

A Compressed Sensing Receiver for UWB Impulse Radio in Bursty Applications like Wireless Sensor Networks

Anand Oka*

Lutz Lampe

Dept. of Electrical and Computer Engineering, University of British Columbia, Canada.

Abstract

We propose a novel receiver for Ultra-Wide-band Impulse-Radio communication in Wireless Sensor Networks, which are characterized by bursty traffic and severe power constraints. The receiver is based on the principle of Compressed Sensing, and exploits the sparsity of the transmitted signal to achieve reliable demodulation from a relatively small number of projections. The projections are implemented in an analog front-end as correlations with tractable test-functions, and a joint decoding of the time of arrival and the data bits is done by a DSP back-end using an efficient quadratic program. The proposed receiver differs from extant schemes in the following respects: (i) It needs neither a high-rate analog-to-digital converter nor wide-band analog delay lines, and can operate in a significantly under-sampled regime. (ii) It is robust to large timing uncertainty and hence the transmitter need not waste power on explicit training headers for timing synchronization. (iii) It can operate in a regime of heavy inter-symbol interference (ISI), and therefore allows a very high baud rate (close to the Nyquist rate). (iv) It has a built-in capability to blindly acquire and track the channel response irrespective of line-of-sight/non-line-of-sight conditions. We demonstrate that the receiver's performance remains close to the maximum likelihood receiver under every scenario of under-sampling, timing uncertainty, ISI, and channel delay spread.

Key words: ultra-wide-band, low power physical layer, compressed sensing, maximum likelihood sequence estimation, inter-symbol interference, timing uncertainty, wireless sensor networks

*Corresponding author.

Email addresses: anando@ece.ubc.ca (Anand Oka), lampe@ece.ubc.ca (Lutz Lampe)

URL: www.ece.ubc.ca/~anando/ (Anand Oka), www.ece.ubc.ca/~lampe/ (Lutz Lampe)

1. Introduction

Ultra-Wide-band (UWB) radio [1][2][3] is widely regarded to be a promising candidate for power-constrained applications like Wireless Sensor Networks (WSN) [4], on account of its ability to trade bandwidth for a reduced transmit power, its ability to coexist with extant licensed narrow-band systems, and its localized nature which is ideal for short-haul multi-hop transport. Impulse-radio (IR), in particular, is especially well suited to WSNs due to its low cost, immunity to severe multi-path fading even in indoor environments [3], and potential to provide accurate localization [5].

Notwithstanding these advantages, a UWB-IR physical layer has not been widely adopted due to the relative difficulty of implementing a coherent UWB-IR receiver for WSNs, where the transmitter in each ‘mote’ periodically makes short bursts of transmissions, and goes into a sleep mode in the relatively long inter-burst intervals to save power. In a burst, a small payload is modulated either in the amplitude or temporal position of very narrow (~ 1.0 nanosecond) IR pulses transmitted at a pulsing rate (baud rate) f_{baud} . In indoor settings this radio signal typically encounters a channel having tens or even hundreds of resolved multi-path components and a large temporal dispersion of 10 – 100 nanoseconds [6]. Although a coherent all-digital receiver that implements maximum likelihood sequence estimation (MLSE) would be optimal in terms of the bit error rate (BER), it is impractically complex when there is heavy Inter-Symbol Interference (ISI), and furthermore requires an expensive and power-hungry high-speed analog-to-digital (A/D) converter on account of the large bandwidth [3]. On the other hand, analog equalization of the channel is also a formidable challenge, and results in a significant signal-to-noise ratio (SNR) penalty relative to MLSE. Consequently, a pragmatic solution often used [7] is to avoid ISI all-together by using a sufficiently low baud-rate $f_{baud} \ll 2\Omega = f_{nyquist}$, where Ω is the signal bandwidth. The MLSE then simplifies to a matched filter (MF), which can be implemented entirely in the analog domain (no A/D) in the form of a maximum ratio combining (MRC) rake [3]. Of course, the choice of a low baud-rate translates to a low instantaneous data rate, longer channel occupancy, and a reduced number of supported transmitters.

Even if one avoids high-speed A/D and MLSE complexity by using a small f_{baud} and a rake receiver, one still needs an accurate up-to-date estimate of the channel impulse response, and a timing synchronization that is correct to within a small fraction of the pulse width T_{pulse} . The problem of UWB channel estimation has been investigated in [8] under the assumption of Nyquist rate sampling, and in [9] based on a Compressed Sensing approach. Although the problem of timing synchronization is, in principle, subsumed in the problem of channel estimation, the variations in the time of arrival (TOA) due to the drift of the transmitter’s baud clock and the motion of the transmitter/receiver,¹ are

¹Small scale relative motion only alters the over-all TOA, while leaving the shape of channel response invariant.

fast relative to the changes in the physical environment. Such rapid changes in timing cannot be tracked by the channel estimator and hence there is essentially no timing information available from one burst to the next. Hence timing acquisition has to be done *afresh* for each burst, via techniques like correlation, serial search [3] or ‘dirty template’ [10], and to achieve this we need to modulate a *sufficiently long sequence of training bits* as a preamble to each burst before we modulate the comparatively small set of information carrying bits. This is highly wasteful of power and undermines the very rationale of using UWB-IR.

Alternative non-coherent approaches suggested in literature include energy detecting (ED) receivers [11], transmit-reference (TR) receivers [12] [13], and differential transmit-reference (DTR) receivers [14], all of which in principle need neither the channel response nor accurate timing synchronization. However it is quite difficult to ensure robust operation of such non-coherent schemes in the regime of significant ISI [15], and hence they too usually remain restricted to low baud rates. Moreover, ED receivers suffer a very large SNR penalty relative to coherent systems, as do TR and DTR to a lesser extent. TR/DTR also involve the use of a very long analog delay line (of $1/f_{baud}$ seconds), which is difficult to implement with the requisite accuracy.

In this paper we offer a solution that combines the advantages of MLSE coherent receivers (high system gain, high baud rate, ability to operate in ISI) and non-coherent receivers (low complexity, robustness to timing uncertainty and ignorance about the channel response), while avoiding their respective drawbacks. We propose a flexible and robust receiver architecture that performs a ‘joint’ decoding of timing and amplitude information. This joint decoding is inspired by the principle of compressed sensing (CS) proposed by [16] [17]. The uncertainty in the arrival time of each burst is treated as ‘sparsity’ in the classical sense of [16] [17] and therefore tackled automatically in the reconstruction process which is a variation of the L_1 -minimization used by [16] [17]. Furthermore, the fact that the amplitudes are antipodal $\{+1, -1\}$, and hence the overall transmit signal belongs to a relatively small discrete set rather than being a generic real-valued ultrawide-band signal, is also exploited by the reconstruction process. As a result, the receiver architecture completely bypasses the requirement of high-rate ADC conversion. Instead we use an analog front-end consisting of a bank of correlators with tractable test functions (like square waves), a low-rate A/D converter, and a DSP back-end that utilizes the knowledge of the channel response. The number of correlators can be significantly smaller than the requirement suggested by the Shannon-Nyquist sampling theorem, and nevertheless the performance degrades gracefully with such sub-Nyquist sampling.

The work-horse of the DSP back-end is a computationally efficient quadratic program (QP). The proposed receiver works robustly even in significant ISI, and hence we are not restricted to a low baud-rate. At the same time, the complexity of the receiver is far smaller than a full-fledged MLSE. Moreover, we do not rely on long analog delay lines or any specific modulation format as in TR/DTR. The same architecture can operate with various levels of timing accuracy, ranging from a fraction of T_{pulse} to many multiples of T_{pulse} , and in each case a performance close to the MLSE receiver is attained. Furthermore, as the burst

size becomes moderately large, the receiver implicitly acquires perfect timing ‘on the fly’ and hence the penalty associated with timing uncertainty becomes negligible. Therefore we can send bursts *without* training headers, and yet attain a power efficiency comparable to genie-aided timing. Finally, although the DSP back-end needs to know the channel response, it can blindly acquire and track it based on the same observations that are available for bit demodulation. Unlike [9], who use a matching-pursuit reconstruction and exploit the sparsity of the *received* signal, our channel estimator uses a maximum likelihood stochastic approximation that exploits the much more significant sparsity of the *transmitted* signal.

CS has been used previously by [18] for mitigation of narrow band interference. In an approach analogous to ours, [19][20] and [21] have used CS for direct detection of IR pulses without using a rake or a digital correlator. However note that [19][20] have formulated a generalized likelihood ratio test (GLRT) for the detection of a *single* bit in an ISI-free regime, and they presume accurate timing while doing so. Similarly, although [21] do explicitly address the timing problem, their proposal also assumes an ISI-free regime and involves the exact solution of a set of linear equations that are often ill-conditioned.

Outline of the paper: In Section 2 we describe the system model and the architecture of the proposed receiver. In Section 3 we first formulate and analyze the maximum likelihood (ML) receiver (which is typically intractable), and then propose signal demodulation via a significantly simpler suboptimal QP optimization. Section 4 presents a stochastic recursive algorithm based on ML principles for identifying the channel response. Section 5 presents extensive simulations of the proposed receiver, and in Section 6 we make some concluding remarks. *Convention:* With an abuse of notation, $P(x)$ will denote the density or mass function of a random variable X . $U([a, b])$ will denote a uniform distribution over the interval $[a, b]$ of the real line or of integers, depending on the context. x^T will denote the transpose of a vector or matrix x . When x is a vector, $\|x\|_2$ will denote the L_2 -norm (Euclidean length), $\|x\|_1$ the L_1 -norm (largest absolute value), and $\|x\|_0$ the number of non-zero elements. $H(f)$, $\Phi(f)$ etc will denote Fourier transforms of continuous-time finite-energy signals $h(t)$, $\phi(t)$ etc. $h(t) \star \phi(t)$ denotes a convolution of the signals.

2. System Model and Receiver Architecture

In this section we will describe the overall UWB-IR system under consideration, and then present the architecture of our receiver. The reader is advised to refer to Figure 1.

2.1. Transmitter

The UWB-IR transmitter consists of three main blocks, namely, a timing block that generates a clock signal at a nominal frequency f_{baud} , a payload block that supplies the information bits, and an IR pulse generator. The baud clock provides the timing for the IR pulses within each burst, as well as the

timing for the start of each burst after requisite down-sampling. A total of K pulses are transmitted in each burst after which the transmitter hibernates till the start of the next burst.² At the k -th strobe of the clock within a burst, the IR pulse generator sends on the air a pulse $\phi(t)$, amplitude modulated³ by the bit B^k provided by the payload, drawn equiprobably from $\{+1, -1\}$. The pulse $\phi(t)$ is nominally centered at the frequency f_c with a bandwidth Ω . There is no other RF processing at the transmitter, like heterodyning or filtering, which makes this transmitter very simple, small and inexpensive to build.

For example, consider Figure 2 which displays the Hanning modulated RF pulse of [2] which we used in our simulations, with a center frequency $f_c = 4.0$ GHz and a 6-dB bandwidth $\Omega = 2.0$ GHz. The pulse duration is small, $T_{pulse} = 1.0$ nanosecond. It is well-known that the maximum possible ISI-free baud-rate over an ideal channel of bandwidth $\Omega = 2.0$ GHz is $f_{nyquist} = 2\Omega = 4.0$ GBaud. However, since the temporal dispersion of the UWB channel in indoor environments is often as large as $\tau_{chan} = 100$ nanoseconds, a conventional UWB-IR system needs to choose a much smaller baud-rate, $f_{baud} \leq 1/\tau_{chan} = 10$ MBaud, to avoid ISI. Our receiver, on the other hand, can tolerate significant ISI and therefore we may choose a baud-rate close to the Nyquist frequency, say $f_{baud} = f_{nyquist}/8 = 500$ MBaud. Hence the interval between consecutive pulses is $T_{baud} = 1/f_{baud} = 2.0$ nanoseconds, and a burst of $K = 64$ bits will therefore last for 127 nanoseconds. In contrast, the interval between consecutive bursts may be as large as $T_{burst} = 100$ microseconds. Since a practical inexpensive clock has a significant timing drift of $\rho \sim 40$ parts per million (p.p.m.) caused by random frequency modulation [7, 22], the total drift from the beginning to the end of a burst is limited to $(K - 1)\rho/f_{baud} = 5.1$ pico-seconds, which is negligible considering the fact that a timing error of up to 40 pico-seconds causes an SNR penalty of no more than 1.0 dB for coherent demodulation. Thus if exact timing synchronization is available at the start of a burst, there is no further timing problem. On the other hand, the drift from one burst to the next is very large, ~ 4.0 nanoseconds, resulting in a catastrophic loss in rake receivers unless long training headers are used for re-synchronization per burst.

