Application of Compressive Sensing to Sparse Channel Estimation

Christian R. Berger, Carnegie Mellon University

Zhaohui Wang, Jianzhong Huang, and Shengli Zhou, University of Connecticut

ABSTRACT

Compressive sensing is a topic that has recently gained much attention in the applied mathematics and signal processing communities. It has been applied in various areas, such as imaging, radar, speech recognition, and data acquisition. In communications, compressive sensing is largely accepted for sparse channel estimation and its variants. In this article we highlight the fundamental concepts of compressive sensing and give an overview of its application to pilot aided channel estimation. We point out that a popular assumption — that multipath channels are sparse in their equivalent baseband representation — has pitfalls. There are over-complete dictionaries that lead to much sparser channel representations and better estimation performance. As a concrete example, we detail the application of compressive sensing to multicarrier underwater acoustic communications, where the channel features sparse arrivals, each characterized by its distinct delay and Doppler scale factor. To work with practical systems, several modifications need to be made to the compressive sensing framework as the channel estimation error varies with how detailed the channel is modeled, and how data and pilot symbols are mixed in the signal design.

INTRODUCTION

WHAT IS COMPRESSIVE SENSING?

Since the term compressive sensing was coined a few years ago [1, 2], this subject has been under intensive investigation [3–5]. It has found broad application in imaging, data compression, radar, and data acquisition to name a few (see overview in [4, 5]).

In a nutshell, compressive sensing is a novel paradigm where a signal that is sparse in a known transform domain can be acquired with much fewer samples than usually required by the dimensions of this domain. The only condition is that the sampling process is incoherent with the transform that achieves the sparse representation and sparse means that most weighting coefficients of the signal representation in the transform domain are zero. While it is obvious that a signal that is sparse in a certain basis can be fully represented by an index specifying the basis vectors corresponding to non-zero weighting coefficients plus the coefficients — determining which coefficients are non-zero would usually involve calculating all coefficients, which requires at least as many samples as there are basis functions. The definition of incoherence usually states that distances between sparse signals are approximately conserved as distances between their respective measurements generated by the sampling process. In this sense the reconstruction problem has per definition a unique solution.

Making the compressive sensing formulation practical hinges on two conditions:

- Is the incoherence property achievable with a feasible sampling scheme?
- Are there computationally tractable algorithms that can reconstruct the original signal from these samples?

The answers to these questions created the field of compressive sensing and we will try to review the basics of these answers in this article.

APPLICATIONS OF COMPRESSED SENSING IN COMMUNICATIONS

So far compressive sensing has been successfully applied in several signal-processing fields, specifically in imaging the technology has achieved a certain level of maturity. In communications the range of applications so far has been rather limited, with the exception of channel estimation — although in many variations. To cite a few examples:

- Sparse channel estimation in ultra-wideband, was motivated by the ability to resolve individual arrivals or clusters of arrivals in multipath channels [6].
- Considering mobile radio channels, each path is characterized by a delay and a relative Doppler speed [7, 8].
- Underwater acoustic channels are known to exhibit only few arrivals in a long delay spread with each path having different Doppler speed [9].

A variation on channel estimation is the combination with active user detection in code division multiple access [10] or spectrum sensing for cognitive radios.

Another proposed application of compressive sensing in communications is coding over the

 0163-6804/10/$25.00 © 2010 IEEE

real numbers (vs. finite fields as commonly used in coding theory) under a channel model that produces few very large errors (similar to erasures). Although this leads to direct application of compressive sensing algorithms and performance guarantees [11], it is so far unclear if this will lead to practical applications that would replace current error correction schemes.

THIS ARTICLE

Clearly the motivation to use compressive sensing in channel estimation is the observation that some channels are characterized by sparse multipath — by that we mean that there are much fewer distinct arrivals as there are baseband channel taps. With this in mind compressive sensing promises to estimate the channel with much less pilot overhead or at higher accuracy with a constant number of pilots. The common assumption is that a sparse multipath channel leads to a baseband channel model where most taps are negligible. We take a closer look at this and find that in a channel modeled by specular (point) scatterers the number of nonzero baseband taps depends very much on what one defines as negligible. Using instead an oversampled baseband model, the representation of the channel becomes ambiguous, but also more sparse.

In underwater acoustic (UWA) communications, channels are characterized by long delay spread and significant Doppler effects. The long channel delay spread leads to severe inter-symbol interference (ISI) in single-carrier transmissions, while in multicarrier approaches like orthogonal frequency division multiplexing (OFDM) the aforementioned Doppler effects destroy the orthogonality of the sub-carriers and lead to inter-carrier interference (ICI). On top of high equalization complexity, the ISI or ICI corresponds to a convolution with a time-varying impulse response, leading to a large amount of unknown channel coefficients. While it is well recognized in the community that UWA channels are usually sparse [12], there are major challenges to overcome when applying compressive sensing to exploit channel sparsity.

As an example, we show a block-by-block OFDM receiver that re-estimates the channel for every OFDM symbol. To apply compressive sensing one needs to consider the following points:

- A channel model needs to be established that leads to a sparse representation of the channel coefficients, and is accurate (enough) within the considered time interval.
- When placing the pilots, one needs to ensure that ICI from other pilots can be observed.
- When estimating the channel based on pilots, ICI from the unknown data symbols has to be treated as noise.

After going through the details of applying compressive sensing to channel estimation in UWA multicarrier communications, we illustrate the performance using numerical simulation and experimental data.

The article is organized as follows. In the next section we give a more detailed overview of compressive sensing, and in the following section we describe some of the popular compressive sensing recovery algorithms. We then look at sparse representations of multipath channels and explain in detail the application of compressive sensing to UWA communications. We conclude in the final section.

Notation — We represent matrices and vectors with bold upper and lower case letters respectively, $\mathbf{A}$, $\mathbf{c}$; Superscripts T and H denote the transpose and hermitian respectively, $\mathbf{A}^T$, $\mathbf{A}^H$. With $|\mathbf{c}|$ we denote the Euclidean norm.

Clearly the motivation to use compressive sensing in channel estimation is the observation that some channels are characterized by sparse multipath: by that we mean that there are much fewer distinct arrivals as there are baseband channel taps.

COMPRESSIVE SENSING

SPARSE REPRESENTATION

Consider a signal $\mathbf{y} \in \mathbb{C}^n$ that can be represented in an arbitrary basis, $\{\psi_k\}_{k=1}^n$, with the weighting coefficients x_k. Stacking the coefficients into a vector, $\mathbf{x}$, the relationship with $\mathbf{y}$ is obviously through the transform $\mathbf{y} = \mathbf{\Psi x}$, where $\mathbf{\Psi} = [\psi_1, \psi_2, \ldots, \psi_n]$ is a full rank $n \times n$ matrix. A common n example would be a finite length, discrete time signal that one could represent as discrete sinusoids in a limited bandwidth. The matrix $\mathbf{\Psi}$ would then be the discrete Fourier transform (DFT) matrix.

In compressive sensing one is particularly interested in any basis that allows a *sparse* representation of $\mathbf{y}$, i.e., a basis $\{\psi_k\}_{k=1}^n$ such that most x_k are zero. Obviously if one knows $\mathbf{y}$, one could always choose some basis for which $\mathbf{y} = \psi_{k_0}$ for some k_0; then all x_k, $k \neq k_0$, would be zero. This trivial case is not of interest, instead one is interested in a predetermined basis that will render a sparse or approximately sparse representation of any $\mathbf{y}$ that belongs to some class of signals.

EXACTLY AND APPROXIMATELY SPARSE SIGNALS

A signal is called s-sparse, if it can be exactly represented by a basis, $\{\psi_k\}_{k=1}^n$, and a set of coefficients x_k, where only s coefficients are nonzero. A signal is called approximately s-sparse, if it can be represented up to a certain accuracy using s non-zero coefficients. Since the desired accuracy depends on the application, signals considered as approximately sparse usually have the property that the reconstruction error decreases super-linearly in s, therefore any required accuracy can be achieved by only slightly increasing s.

