UWB Sparse/Diffuse Channels, Part I:
Channel Models and Bayesian Estimators

Nicolò Michelusi, *Student Member, IEEE,* Urbashi Mitra, *Fellow, IEEE,* Andreas F. Molisch, *Fellow, IEEE,*
and Michele Zorzi, *Fellow, IEEE*

Abstract—In this two-part paper, the problem of channel estimation in Ultra Wide-Band (UWB) systems is investigated. Due to the large transmission bandwidth, the channel has been traditionally modeled as sparse. However, some propagation phenomena, *e.g.*, scattering from rough surfaces and frequency distortion, are better modeled by a diffuse channel. Herein, a novel Hybrid Sparse/Diffuse (HSD) channel model is proposed. Tailored to the HSD model, channel estimators are designed for different scenarios that differ in the amount of side information available at the receiver. An Expectation-Maximization algorithm to estimate the power delay profile of the diffuse component is also designed. The proposed methods are compared to un-structured and purely sparse estimators. The numerical results show that the HSD estimation schemes considerably improve the estimation accuracy and the bit error rate performance over conventional channel estimators. In Part II, the new channel estimators are evaluated with more realistic geometry-based channel emulators. The numerical results show that, even when the channel is generated in this manner, the new estimation strategies achieve high performance. Moreover, a Mean-Squared Error analysis of the proposed estimators is performed, in the high and low Signal to Noise Ratio regimes, thus quantifying, in closed form, the achievable performance gains.

Index Terms—Ultra Wideband, Bayesian estimation, channel estimation, channel modeling, sparse approximations

I. Introduction

Ultra Wide-Band (UWB) signaling had been originally proposed as a technology for indoor mobile and multiple-access communications [1]–[3]. Due to its significant band-width, UWB offers high precision localization [4], robustness against multipath fading [5] and immunity to narrow-band interference [6], thus representing a compelling solution for applications such as short-range, high-speed broadband access [7], Wireless Body Area Networks (WBANs) [8], covert communication links, through-wall imaging, high-resolution ground-penetrating radar and asset tracking [9]–[11]. However, the performance of coherent UWB transceivers relies on the availability of accurate channel estimates. Thus, it is important to design channel estimation strategies that exploit the struc-tural and statistical properties of UWB propagation to achieve the best estimation accuracy.

N. Michelusi and M. Zorzi are with the Department of Information Engineering, University of Padova, Italy, e-mail: {michelusi,zorzi}@dei.unipd.it.

U. Mitra and A.F. Molisch are with the Ming Hsieh Department of Electrical Engineering, University of Southern California, USA, e-mail: {ubli,molisch}@usc.edu.

The significant transmission bandwidth of UWB systems enables a fine-grained delay resolution at the receiver, of the order of $1\,\mathrm{ns}$. In many environments, only some of the resolvable delay bins carry significant multipath energy, yielding a sparse channel structure [10], [12]. For this reason, UWB channel estimation strategies based on compressive sensing and sparse approximation techniques [13]–[16] have been proposed in the literature, and they have been shown to outperform conventional unstructured estimators [17], [18]. Also, localization techniques that exploit the information about the specular multipath structure of the UWB channel have been proposed (see, *e.g.*, [19], [20]).

However, recent propagation studies suggest that, for some environments, such as indoor, WBANs and vehicular sce-narios, diffuse (dense) components of the impulse response arise. These are caused by propagation processes such as diffuse scattering [21], or unresolvable MultiPath Components (MPCs). Moreover, UWB channels exhibit a significant fre-quency dispersion [22] due to the large transmission bandwidth employed. While irrelevant for conventional narrow-band sys-tems, this effect results in a pulse broadening and spreading of the MPC energy over multiple resolvable delay bins. These propagation mechanisms are not properly modeled by a purely sparse channel.

Recent work explores these effects. In [23], a geometry-based stochastic UWB model is proposed, consisting of a statistical model for the diffuse component. The model de-veloped in [24] combines a geometric approach to model the resolvable MPCs, and a stochastic approach to model the diffuse tail associated with each MPC. In [21], the spatial structure of the diffuse MPCs is investigated, and its pa-rameters are extracted from the measurements. In [25], the impact of diffuse scattering on the characteristics of vehicular propagation channels in highway environments is evaluated, and the Doppler frequency-delay characteristics of diffuse components are analyzed. In [26], a low-complexity model of diffuse scattering is proposed for vehicular radio channels. While these prior models were targeted towards performance assessment, herein we develop a simplified UWB channel model suitable for channel estimation purposes and estimator analysis.

Exploitation of structure in channel models can lead to estimation strategies with strong performance, in [27], a Max-imum Likelihood (ML) estimator is designed which exploits the clustered structure of the UWB channel. In [28], a joint channel estimation and decoding technique for Bit-Interleaved Coded Orthogonal Frequency Division Multiplexing is de-

signed, based on a two-state Gaussian mixture prior to model the sparse/diffuse structure of the channel, and on an hidden Markov prior to model clustering among the large taps. Therein, more structure is assumed, *e.g.*, clustering of the taps, and further the scheme is semi-blind. In [29], an ML framework is developed for parameter estimation in multi-dimensional channel sounding. Therein, the channel comprises a deterministic component, resulting from specular reflection, and a stochastic component modeling diffuse scattering.

Our contributions are as follows: in Part I, based on the analysis of the propagation mechanisms peculiar to UWB systems, we present a novel Hybrid Sparse/Diffuse (HSD) UWB channel model [30]. In particular, we propose statistical models for the sparse and diffuse components. We identify three physically motivated scenarios that differ in the amount of side information available at the receiver (*e.g.*, channel sparsity level, Power Delay Profile (PDP) of the diffuse or sparse component). For each scenario, Bayesian channel estimators are derived. In particular, we propose the *Generalized MMSE* (GMMSE) and the *Generalized Thresholding* (GThres) estimators, for the scenario where the statistics of the specular coefficients are unknown. We also design an Expectation-Maximization (EM) algorithm for the PDP estimation of the diffuse component, which exploits the structure of the PDP over the channel delay dimension to enhance the estimation accuracy.

The proposed algorithms are compared to unconstrained estimators, which do not exploit the structure of the UWB channel, and conventional sparse estimators, which, on the other hand, ignore the diffuse component of the channel. It is shown, by numerical results, that the new channel estimation methods considerably improve the Mean-Squared Error (MSE) accuracy and the Bit Error Rate (BER) performance, thus suggesting the importance of a proper model for the UWB channel. Specifically, a purely sparse estimator, by ignoring the diffuse component, is not able to capture important phenomena in UWB, *e.g.*, pulse distortion [31] and diffuse scattering [22], thus failing to accurately estimate the channel. Moreover, it is shown that it is beneficial to be conservative in the estimation of the sparse component of the channel, by assuming that the sparse component is sparser than it actually is.

In Part II of this paper [32], we present an asymptotic MSE analysis of the GMMSE and the GThres estimators, in the regions of high and low Signal to Noise Ratios (SNR). Moreover, we validate the simplified HSD channel model and the channel estimation strategies proposed in Part I, based on a realistic UWB channel model developed in [24]. We argue that the HSD model, despite its simplicity, can effectively capture important UWB propagation mechanisms, such as fine delay resolution, scattering from rough surfaces and frequency dispersion. Moreover, due to its hybrid structure, the HSD model is robust and covers a wide range of practical scenarios, where the channel exhibits either a sparse, diffuse or hybrid nature.

Part I of this paper is organized as follows: in Section II, we overview the UWB propagation mechanisms. In Section III, we present the system model and the HSD channel model. In Section IV, we derive channel estimation strategies for the HSD channel. In Section V, we present an EM algorithm for the PDP estimation of the diffuse component. In Section VI, we provide simulation results and we compare the performance of the estimators. Finally, Section VII concludes the paper.

Notation: We use lower-case bold letters for column vectors ($\mathbf{a}$), and upper-case bold letters for matrices ($\mathbf{A}$). The scalar $\mathbf{a}_k$ (or $\mathbf{a}(k)$) denotes the kth entry of vector $\mathbf{a}$, and $\mathbf{A}_{k,j}$ (or $\mathbf{A}(k,j)$) denotes the (k,j)th entry of matrix $\mathbf{A}$. A positive definite (positive semi-definite) matrix $\mathbf{A}$ is denoted by $\mathbf{A} \succ 0$ ($\mathbf{A} \succeq 0$). The transpose, complex conjugate of $\mathbf{A}$ is denoted by $\mathbf{A}^*$. We define the square root of $\mathbf{A} \succeq 0$ with eigenvalue decomposition $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^*$ as $\sqrt{\mathbf{A}} = \mathbf{U}\sqrt{\mathbf{D}}\mathbf{U}^*$. The $K \times K$ unit matrix is defined as $\mathbf{I}_K$. The vector $\mathbf{a} \odot \mathbf{b}$ is the component-wise (Schur) product of vectors $\mathbf{a}$ and $\mathbf{b}$. The indicator function is given by $\mathcal{I}(\cdot)$. We use $p(\cdot)$ to indicate a continuous or discrete probability distribution, and $\Pr(\cdot)$ to indicate the probability of an event. The expectation of random variable x, conditioned on y, is denoted by $\mathbb{E}[x|y]$. The Gaussian distribution with mean $\mathbf{m}$ and covariance $\mathbf{\Sigma}$ is written as $\mathcal{N}(\mathbf{m}, \mathbf{\Sigma})$, whereas the circularly symmetric complex Gaussian distribution is denoted by $\mathcal{CN}(\mathbf{m}, \mathbf{\Sigma})$;[1] the Bernoulli distribution with parameter q is denoted by $\mathcal{B}(q)$.

II. UWB CHANNEL PROPAGATION AND MODELING OVERVIEW

In this section, we overview the state of the art of UWB channel propagation and modeling. The aim is to determine an appropriate UWB channel model, which captures the main UWB propagation mechanisms. Neglecting pulse distortion [31] for simplicity, a time-varying channel in the continuous time can be represented as [33]

$$h(\tau, t) = \sum_l a_l(t)\delta(\tau - \tau_l(t)), \qquad (1)$$

where $\delta(\cdot)$ is the Kronecker delta function, t is the time dimension and τ is the channel delay. The sum is over the MPCs, with time-varying amplitude $a_l(t)$ and delay $\tau_l(t)$. If we consider a UWB system with center frequency f_0 and transmission bandwidth W, the discrete baseband time-varying impulse response of the channel is given by

$$h_{bb}(n, t) = \sum_l a_l(t)e^{-i2\pi f_0 \tau_l(t)}\text{sinc}(n - W\tau_l(t)), \quad (2)$$

where $\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$ is the sinc function, and $n \in \mathbb{Z}$ is the discrete channel delay. Due to the large transmission bandwidth of UWB systems, MPCs arising from reflections and scattering in the environment spaced apart (in the delay domain) by more than $\frac{1}{W}$, which is typically of the order of a fraction of a ns, can be resolved at the receiver. Then, by neglecting leakage effects due to the sampling of the sinc

---

[1]For a vector $\mathbf{x} = \mathbf{x}_R + i\mathbf{x}_I \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{x}_R = \text{Re}(\mathbf{x})$, $\mathbf{x}_I = \text{Im}(\mathbf{x})$ and $i = \sqrt{-1}$, we define the covariance matrices of its real and imaginary parts as $\mathbb{E}[\mathbf{x}_R\mathbf{x}_R^*] = \mathbb{E}[\mathbf{x}_I\mathbf{x}_I^*] = \frac{\text{Re}(\mathbf{\Sigma})}{2}$ and $\mathbb{E}[\mathbf{x}_I\mathbf{x}_R^*] = -\mathbb{E}[\mathbf{x}_R\mathbf{x}_I^*] = \frac{\text{Im}(\mathbf{\Sigma})}{2}$.

function off its peak, (2) is commonly approximated by the following *sparse* discrete baseband representation:

$$h_{bb}(n, t) \simeq \sum_l a_l(t) e^{-i2\pi f_0 \tau_l(t)} \delta\left(n - \mathrm{rd}\left(W\tau_l(t)\right)\right), \quad (3)$$

where $\mathrm{rd}(x)$ returns the closest integer to x.

However, in many practical scenarios of interest (*e.g.*, indoor environments), diffuse components, that cannot be described by the above model, arise. These are created mainly by the following phenomena: a large number of unresolved paths, diffuse scattering [22], pulse distortion resulting from the frequency dependence of the gain and efficiency of the antennas and of the dielectric or conductive materials, and diffraction effects [31]. In [23], the following frequency response has been proposed, modeling the contribution from all these effects:

$$H_{UWB}(f) = \left(\mathcal{S}_{LOS}(f) + \sum_k \mathcal{S}_k(f) + \mathcal{D}(f)\right) \frac{f^{-m}}{F}, \quad (4)$$

where f is frequency. In particular, we recognize in $\mathcal{S}_{LOS}(f)$ and $\sum_k \mathcal{S}_k(f)$ the contributions from the line of sight and the resolvable MPCs, respectively, *i.e.*, the MPCs whose inter-arrival time is larger than $\frac{1}{W}$, giving rise to a *sparse* component in the time domain. The term $\mathcal{D}(f)$ represents the *diffuse* component due to multipath interference, and is associated with the non-resolvable MPCs. Finally, $\frac{f^{-m}}{F}$ models the frequency distortion of the channel, where F is a normalization factor and m is the frequency decay exponent. Note that, in this model, the diffuse component is independent of the realization of the discrete MPCs, while, in contrast, the work in [24] models the diffuse component as a diffuse tail associated with each specular component.