Without loss of generality we can concentrate on the reception of a single burst, and treat the estimated epoch of arrival of that burst as the temporal origin, $t = 0$. The residual error of the coarse timing block is then perceived as a late arrival of the actual burst by an amount v seconds. (By prefixing a sufficient guard interval in the coarse timing estimate, we can ensure that $v > 0$ with high probability, i.e. the true arrival can only be late but never early.) For simplicity suppose that the true arrival time v is distributed over the interval

²Our receiver architecture continues to be applicable without modification even in the scenario where a repetition code is used, that is, one information bit is repeated N_f times in the payload $\{B_k\}$. The effect of the repetition code is simply to improve the BER vs SNR characteristic by a factor $10 \log_{10}(N_f)$ dB, at the cost of a $\frac{1}{N_f}$ rate reduction. Unless otherwise stated we will assume that no repetition code is present ($N_f = 1$).

³A generalization to pulse position modulation is possible, but will not be discussed here.

$[0, \gamma]$ according to a uniform density. From the point of view of the receiver, the output of the transmitter during the burst is then written as

$$S(t) = \sum_{k=0}^{K-1} B^k \phi(t - kT_{baud} - v). \quad (1)$$

Notice that in writing this equation we ignore the negligible timing drift *within* a burst. Our setup also subsumes the case of a low baud rate $f_{baud} \leq 1/\tau_{chan}$ (used to avoid ISI, as in [7]), if we choose $K = 1$ (one pulse in each burst), treat the pulse-to-pulse drift as the burst-to-burst drift v , and demodulate each pulse independently.

2.2. Channel

The UWB channel is known to be linearly dispersive with tens or hundred of resolved multi-path components, depending on the radio environment. In [6], a set of standardized random models has been postulated covering several scenarios like indoor line-of-sight (LOS) in residential environments (CM1), indoor non-line-of-sight (NLOS) in residential environments (CM2), indoor LOS in office environments (CM3), indoor NLOS in office environments (CM4) etc. We will use realizations from these standardized models in our simulations.

The demodulation algorithm to be presented in Section 3 assumes that the total system response is known. Of course, apart from the random TOA v , the shape of the channel impulse response (the set of multi-path amplitudes and relative delays) itself can vary with time because of the large-scale motion of the transmitter/receiver as well as random changes in the radio environment (log-normal shadowing). However, these variations are relatively slow (i.e. Doppler spread is small) and we can assume [3, 23] that in the duration of one burst the shape of the channel impulse response is a time-invariant function $h_c(t)$. In fact the channel coherence time is typically of the order of tens or hundreds of milliseconds [24, 25], which allows us to use an incremental estimator to *acquire and track* the shape of the channel response (cf. Section 4). The fast variations in the TOA will *not* be treated as channel variations, but instead be inferred explicitly from burst-to-burst and provided explicitly to the incremental channel estimator, whose relatively slow dynamics will integrate out the occasional error.

2.3. Receiver

The receiver consists of an analog front-end and a DSP back-end. The defining characteristic of our receiver is that we relieve the analog front-end of difficult tasks like fast A/D conversion and accurate delay lines, and instead compensate by using an elaborate DSP back-end. We keep the DSP back-end tractable by avoiding a full-fledged ML demodulator, and instead use a QP reconstruction. QP is considered an ‘easy’ problem in optimization theory, that can be solved in low-order polynomial time [26] by state-of-the-art interior point (IP) methods. At the same time, we will demonstrate that it gives negligible degradation relative to the ML decoder.

2.3.1. Analog Front-end

Let the received signal at the antenna be denoted by $U(t)$. The first block in the analog front-end is a noise-limiting bandpass-pass filter $g(\cdot)$ centered at f_c , having a bandwidth $\approx \Omega$. The output of this filter is

$$R(t) = \sum_{k=0}^{K-1} B^k h(t - kT_{baud} - v) + W(t), \quad (2)$$

where $h(t)$ denotes the total impulse response, which is the convolution of the transmit pulse $\phi(t)$, the channel $h_c(t)$, and the filter response $g(t)$, and

$$W(t) = \int V(t - \tau)g(\tau)d\tau \quad (3)$$

is band-limited zero-mean additive Gaussian noise, modeled as the response of the filter to a white Gaussian thermal noise process $V(t)$ of power spectral density N_0 .

The signal $R(t)$ is fed to a bank of M parallel analog correlators, followed by M integrators. This module replaces other conventional structures like a rake receiver, a fast A/D converter for subsequent MLSE or digital correlation, an ED receiver or a TR/DTR receiver. The test function used in correlator number m is denoted as $\psi_m(t)$, and the whole ensemble of test functions is denoted by $\{\psi_m(t)\}$. In Section 3.4, we will discuss the criteria for selecting the ensemble. At this point, it suffices to note that we do not need to tune the timing of these test functions (i.e. no analog delay lines), and hence they are relatively easy to implement. All we require is that the ensemble be known to the DSP back-end.

The integrators $m = 0, 1, \dots, M - 1$ are reset to zero at the epoch $t = 0$ and their output is sampled synchronously at the epoch $\lambda_h + \gamma + (K - 1)T_{baud}$ when all of the energy of the burst is known to have arrived with high probability (recall that γ is the uncertainty in the TOA of the burst). Thus we have the M measurements

$$Y_m = \int_0^{\lambda_h + \gamma + (K-1)T_{baud}} R(t)\psi_m(t) dt, \quad m = 0, 1, \dots, M - 1. \quad (4)$$

The vector of measurements $Y = [Y_1, Y_2, \dots, Y_M]^T$ is then fed to the DSP back-end, which recovers the payload bits B^k , $k = 0, 1 \dots, K - 1$ via a tractable QP algorithm.

Extension to the Case of Repetition Coding: Suppose an $N_f > 1$ repetition code is being used, hence payload bits B^j , $j = iN_f, \dots, (i + 1)N_f - 1$ are all copies of the i -th information bit C^i , and the total burst of K bits corresponds to K/N_f information bits. In this case we simply rewrite the received filtered signal as

$$R(t) = \sum_{i=0}^{K/N_f-1} C^i h^{comp}(t - iT_{baud}^{comp} - v) + W(t), \quad (5)$$

where we define $h^{comp}(t)$ to be a *composite* impulse response

$$h^{comp}(t) = \sum_{j=0}^{N_f-1} h(t - jT_{baud}). \quad (6)$$

Since equation (5) has the same mathematical form as equation (2), we can clearly use exactly the same DSP back-end to directly recover the information bits C^i , $i = 1, 2, \dots, K/N_f$ from the measurement vector Y , by appropriately replacing $h(\cdot)$ with $h^{comp}(\cdot)$ in the reconstruction algorithm. An alternative method would be to continue to demodulate based on the representation in equation (2) and then do an algebraic decoding (hard decoding) of the repetition code via majority rule. Note that such hard decoding costs $\sim 1.0 - 2.0$ dB in SNR relative to optimal joint decoding-demodulation [27].

Signal to Noise Ratio. Let $h_U(t) \doteq \phi(t) \star h_c(t)$, and define

$$h_U^l(t) \doteq \sum_{k=0}^{K-1} b_l^k h_U(t - kT_{baud}), \quad (7)$$

$$\xi(f) \doteq \frac{1}{2^K} \sum_{l=0}^{2^K-1} |H_U^l(f)|^2, \quad (8)$$

where $b_l^k \in \{+1, -1\}$ is the k -th bit of the number $l \in \{0, 1, \dots, 2^K - 1\}$. An optimal (but intractable) receiver would replace the front-end filter $g(t)$ with a bank of 2^K matched filters (MFs), one each for the candidate matched signal $h_U^l(-t)$, $l = 0, 1, \dots, 2^K - 1$. Assuming that the *timing is perfectly known*, it would then declare as the estimate of the payload, the index l of the filter which has the maximum output at the sampling time. Such a hypothetical genie-timed MF receiver serves as a reference with which we can compare our suboptimal receiver. The average SNR *per bit* in the MF receiver is therefore given by

$$\text{SNR}_{bit} \doteq \frac{\int \xi(f) df}{K \frac{N_0}{2}}, \quad (9)$$

where $N_0/2$ is the two-sided power spectral density of the zero-mean additive white Gaussian (AWG) thermal noise $V(t)$. It is not difficult to show that since the K bits in the pay-load are i.i.d. Bernoulli($\frac{1}{2}$), (hence all the candidate signals $h_U^l(t)$ are a-priori equiprobable), we have the relation

$$\xi(f) = K \|H_U(f)\|^2. \quad (10)$$

Hence the SNR per bit in the MF receiver is given simply by

$$\text{SNR}_{bit} \doteq \frac{\int \|H_U(f)\|^2 df}{\frac{N_0}{2}}. \quad (11)$$

For consistency with literature, we will use this definition of SNR in all our analysis and simulations.

2.3.2. DSP Back-end

The demodulation of the payload by the DSP back-end relies on a consistent discrete time representation of the signal. Let f_s be a sufficiently large *virtual sampling frequency* [19] for the received UWB-IR signal. We would like to emphasize that this is only a ‘thought-experiment’ construction, and no A/D conversion is done at rate f_s in actuality. Choosing an f_s as large as possible reduces aliasing and timing quantization errors. On the other hand, it also increases the size of the optimization problem, hence a suitable tradeoff must be made. For example, for the IR pulse described in Section 2.1, the choice of $f_s = 2(f_c + \frac{\Omega}{2}) = 10$ GHz practically eliminates aliasing and limits the timing quantization penalty to 1.5 dB. Let $h[n]$ denote the sampled version of the total impulse response $h(t)$ at the rate f_s samples per second, and let h denote a vector representation of $h[n]$. That is, letting $T_s \doteq \frac{1}{f_s}$,

$$h[n] \doteq h(nT_s), \quad n = 0, 1, \dots, \Lambda_h - 1, \quad (12)$$

$$h \doteq [h[0], h[1], \dots, h[\Lambda_h - 1]]^T, \quad (13)$$

where $\Lambda_h = \lceil \lambda_h f_s \rceil$ is the length of the discrete-time finite impulse response $h[n]$. A similar convention will apply to other signals like $g(t)$, $\psi_m(t)$, $W(t)$ etc.

Let γ and T_{baud} be multiples of T_s , which can be achieved by construction. Now, expressed in rate f_s samples, the arrival time uncertainty is $\Gamma \doteq \gamma f_s$ and the baud period (the interval between consecutive pulses) is $N_{baud} = T_{baud} f_s$. Define $\Lambda_X \doteq \Gamma + (K - 1)N_{baud}$. Then the length of the total burst response including the timing uncertainty is

$$N \doteq \Lambda_h + \Lambda_X - 1. \quad (14)$$

Let $\Upsilon = \text{round}(vf_s)$ be the burst arrival time v quantized to a step size of T_s . As remarked earlier, this quantization introduces an extra measurement error which is negligible provided f_s is chosen large enough. Now, the sampled version of $R(t)$ can be written as a vector $R \in \mathbb{R}^N$ given by

$$R = \mathcal{H}X + W. \quad (15)$$

Here the vector $X \in \mathbb{R}^{\Lambda_X}$ is a *virtual* discrete time information signal which has all samples equal to zero except for K non-zero samples. The k -th non-zero sample, for $k = 0, 1, \dots, K-1$, has a random amplitude B^k drawn independently and equiprobably from $\{-1, +1\}$, and has a random location $\Lambda^k = \Upsilon + kN_{baud}$. On account of the modeling assumption made in Section 2.1, it follows that $\Upsilon \sim U([0, \Gamma])$. The vector $W \in \mathbb{R}^N$ is the sampled version of the additive Gaussian noise $W(t)$, and the matrix $\mathcal{H} \in \mathbb{R}^{N \times \Lambda_X}$ is the convolutional matrix

(Toeplitz form) of $h[n]$,

$$\mathcal{H} = \begin{pmatrix} h[0] & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ h[1] & h[0] & 0 & \dots & 0 & 0 & 0 & 0 \\ h[\Lambda_h - 1] & h[1] & h[0] & \dots & 0 & 0 & 0 & 0 \\ h[\Lambda_h - 1] & h[\Lambda_h - 2] & h[1] & h[0] & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & h[\Lambda_h - 1] & h[\Lambda_h - 2] & h[1] & h[0] \\ 0 & 0 & 0 & \dots & 0 & h[\Lambda_h - 1] & h[\Lambda_h - 2] & h[1] \\ 0 & 0 & 0 & \dots & 0 & 0 & h[\Lambda_h - 1] & h[\Lambda_h - 2] \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & h[\Lambda_h - 1] \end{pmatrix}. \quad (16)$$

The rows of the matrix are formed by right shifts of the time-flipped response $h[\Lambda_h - 1], h[\Lambda_h - 2], \dots, h[1], h[0]$.

