WIRELESS COMMUNICATIONS AND MOBILE COMPUTING
Wirel. Commun. Mob. Comput. (2013)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/wcm.2356

RESEARCH ARTICLE

Deterministic process-based generative models for characterizing packet-level bursty error sequences†

Yejun He[1], Omar S. Salih[2], Cheng-Xiang Wang[3,2*] and Dongfeng Yuan[3]

[1] College of Information Engineering, Shenzhen University, Shenzhen, 518060, China
[2] Joint Research Institute for Signal and Image Processing, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, U.K.
[3] School of Information Science and Engineering, Shandong University, Jinan, Shandong, 250100, China

ABSTRACT

Errors encountered in digital wireless channels are not independent but rather form bursts or clusters. Error models aim to investigate the statistical properties of bursty error sequences at either packet level or bit level. Packet-level error models are crucial to the design and performance evaluation of high-layer wireless communication protocols. This paper proposes a general design procedure for a packet-level generative model based on a sampled deterministic process with a threshold detector and two parallel mappers. In order to assess the proposed method, target packet error sequences are derived by computer simulations of a coded enhanced general packet radio service system. The target error sequences are compared with the generated error sequences from the deterministic process-based generative model using some widely used burst error statistics, such as error-free run distribution, error-free burst distribution, error burst distribution, error cluster distribution, gap distribution, block error probability distribution, block burst probability distribution, packet error correlation function, normalized covariance function, gap correlation function, and multigap distribution. The deterministic process-based generative model is observed to outperform the widely used Markov models. Copyright © 2013 John Wiley & Sons, Ltd.

KEYWORDS
generative models; digital wireless channels; burst error statistics; deterministic fading processes; Markov models

***Correspondence**
Cheng-Xiang Wang, Joint Research Institute for Signal and Image Processing, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, EH14 4AS, U.K.
E-mail: cheng-xiang.wang@hw.ac.uk

1. INTRODUCTION

In general, there are two types of wireless fading channels: physical (analog) channels and digital channels. Common parameters of physical channels are noise and/or interference power, received signal strength, Doppler frequency, and so on. Physical channel models mainly characterize fading processes due to the channel impairments and their impact on the received signal. Well-known physical channel models are Rayleigh and Rice models [1]. A digital (time-discrete) channel refers to the entire communication chain, which includes the transmitter, the physical channel, and the receiver in the complex baseband. Errors occurring in digital wireless channels are not independent but instead form clusters or bursts. Common parameters of digital channels are the number and distribution of error events in a sequence of bits or packets. Channel models that are able to characterize the statistical properties of these bursty error events are called error models, which can further be divided into descriptive models [2] and generative models [3]. As the name implies, descriptive models describe burst error statistics of reference (target) error sequences obtained directly from experimental results and are therefore considered as reference models. Generative models, on the other hand, define mechanisms to generate error sequences with bursty error statistics similar to those of target error sequences and are therefore considered as simulation models. Both descriptive and generative models can be applied to the design, optimization, and performance evaluation of error

†This paper was presented in part at the *IEEE International Conference on Communications (ICC 2005)*, Seoul, Korea, May 2005.

control strategies and high-layer wireless communication protocols [4–7]. However, compared with descriptive models, generative models have the advantage of significantly reducing the computational effort required for obtaining long error sequences, and hence reducing the simulation time.

The choice between bit-level and packet-level error sequences depends on the goal of the designer. If the goal is to evaluate error control strategies, then bit-level error sequences are used. On the other hand, if the purpose is to evaluate high-layer protocols such as media access control and automatic repeat request schemes, then packet-level error sequences must be used.

In the literature, five classes of generative models have been developed for generating bit-level error sequences. Nonetheless, they can also be utilized to produce packet-level error sequences by properly tuning the involved parameters. The first class of generative models is based on finite [8–15] or infinite state Markov chains [3]. Simplified Fritchman's models (SFMs) [9,13,14], consisting of only a single error state and a number of error-free states, have been widely used. More advanced finite state Markov models such as bipartite models [10] have very high complexity when a satisfactory accuracy is required. The second class is hidden Markov models (HMMs) [16–18]. As their name implies, these models follow Markov chains but with hidden parameters that can be known through observations. Baum–Welch (BW) algorithm [19] has been widely used to train the observations in order to obtain the hidden parameters. The large number of states required for high-accuracy HMMs increases their complexity, thereby making the performance analysis of high-layer protocols increasingly difficult. The third class of generative models is based on stochastic context-free grammars [20]. These models are only applicable to error bursts with bell-shaped error density. Chaos theory [21–24] has been exploited to construct the fourth class of generative models. These models fail to describe some of the desired burst error statistics, especially the error correlation function.

Lastly, deterministic process-based generative models (DPBGMs) [25–27] are considered as the most promising class of generative models because they can approximate the desired burst error statistics very well. All the aforementioned models involve stochastic processes. However, the main process in the DPBGM is deterministic because the parameters are kept constant during the simulations. The idea of DPBGMs stems from the second-order statistics of fading processes. In [26], the authors proved the superiority of DPBGMs over other investigated models (e.g., SFM) for both hard and soft bit error sequences. In this paper, we extend the idea of Wang and Xu [26] and develop a DPBGM for the simulation of packet-level error sequences [27] and further investigate its performance. An enhanced general packet radio service (EGPRS) system at the radio link layer is utilized to obtain target packet error sequences. It is shown through simulations that the DPBGM is able to generate error sequences having burst error statistics similar to those of target error sequences.

The main contributions of this paper are summarized as follows.

(1) A general design procedure for packet-level DPBGMs is proposed. This is different from the bit-level DPBGM proposed in [26].
(2) A complete set of burst error statistics available in the current literature are thoroughly investigated in this paper in order to fully scrutinize the performance of the proposed packet-level DPBGM. To the best of our knowledge, only a subset of burst error statistics were studied in the previous papers [8–27].
(3) Furthermore, the burst error statistics, complexity, and simulation efficiency of the proposed DPBGM are compared not only with SFMs but also with BWHMMs. In [26], we have only compared the bit-level DPBGM with an SFM in terms of burst error statistics and simulation efficiency, whereas in [27], we have only compared the packet-level DPBGM with an SFM in terms of burst error statistics. It has been shown that the proposed packet-level DPBGM is superior to SFM and BWHMM in terms of all the performance metrics.

The organization of this paper is as follows. Section 2 defines some terms related to binary error sequences. The design methodology of the proposed packet DPBGM is given in Section 3. In Section 4, the burst error statistics of the target error sequences obtained from the EGPRS system are compared with those burst error statistics of the generated error sequences using the DPBGM, an SFM, and a BWHMM. Finally, conclusions are drawn in Section 5.