It is worth noting that the level of channel diffuseness or sparseness depends primarily on two factors: the transmission bandwidth and the environment. In fact, the larger the transmission bandwidth, the finer the delay resolution at the receiver, and the sparser the channel is expected to be. On the other hand, an environment with many scatterers or rough surfaces, *e.g.*, an indoor scenario or WBANs, is more likely to give rise to a dense channel, due to the richer interaction among the MPCs. Dense channels have been observed, *e.g.*, in gas stations [23], industrial [34], office [10] and vehicular environments [25]. We thus expect a dense or hybrid channel representation to be relevant in these or similar scenarios.

Spatio-temporal scale of variation in the UWB channel

We now consider the spatio-temporal variation of the channel, due to the relative motion of the scatterers, receiver and transmitter in the environment. For ease of exposition, we consider movement of the receiver only. Ignoring Doppler effects, which are left for future investigations, the channel time-variations affect the amount of side-information available at the receiver for the purpose of channel estimation, as discussed in Section III-B.

From the discrete baseband model (2), the phase[2] variation of the lth MPC over a time-interval Δt is given by

---

[2]Note that "phase" is a narrow-band concept and can be used only as an approximation in UWB systems, in particular when the lower band edge is at $f = 0$.

$\Delta \phi_l \triangleq 2\pi \frac{c_0}{\lambda_0} |\tau_l(t + \Delta t) - \tau_l(t)|$, where λ_0 is the wavelength at the center frequency, and c_0 is the free space speed of light. Therefore, a significant phase variation (*e.g.*, by more than $\frac{\pi}{2}$) occurs when $\Delta \phi_l > \frac{\pi}{2}$. This quantity corresponds, in the spatial domain, to a wavelength or a fraction of it. Therefore, phase changes are expected to occur on a very small spatio-temporal scale.

Similarly, the variation of the MPC delay, over the same time-interval Δt, is given by $\Delta \tau_l \triangleq |\tau_l(t + \Delta t) - \tau_l(t)|$. Hence, a significant variation (*e.g.*, by more than one channel delay bin, $\frac{1}{W}$) occurs when $\Delta \tau_l > \frac{1}{W}$, *i.e.*, on a spatial scale of $\frac{c_0}{W}$ or roughly a number of wavelengths in the range $[0.5, 5]$, depending on the value of the transmission bandwidth W, relative to the center frequency f_0.

Finally, significant variations of the MPC amplitude $a_l(t)$, due to shadowing effects, typically correspond to a spatial scale of several wavelengths.

Note that, due to mutual interference of the unresolvable MPCs contributing to the same tap location, changes in the amplitude of the diffuse components arise over the same spatio-temporal scale as the phase changes of the MPCs (small scale fading). On the other hand, the amplitude of the resolvable MPCs vary over a much larger spatio-temporal scale (large scale fading).

Remark 1. It is worth noting that the side-lobes of the sinc function in (2) introduce faster time-variations of the amplitude of the resolvable MPCs than the large-scale fading, over the same spatio-temporal scale as the delay variations, and account for the leakage of the MPC energy over nearby channel taps. However, this phenomenon is limited, and can be quantified as follows. The most severe leakage occurs when the MPC arrives exactly in the middle between two sampling times, in which case most of the energy $(2\mathrm{sinc}(0.5)^2 \simeq 80\%)$ is spread equally between two nearby taps (each with amplitude $1 - \mathrm{sinc}(0.5) \simeq 37\%$ smaller than in the no leakage scenario, where the MPC delay is exactly an integer number of the sampling period), and the remaining 20% is leaked among the nearby taps. Therefore, the side-lobes of the sinc function account for at most a 37% variation of the amplitude of the main MPC tap in (2). The problem of MPCs falling in between two sample points can be modeled as a basis mismatch [35].

In the next section, we present the observation and the channel models. In particular, in Section III-A we present the HSD model, which represents a simplification with respect to other models presented in the literature, *e.g.*, (4), but at the same time it captures the main propagation phenomena of the UWB channel discussed in this section: resolvable MPCs, modeled by the sparse vector (3), unresolvable MPCs, diffuse scattering and frequency distortion, modeled by a random, dense vector. Also, based on the analysis of the spatio-temporal scale of variation in the UWB channel, in Section III-B we discuss different practical scenarios, differing in the side-information available at the receiver for the purpose of channel estimation, which enables more accurate estimation techniques.

III. SYSTEM MODEL

We consider a single-user UWB system. The source transmits a sequence of $M = N + L - 1$ pilot symbols, $x(k), k = -(L-1), \ldots, N-1$, over a channel $h(l), l = 0, \ldots, L-1$ with known delay spread $L \geq 1$. The received, discrete time, baseband signal over the corresponding observation interval of duration N is given by

$$y(k) = \sum_{l=0}^{L-1} h(l)x(k-l) + w(k), \quad k = 0, \ldots, N-1, \quad (5)$$

where $w(k) \in \mathcal{CN}(0, \sigma_w^2)$ is i.i.d. noise.

If we collect the N received samples in the column vector $\mathbf{y} = [y(0), y(1), \ldots, y(N-1)]^T$, we have the following matrix representation:

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{w}. \quad (6)$$

Above, $\mathbf{X} \in \mathbb{C}^{N \times L}$ is the $N \times L$ Toeplitz matrix associated with the pilot sequence, having the vector of the transmitted pilot sequence $[x(-k), x(-k+1), \ldots, x(-k+N-1)]^T, k = 0, \ldots, L-1$, as its kth column, $\mathbf{h} = [h(0), h(1), \ldots, h(L-1)]^T \in \mathbb{C}^L$ is the column vector of channel coefficients, and $\mathbf{w} = [w(0), w(1), \ldots, w(N-1)]^T \sim \mathcal{CN}(\mathbf{0}, \sigma_w^2 \mathbf{I}_N)$ is the noise vector.

We assume $\mathbf{X}^*\mathbf{X} \succ 0$, so that the LS estimate $\mathbf{h}_{\text{LS}} = (\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*\mathbf{y}$ is a sufficient statistic [36] for the channel. Therefore, without loss of generality for the purpose of channel estimation, we consider the observation model

$$\mathbf{h}_{\text{LS}} = (\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*\mathbf{y} = \mathbf{h} + (\mathbf{X}^*\mathbf{X})^{-1}\mathbf{X}^*\mathbf{w} = \mathbf{h} + \sqrt{\mathbf{S}}^{-1}\mathbf{n}, \quad (7)$$

where we have defined the SNR matrix $\mathbf{S} = \frac{\mathbf{X}^*\mathbf{X}}{\sigma_w^2} \succ 0$, and $\mathbf{n} = \frac{1}{\sigma_w^2}\sqrt{\mathbf{S}}^{-1}\mathbf{X}^*\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_L)$. With a slight abuse of notation, we will refer to the LS estimate $\mathbf{h}_{\text{LS}}$ as the "observed" sequence. Moreover, we assume that the pilot sequence is orthogonal, so that $\mathbf{S}$ is a diagonal matrix. Then, the noise vector $\sqrt{\mathbf{S}}^{-1}\mathbf{n}$ in the LS estimate has independent entries. This assumption greatly simplifies the channel estimation problem. In fact, when the channel has independent entries over the delay dimension (this is the case for the HSD model we develop), a per-tap estimation approach, rather than a joint one, is optimal. The case with non-orthogonal pilot sequences is considered in Part II of the paper.

A. HSD Channel Model

The channel $\mathbf{h}$ follows the HSD model developed in [30],

$$\mathbf{h} = \mathbf{a}_s \odot \mathbf{c}_s + \mathbf{h}_d, \quad (8)$$

where the terms $\mathbf{a}_s \odot \mathbf{c}_s \in \mathbb{C}^L$ and $\mathbf{h}_d \in \mathbb{C}^L$ represent the *sparse*[3] and the *diffuse components*, respectively.

In particular, $\mathbf{a}_s \in \{0,1\}^L$ is the *sparsity pattern*, which is equal to one in the positions of the specular MPCs, and equal to zero otherwise; its entries are drawn i.i.d. from $\mathcal{B}(q)$, where

$q \ll 1$ so as to enforce sparsity. In the sequel, we refer to the non-zero entries of $\mathbf{a}_s \odot \mathbf{c}_s \in \mathbb{C}^L$ as *active sparse components*.

The vector of sparse coefficients, $\mathbf{c}_s \in \mathbb{C}^L$, is drawn from the continuous probability distribution $p(\mathbf{c}_s)$, with second order moment $\mathbb{E}[\mathbf{c}_s\mathbf{c}_s^*] = \mathbf{\Lambda}_s$, where $\mathbf{\Lambda}_s$ is a diagonal matrix with entries given by the PDP $\mathbf{\Lambda}_s(k,k) = \mathcal{P}_s(k), k = 0, \ldots, L-1$.[4]

Finally, we use the Rayleigh fading assumption for the diffuse component, $\mathbf{h}_d \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Lambda}_d)$, where $\mathbf{\Lambda}_d$ is diagonal, with entries given by the PDP $\mathbf{\Lambda}_d(k,k) = \mathcal{P}_d(k), k = 0, \ldots, L-1$.

Remark 2. The Bernoulli model for $\mathbf{a}_s$ can be interpreted as a discretized Saleh-Valenzuela model [37]. In fact, according to the latter, the inter-arrival times of the specular components have an exponential distribution, whose discrete counterpart is the geometric distribution. This in turn can be interpreted as the inter-arrival time of two consecutive "1"s in a sequence of i.i.d. Bernoulli draws.

Remark 3. In general, the Rayleigh fading assumption does not hold for the distribution of the sparse coefficients $p(\mathbf{c}_s)$ (unlike the diffuse ones), since only very few propagation paths contribute to an active tap in the sparse channel, thus limiting the validity of the central limit theorem. Channel measurement campaigns have shown that the large scale fading, affecting the amplitude of the entries of $\mathbf{c}_s$, can be modeled by a log-normal distribution [23]. However, for the sake of analytical tractability, in the following we either treat $\mathbf{c}_s$ as a deterministic unknown vector, when its second order moment $\mathbf{\Lambda}_s$ is unknown, or we treat it using the Gaussian approximation, when knowledge of $\mathbf{\Lambda}_s$ is available.

Remark 4. Note that in [23] the amplitudes of the diffuse coefficients are modeled by a Weibull distribution, with a delay dependent shape parameter $\sigma < 2$, and approach the Rayleigh fading distribution ($\sigma = 2$) only for large excess delays. This distribution represents a fading worse than Rayleigh. However, we adopt the Rayleigh fading approximation for simplicity and tractability. Also, the side-lobes of the sinc function in (2) introduce correlation in the delay domain, which is not accounted for under the Rayleigh fading model. This is a common assumption in standard cellular channel models, where measurements have well established the independence of fading on different taps [38].

Despite its simplicity, we argue that the HSD model is able to capture the main UWB propagation mechanisms discussed in Section II. In fact, the resolvable specular components and the fine delay resolution are appropriately modeled by the sparse vector $\mathbf{a}_s \odot \mathbf{c}_s$, whereas diffuse scattering, multipath interference and the frequency distortion are approximated by the diffuse component $\mathbf{h}_d$. This is confirmed by simulation results in Part II of the paper, where we validate the proposed HSD model based on a realistic channel emulator [24].

---

[3]In the following, we use the terms *sparse*, *specular* and *resolvable MPCs* interchangeably. In fact, the physical specular components (resolvable MPCs) of the channel can be modeled and represented by a sparse vector (3).

[4]It is worth noting that this is not a PDP in the traditional sense, but rather represents the power profile of the active sparse components, as a function of the delay.

B. Channel Estimation scenarios

The HSD model is described by a number of deterministic parameters, namely, the sparsity level q, the PDP of the diffuse component $\mathcal{P}_d$ and the PDP of the sparse component $\mathcal{P}_s$. Accurate knowledge about some or all of these parameters may not be available at the receiver, depending on a number of factors, most importantly the length of the interval over which the channel is observed, and the dynamics of the environment.

Let $\left\{ \mathbf{h}^{(j)} = \mathbf{a}_s^{(j)} \odot \mathbf{c}_s^{(j)} + \mathbf{h}_d^{(j)}, j = 0, \ldots, N_{\mathrm{ch}} - 1 \right\}$ be a sequence of N_{ch} channel realizations, spaced apart in time by Δt, corresponding to a spatial separation by $\simeq \lambda_0$, resulting from the relative motion of the receiver with respect to the scatterers and the transmitter position. Under this assumption, the samples of the diffuse component $\left\{ \mathbf{h}_d^{(j)}, j \geq 0 \right\}$ can be approximated as drawn independently from $\mathcal{CN}\left(\mathbf{0}, \mathbf{\Lambda}_d\right)$, due to multipath interference (Section II).