As an example of an s-sparse signal, consider the class of signals that are the sum of s sinusoids chosen from the n harmonics of the observed time interval. Now obviously the DFT basis will render an s-sparse representation of any such $\mathbf{y}$, i.e., taking the DFT of any such signal would render only s non-zero values x_k.

An example of approximately sparse signals is when the coefficients x_k, sorted by magnitude, decrease following a power law. This includes smooth signals or signals with bounded variations [4]. In this case the sparse approximation constructed by choosing the s largest coefficients is guaranteed to have an approximation error that decreases with the same power law as the coefficients.

What all these algorithms have in common is that they lead to convex optimization problems, which can be solved efficiently with advanced techniques, such as interior-point methods, projected gradient methods, or iterative thresholding.

SENSING

So far it was assumed that $\mathbf{y}$ is available, and that one can simply apply the transform into the domain of $\{\psi_k\}_{k=1}^n$ to determine which x_k are relevant (non-zero). Although this case does exist and is important for some forms of data-compression, the real application of compressive sensing is the acquisition of the signal from m, possibly noisy, measurements $z_l = \phi_l^H \mathbf{y} + v_l$ for $l = 1, \ldots, m$, where here it is assumed that v_l is zero-mean complex Gaussian distributed with variance N_0 and the noiseless case is included for $N_0 \to 0$. The signal acquisition process can now be written using the $m \times n$ matrix $\mathbf{A}$,

$$\mathbf{z} = \mathbf{\Phi}^H \mathbf{y} + \mathbf{v} = \underbrace{\mathbf{\Phi}^H \mathbf{\Psi}}_{\mathbf{A}} \mathbf{x} + \mathbf{v}$$

where $\mathbf{\Phi} = [\phi_1, \phi_2, \ldots, \phi_m]$ is an $n \times m$ matrix and $\mathbf{z} = [z_1, z_2, \ldots, z_m]^T$ is the stacked measurement vector. Since this is a simple linear Gaussian model, it is well posed as long as $\mathbf{A}$ is at least of rank n. By well posed we simply mean that there exists some estimator $\hat{\mathbf{x}}$ (or $\hat{\mathbf{y}}$ for that matter), whose estimation error is proportional to the noise variance; therefore as the noise variance approaches zero, the estimation error does as well. This generally requires at least $m \geq n$ measurements if $\mathbf{y}$ is unconstrained in $\mathbb{C}^n$.

SIGNAL RECOVERY AND RIP

The novelty in compressive sensing is that for signals $\mathbf{y}$ that are s-sparse in some $\{\psi_k\}_{k=1}^n$, less measurements are sufficient to make this a well posed problem. The requirement on $\mathbf{A}$ to have at least rank n is replaced by the restricted isometry property (RIP) (first defined in [11]) that we will explain in the following.

For any matrix $\mathbf{A}$ with unit-norm columns one can define the restricted isometry constants δ_s as the smallest number such that, $|\mathbf{Ax}|^2 \geq (1 - \delta_s)|\mathbf{x}|^2$ and $|\mathbf{Ax}|^2 \leq (1 + \delta_s)|\mathbf{x}|^2$ for any $\mathbf{x}$ that is s-sparse. This can be seen as conserving the (approximate) length of s-sparse vectors in the measurement domain and effectively puts bounds on the Eigenvalues of any $s \times s$ submatrix of $\mathbf{A}^H\mathbf{A}$.

Now assuming that under all s-sparse vectors, one chooses the estimate $\hat{\mathbf{x}}$ that has the smallest distance to the observations, $|\mathbf{z} - \mathbf{A}\hat{\mathbf{x}}|^2$, it is easily shown that the estimation error is bounded by $E\{|\mathbf{x} - \hat{\mathbf{x}}|^2\} \leq 2mN_0/(1 - \delta_{2s})$. This uses the fact that the estimation error $\tilde{\mathbf{x}} := \mathbf{x} - \hat{\mathbf{x}}$ is $2s$-sparse. So we see that the signal recovery problem is well-posed as long as $\delta_{2s} < 1$, but since the δ_s are monotonic in s, $\delta_s \leq \delta_{s+1}$, and usually increase gradually, it is commonly said that $\mathbf{A}$ obeys the RIP if δ_s is not too close to one.

In case of approximately sparse signals, the error caused by noisy observations is additive with the error caused by the approximation as s-sparse. Therefore a good choice of s needs to consider the noise level N_0, since a trade-off exists between choosing a smaller s that increases the approximation error, but decreases the error caused by the noise due to the monotonic nature of the δ_s and vice-versa.

SENSING MATRICES

While evaluating the RIP for a particular matrix at hand is (at worst) an NP-hard problem, there are large classes of matrices that obey the RIP with high probability, that is $\delta_s \ll 1$ for any $s \ll m$. Specifically for random matrices like i.i.d. Gaussian or Bernoulli entries, or randomly selected rows of an orthogonal $(n \times n)$ matrix (e.g., the DFT), it can be shown that for $m \geq Cs\log(n/s)$ measurements the probability that $\delta_s \geq \delta$ decreases exponentially with m and δ. With other words, as long as one takes enough measurements, i.e., increase m, the probability of any such matrix obeying the RIP for a given threshold δ can be made arbitrarily small. Although the constant C is only loosely specified for the various types of matrices, the fact that the probability decreases exponentially is encouraging as to the number of required measurements. Furthermore it is important to consider that these bounds are on worst cases, so that on the average much fewer measurements, m, will be sufficient.

ALGORITHMS

Previously we considered the estimator that chooses the solution with minimum distance from the observations between all s-sparse vectors in $\mathbb{C}^n$ to show that the average estimation error is bounded. This is in essence a combinatorial problem, which has exponential complexity. In case s is not known, or for an approximately sparse signal, a joint cost function has to be used that penalizes less sparse solutions vs. a better fit of the observations. This can be achieved using a Lagrangian formulation adding a penalty proportional to s, which is usually formulated using the *zero-norm*, $\|\mathbf{x}\|_0$ that counts the non-zero elements in $\mathbf{x}$. This further increases the size of the combinatorial problem as all s-sparse vectors for various values of s have to be considered now.

Other algorithms that reconstruct a signal taking advantage of its sparse structure have been used well before the term compressive sensing was coined. The surprising discovery is that it can be shown that several of these algorithms will — under certain conditions — render the same solution as the combinatorial approach. These conditions largely amount to tighter constraints on the sparsity of $\mathbf{x}$ beyond identifiability. We briefly introduce the two main types of algorithms.

CONVEX/L_1-BASED

Since the exact formulation using the zero-norm $\|\mathbf{x}\|_0$ is not amenable to efficient optimization, an immediate choice is its convex relaxation, leading to the following Lagrangian formulation,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} |\mathbf{Ax} - \mathbf{z}|^2 + \xi \|\mathbf{x}\|_{l_1}$$

where the l_1-norm is defined as $\|\mathbf{x}\|_{l_1} = \sum_{k=1}^n |x_k|$. While the l_1-norm has been used in various applications to promote sparse solutions in the past [4, references therein], it is now largely popular under the name Basis Pursuit (BP), as introduced in [13]. While originally the term BP was used to designate the case of noiseless measurements and the qualifier Basis Pursuit De-Noising to refer to the case of noisy measurements [13], we will generally refer to both cases simply by BP. In statistics the Lasso algorithm is well-known [14], which can be shown to be equivalent to BP under appropriate parameterization.

What all these algorithms have in common is that they lead to convex optimization problems, which can be solved efficiently with advanced techniques, such as interior-point methods, projected gradient methods, or iterative thresholding. Due to the relaxation and numerical accuracy the solutions will not be exactly sparse, but will exhibit numerous small values that do not contribute significantly to the estimation error. If an exactly sparse solution is sought, an additional thresholding or de-biasing stage can remove the small components.

