Pilot-Assisted Wireless Transmissions

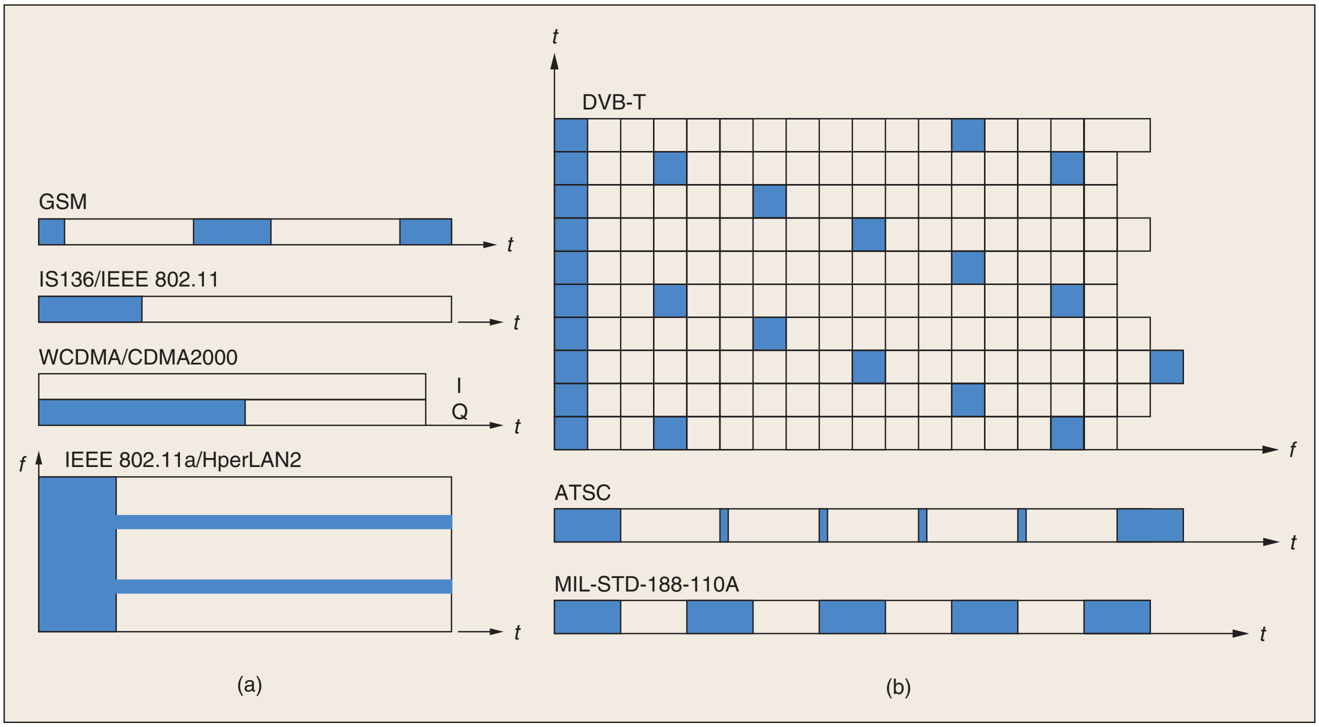
Lang Tong,
Brian M. Sadler,
and Min Dong

General model, design criteria, and signal processing

Pilot-assisted transmission (PAT) multiplexes known symbols with information bearing data. These pilot symbols and the specific multiplexing scheme are known at the receiver and can be exploited for channel estimation, receiver adaptation, and optimal decoding. Some of the earliest studies on PAT focused on fast varying channels [1]–[5]. It was Cavers who coined the now widely used term pilot symbol assisted modulation (PSAM) and presented an analytical approach to the design of PATs [4]. Since then, there has been continuing interest from the signal processing and communications communities in the design of PATs for wireless systems. We use here a slightly broader view of PAT to address issues beyond modulations, especially in topics related to parameter estimation and information theory.

PAT is prevalent in modern communication systems; some of the standardized schemes are illustrated in Figure 1. The GSM system [6], for example, includes 26 pilot bits placed in the middle of every packet along with a small number of starting and tail bits. The North America TDMA standard [7] places pilot symbols at the beginning of each packet. Third-generation systems such as WCDMA [8] and CDMA-2000 [9] transmit pilots and data simultaneously using separate spreading codes. Pilots are also used in broadband systems such as HyperLAN II [10]

1053-5888/04/$20.00©2004IEEE



▲ 1. *Pilot placement patterns in existing wireless systems. (a) Packet transmissions and (b) continuous transmissions. Shaded areas are pilot symbols.*

and by the IEEE 802.11 [11]–[13] family, all considered part of the fourth-generation wireless systems. Wireless broadcast also relies on PAT. The DVB-T [14] inserts different types of pilots in a doubly periodic manner whereas the single carrier ATSC [15] sends pilots in chunks and sprinkles the data stream with synchronization pilots. PAT is common in wire-line systems such as DSL, cable, and voiceband modems.

To practical designers, PAT makes sense: it simplifies the challenging task of receiver design for unknown channels. The presence of pilots also offers the possibility of link recovery from outages. The more frequently pilot symbols are transmitted, the better the estimation and tracking, and the more robust the receiver. The placement of pilot symbols also appears to be straightforward: they should be placed in clusters to avoid interference from unknown data symbols. If the channel varies rapidly, more pilot symbols are needed and transmitted more frequently. For the most part, the design of PAT is based on engineering intuition, practical experience, heuristic analysis, and simulation.

At the theoretical level, however, the design of pilot symbols and the way that they are multiplexed with data are far from trivial. Pilot symbols carry no information about the data; the time spent on sending pilot symbols is time missed for transmitting information. The power allocated to pilots is power taken away from data. Indeed, Shannon theory neither requires nor disqualifies the use of PAT when the channel is unknown. For example, the structure of capacity achieving codes for ergodic block fading channels *excludes* the use of PAT for certain cases [16], [17]. See [19] for a survey of information theoretical approaches to fading channels. From a detection and estimation perspective, blind signal processing techniques provide means for detection and estimation without embedding training into data transmission [18].

Even if PAT is used for practical reasons, there remains a need for optimal design. The amount of pilots, the power allocated for pilot symbols, and the locations of these pilots in the data stream all affect the system performance measured by the reliable transmission rate, bit error rate (BER), or the mean square error of the estimator (by reliable transmission rate we mean the rate for which bit error probability can be made arbitrarily small). Theory and methodology for design of optimal PAT are emerging, although much remains unknown. Do we need optimal PAT? How much gain can one expect from an optimal scheme over an ad hoc but simple one? Is the optimal PAT so complicated that no receiver of reasonable complexity can be implemented to take advantage of its optimal design? Indeed, one may even question the existence of optimal PAT. No clear answers yet exist. In the past few years, however, partial results have emerged, which suggest that significant potential gain can be realized in some cases if pilot symbols are designed optimally and placed judiciously.

Optimal PAT Design: A Framework

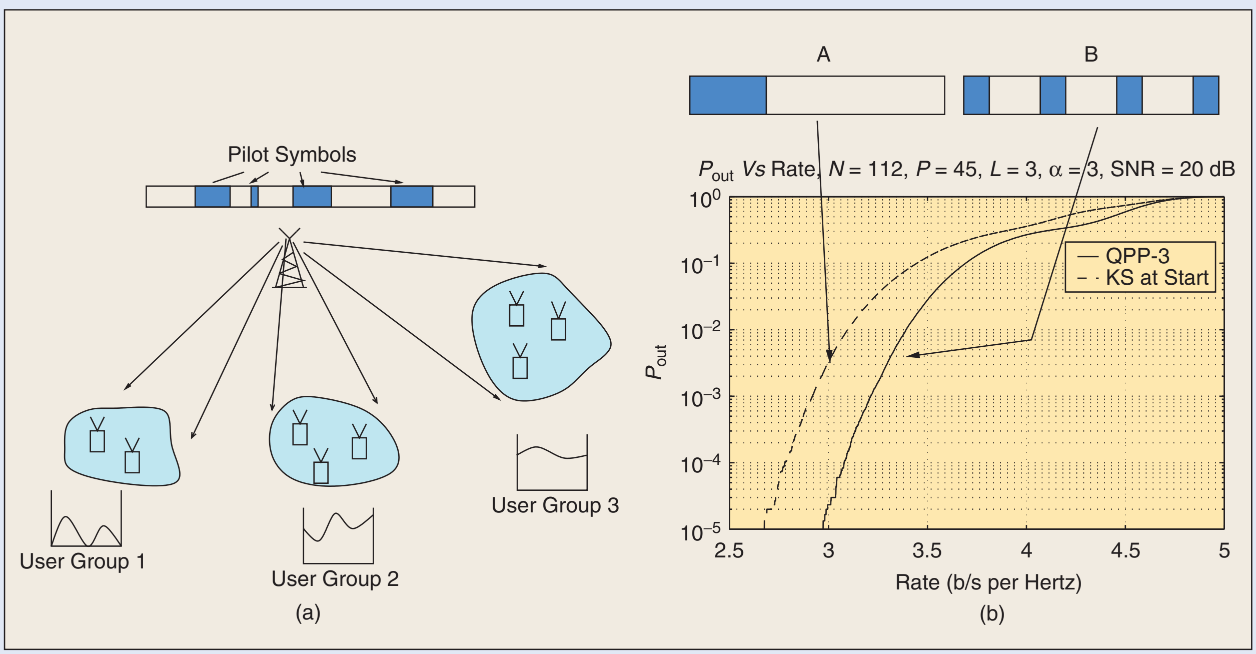
An optimal design depends critically on a carefully chosen model and a well-reasoned criterion. In this section, we lay out the framework on which various optimal PAT schemes are to be developed.

