Spreading Signals in the Wideband Limit

Elchanan Zwecher and Dana Porrat
The Hebrew University of Jerusalem
Jerusalem, Israel 91904
dana.porrat@huji.ac.il, elchanzw@cs.huji.ac.il

***Abstract*— Wideband communications are asymptotically impossible with signals that are spread over a very wide band and are transmitted over a multipath channel unknown ahead of time, if the signal to noise ratio (snr) during communications is too low. This work exploits the recently discovered connection between mutual information and minimum mean square error (I-mmse) to bound the achievable data-rate of spreading signals in wideband settings, and to conclude that the achievable data-rate diminishes as the bandwidth increases due to channel uncertainty if the signal to noise ratio during active transmission is too low. The result applies to all spreading modulations, i.e. signals that are spread almost uniformly over the bandwidth available to the communication system, with snr during active transmission that is $o\left(\frac{\log(W/L)}{W/L}\right)$. The result holds for communications over channels where the number of paths L is unbounded but sub-linear in the bandwidth W.**

I. Introduction and Discussion

This work analyzes the performance of wideband communication systems where the signal to noise ratio (snr) per degree of freedom is low. Previous work [1], [2], [3] investigated systems employing signals that are spread over both time and frequency, i.e. transmit continuously in time and spread their power over the entire bandwidth available to them. This type of signaling is not effective in the limit of large bandwidth, under the assumptions of limited power and a multipath channel with a large number of components. In fact, if the number of effective multipath components increases without bound as the bandwidth increases, then the achievable datarate converges to zero [1].

A *degree of freedom* of the channel holds a small time-frequency block, of the order of $1/W$ time units over $1/t_c$ frequency units, where W is the system bandwidth and t_c the coherence time of the channel. In the setting analyzed by [1], i.e. power limited systems with signals that are spread over the entire time and bandwidth available to them, the signal to noise ratio snr per degree of freedom is inverse to the bandwidth, i.e. snr $= \theta\left(1/W\right)$. Our result generalizes by providing a range of snr where the datarate diminishes in the limit. We show that if burstiness is used, the snr during active bursts must not be too low, i.e.

$$o\left(\frac{\log W/L}{W/L}\right) \quad (1)$$

because in this regime communications are not possible in the limit.

Two wideband systems that are peaky (bursty) over time and employ a duty cycle mechanism were investigated in [4]. Although the setting of [4] maintained an average power limitation, the snr per degree of freedom during active transmission was increased by the usage of a duty cycle mechanism, that concentrated the signal's energy over short periods of time. The snr per degree of freedom during active bursts of transmission, that enables datarates of DSSS systems to approach capacity in the wideband limit was lower bounded in [4] by

$$\text{snr} = \omega\left(\frac{\log W}{W/L}\right) \quad (2)$$

and snr $= o(1)$, i.e. snr $\xrightarrow[W\to\infty]{} 0$; L is the number of apparent multipath components. Our new results complements [4] by showing that snr values during active bursts that are only slightly lower than (2) prevent communications in the limit.

The 'low snr regime' includes all signal to noise ratios that approach zero as the bandwidth increases. The new result basically determines that low snr does not ensure efficient communications. In systems that do not possess knowledge of the multipath channel realization, i.e. *incoherent systems,* the low snr regime is divided as seen in Figure 1

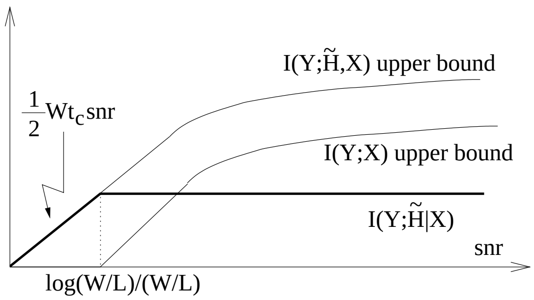

Fig. 1. A sketch of the mutual information – snr relationship for spread signals for a fixed but large bandwidth and diminishing snr. The coherent datarate upper bound (top graph) is linear in the low snr regime, and concave $\cap$. The channel uncertainty penalty (in a bold line) consumes the entire coherent rate at low snr values (below $\frac{\log W/L}{W/L}$), and asymptotically saturates at this snr because it reaches the channel uncertainty penalty. For finite bandwidths the path gains cause a further increase in the channel uncertainty penalty. The incoherent datarate $I(Y;X)$ is not concave.

into a non efficient regime where communications are not possible in the limit and the asymptotic rate is zero, and an efficient regime where the rate is penalized by the channel's entropy. The transition between the two regimes is asymptotically at snr $= \theta\left(\frac{\log W/L}{W/L}\right)$. *Coherent systems* present a different asymptotic behavior. Their rate is concave $\cap$ in snr, and linear in the entire low snr regime, up to the spectral efficiency of the modulation.