The discovery that there are conditions under which convex relaxation will render the same result as the combinatorial formulation was the birth of compressive sensing [1, 2]. These conditions usually consider the minimum number of measurements m required to identify an s-sparse signal with high probability, given a certain measurement matrix. For example, in [1] it is shown for m noiseless measurements taken using random rows of the DFT matrix, that if $m > C_M s \log(n)$, any s-sparse signal can be recovered with at least probability $1 - \mathcal{O}(n^{-M})$, where the constant C_M is roughly linear in the parameter M. One immediately notices that this formulation closely resembles the criterion for identifiability, but the constants will take different values.

Greedy Pursuits

Another approach to the combinatorial problem is based on dynamic programming. In this type of approach the combinatorial problem is circumvented by heuristically choosing which values of $\mathbf{x}$ are non-zero and solving the resulting constrained least-squares problem. The most popular algorithms of this type are greedy algorithms, like Matching Pursuit (MP) or Orthogonal Matching Pursuit (OMP), that identify the nonzero elements of $\mathbf{x}$ in an iterative fashion. A short algorithmic description of OMP would be:

1. Initialize the set of non-zero elements as empty, the observations are set as the residual, $\mathbf{r} = \mathbf{z}$.
2. Correlate all columns of $\mathbf{A}$ with the residual, $\mathbf{A}^H\mathbf{r}$, choose the largest element by magnitude and add its index to the set of non-zero elements.
3. With the constraint that only elements of $\mathbf{x}$ are non zero that have been added to the set previously, find an estimate $\hat{\mathbf{x}}$ that minimizes $|\mathbf{z} - \mathbf{A}\hat{\mathbf{x}}|^2$.
4. Update the residual as $\mathbf{r} = \mathbf{z} - \mathbf{A}\hat{\mathbf{x}}$.
5. Repeat steps 2–4 until either a known s is reached or the norm of the residual $|\mathbf{r}|^2$ falls below a predetermined threshold.

This type of algorithm has been popular mainly because it can be easily implemented and has low computational complexity, but recently it has been shown that this algorithm will also render the optimal solution [15], whereby the constraints are somewhat stronger. This has lead to renewed interest in dynamic programming based solutions, leading to new greedy pursuit algorithms [16, references therein].

After reviewing the theory of compressive sensing, we will next illustrate how this matches to the task of channel estimation. To this end we will first look at sparse representations of the channel frequency response and then study the specific case of underwater acoustic multicarrier communication.

How Sparse are Multipath Channels?

Multipath Model

Channel estimation is in essence a problem of system identification; a known signal $s(t)$ is transmitted and we receive the signal $r(t)$ that has gone through the unknown system H. After H has been estimated with sufficient accuracy, its effect can be accounted for in the following data transmission.

For simplicity, let us consider a linear time-invariant system, which can be completely characterized by its impulse response $h(\tau)$ or its frequency response $H(f)$, and neglect any random noise. The frequency spectrum $R(f)$ of the received signal will then simply be the product of the transmitted spectrum $S(f)$and the channel frequency response $H(f)$. Without loss of generality, assume that the signal $s(t)$ is a multicarrier signal defined in the frequency domain by the complex symbols $S(f_k) = s[k]$, transmitted on the K subcarriers $f_k = f_c + k\Delta f$, with k taking values between $\pm K/2$. The receiver samples the waveform and applies a DFT; the outputs will correspond to sampling the waveform at frequencies f_k,

$$R(f_k) = H(f_k)s[k]$$

We see that to recover the transmitted signal one will need the channel frequency response at corresponding frequencies $H(f_k)$ (Fig. 1, top). The underlying assumption of applying compressed sensing to channel estimation is that the channel frequency response $H(f_k)$ is sparse in some basis, or at least approximately so. This is usually based on the model that the impulse response $h(\tau)$ consists of P specular (point) scatterers (Fig. 1, middle). The complex amplitudes ξ_p include attenuation and initial phase, and the delays τ_p are assumed to be less than some maximum delay spread.

The vector of interest, $\mathbf{y}$, consists of the stacked frequency response at the K subcarriers $H(f_k)$. Now each entry of $\mathbf{y}$ is a linear combination of P complex phases, $\exp(-j2\pi f_k\tau_p)$, with complex weighting coefficients ξ_p. Arranging the complex phases in length K vectors ψ_p renders a P-sparse representation of $\mathbf{y}$, with $x_p = \xi_p$. Since the τ_p are random values from a continuous distribution, one cannot choose a (finite) basis that will include all possible basis vectors ψ_p that correspond to the random τ_p. In the following several fixed choices of $\mathbf{\Psi}$ will be considered that lead to more or less sparse representations of the channel frequency response.

Another approach to the combinatorial problem is based on dynamic programming. In this type of approach the combinatorial problem is circumvented by heuristically choosing which values of **x** *are non-zero and solving the resulting constrained least-squares problem.*

SPARSE APPROXIMATION

Commonly the equivalent baseband model is used; this basis simply limits the τ_p to be multiples of the sampling time ℓT_s, which is the inverse of the bandwidth $1/T_s = B = K\Delta f$. With potential delays $\ell = 0, \ldots, K - 1$ the matrix $\mathbf{\Psi}$ turns out to be the $K \times K$ DFT matrix and $\mathbf{x}$ turns out to be samples at baseband rate of the bandpass filtered version of $h(\tau)$ (the bandpass is from $f_c - B/2$ to $f_c + B/2$ due to the transmitted signal). Since the DFT matrix is a unitary transform, one can calculate $\mathbf{x}$ by taking the inverse DFT of $\mathbf{y}$, which is generally not sparse. We purposefully consider this simple case, because the optimum s-sparse approximation of $\mathbf{y}$ (in the mean squared error sense) using this basis can be trivially determined by keeping the s largest values of the inverse DFT of $\mathbf{y}$. The first 64 values of $h(\ell)$ corresponding to $h(\tau)$ are plotted in Fig. 1 (bottom). Although there are only few large values, there seem to be a substantial number of smaller, but maybe not negligible values.