In a similar vein we can further relate the actually sampled measurements Y at the output of the integrators to the virtual information signal X . Define the $M \times N$ *measurement matrix* Ψ to be

$$\Psi \doteq \frac{1}{f_s} [\psi_0, \psi_1, \dots, \psi_{M-1}]^T, \quad (17)$$

where, for all $i = 0, 1, \dots, M - 1$,

$$\psi_i \doteq [\psi_i[0], \psi_i[1], \dots, \psi_i[N - 1]]^T. \quad (18)$$

The sampling lemma [27] tells us that for any signals $x(t), y(t)$ band-limited to $\frac{f_s}{2}$, sampling at rate f_s leaves the inner product invariant up to a scaling factor. That is, $\int x(\tau)y(\tau)d\tau = \frac{1}{f_s} \sum_n x[n]y[n]$. Hence we can write the *measurement equation*

$$Y = \Psi R = \Psi \mathcal{H} X + \Psi W. \quad (19)$$

Let $B \doteq [B^0, B^1, \dots, B^{K-1}]^T$. Then the aim of the DSP back-end is to optimally estimate B, Υ from the measurement Y , based on the relation in equation (19) and the a-priori statistical knowledge about B, Υ . Note that B contains the payload which is of primary interest, while the quantity Υ is a ‘nuisance’ parameter.⁴ As we shall see in the next section, for optimal performance we need to maximize the observation likelihood jointly over the informative parameter as well as the nuisance parameter.

3. Bit Demodulation Based on Incomplete Measurements

The maximum likelihood (ML) demodulation of B , based on the measurement Y given by equation (19), will be described in Section 3.1. It involves the maximization of the likelihood $P(Y|B, \Upsilon)$ over all the valid values of payload B

⁴In PPM, Υ carries the payload, while B is deterministic. In any case, Υ will always be informative in the context of localization.

and the nuisance timing parameter Υ . Since this can be complex to implement under a large timing uncertainty Γ and even moderately large burst length K , in Section 3.2 we propose an alternative computationally efficient reconstruction via a QP. We will see in simulation results discussed in Section 5, that the QP reconstruction gives only a small loss compared to ML demodulation.

3.1. ML Demodulation and BER Analysis

Let us define the set \mathcal{X} as the set of all signals $x \in \mathbb{R}^{\Lambda_x}$ that satisfy the following properties: (i) $\|x\|_0 = K$ (sparsity). (ii) The first nonzero sample is located at $\ell^0 \in [0, \Gamma]$. The subsequent non-zero samples are located at positions $\ell^k = \ell^0 + kN_{baud}$, $\forall k = 1, 2, \dots, K - 1$ (timing). (iii) The amplitudes of all the nonzero samples are from $\{-1, +1\}$ (signaling alphabet).

Clearly, \mathcal{X} is the finite equiprobable alphabet of the random information signal X (cf. Section 2.3.2), of cardinality $|\mathcal{X}| = 2^K(\Gamma + 1)$, and there is a one-to-one mapping

$$\{-1, +1\}^K \times \{0, 1, \dots, \Gamma\} \rightarrow \mathcal{X} \quad (20)$$

$$(B, \Upsilon) \mapsto X(B, \Upsilon). \quad (21)$$

Hence we can write $P(Y|B, \Upsilon) = P(Y|X)$, which implies that, without losing optimality, we may first make the ML estimate \hat{X} of the information signal X , and then map it to the optimal payload estimate $\hat{B}(\hat{X})$ and TOA estimate $\hat{\Upsilon}(\hat{X})$.

It is easy to see that the noise term ΨW in the measurement equation (19) is a zero mean multivariate Gaussian random variable with a covariance matrix $\sigma^2 \Psi \mathcal{G} \mathcal{G}^T \Psi^T$, where \mathcal{G} is the Toeplitz form of the front-end filter $g[n]$, analogous to the definition in equation (16), and $\sigma^2 = N_0/(2f_s)$. Hence, the likelihood of a candidate signal $x \in \mathcal{X}$ conditioned on the observation Y is given, up to a normalization factor, by

$$P(Y|x) \propto \exp \left\{ \frac{-1}{2\sigma^2} (Y - \Psi \mathcal{H}x)^T (\Psi \mathcal{G} \mathcal{G}^T \Psi^T)^{-1} (Y - \Psi \mathcal{H}x) \right\}. \quad (22)$$

Therefore, the ML demodulator declares the estimated signal as

$$\hat{X} = \underset{x \in \mathcal{X}}{\operatorname{argmax}} P(Y|x) = \underset{x \in \mathcal{X}}{\operatorname{argmin}} (Y - \Psi \mathcal{H}x)^T (\Psi \mathcal{G} \mathcal{G}^T \Psi^T)^{-1} (Y - \Psi \mathcal{H}x). \quad (23)$$

Since B and Υ are drawn equiprobably from their alphabets, they do not have informative priors, and the ML estimate is also the Bayesian estimate, which is optimal in terms of the error rate. Suppose that $x^0 \in \mathcal{X}$ was the true information signal, hence

$$Y = \Psi \mathcal{H}x^0 + \Psi W. \quad (24)$$

Let $x^1 \neq x^0$, $x^1 \in \mathcal{X}$ be some other information signal. Then, under ML demodulation, the pair-wise error probability (PEP) is given by

$$\Pr(x^0 \rightarrow x^1) \doteq \Pr\{P(x^1|Y) > P(x^0|Y)\}. \quad (25)$$

With some straightforward manipulation it can be shown that

$$P(x^0 \rightarrow x^1) = \mathbf{Q} \left(\frac{\sqrt{(x^0 - x^1)^T \mathcal{H}^T \Psi^T (\Psi \mathcal{G} \mathcal{G}^T \Psi^T)^{-1} \Psi \mathcal{H} (x^0 - x^1)}}{2\sigma} \right), \quad (26)$$

where $\mathbf{Q}(a) = \int_a^\infty \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx$ is the area under the tail of a standard normal distribution. Since we have $|\mathcal{X}| = 2^K(\Gamma + 1)$ equiprobable candidates for the selecting the transmitted signal x^0 , and the pair-wise error event $x^0 \rightarrow x^1$ leads to $\|\hat{B}(x^0) - \hat{B}(x^1)\|_0$ bit errors, we can write the following union bound on the BER,

$$P_e \leq \sum_{x^1, x^0 \in \mathcal{X}} \frac{\|\hat{B}(x^0) - \hat{B}(x^1)\|_0}{K 2^K (\Gamma + 1)} \mathbf{Q} \left(\frac{\sqrt{(x^0 - x^1)^T \mathcal{H}^T \Psi^T (\Psi \mathcal{G} \mathcal{G}^T \Psi^T)^{-1} \Psi \mathcal{H} (x^0 - x^1)}}{2\sigma} \right). \quad (27)$$

Note that $\mathcal{H} = T_s \mathcal{G} \mathcal{H}_U$, where \mathcal{H}_U is the convolutional matrix of the response $h_U[n]$. As a sanity check, notice that if

1. there is no under-sampling (i.e. $M = N$),
2. Ψ is invertible (i.e. the ensemble $\{\Psi_m\}_{m=0}^{M-1}$ are linearly independent)
3. \mathcal{G} is invertible (i.e. the translations of the pulse $g(t)$ in steps of T_s are linearly independent)
4. the timing is ideal (i.e. $\Gamma = 0$),
5. there is only one bit per burst (i.e. $K = 1$),

expression (27) reduces to the familiar expression for a perfectly timed MF,

$$P_e = \mathbf{Q} \left(\sqrt{\frac{\int |h_U(t)|^2 dt}{\frac{N_0}{2}}} \right) = \mathbf{Q} \left(\sqrt{\text{SNR}_{bit}} \right), \quad (28)$$

where we used the definition of the SNR per bit from equation (11).

3.2. Suboptimal Computationally Efficient Demodulation Via QP

3.2.1. Motivation For a Sub-optimal Tractable Demodulator

The ML demodulation problem in equation (23) clearly becomes cumbersome when the timing uncertainty Γ or the burst length K is large. Even with a dynamic program like the Viterbi algorithm [27] for MLSE, the complexity is exponential in K or the channel memory, whichever is smaller. With our exemplary choice of $f_{baud} = 500$ MBaud, the channel memory will extend to at least 50 pulses and so the complexity will scale as 2^K for K up to 50. In light of this difficulty, now we will propose an alternative suboptimal demodulation technique whose complexity is $O(K^3)$. The technique is inspired by the philosophy of CS for sparse signal reconstruction under incomplete measurements [17, 16].

3.2.2. QP Demodulation

Let the vector $\xi(a, \ell_1, \ell_2)$ be a positive penalty vector for the candidate information signals $x \in \mathcal{X}$. It incorporates the available timing information by giving more penalty to those locations of x where the occurrence of the non-zero samples is unlikely. That is, for all $n = 0, 1, \dots, \Lambda_X - 1$,

$$\xi(a, \ell_1, \ell_2)[n] \doteq \begin{cases} 1.0, & n = \ell + kN_{baud}, \ell \in [a + \ell_1, a + \ell_2], k = 0, 1, \dots, K - 1 \\ \mathcal{U}, & \text{otherwise,} \end{cases} \quad (29)$$

where \mathcal{U} is some suitable large number like 10^3 . Also define a corresponding diagonal penalty matrix as $\Xi(a, \ell_1, \ell_2) = \text{diag}(\xi(a, \ell_1, \ell_2))$.

Now consider the following relaxation of the ML demodulation problem (23):

$$\tilde{X} = \underset{x \in \mathbb{R}^N: \|\Xi(a, \ell_1, \ell_2)x\|_1 = K}{\text{argmin}} (Y - \Psi\mathcal{H}x)^T (\Psi\mathcal{G}\mathcal{G}^T\Psi^T)^{-1} (Y - \Psi\mathcal{H}x). \quad (30)$$

Notice that the new constraint set $\{x \in \mathbb{R}^N : \|\Xi(a, \ell_1, \ell_2)x\|_1 = K\}$ is not a discrete set, but rather a continuous set of signals of adequately small L_1 norm. Therefore notice that $\mathcal{X} \subset \{x \in \mathbb{R}^N : \|\Xi(0, 0, \Gamma)x\|_1 = K\}$. We can further re-write the problem (30) in a more amenable form [16] by defining

$$x^+ \doteq \max(x, 0) \quad (31)$$

$$x^- \doteq \max(-x, 0) \quad (32)$$

$$z \doteq [x^{+T}, x^{-T}]^T. \quad (33)$$

Then we have the identities $x = x^+ - x^-$ and $\|x\|_1 = x^+ + x^-$. We can now rewrite the problem (30) as

$$\begin{aligned} \tilde{X}_n &= \tilde{Z}_n - \tilde{Z}_{n+N}, \quad n = 0, 1, 2, \dots, N, \\ \tilde{Z} &= \underset{z \geq 0, [\xi(a, \ell_1, \ell_2)^T, \xi(a, \ell_1, \ell_2)^T]z = K}{\min} f^T z + \frac{1}{2} z^T Q z, \end{aligned} \quad (34)$$

where

$$Q = \begin{pmatrix} \mathcal{H}^T \Psi^T (\Psi\mathcal{G}\mathcal{G}^T\Psi^T)^{-1} \Psi\mathcal{H} & -\mathcal{H}^T \Psi^T (\Psi\mathcal{G}\mathcal{G}^T\Psi^T)^{-1} \Psi\mathcal{H} \\ -\mathcal{H}^T \Psi^T (\Psi\mathcal{G}\mathcal{G}^T\Psi^T)^{-1} \Psi\mathcal{H} & \mathcal{H}^T \Psi^T (\Psi\mathcal{G}\mathcal{G}^T\Psi^T)^{-1} \Psi\mathcal{H} \end{pmatrix}, \quad (35)$$

$$f = [-Y^T (\Psi\mathcal{G}\mathcal{G}^T\Psi^T)^{-1} \Psi\mathcal{H}, Y^T (\Psi\mathcal{G}\mathcal{G}^T\Psi^T)^{-1} \Psi\mathcal{H}]. \quad (36)$$

(34) is now a standard QP, which has several efficient large-scale techniques of solution like active set, conjugate gradient and interior point methods, of which the last is generally regarded as the fastest [26].