What Do We Gain from an Optimal PAT Scheme?

A telling example is related to broadcasting systems as illustrated in (a) below. The broadcasting station serves users at various geographical locations with different channel conditions. Some users experience little fading and enjoy high signal-to-noise ratios (SNRs), while others suffer from severe intersymbol interference and poor SNR. For broadcasting applications, users may turn on their receivers at random times. Therefore, pilot symbols are needed throughout the transmission so that a new user can acquire the channel state information and gain synchronization. This implies that a fixed percentage of pilot symbols should be embedded in the data stream.

Outage probability is a key performance measure for broadcasting. Given a transmission rate R, an optimal PAT will minimize the probability that a user fails to receive at rate R reliably. Part (b) shows the outage probability versus transmission rate for two different PATs operating at SNR $= 20$ dB. The nonergodic channel model is used for the intersymbol interference channels, with order L. Specifically, the channel between the transmitter and a user is governed by a probability density, and the channel stays constant for the entire use of the channel. The receiver is assumed to belong to one of three different equally likely geographical locations, each has a different multipath structure, i.e., different fading statistics. PAT A is a widely used scheme that puts all pilot symbols at the beginning of the packet. PAT B uniformly spreads the pilot symbols in small clusters of sizes α equal to the delay spread L. We see that PAT B is better than PAT A for all rates. At the rate of 3 b/s/Hz, the outage probability of PAT B is two orders of magnitude smaller than that of PAT A. For a 6 MHz broadcasting channel and at the outage probability level between 10^{-3} and 10^{-5}, the 0.3 b/s/Hz gain of PAT B over PAT A would translate to a gain of 1.8 Mb/s, which is more than enough to add a CD-quality audio channel or ten MP3 broadcasts.

The superiority of PAT B is not a fluke; it can be shown that PAT B is optimal among all placement schemes that have pilot clusters larger than the delay spread of the channel. Most surprisingly, the optimality is uniform across all channel distributions [20]. There are intuitive reasons that PAT B is better than PAT A for intersymbol interference (ISI) channels. Suppose that the ISI channel has two taps, i.e., every symbol is interfered by its predecessor. If the predecessor is a pilot symbol and if the channel taps are known, the interference to the unknown data symbol can be subtracted. If there are as many pilots as data, we can interlace pilots and data so that ISI is completely removed. On the other hand, a standard approach that groups all pilot symbols at the beginning will perform much worse, because only the first data symbol benefits from the knowledge of the pilot.



▲ *Outage probability of a broadcast system. (a) A PAT packet is broadcast to three user groups with different channel distributions. (b) Outage probability versus rate for two different PAT schemes. In PAT B, the number of data and pilot symbols is* N *and* P*, respectively, and* α *is the pilot cluster size. The channel order is* L*. See [20] for details.*

PAT Models

Pilot symbols are traditionally time multiplexed. The use of an antenna array extends the multiplexing to the spatial dimension. Multicarrier transmissions and code division multiple access (CDMA) add frequency and code dimensions to the mix, respectively. Power allocation among data and pilots is another factor. In addition to interleaving training and data by time division multiplexing (TDM), we may consider superimposing pilots and data, an idea proposed as early as 1987 [22] that has also received more recent attention [23]–[29]. The combination of these factors multiplies the possible scenarios to be examined and motivates the model described next.

The Multidimensional PAT Model

The key to unifying various schemes is to view the problem of PAT design as one of power allocation. Specifically, in each design dimension, such as time, frequency, and space, a pair of power allocation parameters are used to model PAT. The simplest case is single carrier transmission over a single input and (possibly) multiple output channel where each transmitted symbol s_t can be modeled as a linear combination of a known pilot s_t^p and an information bearing data symbol s_t^d. Specifically

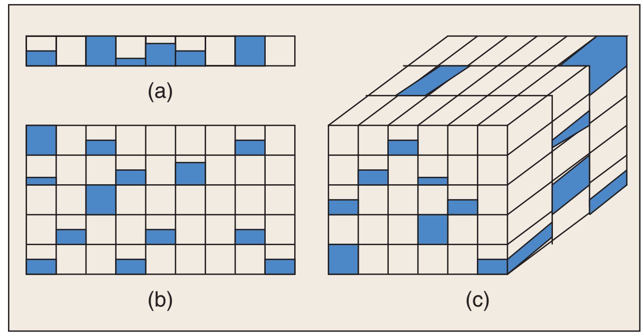
$$s_t = \sqrt{\phi_t} s_t^p + \sqrt{\gamma_t} s_t^d, \quad t = 1, \ldots, B, \tag{1}$$

where s_t^p satisfying $|s_t^p| = 1$ is known with allocated power $\phi_t \geq 0$, and s_t^d is unknown data with zero mean, unit variance, and average power $\gamma_t \geq 0$. For the transmission of a packet of size B, a PAT scheme is defined by the B-dimensional pilot vector $\mathbf{s}_p = [s_t^p]$, and two power allocation vectors $\boldsymbol{\phi} = [\phi_t]$, and $\boldsymbol{\gamma} = [\gamma_t]$. A graphical illustration of the one dimensional PAT scheme is shown in Figure 2(a) where a partially shaded square indicates superimposed pilot and data symbols.

If the spatial domain is added, a two-dimensional description is necessary, as shown in Figure 2(b). A block coded space-time transmission, for example, sends the data symbols in blocks, and each transmitted symbol s_{it} is indexed by the block number t and the position i within the block. If we assume that pilots may be superimposed in any position, we have

$$s_{it} = \sqrt{\phi_{it}} s_{it}^p + \sqrt{\gamma_{it}} s_{it}^d, \\ i = 1, \cdots, N, \quad t = 1, \cdots, B, \tag{2}$$

and the PAT scheme is parameterized by the $N \times B$ pilot matrix $\mathbf{S}_p = [s_{it}^p]$, and nonnegative power allocation matrices $\boldsymbol{\Phi} = [\phi_{it}]$ and $\boldsymbol{\Gamma} = [\gamma_{it}]$. The same formulation naturally applies to orthogonal frequency division multiplexing (OFDM) by treating i as the frequency index. The generalization to higher dimensions is straightforward; the idea is illustrated in Figure 2(c).



▲ 2. An illustration of the multidimensional PAT model (with a per-symbol average power constraint). Each square is a symbol boundary, and the pilot symbol at each square has power proportional to the volume of the shaded region.

Power Constraints

All transmissions are subject to power constraints, and there are many ways such constraints can be imposed on PAT. It is sufficient to consider the two-dimensional case. Given $N \times B$ matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Gamma}$, the average power constraint is given by

$$\frac{1}{NB} \sum_{t=1}^{B} \sum_{i=1}^{N} E\left\{|s_{it}|^2\right\} = \frac{1}{NB} \sum_{t=1}^{B} \sum_{i=1}^{N} (\gamma_{it} + \phi_{it}) = P. \tag{3}$$

As a special case, the per-symbol average power constraint imposes a more stringent condition:

$$E\left\{|s_{it}|^2\right\} = \gamma_{it} + \phi_{it} = P. \tag{4}$$

In this case, power allocation matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Gamma}$ are complementary, and one power allocation matrix is sufficient.

PAT Transceiver Structures

The presence of pilots naturally implies that they will be used at the PAT receiver explicitly or implicitly. Parametric approaches, as illustrated in Figure 3, estimate channel parameters and use the estimated channel for demodulation and decoding. The channel estimator takes the pilot vector $\mathbf{s}_p$ (and possibly the entire observation $\mathbf{y}$), produces a channel estimate $\hat{\boldsymbol{\theta}}$, and feeds the estimate to the decoder. A practical decoder may assume that the estimated channel parameters are perfect. Such an assumption is of course not valid, and the corresponding scheme is referred to as a mismatched decoder [30], [31]. An alternative is to treat the estimated channel parameters as part of the observation. The decoder exploits the joint statistics of $(\hat{\boldsymbol{\theta}}, \mathbf{y})$.

Nonparametric approaches, in contrast, treat the

It was Cavers who coined the now widely used term pilot symbol assisted modulation and presented an analytical approach to the design of PATs.