On the other hand, the positions of the active sparse coefficients $\left\{ \mathbf{a}_s^{(j)}, j = 0, \ldots, N_{\mathrm{ch}} - 1 \right\}$ exhibit correlation with each other. In fact, as pointed out in Section II, a variation of the delay associated with a specular MPC by one channel delay bin occurs over a spatial scale of the order of $\frac{c_0}{W\lambda_0} \in [0.5, 5]$ wavelengths. Therefore, the positions of the "1"s observed in subsequent realizations of the sparsity pattern $\mathbf{a}_s^{(j)}$ are bound not to vary appreciably over a large spatial scale, relative to the wavelength.

A similar consideration holds for the amplitudes of the specular components (i.e., the active sparse components in the vector $\mathbf{a}_s^{(j)} \odot \mathbf{c}_s^{(j)}$), which vary according to the large scale fading, i.e., over a relatively large spatial scale, compared to the rate of variation of the diffuse component (however, the side-lobes of the sinc function account for a 37% variation in the amplitude on same spatial scale as the delay variations, as discussed in Remark 1 of Section II).

This correlation structure, i.e., slow amplitude and delay variations, may be exploited to enhance the estimation accuracy of the sparse component $\mathbf{a}_s^{(j)} \odot \mathbf{c}_s^{(j)}$, by tracking the position and amplitude of the resolvable MPCs over subsequent observation windows. However, in this work we consider estimation of $\mathbf{a}_s^{(j)} \odot \mathbf{c}_s^{(j)}$ based on either only one channel realization, or the statistics of the ensemble of realizations that ignores the information about the temporal sequence in which the realizations occur.

We consider three different physical scenarios, dictated by the length of the observation window N_{ch}.

C. Single Snapshot of the channel

If a very short observation window is available ($N_{\mathrm{ch}} = 1$, or less than a wavelength in the spatial domain), averaging over the small scale and the large scale fading is not possible. Under this assumption, statistical information about the channel cannot be reliably collected, and the channel can reasonably be considered a deterministic and unknown vector. In this case, an LS estimate $\mathbf{h}_{\mathrm{LS}}$ may be employed. In the absence of prior information about the channel, this is a robust approach for channel estimation.

	Scenario	sparsity q	PDP $\mathbf{\Lambda}_s$	PDP $\mathbf{\Lambda}_d$
S0	Single snapshot (unstructured)	unknown	unknown	unknown
S1	Single snapshot (PDP structure exploited)	unknown	unknown	known
S2	Avg. over Small scale fading	known	unknown	known
S3	Avg. over Small&Large scale fading	known	known	known

Alternatively, we may exploit further structure of the channel, e.g., exponential PDP of the diffuse component, to average the fading over the delay dimension rather than over time. As shown in Section V, under this assumption, an accurate PDP estimate of $\mathbf{h}_d^{(j)}$ is possible even in the extreme case $N_{\mathrm{ch}} = 1$. We may then assume that the PDP of $\mathbf{h}_d^{(j)}$ is known at the receiver, whereas the vector $\mathbf{c}_s^{(j)}$ is modeled as deterministic and unknown.

As to the sparsity level q, letting N_{sc} be the number of resolvable scatterers, we have $q \simeq \frac{N_{\mathrm{sc}}}{L}$. This number is not expected to vary appreciably over a relatively long observation interval, and can be estimated by counting the number of resolvable MPCs which can be distinguished from the noise plus diffuse background. However, an accurate estimate of N_{sc} is obtained by averaging the small-scale fading and the noise over subsequent channel realizations. Hence, we model q as a deterministic and unknown parameter.

D. Averaging over the Small scale fading

When a larger observation window is available (corresponding, in the spatial domain, to a few wavelengths, $N_{\mathrm{ch}} > 1$), averaging over the small scale fading (amplitude and phase of the diffuse component) may be possible. In this case, the PDP of $\mathbf{h}_d^{(j)}$ can be estimated accurately by averaging over subsequent realizations of the fading process.

In this scenario, we assume that $\mathbf{\Lambda}_d$ is perfectly known at the receiver. This knowledge can be exploited by performing a Minimum MSE (MMSE) estimate of $\mathbf{h}_d^{(j)}$, which achieves a better accuracy than LS. On the other hand, due to the inability to average over the large-scale fading, which affects the variation of the amplitude of the resolvable MPCs, $\mathbf{c}_s^{(j)}$ is treated as deterministic and unknown.

E. Averaging over the Small scale and the Large scale fading

Finally, when the observation interval spans several wavelengths ($N_{\mathrm{ch}} \gg 1$), averaging over the large scale, other than the small scale fading, is possible.

In this scenario, we assume that $\mathbf{\Lambda}_d$, $\mathbf{\Lambda}_s$ and q are known at the receiver. This information can be exploited to compute a linear-MMSE estimate of $\mathbf{c}_s^{(j)}$ and $\mathbf{h}_d^{(j)}$, thus enhancing the estimation accuracy over an unstructured estimate (e.g., LS).

The main scenarios of interest, and the side information at the receiver, are listed in Table I. Scenario S0 will not be further considered, since the channel is estimated via LS. In the next section, we design channel estimators for the other scenarios.

IV. Hybrid Sparse/Diffuse channel estimation

In this section, we design channel estimation strategies based on the HSD model. In particular, for scenario S3, we propose an MMSE estimator in Section IV-A. For scenarios S1 and S2, we propose the GMMSE and GThres estimators in Section IV-B.

A. MMSE Estimator

When $\mathbf{\Lambda}_d$, $\mathbf{\Lambda}_s$ and q are known, we can devise an MMSE estimator. By exploiting the orthogonality of the pilot sequence, we can use a per-tap estimation approach. The MMSE estimate of the kth delay bin is given by the posterior mean of the channel, given the observed channel sample $\mathbf{h}_{\mathrm{LS}}(k)$ [36],

$$\hat{\mathbf{h}}_{\mathrm{MMSE}}(k) = \Pr(\mathbf{a}_s(k)\!=\!0|\mathbf{h}_{\mathrm{LS}}(k))\mathbb{E}\left[\mathbf{h}_d(k)|\mathbf{h}_{\mathrm{LS}}(k), \mathbf{a}_s(k)\!=\!0\right]$$
$$+ \Pr(\mathbf{a}_s(k)\!=\!1|\mathbf{h}_{\mathrm{LS}}(k))\mathbb{E}\left[\mathbf{c}_s(k) + \mathbf{h}_d(k)|\mathbf{h}_{\mathrm{LS}}(k), \mathbf{a}_s(k)\!=\!1\right], \tag{9}$$

where we have conditioned on the realization of the sparsity bit $\mathbf{a}_s(k)$. In particular, the sum is over the posterior mean under the two hypotheses $\mathbf{a}_s(k) = 1$ and $\mathbf{a}_s(k) = 0$, weighted by their posterior distribution $\Pr(\mathbf{a}_s(k) = 1|\mathbf{h}_{\mathrm{LS}}(k))$ and $\Pr(\mathbf{a}_s(k) = 0|\mathbf{h}_{\mathrm{LS}}(k))$, respectively.

In order to compute (9), we use the circular Gaussian approximation for $\mathbf{c}_s(k)$.[5] Under this assumption, $\mathbf{h}_{\mathrm{LS}}(k)|\{\mathbf{a}_s(k) = a, \mathbf{h}(k)\} \sim \mathcal{CN}(\mathbf{h}(k), 1/\mathbf{S}_{k,k})$, whereas the channel sample $\mathbf{h}(k)$, conditioned on $\mathbf{a}_s(k) = a$, is distributed as $\mathbf{h}(k)|\mathbf{a}_s(k) = a \sim \mathcal{CN}(0, \mathbf{a}_s(k)\mathcal{P}_s(k) + \mathcal{P}_d(k))$. Then, $\mathbf{h}(k)|\{\mathbf{h}_{\mathrm{LS}}(k), \mathbf{a}_s(k) = a\} \sim \mathcal{CN}(\mathbf{m}(a), \Sigma)$, with posterior mean $\mathbf{m}(a) = \mathbb{E}[\mathbf{h}(k)|\mathbf{h}_{\mathrm{LS}}(k), \mathbf{a}_s(k) = a]$ given by

$$\mathbf{m}(a) = \frac{a\mathcal{P}_s(k) + \mathcal{P}_d(k)}{1/\mathbf{S}_{k,k} + a\mathcal{P}_s(k) + \mathcal{P}_d(k)}\mathbf{h}_{\mathrm{LS}}(k). \tag{10}$$

From (9), we finally obtain

$$\hat{\mathbf{h}}_{\mathrm{MMSE}}(k) = \Pr(\mathbf{a}_s(k) = 0|\mathbf{h}_{\mathrm{LS}}(k))\frac{\mathbf{S}_{k,k}\mathcal{P}_d(k)}{1 + \mathbf{S}_{k,k}\mathcal{P}_d(k)}\mathbf{h}_{\mathrm{LS}}(k)$$
$$+ \Pr(\mathbf{a}_s(k) = 1|\mathbf{h}_{\mathrm{LS}}(k))\frac{\mathbf{S}_{k,k}(\mathcal{P}_s(k) + \mathcal{P}_d(k))}{1 + \mathbf{S}_{k,k}(\mathcal{P}_s(k) + \mathcal{P}_d(k))}\mathbf{h}_{\mathrm{LS}}(k),$$

where, from Bayes' rule and $\mathbf{a}_s(k) \sim \mathcal{B}(q)$, letting $Q_k = \frac{\mathbf{S}_{k,k}\mathcal{P}_s(k)}{1+\mathbf{S}_{k,k}\mathcal{P}_d(k)}$, we have

$$\Pr(\mathbf{a}_s(k) = 1|\mathbf{h}_{\mathrm{LS}}(k)) = \left(1 + \frac{1-q}{q}\frac{p(\mathbf{h}_{\mathrm{LS}}(k)|\mathbf{a}_s(k) = 0)}{p(\mathbf{h}_{\mathrm{LS}}(k)|\mathbf{a}_s(k) = 1)}\right)^{-1}$$
$$= \frac{1}{1 + \frac{1-q}{q}(1 + Q_k)\exp\left\{-\frac{Q_k}{1+Q_k}\frac{\mathbf{S}_{k,k}|\mathbf{h}_{\mathrm{LS}}(k)|^2}{1+\mathbf{S}_{k,k}\mathcal{P}_d(k)}\right\}}. \tag{11}$$

[5]As discussed in Remark 3 in Section III, the large scale fading is commonly modeled by a log-normal prior; however, due to the difficulty in handling it, the Rayleigh fading approximation is used, thus leading to the classical linear MMSE estimator. We have numerically evaluated the performance loss incurred by using the linear MMSE estimator over an MMSE estimator based on the log-normal prior, for the simple scalar model $y = c_s + n$, where $c_s = e^{\nu_s + i\theta_s}$, with $\nu_s \sim \mathcal{N}(0, 1)$ and θ_s uniform in $[0, 2\pi]$, is the channel coefficient with log-normal amplitude, $n \sim \mathcal{CN}(0, \sigma_w^2)$ is the noise; we found that the performance loss is at most $1.67\,\mathrm{dB}$, at $0\,\mathrm{dB}$ SNR level.

B. Generalized MMSE (GMMSE) and Generalized Thresholding (GThres) Estimators

In this section, we develop estimators for scenarios S1 and S2. In particular, $\mathbf{\Lambda}_d$ is assumed to be known at the receiver, whereas $\mathbf{c}_s$ is treated as a deterministic and unknown vector. The case where $\mathbf{\Lambda}_d$ is unknown and is estimated from the observed sequence is treated in Section V.

For generality, we assume that the sparsity level q is unknown, and an estimate $\tilde{q}$ of q, which might be different from the real q, is used in the estimation phase. This choice represents a generalization with respect to [30], where the true sparsity level q is used. We will show by simulation, and by analysis in Part II of the paper, that assuming a sparsity level $\tilde{q} < q$ often improves the estimation accuracy, thus implying that knowledge of this parameter is not crucial to the performance of the estimators.

We proceed as follows. $\mathbf{c}_s$ is estimated by Maximum Likelihood (ML). Then, the estimate $\hat{\mathbf{c}}_s$ is used to perform either an MMSE or a Maximum A Posteriori (MAP) estimate of the sparsity pattern $\mathbf{a}_s$, denoted by $\hat{\mathbf{a}}_s$, assuming the prior $\mathbf{a}_s \sim \mathcal{B}(\tilde{q})^L$. We refer to these estimators as the GMMSE and GThres estimators, respectively. Finally, the diffuse component $\mathbf{h}_d$ is estimated via MMSE, based on the residual estimation error $\mathbf{h}_{\mathrm{LS}} - \hat{\mathbf{a}}_s \odot \hat{\mathbf{c}}_s$.

The ML estimate of $\mathbf{c}_s(k)$ is given by

$$\hat{\mathbf{c}}_s(k) = \arg\min_{\mathbf{c}_s(k)\in\mathbb{C}}\{-\ln p(\mathbf{h}_{\mathrm{LS}}(k)|\mathbf{c}_s(k), \mathbf{a}_s(k) = 1)\}$$
$$= \mathbf{h}_{\mathrm{LS}}(k), \tag{12}$$

where we have used the fact that, when conditioned on $\mathbf{a}_s(k) = 0$, the observation $\mathbf{h}_{\mathrm{LS}}(k)$ does not depend on $\mathbf{c}_s(k)$, and $\mathbf{h}_{\mathrm{LS}}(k)|\{\mathbf{c}_s(k), \mathbf{a}_s(k) = 1\} \sim \mathcal{CN}\left(\mathbf{c}_s(k), [\mathbf{S}_{k,k}]^{-1} + \mathcal{P}_d(k)\right)$. We thus obtain $\hat{\mathbf{c}}_s = \mathbf{h}_{\mathrm{LS}}$.