We perform the demodulation in two stages. In the first stage we solve the QP in (34) using $\xi(a = 0, \ell_1 = 0, \ell_2 = \Gamma)$, corresponding to the full TOA uncertainty Γ . The result of this stage, $\tilde{X}^{(1)}$, is then used to extract an estimate $\hat{\Upsilon}$ of the arrival time via correlation with the template $\xi(0, 0, 0)[n]$ as follows:

$$\hat{\Upsilon} = \underset{n' \in \{0, 1, \dots, \Gamma\}}{\text{argmax}} \sum_n |\tilde{X}^{(1)}[n - n']| \xi(0, 0, 0)[n]. \quad (37)$$

We then solve the QP in (34) again, using $\xi(a = \hat{\Upsilon}, \ell_1 = 0, \ell_2 = 0)$, which corresponds to the assumption that the $\hat{\Upsilon}$ is exactly correct and there is no residual TOA uncertainty. The result of this stage, $\tilde{X}^{(2)}$, is not necessarily in the set \mathcal{X} . Hence, we cannot consistently map it back into an estimate \hat{B} for the payload. To overcome this difficulty, we must implement a further simple *decision rule*: Once $\tilde{X}^{(2)}$ has been delivered, demodulate the payload as

$$\hat{B}^k = \text{sign}(\tilde{X}^{(2)}[\hat{\Upsilon} + k N_{\text{baud}}]), \quad k = 0, 1, \dots, K - 1. \quad (38)$$

In summary, in lieu of the ML demodulation problem, which involves maximization over a large discrete set \mathcal{X} , we have formulated a relaxed continuous QP which jointly solves for the best sparsity and timing *without explicitly checking each timing epoch and bit pattern individually*. Since the optimization problem size is $\Lambda_X = \Gamma + (K - 1)N_{\text{baud}}$, and interior point methods can solve a QP with polynomial complexity of degree-3 [26], the demodulation complexity is now only $O(K^3)$.

3.3. On The Relation Between Relaxed QP Demodulation and L_1 -minimization

While we have proposed the QP reconstruction as an inexpensive suboptimal substitute for ML demodulation, it is also worthwhile to briefly discuss its relationship with the classical CS reconstruction method based on L_1 -norm minimization.

Recall that the information signal X satisfies the property $\|X\|_0 \leq K$. Actually the constraint $\|X\|_0 \leq K$ by itself allows up to K nonzero samples to be placed at *arbitrary* locations within the signal and they can have *arbitrary* amplitudes, while in reality our information signal has considerably more structure. But let us ignore the extra structure for the time being. Let $\Psi_r \doteq \Psi\mathcal{H}$, and recall that since typically $M \ll N$, the system of equalities $Y = \Psi_r X$ is highly *under-determined*, and the classical least squares approach fails badly. In the CS literature [16, 28] the problem of sparse signal reconstruction from incomplete measurements is instead formulated as a *basis pursuit*:

$$\hat{X} = \begin{array}{l} \min \|x\|_1 \\ \text{s.t. } Y = \Psi_r x \end{array}, \quad (39)$$

which is a relaxation of the intractable L_0 -minimization problem. The central tenet of CS theory is that, if $M \geq \xi \log(\Gamma_X)K$, then perfect reconstruction of X with high probability is assured via (39), provided an appropriate ‘decoherent’ measurement ensemble is used. The factor ξ is a constant that depends on the choice of the ensemble, and is called the *over-sampling* factor.

The practical advantage of formulation (39) is that it can be re-cast as a linear program (LP) and hence can be solved very efficiently by interior point methods. Unfortunately, (39) is known to be very fragile to perturbations of the measurements, and we have verified that it performs poorly in even moderate amounts of noise. In light of this problem, it has been proposed that a regularized optimization in the form of a LASSO [29], a Dantzig selector [30],

or a penalty function [31] would be a better choice. For example, the LASSO optimization is written as

$$\hat{X} = \begin{array}{l} \min \|x\|_1 \\ \text{s.t. } \|Y - \Psi_r x\|_2^2 \leq \epsilon. \end{array} \quad (40)$$

Unfortunately, these regularized optimizations are considerably more complex than the linear program in (39), and can also suffer from problems of local optima and non-convergence.

However, as remarked earlier, classic CS reconstruction as well as regularized approaches like the LASSO exploit only generic sparsity $\|X\|_0 \leq K$. In contrast, in our application we know that the signal X has *exactly* K non-zero samples and they are spaced *exactly* N_{baud} samples apart. Hence the knowledge of the timing of the first sample fixes the locations of all the other samples. *In this sense the sparsity of X is not K but just 1.* Moreover, we have the following pieces of side-information: (i) the non-zero samples are always from a known fixed alphabet, and (ii) the measurement noise is not white and its covariance matrix is known. All these extra pieces of information mean that we can improve upon a generic LASSO type reconstruction. Specifically, we can switch the cost and the constraints of the LASSO problem of (40) and recast it as a QP, which was precisely what was done in Section 3.2.2. This establishes the connection of the QP reconstruction to classical CS reconstruction.

The QP receiver gives a better performance than techniques like LASSO, owing to the side-information. Additionally, QP has the important advantage of being computationally much cheaper and more stable than the LASSO. Lastly, as we shall demonstrate in Section 5.2, the performance of the QP receiver is essentially *invariant* w.r.t. the number of bits per packet K , provided the number of front-end correlators, M , scales *linearly* with K .

3.4. Choice of Measurement Ensemble

The choice of the measurement ensemble needs to be made in such a way that M can be kept as small as possible while achieving an acceptable performance. Moreover, the ensemble should be easy to generate practically and the demodulation should be insensitive to imperfections in signal generation.

3.4.1. Canonical Nyquist and other Orthonormal Ensembles

At this point it is worthwhile noting that if we set $M = N$, idealize the front-end filter be a low-pass Nyquist filter of bandwidth $f_s/2$ so that $g(t) = \frac{\sin \pi t f_s}{\pi t f_s}$, and let the test functions be the canonical functions $\psi_m(t) = \delta(t - mT_s)$, $m = 0, 1, \dots, M - 1$, the correlator bank simply provides uniform time domain Nyquist samples of the incoming signal i.e. an A/D conversion. The resulting samples can then be used for a digital MLSE or MF as the case may be. Due to the Nyquist criterion, we know that the this time domain signal vector of length N will already be oversampled by a factor $f_s/(2\Omega)$ and hence the samples will not be uncorrelated over the ensemble of all bursts. For *critical sampling*, where all signal energy is captured, we need only $2\Omega N/f_s$ uncorrelated samples,

and moreover they need not necessarily be made in the time domain but could be made in any other domain reached by a linear orthonormal transformation. Digital MLSE/MF demodulation can then, in principle, be implemented in any such domain because of the invariance of the inner product.

Of course, if we have $M < 2\Omega N/f_s$ samples, i.e. under-sampling, then choice of ensemble does become important. This is because if the energy of the signal happens to fall in the null-space of the ensemble with high probability, there is no hope of reconstructing the signal. The classical approach to reduced rate sampling and compression is the so called *transform method*, where we *a priori identify* the subspace in which the signal energy is concentrated and then take projections only on basis elements that span that subspace. This, however precludes a *universal receiver*. In particular, in our application this sparsity subspace is spanned by the signals in \mathcal{X} and therefore depends on *the channel, the payload size and the timing uncertainty*.

3.4.2. Uniformly Decoherent Ensembles

This leads us to the central question: Is it possible to devise *universal ensembles* that allow reliable reconstruction of any under-sampled *sparse* signal, provided the under-sampling is not too severe relative to the sparsity? Moreover, can they allow a graceful SNR penalty in the presence of receiver noise? The surprising answer to the first question is known to be in the affirmative, as was shown in the ground breaking work of [17, 16]. In this paper, we show through the ML demodulator analysis of Section 3.1 and extensive simulations in Section 5, that the answer to the second question also seems to be affirmative. These ‘universal’ ensembles are known to be sets of randomly generated noise-like signals. One example is that of binary pseudo-noise (PN) signals that transit independently and equiprobably between levels $\{\frac{+1}{\sqrt{N}}, \frac{-1}{\sqrt{N}}\}$ at intervals of T_s seconds.

The reason why such noise-like ensembles perform well is that [17, 16] (i) they are *uniformly decoherent* w.r.t any family of sparse signals, not just those that are temporally so (it is notable that the philosophy of choosing measurement signals having such a decoherence property is the exact antithesis of the philosophy of transform coding), and (ii) any M such signals have a full rank M with high probability, for every $M \leq N$. In other words, $\Psi\Psi^T$ is invertible with high probability. In fact, though they are not necessarily exactly orthonormal, they are asymptotically so, i.e. $\Psi\Psi^T \rightarrow I_M$ as $N \uparrow \infty$.

In order to reject the out-of-band noise we must use a non-trivial front-end band-pass filter $g(t)$, and hence the *apparent* measurement ensemble matrix becomes $\Psi\mathcal{G}$. If we idealize $g(t)$ to be an ideal band-pass Nyquist filter of bandwidth Ω , we are assured that $\mathcal{G}\mathcal{G}^T = I$. Furthermore, if we pretend that $\Psi\Psi^T = I$ also holds, we have $\Psi\mathcal{G}\mathcal{G}^T\Psi^T = I$. The $M \leq \frac{2N\Omega}{f_s}$ measurements made by the CS front end are roughly uncorrelated due to the decoherence property, and hence the under sampling factor $\frac{Mf_s}{2N\Omega}$ is also the fraction of the signal energy they will capture (Parseval’s theorem). This will be true no matter which M measurement signals we choose from the underlying ensemble. This implies that reliable demodulation of the UWB-IR signal is possible only after

paying an under-sampling penalty of at least $10 \log_{10} \frac{2\Omega N}{M f_s}$ dB in SNR (which we will call the ‘energy-loss penalty’), and this penalty will (on an average) decrease monotonically and vanish as $M \uparrow N \frac{2\Omega}{f_s}$. What is significant is that since a CS receiver exploits the sparsity of the signal, we *do not pay* any extra penalty on top of this unavoidable energy-loss penalty. Another point worth noting is that in a hypothetical noiseless case (with $\text{SNR}_{bit} = \infty$), error-free demodulation is possible with M as small as 4 – 8 (depending on the pulse and channel response) which agrees with classical CS results [17, 16]. In contrast, when doing *direct temporal under-sampling* with any under-sampling factor $\frac{M f_s}{2N\Omega} < 1$, there typically is a catastrophic loss in performance, far more than $10 \log_{10} \frac{2\Omega N}{M f_s}$ dB, and there is no graceful degradation. Similarly, error-free demodulation is impossible even in the noiseless case.

Although in practice $\Psi \mathcal{G} \mathcal{G}^T \Psi^T$ is not exactly an identity due to a non-ideal filter $g(t)$ and a finite N , we will nevertheless see in extensive simulation results in Section 5 that the above described robustness to under-sampling does hold in all practical settings *irrespective of the timing uncertainty, the size of payload and the amount of ISI*.

3.4.3. Fourier and Square Wave Ensembles

Actually we do not need a strictly universal measurement ensemble since we know that our signal sparsity is always in the temporal domain. It is known [32, 28] that the *Fourier* ensemble, M sinusoids of random frequencies drawn uniformly from the band $[f_c - \frac{\Omega}{2}, f_c + \frac{\Omega}{2}]$, is *maximally decoherent* with respect to such signals (the renowned Heisenberg uncertainty principle), and would be the optimal ensemble for our signals in a noiseless setting. However, since we also need to deal with noise, the optimality in terms of BER performance is not guaranteed. Our simulations indicate that a Fourier ensemble with proper windowing and frequencies selected *deterministically and uniformly* from the signal band $[f_c - \frac{\Omega}{2}, f_c + \frac{\Omega}{2}]$ performs only slightly worse than the PN-ensemble, presumably because of the point-like support of the test functions in the frequency domain. Note that the Fourier ensemble may still be desirable from the point of view of robustness to narrow-band interference, an issue which we have discussed elsewhere [33]. Finally, the ensemble of *square waves* of amplitude $1/\sqrt{N}$ and frequencies selected deterministically and uniformly from the signal band is also seen to perform as well as the PN ensemble. From a practical perspective the square wave and Fourier ensembles are perhaps more attractive than the PN ensemble because we do not need any pseudo-random generators.