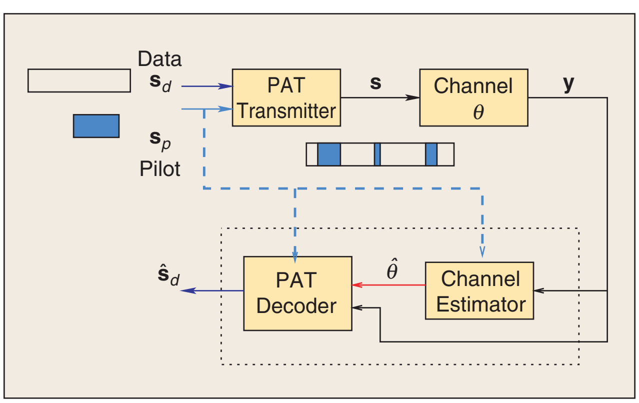
pilot symbols as side information. The channel estimator in Figure 3 is bypassed, and pilots are used to tune the receiver directly. In voiceband modems, for example, explicit channel estimates are not obtained; rather, the training is used for adaptively updating an equalizer.

Figures of Merit

PAT design is primarily a transmitter technique, although receiver characteristics must also be taken into account. Once a PAT scheme is chosen, it may be standardized for a specific application. It is therefore important that a PAT scheme is optimal or near optimal for a wide range of channel conditions. Furthermore, since designers may have different design constraints and objectives, it is preferable that the PAT scheme is optimal for different design criteria. We outline next a few commonly used design criteria.

Information Theoretic Metrics

Reliable transmission at rate R requires the existence of encoding and decoding schemes that make the detection error probability arbitrarily small when the code length is sufficiently long. Shannon capacity measures the maximum rate of reliable transmission among all transceiver designs. The information theoretical metrics for PAT apply to the class of systems constrained to using pilots in specific ways. In other words, we are interested in the PAT design with some fixed transceiver structure (such as that shown in Figure 3) while allowing the design of optimal signaling and codes with long code words.



▲ 3. The structure of PAT transceivers.

Shannon's characterization of the reliable rate of transmission (the achievable rate) is through the use of mutual information [32]. The optimal PAT design that maximizes channel capacity requires expressing mutual information as a function of PAT parameters and maximizing the mutual information with respect to these parameters and the channel input distribution. Unfortunately, the required mutual information expressions are often difficult to obtain. In some cases, however, bounds [33], [34] on the achievable rate can be obtained and optimized with respect to PAT parameters [35]–[37].

For certain fading channels, codes that ensure reliable transmission do not exist. Sometimes, practical constraints on decoding delay make the Shannon capacity zero. In such cases, the outage probability may be used. Specifically, given a transmission rate R, an optimal PAT scheme maximizes the probability that rate R can be achieved reliably.

The channel reliability function, random coding exponent, and cutoff rate [38], [39] are explicit measures that relate detection error probability with data rate and code word length. For PAT systems, they are all functions of PAT parameters. These information theoretic metrics allow us to quantify the decay rate of error probability with respect to the code length.

Channel Estimation: Mean Square Error and Cramér-Rao Bound

The information-theoretic metrics are global measures of a communication system. Often, we are interested in the optimality of specific configurations or system components. For the receiver structure shown in Figure 3, one may be interested in the PAT scheme that minimizes the channel estimator error. A sensible measure is the mean square error (MSE) of the estimator. Because it is desirable that the design of optimal PAT does not depend on the specific algorithm used at the receiver, the Cramér-Rao bound (CRB) is a natural choice as a figure of merit. Specifically, the MSE of an unbiased estimator $\hat{\boldsymbol{\theta}}$, under regularity conditions, is lower bounded by

$$E\left\{\left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right\|^2\right\} \geq \operatorname{trace}\left\{\mathbf{F}^{-1}(\boldsymbol{\Gamma}, \boldsymbol{\Phi}, \mathbf{S}^p)\right\}, \tag{5}$$

where $\mathbf{F}(\boldsymbol{\Gamma}, \boldsymbol{\Phi}, \mathbf{S}^p)$ is the Fisher information matrix [40], [41], and $\mathbf{S}^p$ is the matrix consisting of pilot s_{it}^p. (For deterministic parameters, $\mathbf{F}(\boldsymbol{\Gamma}, \boldsymbol{\Phi}, \mathbf{S}^p)$ is also a function of the unknown parameters. For random parameters, the Fisher information is a function of the (prior) parameter distribution.) Note that, although the CRB can be achieved with finite data samples in some estimation problems, such as in the case of linear system model, the achievability is not always guaranteed with finite data samples. However, the existence of asymptotically efficient (asymptotic efficiency is defined as that the variance of the estimator approaches to the

Information Theoretic Formulation

We illustrate here a simple case of PAT design from the information theoretic point of view, which is germane to many scenarios considered in the literature. Consider the flat fading time-varying channel model

$$y_t = h_t s_t + w_t, \tag{6}$$

where h_t is a stationary ergodic channel process, s_t the channel input, y_t the output, and w_t is additive white Gaussian noise. We assume that a fraction η of the channel resource is used for transmitting pilots using a fraction ξ of the total power P_s. For OFDM with B carriers, this may correspond to ηB carriers being used for pilot tones. For single carrier transmissions, this may correspond to inserting pilots periodically with period $1/\eta$.

With the estimated channel process $\hat{h}_t$ available at the decoder, the model in (6) becomes

$$y_t = \hat{h}_t s_t + z_t, \quad z_t = \tilde{h}_t s_t + w_t \tag{7}$$

where $\tilde{h}_t = h_t - \hat{h}_t$ is the estimation error. The above equation resembles the model when the channel ($\hat{h}$) is known at the receiver. The difference is that noise z_t is not necessarily Gaussian or independent of s_t.

If the decoder takes $\hat{h}_t$ as part of the observation and uses it along with y_t to decode s_t, the capacity of such a scheme is lower bounded by the mutual information $I(\cdot;\cdot)$ with any fixed i.i.d. distribution of s_t

$$\begin{aligned} C &\geq \liminf_{n\to\infty} \frac{1-\eta}{n} \sum_{i=1}^{n} I\left(s_i; y_i, \hat{h}_i\right) \\ &= \liminf_{n\to\infty} \frac{1-\eta}{n} \sum_{i=1}^{n} I\left(s_i; y_i | \hat{h}_i\right), \end{aligned}$$

where the equality is the result of applying the chain rule under the assumption that the data source s_t is independent of the channel estimate $\hat{h}_t$. If we choose s_t to be zero mean Gaussian with variance $(1-\xi)P_s$, i.e., $s_t \sim \mathcal{N}(0, (1-\xi)P_s)$, and if the estimator $\hat{h}_t$ is unbiased conditioned on $\hat{h}_t$, i.e., $E(h_t - \hat{h}_t|\hat{h}_t) = 0$, then the conditional mutual information $I(s_i; y_i|\hat{h}_i)$ can be lower bounded by

$$I(s_i; y_i|\hat{h}_i) \geq E_{\hat{h}_i}\left(\log\left(1 + \frac{(1-\xi)|\hat{h}_i|^2 P_s}{\sigma_w^2 + (1-\xi)P_s E(|h_i - \hat{h}_i|^2|\hat{h}_i)}\right)\right). \tag{8}$$

The proof and a more general bound can be found in [34]. Notice that, conditioned on $\hat{h}_t$, the noise term z_t in (7) has variance

$$E\left(|z_t|^2|\hat{h}_t\right) = (1-\xi)P_s E\left(|h_i - \hat{h}_i|^2|\hat{h}_i\right).$$

Therefore, the lower bound in (8) appears as if the noise term z_t in (7) were independent Gaussian. This comes from the fact that Gaussian noise is the worst uncorrelated additive noise for the Gaussian model [33], [34], [44].

The problem of PAT design can be formulated as maximizing the lower bound on capacity. Let $\mathcal{P}$ denote the set of PAT parameters that specify η, ξ and the placement of pilot symbols. For any conditionally unbiased estimator $\hat{h}_t$, the optimal PAT scheme is given by

$$\sup_{\mathcal{P}} \liminf_{n\to\infty} \frac{1-\eta}{n} \sum_{i=1}^{n} E_{\hat{h}_i}\left(\log\left(1 + \frac{(1-\xi)|\hat{h}_i|^2 P_s}{\sigma_w^2 + (1-\xi)P_s \mathcal{E}_{i,\mathcal{P}}}\right)\right),$$

where we explicitly indicate the dependencies of the channel estimation error $\mathcal{E}_{i,\mathcal{P}} \triangleq E(|h_i - \hat{h}_i|^2|\hat{h}_i; \mathcal{P})$ on the PAT parameters $\mathcal{P}$. If h_i is stationary Gaussian and the MMSE estimator is used, the estimation error $\mathcal{E}_{i,\mathcal{P}}$ is independent of $\hat{h}_i$. In some cases, the MMSE estimator is itself stationary, which leads to a simpler optimization

$$\sup_{\mathcal{P}} (1-\eta) E_{\hat{h}}\left(\log\left(1 + \frac{(1-\xi)|\hat{h}|^2 P_s}{\sigma_w^2 + (1-\xi)P_s \mathcal{E}_{\mathcal{P}}}\right)\right), \tag{9}$$

where $\mathcal{E}_{\mathcal{P}}$ is the channel MMSE for a fixed pilot scheme. The above optimization is performed with respect to the training percentage η, power allocation ξ, and pilot placement that affects the MMSE.