Using the estimate $\hat{\mathbf{c}}_s(k) = \mathbf{h}_{\mathrm{LS}}(k)$ and conditioning on $\mathbf{a}_s(k) = a$, $a \in \{0, 1\}$, the MMSE estimate of the diffuse component $\mathbf{h}_d(k)$ is given by

$$\hat{\mathbf{h}}_d^{(a)}(k) = \mathbb{E}[\mathbf{h}_d(k)|\mathbf{h}_{\mathrm{LS}}(k), \hat{\mathbf{c}}_s(k), \hat{\mathbf{a}}_s(k) = a]$$
$$= \frac{\mathbf{S}_{k,k}\mathcal{P}_d(k)}{1 + \mathbf{S}_{k,k}\mathcal{P}_d(k)}(1 - a)\mathbf{h}_{\mathrm{LS}}(k). \tag{13}$$

Finally, by combining the estimates $\hat{\mathbf{a}}_s$, $\hat{\mathbf{c}}_s$ and $\hat{\mathbf{h}}_d^{(a)}$, the overall HSD estimate is given by

$$\hat{\mathbf{h}}(k) = \hat{\mathbf{a}}_s(k)\mathbf{h}_{\mathrm{LS}}(k) + (1 - \hat{\mathbf{a}}_s(k))\frac{\mathbf{S}_{k,k}\mathcal{P}_d(k)}{1 + \mathbf{S}_{k,k}\mathcal{P}_d(k)}\mathbf{h}_{\mathrm{LS}}(k). \tag{14}$$

We now develop the MMSE and MAP estimates of $\mathbf{a}_s(k)$.

1) Generalized MMSE Estimator:
The MMSE estimate of the sparsity bit $\mathbf{a}_s(k)$ is given by

$$\hat{\mathbf{a}}_s^{(\mathrm{GMMSE})}(k) = \mathbb{E}[\mathbf{a}_s(k)|\mathbf{h}_{\mathrm{LS}}(k), \hat{\mathbf{c}}_s(k)]$$
$$= \Pr(\mathbf{a}_s(k) = 1|\mathbf{h}_{\mathrm{LS}}(k), \hat{\mathbf{c}}_s(k)). \tag{15}$$

Using Bayes' rule, $\hat{\mathbf{c}}_s(k) = \mathbf{h}_{\mathrm{LS}}(k)$, and assuming $\mathbf{a}_s(k) \sim \mathcal{B}(\tilde{q})$, we have

$$\hat{\mathbf{a}}_s^{(\mathrm{GMMSE})}(k) = \frac{1}{1 + e^\alpha \exp\left\{-\frac{\mathbf{S}_{k,k}|\mathbf{h}_{\mathrm{LS}}(k)|^2}{1+\mathbf{S}_{k,k}\mathcal{P}_d(k)}\right\}}, \tag{16}$$

where we have defined $\alpha = \ln\left(\frac{1-\tilde{q}}{\tilde{q}}\right)$.

2) Generalized Thresholding Estimator:

Using Bayes' rule and the ML estimate $\hat{\mathbf{c}}_s(k) = \mathbf{h}_{\text{LS}}(k)$, the MAP estimate of $\mathbf{a}_s$ is given by

$$\hat{\mathbf{a}}_s^{(\text{GThres})}(k) = \arg\max_{a\in\{0,1\}} \{\ln\Pr\left(\mathbf{a}_s(k) = a|\mathbf{h}_{\text{LS}}(k), \hat{\mathbf{c}}_s(k)\right)\}$$

$$= \arg\min_{a\in\{0,1\}} \left\{(1-a)\frac{\mathbf{S}_{k,k}|\mathbf{h}_{\text{LS}}(k)|^2}{1+\mathbf{S}_{k,k}\mathcal{P}_d(k)} + a\ln\left(\frac{1-\tilde{q}}{\tilde{q}}\right)\right\}$$

$$= \mathcal{I}\left(|\mathbf{h}_{\text{LS}}(k)|^2 \geq \alpha\left(1/\mathbf{S}_{k,k} + \mathcal{P}_d(k)\right)\right). \tag{17}$$

This solution consists in a thresholding of the LS estimate, hence the name *Generalized Thresholding estimator*, where the diffuse component represents noise for the estimation of the sparse coefficients. For this reason, the threshold is proportional, by a factor α, to the sum of the noise strength $1/\mathbf{S}_{k,k}$ and the power of the diffuse component $\mathcal{P}_d(k)$.

It is worth noting that, if $\alpha \leq 0$ (*i.e.*, $\tilde{q} \geq \frac{1}{2}$), then $\hat{\mathbf{a}}_s^{(\text{GThres})}(k) = 1$, and the GThres estimator trivially reduces to the LS solution.

V. STRUCTURED PDP ESTIMATION OF THE DIFFUSE COMPONENT

In the derivation of the GMMSE and GThres estimators in the previous section, we have assumed that the PDP of the diffuse component $\mathbf{h}_d$ is perfectly known at the receiver. However, in a practical system, this is unknown, and therefore needs to be estimated.

Herein, we develop a structured estimate of the PDP $\mathcal{P}_d$, when the observation interval is too short to allow time-averaging over the small scale fading. By exploiting prior information about the structure of the PDP, we can average the small scale fading over the delay dimension, rather than over subsequent realizations of the fading process, thus enhancing the estimation accuracy.

We assume an exponential PDP model [23], [38], [39] $\mathcal{P}_d(k) = \beta e^{-\omega k}$, $k = 0,\ldots,L-1$, where the deterministic, unknown parameters $\beta \geq 0$ and $\omega \geq 0$ represent the relative power and the decay rate of the PDP, respectively. We derive an ML estimate of these parameters, using the EM algorithm (the general EM framework is presented in, *e.g.*, [40]). For simplicity, we assume a single channel snapshot. However, the following derivation can be extended to include a sequence of channel realizations. Moreover, we treat the vector $\mathbf{c}_s$ as a deterministic unknown parameter, and we assume a sparsity level $\tilde{q}$ (possibly, $\neq q$), which is consistent with the design choice of the GMMSE and GThres estimators.

Let the HSD channel and the observed sequence be given by (8) and (7), respectively. From (8), if $\mathbf{a}_s(k) = 1$, then $\mathbf{h}_{\text{LS}}(k) = \mathbf{c}_s(k) + \mathbf{h}_d(k) + \sqrt{\mathbf{S}_{k,k}}^{-1}\mathbf{n}(k)$. In this case, since $\mathbf{c}_s(k)$ is a deterministic, unknown parameter, the observed sample $\mathbf{h}_{\text{LS}}(k)$ does not provide statistical information to estimate the diffuse component (hence, its power). In fact, the ML estimate of $\mathbf{c}_s(k)$ is $\hat{\mathbf{c}}_s(k) = \mathbf{h}_{\text{LS}}(k)$ (12). The estimated contribution from the noise and the diffuse component is then $\mathbf{h}_{\text{LS}}(k) - \hat{\mathbf{c}}_s(k) = 0$, and the estimate of $\mathbf{h}_d(k)$, given by (13), is forced to zero. Therefore, the observations corresponding to the active sparse components should be neglected in the estimation process.

Conversely, all the statistical information to estimate the PDP parameters ω and β is contained in the vector $(\mathbf{1}-\mathbf{a}_s)\odot \mathbf{h}_{\text{LS}} = (\mathbf{1}-\mathbf{a}_s)\odot(\mathbf{h}_d + \sqrt{\mathbf{S}}^{-1}\mathbf{n})$, which is obtained by zeroing the contribution from the active sparse components. Unfortunately, $\mathbf{a}_s$ is unknown in advance, hence it needs to be estimated from the observed sequence.

In employing the EM algorithm to estimate the PDP parameters β and ω, we assume $\mathbf{a}_s$ and $(\mathbf{1}-\mathbf{a}_s)\odot\mathbf{h}_d$ as the hidden variables. Moreover, we discard the contribution of the active sparse components to the observed sequence, as justified above. Then, letting $\hat{\beta}, \hat{\omega}$ be the current estimates of the deterministic unknown parameters β and ω, respectively, in the E-step we compute (18), where, in the last step, we have defined the posterior probability of an active sparse component

$$\hat{q}_{\text{post}}(k) = \Pr\left(\mathbf{a}_s(k) = 1|\mathbf{h}_{\text{LS}}(k), \hat{\beta}, \hat{\omega}, \hat{\mathbf{c}}_s(k) = \mathbf{h}_{\text{LS}}(k)\right)$$

$$= \frac{1}{1 + \frac{1-\tilde{q}}{\tilde{q}}\exp\left\{-\frac{\mathbf{S}_{k,k}|\mathbf{h}_{\text{LS}}(k)|^2}{1+\mathbf{S}_{k,k}\hat{\beta}e^{-\hat{\omega}k}}\right\}}. \tag{19}$$

In particular, in step (a) we have expressed the likelihood function in terms of its conditional probabilities. Moreover, we have used that fact that the term $(\mathbf{1}-\mathbf{a}_s)\odot\mathbf{h}_{\text{LS}} = (\mathbf{1}-\mathbf{a}_s)\odot(\mathbf{h}_d + \sqrt{\mathbf{S}}^{-1}\mathbf{n})$ is independent of the PDP parameters β, ω, when conditioned on $(\mathbf{1}-\mathbf{a}_s)\odot\mathbf{h}_d$ and $\mathbf{a}_s$, and the prior distribution of $\mathbf{a}_s$ is independent of β, ω. In step (b), we have neglected the terms which are independent of the optimization parameters β, ω. In step (c), the expectation is first conditioned on $\mathbf{a}_s = \mathbf{x}$, and then averaged over the posterior probability of $\mathbf{a}_s \in \{0,1\}^L$. The conditional expectation of $|\mathbf{h}_d(k)|^2$ is given by

$$\mathbb{E}\left[|\mathbf{h}_d(k)|^2 \,\middle|\, \mathbf{h}_{\text{LS}}(k), \mathbf{a}_s(k) = 0, \hat{\beta}, \hat{\omega}\right] \tag{20}$$

$$= \frac{\hat{\mathcal{P}}_d(k)^2}{(\hat{\mathcal{P}}_d(k) + 1/\mathbf{S}_{k,k})^2}|\mathbf{h}_{\text{LS}}(k)|^2 + \frac{\hat{\mathcal{P}}_d(k)}{1 + \hat{\mathcal{P}}_d(k)\mathbf{S}_{k,k}},$$

where $\hat{\mathcal{P}}_d(k) = \hat{\beta}e^{-\hat{\omega}k}$ is the current estimate of the prior variance of $\mathbf{h}_d(k)$.

In the M-step, the term $\mathcal{L}(\beta,\omega;\hat{\beta},\hat{\omega})$ is minimized with respect to the optimization parameters β, ω. We obtain

$$\left\{\tilde{\beta}, \tilde{\omega}\right\} = \arg\min_{\beta\geq 0,\omega\geq 0}\mathcal{L}(\beta,\omega;\hat{\beta},\hat{\omega}) = \arg\min_{\beta\geq 0,\omega\geq 0}\mathcal{R}(\beta,\omega;\hat{\beta},\hat{\omega})$$

$$= \arg\min_{\beta\geq 0,\omega\geq 0}\sum_{k=0}^{L-1}(1-\hat{q}_{\text{post}}(k))\ln\left(\beta e^{-\omega k}\right) \tag{21}$$

$$+ \sum_{k=0}^{L-1}(1-\hat{q}_{\text{post}}(k))\frac{\mathbb{E}\left[|\mathbf{h}_d(k)|^2\,\middle|\,\mathbf{h}_{\text{LS}}(k), \mathbf{a}_s(k) = 0, \hat{\beta}, \hat{\omega}\right]}{\beta e^{-\omega k}}.$$