3.4.4. Robustness to Non-Ideal Test Functions

Another important robustness property inherent to compressed sensing is that the generated test functions do not need to have an ideal waveform. For example the PN ensemble or the square wave ensemble need not have rectangular level transitions. Imperfections like ringing and non-ideal rise time are well-tolerated, provided we know these effects in advance so that we can compensate for them by choosing an appropriately modified Ψ in the reconstruction algorithm. Evidence for this property will be presented in Section 5.4.

4. Channel Identification

In the discussion so far we have assumed that the total system impulse response $h(\cdot)$ is available to the receiver. We will now describe a technique to estimate the channel response via a stochastic recursive approximation [34, 35] of the Expectation Maximization (EM) algorithm [36]. It is noteworthy that our estimator uses only the observations Y and the demodulated virtual information signal $\hat{X}(\hat{B}, \hat{Y})$, and hence does not need any extra sensing hardware. Moreover, due to its simplicity, it can be easily accommodated in the DSP back-end without any significant increase in complexity.

Let $\hat{h} \in \mathbb{R}^{\Lambda_h}$ denote the current estimate of the total channel impulse response. Let $\hat{\mathcal{H}}$ be its Toeplitz matrix representation as in equation (16). We will rewrite $P(Y|x)$ from equation (22) as $P(Y|x, h)$, to make explicit its dependence of the total channel response h . Let $\epsilon[n]$ be a suitably chosen time dependent step size, $\hat{X} = \hat{X}(\hat{B}, \hat{Y})$ the information signal estimated by the QP algorithm by demodulating the burst, and $e = [e_0, e_1, \dots, e_{M-1}]^T$ the estimated measurement error given by

$$e \doteq Y - \Psi \hat{\mathcal{H}} \hat{X}. \quad (41)$$

Let $\Psi_{m, b:b+\Lambda_h-1}$ denote the elements on the m -th row of Ψ , from column b through $b+\Lambda_h-1$. With these conventions in place, we implement the following update upon the arrival of each burst:

$$\begin{aligned} \hat{h} &\longleftarrow \hat{h} + \epsilon[n] \left. \frac{\partial \log P(Y|\hat{X}, h)}{\partial h} \right|_{h=\hat{h}} \\ &= \hat{h} + \epsilon[n] \frac{1}{\sigma_W^2} \sum_{m=0}^{M-1} e_m \sum_{b=0}^{\Gamma_X-1} \hat{X}_b \Psi_{m, b:b+\Lambda_h-1}^T. \end{aligned} \quad (42)$$

The starting point of this algorithm can be simply chosen to be an all-zero response. Notice that we are updating the response based on bits \hat{X} , which are demodulated under the assumption that the current estimate \hat{h} is the correct one. Therefore we have a totally blind algorithm. We will demonstrate with simulations in Section 5.5 that, in spite of this blindness, it acquires and tracks the total channel response very robustly. Of course we can also accommodate the case of training bits by simply replacing \hat{X} by the true bits X in the recursion (42), which typically improves the convergence speed and steady state characteristic of the estimator. However our simulations suggest that such training bits are not necessary in typical practical scenarios. Note that while the proposed estimator has similarities to other algorithms like decision feedback equalizers [27], its innovation is based not directly on the fully sampled received signal R but on under-sampled linear functionals Y thereof, and is made on a per-burst rather than per-symbol basis. The analysis of the almost sure convergence of such stochastic EM algorithms based on averaged gradient methods has been investigated in literature [37, 38, 39], and will not be pursued here.

5. Simulations

In this section we will describe the results of simulations that investigate the performance of the proposed receiver under practical conditions. Let ‘CS-ML’ denote a receiver having the CS analog front-end and ML demodulation in the DSP back-end, as described in Section 3.1. Similarly, let ‘CS-QP’ denote a receiver having the CS analog front-end and a QP demodulation in the DSP back-end, as described in Section 3.2. The performance of CS-ML is always an achievable lower bound with which we will compare the performance of the practical CS-QP receiver. Let ‘Genie-MF’ denote a receiver implementing a perfectly timed matched filter in an ISI-free environment. Clearly Genie-MF performance represents an ultimate lower bound, but it is not necessarily achievable when there is ISI, timing uncertainty or under-sampling.

Our discussion is divided into four parts. First, in Section 5.1, we will show an example of QP reconstruction of a transmitted burst. Then, in Section 5.2, we will investigate the effect of incomplete measurements, timing error and burst length on the BER of CS-ML and CS-QP receivers. In Section 5.3 we will demonstrate the robustness of CS-ML and CS-QP receivers to channels models and their random realizations, and in Section 5.4 we will discuss robustness to variations in the shape of the test functions. In Sections 5.1, 5.2, 5.3 and 5.4 the total system impulse response $h[\cdot]$ is assumed to be perfectly known. In Section 5.5 we will demonstrate the performance of the blind incremental algorithm that identifies the total system response.

All simulations were performed with $f_s = 10$ GHz and the IR pulse described in Section 2.1 (Figure 2). The baud rate is $f_{baud} = 500$ MBaud. Hence note that there is significant ISI lasting up-to $\sim 25 - 100$ symbols. N ranged from 300 to 1000 samples, depending on the channel type and realization. In all cases (except a part of Section 5.4) the measurement ensemble used was the square wave ensemble described in Section 3.4.3. The front-end filter $g(t)$ was chosen to be an ideal bandpass Nyquist-filter response truncated to $\pm \frac{5}{\Omega}$ seconds, and delayed by $\frac{5}{\Omega}$ seconds for causality. No repetition code was used in any of the simulations ($N_f = 1$). In the following, the quantity $\frac{Mf_s}{2\alpha\Omega N}$ will be called the *under-sampling factor*, where α is a constant. For an ideally band-limited signal we would set $\alpha = 1.0$ and specialize to the case discussed in Sections 3.4.1 and 3.4.2. However, in practice the pulse is not strictly band-limited (see Figure 2) and the Nyquist theorem is not directly applicable. Hence we need to define a practical measure of band-width by accounting for the ‘roll-off’ in the signal spectrum. This is accomplished by α , so that $(\alpha - 1)/\alpha$ is analogous to the roll-off factor used in communications literature. We have empirically found that, for the chosen pulse, a fixed value $\alpha = 1.5$ ensures that $\frac{Mf_s}{2\alpha\Omega N} = 1.0$ achieves a performance indistinguishable from MLSE under Nyquist rate sampling. Hence the case $\frac{Mf_s}{2\alpha\Omega N} = 1.0$ will be called *adequate sampling*, and the case $\frac{Mf_s}{2\alpha\Omega N} < 1.0$ will be called *under-sampling*.

5.1. An Example of QP Reconstruction

Consider the illustration in the panel of plots in Figure 3, which shows the various signals in the processing stream of the receiver. The simulation was done under the following conditions: $\text{SNR}_{bit} = 10$ dB, a CM1 channel, $N = 599$, $\Lambda_X = 151$, $M = 363$, $\Lambda_h = 449$, $\Gamma = 10$ samples ($\gamma = 1.0$ nanosecond), $K = 8$ bits per burst. The first (top) sub-plot shows the virtual information signal $X[n]$, that has only K non-zero sample of amplitudes B , with a random arrival time in the range $0 - 10$ samples. The second sub-plot shows the net impulse response of the channel and the pulse, $\psi[n] \star h_c[n]$. The third sub-plot shows the noiseless signal $U[n]$ impinging on the receiver antenna after passing through the linear channel $\psi[n] \star h_c[n]$, while the fourth sub-plot shows the noise contaminated signal $R[n]$ after the front-end filter. The final sub-plot displays the reconstruction \tilde{X} made by the QP optimization. Notice that the CM1 channel realization has a very wide temporal dispersion ~ 40 nanosecond, yet the reconstruction \hat{X}^k correctly estimates the location and sign of the impulses in X . As explained in Section 3.2, \tilde{X} is not necessarily in the set \mathcal{X} , and hence we must use a further hard decision rule (cf. Section 3.2) to declare the bit estimates \hat{B} .

5.2. Effect of Under-Sampling, Timing Uncertainty and Multiple Interfering Symbols

Now consider Figure 4 which shows, again for a fixed CM1 channel, the effect of under-sampling (via M), timing uncertainty (via Γ) and the burst length K . Figures 4(a),(b) correspond to $\frac{Mf_s}{2\alpha\Omega N} = 1.0, 0.25$ under ideal timing $\Gamma = 0$, and Figures 4(c),(d) correspond to $\frac{Mf_s}{2\alpha\Omega N} = 1.0, 0.25$ under uncertain timing $\Gamma = 10$. In each sub-figure we simulate CS-QP with $K = 1, 2, 4, 8, 16$ bits per burst and plot it with dashed lines with circle markers. We plot with solid blue lines the analytical performance of CS-ML given by equation (27), for $K = 1, 2, 4, 8$. Note that we do not give CS-ML performance for $K = 16$ because the calculation seems intractable. Finally we also plot the Genie-MF curve (dotted black line) for reference. The figure is very informative, and we can make several interesting observations:

- (i) We see that with ideal timing $\Gamma = 0$ and various amounts of under-sampling in sub-plots (a),(b), the CS-QP receiver performance is very close to CS-ML, for all K . This demonstrates that we can indeed recover the performance of an ideal coherent receiver with the proposed architecture. Furthermore with adequate sampling, all the CS-ML and CS-QP curves for various K coincide with the Genie-MF curve, implying that there is negligible loss due to the ISI. *Therefore there is no inherent justification for avoiding ISI by using a low baud rate, because it does not appreciably affect the distance spectrum of the modulation.* With under-sampling $\frac{Mf_s}{2\alpha\Omega N} = 0.25$, the curves of CS-QP and CS-ML for all K stay bunched together and have a consistent penalty of about 6.0 dB. w.r.t. the adequate sampling case, as predicted in Section 3.4.2.
- (ii) Even with non-ideal timing $\Gamma = 10$, the CS-QP receiver performance is reasonably close to CS-ML, for each K respectively. The loss in performance with

adequate sampling is less than one dB, while it is 1–2.5 dB with under-sampling $\frac{Mf_s}{2\alpha\Omega N} = 0.25$. Note that now even in the adequate sampling case, the $K = 1$ curve of CS-ML suffers a penalty of ~ 7.0 dB w.r.t. the corresponding curve of ideal timing from sub-plot (a). While this penalty is big, it is not catastrophic like the rake receiver which suffers a loss of 20 dB or so in performance. (This can be inferred from the auto-correlation of $h(t)$). More interestingly, as the number of bits in a burst K increases, the CS-ML curves start paring the loss and approach the ideal timing curve. This makes sense heuristically, because as we have multiple bits in a burst we can acquire timing ‘on the fly’. Asymptotically the timing acquisition will obviously become perfect. What is surprising is that with only $K = 8 - 16$ pulses we can practically eliminate the timing penalty.

(iii) Finally notice that in sub-plot (d), where we have both under-sampling as well as non-ideal timing, the penalty suffered by the CS-ML receiver is approximately the *additive composition* of the two individual penalties, and this is seen to consistently hold for all K . The CS-QP performance is also seen to mimic this behavior.

In summary, with CS-ML demodulation, the effects of under-sampling and timing uncertainty are approximately de-coupled. The loss due to under-sampling is consistently $10 \log_{10} \frac{Mf_s}{2\alpha\Omega N}$ dB, and unavoidable in principle. For lossless sampling we need only $M = \frac{2\alpha\Omega}{f_s} N$ projections, rather than N samples as in direct A/D. Thus we are inherently exploiting the bandpass nature of the signal. Timing uncertainty can be combated by using sufficiently many bits per burst, and the associated penalty can thus be practically eliminated. All these observations hold, with minor caveats, for the tractable CS-QP receiver too.

5.3. Robustness to Stochasticity of Channel Realizations

In the preceding discussion, we used one fixed realization from the CM1 channel model. Now we will study the effect of the stochasticity of the channel realizations and variations in the channel models. In Figure 5(a)-(d) we draw six random realizations from each model CM1 through CM4, and respectively plot the BER-vs-SNR_{bit} characteristic of the various receivers CS-QP, CS-ML and Genie-MF. In all cases we use a *constant number of projections* $M = 128$ which corresponds to a significant amount of under sampling ranging from $\frac{Mf_s}{2\alpha\Omega N} = 0.15$ to 0.25, depending of the channel model and realization. The timing uncertainty is $\Gamma = 10$ samples and the number of bits in a burst is $K = 8$.