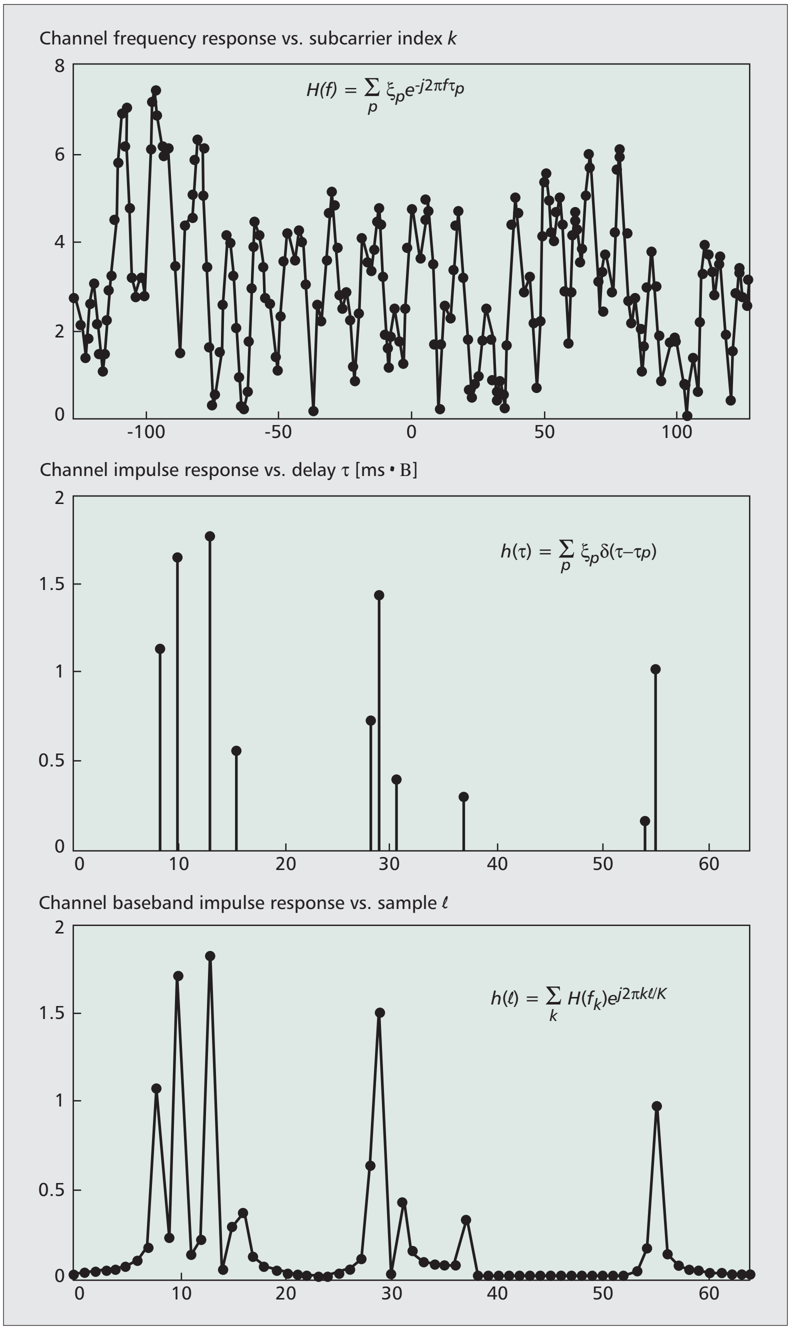

Figure 1. *The channel frequency response* H(f) *maps to the impulse response* h(τ), *but from a limited number of samples* H(f_k) *only the baseband model* h(ℓ) *can be determined unambiguously; in this example there are* P = *10 discrete paths, and* K = *256 frequency samples; all plots are magnitude only.*

Next, a redundant basis is considered (which we often refer to as dictionary), generated by delays at a finer grained resolution of $\ell T_s/\lambda$, where λ is the oversampling factor relative to baseband sampling. This is a quite natural approximation of the continuous time τ_p, but leads to an uncommon case for the compressive sensing theory. One might think that this is the baseband model of a system with a λ-times larger bandwidth and that K samples were chosen deterministically. The corresponding matrix $\mathbf{A}$ is a $K \times \lambda K$ partial DFT matrix, that turns out to have quite bad sensing properties $\delta_s \approx 1$, since neighboring columns will be highly correlated. This is obvious, since when taking samples within the actual system bandwidth, it is hard to interpolate to the frequency response outside.

On the other hand, the goal is to approximate the frequency response only within the system bandwidth, this time using a redundant dictionary. Therefore one seeks the basis that leads to the smallest approximation error within the signal bandwidth using a limited number of nonzero weighting coefficients. If within this basis there are several possible sparse representations leading to similar approximation errors, the sparse approximation in this basis might be ambiguous, but the same guarantees on the approximation error will hold.

NUMERICAL EXAMPLE

Let us consider the same simple scenario for a numerical study (Fig. 2). The signal $\mathbf{y}$ is approximated using s basis vectors as $\hat{\mathbf{y}}_s$; the corresponding mean squared error (MSE) is $E[|\mathbf{y} - \hat{\mathbf{y}}_s|^2]$. Naturally when using a less sparse approximation (increasing s) the MSE will decrease. For example, for known delays the error will reach zero for $s = P$. In general, how fast the MSE reduces with s will indicate how sparse the corresponding basis can approximate $\mathbf{y}$. While for the baseband model ($\lambda = 1$) there is a trivial way to determine the optimum s-sparse approximation, this is not the case for the redundant dictionaries. Therefore the OMP algorithm is used to find s-sparse approximations, which are not necessarily optimal.

In Fig. 2 the MSE decreases similarly for all cases up to about 10^{-1}, this means that for this multipath channel about 90 percent of the channel energy is concentrated in the ten strongest channel taps. On the other hand, the baseband model will need more than 30 non-zero channel taps to approximate the frequency response with a MSE of 10^{-2}, while using a redundant basis one needs about half that. This points towards an interesting fact, that the baseband channel taps are not approximately sparse in terms of a power-law; if they would follow a power law, the slope in the plot would be constant, while in fact it levels off.

Sparse Channel Estimation in Underwater Acoustic Multicarrier Communications

In this section, we present sparse channel estimation for multicarrier underwater acoustic (UWA) communications as a concrete example of the application of compressive sensing techniques.

Underwater Acoustic Channel

UWA channels are different from radio channels, due to the fundamental differences between acoustic waves and radio waves. For once, the practical bandwidths in UWA channels are limited, due to the absorption of acoustic energy at high frequencies. Also, the speed of sound is only about 1500 m/s in water, while electromagnetic waves propagate at the speed of light in air (3×10^8 m/s). As a result, UWA channels usually have a long delay spread, even in relation to their (low) sampling rate, for example about 20 ms in typical shallow water environments along with a 10 kHz bandwidth, leading to 200 taps in the baseband channel. While channel variations happen at a similar rate to urban radio environments (tens or hundreds of milliseconds), the symbol duration in UWA systems is orders of magnitudes larger than that in radio systems. Also Doppler effects caused by the slow movement of seaborne vessels (or even the sea surface) are magnified by the much lower sound speed.

The key obstacle hindering satisfactory performance in UWA channels is the combination of the long delay spread and (effectively) quick channel variations. This impacts the achievable data rate two-fold:

- Channel estimation has to capture many parameters due to the long delay spread and the estimates have to be frequently updated, consuming a large amount of already scarce frequency/power resources.
- The assumption of a linear time-invariant channel model holds only over a very short time span (on the order of the delay spread), impacting the use of efficient frequency domain equalization or multicarrier systems.

UWA channels therefore can be characterized as doubly (time- and frequency-) selective channels.

As a concrete example, we will consider a multicarrier system, specifically orthogonal frequency division multiplexing (OFDM). The OFDM symbol length needs to be larger than the delay spread to avoid inter-symbol interference (ISI), which in turn make the symbols too long to approximate the channel as fully time-invariant. Although the rate of change is small, the low speed of sound magnifies these changes to result in significant Doppler effects in the received signal. This impairs the orthogonality of the OFDM subcarriers, leading to inter-carrier interference (ICI). When taking samples in the frequency domain, as discussed earlier, every DFT output is now potentially affected by all K transmitted symbols $s[k]$,

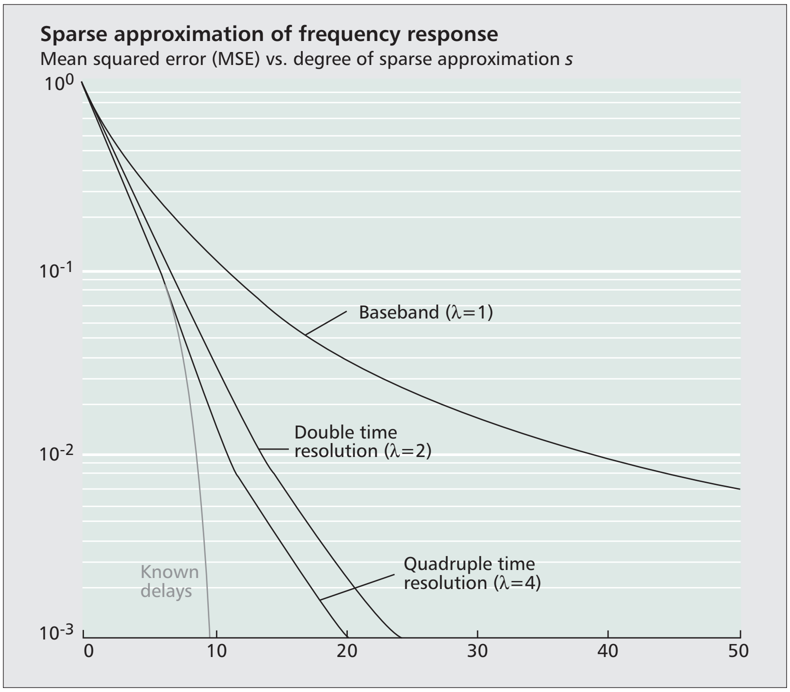

Figure 2. *Comparison of MSE of sparse approximations of* $H(f_k)$ *using s terms with various basis models; all cases lead to better approximation with larger s, but using a redundant basis leads to significantly fewer terms; the parameters are* P = *10,* K = *256.*