By defining, for $k = 0,\ldots,L-1$,

$$\begin{cases} A_k = \frac{L(1-\hat{q}_{\text{post}}(k))\mathbb{E}\left[|\mathbf{h}_d(k)|^2\,\middle|\,\mathbf{h}_{\text{LS}}(k),\mathbf{a}_s(k)=0,\hat{\beta},\hat{\omega}\right]}{\sum_{p=0}^{L-1}(1-\hat{q}_{\text{post}}(p))}, \\ Z = \frac{\sum_{p=0}^{L-1}p(1-\hat{q}_{\text{post}}(p))}{\sum_{p=0}^{L-1}(1-\hat{q}_{\text{post}}(p))}, \end{cases} \tag{22}$$

the M-step (21) is equivalent to

$$\left\{\tilde{\beta}, \tilde{\omega}\right\} = \arg\min_{\beta\geq 0,\omega\geq 0}\ln\beta - \omega Z + \frac{1}{\beta L}\sum_{k=0}^{L-1}A_k e^{\omega k}. \tag{23}$$

$$\mathcal{L}(\beta,\omega;\hat{\beta},\hat{\omega}) \triangleq -\mathbb{E}\left[\ln p\left((\mathbf{1}-\mathbf{a}_s)\odot\mathbf{h}_{\text{LS}},(\mathbf{1}-\mathbf{a}_s)\odot\mathbf{h}_d,\mathbf{a}_s\,|\,\beta,\omega\right)|\,\mathbf{h}_{\text{LS}},\hat{\beta},\hat{\omega}\right] \tag{18}$$

$$\overset{(a)}{=} -\mathbb{E}\left[\ln p\left((\mathbf{1}-\mathbf{a}_s)\odot\mathbf{h}_{\text{LS}}\,|\,(\mathbf{1}-\mathbf{a}_s)\odot\mathbf{h}_d,\mathbf{a}_s\right)|\,\mathbf{h}_{\text{LS}},\hat{\beta},\hat{\omega}\right] - \mathbb{E}\left[\ln p\left((\mathbf{1}-\mathbf{a}_s)\odot\mathbf{h}_d\,|\,\mathbf{a}_s,\beta,\omega\right)|\,\mathbf{h}_{\text{LS}},\hat{\beta},\hat{\omega}\right]$$

$$-\mathbb{E}\left[\ln p\left(\mathbf{a}_s\right)|\,\mathbf{h}_{\text{LS}},\hat{\beta},\hat{\omega}\right] \overset{(b)}{\propto} -\mathbb{E}\left[\ln p\left((\mathbf{1}-\mathbf{a}_s)\odot\mathbf{h}_d\,|\,\mathbf{a}_s,\beta,\omega\right)|\,\mathbf{h}_{\text{LS}},\hat{\beta},\hat{\omega}\right]$$

$$\overset{(c)}{=} -\sum_{\mathbf{x}\in\{0,1\}^L}\Pr\left(\mathbf{a}_s=\mathbf{x}\,|\,\mathbf{h}_{\text{LS}},\hat{\beta},\hat{\omega},\hat{\mathbf{c}}_s=\mathbf{h}_{\text{LS}}\right)\mathbb{E}\left[\ln p\left((\mathbf{1}-\mathbf{a}_s)\odot\mathbf{h}_d\,|\,\mathbf{a}_s=\mathbf{x},\beta,\omega\right)|\,\mathbf{h}_{\text{LS}},\mathbf{a}_s=\mathbf{x},\hat{\beta},\hat{\omega}\right]$$

$$=\sum_{k=0}^{L-1}\left(1-\hat{q}_{\text{post}}(k)\right)\left[\ln\left(\beta e^{-\omega k}\right)+\frac{\mathbb{E}\left[|\mathbf{h}_d(k)|^2\,\Big|\,\mathbf{h}_{\text{LS}}(k),\mathbf{a}_s(k)=0,\hat{\beta},\hat{\omega}\right]}{\beta e^{-\omega k}}\right] \triangleq \mathcal{R}(\beta,\omega;\hat{\beta},\hat{\omega})$$

We have the following theorem, whose proof is provided in the Appendix.

Theorem 1. *There is a unique solution $\left\{\tilde{\beta},\tilde{\omega}\right\}$ to*

$$\left\{\tilde{\beta},\tilde{\omega}\right\} = \arg\min_{\beta\geq 0,\omega\geq 0}\ln\beta-\omega Z+\frac{1}{\beta L}\sum_{k=0}^{L-1}A_k e^{\omega k}. \tag{24}$$

If $\sum_{k=0}^{L-1}(Z-k)A_k > 0$, then $\tilde{\omega}$ is the unique solution in $(0,+\infty)$ of

$$\sum_{k=0}^{L-1}(Z-k)A_k e^{\tilde{\omega}k}=0. \tag{25}$$

Otherwise, $\tilde{\omega}=0$. In both cases, $\tilde{\beta}=\frac{1}{L}\sum_{k=0}^{L-1}A_k e^{\tilde{\omega}k}$.

Note that, when $\sum_{k=0}^{L-1}(Z-k)A_k > 0$, the solution is a zero of a Lth order polynomial, therefore we must recur to approximate solutions. Since the solution we seek satisfies $e^{-\tilde{\omega}}\in(0,1]$, and we have proved that it is unique, we recur to the *bisection method* [41] to determine an approximate zero $\tilde{x}=e^{-\tilde{\omega}}$ of (25).

Finally, the overall EM algorithm consists in the iterations of the E-step (19), (22) and the M-step (23). The algorithm may be initialized by neglecting the noise and the sparse component, *i.e.*, assuming $\mathbf{S}_{k,k}\to+\infty$ and $\tilde{q}=0$ in the first stage. In this case, we have $\hat{q}_{\text{post}}(k)=0,\forall k$ in (19) and the parameters of the E-step (22) are given by

$$\begin{cases} A_k=|\mathbf{h}_{\text{LS}}(k)|^2, & k=0,\ldots,L-1 \\ Z=\frac{L-1}{2}. \end{cases} \tag{26}$$

It is worth noting that, if we had assumed the diffuse component $\mathbf{h}_d$, rather than $(\mathbf{1}-\mathbf{a}_s)\odot\mathbf{h}_d$, as the hidden variable, and we had used all the observed sequence $\mathbf{h}_{\text{LS}}$ to estimate the unknown PDP parameters instead of $(\mathbf{1}-\mathbf{a}_s)\odot\mathbf{h}_{\text{LS}}$, then

in the M-step we would have

$$\left\{\tilde{\beta},\tilde{\omega}\right\}=\arg\min_{\beta\geq 0,\omega\geq 0}\sum_{k=0}^{L-1}\left(1-\hat{q}_{\text{post}}(k)\right)\ln\left(\beta e^{-\omega k}\right) \tag{27}$$

$$+\sum_{k=0}^{L-1}\left(1-\hat{q}_{\text{post}}(k)\right)\frac{\mathbb{E}\left[|\mathbf{h}_d(k)|^2\,\Big|\,\mathbf{h}_{\text{LS}}(k),\mathbf{a}_s(k)=0,\hat{\beta},\hat{\omega}\right]}{\beta e^{-\omega k}}$$

$$+\sum_{k=0}^{L-1}\hat{q}_{\text{post}}(k)\left[\ln\left(\beta e^{-\omega k}\right)+\frac{\hat{\beta}e^{-\hat{\omega}k}}{\beta e^{-\omega k}\left(1+\mathbf{S}_{k,k}\hat{\beta}e^{-\hat{\omega}k}\right)}\right],$$

where we have used the fact that, since $\hat{\mathbf{c}}_s=\mathbf{h}_{\text{LS}}$,

$$\mathbb{E}\left[|\mathbf{h}_d(k)|^2\,\Big|\,\mathbf{h}_{\text{LS}}(k),\hat{\mathbf{c}}_s(k),\mathbf{a}_s(k)=1,\hat{\beta},\hat{\omega}\right]=\frac{\hat{\beta}e^{-\hat{\omega}k}}{1+\mathbf{S}_{k,k}\hat{\beta}e^{-\hat{\omega}k}}.$$

By comparing this expression with (21), we note one additional term. In particular, the observations associated with high probability $\hat{q}_{\text{post}}(k)\to 1$ with an active sparse component give a significant contribution to the log-likelihood function. However, these observations do not provide information about the diffuse component $\mathbf{h}_d$, since $\mathbf{c}_s$ is a deterministic, unknown vector. Conversely, in (21), these observations yield a negligible contribution.

Choice of the sparsity level $\tilde{q}$

We next discuss the choice of the parameter $\tilde{q}$ used to estimate the parameters β,ω. Since the EM algorithm solves the ML problem [40], we consider the general problem of maximizing the likelihood function. Assuming the sparsity level $\tilde{q}$, the ML estimate of β, ω and $\mathbf{c}_s$ is defined as

$$\{\hat{\beta},\hat{\omega},\hat{\mathbf{c}}_s\}=\arg\max_{\beta\geq 0,\omega\geq 0,\mathbf{c}_s}p(\mathbf{h}_{\text{LS}}|\beta,\omega,\mathbf{c}_s)$$

$$=\arg\max_{\beta\geq 0,\omega\geq 0,\mathbf{c}_s}-\sum_{k=0}^{L-1}\ln\left(1/\mathbf{S}_{k,k}+\mathcal{P}_d(k)\right)$$

$$+\sum_{k=0}^{L-1}\ln\left(\tilde{q}\exp\left\{-\frac{|\mathbf{h}_{\text{LS}}(k)-\mathbf{c}_s(k)|^2}{1/\mathbf{S}_{k,k}+\mathcal{P}_d(k)}\right\}\right.$$

$$\left.+(1-\tilde{q})\exp\left\{-\frac{|\mathbf{h}_{\text{LS}}(k)|^2}{1/\mathbf{S}_{k,k}+\mathcal{P}_d(k)}\right\}\right),$$

where we have used the fact that $\mathbf{h}_{\text{LS}}(k)|\mathbf{a}_s(k)=a\sim\mathcal{CN}(a\mathbf{c}_s(k),\mathcal{P}_d(k)+1/\mathbf{S}_{k,k})$ and $\mathcal{P}_d(k)=\beta e^{-\omega k}$. By

maximizing over $\mathbf{c}_s$, we obtain $\hat{\mathbf{c}}_s = \mathbf{h}_{\text{LS}}$. Then, letting $t_k(\mathcal{P}_d(k)) = \frac{|\mathbf{h}_{\text{LS}}(k)|^2}{1/\mathbf{S}_{k,k} + \mathcal{P}_d(k)}$, $s(\tilde{q}, t) = \ln\left(t + \frac{1-\tilde{q}}{\tilde{q}} t e^{-t}\right)$ and $\mathcal{F}(\tilde{q}, \beta, \omega) = \sum_{k=0}^{L-1} s(\tilde{q}, t_k(\mathcal{P}_d(k)))$, we obtain

$$
\begin{aligned}
\{\hat{\beta}, \hat{\omega}\} &= \arg\max_{\beta \geq 0, \omega \geq 0} \sum_{k=0}^{L-1}\left[\ln t_k(\mathcal{P}_d(k)) + \ln\left(1 + \frac{1-\tilde{q}}{\tilde{q}} e^{-t_k(\mathcal{P}_d(k))}\right)\right] \\
&= \arg\max_{\beta \geq 0, \omega \geq 0} \sum_{k=0}^{L-1} s(\tilde{q}, t_k(\mathcal{P}_d(k))) = \arg\max_{\beta \geq 0, \omega \geq 0} \mathcal{F}(\tilde{q}, \beta, \omega),
\end{aligned}
$$

where we have added the term $\sum_{k=0}^{L-1} \ln(|\mathbf{h}_{\text{LS}}(k)|^2) - L \ln \tilde{q}$, which does not affect the maximization.

Consider a given pair of parameters (β, ω), and let

$$
s'(\tilde{q}, t) \triangleq \frac{\mathrm{d}s(\tilde{q}, t)}{\mathrm{d}t} = \frac{\tilde{q} - (1-\tilde{q})e^{-t}(t-1)}{\tilde{q}t + (1-\tilde{q})te^{-t}}, \tag{28}
$$

$$
\mathcal{F}'_\beta(\tilde{q}, \beta, \omega) \triangleq \frac{\mathrm{d}\mathcal{F}(\tilde{q}, \beta, \omega)}{\mathrm{d}\beta} = \sum_{k=0}^{L-1} s'(\tilde{q}, t_k(\mathcal{P}_d(k))) \frac{\mathrm{d}t_k(\mathcal{P}_d(k))}{\mathrm{d}\beta}.
$$

Similarly, we define $\mathcal{F}'_\omega(\tilde{q}, \beta, \omega)$ as the derivative with respect to ω. Note that, if $\mathcal{F}'_\beta(\tilde{q}, \beta, \omega) > 0 \ (< 0)$, then there is an incentive to augment (diminish) β so as to increase the log-likelihood function $\mathcal{F}(\tilde{q}, \beta, \omega)$ (the same consideration holds for $\mathcal{F}'_\omega(\tilde{q}, \beta, \omega)$). We now prove that this derivative is a decreasing function of $\tilde{q}$, so that, the larger $\tilde{q}$, the smaller the incentive to increase β (and, possibly, the larger the incentive to decrease it, if the derivative becomes negative). In fact,

$$
\begin{aligned}
\frac{\mathrm{d}s'(\tilde{q}, t)}{\mathrm{d}\tilde{q}} &= \frac{1}{\tilde{q}^2} \exp\{-2s(\tilde{q}, t)\} t^2 e^{-t} > 0, \\
\frac{\mathrm{d}t_k(\mathcal{P}_d(k))}{\mathrm{d}\beta} &= -\frac{1}{\beta} t_k(\mathcal{P}_d(k)) \frac{\mathcal{P}_d(k)}{1/\mathbf{S}_{k,k} + \mathcal{P}_d(k)} < 0, \tag{29}
\end{aligned}
$$

and therefore

$$
\frac{\mathrm{d}\mathcal{F}'_\beta(\tilde{q}, \beta, \omega)}{\mathrm{d}\tilde{q}} = \sum_{k=0}^{L-1} \frac{\mathrm{d}s'(\tilde{q}, t_k(\mathcal{P}_d(k)))}{\mathrm{d}\tilde{q}} \frac{\mathrm{d}t_k(\mathcal{P}_d(k))}{\mathrm{d}\beta} < 0.
$$

Similarly, we can prove that $\mathcal{F}'_\omega(\tilde{q}, \beta, \omega)$ is an increasing function of $\tilde{q}$, so that, the larger $\tilde{q}$, the smaller the incentive to decrease ω (and, possibly, the larger the incentive to increase it, if the derivative becomes negative).