CRB as the number of samples goes to infinity) algorithms (e.g., the maximum-likelihood estimation (MLE) in many cases) justifies the use of CRB as a design criterion. (This contrasts with information theoretic metrics where there is a coding theorem that ensures the achievability of capacity, although the code length may be very long.)

The CRB may be formulated with both random and deterministic parameter models; random models lead to useful insights into ensemble behavior whereas deterministic models provide means for assessing specific realizations of channels and sources. The channel CRB may also be a function of the unknown data transmitted simultaneously with the pilots. When the unknown data are treated as random parameters, they need to be marginalized to obtain the likelihood function. If, on the other hand, these unknown data are treated as deterministic unknown parameters, then the data may be viewed as nuisance parameters that affect the CRB of the channel estimator. Incorporation of known pilots into the bounds yields performance limits for semiblind estimators.

To practical designers, PAT makes sense: it simplifies the challenging task of receiver design for unknown channels.

Source Estimation: BER, Error-Exponent Function, and MSE

For detection, BER (or symbol error rate) is the most appropriate performance metric; it is also one of the most difficult to characterize precisely. A more tractable approach is to use BER bounds as the figure of merit. To that end, Bhattacharyya and random coding bounds [39], [42] can be considered. Also relevant is the error-exponent function that measures the decay rate of the error probability [38]. One can also treat symbol detection as one of parameter estimation, and use the MSE as the metric for optimization. For example, MSE is widely used in the design of equalizers.

A Tour of the Field

In reviewing the literature, we have two fundamental questions in mind. The first is how much training is needed. This problem is not well posed without constraints on data rates. The appropriate figures of merit are of information theoretic nature, where there is a tradeoff between having more training for better estimation and more channel uses for higher rates. The second question deals with how training symbols are multiplexed into the data stream—the problem of optimal placement. Here we can fix the percentage of training and optimize the placement of training symbols. Both information theoretic and detection-estimation measures can be used. With an extensive list of references (see also [21] for full references organized according to a few broad categories), our goal in this section is to provide a few pointers with respect to the two basic questions mentioned previously.

Information Theoretic Approaches to PAT

The use of information theoretic metrics is crucial to reveal tradeoffs among PAT designs. In such settings a PAT scheme provides side information about unknown channels. Along this line, there is extensive literature on reliable communications under channel uncertainty (see [31] and references therein), which provides useful tools for obtaining achievable rate expressions that may be optimized with respect to PAT parameters. Specifically, when channel estimates are available to the receiver, bounds on mutual information [31], [33], [34], [43] are the primary equations used in many of the optimal PAT designs. These expressions, however, do not incorporate the resources required to obtain channel estimates. For the analysis of PAT systems, one must take into account the resources allocated for pilot transmission.

How Much Training Is Necessary?

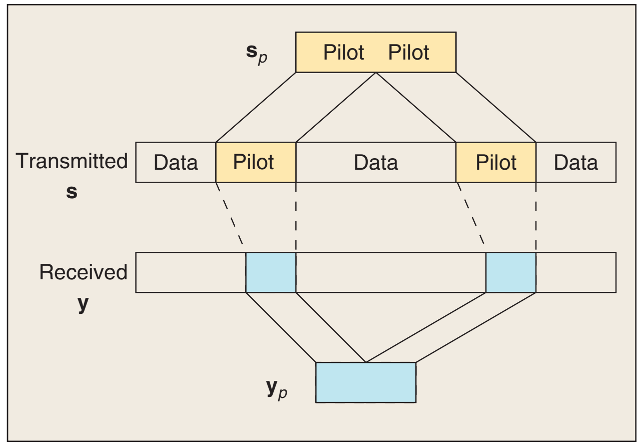
A direct attack on the problem of training design using Shannon capacity as the metric was made by Marzetta [45], and later by Hassibi and Hochwald [36], where the class of ergodic block fading multi-input, multi-output (MIMO) channels were considered. The channel considered there was memoryless, hence the problem of placing pilot symbols uninteresting. Assuming time division multiplexed training, Hassibi and Hochwald maximized a lower bound on channel capacity, a MIMO version of (9), with respect to the number of pilots used in the block and the power allocated to pilots. The work of Hassibi and Hochwald provided several interesting insights. They identified that, in the low SNR regime and when the coherence time (block length) was short, the optimal PAT scheme incurred a substantial penalty; training can lead to bad channel estimates, and no training may be preferable. On the other hand, PAT was close to being optimal in high SNR and long coherence time regimes. This is consistent with the intuition that, with a negligible price paid to obtaining high quality estimates, we can assume that the channel is approximately known at the receiver.

The problem of PAT design for time varying channels was considered in [35] and [46]. Periodic placements of pilot symbol of cluster size one were used, although it was not obvious that one should not group pilot symbols. Mutual information using a binary input was discussed in [46], where a capacity lower bound of the type in (9) was maximized with respect to the percentage of pilot symbols as well as the power allocation in [35]. Cutoff rate is another information theoretic measure that can be used for PAT design. In [47]–[50], cutoff rate was used for optimizing PAT design. Unlike the mutual information measure, the use of cutoff rate leads to a more analytically tractable framework that gives, for some cases, closed-form power allocation [49], [50].

Optimal Placement of Pilot Symbols

When the channel has memory, the positions of the pilot symbols can affect the performance significantly. For intersymbol interference channels, for example, a transmitted symbol is interfered by its predecessors, and the effect of interference from data is different from that from pilots. Our starting point here is to fix the percentage (in power or in the number of channel uses) of pilot symbols and optimize the pilot symbol placement.

Using a capacity lower bound and a version of (9) that incorporates frequency selective fading, the optimal pilot placement for ergodic block frequency-selective fading channel of memory order L for both single carrier and multicarrier transmissions are obtained in [37]. For OFDM transmissions over block ergodic fading channels, pilot symbols are placed in frequency. It is shown in [37] that the optimal placement that maximizes the capacity lower bound is the periodic placement in frequency. For the single carrier system with TDM training, the optimal pilot placement is the so-called quasi-periodic placement. Specifically, given a packet with P pilot symbols and N unknown data symbols, the optimal placement breaks P pilot symbols into as small cluster size as required, and these clusters are placed as evenly in the packet as possible. The minimum cluster size may be dictated by a specific estimation scheme. It turns out that the same strategy also maximizes the outage capacity when the channel fading is nonergodic [51]. The optimal PAT design for time varying channels with memory is, in general, difficult. Under certain conditions, the use of a basis expansion technique can convert time and frequency selective channels to a block fading model that allows the optimization of PAT parameters [52].



▲ 4. *Training-based estimators rely on $\mathbf{y}_p$, while semiblind estimators make use of the entire received block $\mathbf{y}$.*

Signal Processing Perspectives on PAT

We now consider PAT design when there is a fixed allocation of training, focusing on performance measures relating to detection and estimation in PAT systems. There is a vast amount of literature in this area. Typically, the analysis of a specific PAT structure is considered, and algorithms are developed that exploit this particular structure. Readers will find a snapshot of these results in the reference section and [21]. We focus here on a few interference models and their PAT designs, especially on the optimal placement of pilot symbols.

PAT for Block Fading Channels

Block fading channels are time invariant for the coherence time, and then change to a different fading state. This, of course, is an approximation but a reasonable one for many applications. Channel estimation and detection under such models are often performed within each block with the help of pilot symbols.

For the block-fading (quasi-static) model, pilots are used in each packet, and they can be exploited in different ways (see Figure 4). The classic training based estimator, $\hat{h} = F_t(y_p, s_p)$, forms a channel estimate based only on observations of the pilot symbols. A more sophisticated class of estimators, referred to as semiblind, use all available observations (i.e., pilot and data) in forming a channel estimate $\hat{h} = F(y, s_p)$. This makes the unknown data s_d a nuisance parameter that can either be marginalized or estimated jointly with $\hat{h}$.

Similar strategies also apply to detection. A mismatched detector assumes the estimated channel is correct and ignores those observations unrelated to the symbols to be detected. A more sophisticated approach is to view the channel estimates as part of the observation, so that the effective observation vector is $[y, \hat{h}]$. A noncoherent approach may bypass channel estimation, directly using the observation data and the knowledge of pilots.

The apparent first attempts for designing pilot placements optimally were made by Rinne and Renfors [53] and Negi and Cioffi [54] for OFDM in single-input, single-output (SISO) systems. In [54], the authors optimized pilot tone spacing for training-based MMSE estimation. For a channel with order L, and for which $L+1$ tones are selected for training, they showed that selecting pilot tones periodically (in frequency) results in the minimum MSE. Recall that this is precisely the same placement that maximizes the Shannon capacity of an ergodic block fading channel. For multicarrier systems, estimating carrier frequency offset (CFO) is important. It has been shown in [55] that placing equally spaced nulls (in frequency) minimizes the CRB.