2. PACKET ERROR SEQUENCE

Packet error sequence is a string of '0s' and '1s'. A correctly received packet is denoted by '0' in the error sequence. Otherwise, the packet is deemed to be in error and represented by '1'. In order to characterize the error sequence efficiently, it is necessary to define some terms. A *gap* is a series of consecutive zeros between two ones, having a length equal to the number of zeros [13,28]. An *error cluster* is a region where the errors occur successively and has a length equal to the number of ones [9]. An *error-free burst* is defined as an all-zero sequence with a minimum length of η packets, where η is a positive integer [10,17]. An *error burst* is a chain of zeros and ones, starting and ending with a '1', and separated from neighboring error bursts by error-free bursts [10,17].

Obviously, a packet error sequence is composed of consecutive error bursts and error-free bursts, whereas an error burst is composed of successive error clusters separated by gaps of length less than η. The features of interest, in our study, are the error burst and error-free burst lengths. Therefore, we construct an error burst recorder $\mathbf{EB}_{\text{rec}}$ and error-free burst recorder $\mathbf{EFB}_{\text{rec}}$ to record the consecutive error burst and error-free burst lengths, respectively.

The minimum and maximum values in $\mathbf{EB}_{\text{rec}}$ are denoted as m_{B1} and m_{B2}, respectively. Similar notations can be applied to $\mathbf{EFB}_{\text{rec}}$.

3. THE DETERMINISTIC PROCESS BASED GENERATIVE MODEL

Fading processes can be described using second-order statistics such as level-crossing rate (LCR) and average duration of fades (ADF). It has been observed that second-order statistics are related to burst error statistics. This is the motivation for proposing a generative model based on fading processes. A packet error sequence has consecutive error bursts and error-free bursts. Similarly, each fading interval is followed by an inter-fading interval. Therefore, we can extract an error burst length generator and an error-free burst length generator directly from a deterministic process in order to build up a packet-level generative model.

The utilized deterministic fading process $\tilde{\zeta}(t)$ is sampled with a reliable sampling period T_A. This is natural when considering block or packet transmissions, especially when the packet is short and the data rate is high; that is, data rate is much greater than Doppler frequency. In this case, it is reasonable to assume that the various bits of a same packet experience approximately the same channel conditions [4]. For this purpose, a threshold detector with a chosen threshold value r_{th} is then applied to the sampled deterministic process $\tilde{\zeta}(kT_A)$, where k is a non-negative integer. During the simulation, the level of the deterministic process fluctuates and crosses the given threshold r_{th} along the time axis. If the level of $\tilde{\zeta}(kT_A)$ is larger than r_{th}, the model's output produces an error-free sample, whereas error sample occurs when the level of $\tilde{\zeta}(kT_A)$ is less than r_{th}. The counts of consecutive error samples or error-free samples, in the corresponding fading and inter-fading intervals of $\tilde{\zeta}(t)$, are the lengths of the error bursts and error-free bursts. This is the mechanism for obtaining the error burst length generator $\widetilde{\mathbf{EB}}_{\text{rec}}$ and error-free burst length generator $\widetilde{\mathbf{EFB}}_{\text{rec}}$. Note that any symbol that has a tilde (˜) sign is related to the generative model.

3.1. The parametrization of the sampled deterministic process

A reference packet error sequence is essentially required to find out the parameters of the underlying deterministic process used in the packet generative model. The LCR $\tilde{N}_\zeta(r_{\text{th}})$ at the chosen threshold r_{th} is fitted to the desired occurrence rate $R_{\text{EB}} = N_{\text{p}}\mathcal{N}_{\text{EB}}/T_{\text{t}}$ of error bursts. Here, N_{p} stands for the packet size, $\mathcal{N}_{\text{EB}}$ is the total number of error bursts, and T_{t} denotes the total transmission time of the reference transmission system, from which the reference packet error sequence of length N_{t} bits is obtained. The ratio $\tilde{\mathcal{R}}_B$ of the ADF $\tilde{T}_{\zeta_-}(r_{\text{th}})$ at r_{th} to the average duration of inter-fades (ADIF) $\tilde{T}_{\zeta_+}(r_{\text{th}})$ at r_{th} is approximated to the desired ratio $\mathcal{R}_B = M_{\text{EB}}/M_{\text{EFB}}$, where M_{EB} and M_{EFB} are the mean values of the error burst and error-free burst lengths, respectively in the reference packet error sequence. Moreover, the sampling interval T_A is chosen carefully, as specified below, in order to detect most of the level crossings and fading intervals at deep levels.

For the deterministic process, we choose a simple continuous-time deterministic process [26,29,30] as

$$\tilde{\zeta}(t) = |\tilde{\mu}_1(t) + j\tilde{\mu}_2(t)| \tag{1}$$

with

$$\tilde{\mu}_i(t) = \sum_{n=1}^{N_i} c_{i,n}\cos(2\pi f_{i,n}t + \theta_{i,n}), \quad i = 1, 2 \tag{2}$$

The number of sinusoids is N_i. The phases $\theta_{i,n}$ are considered as the realizations of a random generator uniformly distributed over $(0, 2\pi]$. The gains $c_{i,n} = \sigma_0\sqrt{\frac{2}{N_i}}$ where σ_0 is the square root of the mean power of $\tilde{\mu}_i(t)$. The discrete frequencies $f_{i,n} = f_{\max}\sin\left[\frac{\pi}{2N_i}\left(n - \frac{1}{2}\right)\right]$ where $f_{\max}$ represents the maximum Doppler frequency.

When using the method of exact Doppler spread (MEDS) with $N_i \geq 7$, the LCR $\tilde{N}_\zeta(r_{\text{th}})$ of $\tilde{\zeta}(t)$ approximately fits the LCR $N_\zeta(r_{\text{th}})$ of a Rayleigh process. Moreover, the ADF $\tilde{T}_{\zeta_-}(r_{\text{th}})$ and the ADIF $\tilde{T}_{\zeta_+}(r_{\text{th}})$ of $\tilde{\zeta}(t)$ approximate very well the corresponding quantities of a Rayleigh process $T_{\zeta_-}(r_{\text{th}}) = \sqrt{\frac{2\pi}{\beta}}\frac{\sigma_0^2}{r_{\text{th}}}\left[\exp\left(\frac{r_{\text{th}}^2}{2\sigma_0^2}\right) - 1\right]$ and $T_{\zeta_+}(r_{\text{th}}) = \sqrt{\frac{2\pi}{\beta}}\frac{\sigma_0^2}{r_{\text{th}}}$, respectively. Therefore, $\tilde{\mathcal{R}}_B$ can be approximated as

$$\tilde{\mathcal{R}}_B \approx \frac{T_{\zeta_-}(r_{\text{th}})}{T_{\zeta_+}(r_{\text{th}})} = \exp\left(\frac{r_{\text{th}}^2}{2\sigma_0^2}\right) - 1 \tag{3}$$

