A Simple Recursively Computable Lower Bound on the Noncoherent Capacity of Highly Underspread Fading Channels

Steven Herbert, Ian Wassell, Tian-Hong Loh, *Senior Member, IEEE,*

***Abstract*—Real-world wireless communication channels are typically highly underspread: their coherence time is much greater than their delay spread. In such situations it is common to assume that, with sufficiently high bandwidth, the capacity without Channel State Information (CSI) at the receiver (termed the *noncoherent* channel capacity) is approximately equal to the capacity with perfect CSI at the receiver (termed the *coherent* channel capacity). In this paper, we propose a lower bound on the noncoherent capacity of highly underspread fading channels, which assumes only that the delay spread and coherence time are known. Furthermore our lower bound can be calculated recursively, with each increment corresponding to a step increase in bandwidth. These properties, we contend, make our lower bound an excellent candidate as a simple method to verify that the noncoherent capacity is indeed approximately equal to the coherent capacity for typical wireless communication applications.**

We precede the derivation of the aforementioned lower bound on the information capacity with a rigorous justification of the mathematical representation of the channel. Furthermore, we also provide a numerical example for an actual wireless communication channel and demonstrate that our lower bound does indeed approximately equal the coherent channel capacity.

***Index Terms*—Gauss-Markov processes, Hidden Markov processes, Kalman filters, Noncoherent information capacity, Underspread channels, Vehicular communications.**

I. INTRODUCTION

ACTUAL communication channels are typically underspread: their delay spread is much smaller than their coherence time. A more exact definition can be made by describing the action of a wireless channel as a linear operator $\mathbb{H} : \mathcal{L}^2 \rightarrow \mathcal{L}^2$. The action of $\mathbb{H}$ can be expressed in terms of the scattering function, $\mathcal{C}_{\mathbb{H}}(\nu, \tau)$, where ν is the Doppler shift, and τ is the time delay [1]. For a Wide-Sense Stationary Uncorrelated Scattering (WSSUS) channel the non-zero region of the scattering function is defined: $\mathcal{C}_{\mathbb{H}}(\nu, \tau) = 0$ for all $(\nu, \tau) \notin [-\nu_0, \nu_0] \times [-\tau_0, \tau_0]$. Letting $\Delta_{\mathbb{H}} = 4\nu_0\tau_0$, the channel is said to be underspread if $\Delta_{\mathbb{H}} < 1$ [2]. For land-mobile channels $\Delta_{\mathbb{H}} \approx 10^{-3}$, for indoor channels $\Delta_{\mathbb{H}} \approx 10^{-7}$ [2], and for in-vehicle channels $\Delta_{\mathbb{H}}$ is typically of the order 10^{-5} [3]. Durisi *et al* [2] argue that the WSSUS model is appropriate for many scenarios and our previous work shows that, for in-vehicle channels (which are of particular interest to us [4]), the channel can be assumed to be wide-sense stationary [3, Assumption 1] and to have uncorrelated scattering [5, Assumptions 1,2]. An important metric for such channels is their *noncoherent* information capacity – i.e., their capacity where it is assumed that neither the transmitter or receiver have *a priori* Channel State Information (CSI). As identified by Durisi *et al* [6], the noncoherent capacity of underspread fading channels has been the subject of research for a long time, with early work typically focussing on characterising the noncoherent capacity in the infinite bandwidth limit [7]–[10]. It is shown that in this situation, the capacity tends to that of a fading channel with perfect CSI available at the receiver in the presence of Additive White Gaussian Noise (AWGN), known as the *coherent* capacity, [2], [6], [11]–[13], which is a generalisation of the AWGN channel capacity derived by Shannon [14].

In this paper, we derive a lower bound on the noncoherent capacity of highly underspread fading channels in terms of parameters that would typically be available in actual wireless communication systems, namely the coherence time and delay spread (if available, a statistical model for the impulse response can be used to tighten the bound – but isn't strictly necessary). We work from the intuitive starting point that CSI can be learnt increasingly accurately at the receiver by increasing the bandwidth. This leads to a lower bound on the channel capacity in a recursive form, with each iteration corresponding to a step increase in bandwidth. This is shown to be a monotonically non-decreasing function, which not only illustrates the role of bandwidth in the deployment of effective wireless communication systems, but also potentially reduces the computational load, for when the lower bound value becomes sufficiently large (i.e., for the application in question) then the recursive calculation can be stopped. We show that the Gauss-Markov scenario is the worst-case (i.e., the loosest lower bound) and hence our recursive computational method uses a *Kalman* filter [15]. This may lead to our analysis having a wider application than just the capacity of underspread channels considered here.

A. Related work and novel contributions

The noncoherent capacity of underspread fading channels has been the subject of a great deal of previous research. In

Steven Herbert is with the Computer Laboratory, University of Cambridge, CB3 0FD, UK. Contact: sjh227@cam.ac.uk

Ian Wassell is with the Computer Laboratory, University of Cambridge, CB3 0FD, UK. Contact: ijw24@cam.ac.uk

Tian Hong Loh is with the National Physical Laboratory, TW11 0LW, UK. Contact: tian.loh@npl.co.uk

The work of T. H. Loh was supported by the 2013 - 2017 Electromagnetics and Time Metrology Programme of the National Measurement Office, an Executive Agency of the U.K. Department for Business, Innovation and Skills, under Projects EMT13018.

particular Durisi *et al* investigate the noncoherent capacity of underspread channels from a general starting point, where only the non-zero region of the scattering function need be known. The approach taken by Durisi *et al* is to transmit symbols on carriers which are localised in both time and frequency and subsequently they derive a number of lower and upper bounds on the channel capacity, which have been optimised either for the low bandwidth or high bandwidth regimes [2], [6], [11], [12]. In a further paper, Durisi *et al* assess the impact of interference terms (i.e., arising from imperfect modelling of the scattering function) on their capacity bounds [13]. In this paper we assume that the channel coherence time and delay spread are known, and since the coherence time is itself deterministically related to the Doppler spread (for an appropriate definition of coherence time), our channel model is equivalent to that investigated by Durisi *et al*. We do, however, maintain that our work complements this prior literature and is especially relevant to the field of wireless communications (i.e., to aid engineers assessing the capacity of actual real-world wireless communication channels), and we make the following specific claims: the expression in this paper of the lower bound as a straightforward recursively calculable term makes it computationally simple; the property of our bound that the recursive calculation iterations correspond to equal step increases in the bandwidth means that a sufficient bandwidth to achieve a specified fraction of the coherent capacity can be calculated explicitly; and finally, the derivation in this paper is an important formalisation of an intuitive way of understanding why the capacity of underspread channels is approximately the coherent capacity – namely that the symbol carriers can be chosen to be pure frequency tones which will be highly correlated, and thus CSI can be learnt.

Another intuitive way to understand the underspread channel is that it can be formalised as a highly correlated time series, and on this subject there is also much existing literature. In particular, low Signal to Noise Ratio (SNR) [16] and high SNR [17] bounds are derived for such channels. The method for lower bounding the channel capacity concerns the estimation of the channel response as a time series in the presence of noise, for example [17], [18], that ostensibly appears to be analogous to our approach of estimating the channel as a frequency response series. However, we wish to point out that our method uses a Kalman filter for estimation, in contrast to a general Minimum Mean Squared Error estimate employed in [17], [18]. These papers do, however, highlight an important property which also applies to our channel model, specifically that the evolving time-series is either *regular* or *nonregular* (deterministic), the former representing the case where the channel variation in time is not perfectly predictable, the latter where it is, hence 'deterministic'. This categorisation as either regular or nonregular highlights an important factor in the investigation of the capacity of underspread channels, namely that in order for a continuous channel to have a capacity approximately equal to the coherent capacity at all SNR, it must be capable of representation (in some basis) as a nonregular process. As the channel model that we use in this paper is necessarily regular, we can see that the capacity can only be approximated to the coherent capacity for a limited SNR range (although our numerical example, in Section IV, suggests that the approximation is valid for typical wireless systems).

One common approach to bounding capacity of channels represented as correlated time series is to express the channel capacity as the coherent capacity minus a penalty term which is related to the lack of perfect CSI, for example [18, Equation (9)]. A particularly apt example of this approach can be seen in the paper by Deng and Haimovich [19], where the evolution of their channel in the time domain is analagous to the evolution of our channel in the frequency domain. Furthermore, [19, Equation (8)] gives an upper bound on the penalty term (which can therefore be related to a lower bound on the capacity), and the intuitive explanation given for this bound, that the CSI knowledge increases as the number of symbols increases, bears more than a passing resemblence to our own interpretation of the bound derived in this paper. The bounding method in [19] relies on finding the eigenvalues of a *Toeplitz* matrix with the number of elements in each dimension equal to the number of symbols, and therefore, including a greater number of symbols in order to tighten the bound (i.e., by reducing the uncertainty in the CSI) necessarily increases the computational load, in general, with computational complexity $\mathcal{O}(n^2 \log n)$ for n symbols [20]. Liang and Veeravalli [21], [22] use a similar method. This is quite distinct from our approach, where the Gauss-Markov property of the channel means that each additional symbol adds an equal computational load, and thus has computational complexity $\mathcal{O}(n)$. It should, however, be noted that we have implicitly sacrficed some tightness in the bound by lower bounding the capacity using that of a Gauss-Markov channel, and therefore for any given application either our bound, or one based on [19, Equation (8)] may be more appropriate.

On the subject of the capacity of Gauss-Markov processes there is also some existing prior literature. Etkin and Tse [23] use a Kalman filter to estimate the channel, and calculate the information capacity in terms of an estimation error, whilst Médard takes a similar approach in [24]. Chen *et al* [25] also investigate the capacity of Gauss-Markov channels, but produce bounds without using a Kalman filter. Our analysis differs from these as we explicitly show that the Gauss-Markov channel model lower bounds the capacity of the actual channel – and hence the Kalman filter may be used to calculate a lower bound on the channel capacity.

B. Paper Organisation

In Section II a general channel model is defined in terms of its frequency response and it is explained how this can be used in an Orthogonal Frequency Division Multiplexing (OFDM) scheme to evaluate a lower bound on the channel capacity. In Section III the main results are presented (with part of the proof in the appendix) in the form of the aforementioned lower bound, whilst in Section IV a numerical worked example is given and finally in Section V conclusions are drawn.

C. Notation

Italicised and non-italicised symbols are used for frequency and time domain variables respectively. Scalars are non-bold lower case, as in general are functions (i.e., x for the time domain, x for the frequency domain), vectors are bold lower-case (i.e., $\mathbf{x}$ for the time domain, $\boldsymbol{x}$ for the frequency domain), a single element from a vector or matrix is non-bold lower-case with subscript to denote its index (i.e., $\mathrm{x_i}$ for a vector and $\mathrm{x_{i,j}}$ for a matrix in the time domain; and x_i for a vector and $x_{i,j}$ for a matrix in the frequency domain), truncated vectors are bold lower-case with subscript to denote first element and superscript to denote final element (i.e., $\mathbf{x_i^j}$ for the time domain, $\boldsymbol{x_i^j}$ for the frequency domain) and matrices are upper-case (i.e., X for the time domain, X for the frequency domain). Convolution is denoted $*$, $(.)^*$ is used to denote complex conjugation and $(.)^T$ to denote the transpose. $\mathcal{N}$ denotes the normal distribution, $\mathcal{CN}$ denotes the complex normal distribution, $\mathcal{FT}$ denotes the Fourier transform and $\odot$ denotes the *Hadamard* (element-wise) product. The magnitude of a complex number is denoted $|.|$, as is the determinant of a matrix, however it is always clear in context which is meant. Finally, it is convenient to represent complex numbers as vectors, and when multiplied together as a matrix acting on a vector. Letting x be a number, which in general may be complex:

$$\underline{x} = \begin{bmatrix} \mathrm{Re}(x) \\ \mathrm{Im}(x) \end{bmatrix}, \quad \underline{X} = \begin{bmatrix} \mathrm{Re}(x) & \text{-}\mathrm{Im}(x) \\ \mathrm{Im}(x) & \mathrm{Re}(x) \end{bmatrix},$$

for example (letting z also be a number which may in general be complex):

$$z \times x = \underline{Z}\,\underline{x} = \underline{X}\,\underline{z}.$$

The notation can also be generalised to complex vectors: $\underline{\boldsymbol{x}} = [\underline{x}_1; \underline{x}_2; ...; \underline{x}_n]$. Whilst this notation may ostensibly appear to be unwieldy and uneccessary, it is crucial throughout the paper, in particular for Proposition 1 to show why the channel, which is in general complex Gaussian, can be respresented as bivariate Gaussian, with the two elements corresponding to the real and imaginary parts.