When training-based estimators are used for single carrier systems over intersymbol interference channels, the PAT design is simple. All pilots should be clustered into a single block. The only design left is the choice of pilot sequence. Designing the optimal sequence is an old problem dating back to the early 1960s [56], and the work continues for different settings and objectives (see a version of this article with full references [21]). For semiblind channel estimation, on the other hand, the positions of pilot symbols do make a difference as pilots and data are both part of the estimation model. Using the Cramér-Rao bound as the design metric, and under a random channel model, optimal placement of pilot symbols and power allocations are derived for SISO and MIMO intersymbol interference channels in [57]. It is shown that the optimal PAT scheme that minimizes the CRB is independent of the prior channel distribution and the SNR. The placement does depend on the power allocation and the number of pilot symbols in a block.

The CRB has also been employed to study the impact of side information, including training, with MIMO channels in [58] and [59]; see also the discussion in "Beyond PAT: Some Generalizations." Identifiability of the unknown channel(s) may also be

Significant potential gain can be realized in some cases if pilot symbols are designed optimally and placed judiciously.

studied by considering the minimum conditions under which the CRB exists (that is, the minimum conditions under which the Fisher information matrix (FIM) becomes full rank, which is referred to as FIM identifiability). Necessary and sufficient conditions for this to occur are provided in [60] and [61] covering cases from SISO to convolutive MIMO channels. Minimum conditions include the length of the training. For single carrier with cyclic prefix, the single carrier counterpart of OFDM, the idea of replacing the cyclic prefix by a fixed training sequence, referred to as "known symbol padding," is proposed [62]. By inserting a short cluster of training sequence at the beginning of each transmission block to maintain the cyclic structure, such a scheme resembles the PAT scheme and allows frequency domain equalization with low complexity.

Channel estimation using a superimposed periodic pilot sequence was considered in [63]. This idea of exploiting the underlying cyclostationary statistics induced by the periodic training sequence was further explored and analyzed in [29]. Although the superimposed training may not be the optimal approach for block fading frequency-selective channels, periodic superimposed training leads to a low complexity channel estimation algorithm [29]. A superimposed pilot scheme for space-time coded transmission over flat block fading was considered in [25], where the problem setting is general but the optimal placement was not found. The analysis, however, revealed the weakness of superimposed training in the block stationary case, showing that TDM training had lower CRB than that of superimposed training. On the other hand, if training must be included in every block and the channel estimation is accurate, the superimposed scheme gives higher mutual information.

PAT for Fast-Fading Channels

By fast-fading channels we mean that the channel varies within a slot for transmission, or within a packet, or from symbol to symbol. We typically assume that channel variations are highly correlated, at least for a short time, which is consistent with mobile channel measurements [64]. Indeed, if the channel process behaves independently from sample to sample, then training is required for every sample and training placement is not the issue. A practical model is the first order autoregressive (AR) model of the channel process $\{h_t\}$ that leads to a state-space representation

$$h_{t+1} = \alpha h_t + u_t, \quad y_t = h_t s_t + w_t, \tag{10}$$

where α characterizes the fading rate and u_t is the driving noise. When u_t is a white Gaussian sequence, the previous model is also referred to as the Gauss-Markov model. Higher order AR models have also been employed for mobile channel modeling. AR models provide a reasonable fit to the widely used Jake's model that characterizes the power spectral density of the channel process h_t.

In some applications, one must model the channel as doubly selective, i.e., it is both time varying and frequency selective. Both intersymbol interference and fading correlation introduce memory to the model. Thus PAT design will again affect the performance and is also more challenging. A useful technique is to convert the doubly selective channel to either a time or frequency selective model. For example, the idea of a basis expansion model [65] allows us to absorb the time-varying part of the channel into a set of known basis functions and therefore converts a general time varying model to a block frequency-selective fading model or a flat fading model. The number of basis coefficients grows with the packet size, so the method is generally appropriate for short blocks. See, for example, [52].

Single Carrier System

For fast flat-fading channels, when pilots are time-division multiplexed, it is intuitive that pilots should be inserted periodically, although no simple proof exists. For Markov channels with a causal linear MMSE estimator, the optimality of periodic placement of cluster size one was established in [28]. For bandlimited fading models, the optimality of such PAT scheme, though expected, has not been established.

The superimposed pilot scheme was compared with the TDM scheme for time varying flat fading channels in [28], using the Gauss-Markov channel model in (10). Under the same overall power allocation, it was shown that the superimposed scheme performs better for fast fading channels, which confirms the intuition that the constant presence of training in the superimposed scheme has considerable benefit. When the channel is fading slowly and the SNR is high, there is a penalty for using superimposed training because data transmission interferes with channel estimation. An example from [28] is shown in Figure 5, comparing TDM and superimposed training, plotting BER against α. For the superimposed case, a Kalman filter was employed to track the channel and estimate the data; steady state was typically achieved after 20 symbols. Notice that, except for very slow time variation ($\alpha \to 1$), the superimposed scheme is preferred.

OFDM

Pilot placement for fast fading OFDM was considered in [66]–[68]. When the channel experiences symbol-by-symbol variation, the orthogonality of OFDM is destroyed causing intercarrier interference (ICI), complicating channel estimation. In [68], the authors analyzed the effect of intercarrier interference on MIMO-OFDM. They proposed an ICI-mitigating linear filter, as well as a channel estimation and tracking scheme. The authors showed that grouping pilot tones into equally spaced clusters is more effective for time-varying channels than equally spaced pilot tones, while the later is shown to be optimal for time-invariant channels. Garcia et al. considered hexagonal placement in [66], with the intuition that hexagonal placement has the best coverage in the time-frequency plane.

CDMA and Diversity Systems

The pilot channel design for CDMA systems has been studied by many. The optimal power ratio of pilot and data channels was obtained in [69]. Under the optimal power ratio, the loss due to imperfect channel estimation was calculated. In [70], optimal pilot symbol spacing for pilot symbol assisted binary phase shift (BPSK) over Rayleigh fading channels with L diversity paths is obtained. The loss due to imperfect channel estimation

Channel Estimation Under TDM PAT

A widely considered PAT scheme is one in which a single pilot is inserted once every T transmission symbols. Specifically, s_{nT} are pilots for all n, and s_t are data otherwise. The linear minimum mean square error (L-MMSE) is typically used. If the fading process h_t can be modeled as a first order AR process as in (10), a low complexity causal estimator for flat fading channels using all previous pilots is the Kalman predictor based on (10). The L-MMSE estimator, which is also the MMSE estimator when h_t is Gaussian, has the form

$$\hat{h}_{nT} = \hat{h}_{(n-1)T} + K_n\left(y_{nT} - \alpha^T \hat{h}_{(n-1)T} s_{nT}\right) \quad (11)$$

$$\hat{h}_{nT+i} = \alpha^i \hat{h}_{nT} \quad (12)$$

where K_n is the Kalman gain [78]. A closed-form expression for the MSE at each time instance can be obtained. For this model, the further away from a pilot position, the higher the MSE of the estimator, which suggests the possibility of a modulation and coding strategy that takes into account the quality of the channel estimates (see [46] and [79]).

The frequency domain fading models (such as the Jake's model) make the design of a causal channel estimator more difficult, although the noncausal MMSE estimator that uses both future and past pilots is easy to derive (for example, see [4] and [35]).

Let the observation space $\mathcal{Y} = \{y_t\}$ be decomposed into the pilot subspace $\mathcal{Y}_p = \{y_{nT}\}$ and the data subspace $\mathcal{Y}_d = \mathcal{Y} - \mathcal{Y}_p$. Assuming that the data sequence is white and independent of the noise, we see that $\mathcal{Y}_p$ is orthogonal to $\mathcal{Y}_d$, i.e., $E(y_d y_p^*) = 0$ for any $y_d \in \mathcal{Y}_d$ and $y_p \in \mathcal{Y}_p$. Furthermore, the channel process h_t is also orthogonal to the data subspace $\mathcal{Y}_d$. This implies that those samples corresponding to data transmission are not useful in channel estimation and can be discarded. Next, to estimate h_{nT+i}, we only need $\mathcal{Y}_p$, and the standard noncausal Wiener filter in the frequency domain is given by $F_i(\omega) = (S^{(i)}_{h,y_p})/((\omega)S_{y_p}(\omega))$, where $S_{y_p}(\omega)$ is the power spectral density (PSD) of $\{y_{nT}\}$, and $S^{(i)}_{h,y_p}(\omega)$ is the cross spectral density between h_{nT+i} and $\{y_{nT}\}$. The MMSE is given by

$$\mathcal{E}_i = E\left(|h_{nT+i}|^2\right) - \frac{1}{2\pi}\int \frac{\left|S^{(i)}_{h,y_p}(\omega)\right|^2}{S_{y_p}(\omega)}\, d\omega. \quad (13)$$