As observed, the second-order statistics of $\tilde{\zeta}(kT_A)$ are determined by the parameter vector $\mathbf{\Psi} = (N_1, N_2, r_{\text{th}}, \sigma_0, f_{\max}, T_A)$. The parameter r_{th} can be assigned such that it is much less than 1. In order to find other parameters, we apply $\mathcal{R}_B = \tilde{\mathcal{R}}_B$ and $R_{\text{EB}} = N_\zeta(r_{\text{th}})$ to get $f_{\max} = \frac{N_{\text{p}}\mathcal{N}_{\text{EB}}(1+\mathcal{R}_B)}{T_{\text{t}}\sqrt{2\pi\ln(1+\mathcal{R}_B)}}$. $T_A \approx \frac{4\sigma_0\left[\exp\left(\frac{r_{\text{th}}^2}{2\sigma_0^2}\right)-1\right]}{\sqrt{5}\pi r_{\text{th}} f_{\max}}\sqrt{-1 + \sqrt{1 + 10q_{\text{s}}/3}}$, where q_{s} determines the maximum measurement error of the LCR and chosen as 0.01.

Finally, the sampled deterministic process $\tilde{\zeta}(kT_A)$ can be simulated within the necessary time interval $[0, \tilde{T}_{\text{t}}]$. Here, $\tilde{T}_{\text{t}} = T_{\text{t}}\tilde{N}_{\text{t}}/N_{\text{t}}$ with $\tilde{N}_{\text{t}}$ denoting the required length of the generated packet error sequence. The total numbers of the generated error bursts $\tilde{\mathcal{N}}_{\text{EB}}$ and error-free bursts $\tilde{\mathcal{N}}_{\text{EFB}}$ can approximately be estimated from $\tilde{\mathcal{N}}_{\text{EB}} = \left\lfloor \frac{\tilde{N}_{\text{t}}}{N_{\text{t}}}\mathcal{N}_{\text{EB}} \right\rfloor$ and $\tilde{\mathcal{N}}_{\text{EFB}} = \left\lfloor \frac{\tilde{N}_{\text{t}}}{N_{\text{t}}}\mathcal{N}_{\text{EFB}} \right\rfloor$, respectively. Here, $\lfloor x \rfloor$ stands for the floor of x. Consequently, an error burst length generator $\widetilde{\mathbf{EB}}_{\text{rec}}$ with $\tilde{\mathcal{N}}_{\text{EB}}$ entries and an error-free burst length generator $\widetilde{\mathbf{EFB}}_{\text{rec}}$ with $\tilde{\mathcal{N}}_{\text{EFB}}$ entries are derived.

3.2. The mappers

The lengths of the generated error bursts and error-free bursts do not match the desired lengths properly. Therefore, two mappers are designed in order to achieve good fit to the desired burst error statistics. The number of error bursts of length m_{e} in $\mathbf{EB}_{\text{rec}}$ is denoted by $N_{\text{EB}}(m_{\text{e}})$ and the number of error-free bursts of length $m_{\bar{\text{e}}}$ in $\mathbf{EFB}_{\text{rec}}$ is denoted by $N_{\text{EFB}}(m_{\bar{\text{e}}})$. A modification to $\widetilde{\mathbf{EB}}_{\text{rec}}$ and $\widetilde{\mathbf{EFB}}_{\text{rec}}$ must be done such that $\tilde{N}_{\text{EB}}(m_{\text{e}}) = \hat{N}_{\text{EB}}(m_{\text{e}})$ and $\tilde{N}_{\text{EFB}}(m_{\bar{\text{e}}}) = \hat{N}_{\text{EFB}}(m_{\bar{\text{e}}})$ hold, respectively. Here, $\hat{N}_{\text{EB}}(m_{\text{e}})$ equals $\left\lfloor \frac{\tilde{N}_{\text{t}}}{N_{\text{t}}} N_{\text{EB}}(m_{\text{e}}) \right\rfloor$ or $\left\lfloor \frac{\tilde{N}_{\text{t}}}{N_{\text{t}}} N_{\text{EB}}(m_{\text{e}}) \right\rfloor + 1$ for different error burst lengths m_{e} in order to fulfill $\sum_{m_{\text{e}}=m_{B1}}^{m_{B2}} \hat{N}_{\text{EB}}(m_{\text{e}}) = \tilde{\mathcal{N}}_{\text{EB}}$. Similarly, $\hat{N}_{\text{EFB}}(m_{\bar{\text{e}}})$ equals $\left\lfloor \frac{\tilde{N}_{\text{t}}}{N_{\text{t}}} N_{\text{EFB}}(m_{\bar{\text{e}}}) \right\rfloor$ or $\left\lfloor \frac{\tilde{N}_{\text{t}}}{N_{\text{t}}} N_{\text{EFB}}(m_{\bar{\text{e}}}) \right\rfloor + 1$ for different error-free burst lengths $m_{\bar{\text{e}}}$ to satisfy $\sum_{m_{\bar{\text{e}}}=m_{\bar{B}1}}^{m_{\bar{B}2}} \hat{N}_{\text{EFB}}(m_{\bar{\text{e}}}) = \tilde{\mathcal{N}}_{\text{EFB}}$.

In order to modify $\widetilde{\mathbf{EB}}_{\text{rec}}$, we first find the corresponding values $\ell_{m_{\text{e}}}^1$ and $\ell_{m_{\text{e}}}^2$ in $\widetilde{\mathbf{EB}}_{\text{rec}}$ to assure the following conditions

$$\sum_{l=\ell_{m_{\text{e}}}^1}^{\ell_{m_{\text{e}}}^2-1} \tilde{N}_{\text{EB}}(l) < \hat{N}_{\text{EB}}(m_{\text{e}})$$

$$\text{and} \qquad \sum_{l=\ell_{m_{\text{e}}}^1}^{\ell_{m_{\text{e}}}^2} \tilde{N}_{\text{EB}}(l) \geq \hat{N}_{\text{EB}}(m_{\text{e}}) \tag{4}$$

Then,

$$\sum_{l=\ell_{m_{\text{e}}}^1}^{\ell_{m_{\text{e}}}^2-1} \tilde{N}_{\text{EB}}(l) + N_{\ell_{m_{\text{e}}}^2} = \hat{N}_{\text{EB}}(m_{\text{e}}) \tag{5}$$

The same idea applies to $\widetilde{\mathbf{EFB}}_{\text{rec}}$. It is clear that the mappers renovate the l $\left(\ell_{m_{\text{e}}}^1 \leq l \leq \ell_{m_{\text{e}}}^2 - 1\right)$ samples of the fading process in each interval to m_{e}, which is the required burst length.