When handling probabilities, we make some simplifications to the notation, these do not affect the validity of what follows and are introduced purely for the purpose of brevity and clarity. For example, letting $\mathfrak{x}_1$ and $\mathfrak{x}_2$ be random variables, formally we have

$$P(\mathfrak{x}_1 = x_1) = f_{\mathfrak{x}_1}(x_1) = \mathcal{N}(x_1; 0, \sigma^2),$$

which we would write

$$P(x_1) = \mathcal{N}(x_1; 0, \sigma^2).$$

We can also extend this to conditional distributions. Again, a formal example would be

$$P(\mathfrak{x}_1 = x_1 | \mathfrak{x}_2 = x_2) = f_{\mathfrak{x}_1|\mathfrak{x}_2}(x_1) = \mathcal{N}(x_1; ax_2, \sigma_1^2),$$

which we would write

$$P(x_1|x_2) = \mathcal{N}(x_1; ax_2, \sigma_1^2),$$

where in this example a is an arbitrary scaling factor of the previous value x_2.

II. Channel Model

The action of the channel can be defined:

$$\mathrm{y(t)} = \mathrm{z_{\mathbb{H}}(t)} * \mathrm{x(t)} + \mathrm{n(t)}, \tag{1}$$

where all terms are functions in time, t, with $\mathrm{y(t)}$ as the channel output, $\mathrm{z_{\mathbb{H}}(t)}$ the channel time varying impulse response, $\mathrm{x(t)}$ the input and $\mathrm{n(t)}$ AWGN. Shannon showed that the capacity of this channel is the mutual information between $\mathrm{y(t)}$ and $\mathrm{x(t)}$ maximised over all permissible inputs $\mathrm{x(t)}$ [14], we denote this C:

$$C = \sup_{\mathrm{x(t)}} \mathcal{I}(\mathrm{x(t)}; \mathrm{y(t)}), \tag{2}$$

where $\mathcal{I}(.;.)$ is mutual information. It should be noted that it is usual to think of mutual information in terms of some discrete representation of the channel, rather than the channel itself – as indeed we shall do in this paper.

From (1), it is convenient to use the property of linearity to consider $\mathrm{z_{\mathbb{H}}(t)}$ to be the sum of two parts: $\mathrm{z_1(t)}$ being a part with impulse response truncated after τ_t seconds and also not varying within successive time-blocks of duration τ_B; $\mathrm{z_2(t)}$ is the remainder of $\mathrm{z_{\mathbb{H}}(t)}$ (i.e., $\mathrm{z_{\mathbb{H}}(t)} = \mathrm{z_1(t)} + \mathrm{z_2(t)}$). Note, '$\mathbb{H}$' has been dropped for the subscript to simplify the notation. This leads to:

$$\mathrm{y(t)} = \mathrm{z_1(t)} * \mathrm{x(t)} + \mathrm{z_2(t)} * \mathrm{x(t)} + \mathrm{n(t)}. \tag{3}$$

From (3) we can define a second channel, where $\mathrm{z_2(t)} * \mathrm{x(t)}$ is treated as white noise of the same power (denoted $\mathrm{n_1(t)}$):

$$\mathrm{y_1(t)} = \mathrm{z_1(t)} * \mathrm{x(t)} + \mathrm{n_1(t)} + \mathrm{n(t)}, \tag{4}$$

for which we define the capacity C_1:

$$C_1 = \sup_{\mathrm{x(t)}} \mathcal{I}(\mathrm{x(t)}; \mathrm{y_1(t)}). \tag{5}$$

In the OFDM scheme that we will now describe, $\mathrm{z_2(t)}$ is essentially the inter-symbol interference (ISI) plus the inter-carrier interference (ICI). Treating this interference as AWGN is consistent with the initial assumption of treating noise as AWGN (which itself is likely to comprise of interference from signals transmitted on other channels in the vicinity). As long as the power in $\mathrm{n_1(t)}$ is small relative to the power in the actual signal and the noise, then we can assert that

$$C \approx C_1, \tag{6}$$

and we can reason that this is in-fact a worst case scenario where the interference is not correlated with the signal and so cannot be used to improve the information transfer capability, and that when combined with the other interference and noise sources that it manifests itself as AWGN. We deal with the specific process of separating $\mathrm{z_{\mathbb{H}}(t)}$ into $\mathrm{z_1(t)}$ and $\mathrm{z_2(t)}$ for actual channels in Section IV.

A. Using orthogonal frequency division multiplexing to find an achievable rate

The previous deductions lead to a block fading model [26], [27]. To find an achievable rate, an OFDM [26], [27] scheme is used, with block length $T_B \approx \tau_B$ and cyclic prefix length $T_t > \tau_t$. T_B and T_t must be chosen such that WT_B and WT_t are integers (where W is the bandwidth, and $N = WT_B$ is the total number of subcarriers) – and indeed it is sensible to choose the smallest value of T_t which satisfies this condition. According to the Sampling Theorem [7], [14], the waveform in one block can be reconstructed from samples spaced $1/2W$ s apart. Performing an Inverse Discrete Fourier Transform (IDFT) on the resulting vector of samples (i.e., in the time domain) yields a vector of samples in the frequency domain. These are spaced $1/T_B$ Hz apart, which we define as Δf. Choosing to define the input signal in the frequency domain, the channel output can be expressed:

$$\boldsymbol{y} = \boldsymbol{z} \odot \boldsymbol{x} + \boldsymbol{n}, \tag{7}$$

where $\boldsymbol{y}$ is a vector of outputs, $\boldsymbol{z}$ is a vector of the channel frequency response, $\boldsymbol{x}$ is a vector of the channel symbols, $\boldsymbol{n}$ is a vector of AWGN samples and $\odot$ denotes element-wise multiplication. All these vectors are of size $N = WT_B$. Since all the elements of the vectors are complex, (7) can be expressed:

$$\underline{\boldsymbol{y}} = \underline{\boldsymbol{z}} \odot \underline{\boldsymbol{x}} + \underline{\boldsymbol{n}}. \tag{8}$$

Noting that the cyclic prefix means that cyclic convolution is identical to linear convolution performed on a Linear Time Invariant channel. This means that $\boldsymbol{z}$ is a vector of the channel frequency responses evaluated at intervals of Δf Hz.

It is important to note that in order to find a capacity, we must consider an infinite number of these time blocks (as for any single time-block there is a non-zero probability that the noise power will be such that the channel is in outage), and also owing to the power constraint we will impose on $\mathrm{x(t)}$, it is necessary to consider this over an infinite number of blocks as for any finite number of blocks it may be the case that the power constraint is not satisfied. With these two issues taken into consideration, we can define the achievable rate, R, in terms of the mutual information between the discretised $\boldsymbol{x}$ and $\boldsymbol{y}$ (i.e., for a single block):

$$\mathrm{R} = \frac{1}{N}\mathcal{I}(\boldsymbol{x}_0^{N-1}; \boldsymbol{y}_0^{N-1}) \ \mathrm{bit\,s^{-1}\,Hz^{-1}}, \tag{9}$$

which in turn can be related to the capacity of the original channel in (1):

$$C \approx C_1 \geq \frac{T_B}{T_B + T_t}\mathrm{R} \ \mathrm{bit\,s^{-1}\,Hz^{-1}}. \tag{10}$$

B. Channel frequency response

Let $\mathrm{P}_{\mathbb{H}}(\tau)$ be the instantaneous channel Power Delay Profile (PDP), from which a truncated version is defined:

$$\mathrm{P}'_{\mathbb{H}}(\tau) = \begin{cases} \mathrm{P}_{\mathbb{H}}(\tau) & \text{if } 0 \leq \tau < \tau_t \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

The bounding method requires information regarding the channel characterisation in the frequency domain. At a randomly chosen frequency, over a time period which is short compared to the delay spread of the signal, the distribution of the phase of the various multipath components will be uniform, and thus the frequency response will be a Zero Mean Circularly Symmetric (ZMCS) Gaussian random variable; defining this as $z(\omega)$, let:

$$P(\underline{z}(\omega)) = \mathcal{N}\left(\underline{z}(\omega); \underline{0}, \Sigma_z\right), \tag{12}$$

where $\mathcal{N}$ is the Gaussian distribution and ω is angular frequency (i.e., $\omega = 2\pi f$ where f is frequency in Hz) :

$$\Sigma_z = \begin{bmatrix} \sigma_z^2 & 0 \\ 0 & \sigma_z^2 \end{bmatrix}, \tag{13}$$

where σ_z^2 is the variance.

The bounding method also requires the conditional distribution of the frequency response, given the frequency response at a known separation, ($\Delta\omega = 2\pi\Delta f$), i.e., $P(z(\omega)|z(\omega - \Delta\omega))$. To formally derive this conditional probability distribution, it is necessary to make the assumption that the temporal length of $\mathrm{z_1(t)}$ is divisible into an integer, k, of arbitrarily short time intervals of duration, $\Delta\tau$, each of which is a ZMCS Gaussian random variable, with power equal to the integral of $\mathrm{P}'_{\mathbb{H}}(\tau)$ over its duration. The joint distribution of the signal from these intervals can thus be expressed as a multivariate complex Gaussian distribution: The discrete signal vector, $\mathbf{z}$, is of size K where $\mathrm{K} = \tau_t/\Delta\tau$ and the $\mathrm{k^{th}}$ element occurs at $\tau = \mathrm{k}\Delta\tau$:

$$\begin{aligned} P(\mathbf{z}) &= \mathcal{CN}(\mathbf{z}; \mathbf{0}, \Gamma, 0), && (14) \\ \Gamma &= \mathbb{E}\left(\mathbf{z}(\mathbf{z}^{*T})\right) \\ &= \mathrm{diag}(2\sigma_{\mathrm{k}}^2), && (15) \end{aligned}$$

where $\mathcal{CN}(.; \mu, \Gamma, \mathrm{C})$ is the complex Gaussian distribution with mean μ, covariance matrix Γ and relation matrix C (that is zero here, and for all subsequent occurrences), also for small $\Delta\tau$:

$$\sigma_{\mathrm{k}}^2 \approx \left(\mathrm{P}'_{\mathbb{H}}(\mathrm{k}\Delta\tau)\right)\Delta\tau. \tag{16}$$

Whilst this may seem to be an unreasonable assumption, it is actually the case that the following analysis will be a good model if $\mathrm{P}'_{\mathbb{H}}(\mathrm{t})$ can be split into time intervals in which many rays arrive, but $\mathrm{P}'_{\mathbb{H}}(\mathrm{t})$ itself only negligibly changes in value. By the WSSUS assumption these arriving rays will be independent, and with sufficiently high carrier frequency their phase can be considered to be independently drawn from a uniform distribution.

Proposition 1

For the channel with PDP defined in (11), the conditional distribution of the frequency response, given the frequency response at a known separation can be expressed:

$$P(\underline{z}(\omega)|\underline{z}(\omega - \Delta\omega)) = \mathcal{N}\left(\underline{z}(\omega); \underline{\mu}_a, \Sigma_a\right), \tag{17}$$

where:

$$\begin{aligned} \underline{\mu}_a &= \underline{A}\underline{z}(\omega - \Delta\omega), && (18) \\ \Sigma_a &= \sigma_z^2 \begin{bmatrix} 1 - |a|^2 & 0 \\ 0 & 1 - |a|^2 \end{bmatrix}, && (19) \end{aligned}$$

where $\underline{A}$ is the matrix version of the complex number a (i.e., according to the notation in Section I-C) and:

$$a = \frac{\int_0^{\tau_t} \mathrm{P}'_{\mathbb{H}}(\tau)e^{-j\Delta\omega\tau}\,\mathrm{d}\tau}{\int_0^{\tau_t} \mathrm{P}'_{\mathbb{H}}(\tau)\,\mathrm{d}\tau}. \tag{20}$$

Proof:

In the appendix. Experimental evidence using the frequency responses measured in [5] show that the theory developed in Proposition 1 is a good fit for actual channels.