First notice that, for every channel model, the CS-ML receiver performance does not vary by more than 2–3 dB no matter what the realization of the channel is. This demonstrates the *universality* of the ensemble used in the analog front-end even in the under-sampled case. (The reader will recall that the ensemble is not tuned to any particular channel model or realization.). Obviously if we used adequate sampling (large enough M), all the curves would bunch together with no appreciable variation. Secondly, observe that similar remarks continue to apply to the CS-QP performance too. The stochastic spread in the curves is

slightly more, say an additional dB, but otherwise it mimics the performance of CS-ML. The loss of CS-QP relative to CS-ML is the result of the sub-optimality of CS-QP and is in line with the loss observed in Figure 4(d). Finally, the Genie-MF performance obviously is invariant w.r.t. the channel model and realization, and is only an optimistic (unachievable) benchmark. We can thus conclude that the proposed receiver is indeed very robust to various channel models and the stochasticity of their realizations.

5.4. Robustness to Shape of Test Functions

To investigate the robustness of the CS-QP receiver to test functions of different shapes, we will now compare the performance of the receiver with the Square-wave ensemble as well as a Hanning-Windowed Fourier ensemble, under identical conditions of test-frequencies, under-sampling, timing, and a fixed CM1 channel impulse response. Figure 6 shows that the performance of the CS-QP receiver under the Fourier ensemble is only slightly worse (about a dB) than the Square-wave ensemble for conditions of ideal timing and adequate sampling, and for the case of non-ideal timing and under-sampling the difference is even smaller. This suggests that the shape of the test functions is not too critical as long they satisfy the decoherence property (which, in this case, means that they need to sufficiently sparse in the *frequency* domain).

5.5. Channel Acquisition and Tracking

Finally, we will investigate the performance of the incremental channel estimator proposed in Section 5.5. Figure 7(a) illustrates the Mean Squared Error (MSE) of the estimator under a realization from the CM1 model. Figure 7(b) shows the corresponding BER of the CS-QP receiver using the latest estimate of the channel. (We calculate the ‘instantaneous’ BER by performing a temporal averaging of the bit errors using an adequate IIR filter.) We also plot, with a dashed horizontal line, the BER of the CS-QP receiver operating under ideal channel knowledge, which was calculated by a separate simulation.

The parameters of the simulation are exactly those used in Section 5.3, namely $M = 128, \Gamma = 10, K = 8$. We simulate three values of SNR_{bit} namely 10, 13, 16 dB which are at the very low end of the operating range. (This is the most vulnerable region, where the error rates are significant even with ideal channel knowledge.) In each case we start the estimator from an *all-zeros initial value* \hat{h} . We use the step-size schedule $\epsilon[n] = \max\left(10^{-2}, \frac{10.0}{\sqrt{n}}\right)$, $n = 1, 2, \dots$, where n is the burst number. This schedule allows us to acquire the channel rapidly, and then settle into a steady state with a small MSE error floor. We would like to emphasize that we have simulated a fully blind algorithm where the bit decisions are supplied by the CS-QP demodulator of Section 3.2.2. From the MSE and BER results we conclude that the estimator acquires and tracks blindly and robustly even at low SNRs. We get qualitatively similar results with multiple realizations and with other models CM2-CM4, though we do not display them here for brevity. In every case we observe that the significant part of the acquisition is accomplished in less than a thousand bursts. The steady

state MSE (which depends on the channel model and realization) is sufficiently low so that there is negligible degradation in the BER relative to the case of ideal channel knowledge.

Note that when considering any channel identification algorithm, it is reasonable to separately consider the issues of acquisition and tracking. Acquisition refers to the process of coarsely identifying the channel response from a state of total ignorance. Tracking refers to tuning the estimate in response to small fluctuations of the true channel. Acquisition is typically performed *sporadically*, and hence it is not necessary to have an acquisition time constant of the order of the channel coherence time. For example, with reference to Figure 7, if the burst rate is 10^4 bursts per second, we have an acquisition time of about 100 ms. The tracking time constant, on the other hand, does need to be equal to or smaller than the channel coherence time. Since in steady state the step size is 10^{-2} , the estimator can track channel variations over intervals of around 10^2 bursts. Hence, with a burst rate 10^4 bursts per second, the tracking time constant is $10^{-4} \times 10^2$ seconds, implying that it can track channels with a coherence time of 10 ms or larger. This would be adequate in most practical scenarios. If the channel is expected to vary with a time constant smaller than 10 ms, or an acquisition time smaller than 100 ms is deemed necessary, we can implement one or both of the following options: (i) Increase the burst rate by a commensurate amount. (ii) Use training symbols, which reduce the stochasticity of the gradient and hence allow us to use a larger ϵ . Finally, also recall that we allow the TOA variations to be much faster (orders of magnitude faster) than 10 ms, since they need not be tracked by the estimator - the TOA is directly inferred by the QP demodulator.

6. Concluding Remarks

We have proposed a novel receiver for UWB Impulse Radio transmission based on the principle of Compressed Sensing (CS). It is very robust to timing uncertainty, ISI and under-sampling, and gives a performance that is consistently close to that of an optimal (ML) receiver. It allows the use of baud-rates comparable to the Nyquist rate, and hence large network loading factors. The demodulation procedure is insensitive to the nature of the multi-path channel (CM1, CM2 etc). Finally, although the proposed receiver needs to know the channel response in performing the demodulation of the payload, it also has a built-in ability to blindly identify it based on the CS measurements. The receiver is thus ideally suited to low-power applications with bursty traffic, like wireless sensor networks.

While in this paper we have only considered the problem of single-user UWB-IR demodulation, the proposed receiver architecture could also be used for demodulating multiple non-cooperating users. This would be based on the property that when the incoming signal is a mixture of ‘signature’ waveforms from several transmitters, out of which the CS-QP is matched to one, the receiver reconstructs the data from the matched transmitter while treating the others as

‘noise’ and hence suppressing them. However the non-Gaussianness and temporal structure of the interference may significantly affect this rejection capability, and hence this scenario needs to be studied in some detail before its practicality can be judged. Similarly the possibilities of joint detection or successive interference cancellation also need to be revisited in the CS setting. On a related note, in this work we have not addressed the important issue of narrow-band interference (NBI). We believe the CS-QP receiver can be made robust to NBI by the simple expedient of (a) using frequency selective test functions in the correlators, and (b) implementing a simple *digital notching* mechanism, wherein we identify and drop the few NBI corrupted measurements. We hope to accomplish this in future research.

References

- [1] M. Win, R. Scholtz, Impulse Radio: How It Works, *IEEE Commun. Letters* 2 (2) (1998) 36–38. doi:10.1109/4234.660796.
- [2] S. Roy, J. Foerster, V. Somayazulu, D. Leeper, Ultrawideband Radio Design: The Promise of High-Speed, Short-Range Wireless Connectivity, *Proc. of the IEEE* 92 (2) (2004) 295–311. doi:10.1109/JPROC.2003.821910.
- [3] H. Arslan, Z. N. Chen, M. D. Benedetto, *Ultra Wideband Wireless Communication*, John Wiley & Sons, Inc., 2006.
- [4] D. Culler, D. Estrin, M. Srivastava, Guest Editors’ Introduction: Overview of Sensor Network, *Computer* 37 (8) (2004) 41–49.
- [5] S. Gezici, Z. Tian, G. Giannakis, H. Kobayashi, A. Molisch, H. Poor, Z. Sahinoglu, Localization via ultra-wideband radios: A look at positioning aspects for future sensor networks, *IEEE Signal Processing Magazine* 22 (4) (2005) 70–84. doi:10.1109/MSP.2005.1458289.
- [6] A. F. Molisch, D. Cassioli, C. C. Chong, S. Emami, A. Fort, B. Kanan, J. Karedal, J. Kunisch, H. G. Schantz, K. Siwiak, M. Z. Win, A Comprehensive Standardized Model for Ultrawideband Propagation Channels, *IEEE Trans. on Antennas and Propagation* 54 (11) (2006) 3151–3166. doi:10.1109/TAP.2006.883983.
- [7] IEEE 802.15.4a-2007, Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (WPANs); Amendment 1: Add Alternate PHYs (Mar. 2007).

- [8] V. Lottici, A. D'Andrea, U. Mengali, Channel Estimation for Ultra-Wideband Communications, *IEEE J. Select. Areas Commun.* 20 (9) (2002) 1638–1645. doi:10.1109/JSAC.2002.805053.
- [9] J. Paredes, G. R. Arce, Z. Wang, Ultra-Wideband Compressed Sensing: Channel Estimation, *IEEE J. of Select. Topics in Signal Processing* 1 (3) (2007) 383–395. doi:10.1109/JSTSP.2007.906657.
- [10] L. Yang, G. B. Giannakis, Timing Ultra-Wideband Signals with Dirty Templates, *IEEE Trans. Commun.* 53 (11) (2005) 1952–1963. doi:10.1109/TCOMM.2005.858663.
- [11] A. A. D'Amico, U. Mengali, E. A. de Reyna, Energy-Detection UWB Receivers with Multiple Energy Measurements, *IEEE Trans. Wireless Commun.* 6 (7) (2007) 2652–2659. doi:10.1109/TWC.2007.05974.
- [12] Y.-L. Chao, R. A. Scholtz, Ultra-wideband transmitted reference systems, *IEEE Trans. Veh. Technol.* 54 (5) (2005) 1556–1569. doi:10.1109/TVT.2005.855700.
- [13] T. Q. S. Quek, M. Win, Analysis of UWB Transmitted-Reference Communication Systems in Dense Multipath Channels, *IEEE J. Select. Areas Commun.* 23 (9) (2005) 1863–1874. doi:10.1109/JSAC.2005.853809.
- [14] Y.-L. Chao, R. A. Scholtz, Optimal and Suboptimal Receivers for Ultra-Wideband Transmitted Reference Systems, *Proc. IEEE Global Telecom. Conf. (GLOBECOM)* 2 (2003) 759–763 Vol.2. doi:10.1109/GLOCOM.2003.1258340.
- [15] M. Pausini, G. Janssen, K. Witrisal, Performance Enhancement of Differential UWB Autocorrelation Receivers Under ISI, *IEEE Journal on Selected Areas in Communications* 24 (4) (2006) 815–821. doi:10.1109/JSAC.2005.863845.
- [16] D. L. Donoho, Compressed Sensing, *IEEE Trans. Inform. Theory* 52 (4) (2006) 1289–1306.
- [17] E. J. Candes, T. Tao, Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?, *IEEE Trans. Inform. Theory* 52 (12) (2006) 5406–5425.
- [18] Z. Wang, G. R. Arce, B. M. Sadler, J. L. Paredes, S. Hoyos, Z. Yu, Compressed UWB Signal Detection with Narrowband Interference Mitigation,

- IEEE Int. Conf. on UWB 2 (2008) 157–160. doi:10.1109/ICUWB.2008.4653375.
- [19] Z. Wang, G. R. Arce, B. M. Sadler, J. L. Paredes, X. Ma, Compressed Detection for Pilot Assisted Ultra-Wideband Impulse Radio, IEEE Int. Conf. on UWB (2007) 393–398doi:10.1109/ICUWB.2007.4380976.
- [20] Z. Wang, G. R. Arce, J. L. Paredes, B. M. Sadler, Compressed Detection for Ultra-Wideband Impulse Radio, IEEE Workshop on Sig. Proc. Advances in Wireless Communications (2007) 1–5<http://dx.doi.org/10.1109/SPAWC.2007.4401384> doi:10.1109/SPAWC.2007.4401384.
- [21] J. Kusuma, I. Maravic, M. Vetterli, Sampling with Finite Rate of Innovation: Channel and Timing Estimation for UWB and GPS, Proc. IEEE Int. Conf. Commun. (ICC) 5 (2003) 3540–3544 vol.5. doi:10.1109/ICC.2003.1204112.
- [22] L. Huang, N. E. Ghouti, O. Rousseaux, B. Gyselinckx, Timing Tracking Algorithms for Impulse Radio (IR) Based Ultra Wideband (UWB) Systems, Int. Conf. Wireless Communications, Networking and Mobile Computing (2007) 570–573doi:10.1109/WICOM.2007.148.
- [23] M. Win, R. Scholtz, Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple-access communications, IEEE Trans. Commun. 48 (4) (2000) 679–689. doi:10.1109/26.843135.
- [24] A. Saleh, R. Valenzuela, A Statistical Model for Indoor Multipath Propagation, IEEE Journal on Selected Areas in Communications 5 (2) (1987) 128–137.
- [25] C. C. Chong, S. K. Yong, A Generic Statistical-Based UWB Channel Model for High-Rise Apartments, IEEE Transactions on Antennas and Propagation 53 (8) (2005) 2389–2399. doi:10.1109/TAP.2005.852505.
- [26] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
- [27] J. G. Proakis, Digital Communications, McGraw-Hill, New York, San Francisco, Toronto, London, 2001.
- [28] E. J. Candes, J. Romberg, T. Tao, Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information, IEEE Trans. Inform. Theory 52 (2) (2006) 489–509.