$$R(f_m) = \sum_k H(f_m, f_k)s[k] + V(f_m).$$

The ICI coefficients $H(f_m, f_k)$ specify how the $s[k]$ affect $R(f_m)$, and $V(f_m)$ denotes the additive noise. Stacking the $R(f_m)$, $V(f_m)$, and $s[k]$ from all subcarriers into vectors $\mathbf{z}$, $\mathbf{v}$, and $\mathbf{s}$, leads to a matrix-vector formulation as

$$\mathbf{z} = \mathbf{H}\mathbf{s} + \mathbf{v},$$

where $\mathbf{H}$ is the channel mixing matrix.

For a multicarrier system, $\mathbf{H}$ is the channel that needs to be estimated for the purpose of channel equalization and decoding. However, $\mathbf{H}$ contains K^2 entries, much more than the number of measurements in $\mathbf{z}$.

Sparse Representation and Dictionary Construction

To apply compressive sensing techniques to channel estimation in a practical system, one has to find a suitable sparse representation of the channel. This is helped by the unique properties of the UWA channel: Consider two propagation paths that differ by 1.5 meters; the corresponding delay difference is 1 ms, which is already 10 times that of the baseband sampling interval with a 10 kHz bandwidth. Hence, one expects that in the sampled channel impulse response, many entries will be close to zero. This makes UWA channels intuitively sparse [12].

Let us therefore consider a channel model that consists of P discrete paths, similar to the one in Fig. 1, but now with *time-varying* amplitude and delay. The block-by-block receiver in [9] aims to estimate the channel based on each received

Assuming a reasonable SNR, all observations related to data subcarriers are discarded; alternatively a colored noise model could be considered. Furthermore the effect of the unknown data on channel estimation should also be taken under consideration in pilot design.

OFDM symbol individually, such that the receiver is robust to rapid channel changes across OFDM symbols. This motivates the following assumptions on the time variability of the channel:

- The amplitude ξ_p of each path remains approximately constant during each block.
- The path delays vary approximately linear with time, $\tau_p(t) \approx \tau_p - a_p t$ where τ_p is the delay at the start of the block and a_p is the Doppler scale factor. This means that the signal components propagating along the pth path will experience a Doppler shift where frequency f_k will be translated to $(1 + a_p)f_k$.

Now, the channel matrix **H** is characterized by P triplets $\{\xi_p, a_p, \tau_p\}$; see [9] for the exact formulation. However, the exact number of paths P is unknown and the relationship of **H** with a_p and τ_p is nonlinear, complicating the estimation task. Sampling the delay-Doppler plane on a grid, a linear and sparse representation of the channel matrix can be formulated. Specifically:

- The delay dimension is discretized at a multiple of the baseband sampling rate, $\ell T_s/\lambda$, where ℓ can take N_τ values to cover the maximum possible delay.
- The Doppler scale dimension is similarly sampled using N_a values within some interval $|a_p| \le a_{\max}$, with step-size $\Delta a = 2a_{\max}/N_a$.

With this the channel model can be expressed as,

$$h(\tau;t) = \sum_{p=1}^{N_a} \sum_{q=1}^{N_\tau} \xi_{p,q} \delta\Big(\tau - \big(\tau_q - a_p t\big)\Big).$$

The received signal will then be a linear combination of up to $N_\tau N_a$ delayed and Doppler scaled copies of the transmitted signal with complex weights $\xi_{p,q}$.

Now let the vector **x** contain the complex amplitudes of all the $N_\tau N_a$ possible paths on the discretized delay-Doppler plane, of which many entries shall be close to zero. With this **H** is a linear function of **x**, and the channel to be estimated has found a sparse representation in the delay-Doppler domain after a series of approximations. One can write

$$\mathbf{z} = \underbrace{\left(\mathbf{s}^T \otimes \mathbf{I}_K\right)}_{\mathbf{\Phi}^H} \underbrace{\mathrm{vec}\left(\mathbf{H}\right)}_{\mathbf{\Psi x}} + \mathbf{v} = \underbrace{\mathbf{\Phi}^H \mathbf{\Psi}}_{\mathbf{A}} \mathbf{x} + \mathbf{v},$$

which reveals the connection with the compressive sensing formulation presented earlier.

Practical Issues

For the block-by-block multicarrier receiver, the following two facts are not considered in the compressive sensing theory:

- Most elements of the matrix **H** are (generally) negligible in magnitude; specifically, most energy is concentrated on the main diagonal and a few off-diagonals (the magnitude decreases with distance from the main diagonal).
- Only part of the vector **s** is known (the pilots).

Both facts will also affect pilot design.

Structure of Channel Matrix — Since the channel estimation error is determined by the relationship between the sparse estimate and the channel coefficients, $\hat{\mathbf{y}} = \mathbf{\Psi}\hat{\mathbf{x}}$, the estimation error on each element of $\hat{\mathbf{y}}$ is generally of similar variance. This means that on far off-diagonal values of **H**, the estimation error will surely be much larger than the actual value. Therefore, one can reduce the channel estimation error by approximating **H** as a banded matrix with D off-diagonals on each side. This is equivalent to a shorter **y** or removing rows of $\mathbf{\Psi}$. How many off-diagonals to keep will depend on the estimation accuracy of $\hat{\mathbf{x}}$ and on the rate with which the magnitude of the off-diagonal values of **H** decrease.

Influence of Unknown Data — Since the symbols that convey data are unknown to the receiver, one has to treat them as additional noise with a known mean and variance. Therefore the dictionary is constructed by setting all values $s[k] = 0$ if f_k corresponds to a data subcarrier. Due to the known structure of **H**, it is clear that the impact of the noise caused by the unknown data symbol $s[k]$ will be the strongest on the kth entry of **z**. Assuming a reasonable signal-to-noise ratio (SNR), all observations related to data subcarriers are discarded; alternatively a colored noise model could be considered. Furthermore the effect of the unknown data on channel estimation should also be taken under consideration in pilot design.

Pilot Design — As paths with non-zero Doppler scale a_p need to be identified based on their ICI pattern, one needs pilots on adjacent subcarriers. Conversely if one selects pilots adjacent to data symbols, the ICI from these unknown symbols will be stronger. Therefore a random pilot assignment, as would be expected from compressed sensing theory, will very likely be suboptimal due to the specific structure of the dictionary $\mathbf{A} = \mathbf{\Phi}^H\mathbf{\Psi}$. On the other hand, iterative receivers are of great interest, as the data symbols estimated in the previous round can serve as pilot symbols for channel estimation.

Next, we will look at two specific examples of receivers, where the first will be based on negligible time-variation, while the second will assume significant time-variation.