III. A Lower Bound on the Channel Capacity

Recall from (9), and introducing a power constraint:

$$\mathrm{R} = \frac{1}{N}\mathcal{I}(\boldsymbol{x}_0^{N-1};\boldsymbol{y}_0^{N-1})\ \mathrm{bit\,s^{-1}\,Hz^{-1}} \tag{21}$$

subject to:

$$\mathrm{P}_{\text{ave}} \geq \lim_{N'\to\infty}\frac{1}{N'}\sum_{i=0}^{N'-1}|x_i'|^2, \tag{22}$$

where $\boldsymbol{x}'$ is a vector comprising $\boldsymbol{x}$ across an infinite number of times blocks and P_{ave} is the average power constraint. Note that this average is defined across an infinite number of input symbols, x_i, and thus an infinite number of time blocks.

The input distribution on x is chosen to be a series of Independent Identically Distributed (IID) ZMCS Gaussian random variables:

$$P(\underline{x}_i) = \mathcal{N}(\underline{x}_i;\underline{0},\Sigma_x), \tag{23}$$

where:

$$\Sigma_x = \begin{bmatrix}\sigma_x^2 & 0\\ 0 & \sigma_x^2\end{bmatrix}. \tag{24}$$

Likewise, the additive white noise is modelled as IID ZMCS Gaussian random variables:

$$P(\underline{n}_i) = \mathcal{N}(\underline{n}_i;\underline{0},\Sigma_n), \tag{25}$$

where:

$$\Sigma_n = \begin{bmatrix}\sigma_n^2 & 0\\ 0 & \sigma_n^2\end{bmatrix}. \tag{26}$$

A. Bounding idea

The aim is to lower bound (21). In essence, the bounding method is similar to a Kalman filter, where the ‘state’ is the channel response at the discrete frequencies corresponding to the input (i.e., where successive frequency responses are correlated, as shown in Proposition 1) and the noisy measurement is formed by the information signal input-output pair. For each successive discrete frequency response, some CSI is learned, and thus eventually (assuming the frequency separation between discrete frequency responses is small) the CSI approaches perfect CSI.

There is no feedback in the system, i.e., because successive values of x_i have no dependence on previous values of y_i, z_i or n_i. This property is used throughout the bounding process to simplify various expressions.

Theorem 2

There exists a lower bound, L_1, on the achievable rate:

$$\mathrm{R} \geq \mathrm{L}_1 = \frac{1}{N}\sum_{i=0}^{N-1}\mathcal{I}_i\ \mathrm{bit\,s^{-1}\,Hz^{-1}}, \tag{27}$$

where:

$$\mathcal{I}_i = \mathbb{E}\left(\log_2\left(\frac{|\mu_i'|^2\sigma_x^2+\sigma_n^2}{|x_i|^2\sigma_i^2+\sigma_n^2}\right)\right), \tag{28}$$

and:

$$\sigma_i^2 = \begin{cases}\sigma_z^2 & \text{if } i=0,\\ (1-|a|^2)\sigma_z^2 & \\ \quad +|a|^2(\sigma_{i-1}^{-2}+|x_{i-1}|^2\sigma_n^{-2})^{-1} & \text{if } i>0,\end{cases} \tag{29}$$

$$P(\underline{\mu}_i') = \mathcal{N}\left(\underline{\mu}_i';\underline{0},\begin{bmatrix}\sigma_z^2-\sigma_i^2 & 0\\ 0 & \sigma_z^2-\sigma_i^2\end{bmatrix}\right). \tag{30}$$

The terms σ_i^2 and $\underline{\mu}_i'$ represent the variance and mean of the estimate of the i^{th} frequency response respectively. Notice that we have the coherent setting when $\sigma_i^2=0$, and it is reassuring to notice that in this situation, $\mathcal{I}_i = \mathbb{E}\left(\log_2\left(1+|z_i|^2\sigma_x^2/\sigma_n^2\right)\right)$ (where z_i is equivalent to $z(\omega)$ as defined in (12) and (13)) i.e., the coherent channel capacity. Another important observation is that $\mathcal{I}_i$ is a monotonically non-decreasing function with i, which is proved in Lemma 3 (in the appendix).

Proof (of Theorem 2):

The chain rule of mutual information is used to lower bound the right-hand side (RHS) of (21):

$$\begin{aligned}\frac{1}{N}\mathcal{I}(\boldsymbol{x}_0^{N-1};\boldsymbol{y}_0^{N-1}) &= \frac{1}{N}\sum_{i=0}^{N-1}\mathcal{I}(x_i;\boldsymbol{y}_0^{N-1}|\boldsymbol{x}_0^{i-1})\\ &= \frac{1}{N}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\mathcal{I}(x_i;y_j|\boldsymbol{x}_0^{i-1},\boldsymbol{y}_0^{j-1})\\ &\geq \frac{1}{N}\sum_{i=0}^{N-1}\mathcal{I}(x_i;y_i|\boldsymbol{x}_0^{i-1},\boldsymbol{y}_0^{i-1}).\end{aligned} \tag{31}$$

Lemma 4

For the channel defined in (7), (12) and (17), the conditional distribution of the frequency response, z_i, given all previous realisations of the input, x_i, and output, y_i is a Gaussian distribution:

$$P(\underline{z}_i|\underline{\boldsymbol{x}}_0^{i-1},\underline{\boldsymbol{y}}_0^{i-1}) = \mathcal{N}(\underline{z}_i;\underline{\mu}_i,\Sigma_i-\Sigma_\epsilon), \tag{32}$$

where :

$$\Sigma_i = \begin{bmatrix}\sigma_i^2 & 0\\ 0 & \sigma_i^2\end{bmatrix}, \tag{33}$$

$$\sigma_i^2 = \begin{cases}\sigma_z^2 & \text{if } i=0,\\ (1-|a|^2)\sigma_z^2 & \\ \quad +|a|^2(\sigma_{i-1}^{-2}+|x_{i-1}|^2\sigma_n^{-2})^{-1} & \text{if } i\neq 0,\end{cases} \tag{34}$$

$$\leq \sigma_z^2, \tag{35}$$

$$P(\underline{\mu}_i) = \mathcal{N}(\underline{\mu}_i;\underline{0},\Sigma_z-(\Sigma_i-\Sigma_\epsilon)) \tag{36}$$

and Σ_ϵ is some Positive Definite Symmetric (PDS) matrix or zero. Note that the evolution of σ_i^2 is essentially in the form of a Kalman filter, with $\sigma_n^2|x_{i-1}|^{-2}$ as the noisy measurement variance.

Proof:

In the appendix.

Lemma 5

Using Lemma 4, for the channel defined in (7), (12) and (17), the mutual information can be evaluated thus:

$$\mathcal{I}(x_i; y_i|\boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1}) = \mathcal{I}(x_i; y_i|\mu_i, (\Sigma_i - \Sigma_\epsilon)). \tag{37}$$

Lemma 5 proves that μ_i and $(\Sigma_i - \Sigma_\epsilon)$ are sufficient to calculate the mutual information.

Proof:

In the appendix.

Proof of Theorem 2 (continued):

Using Lemma 5 to consider only the i^{th} frequency response, z_i, which is itself a random variable, from (8):

$$\underline{y}_i = \underline{Z}_i \underline{x}_i + \underline{n}_i, \tag{38}$$

and as shown in Lemma 4, z_i is a circularly symmetric Gaussian random variable, thus decomposing z_i such that:

$$P(\underline{z}_i') = \mathcal{N}\left(\underline{z}_i'; \underline{0}, (\Sigma_i - \Sigma_\epsilon)\right), \tag{39}$$

it follows that:

$$\underline{y}_i = \underline{M}_i \underline{x}_i + \underline{Z}_i' \underline{x}_i + \underline{n}_i, \tag{40}$$

where M is the capitalised version of μ, i.e., for the purposes of representing complex multiplication as a matrix operation. Further decomposing μ, such that:

$$P(\underline{\mu}_i') = \mathcal{N}(\underline{\mu}_i'; \underline{0}, \Sigma_z - \Sigma_i), \tag{41}$$

$$P(\underline{\mu}_i'') = \mathcal{N}(\underline{\mu}_i''; \underline{0}, \Sigma_\epsilon), \tag{42}$$

(40) can be expressed:

$$\underline{y}_i = \underline{M}_i' \underline{x}_i + \underline{M}_i'' \underline{x}_i + \underline{Z}_i' \underline{x}_i + \underline{n}_i. \tag{43}$$

Consider the mutual information:

$$\begin{aligned}\mathcal{I}(x_i; y_i|\mu_i, (\Sigma_i - \Sigma_\epsilon)) =& \mathcal{H}(y_i|\mu_i, (\Sigma_i - \Sigma_\epsilon)) \\ &- \mathcal{H}(y_i|x_i, \mu_i, (\Sigma_i - \Sigma_\epsilon)).\end{aligned} \tag{44}$$

Now consider the first term of the RHS of (44):

$$\begin{aligned}\mathcal{H}(y_i|, \mu_i, (\Sigma_i - \Sigma_\epsilon)) &\geq \mathcal{H}(y_i|\mu_i, (\Sigma_i - \Sigma_\epsilon), z_i') \\ &= \mathbb{E}(\log_2(2\pi e|\underline{M}_i' \Sigma_x (\underline{M}_i')^T \\ &\quad + \underline{M}_i'' \Sigma_x (\underline{M}_i'')^T \\ &\quad + \underline{Z}_i' \Sigma_x (\underline{Z}_i')^T + \Sigma_n|^{1/2})) \\ &\geq \mathbb{E}(\log_2(2\pi e||\mu_i'|^2 \Sigma_x + \Sigma_n|^{1/2})),\end{aligned} \tag{45}$$

and also consider the second term of the RHS of (44):

$$\begin{aligned}&\mathcal{H}(y_i|x_i, \mu_i, (\Sigma_i - \Sigma_\epsilon)) \\ &\quad = \mathbb{E}(\log_2(2\pi e|\underline{X}_i (\Sigma_i - \Sigma_\epsilon) \underline{X}_i^T + \Sigma_n|^{1/2})) \\ &\quad = \mathbb{E}\left(\log_2\left(2\pi e||x_i|^2(\Sigma_i - \Sigma_\epsilon) + \Sigma_n|^{1/2}\right)\right) \\ &\quad = \mathbb{E}\left(\log_2\left(2\pi e||x_i|^2\Sigma_i + \Sigma_n|^{1/2}|I - \Sigma_{\epsilon\,2}|^{1/2}\right)\right) &(46) \\ &\quad \leq \mathbb{E}(\log_2(2\pi e||x_i|^2\Sigma_i + \Sigma_n|^{1/2})), &(47)\end{aligned}$$

where I is the identity matrix and (47) is derived by noticing that Σ_i and Σ_n are proportional to the identity, and Σ_ϵ is PDS (as $(\Sigma_i - \Sigma_\epsilon)$ is a covariance matrix). The expression is then normalised, with the term $\Sigma_{\epsilon\,2}$ introduced to equal Σ_ϵ divided by the appropriate scale factor. The resultant term $|I - \Sigma_{\epsilon\,2}|$ is shown to be less than one in Lemma 8, hence the lower bound.

Substituting (45) and (47) into (44):

$$\begin{aligned}\mathcal{I}(x_i; y_i|\mu_i, (\Sigma_i - \Sigma_\epsilon)) =& \mathcal{H}(y_i|\mu_i, (\Sigma_i - \Sigma_\epsilon)) \\ &- \mathcal{H}(y_i|x_i, \mu_i, (\Sigma_i - \Sigma_\epsilon)) \\ \geq& \mathbb{E}(\log_2(2\pi e||\mu_i'|^2\Sigma_x + \Sigma_n|^{1/2})) \\ &- \mathbb{E}(\log_2(2\pi e||x_i|^2\Sigma_i + \Sigma_n|^{1/2})) \\ =& \mathbb{E}\left(\log_2\left(\frac{|\mu_i'|^2\sigma_x^2 + \sigma_n^2}{|x_i|^2\sigma_i^2 + \sigma_n^2}\right)\right), \\ =& \mathcal{I}_i,\end{aligned} \tag{48}$$

which proves Theorem 2.