Wideband PPM systems were analyzed in [5] and [6]; Correct estimation of the channel path delays is not possible in the limit of large bandwidth for PPM systems with limited average power and a non-vanishing data rate. The results in [5] and [6] use different detectors of the multipath components, a threshold detector in the first and a maximum likelihood detector in the second. Our new result applies to PPM systems, but translates the inability of the system to estimate the channel correctly into an upper bound on the datarate. When the per burst snr is too low, communications are not possible in the limit of large bandwidth. Our previous PPM results employed the low spectral efficiency of this orthogonal modulation, by effectively calculating the lowest duty cycle that allows a positive datarate as the bandwidth grows. The spectral efficiency of PPM with a non-diminishing symbol period is upper bounded by $\theta\left(\frac{\log Wt_c}{Wt_c}\right)$. Our new result basically shows that with the low snr required to maintain spectral efficiency (i.e. snr $= O\left(\frac{\log W}{W}\right)$), communications are asymptotically impossible. The new result applies to a much wider family of wideband modulations, that may be spectrally efficient (such as direct sequence spread spectrum or iid Gaussian signaling).

Our new result applies to a wide family of signals with energy Wt_c per coherence period, and a shifted auto-correlation that may be as high as $O\left(\sqrt{W}\right)$. The auto-correlation is significant because spreading signals generated by iid or pseudo-random sequences have an empirical correlation that is linear with $\sqrt{W}$. [1] considered signals with a very low correlation, and analyzed separately iid signals.

The underlying reason for the inability of wideband systems to communicate is channel uncertainty, and our result applies to channels where the number of multipath components is unbounded and sub-linear in the bandwidth. Our proof is based on the connection between mutual information and the minimum mean square error (I-mmse) [7] and hinges on a calculation of the minimum mean square error (mmse) estimate of the unknown channel. A similar technique is used by [8], to calculate an upper bound on datarates in systems of a different type.

We consider signals that spread their power over the entire available bandwidth, such as PPM or impulse radio [9] where the pulse shape and duration determine the bandwidth, or direct sequence spread spectrum where the chip duration determines the frequency spread of the signal. Our result applies also to OFDM-type signals, if the entire available bandwidth is used uniformly and concurrently. Examples of signals that are not spread over bandwidth are FSK and multi-tone FSK [10], [11], where each symbol concentrates power on a small span of frequencies, although the entire range of symbols may span a very large bandwidth.

The new result shows that channel uncertainty is detrimental to low snr spreading systems operating over multipath channels where the number of apparent paths is unbounded but sub-linear with the system bandwidth [12], [13], essentially because the signal uses too many eigen-modes of the channel. An eigen-mode in this contect is a signal that only goes through scaling and delay as a result of being transmitted over the noiseless channel. Our model of channel variation in time is a block coherent one, where the channel is fixed for known lengths of time (coherence periods) and realizations over different coherence periods are iid. This channel model offers the advantage of eigen-modes that are particularly simple, as harmonic signals are eigen-modes of any linear time invariant channel. The channel uncertainty a communication system faces when operating over the block–coherent channel is thus limited to the eigen-values of the channel, or in other words to the complex channel gain over the frequency band the system uses.

The essential feature of spreading signals that renders them ineffective over wide bands, is that they use the entire range of channel eigen-modes concurrently, and are thus exposed to uncertainty of a large number of parameters (channel eigen-values). Modulation schemes of the FSK type, that exploit a small number of channel eigen-modes per symbol, are exposed to uncertainty in only a small number of parameters.

Our result can be extended to more complex channels, where the variation in time is described using the Doppler spectrum rather than block–coherence [14], [15], [16]. The eigen-modes of such channels are approximately given by orthogonal Weyl-Heisenberg bases [17] in the under-spread case, i.e. when the channel's response is highly concentrated in the delay–Doppler plane. The essential feature that determines whether communications are possible in the wideband limit is the spreading of symbol power over the eigen-modes of the unknown channel.

The channel uncertainty penalty is in fact upper bounded by $\theta\left(L\log W/L\right)$ in our model because this is the asymptotic entropy of the channel. The snr bound (1) corresponds to this channel uncertainty by equating the maximal coherent datarate $\theta\left(W\text{snr}\right)$ to the channel uncertainty bound. We conclude that channel uncertainty prevents asymptotic wideband communications in all situations where the coherent rate is lower than the channel entropy, as seen in Figure 1.

Outline of the rest of the paper: after presenting the model in Section II, the main result is presented in Section III and discussed in Section IV. A truncated proof is presented in Section V, and a central lemma required for the proof is proven in an appendix.

Notation: vectors are marked by upper case symbols (X) and matrices by bold upper case ($\mathbf{X}$). Scalars are marked by either lower or upper case, upper case symbols are used for the important parameters, i.e. the bandwidth W and number of channel paths L.