Receiver for Time-Invariant Channels

The general parameters of the considered OFDM system will be the same for both receivers; the bandwidth of 9.8 kHz is centered around f_c = 13kHz, and is divided into K = 1024 subcarriers, leading to a subcarrier spacing of Δf = 9.5 Hz. This leads to an OFDM symbol length of $1/\Delta f$ = 105 ms, during which the channel is approximated as constant, and followed by a guard interval of 25 ms to avoid ISI. The symbol length is chosen as a trade-off between increasing the symbol length to minimize the overhead caused by the guard interval, and a short symbol length ensuring that the assumptions on the channel will hold. Out of the 1024 subcarriers 96 will be null subcarriers, half at the edges of the signal band and half evenly spaced among the data subcarriers. The data subcarriers are modulated using 16-QAM, and each OFDM symbol is separately encoded using a rate-1/2 nonbinary LDPC code. We will use the block error rate (BLER) after LDPC decoding plotted vs. received signal-to-noise ratio (SNR) as the bottom line performance metric throughout.

When assuming that the rate-of-change of the multipath delays is negligible within an OFDM symbol duration, $a_p \triangleq 0$ the channel is simply linear time-invariant, and matches the model in the previous section. The corresponding receiver uses a dictionary consisting only of delays for channel estimation ($N_a = 1$), and since the channel matrix **H** will be diagonal by definition we set $D = 0$, which makes equalization simply a scalar multiplication. To sense the channel 256 pilot subcarriers are evenly distributed between the data (no ICI is expected). Including all overhead, for guard interval, pilots, and coding, the achieved data rate is 10.4 kb/s.

Earlier we saw that overcomplete dictionaries lead to a sparser representation of the multipath channel. We would expect that this will lead to a reduced channel estimation error and in turn a better bit error rate performance. We will use a similar simulation as previously, but now with $P = 15$ paths, where the inter-arrival times are distributed exponentially with mean 1 ms, leading to an average channel delay spread of 15 ms or about 150 baseband channel taps. The amplitudes are Rayleigh distributed with the average power decaying exponentially with the delay. The channel parameters are constant within the duration of one OFDM symbol and independent between symbols.

The simulation results are shown in Fig. 3, where we consider both OMP and BP, as well as a conventional least-squares (LS) channel estimator, which does not take advantage of channel sparsity. We also consider the BLER performance of a receiver with full channel state information (CSI), which can be seen as a bound. First, note that both compressive sensing recovery algorithms gain more than 1 dB over the LS estimator. Second, matching our previous observations that an overcomplete dictionary, $\lambda > 1$, leads to a sparser representation, we see that both OMP and BP benefit and translate this gain into improved BLER performance. We should note that this gain comes at increased computational complexity and further studies revealed a strongly diminishing return for even larger values of λ. We fix $\lambda = 2$ for BP and $\lambda = 4$ for OMP in the following.

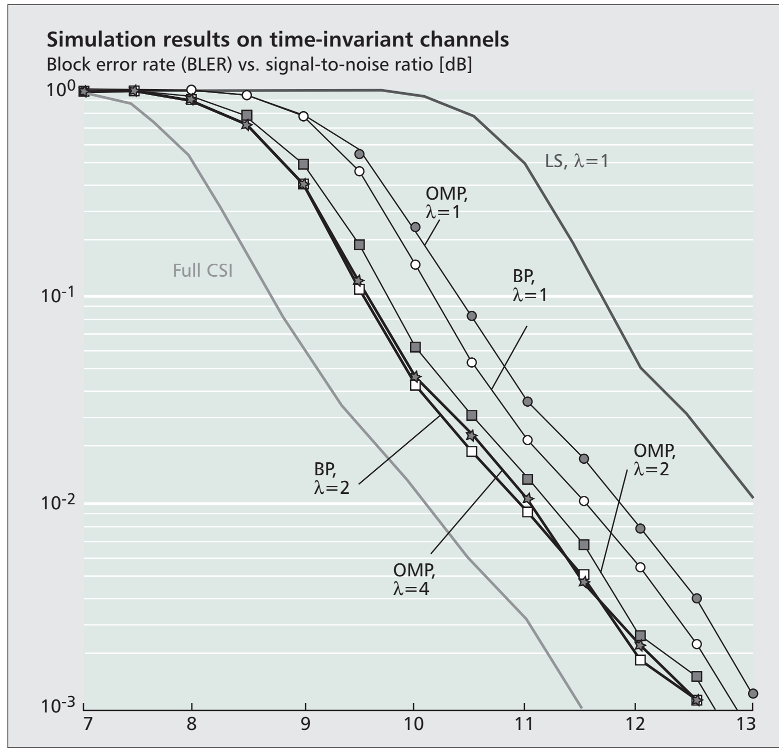

Figure 3. *When considering a coded OFDM system, sparse approximation of the channel frequency response leads to reduced channel estimation error and in turn to improved BLER performance.*

Receiver for Time-Varying Channels

The significant changes for operation on time-varying channels are three-fold: 352 pilots are used that form clusters of four consecutive pilots, channel estimation uses larger dictionaries with $N_a = 15$, and channel equalization involves matrix inversion of a banded matrix with $D > 0$. Due to the increased number of pilots, the data rate is now 7.4 kb/s.

In the simulation, the Doppler rate of each path is drawn from a zero mean uniform distribution, with maximum value $\sqrt{3}\sigma_v/c$, in which σ_v corresponds to the standard deviation of the platform velocity, and c is the sound speed being set to 1500 m/s. We set $\sigma_v = 0.25$ m/s to model significant Doppler spread and keep the other settings of the simulation as before.

We plot simulation results in Fig. 4, where we investigate the trade-off using banded matrices ($D = 1, 3$) and the case where we assume that the channel is time-invariant ($D = 0$). In the latter case we set again $N_a = 1$ and also the conventional LS estimator can be used. When increasing D, there is a point when the performance stops improving because we introduce more channel estimation error than gained by the more precise model. For example in Fig. 4, the performance of BP and OMP are similar for $D = 0$, but when increasing D to account for the ICI, BP considerably outperforms OMP for larger D. We conclude that the break even point for OMP is around $D = 3$ as the gain diminishes quickly, while for BP the estimation error is lower leading to significant gains for $D = 3$ (the break even point is reached at $D = 5$ [9]).

In summary, on significantly time-varying channels receivers that do not account for time-variation ($D = 0$) perform poorly. When estimating also the rate-of-change of the channel delays, we can reconstruct the ICI pattern and use MMSE equalization to suppress it. In this case we need to find a suitable level of modeled ICI by using a banded matrix **H** that has the minimum channel estimation error.

Experimental Validation

We now will use experimental data to validate the simulation results, which was recorded at the Surface Processes and Acoustic Communications Experiment (SPACE08). The experiment was carried out off the coast of Martha's Vineyard, Massachusetts, from October 14 to November 1, 2008. The water depth was about 15 meters. Among the total six receivers, we only consider the data collected by three receivers, labeled as S1, S3, and S5, which were 60 m, 200 m, and 1000 m away from the transmitter respectively, with each receiver array consisting of twelve hydrophones. To show performance differences, we plot the performance based on combining a variable number of phones, as this multiphone combining will increase the effective SNR.