Corollary 9

A very important corollary to Theorem 2 is:

$$\mathcal{I}_i = \mathbb{E}\left(\log_2\left(\frac{\text{SNR}|\mu_i''''|^2 + 1}{\text{SNR}|x_i'''|^2(\sigma_i''')^2 + 1}\right)\right). \tag{49}$$

where:

$$P(\underline{x}_i''') = \mathcal{N}\left(\underline{x}_i'''; \underline{0}, \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\right), \tag{50}$$

$$(\sigma_i''')^2 = \begin{cases} 1 & \text{if } i = 0, \\ (1 - |a|^2) & \\ \quad + |a|^2\left((\sigma_{i-1}''')^{-2} + \text{SNR}|x_{i-1}'''|^2\right)^{-1} & \text{if } i > 0, \end{cases} \tag{51}$$

$$P(\underline{\mu}_i'''') = \mathcal{N}\left(\underline{\mu}_i''''; \underline{0}, \begin{bmatrix}1 - (\sigma_i''')^2 & 0\\ 0 & 1 - (\sigma_i''')^2\end{bmatrix}\right). \tag{52}$$

$$\text{SNR} = \frac{\sigma_z^2\sigma_x^2}{\sigma_n^2}. \tag{53}$$

That is, the bound can be expressed such that it only relies on the SNR and not the individual realisations of σ_z^2, σ_x^2 and σ_n^2.

Proof:

In the appendix.

Corollary 10

There exists a lower bound, L_2, on the achievable rate:

$$\text{R} \geq \text{L}_2 = \frac{1}{N}\sum_{i=0}^{N-1} \mathcal{I}_i' \ \text{bit}\,\text{s}^{-1}\,\text{Hz}^{-1}, \tag{54}$$

where:

$$\mathcal{I}_i' = \mathbb{E}\left(\max\left(\log_2\left(\frac{|\mu_i'|^2\sigma_x^2 + \sigma_n^2}{|x_i|^2\sigma_i^2 + \sigma_n^2}\right), 0\right)\right). \tag{55}$$

This corollary simply allows $\mathcal{I}_i$ to be replaced with zero if it is negative. This is useful for computation, and slightly tightens the bound.

Proof:

This corollary is obviously true, as mutual information is always non-negative and therefore any values of $(x_i, \mu_i', \sigma_i^2)$ for which our estimate happens to be negative can be replaced with zero, without affecting the validity of the bound.

Corollary 11

There exist lower bounds $\mathrm{L}_{1\mathrm{A}}$ and $\mathrm{L}_{2\mathrm{A}}$ such that:

$$\mathrm{R} \geq \mathrm{L}_1 \geq \mathrm{L}_{1\mathrm{A}} = \frac{1}{N}\left(\sum_{i=0}^{N''-1} \mathcal{I}_i + (N - N'')\mathcal{I}_{N''}\right), \quad (56)$$

$$\mathrm{R} \geq \mathrm{L}_2 \geq \mathrm{L}_{2\mathrm{A}} = \frac{1}{N}\left(\sum_{i=0}^{N''-1} \mathcal{I}_i' + (N - N'')\mathcal{I}_{N''}'\right), \quad (57)$$

where:

$$0 < N'' \leq N. \quad (58)$$

This is useful as the lower bounds L_1 and L_2 can themselves be lower bounded by $\mathrm{L}_{1\mathrm{A}}$ and $\mathrm{L}_{2\mathrm{A}}$ respectively. This means computation of the mutual information could be halted when a sufficiently tight lower bound has been achieved.

Proof:

The first part of Corollary 11 (56) arises trivially from the monotonically non-increasing form of $\mathcal{I}_i$, as shown in Lemma 3. The second part of Corollary 11 follows from this using the same argument as Corollary 10.

IV. Numerical Example

As identified in Section I, in-vehicle channels are of particular interest, which have PDPs that decay exponentially with time [5], [28]. For such channels, the delay spread is infinite, and therefore the underspread property is only approximate. At sufficiently large SNR, the fact that this is only an approximation becomes significant, as identified by Durisi *et al* [6], and further supported by Koch and Lapidoth [29] who show that, for a discrete exponentially decaying channel, the capacity is bounded in the SNR. It is therefore important not to over-generalise the applicability of our bound, and thus it is evaluated for a typical example application. A suitable example application is a Wireless Sensor Network operating using *Zigbee* [30] – note that even though this does not use OFDM, the principle of lower bounding the capacity using an arbitrary coding scheme to find an achievable rate still applies. To evaluate the lower bound, it is necessary to find appropriate parameters to substitute into the expressions (49) and (51). These parameters derive from the fundamental parameters (i.e., the cavity time constant and system SNR) via the parameters required to model the channel as a block fading system (i.e., the block length, cyclic prefix length and the adjustment to the SNR to account for the ISI and ICI).

A. Parameters

The coherent capacity can be expressed:

$$\begin{aligned} C =& \mathbb{E}\left(\log_2\left(1 + |z|^2 \frac{\sigma_x^2}{\sigma_n^2}\right)\right), \\ =& \mathbb{E}\left(\log_2\left(1 + |z''|^2 \frac{\sigma_z^2 \sigma_x^2}{\sigma_n^2}\right)\right), \\ =& \mathbb{E}\left(\log_2\left(1 + |z''|^2 \mathrm{SNR}\right)\right), \end{aligned} \quad (59)$$

where, by definition:

$$P(\underset{\sim}{z}'') = \mathcal{N}\left(\underset{\sim}{z}''; \underset{\sim}{0}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right). \quad (60)$$

It can be shown that the specified 250 kbit/s data rate for a single Zigbee channel (i.e., occupying a frequency band of width 5 MHz) can be achieved at an SNR of 0.0180 (for this analysis it is irrelevant that actual Zigbee systems would typically have a much higher SNR).

From our previous measurements [5], the time constant, τ_c, of the exponential decay in a typical vehicle cavity is 17.2 ns. Regarding the choice of cyclic prefix length, T_t, and noting that this must correspond to an integer number of samples in the time domain, consider choosing just a single time sample duration to be the cyclic prefix, which is equal to 200 ns. From this, the power contained in the ISI, denoted $\mathrm{P}_{\mathrm{ISI}}$ as a portion of the signal power, denoted P_{S}, can be found

$$\frac{\mathrm{P}_{\mathrm{ISI}}}{\mathrm{P}_{\mathrm{S}}} = \frac{\mathrm{k}_1 \int_{T_t}^{\infty} e^{-\frac{\tau}{\tau_c}} \, \mathrm{d}\tau}{\mathrm{k}_1 \int_0^{\infty} e^{-\frac{\tau}{\tau_c}} \, \mathrm{d}\tau}, \quad (61)$$

Substituting $T_t = 200$ ns and $\tau_c = 17.2$ ns into (61) yields $\mathrm{P}_{\mathrm{ISI}}/\mathrm{P}_{\mathrm{S}} = 8.91 \times 10^{-6}$, that we deem to be sufficiently small. Regarding the choice of block length, T_B, an appropriate criteria (i.e., for this example) is the time duration during which 0.99 of the energy is expected to remain undisturbed. This is equivalent to stating that the power contained in the ICI, $\mathrm{P}_{\mathrm{ICI}}$, is 0.01 of P_{S} which can be found from the work in [3] in which it was demonstrated that the autocorrelation function at time separation, t, of the in-vehicle channel could be modelled in the form $\sigma^2 e^{-\theta t}$, and that an appropriate value for the parameter θ is 0.475. Letting $t = T_B$:

$$\begin{aligned} 0.99 =& \frac{\sigma^2 e^{-\theta T_B}}{\sigma^2} \\ \implies \quad T_B =& -\frac{\log_e(0.99)}{\theta} \\ =& 0.02 \text{ s}. \end{aligned} \quad (62)$$

Upon closer inspection of [3] it may seem a little unreasonable to extrapolate our time correlation model to such a short time interval when our measurements were spaced 0.125 s apart, however more sophisticated analysis based around fitting a curve to the Power Spectral Density plots in [3, Fig. 5] lead to an estimation of $T_B = 0.0053$ s with the same criteria that 0.99 of the energy does not vary. This provides further evidence that our value of T_B is of the correct order of magnitude, and we shall continue using the more conservative value of $T_B = 0.0053$ s. The criteria of 0.99 of the energy not varying is equivalent to defining the variation owing to ICI as $\mathrm{P}_{\mathrm{ICI}}/\mathrm{P}_{\mathrm{S}} = 0.01$. According to the definitions in Section II-A, we have

$\Delta f = 1/T_B = 189$ Hz and $N = WT_B = 26500$. These deductions allow us to find an adjusted value of the SNR to account for the interference, we denote this SNR$'$. Note that the following expression is valid for the worst case, where all of the ICI manifests itself as interference on other subcarriers (in reality some of it will experience a Doppler shift which will move it out of the band of interest). Also, it has been assumed that all of the channel variation occurs in non-truncated part of the PDP:

$$\begin{aligned}\mathrm{SNR}' =& \frac{\mathrm{P_S} - \mathrm{P_{ISI}} - \mathrm{P_{ICI}}}{\mathrm{P_N} + \mathrm{P_{ISI}} + \mathrm{P_{ICI}}} \\ =& \frac{1 - \frac{\mathrm{P_{ISI}}}{\mathrm{P_S}} - \frac{\mathrm{P_{ICI}}}{\mathrm{P_S}}}{\mathrm{SNR}^{-1} + \frac{\mathrm{P_{ISI}}}{\mathrm{P_S}} + \frac{\mathrm{P_{ICI}}}{\mathrm{P_S}}} \\ =& 0.0178, \end{aligned} \tag{63}$$

where $\mathrm{P_N}$ is the noise power.

The final parameter that it is necessary to find is $|a|^2$. This is, however, complicated by the fact that we can no longer assume that $\mathrm{z_1(t)}$ has the same PDP as $\mathrm{z_{\mathbb{H}}(t)}$. We can mitigate this issue by finding the worst case of $|a|^2$, given T_t and T_B. This is a useful result in its own right, as it allows the lower bounding of the capacity even if only the coherence time and delay spread are known. First, we must establish how $\mathrm{L_1}$, $\mathrm{L_2}$, $\mathrm{L_{1A}}$ and $\mathrm{L_{2A}}$, vary with $|a|^2$.

Lemma 12

$$\mathcal{I}_i = \mathrm{F_1}(|a|^2), \tag{64}$$

for all permissible values of $|a|^2$ and where $\mathrm{F_1}$ is a monotonically non-decreasing function. This is sufficient for $\mathrm{L_1}$ and $\mathrm{L_{1A}}$ to be monotonically non-decreasing functions of $|a|^2$.

Proof:

In the appendix.

Corollary 13

An important corollary to Lemma 12 is:

$$\mathcal{I}'_i = \mathrm{F_2}(|a|^2), \tag{65}$$

for all permissible values of $|a|^2$ and where $\mathrm{F_2}$ is a monotonically non-decreasing function. This is sufficient for $\mathrm{L_2}$ and $\mathrm{L_{2A}}$ to be monotonically non-decreasing functions of $|a|^2$.

Proof:

Corollary 13 can be trivially shown to be true by considering that $\mathcal{I}'_i$ only differs from $\mathcal{I}_i$ as defined in (132) because for some values of $\boldsymbol{x}$ the term in the braces in (132) will be replaced by zero. For this variation, the analysis in Lemma 12 remains valid.

From Lemma 12 and Corollary 13, we can see that we must choose $\mathrm{P}'_{\mathbb{H}}(\tau)$ to minimise $|a|^2$ in order to establish the worst case scenario. Note that as the coding scheme described in the analysis does not depend on $|a|^2$, this is sufficient to propose a lower bound, and thus show how the capacity can be bounded even if only the delay spread and coherence time are known.

Lemma 14

For underspread channels by definition $T_t < T_B$, for highly underspread channels it is not unreasonable to allow only channels where $T_t < 0.5T_B$. In this case $\mathrm{P}'_{\mathbb{H}}(\tau)$ which lower bounds $|a|^2$ according to (20) is:

$$\mathrm{P}'_{\mathbb{H}}(\tau) = \frac{1}{2}\left(\delta(\tau) + \delta(\tau - T_t)\right), \tag{66}$$

where $\delta(.)$ is the Dirac delta function. Substituting (66) into (20) and noting that $\Delta\omega = 2\pi/T_B$ yields:

$$|a|^2 = \frac{1}{4}\left((1 + \cos(2\pi T_t/T_B))^2 + (\sin(2\pi T_t/T_B))^2\right). \tag{67}$$

Proof:

This Lemma arises trivially from a geometric argument. Notice that $\mathrm{P}'_{\mathbb{H}}(\tau)e^{-j2\pi T_t/T_B}$ can be considered to consist of several complex vectors, each of whose direction must be at an angle of less than π radians by the inequality $T_t < 0.5T_B$. In this case, the magnitude of the sum of these vectors (and hence the magnitude squared) is minimised when $\mathrm{P}'_{\mathbb{H}}(\tau)$ consists of two equal length vectors separated by the greatest possible angle. This is expressed mathematically in Lemma 14.