3.3. The generation of packet error sequences

The mapping process yields the correct $\widetilde{\mathbf{EB}}_{\text{rec}}$ and $\widetilde{\mathbf{EFB}}_{\text{rec}}$, from which the corresponding error burst and error-free bursts, respectively, can be produced. Because error bursts consist of clusters and gaps combined in sequence, it is convenient to create parameter vectors $\widetilde{\mathbf{ECG}}_j$ ($j = 1, 2, \ldots, \tilde{\mathcal{N}}_{\text{EB}}$), which reflect the construction of each error burst from $\widetilde{\mathbf{EB}}_{\text{rec}}$ as error cluster and gap lengths. Therefore, all vectors $\mathbf{ECG}_i$ corresponding to error bursts with length m_{e} are found in $\mathbf{EB}_{\text{rec}}$. Thereafter, for all error bursts with the same length m_{e} in $\widetilde{\mathbf{EB}}_{\text{rec}}$, random $\widetilde{\mathbf{ECG}}_j$ are allocated from all possible vectors $\mathbf{ECG}_i$. This is the procedure for generating the error bursts. Error-free bursts, on the other hand, consist of zeros only. Therefore, they are obtained by generating a series of zeros for each length in $\widetilde{\mathbf{EFB}}_{\text{rec}}$. By combining generated error bursts with error-free bursts in succession, an entire packet error sequence is constructed.

Note that in the aforementioned three design steps of the DPBGM, the first two steps (parametrization and mapping) are called the simulation set-up phase, and the last step (generation of packet error sequences) is called the simulation run phase.

4. SIMULATION RESULTS AND DISCUSSIONS

In order to validate the efficiency and accuracy of the DPBGM, we apply its mechanism to reference packet error sequences. These error sequences were obtained from an EGPRS simulator. The EGPRS system that we have used is composed of a convolutional encoder; a burst interleaver and a Gaussian minimum shift keying modulator at the transmitter side; and a Gaussian minimum shift keying demodulator, a viterbi equalizer, a burst de-interleaver, a convolutional decoder, and a cyclic redundancy check for error detection at the receiver side. The convolutional decoder uses the coding scheme 3 (CS3) which is specific for the EGPRS [31]. According to [31], the utilized propagation channel can be expressed as NAMEx, where x represents the vehicle speed in km/h. NAME here represents the name of the underlying channel (e.g., a rural area (RA) channel and a typical urban (TU) channel). Moreover, the EGPRS supports both ideal frequency-hopping (IFH) and non-FH (NFH). IFH introduces good decorrelation between time-division multiple access bursts. When employing NFH, the data are transmitted using a GSM carrier within a bandwidth of 200 kHz; therefore, the system operates in narrowband channel. On the other hand, when employing the IFH, the bandwidth could be 5 MHz or more; hence, the channel is wideband. In this paper, the following channels are utilized: RA 275 NFH, TU3 IFH, TU3 NFH, and TU50 NFH [31]. The data were transmitted as packets of $N_{\text{p}} = 456$ bits with a transmission rate of $F_{\text{s}} = 270.8$ kb/s. Each frame contains four time-division multiple access bursts of 114 bits. The target packet error sequences of length $N_{\text{t}} = 1 \times 10^6$ were produced at carrier-to-interference ratios of 5, 7, 8, 9, 11, 13, 15, and 17 dB. They are formulated by allotting a '0' to a correctly decoded packet and a '1' to an error packet, which contains at least one undecodable error. The total transmission time is therefore $T_{\text{t}} = N_{\text{p}} N_{\text{t}} / F_{\text{s}} = 1684$ s, which is about 28 min.

The performance criteria are evaluated by working out burst error statistics. One generative model is preferred over others if its burst error statistics fit very well those of the reference packet error sequences obtained directly from the EGPRS system. The burst error statistics that are illustrated in this paper are defined as follows.

(1) $G(m_g)$: the gap distribution (GD), which is defined as the cumulative distribution function (CDF) of gap lengths m_g [28].
(2) $P(0^{m_0}|1)$: the error-free run distribution (EFRD), which is the probability that an error is followed by at least m_0 error-free packets [9]. The EFRD can be calculated from the GD [28]. Obviously, $P(0^{m_0}|1)$ is a monotonically decreasing function of m_0 such that $P(0^0|1) = 1$ and $P(0^{m_0}|1) \to 0$ as $m_0 \to \infty$.
(3) $P(1^{m_c}|0)$: the error cluster distribution (ECD), which is the probability that a correct packet is followed by m_c or more consecutive packets in error [9].
(4) $P_{EB}(m_e)$: the error burst distribution (EBD), which is the CDF of error burst lengths m_e.
(5) $P_{EFB}(m_{\bar{e}})$: the error-free burst distribution (EFBD), which is the CDF of error-free burst lengths $m_{\bar{e}}$.
(6) $P(m,n)$: the block error probability distribution (BEPD), which is the probability that at least m out of n packets are in error. This statistic is important for determining the performance of automatic repeat request protocols [13].
(7) $Q(l,n)$: the block burst probability distribution (BBPD), which is the probability of an error burst of length l occurring in a block of length n. For only this statistic, the length of a burst in a block of n digits is the number of zeros and ones between the first error to the last error in the block (both errors included) irrespective of the nature of the digits in between [3].
(8) $Cov(l)$: the normalized covariance function [20].
(9) $\rho(\Delta k)$: the packet error correlation function (PECF), which is the conditional probability that the Δkth packet following an error packet is also in error. The PECF represents the burstiness of the channel [2,3].
(10) GCF: the gap correlation function, which is the conditional probability that the Δrth gap following a short (long) gap is also short (long) [3].
(11) MGD: the multigap distribution, which is the CDF of r consecutive gaps, considered as a single parameter, which are separated by one or more consecutive errors [3]. The gap here is different from the one adopted before, it is defined here as a string of consecutive zeros between two errors and having a length equal to one plus the number of zeros between the two errors. It can be seen that the minimum value for a gap length is one, occurring in case of consecutive errors.

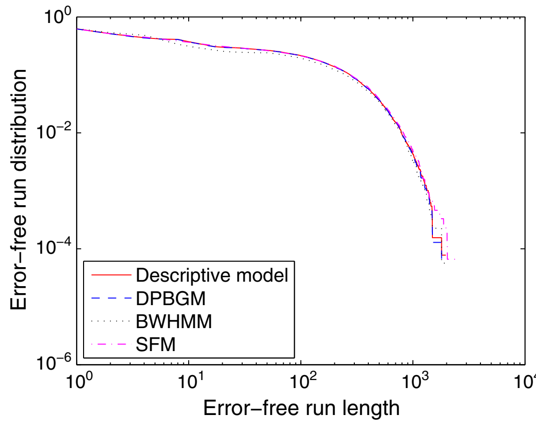

Figure 1. The error-free run distributions of the descriptive model and three generative models. DPBGM, deterministic process-based generative model; BWHMM, Baum–Welch-based hidden Markov model; SFM, simplified Fritchman's model.