Moreover, note that, if $\tilde{q} \geq \frac{1}{1+e^2} \simeq 0.12$, then we have $e^{-t}(t-1) \leq e^{-2} \leq \frac{\tilde{q}}{1-\tilde{q}}$ (since the left hand side is maximized for $t = 2$), which implies $s'(\tilde{q}, t) \geq 0, \forall t$. We conclude that, when $\tilde{q} \geq \frac{1}{1+e^2}$, the derivatives $\mathcal{F}'_\beta(\tilde{q}, \beta, \omega) < 0, \forall \beta \geq 0, \omega \geq 0$ and $F'_\omega(\tilde{q}, \beta, \omega) > 0, \forall \beta \geq 0, \omega \geq 0$. Therefore, the ML estimate of β, ω gives $\hat{\beta} = 0, \hat{\omega} \to +\infty$, and the PDP estimate is forced to zero.

Conversely, if we let $\tilde{q} \to 0^+$, then the contribution of the sparse component $\mathbf{a}_s \odot \mathbf{c}_s$ is neglected, and the channel is treated as being purely diffuse.

This analysis proves that the prior sparsity level $\tilde{q} \geq 0.12$ should never be used, and suggests the existence of a trade-off in the optimal algorithm parameter $\tilde{q}$, which is confirmed by simulation in Section VI: in order not to force the PDP estimate to zero, $\tilde{q}$ should be "small"; however, in order to take into account the presence of the sparse component in the observations, $\tilde{q}$ should not be "too small". A further investigation on the optimal value of $\tilde{q}$ is left for future work.

VI. SIMULATION RESULTS

In this section, we present the simulation results and evaluate the MSE and BER performance achievable with the above estimation strategies for a channel following the HSD model. In particular, the HSD model allows us to control the parameters (*e.g.*, sparsity level $\tilde{q}$, PDP profiles $\mathcal{P}_d$, $\mathcal{P}_s$) and to evaluate the performance of the estimators developed in Section IV in an ideal setting, *i.e.*, where the channel realizations follow exactly the HSD model, based on which the estimators have been designed. Conversely, in Part II [32], simulation results are given based on a more realistic channel model [24], which allows us to evaluate the robustness of the proposed estimators against deviations from the HSD model.

We define the MSE of the estimator $\hat{\mathbf{h}}$ of the channel $\mathbf{h}$ as

$$
\text{MSE} = \frac{1}{L} \sum_{k=0}^{L-1} \mathbb{E}\left[\left|\hat{\mathbf{h}}(k) - \mathbf{h}(k)\right|^2\right]. \tag{30}
$$

In the simulations, the channel $\mathbf{h} \in \mathbb{C}^L$ has delay spread $L = 100$. The sparsity pattern $\mathbf{a}_s \in \{0, 1\}^L$ has i.i.d. Bernoulli entries with parameter $q = 0.1$. The vector $\mathbf{c}_s \in \mathbb{C}^L$ is drawn from $\mathcal{CN}(\mathbf{0}, \Lambda_s)$, with exponential PDP $\Lambda_s(k, k) = \mathcal{P}_s(k) = P_s e^{-\omega k}$, where $\omega = 0.05$. The diffuse component $\mathbf{h}_d \in \mathbb{C}^L$ is drawn from $\mathcal{CN}(\mathbf{0}, \Lambda_d)$, with exponential PDP $\Lambda_d(k, k) = \mathcal{P}_d(k) = \beta P_s e^{-\omega k}$. Unless otherwise stated, we use $\beta = 0.01$, hence the ratio between the energy of the sparse and diffuse components is given by $[\mathbb{E}[\mathbf{h}_s^* \mathbf{h}_s]/\mathbb{E}[\mathbf{h}_d^* \mathbf{h}_d]]_{\text{dB}} = 10\,\text{dB}$, where $\mathbf{h}_s = \mathbf{a}_s \odot \mathbf{c}_s$ denotes the sparse component. The parameter $P_s > 0$ is a normalization factor, and is chosen so that the average channel energy is L, *i.e.*, $\sum_{k=0}^{L-1} \mathbb{E}\left[|\mathbf{h}(k)|^2\right] = P_s \sum_{k=0}^{L-1} (\beta + q)e^{-\omega k} = L$. We assume an orthogonal pilot sequence, thus $\mathbf{S}$ is diagonal and can be described as $\mathbf{S} = S \cdot \mathbf{I}_L$, where $S > 0$ is the estimation SNR.

We compare the LS estimate, the MMSE estimate (Section IV-A), and the GMMSE and GThres estimators (Sections IV-B1 and IV-B2, respectively), for different values of the assumed sparsity level $\tilde{q} \in \{0.1, 0.01, 0.001\}$, corresponding to $\alpha = \frac{1-\tilde{q}}{\tilde{q}} \in \{2.2, 4.6, 6.9\}$. We also compare these estimators with a purely sparse and a purely diffuse estimators, which ignore the diffuse or sparse components, respectively. Since a per-tap approach is optimal in this case, for the sparse estimator we choose a variation of the GThres estimator which assumes no diffuse component ($\mathbf{h}_d = \mathbf{0}$).

Note that the MMSE estimator in Section IV-A, by assuming perfect knowledge of q, Λ_d and Λ_s, is a lower bound to the estimation accuracy. This is used primarily as a reference.

Figure 1 plots the MSE of the estimators, assuming perfect knowledge of Λ_d, as a function of the average SNR per channel entry, defined as $S\mathbb{E}[\mathbf{h}^*\mathbf{h}]/L$. We observe that, using a more conservative approach, *i.e.*, assuming a Bernoulli prior $\tilde{q} \ll q$, improves the estimation accuracy in the high and low SNR regimes. In fact, the optimal threshold for the GThres estimator represents a balance between the probability of mis-detecting an active sparse component as diffuse contribution and the probability of false alarm (detecting a diffuse contribution as active sparse component). A conservative approach, by employing a small threshold, reduces the false alarm probability (a similar consideration holds for the
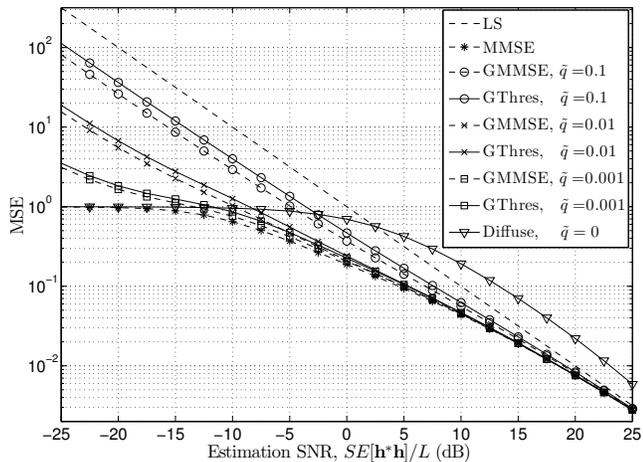
Fig. 1. MSE of the channel estimators, assuming perfect knowledge of the PDP $\mathcal{P}_d(k)$. $\beta = 0.01$, $q = 0.1$

Fig. 2. MSE of the channel estimators as a function of β, assuming perfect knowledge of the PDP of the diffuse component $\mathcal{P}_d(k)$. $[S\mathbb{E}[\mathbf{h}^*\mathbf{h}]/L]_{\text{dB}} = 10\,\text{dB}$, $q = 0.1$

GMMSE estimator). This trend can also be observed in the medium SNR ranges. However, this property does not hold in general. To see that, we also plot the accuracy of the *diffuse* estimator $\hat{\mathbf{h}}^{(\text{Diff})}(k) = \frac{S\mathcal{P}_d(k)}{1+S\mathcal{P}_d(k)}\mathbf{h}_{\text{LS}}(k)$, which ignores the sparse component $\mathbf{a}_s \odot \mathbf{c}_s$. This can be interpreted as a limit case of the GMMSE and GThres estimators, for $\tilde{q} \to 0$, or equivalently $\alpha \to +\infty$. An analytical explanation of this behavior, based on the MSE analysis of the estimators in the asymptotic regimes of high and low SNR, follows in Part II of this paper [32]. Note also that the GMMSE estimator performs better than the GThres estimator with respect to MSE, for a given value of $\tilde{q}$. This is a consequence of the fact that GThres allows only the extreme values $\hat{\mathbf{a}}_s^{(\text{GThres})}(k) \in \{0, 1\}$, whereas GMMSE allows a smoother transition between these two extremes.

In Figure 2, we let vary the ratio between the energies of the sparse and diffuse components, $\mathbb{E}[\mathbf{h}_s^*\mathbf{h}_s]/\mathbb{E}[\mathbf{h}_d^*\mathbf{h}_d] = q/\beta$. The SNR per channel entry is $[S\mathbb{E}[\mathbf{h}^*\mathbf{h}]/L]_{\text{dB}} = 10\,\text{dB}$. The MSE of the purely sparse estimator is also plotted in this case. Similarly to Figure 1, we note that a conservative approach is beneficial from an MSE perspective. As expected, the sparse estimator performs worse than the GThres estimator, due to its inability to exploit the diffuse component of the channel. In particular, it performs closely to the GThres estimator for small values of β (*i.e.*, large values of $\mathbb{E}[\mathbf{h}_s^*\mathbf{h}_s]/\mathbb{E}[\mathbf{h}_d^*\mathbf{h}_d]$), where the diffuse component is negligible with respect to the sparse one, and incurs a performance degradation for large values of β, where the diffuse component becomes significant. Moreover, as expected, the only diffuse estimator achieves good performance for large values of β. However, it performs poorly for small values of β, where the sparse component yields a significant contribution. Note that, excluding the MMSE estimator, the GThres estimator with $\tilde{q} = 0.001$ achieves the best performance over the entire range of values considered, very close to the MMSE lower bound. This proves that the proposed methods are robust, and adapt to a wide range of estimation scenarios, where the channel exhibits either a sparse, diffuse or hybrid nature (corresponding to large, small and moderate
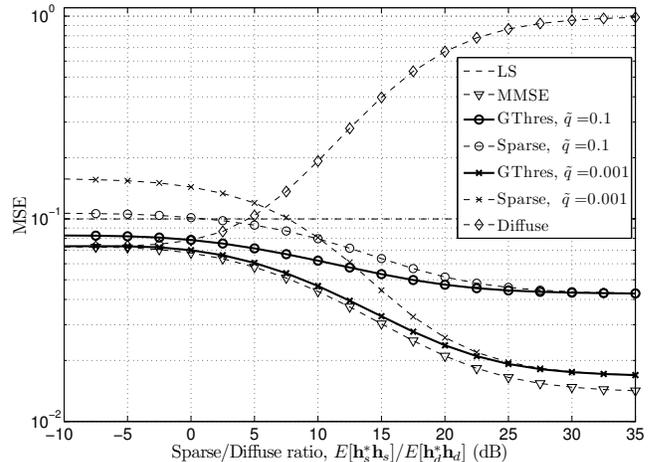
values of $\mathbb{E}[\mathbf{h}_s^*\mathbf{h}_s]/\mathbb{E}[\mathbf{h}_d^*\mathbf{h}_d]$, respectively).

Figure 3 compares the MSE of the GMMSE estimator, for the two cases where $\mathbf{\Lambda}_d$ is perfectly known at the receiver, and where it is estimated from the observed sequence using the EM algorithm (Section V), based on only one realization of the channel. We notice that, in general, there is a small performance loss due to the unknown $\mathbf{\Lambda}_d$, mainly in the low SNR range and for small values of $\tilde{q}$ (however, no performance degradation is observed for $\tilde{q} = 0.1$). This behavior is explained by the fact that the MMSE estimate of $\mathbf{h}_d$ in (14) is more sensitive to errors in the estimation of $\mathbf{\Lambda}_d$ in the low SNR than in the high SNR regime. In fact, for high SNR values, it approaches the LS solution. On the other hand, for small values of $\tilde{q}$ we have the following. The posterior probability of the entries of the sparsity pattern $\mathbf{a}_s$, as a function of the factor $\alpha = \left(\frac{1-\tilde{q}}{\tilde{q}}\right)$, is given by (16) with $\mathbf{S}_{k,k} = S$. This is a decreasing function of α (*i.e.*, increasing function of $\tilde{q}$). As a consequence, the smaller $\tilde{q}$ the more the weight given to the right-hand term of (14), associated with the MMSE estimate of $\mathbf{h}_d(k)$, which is sensitive to errors in the estimate of $\mathcal{P}_d(k)$, compared to the left-hand term, associated with the LS estimate of $\mathbf{c}_s(k)$, which is independent of the PDP estimate. As a consequence, a smaller value of $\tilde{q}$ results in an overall estimate that is more sensitive to errors in the PDP estimate of $\mathbf{h}_d$. Similar considerations hold for the GThres estimator.