Notice from (67) that as $T_t/T_B \to 0$ (i.e., the channel becomes more underspread) then $|a|^2 \to 1$. This corresponds to the case where the channel response at a certain frequency is known deterministically given the response at the preceding frequency. In this instance, perfect estimation of the channel state information is limited only by the lack of knowledge concerning the value of initial frequency response and the noise. Observing (29) with $|a|^2 \to 1$ we can see that the effects of both of these factors, i.e., the lack of knowledge concerning the value of initial frequency response and the noise, on the channel state information estimation diminish to zero with an infinite number of samples. Thus in the infinite bandwidth limit the channel information capacity will approach coherent capacity. This result explicitly demonstrates an intuitively obvious property of the underspread channel. Substituting T_B and $\tau_t = T_t$ into (67) yields $|a|^2 = 1 - 1.45 \times 10^{-8}$.

B. Results

Our lower bounds $\mathrm{L_1}$, $\mathrm{L_2}$, L_{1A} and L_{2A}, in (27), (54), (56) and (57) are on R, using (10) we define:

$$C_1 \geq \mathrm{L}_{C2} = \frac{T_B}{T_B + T_t}\mathrm{L_2}. \tag{68}$$

Fig. 1(a) shows the variation of L_{C2} with N, demonstrating the learning process (recall that each increment in N corresponds to a step increase in bandwidth). Fig. 1(b) shows the capacity bound L_{C2} as a function of frequency for the entire 5 MHz band of interest. Note that for both subplots in Fig. 1 the capacity has been normalised by dividing it by the coherent capacity. Whilst performing these simulations we noticed that at low SNR the capacity bound is sensitive to the value of $|a|^2$. One way to understand this is to consider that $|a|^2$ characterises how much the channel has changed between successive subcarriers (where $|a|^2 = 1$ means no change at

Lower Bound	$\boldsymbol{\Delta}_{\mathbb{H}}$	SNR/ dB	Bandwidth/ Hz	$\mathbf{C/C_{coh}}$
L_{C2}	2×10^{-5}	-17.5	37.8×10^{3}	0.92
L_{C2}	2×10^{-5}	-17.5	5×10^{6}	>0.99
[6, Fig. 2.4]	10^{-5}	-17.5	1.4×10^{9}	0.99
[6, Fig. 2.10a]	10^{-4}	-17.5	-	0.42
[6, Fig. 2.10b]	10^{-6}	-17.5	-	0.84

TABLE I
COMPARISON OF OUR LOWER BOUND WITH THOSE IN THE LITERATURE

all). At low SNR the Kalman filter estimation is buried in noise, so even with relatively little change between successive subcarriers it may be the case that the channel evolves more rapidly than it can be estimated (for the majority of realisations of $\boldsymbol{x}$ – noting that the mutual information is averaged over all realisations of $\boldsymbol{x}$). For low SNR channels it may be that an alternative characterisation of the capacity is more appropriate, for example Kennedy [8]. It should, however, be noted that for low SNR channels which are extremely stable (i.e., $|a|^2 \approx 1$) our lower bound demonstrates explicitly the bandwidth which is sufficient to achieve a certain proportion of the coherent capacity – as the Kalman filter could take many subcarriers to learn the channel sufficiently well to estimate receiver CSI. In such cases, we contend that our bound is useful for wireless applications as it demonstrates the role that bandwidth has in estimating the channel in an intuitive way, which also explicitly allows the calculation of a sufficient bandwidth according to a user defined criteria. This contrasts with some of the other bounds which are only valid as SNR$\rightarrow\infty$.

Our motivation for this work is not specifically to improve upon the existing bounds, but to provide an intuitive and computationally simple method of verifying that the noncoherent capacity is approximately equal to the coherent capacity for typical real-world underspread fading channels. Nonetheless, for greater context, we now present a comparison as best we can with the lower bounds found in the literature, where it appears that the work by Durisi *et al* [6] provides the most appropriate numerical results. These results are presented in Table I, with the first two rows corresponding to our lower bound, L_{C2} evaluated at the right-hand extremities of the two plots in Fig. 1, and the remaining three rows corresponding to the most comparable bounds in the literature at the same SNR (0.0178 = -17.5 dB). Note that the final column of Table I refers to the respective lower bounds normalised by the coherent capacity, also note that the value of 2×10^{-5} for $\boldsymbol{\Delta}_{\mathbb{H}}$ of our channel comes from the specific values used, and is consistent with our previous statement that $\boldsymbol{\Delta}_{\mathbb{H}}$ is of the order 10^{-5} for in-vehicle channels. We can see that our lower bound performs comparably to those in the literature and, importantly, it becomes tight at a much lower bandwidth (note that for the [6, Fig. 2.10] no value was given for bandwidth). This provides evidence that strengthens our assertion that our bounding method is suitable for finding a sufficient bandwidth to achieve a certain fraction of the coherent capacity.

V. CONCLUSIONS

The major contribution that we have made in this paper is a lower bound on the noncoherent capacity of underspread fading channels, where only the channel coherence time, delay spread and bandwidth are known. Our approach is firmly rooted in our previous experience gained obtaining channel measurements, and our observation that most real-world channels are highly underspread and it is therefore useful to provide a computationally simple technique to verify that for any given channel the noncoherent capacity is approximately equal to the coherent capacity. The lower bound is also expressed as a non-decreasing recursive function, with each iteration corresponding to a discrete increase in bandwidth, which we contend makes it suitable for wireless communication engineering applications, as it explicitly provides a method of calculating the bandwidth which is sufficient to attain a user-defined percentage of the coherent capacity. It can therefore be concluded that our lower-bound complements the existing work which has largely been undertaken from a pure information theoretical perspective.

For real-world channels, typically the impulse response will have infinite time duration and thus the underspread property is actually an SNR dependent assumption. An important open problem is evaluating or bounding the noncoherent capacity of fading channels which may in general be overspread.

APPENDIX A
PROOF OF LEMMAS, PROPOSITIONS AND COROLLARIES

Proposition 1 proof:

By a Fourier transform, (14) can be used to approximate the frequency response at ω and $(\omega-\Delta\omega)$, for sufficiently small values of $\Delta\tau$:

$$P\left(\begin{bmatrix} z(\omega) \\ z(\omega-\Delta\omega)\end{bmatrix}\right) \approx \mathcal{CN}\left(\begin{bmatrix} z(\omega) \\ z(\omega-\Delta\omega)\end{bmatrix}; \mathbf{0}, \begin{bmatrix}\boldsymbol{f}\\ \boldsymbol{g}\end{bmatrix}\Gamma\begin{bmatrix}\boldsymbol{f}\\ \boldsymbol{g}\end{bmatrix}^{*T}, 0\right) \tag{69}$$

where $\boldsymbol{f}$ and $\boldsymbol{g}$ are row vectors, each of size K, with k^{th} elements:

$$f_{\mathrm{k}} = e^{-j\omega\mathrm{k}\Delta\tau}, \tag{70}$$

$$g_{\mathrm{k}} = e^{-j(\omega-\Delta\omega)\mathrm{k}\Delta\tau}. \tag{71}$$

Let:

$$\begin{bmatrix}\boldsymbol{f}\\ \boldsymbol{g}\end{bmatrix}\Gamma\begin{bmatrix}\boldsymbol{f}\\ \boldsymbol{g}\end{bmatrix}^{*T} = \begin{bmatrix}\gamma_1 & \gamma_2\\ \gamma_3 & \gamma_4\end{bmatrix}, \tag{72}$$

then, noting that $\tau = \mathrm{k}\Delta\tau$:

$$\gamma_1 = \sum_{\mathrm{k}=0}^{\mathrm{K}-1} e^{-j\omega\tau}\left(2\mathrm{P}'_{\mathbb{H}}(\tau)\Delta\tau\right)e^{j\omega\tau}, \tag{73}$$

$$\gamma_2 = \sum_{\mathrm{k}=0}^{\mathrm{K}-1} e^{-j\omega\tau}\left(2\mathrm{P}'_{\mathbb{H}}(\tau)\Delta\tau\right)e^{j(\omega-\Delta\omega)\tau}, \tag{74}$$

$$\gamma_3 = \sum_{\mathrm{k}=0}^{\mathrm{K}-1} e^{-j(\omega-\Delta\omega)\tau}\left(2\mathrm{P}'_{\mathbb{H}}(\tau)\Delta\tau\right)e^{j\omega\tau}, \tag{75}$$

$$\gamma_4 = \sum_{\mathrm{k}=0}^{\mathrm{K}-1} e^{-j(\omega-\Delta\omega)\tau}\left(2\mathrm{P}'_{\mathbb{H}}(\tau)\Delta\tau\right)e^{-j(\omega-\Delta\omega)\tau}. \tag{76}$$

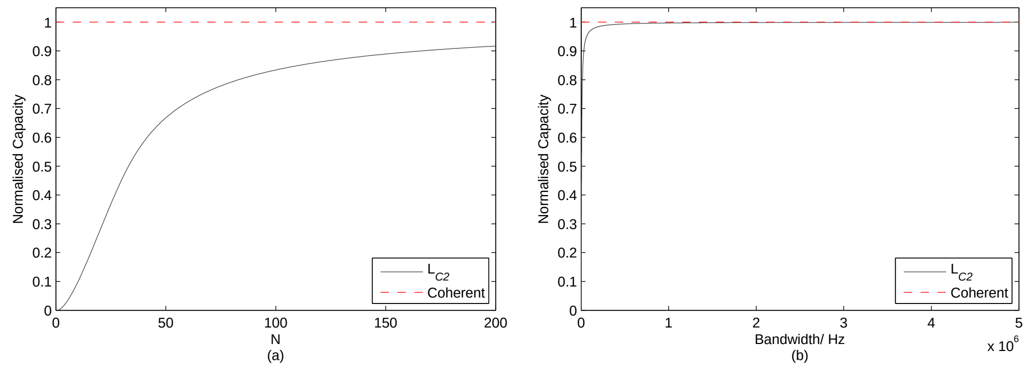


Fig. 1. Capacity lower bound: (a) with N; (b) with bandwidth.

Let: $\Delta\tau \rightarrow 0$ (and thus adjusting K such that τ_t does not vary):

$$\gamma_1 = \gamma_4 = \int_0^\infty 2\mathrm{P}'_{\mathbb{H}}(\tau)\,\mathrm{d}\tau, \tag{77}$$

$$\gamma_2 = \int_0^\infty 2\mathrm{P}'_{\mathbb{H}}(\tau)e^{-j\tau\Delta\omega}\,\mathrm{d}\tau, \tag{78}$$

$$\gamma_3 = \int_0^\infty 2\mathrm{P}'_{\mathbb{H}}(\tau)e^{j\tau\Delta\omega}\,\mathrm{d}\tau. \tag{79}$$

Noticing that $[z(\omega); z(\omega-\Delta\omega)]$ is ZMCS complex Gaussian, it can be expressed as a multivariate Gaussian:

$$P\left(\begin{bmatrix} \underset{\sim}{z}(\omega) \\ \underset{\sim}{z}(\omega-\Delta\omega)\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix} \underset{\sim}{z}(\omega) \\ \underset{\sim}{z}(\omega-\Delta\omega)\end{bmatrix}; \underline{\mathbf{0}}, \Sigma_{\mathrm{t}}\right), \tag{80}$$

where:

$$\Sigma_t = \sigma_z^2 \begin{bmatrix} 1 & 0 & \mathrm{Re}(a) & -\mathrm{Im}(a) \\ 0 & 1 & \mathrm{Im}(a) & \mathrm{Re}(a) \\ \mathrm{Re}(a) & \mathrm{Im}(a) & 1 & 0 \\ -\mathrm{Im}(a) & \mathrm{Re}(a) & 0 & 1 \end{bmatrix}, \tag{81}$$

where a is as defined in (20) and:

$$\sigma_z^2 = \int_0^\infty \mathrm{P}'_{\mathbb{H}}(\tau)\,\mathrm{d}\tau. \tag{82}$$

From this multivariate distribution, the conditional distribution of $(z(\omega)|z(\omega-\Delta\omega))$ can be expressed in the form given in (17), (18) and (19), thus proving Proposition 1.