The obtained packet error sequences are applied to the DPBGM to generate error sequences with length $\tilde{N}_t = 1.2 \times 10^6$. In this paper, we show only the results of TU3 IFH at carrier-to-interference ratio of 11 dB. The corresponding FER is 0.013. The value of η can be found from Figure 1 when the curve becomes slightly constant. The value of η is chosen to be 50. Consequently, the value of $\mathcal{N}_{EB} = 3429$ and $\mathcal{R}_B = 0.052$. The value of q_s is 0.01; hence, the parameter vector for the underlying sampled deterministic process is $\boldsymbol{\Psi} = (9, 10, 0.09, 0.2255, 300.85 \text{ Hz}, 0.0394 \text{ ms})$.

For comparison purposes two other generative models, namely SFM and BW-based HMM (BWHMM) have been implemented. For the SFM, The transition probability

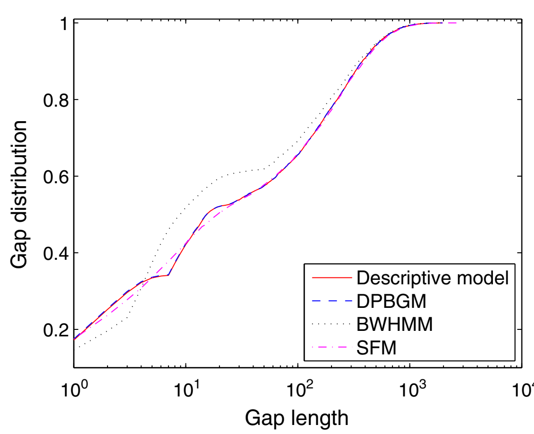

Figure 2. The GDs of the descriptive model and three generative models. DPBGM, deterministic process-based generative model; BWHMM, Baum–Welch-based hidden Markov model; SFM, simplified Fritchman's model.

Wirel. Commun. Mob. Comput. (2013) © 2013 John Wiley & Sons, Ltd.
DOI: 10.1002/wcm

matrix of a K-state is calculated by expressing the EFRD $P(0^{m_0}|1)$ as the sum of $K-1$ exponentials with suitable weighting coefficients [9]. This procedure has to involve curve fitting techniques and is called the simulation set-up phase of an SFM. From the transition probability matrix of an SFM, packet error sequences can be generated for any desired length, which is considered as the simulation run phase of an SFM. In our case, the fitting of $P(0^{m_0}|1)$ is achieved by using five exponentials. Our experiments have shown that no better performance can be obtained from SFMs with more than six states.

In the case of BWHMM, the simulation set-up phase involves extracting the error bursts from the error sequence. The error bursts are then divided into smaller blocks of length L. With the maximum number of errors in L, the error bursts are classified. The symbols represent the number of errors in each block, which are considered as subclasses. These symbols are used in the training process with the BW algorithm in order to find the optimal values of the transition probabilities between the subclasses. Afterwards, error bursts can be generated in each class. Finally, the classes of error bursts are combined with one class for error-free bursts. The last phase is called the simulation run phase. In our simulations, the number of classes is 3, and the total number of states is 100. The chosen value of L is 2.

Figures 1–12 depict the behavior of the descriptive model and the generative models of the DPBGM, BWHMM, and SFM in terms of EFRDs, GDs, ECDs, EBDs, EFBDs, BEPDs, BBPDs, NCs, PECFs, GCFs,

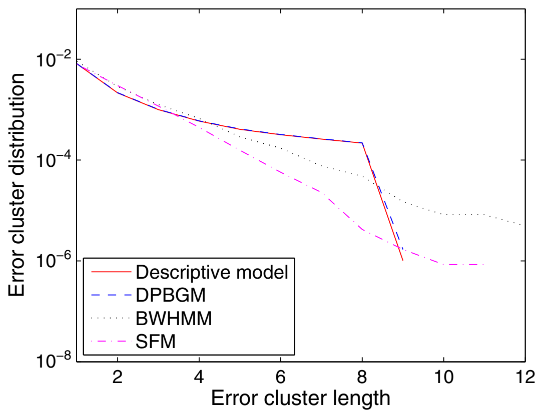

Figure 3. The error cluster distributions of the descriptive model and three generative models. DPBGM, deterministic process-based generative model; BWHMM, Baum–Welch-based hidden Markov model; SFM, simplified Fritchman's model.

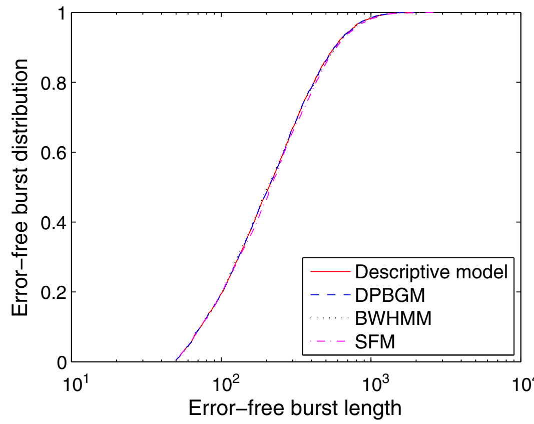

Figure 5. The error-free burst distributions of the descriptive model and three generative models. DPBGM, deterministic process-based generative model; BWHMM, Baum–Welch-based hidden Markov model; SFM, simplified Fritchman's model.

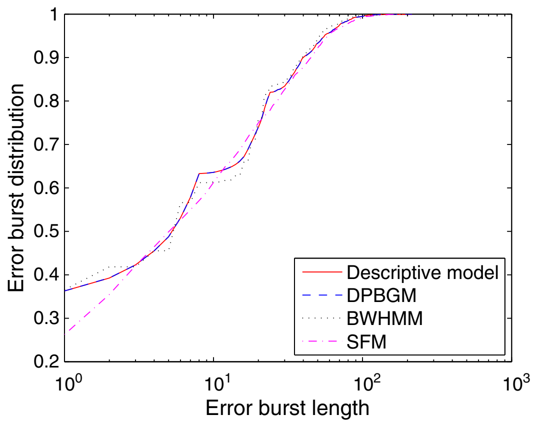

Figure 4. The error burst distributions of the descriptive model and three generative models. DPBGM, deterministic process-based generative model; BWHMM, Baum–Welch-based hidden Markov model; SFM, simplified Fritchman's model.