Figure 4 plots the MSE of PDP estimator of the diffuse component developed in Section V, for different values of $\tilde{q}$ and of the number of iterations of the EM algorithm, based on only one channel realization, as a function of the SNR per diffuse channel entry $S\mathbb{E}[\mathbf{h}_d^*\mathbf{h}_d]/L$. In particular, letting $\hat{\mathcal{P}}_d(k), k = 0, \ldots, L-1$ be an estimate of $\mathcal{P}_d(k) = \beta e^{-\omega k}$, we compute the following MSE metric:

$$\text{MSE}_{\text{PDP}} = \frac{1}{L} \sum_{k=0}^{L-1} \mathbb{E}\left[\left(\ln \hat{\mathcal{P}}_d(k) - \ln \mathcal{P}_d(k)\right)^2\right]. \quad (31)$$

The performance is compared also with an oracle estimator, which assumes perfect knowledge of $\mathbf{a}_s \odot \mathbf{c}_s$, thus being able to
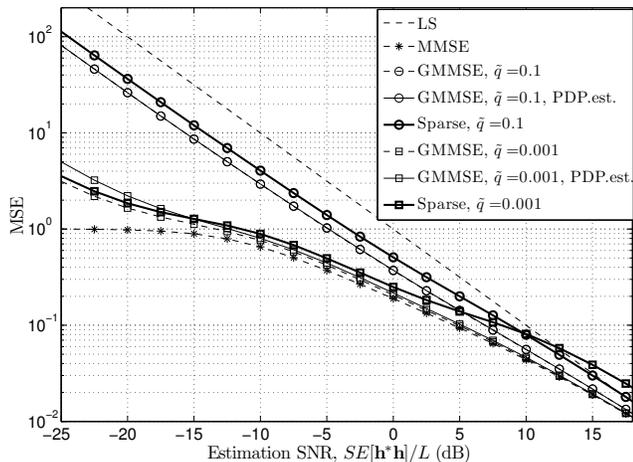
Fig. 3. MSE of the GMMSE estimators, comparison between the cases where the PDP of the diffuse component is known and estimated from the data, respectively. $\beta = 0.01$, $q = 0.1$. The two curves of the GMMSE estimator with $\tilde{q} = 0.1$ where the PDP is known and estimated overlap.

Fig. 4. MSE of the PDP estimator of $\mathbf{h}_d$. $\beta = 0.01$, $q = 0.1$.

perfectly remove the interference from the sparse component (in particular, we use the EM estimator with $\tilde{q} = 0$). In the Figure, the MSE floor refers to the ML estimator of β, ω in the noiseless scenario with no sparse component. It can be shown that, in this case, the ML estimator is obtained by setting $A_k = |\mathbf{h}_d(k)|^2$ and $Z = \frac{L-1}{2}$ in the E-step (22), and by solving (24) using the results of Theorem 1. As expected, the Oracle estimator achieves the best performance, and approaches the MSE floor in the high SNR. Remarkably, the EM estimator with $\tilde{q} = 0.001$ and 300 iterations approaches the performance of the Oracle estimator, although it cannot take advantage of prior knowledge of $\mathbf{a}_s \odot \mathbf{c}_s$. This proves that the proposed method effectively removes the interference from the sparse component, by discarding the observations associated, with high probability, to the active sparse components. Interestingly, the case $\tilde{q} = 0.001$ with 20 iterations incurs a small performance degradation compared to the MSE achievable after 300 iterations, which becomes negligible for moderate and large SNR values. On the other hand, when $\tilde{q} = 0$ is used, the presence of the sparse component is neglected and the channel is treated as being purely diffuse. In this case, a significant performance degradation is incurred. Finally, we notice that the case $\tilde{q} = 0.15$ incurs a performance degradation, compared to the case $\tilde{q} = 0.001$, which confirms our analysis in Section V. In fact, we have verified that the estimate of the PDP parameter ω diverges to $+\infty$ as the EM algorithm is iterated, so that the PDP estimate is forced to zero and the overall MSE diverges to $+\infty$.

Finally, in Figure 5 we plot the BER induced by channel estimation errors, for the case where the PDP of $\mathbf{h}_d$ is known. To this end, we define an OFDM-UWB system, employing $N_{\text{dft}} = 512$ sub-carriers and a 4-QAM constellation. Our observation for channel estimation has noise; in contrast, we assume no noise when evaluating the BER. As a result, the BER curves reflect the errors induced by channel estimation versus additive channel noise. In particular, let $X(n)$ be the 4-QAM symbol transmitted on the nth sub-carrier, and $H(n) =$
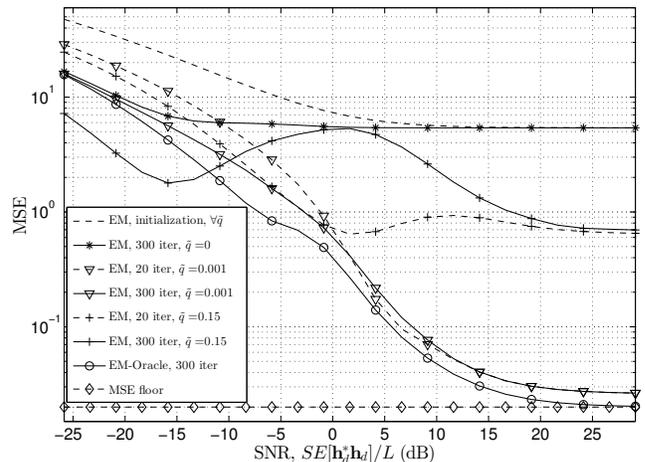
$\sum_{l=0}^{L-1} \mathbf{h}(n) e^{-i2\pi \frac{ln}{N_{dft}}}$ be the DFT of the channel. Then, the received symbol is $Y(n) = H(n)X(n)$. This is equalized by using the estimate $\hat{H}(n)$ of $H(n)$ (using either of the estimators developed in Section IV), i.e., $\tilde{X}(n) = \frac{H(n)}{\hat{H}(n)}X(n)$, and the decision is based on a minimum distance criterion, i.e., $\hat{X}(n) = \min_{x \in 4-\text{QAM}} |\tilde{X}(n) - x|^2$. Moreover, the BER is averaged over the "good" sub-carriers only, which are chosen based on the heuristic carrier selection scheme

$$\left\{ k : |H(k)|^2 \geq \lambda \max_n |H(n)|^2 \right\}, \qquad (32)$$

where $\lambda \in (0, 1)$ is a threshold value. In particular, λ is chosen so that 30% of the sub-carriers are classified as "good". The SNR is referred to the output of an ideal Rake receiver with perfect channel knowledge, where the estimation noise is treated as additive Gaussian noise at the receiver. This is defined as $\text{SNR}_{\text{rake}} = S\mathbf{h}^*\mathbf{h}$. We notice that GMMSE estimator with $\tilde{q} = 0.001$ performs very closely to the lower bound, represented by the BER induced by the MMSE estimator, defined in Section IV-A. On the other hand, both the diffuse and the purely sparse estimators perform poorly, due to their inability to exploit both the sparse and the diffuse components jointly.

VII. CONCLUSIONS

In this paper, we have investigated channel estimation for UWB systems. In particular, we have proposed a novel sparse/diffuse model for the UWB channel, which is able to capture the main UWB propagation mechanisms: fine delay resolution capability, scattering from rough surfaces, frequency dispersion. We have then identified four scenarios of interest in practical systems, differing in the amount of side information available at the receiver for the purpose of channel estimation, and we have proposed channel estimators exploiting the channel structure and the side-information to enhance the estimation accuracy.

Of particular interest is the scenario where the PDP of the diffuse component is known at the receiver, and the statistics of the specular component are unknown. This is relevant when the observation interval is large enough to allow averaging
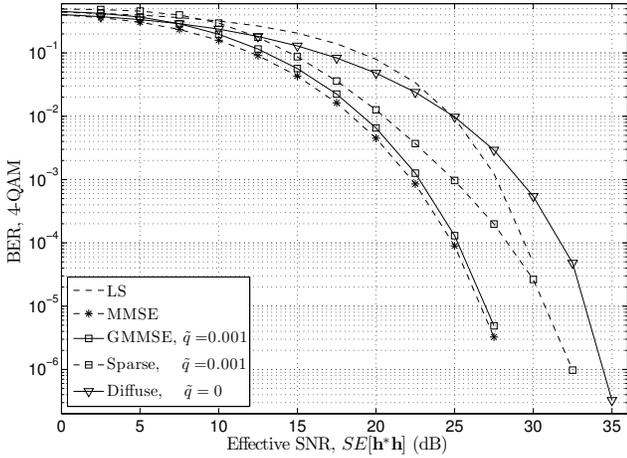
Fig. 5. BER induced by channel estimation errors, with known PDP of $\mathbf{h}_d$. $\beta = 0.01$, $q = 0.1$.

over the small scale fading, but not over the large scale fading. For this scenario, we have proposed the *Generalized MMSE* and *Generalized Thresholding Estimators*. Moreover, we have proposed an EM algorithm for the PDP estimation of the diffuse component, which exploits the exponential structure of the PDP to average the fading over the channel delay dimension, rather than over subsequent independent realizations of the fading process.

We have compared these estimators to the unconstrained LS estimator, and to conventional purely sparse or diffuse estimators, which, on the other hand, ignore either the diffuse or the sparse component. The numerical results show that, when the channel follows the hybrid sparse/diffuse model, the proposed estimators considerably improve the performance over LS and conventional sparse or diffuse estimators, from both an MSE and a BER perspective. Moreover, we have observed that it is beneficial to be conservative in the estimation of the sparse component of the channel, *i.e.*, to assume that the sparse component is sparser than it actually is. In Part II, we develop an MSE analysis of these estimators, proving this conjecture in the asymptotic high and low SNR regimes, and we validate the proposed sparse/diffuse model and estimation strategies based on a more realistic UWB channel emulator.

ACKNOWLEDGMENTS

APPENDIX

Proof of Theorem 1 in Section V. Let $f(x, \beta) = \ln \beta + Z \ln x + \frac{1}{\beta L} \sum_{k=0}^{L-1} A_k x^{-k}$, where we have defined $x = e^{-\omega} \in (0, 1]$ in the argument of the minimization in (24). By minimizing with respect to $\beta \geq 0$, for a fixed x, we have

$$\tilde{\beta}(x) = \arg\min_{\beta \geq 0} \left\{ \ln \beta + \frac{1}{\beta L} \sum_{k=0}^{L-1} A_k x^{-k} \right\} = \frac{1}{L} \sum_{k=0}^{L-1} A_k x^{-k}.$$

Substituting into $f(x, \beta)$, we obtain $f(x, \tilde{\beta}(x)) = 1 + \ln \tilde{\beta}(x) + Z \ln x$. We now minimize $f(x, \tilde{\beta}(x))$ with respect to $x \in (0, 1]$. $f(x, \tilde{\beta}(x))$ is an increasing function of $x \in (0, 1]$ if and only if

$$f'(x, \tilde{\beta}(x)) = \frac{\mathrm{d}f(x, \tilde{\beta}(x))}{\mathrm{d}x} = \frac{\tilde{\beta}'(x)}{\tilde{\beta}(x)} + \frac{Z}{x} > 0, \qquad (33)$$

where $\tilde{\beta}'(x) = \frac{\mathrm{d}\tilde{\beta}(x)}{\mathrm{d}x} = -\frac{1}{L} \sum_{k=0}^{L-1} k A_k x^{-(k+1)}$. Equivalently, multiplying both sides by $x^{Z+1} \tilde{\beta}(x) > 0$, $f(x, \tilde{\beta}(x))$ is an increasing function of $x \in (0, 1]$ if and only if

$$g(x) \triangleq x^{Z+1} \tilde{\beta}(x) f'(x, \tilde{\beta}(x)) \qquad (34)$$

$$= \frac{1}{L} \sum_{k=0}^{L-1} A_k x^{Z-k} (Z - k) > 0.$$

Note that $g'(x) = \frac{\mathrm{d}g(x)}{\mathrm{d}x} = \frac{1}{L} \sum_{k=0}^{L-1} A_k x^{Z-k-1} (Z-k)^2 > 0$, $\forall x \in (0, 1]$. Therefore, $g(x)$ is a continue monotone increasing function of x. Moreover, since $Z < L - 1$ from (22) and $\lim_{x \to 0^+} x^m = +\infty$ when $m < 0$, we have $\lim_{x \to 0^+} g(x) = -\infty$. Therefore, if $g(1) > 0$, or equivalently $\sum_{k=0}^{L-1} (Z-k) A_k > 0$, then there exists a unique $\tilde{x} \in (0, 1)$ solution of $g(\tilde{x}) = 0$ such that

$$\begin{cases} g(x) > 0, & \forall x > \tilde{x} \\ g(x) < 0, & \forall x < \tilde{x}. \end{cases} \qquad (35)$$

Equivalently, $\tilde{x} \in (0, 1)$ is the unique solution of $f'(x, \tilde{\beta}(x)) = 0$ such that

$$\begin{cases} f'(x, \tilde{\beta}(x)) > 0, & \forall x > \tilde{x} \\ f'(x, \tilde{\beta}(x)) < 0, & \forall x < \tilde{x}. \end{cases} \qquad (36)$$

As a consequence, $\tilde{x}$ is the unique minimizer of $f(x)$, $x \in (0, 1]$, and $\left\{ \tilde{\beta}(\tilde{x}), \tilde{\omega} = -\ln \tilde{x} \right\}$ uniquely minimizes (24).