Lemma 3

For the definition of $\mathcal{I}_i$ in (28), for $i \geq 0$:

$$\mathcal{I}_{i+1} - \mathcal{I}_i \geq 0. \tag{83}$$

Proof:

Lemma 3 is proven by demonstrating that, for any sequence of inputs $\boldsymbol{x}$, an extra input directly prior to this (i.e., an extra subcarrier usage a single frequency interval lower than that used by x_0) will never decrease the mutual information. Starting from (28):

$$\mathcal{I}_{i+1} = \int_{\boldsymbol{x}} \int_{(|\mu'''|^2)} P(\boldsymbol{x})P(|\mu'''|^2) \left\{ \log_2\left(\frac{|\mu'''|^2(\sigma_z^2 - \sigma_{i+1}^2)\sigma_x^2 + \sigma_n^2}{|x_{i+1}|^2\sigma_{i+1}^2 + \sigma_n^2}\right)\right\} \mathrm{d}(|\mu'''|^2)\,\mathrm{d}\boldsymbol{x}, \tag{84}$$

where

$$P(\mu''') = \mathcal{N}\left(\underline{\mu}'''; \underline{0}, \begin{bmatrix} 1 & 0 \\ 0 & 1\end{bmatrix}\right). \tag{85}$$

As $\boldsymbol{x}$ is simply a variable of integration, we define $\boldsymbol{x}' = \boldsymbol{x}_1^{i+1}$. Using this to re-write (84) yields

$$\mathcal{I}_{i+1} = \int_{\boldsymbol{x}'} \int_{x_0} \int_{(|\mu'''|^2)} P(\boldsymbol{x}')P(x_0)P(|\mu'''|^2) \left\{ \log_2\left(\frac{|\mu'''|^2(\sigma_z^2 - (\sigma_i')^2)\sigma_x^2 + \sigma_n^2}{|x_i'|^2(\sigma_i')^2 + \sigma_n^2}\right)\right\} \mathrm{d}(|\mu'''|^2)\,\mathrm{d}x_0\,\mathrm{d}\boldsymbol{x}', \tag{86}$$

where, from (29)

$$(\sigma_i')^2 = \begin{cases} \sigma_z^2 & \text{if } i = -1, \\ (1-|a|^2)\sigma_z^2 + |a|^2(\sigma_z^{-2} + |x_0|^2\sigma_n^{-2})^{-1} & \text{if } i = 0, \\ (1-|a|^2)\sigma_z^2 + |a|^2((\sigma_{i-1}')^{-2} + |x_{i-1}'|^2\sigma_n^{-2})^{-1} & \text{if } i > 0. \end{cases}$$

Again, noticing that $\boldsymbol{x}'$ and μ''' are simply variables of integration, without loss of generality we can write from (28)

$$\mathcal{I}_i = \int_{\boldsymbol{x}'} \int_{(|\mu'''|^2)} P(\boldsymbol{x}')P(|\mu'''|^2) \left\{ \log_2\left(\frac{|\mu'''|^2(\sigma_z^2 - (\sigma_i'')^2)\sigma_x^2 + \sigma_n^2}{|x_i'|^2(\sigma_i'')^2 + \sigma_n^2}\right)\right\} \mathrm{d}(|\mu'''|^2)\,\mathrm{d}\boldsymbol{x}' \tag{87}$$

$$= \int_{\boldsymbol{x}'} \int_{x_0} \int_{(|\mu'''|^2)} P(\boldsymbol{x}')P(x_0)P(|\mu'''|^2) \left\{ \log_2\left(\frac{|\mu'''|^2(\sigma_z^2 - (\sigma_i'')^2)\sigma_x^2 + \sigma_n^2}{|x_i'|^2(\sigma_i'')^2 + \sigma_n^2}\right)\right\} \mathrm{d}(|\mu'''|^2)\,\mathrm{d}x_0\,\mathrm{d}\boldsymbol{x}', \tag{88}$$

where

$$(\sigma_i'')^2 = \begin{cases} \sigma_z^2 & \text{if } i=0, \\ (1-|a|^2)\sigma_z^2 \\ +|a|^2((\sigma_{i-1}'')^{-2} + |x_{i-1}'|^2\sigma_n^{-2})^{-1} & \text{if } i>0. \end{cases}$$

Noticing that (88) is equivalent to (87) as there is no dependence on x_0 in the expression. We can therefore see that

$$\begin{aligned}\mathcal{I}_{i+1} - \mathcal{I}_i = &\int_{\boldsymbol{x}'}\int_{x_0}\int_{(|\mu'''|^2)} P(\boldsymbol{x}')P(x_0)P(|\mu'''|^2) \\ &\Bigg\{\Bigg(\log_2\left(|\mu'''|^2(\sigma_z^2-(\sigma_i')^2)\sigma_x^2+\sigma_n^2\right) \\ &-\log_2\left(|\mu'''|^2(\sigma_z^2-(\sigma_i'')^2)\sigma_x^2+\sigma_n^2\right)\Bigg) \\ &+\Bigg(\log_2\left(|x_i'|^2(\sigma_i'')^2+\sigma_n^2\right) \\ &-\log_2\left(|x_i'|^2(\sigma_i')^2+\sigma_n^2\right)\Bigg)\Bigg\}\mathrm{d}(|\mu'''|^2)\,\mathrm{d}x_0\,\mathrm{d}\boldsymbol{x}'.\end{aligned} \tag{89}$$

Given that $\log_2$ is a monotonically increasing function, we can see that $(\sigma_i')^2 \le (\sigma_i'')^2$ is a sufficient condition to prove Lemma 3 (as in (89) $\log_2$ always operates on positive terms). We can prove this by induction, consider first $i=0$ (note that even though we have allowed $i=-1$ in (86), however for the expression in question, (89) we are only concerned with $i \ge 0$):

$$\begin{aligned}(\sigma_i')^2 = &(1-|a|^2)\sigma_z^2 + |a|^2(\sigma_z^{-2}+|x_0|^2\sigma_n^{-2})^{-1} \\ \le &\sigma_z^2 \\ = &(\sigma_i'')^2.\end{aligned} \tag{90}$$

We can also see that the evolution of $(\sigma_i')^2$ and $(\sigma_i'')^2$ is defined by the same function, which is of the form

$$\gamma_1' = (1-|a|^2)\sigma_z^2 + |a|^2((\gamma_0')^{-1} + |x_{i-1}|^2\sigma_n^{-2})^{-1} \tag{91}$$

where γ_1' and γ_0' are variables, and $|a|^2$, $|x_{i-1}|^2$ and σ_n^2 are constants. We can see that

$$\begin{aligned}\frac{\mathrm{d}\gamma_1'}{\mathrm{d}\gamma_0'} = &|a|^2(\gamma_0')^{-2}((\gamma_0')^{-1}+|x_{i-1}|^2\sigma_n^{-2})^{-2} \\ \ge &0.\end{aligned} \tag{92}$$

This result can be understood by visualising the plot of γ_1' versus γ_0' as a monotonically increasing curve. As both $((\sigma_i')^2, (\sigma_{i+1}')^2)$ and $((\sigma_i'')^2, (\sigma_{i+1}'')^2)$ are points on this curve, this means that $(\sigma_i')^2 \le (\sigma_i'')^2 \implies (\sigma_{i+1}')^2 \le (\sigma_{i+1}'')^2$, completing the inductive proof, which in turn proves Lemma 3.

Lemma 4 proof:

Lemma 4 is proven using mathematical induction, for $i=0$:

$$P(\underline{z}_0) = \mathcal{N}(\underline{z}_0; \underline{0}, \Sigma_z), \tag{93}$$

which is true by definition, as there are no previous values of x_i and y_i upon which z_0 is conditioned.

Next, it is shown that if Lemma 4 is true for $\underline{z}_{i-1}$ then it is also true for $\underline{z}_i$

$$\begin{aligned}P(z_i|\boldsymbol{x}_0^{i-1},\boldsymbol{y}_0^{i-1}) = &\int_{z_{i-1}} P(z_i|z_{i-1},\boldsymbol{x}_0^{i-1},\boldsymbol{y}_0^{i-1}) \\ &P(z_{i-1}|\boldsymbol{x}_0^{i-1},\boldsymbol{y}_0^{i-1})\,\mathrm{d}z_{i-1}.\end{aligned} \tag{94}$$

Consider the first term in the integrand in (94)

$$P(z_i|z_{i-1},\boldsymbol{x}_0^{i-1},\boldsymbol{y}_0^{i-1}) = P(z_i|z_{i-1},\boldsymbol{x}_0^{i-2},\boldsymbol{y}_0^{i-2}), \tag{95}$$

it is valid to drop x_{i-1} and y_{i-1} from the conditioning, as these are only correlated with z_i through z_{i-1} as $\underline{z}$ are in effect the states of a hidden Markov process. It is known from Proposition 1 that:

$$P(z_i|z_{i-1}) = \mathcal{N}(\underline{z}_i; \underline{A}\underline{z}_{i-1}, \Sigma_a). \tag{96}$$

Consider the multivariate Gaussian:

$$\begin{aligned}P&\left(\begin{bmatrix}\underline{z}_i & | & \underline{z}_{i-1},\boldsymbol{x}_0^{i-2} \\ \boldsymbol{y}_0^{i-2} & | & \underline{z}_{i-1},\boldsymbol{x}_0^{i-2}\end{bmatrix}\right) \\ &= P\left(\begin{bmatrix}\underline{z}_i & | & \underline{z}_{i-1} \\ \boldsymbol{y}_0^{i-2} & | & \underline{z}_{i-1},\boldsymbol{x}_0^{i-2}\end{bmatrix}\right) \\ &= \mathcal{N}\left(\begin{bmatrix}\underline{z}_i \\ \boldsymbol{y}_0^{i-2}\end{bmatrix}; \begin{bmatrix}\underline{A}\underline{z}_{i-1} \\ \underline{\boldsymbol{\alpha}}\end{bmatrix}, \begin{bmatrix}\Sigma_a & \beta \\ \beta^T & \delta\end{bmatrix}\right),\end{aligned} \tag{97}$$

where the values of $\underline{\boldsymbol{\alpha}}$, β and δ are unimportant for this analysis. Therefore:

$$P(\underline{z}_i|\underline{z}_{i-1},\boldsymbol{x}_0^{i-1},\boldsymbol{y}_0^{i-1}) = \mathcal{N}(\underline{z}_i; \underline{A}\underline{z}_{i-1}+\underline{\mu}_\epsilon, \Sigma_a - \Sigma_\epsilon'). \tag{98}$$

An alternative expression is useful for the subsequent analysis:

$$\begin{aligned}P(\underline{z}_i|\underline{z}_{i-1},\boldsymbol{x}_0^{i-1},\boldsymbol{y}_0^{i-1}) = &P(\underline{A}^{-1}\underline{z}_i|\underline{z}_{i-1},\boldsymbol{x}_0^{i-1},\boldsymbol{y}_0^{i-1}) \\ = &\mathcal{N}(\underline{A}^{-1}\underline{z}_i; \underline{z}_{i-1} \\ &+\underline{A}^{-1}\underline{\mu}_\epsilon, \underline{A}^{-1}\Sigma_a\underline{A}^{-T} \\ &-\underline{A}^{-1}\Sigma_\epsilon'\underline{A}^{-T}),\end{aligned} \tag{99}$$

where the value of $\underline{\mu}_\epsilon$ is unimportant for this analysis, as is $\Sigma_\epsilon' = \beta^T\delta^{-1}\beta$ which is a PDS matrix (i.e., because δ is a covariance matrix), or zero if the underlying process is actually a Markov process.

Consider the second term of the integrand in (94), and notice that it can be split into two conditionally independent terms:

$$P(\underline{z}_{i-1}|\underline{x}_{i-1},\underline{y}_{i-1}) = \mathcal{N}\left(\underline{z}_{i-1}; (\underline{X}_{i-1})^{-1}\underline{y}_{i-1}, |x_{i-1}|^{-2}\Sigma_n\right), \tag{100}$$

and:

$$P(\underline{z}_{i-1}|\underline{\boldsymbol{x}}_0^{i-2},\underline{\boldsymbol{y}}_0^{i-2}) = \mathcal{N}\left(\underline{z}_{i-1}; \underline{\mu}_{i-1}, \Sigma_{i-1} - \Sigma_\epsilon''\right), \tag{101}$$

i.e., from the definition of Lemma 4 in (32). This expression is valid for $i=1$, as $\underline{\boldsymbol{x}}_0^{i-2}, \underline{\boldsymbol{y}}_0^{i-2}$ consists of no elements, and thus it represents the unconditional distribution of $\underline{z}_0$, which is valid by the definition in (34), i.e., with $\Sigma_\epsilon'' = 0$. Note that in general Σ_ϵ'' is a PDS matrix, or zero and this notation is simply used to distinguish it from Σ_ϵ.