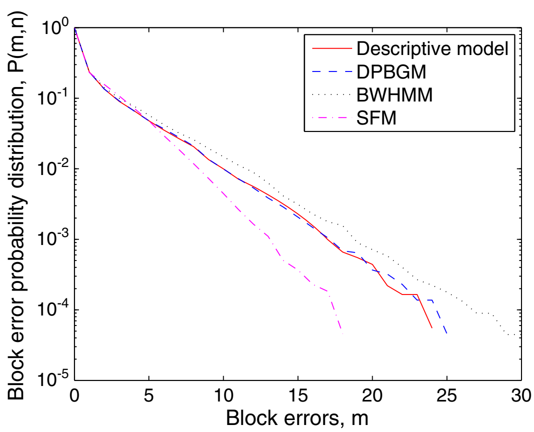

Figure 6. The block error probability distributions of the descriptive model and three generative models with $n = 55$. DPBGM, deterministic process-based generative model; BWHMM, Baum–Welch-based hidden Markov model; SFM, simplified Fritchman's model.

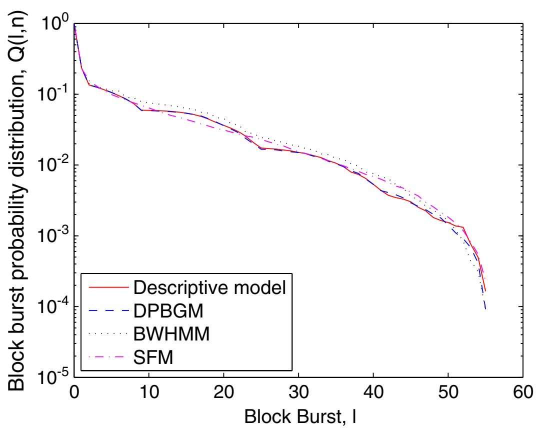

Figure 7. The block burst probability distributions of the descriptive model and three generative models with $n = 55$. DPBGM, deterministic process-based generative model; BWHMM, Baum–Welch-based hidden Markov model; SFM, simplified Fritchman's model.

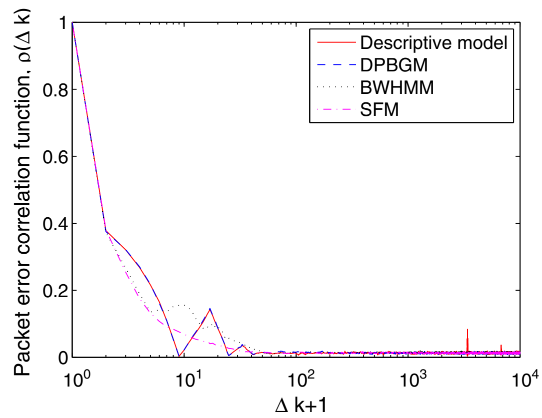

Figure 9. The packet error correlation functions of the descriptive model and three generative models. DPBGM, deterministic process-based generative model; BWHMM, Baum–Welch-based hidden Markov model; SFM, simplified Fritchman's model.

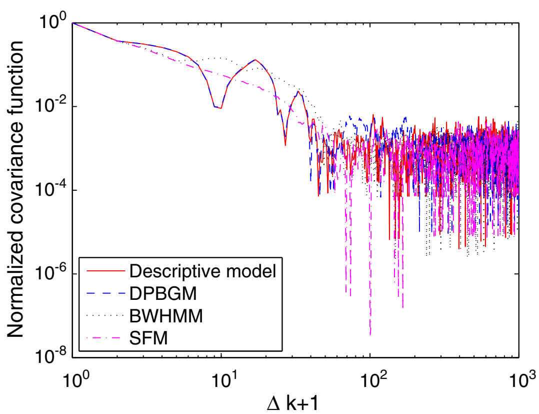

Figure 8. The normalized covariance functions of the descriptive model and three generative models. DPBGM, deterministic process-based generative model; BWHMM, Baum–Welch-based hidden Markov model; SFM, simplified Fritchman's model.

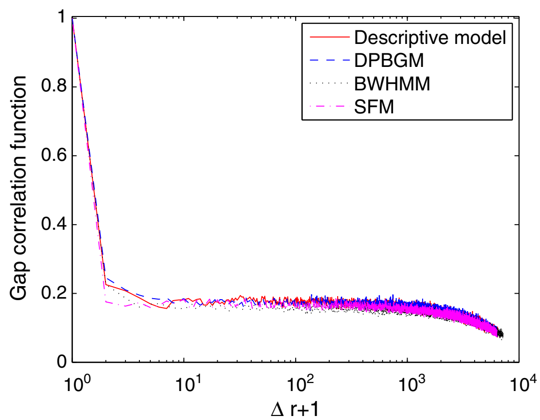

Figure 10. The gap correlation functions of the descriptive model and three generative models. DPBGM, deterministic process-based generative model; BWHMM, Baum–Welch-based hidden Markov model; SFM, simplified Fritchman's model.

MGDs against the multigap length of order 10, and MGDs against the multigap length of order 100, respectively. In general, the DPBGM outperforms the SFM and BWHMM in terms of accuracy by fitting the descriptive model well for all the defined burst error statistics. An exception is the MGD against the multigap length of order 100 (Figure 12), for which the SFM behaves the best. However, the SFM fails to describe some of the desired burst error statistics, as can be seen from the large deviations for ECD (Figure 3), EBD (Figure 4), BEPD (Figure 6), NCF (Figure 8), and PECF (Figure 9), whereas it matches satisfactorily the descriptive model for EFRD (Figure 1), GD (Figure 2), EFBD (Figure 5), BBPD (Figure 7), and MGDs (Figures 11 and 12). The BWHMM fails to characterize the GD (Figure 2), ECD (Figure 3), BEPD (Figure 6), NC (Figure 8), and PECF (Figure 9) adequately, whereas the performance is acceptable for the rest burst error statistics. This is because the BWHMM is best designed to be used for error sequences having error bursts with bell-shaped error density, while this property is difficult to be found in packet error sequences. Nonetheless, the BWHMM shows better performance than the SFM in terms of ECD (Figure 3), EBD (Figure 4), EFBD (Figure 5), BEPD (Figure 6), and PECF (Figure 9), but is no better than the SFM for the rest burst error statistics.