Conversely, if $g(1) \leq 0$, *i.e.*, $\sum_{k=0}^{L-1} (Z-k) A_k \leq 0$, then $g(x) \leq 0$, $\forall x \in (0, 1]$. This is equivalent to $f'(x, \tilde{\beta}(x)) \leq 0$, $\forall x \in (0, 1]$. As a consequence, 1 is the unique minimizer of $f(x, \tilde{\beta}(x))$, and $\left\{ \tilde{\beta}(1), \tilde{\omega} = 0 \right\}$ uniquely minimizes (24). $\qquad \square$

REFERENCES

[1] M. Win and R. Scholtz, "Impulse radio: how it works," *IEEE Communications Letters*, vol. 2, no. 2, pp. 36–38, Feb. 1998.

[2] R. Scholtz, "Multiple access with time-hopping impulse modulation," in *IEEE Military Communications Conference*, vol. 2, Oct. 1993, pp. 447–450.

[3] R. Qiu, H. Liu, and X. Shen, "Ultra-wideband for multiple access communications," *IEEE Communications Magazine*, vol. 43, no. 2, pp. 80–87, Feb. 2005.

[4] S. Gezici, Z. Tian, G. Giannakis, H. Kobayashi, A. Molisch, H. Poor, and Z. Sahinoglu, "Localization via Ultra-Wideband radios: a look at positioning aspects for future sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 70–84, July 2005.

[5] M. Win and R. Scholtz, "On the robustness of ultra-wide bandwidth signals in dense multipath environments," *IEEE Communications Letters*, vol. 2, no. 2, pp. 51–53, Feb. 1998.

[6] M. Chiani and A. Giorgetti, "Coexistence Between UWB and Narrow-Band Wireless Communication Systems," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 231–254, Feb. 2009.

[7] A. Batra, J. Balakrishnan, G. Aiello, J. Foerster, and A. Dabak, "Design of a multiband OFDM system for realistic UWB channel environments," *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 9, Sep. 2004.

[8] T. Zasowski, G. Meyer, F. Althaus, and A. Wittneben, "Propagation effects in UWB body area networks," in *IEEE International Conference on Ultra-Wideband (ICU)*, Sep. 2005, pp. 16–21.

[9] L. Yang and G. Giannakis, "Ultra-Wideband Communications: an idea whose time has come," *IEEE Signal Processing Magazine*, vol. 21, no. 6, pp. 26–54, Nov. 2004.

[10] A. Molisch, D. Cassioli, C.-C. Chong, S. Emami, A. Fort, B. Kannan, J. Karedal, J. Kunisch, H. Schantz, K. Siwiak, and M. Win, "A Comprehensive Standardized Model for Ultrawideband Propagation Channels," *IEEE Transactions on Antennas and Propagation*, vol. 54, no. 11, pp. 3151–3166, Nov. 2006.

[11] G. Aiello and G. Rogerson, "Ultra-Wideband Wireless Systems," *IEEE Microwave Magazine*, vol. 4, no. 2, June 2003.

[12] A. Molisch, "Ultra-Wide-Band Propagation Channels," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 353–371, Feb. 2009.

[13] T. Blumensath and M. Davies, "Iterative Thresholding for Sparse Approximations," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 629–654, 2008, 10.1007/s00041-008-9035-z.

[14] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129–159, 2001.

[15] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[16] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, June 2010.

[17] C. Carbonelli, S. Vedantam, and U. Mitra, "Sparse Channel Estimation with Zero Tap Detection," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1743–1763, May 2007.

[18] J. Paredes, G. Arce, and Z. Wang, "Ultra-Wideband Compressed Sensing: Channel Estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 383–395, Oct. 2007.

[19] P. Meissner, T. Gigl, and K. Witrisal, "UWB sequential Monte Carlo positioning using virtual anchors," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Sep. 2010, pp. 1–10.

[20] Y. Shen and M. Win, "Fundamental Limits of Wideband Localization-Part I: A General Framework," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 4956–4980, Oct. 2010.

[21] F. Quitin, C. Oestges, F. Horlin, and P. De Doncker, "Diffuse multipath component characterization for indoor MIMO channels," in *Proceedings of the Fourth European Conference on Antennas and Propagation (EuCAP)*, Apr. 2010, pp. 1–5.

[22] A. Molisch, "Ultrawideband propagation channels-Theory, Measurement, and Modeling," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 5, pp. 1528–1545, Sep. 2005.

[23] T. Santos, F. Tufvesson, and A. Molisch, "Modeling the Ultra-Wideband Outdoor Channel: Model Specification and Validation," *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, pp. 1987–97, June 2010.

[24] J. Kunisch and J. Pamp, "An ultra-wideband space-variant multipath indoor radio channel model," in *IEEE Conference on Ultra Wideband Systems and Technologies*, Nov. 2003, pp. 290–294.

[25] Y. Zhou, X. Yin, N. Czink, T. Zemen, A. Guo, and F. Liu, "Evaluation of Doppler-Delay Properties of Diffuse Components in Vehicular Propagation Channels," in *2nd IEEE International Conference on Wireless Access in Vehicular Environments*, Dec. 2009.

[26] N. Czink, F. Kaltenberger, Y. Zhou, L. Bernado, T. Zemen, and X. Yin, "Low-Complexity Geometry-Based Modeling of Diffuse Scattering," in *Proceedings of the Fourth European Conference on Antennas and Propagation (EuCAP)*, Apr. 2010.

[27] C. Carbonelli and U. Mitra, "Clustered ML Channel Estimation for Ultra-Wideband Signals," *IEEE Transactions on Wireless Communications*, vol. 6, no. 7, pp. 2412–2416, July 2007.

[28] P. Schniter, "A Message-Passing Receiver for BICM-OFDM Over Unknown Clustered-Sparse Channels," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 8, pp. 1462–1474, Dec. 2011.

[29] R. Thoma, M. Landmann, and A. Richter, "RIMAX-a Maximum Likelihood Framework for Parameter Estimation in Multidimensional Channel Sounding," in *International Symposium on Antennas and Propagation (ISAP)*, Aug. 2004.

[30] N. Michelusi, U. Mitra, and M. Zorzi, "Hybrid Sparse/Diffuse UWB channel estimation," in *IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2011, pp. 201–205.

[31] R. Qiu, "A study of the ultra-wideband wireless propagation channel and optimum UWB receiver design," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 9, pp. 1628–1637, Dec. 2002.

[32] N. Michelusi, U. Mitra, A. Molisch, and M. Zorzi, "UWB Sparse/Diffuse Channels, Part II: Estimator Analysis and Practical Channels," *IEEE Transactions on Signal Processing*, 2012, to be published.

[33] A. Molisch, *Wireless Communications, Second Edition*, ser. Wiley-IEEE. John Wiley & Sons, 2011.

[34] J. Karedal, S. Wyne, P. Almers, F. Tufvesson, and A. Molisch, "Statistical analysis of the UWB channel in an industrial environment," in *IEEE 60th Vehicular Technology Conference*, vol. 1, Sep. 2004, pp. 81–85.

[35] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to Basis Mismatch in Compressed Sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, May 2011.

[36] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. Springer, Aug. 1998.

[37] A. Saleh and R. Valenzuela, "A Statistical Model for Indoor Multipath Propagation," *Journal on Selected Areas in Communications*, vol. 5, no. 2, pp. 128–137, Feb. 1987.

[38] D. Cassioli, M. Win, and A. Molisch, "The ultra-wide bandwidth indoor channel: from statistical model to simulations," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 6, pp. 1247–1257, Aug. 2002.

[39] J. Hansen, "An analytical calculation of power delay profile and delay spread with experimental verification," *IEEE Communications Letters*, vol. 7, no. 6, pp. 257–259, June 2003.

[40] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, 1977.

[41] R. L. Burden and J. D. Faires, *Numerical Analysis, 9th Edition*. Cengage Learning, 2011.

Nicolò Michelusi (S'09) received the B.S. (Electronics Engineering) and M.S. (Telecommunications Engineering) degrees summa cum laude from the University of Padova, Italy, in 2006 and 2009, respectively, and the M.S. degree in Telecommunications Engineering from Technical University of Denmark, Copenhagen, Denmark, in 2009, as part of the T.I.M.E. double degree program. Since January 2009, he is a Ph.D. student at University of Padova, Italy. In 2011, he was on leave at the University of Southern California, Los Angeles, United States, as a visiting Ph.D. student. His research interests include ultrawideband communications, wireless networks, cognitive networks, optimal control, energy harvesting for communications.

Urbashi Mitra received the B.S. and the M.S. degrees from the University of California at Berkeley and her Ph.D. from Princeton University. After a six year stint at the Ohio State University, she joined the Department of Electrical Engineering at the University of Southern California, Los Angeles, where she is currently a Professor. Dr. Mitra has been an Associate Editor for the following IEEE publications: Transactions on Information Theory (2007-2011), Journal of Oceanic Engineering (2006-2011), and Transactions on Communications (1996-2001). She was a member of the IEEE Information Theory Society's Board of Governors (2002-2007) and began a third term in 2012. Dr. Mitra is also a member of the IEEE Signal Processing Society's Technical Committee on Signal Processing for Communications and Networking (2012-2014). She is the recipient of: 2012 NAE Lillian Gilbreth Lectureship, 2011 USC Zumberge Interdisciplinary Innovation Fund (with M. El-Naggar), USC Center for Excellence in Research Fellowship (2010-2013), the Viterbi School of Engineering Dean's Faculty Service Award (2009), USC Mellon Mentoring Award (2008), IEEE Fellow (2007), Texas Instruments Visiting Professor (Fall 2002, Rice University), 2001 Okawa Foundation Award, 2000 Lumley Award for Research (OSU College of Engineering), 1997 MacQuigg Award for Teaching (OSU College of Engineering), and a 1996 National Science Foundation (NSF) CAREER Award. Dr. Mitra has held visiting appointments at: the Delft University of Technology, Stanford University, Rice University, and the Eurecom Institute. She served as co-Director of the Communication Sciences Institute at the University of Southern California from 2004-2007.

Andreas F. Molisch (S'89-M'95-SM'00-F'05) received the Dipl. Ing., Ph.D., and habilitation degrees from the Technical University of Vienna, Vienna, Austria, in 1990, 1994, and 1999, respectively. He subsequently was with AT&T (Bell) Laboratories Research (USA); Lund University, Lund, Sweden, and Mitsubishi Electric Research Labs (USA). He is now a Professor of electrical engineering with the University of Southern California, Los Angeles. His current research interests are the measurement and modeling of mobile radio channels, ultra-wideband communications and localization, cooperative communications, multiple-input-multiple-output systems, and wireless systems for healthcare. He has authored, coauthored, or edited four books (among them the textbook Wireless Communications, Wiley-IEEE Press), 14 book chapters, some 140 journal papers, and numerous conference contributions, as well as more than 70 patents and 60 standards contributions. Dr. Molisch has been an Editor of a number of journals and special issues, General Chair, Technical Program Committee Chair, or Symposium Chair of multiple international conferences, as well as Chairman of various international standardization groups. He is a Fellow of the IET, an IEEE Distinguished Lecturer, and a member of the Austrian Academy of Sciences. He has received numerous awards, most recently the 2011 James Evans Avant-Garde award of the IEEE Vehicular Technology Society, the Donald Fink Prize of the IEEE, and the Eric Sumner Award of the IEEE.

Michele Zorzi (S'89, M'95, SM'98, F'07) was born in Venice, Italy, on December 6th, 1966. He received the Laurea and the PhD degrees in Electrical Engineering from the University of Padova, Italy, in 1990 and 1994, respectively. During the Academic Year 1992/93, he was on leave at the University of California, San Diego (UCSD) as a visiting PhD student, working on multiple access in mobile radio networks. In 1993, he joined the faculty of the Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy. After spending three years with the Center for Wireless Communications at UCSD, in 1998 he joined the School of Engineering of the University of Ferrara, Italy, where he became a Professor in 2000. Since November 2003, he has been on the faculty at the Information Engineering Department of the University of Padova. His present research interests include performance evaluation in mobile communications systems, random access in mobile radio networks, ad hoc and sensor networks, energy constrained communications protocols, broadband wireless access and underwater acoustic communications and networking.

Dr. Zorzi was the Editor-In-Chief of the IEEE WIRELESS COMMUNICATIONS MAGAZINE from 2003 to 2005 and the Editor-In-Chief of the IEEE TRANSACTIONS ON COMMUNICATIONS from 2008 to 2011, and currently serves on the Editorial Board of the WILEY JOURNAL OF WIRELESS COMMUNICATIONS AND MOBILE COMPUTING. He was also guest editor for special issues in the IEEE PERSONAL COMMUNICATIONS MAGAZINE (Energy Management in Personal Communications Systems) IEEE WIRELESS COMMUNICATIONS MAGAZINE (Cognitive Wireless Networks) and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS (Multi-media Network Radios, and Underwater Wireless Communications Networks). He served as a Member-at-large of the Board of Governors of the IEEE Communications Society from 2009 to 2011.