The conditional independence allows (100) and (101) to be fused together as a Kalman filter, i.e., in which the pair

(x_{i-1}, y_{i-1}) forms a measurement, and there exists some prior estimate of the state $P(z_{i-1}|\boldsymbol{x}_0^{i-2}, \boldsymbol{y}_0^{i-2})$. This leads to:

$$P(\underline{z}_{i-1}|\underline{\boldsymbol{x}}_0^{i-1}, \underline{\boldsymbol{y}}_0^{i-1}) = \mathcal{N}(\underline{z}_{i-1}; \underline{\mu}_\alpha, \Sigma_\alpha), \tag{102}$$

where the value of $\underline{\mu}_\alpha$ is unimportant for this analysis, and:

$$\begin{aligned}\Sigma_\alpha =& \big(|x_{i-1}|^2\Sigma_n^{-1} + (\Sigma_{i-1} - \Sigma_\epsilon'')^{-1}\big)^{-1} && (103)\\ =& \big(|x_{i-1}|^2\Sigma_n^{-1} + \Sigma_{i-1}^{-1} + (\Sigma_\epsilon''')^{-1}\big)^{-1} && (104)\\ =& \big(|x_{i-1}|^2\Sigma_n^{-1} + \Sigma_{i-1}^{-1}\big)^{-1} - \Sigma_\epsilon'''', && (105)\end{aligned}$$

where Σ_ϵ''' and Σ_ϵ'''' are PDS matrices. Lemma 6 is applied to $(\Sigma_{i-1} - \Sigma_\epsilon'')^{-1}$ in (103), noticing that Σ_{i-1} is proportional to the identity, to derive (104). Lemma 7 is applied to the RHS of (104), noticing that $(|x_{i-1}|^2\Sigma_n^{-1} + \Sigma_{i-1}^{-1})$ is proportional to the identity, to derive (105). The Lemmas are stated and proved subsequently in this appendix.

Substituting (99) and (102) into (94), and performing the resulting convolution yields:

$$\begin{aligned}P(z_i|\boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1}) =& \mathcal{N}(\underline{A}^{-1}\underline{z}_i; \underline{A}^{-1}\underline{\mu}_\epsilon + \underline{\mu}_\alpha, \underline{A}^{-1}\Sigma_a\underline{A}^{-T}\\ &-\underline{A}^{-1}\Sigma_\epsilon'\underline{A}^{-T} + \Sigma_\alpha)\\ =& \mathcal{N}\left(\underline{z}_i; \underline{\mu}_\epsilon + \underline{A}\underline{\mu}_\alpha, \Sigma_a - \Sigma_\epsilon' + \underline{A}\Sigma_\alpha\underline{A}^T\right)\\ =& \mathcal{N}\left(\underline{z}_i; \underline{\mu}_i, \Sigma_i - \Sigma_\epsilon\right), && (106)\end{aligned}$$

where $\underline{\mu}_i$ is defined later in (111), and by performing substitutions from (19) and (105):

$$\begin{aligned}\Sigma_\epsilon =& \Sigma_\epsilon' + \Sigma_\epsilon''''', && (107)\\ \Sigma_i =& \Sigma_a + |a|^2(\Sigma_{i-1}^{-1} + |x_{i-1}|^2\Sigma_n^{-1})^{-1}\\ =& (1-|a|^2)\Sigma_z + |a|^2(\Sigma_{i-1}^{-1} + |x_{i-1}|^2\Sigma_n^{-1})^{-1}. && (108)\end{aligned}$$

Noticing that if Σ_{i-1} is proportional to the identity, then so is Σ_i, let:

$$\Sigma_i = \begin{bmatrix}\sigma_i^2 & 0\\ 0 & \sigma_i^2\end{bmatrix}, \tag{109}$$

where:

$$\sigma_i^2 = (1-|a|^2)\sigma_z^2 + |a|^2(\sigma_{i-1}^{-2} + |x_{i-1}|^2\sigma_n^{-2})^{-1}. \tag{110}$$

Consider that the overall distribution of z must be preserved, regardless of the input and noise, therefore:

$$P(\underline{\mu}_i) = \mathcal{N}(\underline{\mu}_i; \underline{0}, \Sigma_z - (\Sigma_i - \Sigma_\epsilon)). \tag{111}$$

To prove the final part of Lemma 4, i.e., that $\sigma_z^2 \geq \sigma_i^2$, consider again proof by induction. From (93) it is known that $\sigma_0^2 = \sigma_z^2$, and thus consider (110):

$$\begin{aligned}\sigma_z^2 \geq& \sigma_{i-1}^2\\ \geq& (\sigma_{i-1}^{-2} + |x_{i-1}|^2\sigma_n^{-2})^{-1}\\ \geq& (1-|a|^2)\sigma_z^2 + |a|^2(\sigma_{i-1}^{-2} + |x_{i-1}|^2\sigma_n^{-2})^{-1}\\ =& \sigma_i^2. && (112)\end{aligned}$$

Lemma 5 proof:

$$\begin{aligned}\mathcal{I}(x_i; y_i|\boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1}) =& \mathcal{H}(x_i|\boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})\\ &-\mathcal{H}(x_i|y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})\\ =& \mathcal{H}(x_i)\\ &-\mathcal{H}(x_i|y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1}), && (113)\end{aligned}$$

where the conditioning in the first term of the RHS is dropped as x_i are IID random variables, and there is no feedback in the channel. Regarding the second term of the RHS of (113), consider:

$$\begin{aligned}P(x_i|y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1}) =& \int_{z_i} P(x_i|z_i, y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})\\ &P(z_i|y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})\,\mathrm{d}z_i\\ =& \int_{z_i} P(x_i|z_i, y_i)\\ &P(z_i|y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})\,\mathrm{d}z_i, && (114)\end{aligned}$$

where the conditioning in the first term of the integral has been dropped, because x_i is conditionally independent of $(\boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})$ given (z_i, y_i). Notice also that, for the second term on the RHS of (114), y_i and $(\boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})$ are conditionally independent given z_i, thus:

$$\begin{aligned}P(z_i|y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1}) =& \frac{P(y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1}|z_i)P(z_i)}{P(y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})}\\ =& \frac{P(y_i|z_i)P(\boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1}|z_i)P(z_i)}{P(y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})}\\ =& \frac{P(z_i|y_i)P(y_i)}{P(z_i)}\\ &\times\frac{P(z_i|\boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})P(\boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})}{P(z_i)}\\ &\times\frac{P(z_i)}{P(y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})}\\ =& k\frac{P(z_i|y_i)P(z_i|\boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1})}{P(z_i)}\\ =& \mathrm{F}(z_i, y_i, \mu_i, (\Sigma_i - \Sigma_\epsilon))\\ =& P(z_i|y_i, \mu_i, (\Sigma_i - \Sigma_\epsilon)), && (115)\end{aligned}$$

where k is a constant, and F is a function. Therefore, substituting (115) into (114):

$$\begin{aligned}P(x_i|y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1}) =& \int_{z_i} P(x_i|z_i, y_i)\\ &P(z_i|y_i, \mu_i, (\Sigma_i - \Sigma_\epsilon))\,\mathrm{d}z_i\\ =& \int_{z_i} P(x_i|z_i, y_i, \mu_i, (\Sigma_i - \Sigma_\epsilon))\\ &P(z_i|y_i, \mu_i, (\Sigma_i - \Sigma_\epsilon))\,\mathrm{d}z_i\\ =& P(x_i|y_i, \mu_i, (\Sigma_i - \Sigma_\epsilon))\\ \implies \mathcal{H}(x_i|y_i, \boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1}) =& \mathcal{H}(x_i|y_i, \mu_i, (\Sigma_i - \Sigma_\epsilon)), && (116)\end{aligned}$$

substituting (116) into (113)

$$\begin{aligned}\mathcal{I}(x_i; y_i|\boldsymbol{x}_0^{i-1}, \boldsymbol{y}_0^{i-1}) =& \mathcal{H}(x_i) - \mathcal{H}(x_i|y_i, \mu_i, (\Sigma_i - \Sigma_\epsilon))\\ =& \mathcal{I}(x_i; y_i|\mu_i, (\Sigma_i - \Sigma_\epsilon)), && (117)\end{aligned}$$

which proves Lemma 5.

Lemma 6

For identity matrix, I, and PDS matrix D, there exists a PDS matrix D' such that:

$$(I-D)^{-1} = I + D'. \tag{118}$$

Proof:

From [31, pp. 151]:

$$(I-D)^{-1} = I + (I-D)^{-1}D. \tag{119}$$

Given that $(I-D)$ is a PDS matrix, $(I-D)^{-1}$ is also a PDS matrix. Also, since D is a PDS matrix, then $(I-D)^{-1}D$ must be a PDS matrix, which is renamed D' to prove Lemma 6.

Lemma 7

For identity matrix, I, and PDS matrix D, there exists a PDS matrix D' such that:

$$(I+D)^{-1} = I - D'. \tag{120}$$

Proof:

From [31, pp. 151]:

$$(I+D)^{-1} = I - (I+D)^{-1}D. \tag{121}$$

Given that $(I+D)$ is a PDS matrix, $(I+D)^{-1}$ is also a PDS matrix. Also, since D is a PDS matrix, then $(I+D)^{-1}D$ must be a PDS matrix, which is renamed D' to prove Lemma 7.

Lemma 8

For 2×2 identity matrix, I, and 2×2 PDS matrix, D, with $(I-D)$ also a 2×2 PDS matrix, it follows that:

$$|I-D| < 1. \tag{122}$$

Proof:

Let:

$$D = \begin{bmatrix} d_1 & d_2 \\ d_2 & d_4 \end{bmatrix}, \tag{123}$$

therefore:

$$\begin{aligned} |I-D| &= (1-d_1)(1-d_4) - d_2^2 \\ &= 1 - d_1 - d_4 + d_1 d_4 - d_2^2, \end{aligned} \tag{124}$$

consider $0 < d_1, d_4 < 1$, therefore:

$$d_1, d_4 > d_1 d_4 \tag{125}$$

$$\begin{aligned} \implies \quad 1 &> 1 - d_1 - d_4 + d_1 d_4 - d_2^2 \\ &= |I-D|, \end{aligned} \tag{126}$$

thus proving Lemma 8.

Corollary 9 proof:

Rearranging (29) we can see:

$$(\sigma_i''')^2 = \frac{1}{\sigma_z^2}\sigma_i^2, \tag{127}$$

furthermore, using the location scale property of the Gaussian distribution we can make the following substitutions without loss of generality:

$$x_i''' = \frac{1}{\sigma_x} x_i, \tag{128}$$

$$\mu_i'''' = \frac{1}{\sigma_z}\mu_i', \tag{129}$$

substituting these into (28) yields:

$$\begin{aligned} \mathcal{I}_i &= \mathbb{E}\left(\log_2\left(\frac{\sigma_z^2|\mu_i''''|^2\sigma_x^2 + \sigma_n^2}{\sigma_x^2|x_i'''|^2\sigma_z^2(\sigma_i''')^2 + \sigma_n^2}\right)\right) \\ &= \mathbb{E}\left(\log_2\left(\frac{\frac{\sigma_z^2\sigma_x^2}{\sigma_n^2}|\mu_i''''|^2 + 1}{\frac{\sigma_z^2\sigma_x^2}{\sigma_n^2}|x_i'''|^2(\sigma_i''')^2 + 1}\right)\right), \end{aligned} \tag{130}$$

and thus making the substitution, SNR $= \sigma_z^2\sigma_x^2/\sigma_n^2$, proves Corollary 9.