In terms of model complexity and simulation efficiency, all the three generative models require two phases: simulation set-up phase and simulation run phase. For the set-up phase, all the three models have high complexity and

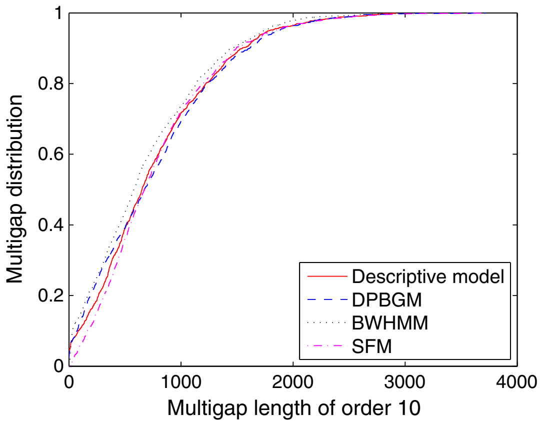

Figure 11. The multigap distributions (multigap length of order 10) of the descriptive model and three generative models. DPBGM, deterministic process-based generative model; BWHMM, Baum–Welch-based hidden Markov model; SFM, simplified Fritchman's model.

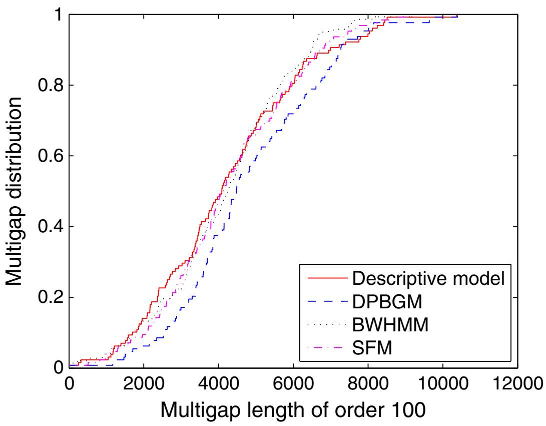

Figure 12. The multigap distributions (multigap length of order 100) of the descriptive model and three generative models. DPBGM, deterministic process-based generative model; BWHMM, Baum–Welch-based hidden Markov model; SFM, simplified Fritchman's model.

require a long simulation time, which depends on individual experiences and are difficult to compare. For the simulation run phase, the DPBGM has the minimum complexity followed by the BWHMM and then the SFM. Using a PC with a 2.4-GHz processor, the DPBGM, BWHMM, and SFM need 0.422, 1.125, and 3.422 s, respectively. Thus, the DPBGM outperforms the BWHMM and SFM in terms of accuracy as well as efficiency.

5. CONCLUSIONS

We have established a general procedure for developing a fast binary packet-level generative model with a properly parameterized and sampled deterministic process followed by a threshold detector and two parallel mappers. It has been demonstrated that in general, the DPBGM exhibits excellent conformity with the descriptive model especially for the most important burst error statistics such as the PECF and BEPD, which are used in the design and performance evaluation of the media access control layer, link control layer, and high-layer wireless communication protocols. The SFM and BWHMM fail to describe most of the burst error statistics and notably the important ones. The SFM outperforms the BWHMM in terms of the EFRD and MGD but is worse than the DPBGM for the same statistics. The BWHMM performs better than the SFM in terms of ECD, EBD, EFBD, BEPD, and PECF, but not better than the DPBGM. The DPBGM has shown its superiority in terms of efficiency as well. The conclusions have been substantiated through performance simulation of coded EGPRS systems producing packet error sequences.

ACKNOWLEDGEMENTS

Y. He acknowledges the support of his work by National Natural Science Foundation of China (grant no. 60972037) and the Fundamental Research Program of Shenzhen City (grant no. JC200903120101A and JC201005250067A). O. Salih and C-X. Wang would like to acknowledge gratefully the sponsorship of this work by the EPSRC and Philips Research Cambridge, UK, and the support from the Opening Project of Key Laboratory of Cognitive Radio and Information Processing (Guilin University of Electronic Technology), Ministry of Education, China (grant no. 2011KF01). The authors would also like to acknowledge the support from the RCUK for the UK–China Science Bridges project: R&D on (B)4G Wireless Mobile Communications.

REFERENCES

1. Jakes WC (ed.). *Microwave Mobile Communications*, 2nd edition. IEEE Press: New Jersey, 1994.
2. Crespo PM, Pelz RM, Cosmas J. Channel error profiles for DECT. *IEE Proceedings—Communications* 1994; **141**(6): 413–420.
3. Kanal LN, Sastry ARK. Models for channels with memory and their applications to error control. *Proceedings of the IEEE* 1977; **66**(7): 724–744.
4. Zorzi M, Rao RR, Milstein LB. ARQ error control for fading mobile radio channels. *IEEE Transactions on Vehicular Technology* 1997; **46**(2): 445–455.
5. Zorzi M, Rao RR. Perspectives on the impact of error statistics on protocols for wireless networks. *IEEE Personal Communications* 1999; **6**(10): 32–40.
6. Pimentel C, Alajaji F. Packet-based modeling of Reed–Solomon block-coded correlated fading channels via a Markov finite queue model. *IEEE*

Transactions on Vehicular Technology 2009; **58**(7): 3124–3136.