Note that $S^{(i)}_{h,y_p}(\omega)$ can be related to the PSD of the channel process. The cross-correlation function between $\{h_{nT+i}\}$ and $\{y_{nT}\}$ is given by

$$R_{h_i,y_p}(k) \triangleq E\left(h_{nT+i} y^*_{(n-k)T}\right) = R_h(i + kT)$$

where $R_h(m)$ is the autocorrelation of the channel process, and we assume that all pilots have the same value. Note that $R_{h_i,y_p}(k)$ is a down-sampled channel autocorrelation function (by a factor T). It is interesting to observe, as in [4] and [35], that when the PSD of the channel process has a low bandwidth so that the sampling of $R_h(n)$ causes no aliasing, we then have

$$S^{(i)}_{h,y_p}(\omega) = \frac{1}{T} S_h\left(\frac{\omega}{T}\right) e^{j\frac{\omega}{T}i}.$$

Substituting the above into (13), we conclude that, when the percentage of pilots is high enough to satisfy the Nyquist sampling theorem and the doubly infinite noncausal Wiener filter is used, then there is no need to consider varying coding and modulation strategy at different positions of data transmissions. When the percentage of pilot symbols drops below the Nyquist rate, however, the MMSE of the channel estimator again varies with the data position. Varying modulation and power allocation according to data position becomes an option. Note that there is no obvious reason, from information theoretical viewpoints, that the pilot symbols need to be inserted at a rate higher than the required Nyquist sampling rate, so that the low pilot rate cases should not be discounted.

was also calculated. In [47], the author provided a comprehensive analysis and performance comparison under two training schemes (pilot channel assisted and pilot symbol assisted schemes) in CDMA systems in terms of channel estimation, detection, and cut-off rate. Training parameters were optimized based on Bhattacharyya bounds and cutoff rates. The optimal pilot spacing of PSAM transmission for an LMMSE channel and data estimator over flat Rayleigh fading channels for multiuser CDMA systems was discussed in [71]. The estimation of both time- and frequency-selective channels for single-user CDMA systems was addressed in [72], where a pilot channel was used in parallel to the data channel.

Ultra Wideband

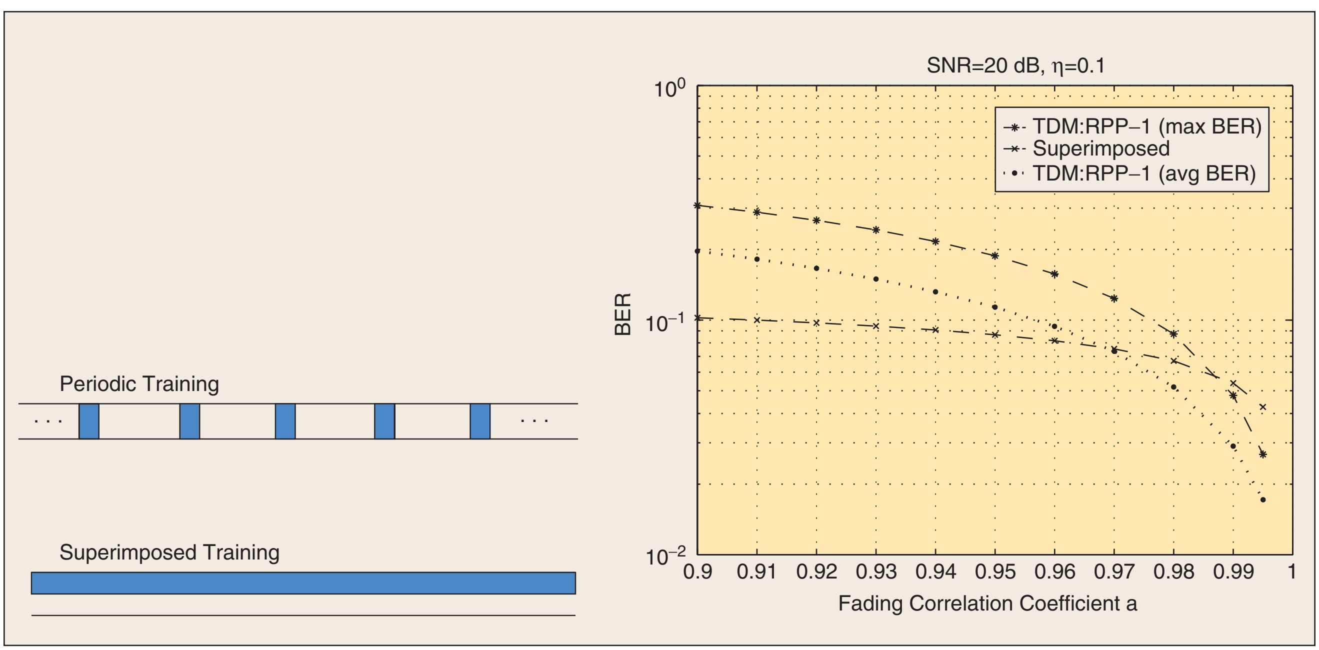
PAT is highly warranted in ultra wideband (UWB) communications, due to the significant challenges of acquisition, synchronization, and equalization in this high bandwidth regime. An interesting way to incorporate training is via a transmitted reference (TR) approach, an old idea that has received attention recently in the UWB context [73]–[75]. TR schemes significantly reduce the receiver complexity but come at a high performance penalty. More generally, optimal training in the UWB context is considered in [76], and the impact of imperfect channel estimates with training in a DS-like UWB approach is analyzed in [77].

Beyond PAT: Some Generalizations

Pilot symbols carry no information. In packet networks, packets have headers that are not completely known, but they are highly protected by error control codes. In applications such as wireless LANs, for example, each user may need to decode the header first and only proceeds to decode the payload if the packet is intended for this user. It is then natural to consider whether the header part can be treated as pilots—an idea that has been used for equalization [80]—and whether the optimal PAT design applies also to the design of the header.

A generalization of PAT design is to consider the transmission of a mixture of low-rate and high-rate sources. While the protocol header can be viewed as a low-rate source, a more interesting case is the mixing of voice transmission with that of data. If the low-rate source is decoded first, it is natural to consider using them as pilots, and the issue of optimal design of packets that contain both low-rate and high-rate sources arises. The new challenge here is to deal with detection errors and analyze how such detection errors affect the overall performance. Given that there are (albeit small) errors in decoding the low rate sources, one may ask whether the PAT strategies apply to the header design. For time varying channels, should the low rate source be periodically inserted in the payload? An initial attempt to this problem is given in [81].

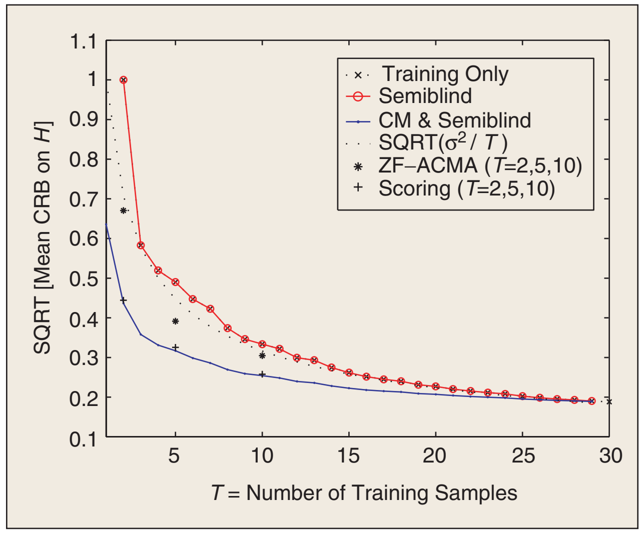
Previously we have noted how training can be viewed as side information and incorporated into semiblind channel estimation and decoding algorithms. Often, additional side information is available at the receiver. Examples include constant modulus signals (or more generally known constellation), known power levels, known angles of arrival, space-time coding, precoding, and others. Exploiting the additional side information can result in significantly enhanced channel estimates and cochannel interference rejection. Alternatively, exploiting the side information can lower the required amount of pilots for a given performance level. The constant modulus (CM) property is particularly powerful in this regard and leads to tractable algorithms. The impact of many forms of side information may be analyzed using the constrained CRB [82], [83]. Examples include MIMO space-time coding



▲ *5. Comparison of TDM and superimposed training with a flat time-varying channel [28].*

[84] and performance of semiblind CM algorithms with cochannel signals [59], [85]. These examples demonstrate the reduced training sizes needed when the CM property is exploited.

An example from [59] is shown in Figure 6, depicting channel estimation performance against training size (block fading, block size = 30, two sensors, two 8-PSK sources with 15° spatial separation, and source SNRs of 15 and 20 dB, respectively). CRBs are shown for training-only, semiblind, and CM semiblind channel estimation. Training-only and semiblind bounds are coincident; in this scenario more sensors than sources are required to obtain improvement with training [86]. The addition of the CM constraint relaxes the amount of training needed. Performance of two algorithms are also shown. First, blind estimation based on a zero-forcing version of ACMA [87], and second, a semiblind CM algorithm based on scoring with ZF-ACMA initialization. The scoring algorithm exploits the training and data (semiblind) as well as the CM property, and comes very close to the CM semiblind CRB.



