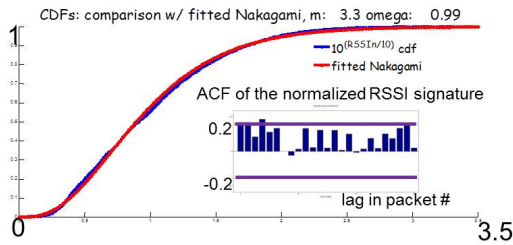


Fig. 8. Empirical and Nakagami CDFs for a block of normalized linear-amplitude samples extracted from 2LH data. The ACF of the samples (inset) shows that they are basically uncorrelated.

(ACF) is the Fourier transform of the Doppler Spectrum. Thus, $T_c \approx \lambda/v$, where the wavelength $\lambda \approx 5$ cm at 6 GHz, and $v$ is the speed. If v = 40 km/h, which is the average vehicle speed in the experiment, the correlation time is $\sim 4.5$ ms; this is a lot less than the 100-ms time between packets at the 10-Hz rate.

The process used was as follows: **(a)** Select long blocks of successive RSSI measurements, with few or no packet losses, for the same TX-RX link. These RSSI signatures are the ones processed to determine fast-fade distributions. **(b)** For each signature, de-log the RSSI values to get a sample sequence of $p_r$. **(c)** Use a P-packet-wide sliding window to smooth out the fast fluctuations $(P >> 1)$, thereby obtaining an approximate local average of $p_r$. **(d)** Use that average to normalize the P power samples, repeat over the next P packets, and continue this way over the entire block. The resulting block of normalized samples should have an ACF indicating low sample-to-sample correlations (see Fig. 8). If this is not the case, a different value of P is selected. **(e)** With the full set of de-logged and properly normalized signature samples, compare different pdfs (from Gamma or Gaussian family) with the empirical one, using both visual sighting and Kolmogorov-Smirnov testing [15].

This procedure is repeated over other long blocks of clean RSSI samples. An example of cumulative distribution functions (CDF), comparing theoretical to empirical CDFs, is shown in Fig 8. Once a satisfactory ACF and a good CDF fit are achieved, it is easy to determine, from any sequence of power samples, the std of the corresponding dB values. This is $\sigma_{mp}$, the standard deviation of $F_{mp}$ in our path loss formula. With $\sigma$ thus determined from the path loss fitting, and $\sigma_{mp}$ computed from the FF procedure above, it is easy to determine

$\sigma_{sh}$, the std of the shadow fading.

## V. CONCLUSION

Despite the fact that RSSI measurements in the field, with many radios transmitting at once, cannot be as clean or as flexible in measurement bandwidth as one would desire, they are still a valuable resource for propagation modeling. The main virtue of such data is that they are obtained at a low cost compared to other methods, hence allowing large-scale experiments illustrative of real-world V2V network phenomena, including fading and interference due to high density of mobile vehicles. We have shown several methods, with examples based on field data, which permit models to be extracted for specified scenarios. We introduced several characterizing factors for each scenario, namely vehicle, transmitter and message densities, allowing collected RSSI measurements to be profiled and adequately processed using our guidelines. In [12], we have also demonstrated good agreement between performance simulations using our processed models and field performance data. A major obstacle is that censored (lost) data is due not only to fading and the median attenuation at high distances, but also due to interference. In [10] we have presented how to decouple interference using stochastic expectation maximization to find the most likely statistical model of the signal given RSSI samples.

## REFERENCES

[1] C. F. Mecklenbraüker et al, "Vehicular Channel Characterization and Its Implications for Wireless System Design and Performance," *Proceedings of the IEEE*, vol. 99, pp. 1189–1212, 2011.

[2] L. Cheng et al, "Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 GHz dedicated short range communication (dsrc) frequency band," *IEEE Journal on Selec. Areas in Communications*, vol. 25, pp. 1501 – 1516, 2007.

[3] C.-X. Wang, X. Cheng, and D. I. Laurenson, "Vehicle-to-vehicle channel modeling and measurements: Recent advances and future challenges," *IEEE Comm. Mag.*, vol. 47, Nov 2009.

[4] S. I. and M. D.W., "VehicleVehicle Channel Models for the 5-GHz Band," *IEEE Trans. on Intell. Transp. Systems*, vol. 9, Jun 2008.

[5] J. Karendal et al., "Path loss modeling for vehicle-to-vehicle communications," *IEEE Trans. on Vehicular Techn.*, vol. 60, Jan 2011.

[6] "Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Spec.: Amendment 7: Wireless Access in Vehicular Environment, Draft 10.0," *IEEE 802.11p/D10.0*, Jan 2009.

[7] Grau, G.P. et al., "Vehicle-2-vehicle communication channel evaluation using the CVIS platform," in *7th Intern. Symp. on Comm. Systems Networks and DSP (CSNDSP)*, Jul. 2010.

[8] A. J. Goldsmith, *Wireless Communications; Ch. 2.* Cambridge University Press, 2005.

[9] A. F. Molisch, *Wireless Communications, (2nd edition); Ch. 4.* Wiley and IEEE Press, 2011.

[10] S. Kokalj-Filipovic and L. Greenstein, "EM-Based Channel Estimation from Crowd-Sourced RSSI Samples Corrupted by Noise and Interference," in *49th Conf. on Information Systems and Sciences (CISS)*, Mar. 2015.

[11] Jijun Yin et al., "DSRC Channel Fading Analysis from Empirical Measurement," in *1st Int. Conf. on Comm. and Networking in China, ChinaCom*, Oct. 2006.

[12] S. Kokalj-Filipovic, L. Greenstein, B. Cheng, and M. Gruteser, "V2V Propagation Modeling with Imperfect RSSI Samples," in *ArXiv*, Sep. 2014.

[13] T. S. Rappaport, *Wireless Communications: Principles and Practice; Ch. 3.* Prentice Hall and IEEE Press, 1996.

[14] W. C. Jakes (ed.), *Microwave Mobile Communications; Ch. 2 (D. O. Reudink).* IEEE Press, 1994.

[15] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures (4th Ed.).* Chapman & Hall/CRC, 2007.