7. Badia L, Baldo N, Levorato M, Zorzi M. A Markov framework for error control techniques based on selective retransmission in video transmission over wireless channels. *IEEE Journal on Selected Areas in Communications* 2010; **28**(3): 488–500.
8. Elliot EO. Estimates of error rates for codes on burst-noise channels. *Bell System Technical Journal* 1963; **42**: 1977–1997.
9. Fritchman BD. A binary channel characterization using partitioned Markov chains. *IEEE Transactions on Information Theory* 1967; **13**(2): 221–227.
10. Willig A. A new class of packet- and bit-level models for wireless channels, In *Proceedings of IEEE PIMRC*, Lisbon, Portugal, September 2002; 2434–2440.
11. Adoul J-PA, Fritchman BD, Kanal LN. Characterization and modeling of real communication channels. *Technical Report*, Department of Electrical Engineering, Lehigh University, and Computer Science Centre, University of Maryland, July 1970.
12. Trafton PJ, Blank HA, McAllister NF. Data transmission network computer-to-computer study, In *Proceedings of the ACM IEEE 2nd Symposium on Problems in the Optimization of Data Communication Systems*, CA, USA, October 1971; 183–191.
13. Swarts F, Ferreira HC. Markov characterization of digital fading mobile VHF channels. *IEEE Transactions on Vehicular Technology* 1994; **43**(4): 977–985.
14. Semmar A, Lecours M, Chouinard JY, Ahern J. Characterization of error sequences in UHF digital mobile radio channels. *IEEE Transactions on Vehicular Technology* 1991; **40**(4): 769–776.
15. Fukawa K, Suzuki H, Tateishi Y. Packet error rate analysis using Markov models of signal-to-interference ratio for mobile packet systems. *IEEE Transactions on Vehicular Technology* 2012; **61**(6): 2517–2530.
16. Turin W. *Digital Transmission Systems: Performance Analysis and Modeling*. McGraw-Hill: New York, 1999.
17. Garcia-Frias J, Crespo PM. Hidden Markov models for burst error characterization in indoor radio channels. *IEEE Transactions on Vehicular Technology* 1997; **46**(6): 1006–1020.
18. Salih OS, Wang C-X, Laurenson DI. Double embedded processes based hidden Markov models for binary digital wireless channels, In *Proceedings of the ISWCS'08*, Reykjavik, Iceland, October 2008; 219–223.
19. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 1989; **77**: 257–286.
20. Zhu W, Garcia-Frias J. Stochastic context-free grammars and hidden Markov models for modeling of bursty channels. *IEEE Transactions on Vehicular Technology* 2004; **53**(3): 666–676.
21. Costamagna E, Favalli L, Gamba P. Multipath channel modeling with chaotic attractors. *Proceedings of the IEEE* 2002; **90**(5): 842–859.
22. Costamagna E, Favalli L, Gamba P, Savazzi P. Block-error probabilities for mobile radio channels derived from chaos equations. *IEEE Communications Letters* 1999; **3**(3): 66–68.
23. Costamagna E, Favalli L, Savazzi P, Tarantola F. Long sequences of error gaps derived from chaotic generators optimized for short ones in mobile radio channels, In *Proceedings of the IEEE VTC'04—Fall*, Los Angeles, USA, September 2004; 4250–4254.
24. Köpke A, Willig A, Karl H. Chaotic maps as parsimonious bit error models of wireless channels, In *Proceedings of IEEE INFOCOM'03*, San Francisco, USA, March 30–April 3, 2003; 513–523.
25. Wang C-X, Pätzold M. A generative deterministic model for digital mobile fading channels. *IEEE Communications Letters* 2004; **8**(4): 223–225.
26. Wang C-X, Xu W. A new class of generative models for burst-error characterization in digital wireless channels. *IEEE Transactions on Communications* 2007; **55**(3): 453–462.
27. Wang C-X, Xu W. Packet-level error models for digital wireless channels, In *Proceedings of the IEEE ICC'05*, Seoul, Korea, May 2005; 2184–2189.
28. Tsai S. Markov characterization of the HF channel. *IEEE Transactions on Communication Technology* 1969; **17**(1): 24–32.
29. Wang C-X, Pätzold M, Yuan D. Accurate and efficient simulation of multiple uncorrelated Rayleigh fading waveforms. *IEEE Transactions on Wireless Communications* 2007; **6**(3): 833–839.
30. Wang C-X, Yuan D, Chen H-H, Xu W. An improved deterministic SoS channel simulator for efficient simulation of multiple uncorrelated Rayleigh fading channels. *IEEE Transactions on Wireless Communications* 2008; **7**(9): 3307–3311.
31. 3GPP TS 45.005. Radio transmission and reception (Rel. 4), 2003.

AUTHORS' BIOGRAPHIES

Yejun He (SM'09) received his PhD degree in Information and Communication Engineering from Huazhong University of Science and Technology (HUST) in 2005, MS degree in Communication and Information System from Wuhan University of Technology (WHUT) in 2002, and his BS degree from Huazhong University of Science and Technology in 1994. From Sept. 2005 to Mar. 2006, he was a research associate with the Department of Electronic and Information Engineering, The Hong Kong Polytechnic University. From April 2006 to Mar. 2007, he was a research associate with the Department of Electronic Engineering, Faculty of Engineering, The Chinese University of Hong Kong. From July 2012 to August 2012, he was a visiting professor with Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. Dr. He is currently a full professor at Shenzhen University, China. His research interests include channel coding and modulation, MIMO-OFDM wireless communication, space–time processing, and smart antennas. Prof. He is a senior member of IEEE and a senior member of China Institute of Communications and China Institute of Electronics. He served as the publicity chair of IEEE PIMRC 2012 and has served as associate editor of Security and Communication Networks journal since 2012. Prof. He is the principal investigator for more than 10 current or finished research projects including NSFC, the integration project of production teaching and research by Guangdong Province and Ministry of Education, and the Science and Technology Program of Shenzhen City. Prof. He has translated three English books into Chinese and authored or co-authored more than 60 research papers as well as applied for 11 patents since 2002.

Omar S. Salih (S'07) is a PhD candidate at the Joint Research Institute for Signal and Image Processing, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, UK. His research interests include digital channel modeling, the long-term evolution system, multi-input multi-output antenna technologies, and propagation channel modeling.

Cheng-Xiang Wang (S'01-M'05-SM'08) received the BSc and MEng degrees in Communication and Information Systems from Shandong University, China, in 1997 and 2000, respectively, and the PhD degree in Wireless Communications from Aalborg University, Denmark, in 2004.

He has been with Heriot-Watt University, Edinburgh, UK, since 2005, first as a lecturer, then as a reader in 2009, and was promoted to a professor in 2011. He is also an honorary fellow of the University of Edinburgh, UK, and a chair/guest professor of Shandong University, Huazhong University of Science and Technology, and Southeast University, China. He was a research fellow at the University of Agder, Grimstad, Norway, from 2001–2005, a visiting researcher at Siemens AG-Mobile Phones, Munich, Germany, in 2004, and a research assistant at Technical University of Hamburg-Harburg, Hamburg, Germany, from 2000–2001. His current research interests include wireless channel modeling and simulation, green communications, cognitive radio networks, vehicular communication networks, Large MIMO, cooperative MIMO, and B4G wireless communications. He has edited one book and published one book chapter and about 170 papers in refereed journals and conference proceedings.

Prof. Wang served or is currently serving as an editor for eight international journals, including *IEEE Transactions on Vehicular Technology* (since 2011) and *IEEE Transactions on Wireless Communications* (2007–2009). He was the leading guest editor for *IEEE Journal on Selected Areas in Communications* and *Special Issue on Vehicular Communications and Networks*. He served or is serving as a TPC member, TPC chair, and general chair for over 70 international conferences. He received the Best Paper Awards from IEEE Globecom 2010 and IEEE ICCT 2011. He is a fellow of the IET, a fellow of the HEA, and a member of EPSRC Peer Review College.

Dongfeng Yuan (SM'01) received his MSc degree from Department of Electrical Engineering, Shandong University, China, in 1988, and got his PhD degree from Department of Electrical Engineering, Tsinghua University, China, in 2000. Currently he is a full professor and dean in School of Information Science and Engineering, Shandong University, China. His research interests are in the area of wireless communications, with a focus on cross-layer design and resource allocation.

DOI: 10.1002/wcm