- [29] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, *J. of the Roy. Stat. Soc. 58 - Series B* (1) (1996) 267–288.
- [30] E. Candes, T. Tao, The Dantzig Selector: Statistical Estimation When p Is Much Larger Than n , *Ann. Statist.* 35 (6) (2007) 2313–2351.
- [31] C. Zhu, Stable Recovery of Sparse Signals Via Regularized Minimization, *IEEE Trans. Inform. Theory* 54 (7) (2008) 3364–3367. doi:10.1109/TIT.2008.924707.
- [32] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, L. Coulot, Sparse Sampling of Signal Innovations, *IEEE Signal Processing Magazine* 25 (2) (2008) 31–40. doi:10.1109/MSP.2007.914998.
- [33] A. Oka, L. Lampe, Compressed Sensing Reception of Bursty UWB Impulse Radio is Robust to Narrow-band Interference Accepted for presentation at the IEEE Global Communications Conference (GLOBECOM) 2009.
- [34] H. Robbins, S. Monro, A Stochastic Approximation Method, *Ann. Math. Stat.* 22 (1951) 400–407.
- [35] R. M. Neal, G. E. Hinton, A View of the EM Algorithm That Justifies Incremental, Sparse, and Other Variants (1998) 355–368.
- [36] A. Dempster, N. Laird, D. Rubin, Maximum Likelihood From Incomplete Data Via the EM Algorithm, *J. of the Roy. Stat. Soc. 39 (Series B)* (1977) 1–38.
- [37] H. J. Kushner, J. Yin, *Stochastic Approximation Algorithms and Applications*, Springer-Verlag, 1997.
- [38] L. Ljung, Analysis of Recursive Stochastic Algorithms, *IEEE Trans. Automatic Control* AC-22 (4) (1977) 205–221.
- [39] M. Metivier and P. Priouret, Applications of a Kushner and Clark Lemma to General Classes of Stochastic Algorithms, *IEEE Trans. Inform. Theory* 30 (2) (1984) 140–151.

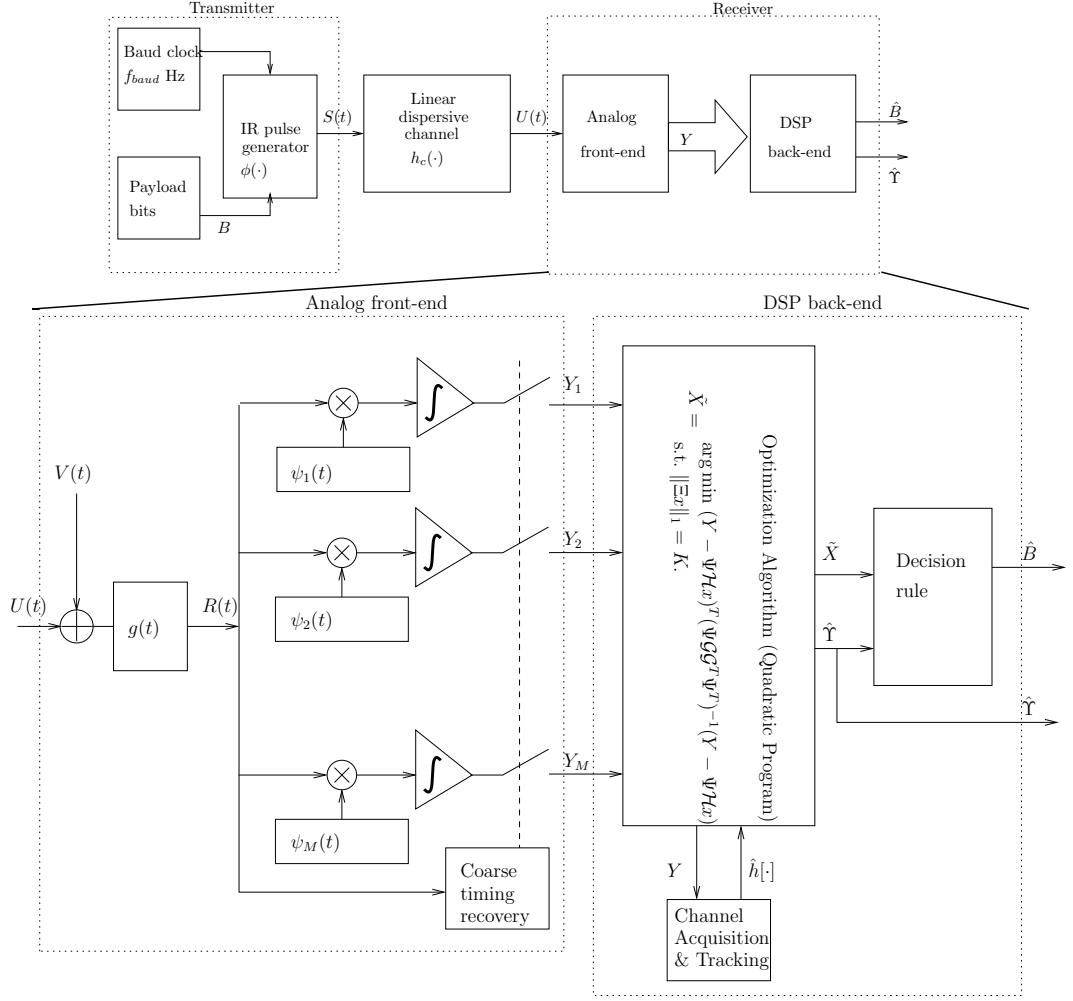


Figure 1: Block diagram of the UWB-IR system.

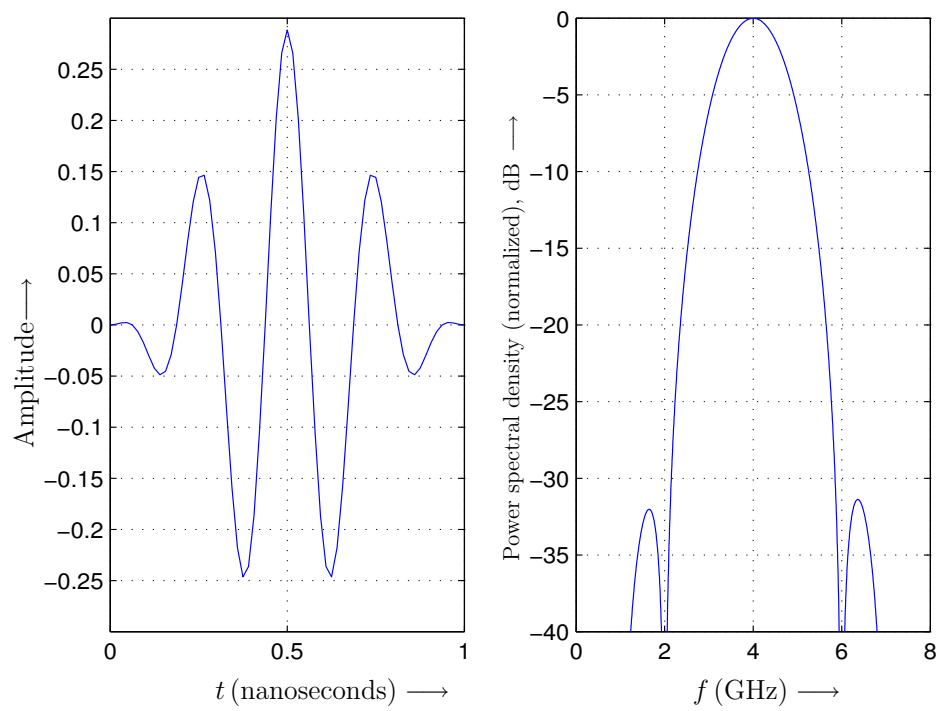


Figure 2: Impulse Radio pulse shape $\phi(t)$, and its power spectrum.

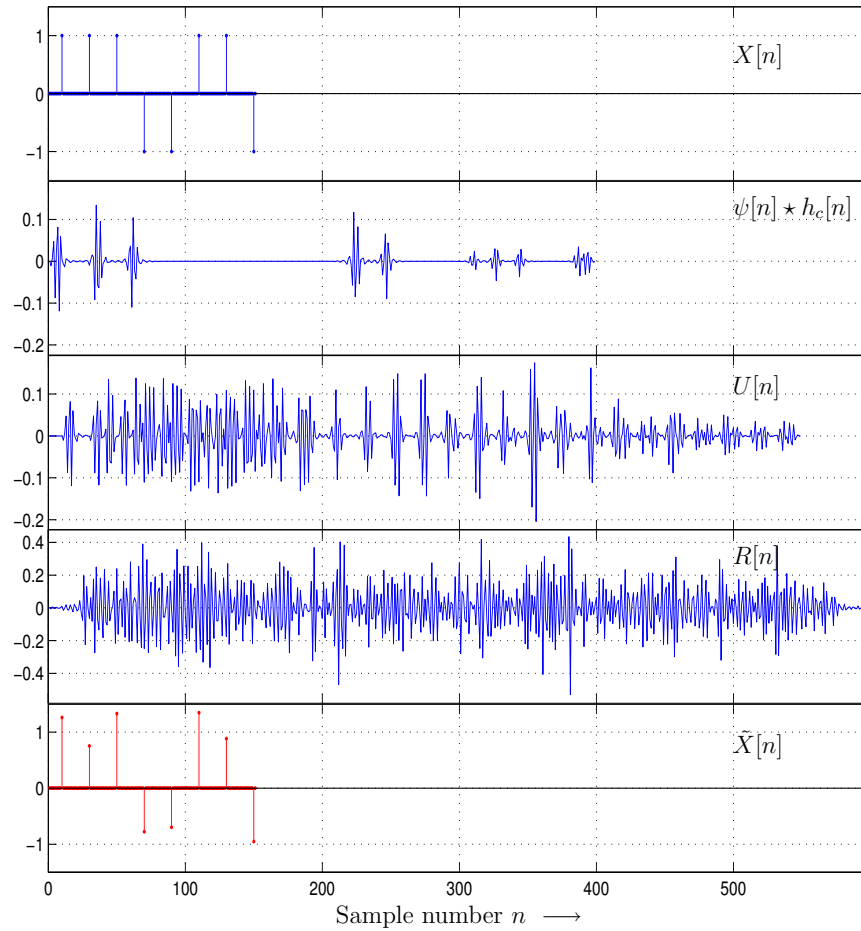


Figure 3: Various signals in the processing stream: the first (top) sub-plot is the virtual information signal $X[n]$, the second sub-plot is the response of the pulse and the channel, $\psi[n] \star h_c[n]$, the third-subplot is the signal impinging on the antenna, $U[n]$, the fourth sub-plot is the signal after the front-end filter, $R[n]$, and the final sub-plot is the reconstructed information signal $\hat{X}[n]$. $f_{baud} = 500$ MBaud, $\text{SNR}_{bit} = 10$ dB, CM1 channel, $N = 599$, $\Lambda_X = 151$, $M = 363$, $\Lambda_h = 449$, $\Gamma = 10$ samples ($\gamma = 1.0$ nanoseconds), $K = 8$ bits per burst.