▲ 6. *Channel estimation bounds and algorithm performance with two constant modulus sources, plotted against number of pilots in a packet of length 30 [59].*

Conclusions

In this article we have presented an overview of PAT. A general PAT model was given, and common design criteria have been reviewed. Also, information theoretic and signal processing issues have been discussed. The optimal design of PAT is application specific and is often dominated by implementation concerns. It is fruitful in pilot designs, however, that the performance limits be part of the design process, with joint consideration of transmitter and receiver issues.

A number of new applications call for careful PAT design. Of particular interest is the pilot design in an asynchronous networking environment, where packets collide and the pilot symbols in the packet may be destroyed [88]. The design of PAT is therefore coupled with the medium access control that requires cross layer considerations. Emerging applications in wireless LANs, ultra wideband communications, and sensor networks all require some form of PAT design.

Acknowledgments

The authors are grateful to S. Adireddy, C. Budianu, Y. Sung, and Z. Yang for their help in the literature search. A. Swami and S. Misra provided valuable detailed comments. This work was supported in part by the Multidisciplinary University Research Initiative (MURI) under the Office of Naval Research Contract N00014-00-1-0564, Army Research Laboratory CTA on Communication and Networks under Grant DAAD19-01-2-0011, and National Science Foundation under Contract CCR-0311055.

Lang Tong is a professor in the School of Electrical and Computer Engineering, Cornell University, Ithaca, New York. He received the B.E. degree from Tsinghua University, Beijing, China, in 1985 and the M.S. and Ph.D. degrees in electrical engineering in 1987 and 1990, respectively, from the University of Notre Dame, Indiana. He was a postdoctoral research affiliate at the Information Systems Laboratory, Stanford University, in 1991. He was also the 2001 Cor Wit Visiting Professor at the Delft University of Technology. He received the Young Investigator Award from the Office of Naval Research in 1996 and the Outstanding Young Author Award from the IEEE Circuits and Systems Society. His areas of interest include statistical signal processing, wireless communications, communication networks and sensor networks, and information theory. He is a Senior Member of the IEEE.

Brian M. Sadler is a senior research scientist at the Army Research Laboratory (ARL) in Adelphi, Maryland, and has been lecturing at Johns Hopkins University since 1994 on statistical signal processing and communications. He was an associate editor for *IEEE Transactions on Signal Processing*, is on the editorial board for the *EURASIP Journal on Wireless Communications and Networking*, and is a guest editor for the *IEEE Journal on Selected Areas of Communications* special issue on military communications. He is a member of the IEEE Technical Committee on Signal Processing for Communications and cochaired the Second IEEE Workshop on Signal Processing Advances in Wireless Communications.

Min Dong received the B.Eng. degree in 1998 in automation from Tsinghua University, Beijing, China, and the M.S. and Ph.D. degree in 2003 and 2004, respectively, in electrical and computer engineering, from Cornell University, Ithaca, New York. She is currently with the Technology & Venture Group, Qualcomm Inc., San Diego, CA. Her research interests include statistical signal processing, wireless communications, and communication networks.

References

[1] J.H. Lodge and M.L. Moher, "Time diversity for mobile satellite channels using trellis coded modulations," in *Proc. IEEE Global Telecommunications,* Tokyo, vol. 3, 1987.

[2] M.L. Moher and J.H. Lodge, "TCMP-a modulation and coding strategy for Rician fading channels," *IEEE J. Select. Areas Commun.,* vol. 7, pp. 1347–1355, 1989.

[3] S. Sampei and T. Sunaga, "Rayleigh fading compensation method for 16QAM in digital land mobile radio channels," in *Proc. IEEE Veh. Technol. Conf.*, San Francisco, CA, 1989, pp. 640–646.

[4] J.K. Cavers, "An analysis of pilot symbol assisted modulation for Rayleigh fading channels," *IEEE Trans. Veh. Technol.*, vol. 40, no. 4, pp. 686–693, 1991.

[5] J.K. Cavers, "Pilot symbol assisted modulation and differential detection in fading and delay spread," *IEEE Trans. Commun.*, vol. 43, no. 7, pp. 2206–2212, 1995.

[6] *Group Speciale Mobile (GSM) Recommendations,* GSM Series 01–12, 1990.

[7] TIA/EIA, *TIA/EIA/IS-136.1:TDMA Cellular/PCS—Radio Interface—Mobile Station—Base Station Compatibility—Digital Control Channel.* Englewood Cliffs, NJ: Prentice Hall, 1997.

[8] *Physical channels and mapping of transport channels onto physical channels (FDD)*, WCDMA, 2001.

[9] *Physical Layer Standard for CDMA2000 Spread Spectrum Systems,* CDMA2000, 2001.

[10] *Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Physical (PHY) Layer*, HIPERLAN II, 2001.

[11] *ANSI/IEEE Std 802.11, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications,* IEEE802.11, 1999.

[12] *IEEE Std 802.11a-1999, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High-speed Physical Layer in the 5 GHz Band,* IEEE802.11a, 1999.

[13] *IEEE Std 802.11b-1999, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-speed Physical Layer Extension in the 2.4 GHz Band,* IEEE802.11b, 1999.

[14] *Digital Video Broadcasting (DVB); Framing structure, channel coding and modulation for digital terrestrial television,* DVB-T, 2001.

[15] ATSC, *ATSC Digital Television Standard (Revision B),* Doc. A/53B. 2001.

[16] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, 1999.

[17] I. Abou-Faycal, M. Trott, and S. Shamai, "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1290–1301, 2001.

[18] L. Tong and S. Perreau, "Multichannel blind channel estimation: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, 1998.

[19] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, 1998.

[20] S. Adireddy and L. Tong, "Optimal placement of known symbols for non-ergodic broadcast channels," in *Proc. CISS 2002*, Princeton, NJ, 2002.

[21] L. Tong, B. Sadler, and M. Dong, "Pilot assisted wireless transmissions—References," [Online]. Available: http://acsp. ece.cornell.edu/pubJ.html

[22] D. Makrakis and K. Feher, "A novel pilot insertion-extraction technique based on spread spectrum affine precoders for reliable communications," presented at *Miami Technicon*, (Miami), 1987.

[23] P. Hoeher and F. Tufvesson, "Channel estimation with superimposed pilot sequence," in *Proc. GLOBECOM 99*, Rio de Janeiro, Brazil, 1999, vol. 4, pp. 2162–2166.

[24] J. Manton, I. Mareels, and Y. Hua, "Affine precoders for reliable communications," in *Proc. ICASSP*, Istanbul, Turkey, 2000, vol. 5, pp. 2749–2752.

[25] C. Budianu and L. Tong, "Channel estimation for space-time block coding systems," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2515–2528, 2002.

[26] S. Ohno and G.B. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless OFDM," *IEEE Trans. Commun.*, vol. 50, no. 10, pp. 2113–2123, 2002.

[27] A.-J. van der Veen and L. Tong, "Packet separation in wireless ad-hoc networks by Known Modulus Algorithms," in *Proc. ICASSP 2002*, Orlando, FL, 2002, pp. 2149–2152.

[28] M. Dong, L. Tong, and B. Sadler, "Optimal placement of training over time-varying channels," *IEEE Trans. Signal Processing*, vol. 52, no. 5, pp. 1403–1418, 2004.

[29] G.T. Zhou, M. Viberg, and T. McKelvey, "A first-order statistical method for channel estimation," *IEEE Signal Processing Lett.*, vol. 10, no. 3, pp. 57–60, 2003.

[30] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 1953–1967, 1994.

[31] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2148–2177, 1998.

[32] T. Cover and J. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[33] M. Medard, "The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel," *IEEE Trans. Inform. Theory*, vol. 46, no. 3, pp. 933–946, 2000.

[34] A. Lapidoth and S. Shamai, "Fading channels: how perfect need 'perfect side information' be," *IEEE Trans. Inform. Theory*, vol. 48, no. 5, pp. 1118–1134, 2002.

[35] J. Baltersee, G. Fock, and H. Meyr, "An information theoretic foundation of synchronized detection," *IEEE Trans. Commun.*, vol. 49, no. 12, pp. 2115–2123, 2001.

[36] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links," *IEEE Trans. Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.

[37] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of known symbols for frequency-selective block-fading channels," *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 2338–2353, 2002.

[38] R.G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[39] A. Viterbi and J. Omura, *Principles of Digital Communication and Coding.* New York: McGraw-Hill, 1979.

[40] H.V. Trees, *Detection, Estimation and Modulation Theory*, vol. 1. New York: Wiley, 1968.

[41] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory.* Englewood Cliffs, NJ: Prentice Hall, 1993.

[42] R. Blahut, *Principles and Practice of Information Theory.* Reading, MA: Addison-Wesley, 1987.

[43] G. Caire and S. Shamai, "On the capacity of some channels with channel state information," *IEEE Trans. Inform. Theory*, vol. 45, no. 6, pp. 2007–2019, 1999.

[44] S. Diggavi and T. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 3072–3081, 2001.

[45] T. Marzetta, "BLAST training: Estimating channel characteristics for high capacity space-Time wireless," in *Proc. 37th Annu. Allerton Conf. Communication, Control and Computing.*, Monticello, IL, 1999.