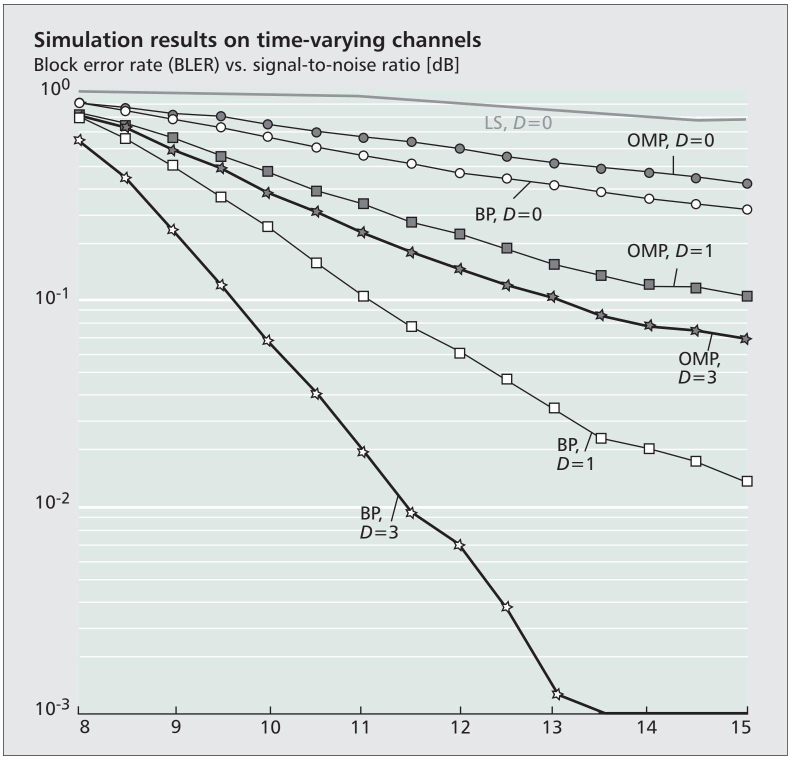

Figure 4. *On time-varying channels, the better estimation accuracy of BP leads to significant gains over OMP for larger* D*;* D = *0 corresponds to a receiver assuming a time-invariant channel.*

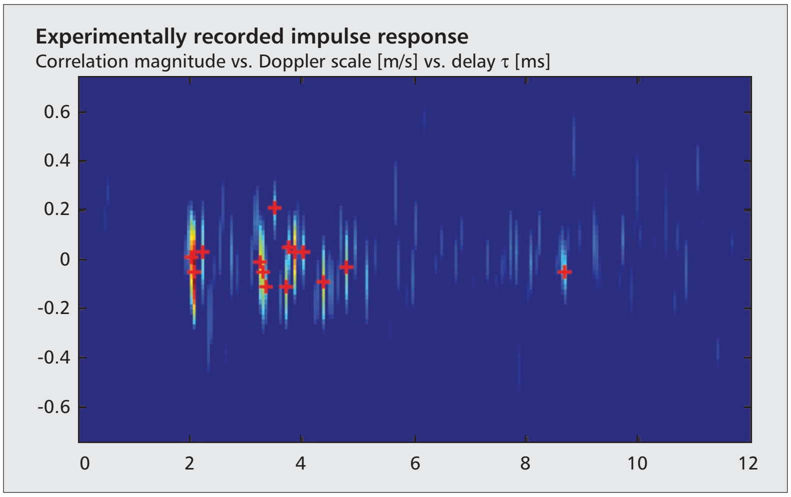

Figure 5. *Example channel from the SPACE08 experiment, where correlation values* $\mathbf{A}^H\mathbf{z}$ *are plotted in a grid according to their delay* τ_q *and Doppler scale* a_p*; we make out about fifteen paths with a delay spread of maybe 10 ms and a Doppler spread of* σ_v *=0.1 m/s.*

We plot a measured channel response in Fig. 5; this plot is based on a matched filter, basically calculating $\mathbf{A}^H\mathbf{z}$ (which is the metric OMP uses to identify non-zero entries). The crosses are simply the strongest peaks and are marked for convenience. We see that there are 10–15 significant peaks, the delay spread is about 10 ms, and the Doppler spread is maybe $\sigma_v = 0.1$ m/s. Also we should note that the correlation between paths of the same delay with different Doppler scales is quite high. In this sense using $N_a = 15$, columns of $\mathbf{A}$ corresponding to the same Doppler scale are even more correlated than columns corresponding to the same delay.

We only report performance results considering the receivers for time-varying channels, more detailed experimental results can be found in [9]. As in the numerical simulation, we also include the conventional LS estimator and versions of OMP and BP with $N_a = 1$ and $D = 0$. The plots include data recorded across six consecutive days (Oct. 22–27), each day a transmission was recorded every two hours, leading to 72 transmissions in total. As each transmission included twenty OFDM symbols, a total of 1440 OFDM symbols are used to calculate the BLER.

Studying the results, see Fig. 6, we see generally the same trends as in our numerical simulation. Note that the differences for nonzero values of D are not as pronounced. This is because the considered interval also included several days of relatively calm weather; see [9] for plots on specific days. Furthermore it is interesting to see that the medium distance of 200 m is the easiest. This is because at the shortest distance the Doppler and delay spread are higher due to the geometry of the reflections at the shallow water bottom, while at the longest distance the received SNR is the weakest.

In summary, the benefits of sparse channel estimation for UWA multicarrier communications are two-fold.

- On (approximately) time-invariant channels, both OMP and BP can reduce the estimation error relative to a conventional LS estimator. Intuitively, the advantage of sparse channel estimation relative to its LS counterpart comes from the fact that by exploiting sparsity in the estimate, sparse channel estimation can effectively reduce the number of unknowns. Therefore a basis that leads to a sparser representation of the channel can further reduce the number of unknowns.
- On a time-varying channel, there are too many unknown channel parameters for a LS estimator to handle with a reasonable amount of pilots. In contrast, using compressive sensing we can identify the relevant parameters and reconstruct a channel matrix with many more unknowns than there are pilots, that is used in equalization. Still, even using compressive sensing it is challenging to estimate the channel with sufficient accuracy, so it can make sense to limit the number of unknowns in the channel matrix using a banded structure. In this case BP seemed to continually outperform OMP.