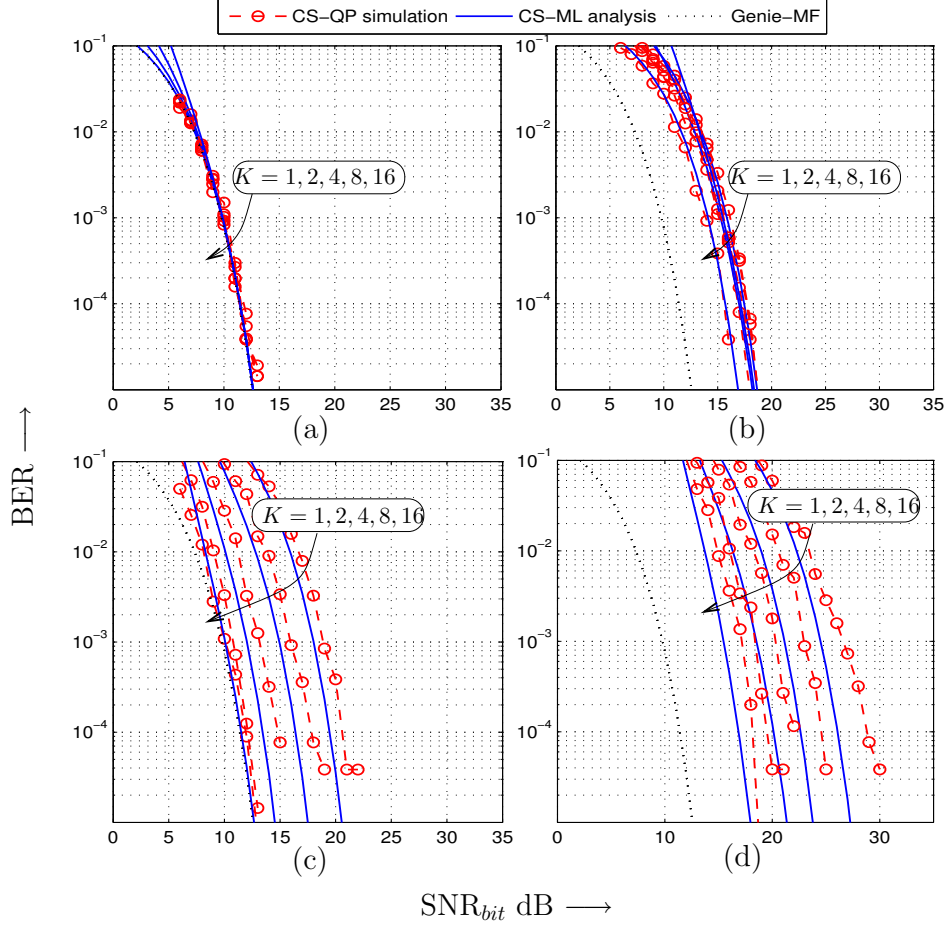


Figure 4: Effect of under-sampling, timing uncertainty and burst length on the receiver performance. Sub-plots (a),(b) correspond to $\frac{Mf_s}{2\alpha\Omega N} = 1.0, 0.25$ under $\Gamma = 0$, and sub-plots (c),(d) correspond to $\frac{Mf_s}{2\alpha\Omega N} = 1.0, 0.25$ under $\Gamma = 10$. In each sub-figure we simulate CS-QP with $K = 1, 2, 4, 8, 16$ bits per burst and plot it with dashed lines with circle markers. We plot with solid blue lines the analytical performance of CS-ML given by equation (27), for $K = 1, 2, 4, 8$. The dotted line is the Genie-MF performance in an ISI free regime.

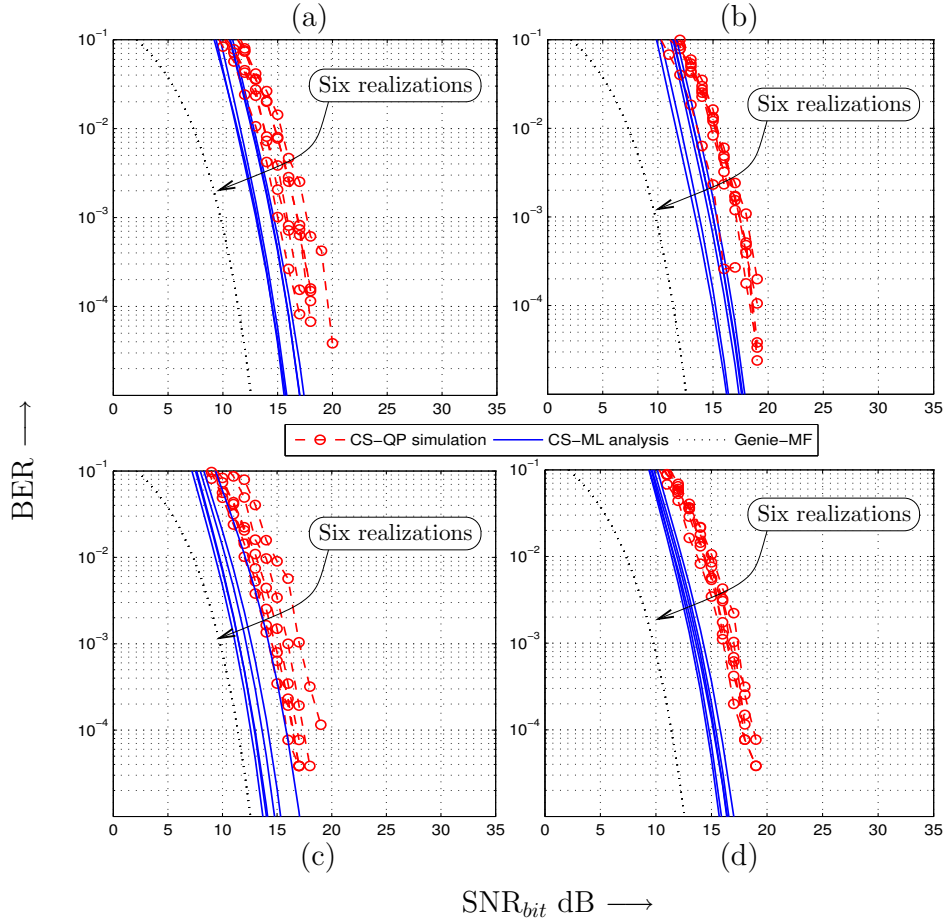


Figure 5: Robustness to stochastic channel realizations. Sub-plots (a) through (d) correspond to channel models CM1 through CM4 respectively. Six stochastic realizations are derived from each model. For each realization the BER vs SNR_{bit} characteristic of CS-ML and CS-QP is provided. The Genie-MF curve is also shown in each sub-plot. In all cases $M = 128$, $\Gamma = 10$ and $K = 8$.

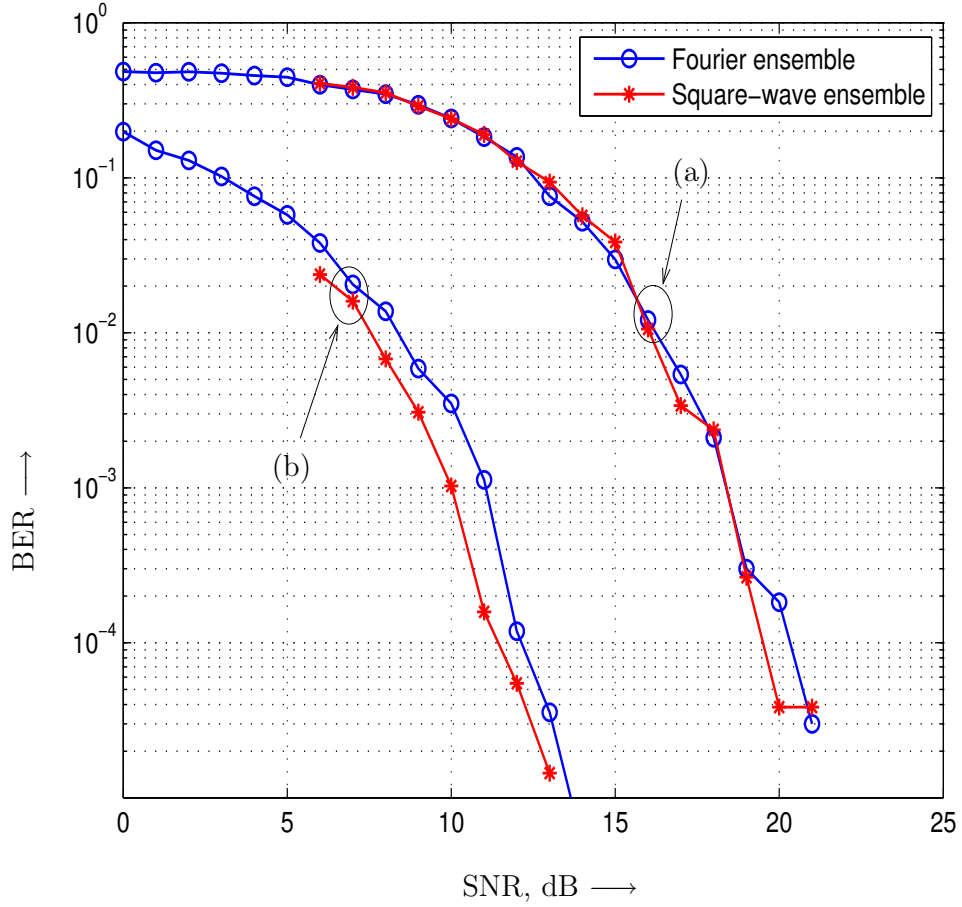


Figure 6: Performance of the CS-QP receiver with Fourier and Square-wave ensembles under identical conditions of test-frequencies and a fixed CM1 channel impulse response. (a) Under-sampling ($\frac{Mf_s}{2\alpha\Omega N} = 0.25$) and poor timing ($\Gamma = 10$), and (b) Adequate sampling ($\frac{Mf_s}{2\alpha\Omega N} = 1.0$) and perfect timing ($\Gamma = 0$).

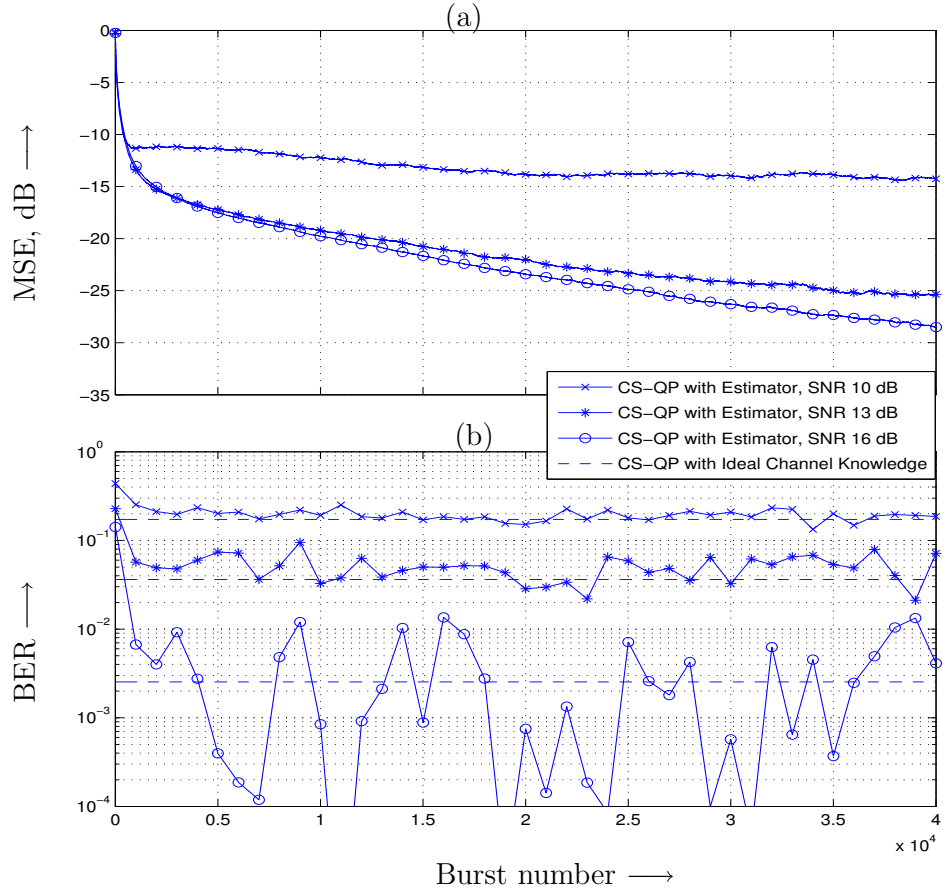


Figure 7: Performance of blind incremental channel acquisition starting from an all zero response. (a) Mean Squared Error (MSE), in dB, of the estimated response relative to the true response, $20 \log_{10} \frac{\|h - \hat{h}\|_2}{\|h\|_2}$. (b) BER of the CS-QP receiver using the latest estimate of the channel, \hat{h} . Three values of SNR_{bit} have been simulated, namely 10, 13, 16 dB. Horizontal red dashed lines are the corresponding BERs of the CS-QP receiver operating under ideal channel knowledge h . $M = 128, \Gamma = 10, K = 8$. The true channel realization, h , is from the CM1 model.