[46] M. Medard, I. Abou-Faycal, and U. Madhow, "Adaptive coding with pilot signals," in *Proc. 38th Annual Allerton Conf. Communication, Control and Computing.*, Allerton, IL, 2000.

[47] F. Ling, "Optimal reception, performance bound, and cutoff rate analysis of reference-assisted coherent CDMA communications with applications," *IEEE Trans.Commun.*, vol. 47, no. 10, pp. 1583–1592, 1999.

[48] W. Phoel and M. Honig, "Performance of coded DS-CDMA with pilot-assisted channel estimation and linear interference suppression," *IEEE Trans. Commun.*, vol. 50, no. 5, pp. 822–832, 2002.

[49] S. Misra, A. Swami, and L. Tong, "Cutoff rate analysis of the Gauss-Markov fading channel with binary inputs and partial CSI at the receiver," in *Proc. 2003 Conf. Info. Sci. and Systems*, Baltimore, MD, 2003.

[50] S. Misra, A. Swami, and L. Tong, "Cutoff rate analysis of the Gauss-Markov fading channel with adaptive energy allocation," in *Proc. 4rd IEEE Signal Processing Workshop Signal Processing Advances in Wireless Communication*, Rome, Italy, 2003, pp. 388–392.

[51] S. Adireddy and L. Tong, "Optimal placement of known symbols for slowly varying frequency selective channels," *IEEE Trans. Wireless Commun.,* to be published.

[52] X. Ma, G. Giannakis, and S. Ohno, "Optimal training for block transmission of doubly selective wireless fading channels," *IEEE Trans. Signal Processing*, vol. 51, no. 5, pp. 1351–1366, 2003.

[53] J. Rinne and M. Renfors, "Pilot spacing in orthogonal frequency division multiplexing systems," *IEEE Trans. Consumer Electron.*, vol. 42, no. 4, pp. 959–962, 1996.

[54] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM System," *IEEE Trans. Consumer Electron.*, vol. 44, no. 3, pp. 1122–1128, 1998.

[55] M. Ghogho, A. Swami, and G. Giannakis, "Optimized null-subcarrier selection for CFO estimation in OFDM over frequency-selective fading channels," in *Proc. IEEE Global Telecommunications Conf.*, 2001, vol. 1, San Antonio, TX, pp. 25–29, 2001.

[56] R.L. Frank, "Phase shift pulse codes with good periodic correlation properties," *IRE Trans. Inform. Theory*, vol. 8, pp. 381–382, 1962.

[57] M. Dong and L. Tong, "Optimal design and placement of pilot symbols for channel estimation," *IEEE Trans. Signal Processing*, vol. 50, no. 12, pp. 3055–3069, 2002.

[58] B.M. Sadler, R.J. Kozick, and T. Moore, "Bounds on MIMO channel estimation and equalization with side information," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, 2001*, vol. 4, Salt Lake City, UT, 2001, pp. 2145–2148.

[59] B.M. Sadler, R.J. Kozick, and T. Moore, "Performance of MIMO: CM and semi-blind cases," in *Proc. IV IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, Rome, Italy, 2003, pp. 309–313.

[60] T.J. Moore, B.M. Sadler, and R.J. Kozick, "Regularity and strict identifiability in MIMO systems," *IEEE Trans. Signal Processing*, vol. 50, no. 8, pp. 1831–1842, 2002.

[61] T.J. Moore and B.M. Sadler, "Sufficient conditions for regularity and strict identifiability in MIMO systems," *IEEE Trans. Signal Processing*, vol. 52, no. 9, pp. 2650–2655, 2004.

[62] L. Deneire, B. Gyselinckx, and M. Engels, "Training sequence versus cyclic prefix—a new look on signal carrier communication," *IEEE Commun. Letters*, vol. 5, no. 7, pp. 292–294, 2001.

[63] F. Mazzenga, "Channel estimation and equalization for M-QAM transmission with a hidden pilot sequence," *IEEE Trans. Broadcasting*, vol. 46, no. 2, pp. 170–176, 2000.

[64] W.C. Jakes, Jr., *Microwave Mobile Communications.* New York: Wiley, 1974.

[65] G. Giannakis and C. Tepedelenlioglu, "Basis expansion models and diversity techniques for blind equalization of time-varying channels," *Proc. IEEE*, vol. 86, no. 10, pp. 1969–1986, 1998.

[66] M.J. Fernandez-Getino Garcia, J.M. Paez Borrallo, and S. Zazo Bello, "Novel pilot patterns for channel estimation in OFDM mobile systems over frequency selective fading channels," in *Proc. 10th Int. Symp. Personal, Indoor and Mobile Radio Communications, PIMRC'99*, Owsaka, Japan, 1999, pp. 363–367.

[67] M. Dong, L. Tong, and B.M. Sadler, "Optimal pilot placement for channel tracking in OFDM," in *Proc. MILCOM*, Anaheim, CA, 2002, pp. 602–606.

[68] A. Stamoulis, S. Diggavi, and N. Al-Dhahir, "Intercarrier interference in MIMO OFDM," *IEEE Trans. Signal Processing*, vol. 50, no. 10, pp. 2451–2464, 2002.

[69] P. Schramm, "Analysis and optimization of pilot-channel-assisted BPSK for DS-CDMA systems," *IEEE Trans. Commun.*, vol. 46, no. 9, pp. 1122–1124, 1998.

[70] P. Schramm and R. Muller, "Pilot symbol assisted BPSK on Rayleigh fading channels with diversity: Performance analysis and parameter optimization," *IEEE Trans. Commun.*, vol. 46, no. 12, pp. 1560–1563, 1998.

[71] J. Evans, "Optimal resource allocation for pilot symbol aided multiuser receivers in Rayleigh faded CDMA channels," *IEEE Trans. Commun.*, vol. 50, no. 8, pp. 1316–1325, 2002.

[72] M.-A. Baissas and A. Sayeed, "Pilot-based estimation of time-varying multipath channels for coherent CDMA receivers," *IEEE Trans. Signal Processing*, vol. 50, no. 8, pp. 2037–2049, 2002.

[73] R. Hoctor and H. Tomlinson, "Delay-hopped transmitted-reference RF communications," in *Proc. IEEE Conf. Ultra Wideband Systems and Technologies, Digest of Papers*, 2002, pp. 265–269.

[74] J.D. Choi and W.E. Stark, "Performance of ultra-wideband communications with suboptimal receivers in multipath channels," *IEEE J. Select. Areas Commun.*, vol. 20, no. 9, pp. 1754–1766, 2002.

[75] S. Hoyos, B.M. Sadler, and G. Arce, "Mono-bit digital receivers for ultra-wideband communications," *IEEE Trans. Wireless Commun.,* to be published.

[76] L. Yang and G. Giannakis, "Optimal pilot waveform assisted modulation for ultra-wideband communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1236–1249, 2004.

[77] B.M. Sadler and A. Swami, "On the performance of episodic UWB and direct-sequence communication systems," *IEEE Trans. Wireless Commun.*, to be published.

[78] T. Kailath, A.H. Sayed, and B. Hassibi, *Linear Estimation.* Englewood Cliffs, NJ: Prentice Hall, 2000.

[79] D.L. Goeckel, "Adaptive coding for time-varying channels using outdated fading estimates," *IEEE Trans. Commun.*, vol. 47, no 6., pp. 844–855, 1999.

[80] Q. Bao and L. Tong, "Protocal-aided channel equalization in wireless ATM," *IEEE J. Select. Areas Comm.*, vol. 18, no. 3, pp. 418–435, 2000.

[81] Z. Yang and L. Tong, "Exploiting protocol information for transmission over unknown fading channels," in *Proc. 2003 Intl. Symp. Inform. Theory*, 2003, p. 388.

[82] P. Stoica and B. Ng, "On the Cramer-Rao bound under parametric constraints," *IEEE Signal Processing Lett.*, vol. 5, no. 7, pp. 177–179, 1998.

[83] B. Sadler, R. Kozick, and T. Moore, "Bounds on bearing and symbol estimation with side information," *IEEE Trans. Signal Processing*, vol. 49, no. 4, pp. 822–834, 2001.

[84] B.M. Sadler and R.J. Kozick, "Bounds on uncalibrated array signal processing," in *Proc. Tenth IEEE Workshop Statistical Signal and Array Processing*, 2000, pp. 73–77.

[85] B.M. Sadler, R.J. Kozick, and T. Moore, "On the performance of source separation with constant modulus signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, 2002, pp. III-2373–III-2376.

[86] R.J. Kozick, B.M. Sadler, and T.J. Moore, "Bounds on MIMO estimation with CM and semi-blind signaling," in *Proc. 37th Conf. Information Sciences and Systems (CISS)*, Baltimore, MD, 2003.

[87] A.-J. van der Veen, "Asymptotic properties of the algebraic constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 49, no. 8, pp. 1796–1807, 2001.

[88] C. Budianu and L. Tong, "Channel estimation under asynchronous packet interference," *IEEE Trans. Signal Processing*, to be published.