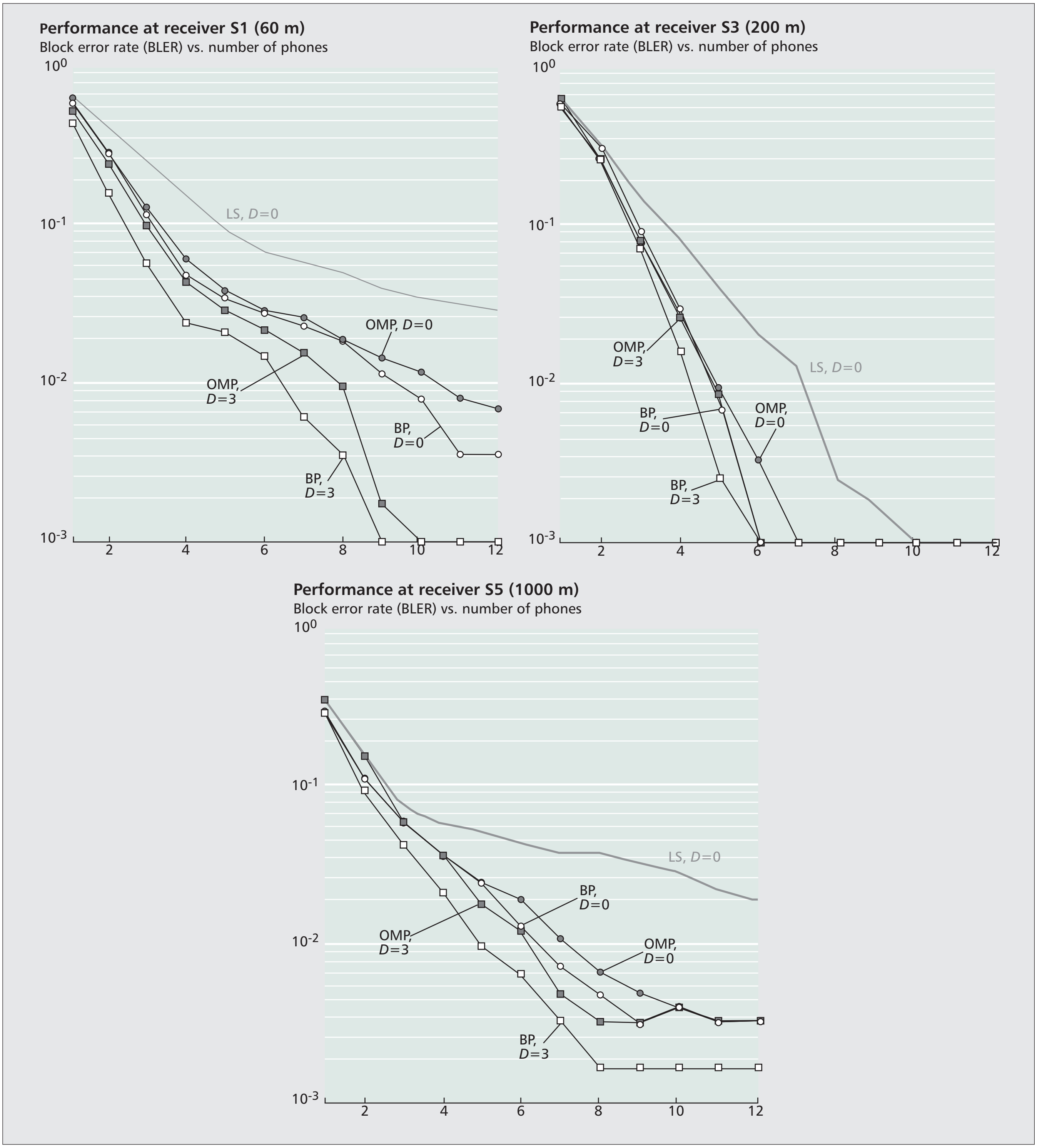

Figure 6. *BLER performance for SPACE08 data; the results are averaged over six consecutive days that include calm weather as well as two storm cycles.*

CONCLUSION

Compressive sensing has made a lasting impression in the signal processing community, where besides an intriguing theory it offers versatile applicability to many challenging problems. In the communications community the application of compressive sensing has been mainly on sparse channel estimation for various types of channels, with extensions to multiuser and cognitive radio systems. In this article, we illustrated the application of the compressive sensing techniques using a concrete example of multicarrier underwater acoustic communications. We showed that an overcomplete dictionary leads to much sparser representation of a multipath channel relative to the baseband tap-based channel model. Numerical simulations and field results demonstrated the substantial benefits of compressive sensing for underwater acoustic

Even using compressive sensing it is challenging to estimate the channel with sufficient accuracy, so it can make sense to limit the number of unknowns in the channel matrix using a banded structure. In this case BP seemed to continually outperform OMP.

communications over long dispersive channels with large Doppler spread.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research and the National Science Foundation.

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information," *IEEE Trans. Info. Theory*, vol. 52, no. 2, Feb. 2006, pp. 489–509.

[2] D. Donoho, "Compressed Sensing," *IEEE Trans. Info Theory*, vol. 52, no. 4, Apr. 2006, pp. 1289–1306.

[3] R. Baraniuk, "Compressive Sensing," *IEEE Signal Process. Mag*., vol. 24, no. 4, July 2007, pp. 118–21.

[4] E. J. Candès and M. B. Wakin, "An Introduction to Compressive Sampling," *IEEE Signal Process. Mag*., vol. 25, no. 2, Mar. 2008, pp. 21–30.

[5] R. Chartrand *et al*., "Introduction to the Issue on Compressive Sensing," *IEEE J. Selected Topics Signal Process*., vol. 4, no. 2, Feb. 2010, pp. 241–43.

[6] J. L. Paredes, G. R. Arce, and Z. Wang, "Ultra-Wideband Compressed Sensing: Channel Estimation," *IEEE J. Selected Topics Signal Process*., vol. 1, no. 3, Oct. 2007, pp. 383–95.

[7] G. Tauböck *et al*., "Compressive Estimation of Doubly Selective Channels in Multicarrier Systems: Leakage Effects and Sparsity-Enhancing Processing," *IEEE J. Selected Topics Signal Process*., vol. 4, no. 2, Apr. 2010, pp. 255–71.

[8] W. U. Bajwa *et al*., "Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels," *Proc. IEEE*, vol. 98, no. 6, June 2010, pp. 1058–76.

[9] C. R. Berger *et al*., "Sparse Channel Estimation for Multicarrier Underwater Acoustic Communication: From Subspace Methods to Compressed Sensing," *IEEE Trans. Signal Process*., vol. 58, no. 3, Mar. 2010, pp. 1708–21.

[10] D. Angelosante *et al*., "Sparsity-Aware Estimation of CDMA System Parameters," *IEEE Wksp. Signal Process. Advances in Wireless Commun*., Perugia, Italy, June 2009.

[11] E. J. Candès and T. Tao, "Decoding by Linear Programming," *IEEE Trans. Info. Theory*, vol. 51, no. 12, Dec. 2005, pp. 4203–15.

[12] M. Stojanovic and J. Preisig, "Underwater Acoustic Communication Channels: Propagation Models and Statistical Characterization," *IEEE Commun. Mag*., vol. 47, no. 1, Jan. 2009, pp. 84–89.

[13] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," *SIAM J. Scientific Comp*., vol. 20, no. 1, 1999, pp. 33–61.

[14] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. Royal Statistical Society*, ser. B, vol. 58, no. 1, 1996v, pp. 267–88.

[15] J. A. Tropp and A. C. Gilbert, "Signal Recovery from Random Measurements via Orthogonal Matching Pursuit," *IEEE Trans. Info. Theory*, vol. 53, no. 12, Dec. 2007, pp. 4655–66.

[16] D. Needell and J. A. Tropp, "CoSaMP: Iterative Signal Recovery from Incomplete and Inaccurate Samples," *Applied Comp. Harmonic Analysis*, vol. 26, 2009, pp. 301–21.

BIOGRAPHIES

CHRISTIAN R. BERGER [S'05, M'09] (crberger@ece.cmu.edu) was born in Heidelberg, Germany, on September 12, 1979. He received his Dipl.-Ing. degree from the Universität Karlsruhe (TH), Karlsruhe, Germany in 2005, and his Ph.D. degree from the University of Connecticut, Storrs, in 2009, both in electrical engineering. In the summer of 2006, he was as a visiting scientist at the Sensor Networks and Data Fusion Department of the FGAN Research Institute, Wachtberg, Germany. He is currently a post-doctoral researcher at the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, USA. His research interests lie in the areas of communications and signal processing, including distributed estimation in wireless sensor networks, wireless positioning and synchronization, and underwater acoustic communications and networking.

ZHAOHUI WANG [S'10] received her B.S. degree in 2006, from the Beijing University of Chemical Technology, and her M.Sc. degree in 2009, from the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, both in signal processing. She is currently working toward her Ph.D. degree in the Department of Electrical and Computer Engineering at the University of Connecticut, Storrs. Her research interests lie in the areas of communications and signal processing, currently focusing on multi-carrier modulation algorithms and signal processing for underwater acoustic communications.

JIANZHONG HUANG [S'09] received his B.S. degree in 2003 and his M.Sc. degree in 2006, from Xidian University, Xi'an, Shaanxi China, both in communication engineering. He is currently working toward his Ph.D. degree in the Department of Electrical and Computer Engineering at the University of Connecticut, Storrs. His research interests lie in the areas of communications and signal processing, currently focusing on multicarrier modulation algorithms and channel coding for underwater acoustic communications.

SHENGLI ZHOU [M'03] received his B.S. degree in 1995 and the M.Sc. degree in 1998, from the University of Science and Technology of China (USTC), Hefei, both in electrical engineering and information science. He received his Ph.D. degree in electrical engineering from the University of Minnesota (UMN), Minneapolis, in 2002. He was an assistant professor with the Department of Electrical and Computer Engineering at the University of Connecticut, Storrs (2003–2009) and now is an associate professor. He holds a United Technologies Corporation (UTC) Professorship in Engineering Innovation (2008–2011). His general research interests lie in the areas of wireless communications and signal processing. His recent focus is on underwater acoustic communications and networking. He has served as an associate editor for *IEEE Transactions on Wireless Communications* (2005–2007) and for *IEEE Transactions on Signal Processing* (2008–2010). He is now an associate editor for *Journal of Oceanic Engineering*. He received the 2007 ONR Young Investigator award and the 2007 Presidential Early Career Award for Scientists and Engineers (PECASE).