Lemma 12 proof:

A sufficient condition to prove the Lemma is that the differential, $\mathrm{d}(\mathcal{I}_i)/\mathrm{d}(|a|^2)$, is non-negative in the region of interest. Thus by the chain rule, it is sufficient that both $\mathrm{d}(\mathcal{I}_i)/\mathrm{d}(\sigma_i^2)$ and $\mathrm{d}(\sigma_i^2)/\mathrm{d}(|a|^2)$ are non-positive in the region of interest:

$$\frac{\mathrm{d}(\mathcal{I}_i)}{\mathrm{d}(|a|^2)} = \frac{\mathrm{d}(\mathcal{I}_i)}{\mathrm{d}(\sigma_i^2)} \times \frac{\mathrm{d}(\sigma_i^2)}{\mathrm{d}(|a|^2)}. \tag{131}$$

Consider the first term in the RHS of (131), starting from

(28), and using the definition of μ''' in (85):

$$\begin{aligned}\mathcal{I}_i=&\int_{\boldsymbol{x}} P(\boldsymbol{x})\int_{(|\mu'''|^2)} P(|\mu'''|^2)\\&\left\{\log_2\left(\frac{|\mu'''|^2(\sigma_z^2-\sigma_i^2)\sigma_x^2+\sigma_n^2}{|x_i|^2\sigma_i^2+\sigma_n^2}\right)\right\}\mathrm{d}(|\mu'''|^2)\,\mathrm{d}\boldsymbol{x}\\=&\int_{\boldsymbol{x}} P(\boldsymbol{x})\int_{(|\mu'''|^2)} P(|\mu'''|^2)\\&\left\{\log_2\left(|\mu'''|^2(\sigma_z^2-\sigma_i^2)\sigma_x^2+\sigma_n^2\right)\right.\\&\left.-\log_2\left(|x_i|^2\sigma_i^2+\sigma_n^2\right)\right\}\mathrm{d}(|\mu'''|^2)\,\mathrm{d}\boldsymbol{x} \qquad (132)\end{aligned}$$

$$\begin{aligned}\Longrightarrow&\\\frac{\mathrm{d}\mathcal{I}_i}{\mathrm{d}\sigma_i^2}=&\int_{\boldsymbol{x}} P(\boldsymbol{x})\int_{(|\mu'''|^2)} P(|\mu'''|^2)\\&\left\{\frac{\mathrm{d}\left(\log_2\left(|\mu_i'''|^2(\sigma_z^2-\sigma_i^2)\sigma_x^2+\sigma_n^2\right)\right)}{\mathrm{d}\sigma_i^2}\right.\\&\left.-\frac{\mathrm{d}\left(\log_2\left(|x_i|^2\sigma_i^2+\sigma_n^2\right)\right)}{\mathrm{d}\sigma_i^2}\right\}\mathrm{d}(|\mu'''|^2)\,\mathrm{d}\boldsymbol{x}\\=&\int_{\boldsymbol{x}} P(\boldsymbol{x})\int_{(|\mu'''|^2)} P(|\mu'''|^2)\\&\left\{-\frac{|\mu'''|^2\sigma_x^2}{\log_e(2)\left(|\mu'''|^2(\sigma_z^2-\sigma_i^2)\sigma_x^2+\sigma_n^2\right)}\right.\\&\left.-\frac{|x_i|^2}{\log_e(2)\left(|x_i|^2\sigma_i^2+\sigma_n^2\right)}\right\}\mathrm{d}(|\mu'''|^2)\,\mathrm{d}\boldsymbol{x}\\\leq&0. \qquad (133)\end{aligned}$$

Consider the second term in the RHS of (131). As σ_i^2 forms a series, we use proof by induction. Consider σ_0^2 which is equal to σ_z^2:

$$\begin{aligned}\frac{\mathrm{d}(\sigma_0^2)}{\mathrm{d}(|a|^2)}=&0\\\leq&0. \qquad (134)\end{aligned}$$

Now consider σ_i^2, from (29) assuming that $\mathrm{d}(\sigma_i^2)/\mathrm{d}(|a|^2)\leq 0$:

$$\begin{aligned}\frac{\mathrm{d}(\sigma_i^2)}{\mathrm{d}(|a|^2)}=&-\sigma_z^2+(\sigma_{i-1}^{-2}+|x_{i-1}|^2\sigma_n^{-2})^{-1}\\&+\frac{|a|^2}{(\sigma_{i-1}^{-2}+|x_{i-1}|^2\sigma_n^{-2})^2(\sigma_{i-1}^2)^2}\frac{\mathrm{d}(\sigma_{i-1}^2)}{\mathrm{d}(|a|^2)} \qquad (135)\\\leq&0, \qquad (136)\end{aligned}$$

which is obtained by noticing that (35) means that the magnitude of the second term on the RHS of (135) is smaller than the magnitude of the first term on the RHS of (135), and thus these two terms sum together to give a non-positive number. The remainder of the proof follows by the induction. Substituting (133) and (136) into (131) proves Lemma 12.

Acknowledgment

The authors would like to thank Tobi Koch and Jossy Sayir for generously lending their expertise in a series of introductory discussions. Also to Edward Mottram for his help with Lemma 14, and to Ramji Venkataramanan for proof reading and internally reviewing the paper.

References

[1] P. Bello, "Characterization of randomly time-variant linear channels," *Communications Systems, IEEE Transactions on*, vol. 11, no. 4, pp. 360–393, 1963.

[2] G. Durisi, U. Schuster, H. Bolcskei, and S. Shamai, "Noncoherent capacity of underspread fading channels," *Information Theory, IEEE Transactions on*, vol. 56, no. 1, pp. 367–395, 2010.

[3] S. Herbert, I. Wassell, T.-H. Loh, and J. Rigelsford, "Characterizing the spectral properties and time variation of the in-vehicle wireless communication channel," *Communications, IEEE Transactions on*, vol. 62, no. 7, pp. 2390–2399, July 2014.

[4] J. Dawson, D. Hope, M. Panitz, and C. Christopoulos, "Wireless networks in vehicles," in *Electromagnetic Propagation in Structures and Buildings, 2008 IET Seminar on*, dec. 2008, pp. 1 – 6.

[5] S. Herbert, T. Loh, and I. Wassell, "An impulse response model and Q factor estimation for vehicle cavities," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2013.

[6] G. Durisi, V. I. Morgenshtern, H. Bolcskei, U. G. Schuster, and S. Shamai (Shitz), *Information theory of underspread WSSUS channels*, 2011, pp. 65–116. [Online]. Available: http://publications.lib.chalmers.se/records/fulltext/local_139143.pdf

[7] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: John Wiley & Sons, Inc., 1968.

[8] R. S. Kennedy, *Fading Dispersive Communication Channels*. Wiley, 1969.

[9] J. Pierce, "Ultimate performance of M -ary transmissions on fading channels," *Information Theory, IEEE Transactions on*, vol. 12, no. 1, pp. 2–5, Jan 1966.

[10] A. Viterbi, "Performance of an m -ary orthogonal communication system using stationary stochastic signals," *Information Theory, IEEE Transactions on*, vol. 13, no. 3, pp. 414–422, July 1967.

[11] G. Durisi, H. Bolcskei, and S. Shamai, "Capacity of underspread wssus fading channels in the wideband regime," in *Information Theory, 2006 IEEE International Symposium on*, 2006, pp. 1500–1504.

[12] G. Durisi and H. Bolcskei, "High-SNR capacity of wireless communication channels in the noncoherent setting: A primer," *AEU - International Journal of Electronics and Communications*, vol. 65, no. 8, pp. 707 – 712, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1434841111000264

[13] G. Durisi, V. Morgenshtern, and H. Bolcskei, "On the sensitivity of continuous-time noncoherent fading channel capacity," *Information Theory, IEEE Transactions on*, vol. 58, no. 10, pp. 6372–6391, Oct 2012.

[14] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October 1948. [Online]. Available: http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf

[15] R. E. Kalman, "A new approach to linear filtering and prediction problems," 1960. [Online]. Available: http://www.cs.unc.edu/ welch/kalman/media/pdf/Kalman1960.pdf

[16] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidoth, "Low-SNR capacity of noncoherent fading channels," *Information Theory, IEEE Transactions on*, vol. 55, no. 4, pp. 1555–1574, April 2009.

[17] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *Information Theory, IEEE Transactions on*, vol. 51, no. 2, pp. 437–446, Feb 2005.

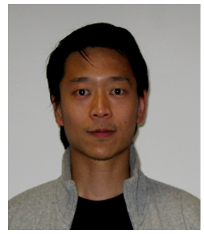
[18] M. Dorpinghaus, H. Meyr, and R. Mathar, "On the achievable rate of stationary Rayleigh flat-fading channels with Gaussian inputs," *Information Theory, IEEE Transactions on*, vol. 59, no. 4, pp. 2208–2220, April 2013.

[19] X. Deng and A. Haimovich, "Information rates of time varying Rayleigh fading channels," in *Communications, 2004 IEEE International Conference on*, vol. 1, June 2004, pp. 573–577 Vol.1.

[20] V. Y. Pan and Z. Q. Chen, "The complexity of the matrix eigenproblem," in *Proceedings of the Thirty-first Annual ACM Symposium on Theory of Computing*, ser. STOC '99. New York, NY, USA: ACM, 1999, pp. 507–516. [Online]. Available: http://doi.acm.org/10.1145/301250.301389

[21] Y. Liang and V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *Information Theory, IEEE Transactions on*, vol. 50, no. 12, pp. 3095–3110, Dec 2004.

[22] ——, “Capacity of noncoherent time-selective block Rayleigh flat-fading channel,” in *Information Theory, 2002. Proceedings. 2002 IEEE International Symposium on*, 2002, p. 166.
[23] R. Etkin and D. Tse, “Degrees of freedom in some underspread MIMO fading channels,” *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1576–1608, April 2006.
[24] M. Medard, “The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel,” *Information Theory, IEEE Transactions on*, vol. 46, no. 3, pp. 933–946, May 2000.
[25] R.-R. Chen, B. Hajek, R. Koetter, and U. Madhow, “On fixed input distributions for noncoherent communication over high-SNR Rayleigh-fading channels,” *Information Theory, IEEE Transactions on*, vol. 50, no. 12, pp. 3390–3396, Dec 2004.
[26] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. New York, NY, USA: Cambridge University Press, 2005.
[27] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
[28] J. Andersen, K. L. Chee, M. Jacob, G. Pedersen, and T. Kurner, “Reverberation and absorption in an aircraft cabin with the impact of passengers,” *Antennas and Propagation, IEEE Transactions on*, vol. 60, no. 5, pp. 2472 –2480, may 2012.
[29] T. Koch and A. Lapidoth, “On multipath fading channels at high SNR,” *Information Theory, IEEE Transactions on*, vol. 56, no. 12, pp. 5945–5957, 2010.
[30] [Online]. Available: http://www.zigbee.org/
[31] S. R. Searle, *Matrix Algebra Useful for Statistics*. New York, NY, USA: John Wiley & Sons, Inc., 1982.



Tian-Hong Loh (S03M05-SM’15) received the B.Eng. degree (first class) from Nottingham Trent University, Nottingham, U.K., in 1999 and the Ph.D. degree from the University of Warwick, Coventry, U.K., in 2005, both in electrical and electronic engineering. In 2005, he joined the National Physical Laboratory, Teddington, U.K., as a Higher Research Scientist, and since 2009, he has been a Senior Research Scientist. His responsibilities include carry out fundamental research and develop measurement technologies in support of the electronics and communication industry. Since 2011, he has been appointed as RF and Microwave Technical Theme Leader, involved in physical program formulation and strategy development. He leads work at NPL on applied and computational electromagnetic metrology research and hold five patents. He is currently visiting industrial fellow at Cambridge University, visiting reader at Surrey University, visiting senior lecturer at Queen Mary University of London, Committee member of UK URSI Commission A (Electromagnetic Metrology), and project coordinator of an EURAMET (European Association of National Metrology Institutes) EMPIR (European Metrology Programme for Innovation and Research) project on Metrology for 5G Communications. His current research interests include metamaterials, computational electromagnetics, small antenna, smart antennas, multiple-inputmultiple-output antennas, electromagnetic compatibility, body-centric, wireless sensor network and 5G communications.

Steven Herbert received the B.A. degree in 2010 [subsequently promoted to M.A. (Cantab) in 2013], the M.Eng. degree in 2010 and the Ph.D. degree in 2015 all from the University of Cambridge, Cambridge, U.K. He is currently a visitor at the University of Cambridge Computer Laboratory. He is the author of a number of publications and patent applications in wireless communications, signal processing, information networks and electromagnetic wave propagation. His current research interests include mesh network design and routing and all aspects of classical and quantum information theory.

Ian Wassell received his B.Sc. and B.Eng. degrees from the University of Loughborough in 1983, and his Ph.D. degree from the University of Southampton in 1990. He is a Senior Lecturer at the Computer Laboratory, University of Cambridge and has experience in excess of 25 years in the simulation and design of radio communication systems gained via a number of positions in industry and higher education. He has published more than 190 papers and his current research interests include: wireless sensor networks, cooperative wireless networks, propagation modelling, cognitive radio and compressive sensing. He is a member of the IET and a Chartered Engineer.