II. Model

Consider communication systems with a (single sided) bandwidth W, operating over block-coherent multipath channels. We use a real discrete model of the system, after sampling at the receiver at rate W. The model over a single coherence period is given by

$$Y = \sqrt{\text{snr}}\,X \star \tilde{H} + Z \tag{3}$$

where Y is the received signal over an entire coherence period of length t_c, this is a vector with $k_c = t_c W$ entries. The vector X of length k_c represents the transmitted signal, the multipath channel is represented by the vector $\tilde{H}$ of length k_c and $\star$ marks a convolution. Z is white standard Gaussian noise (iid with zero mean and variance one) and snr is the signal to noise ratio, per frequency resolution bin or per degree of freedom.

The transmitted signal may be impulsive, we consider the signal to noise ratio during active transmission so there is no need to explicitly address the impulsiveness used by the system (i.e. the duty cycle ratio). We assume that the transmitter does not use information on the channel realization.

The model (3) neglects edge effects at the beginning of the coherence period.

The transmitted signal satisfies and energy constraint

$$E\left[\|X\|_2^2\right] = k_c \tag{4}$$

and in addition we impose a probabilistic energy constraint:

$$P\left(\|X\|_2^2 > (1+o(1))\,k_c\right) \xrightarrow[W\to\infty]{} 0 \tag{5}$$

Note that iid signaling satisfies this assumption.

The transmitted signal X is wideband: its empirical auto-correlation is upper bounded

$$\left|\left\langle \overline{X^i}, \overline{X^j} \right\rangle\right| = O\left(\frac{\|X\|_2^2}{\sqrt{k_c}}\right) = O\left(\sqrt{k_c}\right) \tag{6}$$

$$i \neq j \quad i,\ j = 1,\ldots,k_c$$

The notation $<\ ,\ >$ is used for the inner product of vectors, and the notation $\overline{X^i}$ is used for a vector X that is cyclicly shifted by $i-1$ positions, i.e.

$$\overline{X^i} = \begin{pmatrix} X_{(2-i)} \\ X_{(3-i)} \\ \vdots \\ X_{(k_c+1-i)} \end{pmatrix} \tag{7}$$

and $(\ -\ )$ indicates a mod k_c difference. Condition (6) essentially requires a vector X with an autocorrelation that is similar to that of iid noise.

The channel is composed of L paths, each with a discrete delay in the range $0, 1, \ldots, t_d W$ where t_d is the delay spread. It is represented in a vector $\tilde{H}$ where most (all but L) entries are zeros, and it obeys

$$E\left[\left\|X\tilde{H}\right\|_2^2\right] = k_c \tag{8}$$

The channel is block-constant with coherence time t_c, i.e. it has iid realizations over different coherence periods. We assume $t_d \ll t_c$ and thus justify to an extent our loose treatment of edge effects at the beginning of each coherence period, and approximate (3) with a circularly-shifted matrix:

$$Y = \sqrt{\text{snr}}\,\mathbf{X}\tilde{H} + Z \tag{9}$$

where

$$\begin{aligned} \mathbf{X} &= \begin{pmatrix} X_1 & X_{k_c} & \ldots & X_2 \\ X_2 & X_1 & X_{k_c} & \ldots \\ \vdots & & \ddots & \ddots \\ X_{k_c} & \ldots & X_2 & X_1 \end{pmatrix} \\ &= \begin{pmatrix} \overline{X^1} & \overline{X^2} & \ldots & \overline{X^{k_c}} \end{pmatrix} \end{aligned} \tag{10}$$

The channel model is real, L channel gains are iid and zero mean, with variance $1/L$, so the energy in the channel's impulse response equals one on average. The choice of the L non zero tap delays is uniform over the $\binom{Wt_d}{L}$ possibilities.

The number of paths L diverges as the bandwidth increases in a sub-linear manner [12], [13], i.e. $L \xrightarrow[W\to\infty]{} \infty$ and $L = o(W)$.

We make a technical probabilistic assumption on the channel's response

$$P\left(\left|\sum_{j=1,\ j\neq i}^{k_c} \tilde{H}_j \left\langle \overline{X^i}, \overline{X^j} \right\rangle\right| > \beta\sqrt{k_c}\right) \xrightarrow[W\to\infty]{} 0$$
$$i = 1, 2, \ldots, k_c \tag{11}$$

with a constant β that does not depend on the bandwidth. This assumption holds for all reasonable distributions of the channel gains, i.e. distributions with a tail that is not exceptionally heavy. Assumption (11) ensures that the probability of a channel realization that can be estimated easily is small.

III. RESULT

Theorem 1: Consider a communication system of the form

$$Y = \sqrt{\text{snr}}\,\mathbf{X}\tilde{H} + Z \tag{12}$$

where the transmitted signal $\mathbf{X}$ is generated (using (7) and (10)) from a k_c long vector X that satisfies:

$$E\left[\|X\|_2^2\right] = k_c \tag{13}$$

$$P\left(\|X\|_2^2 > (1+o(1))\,k_c\right) \xrightarrow[W\to\infty]{} 0 \tag{14}$$

$$\left|\left\langle \overline{X^i}, \overline{X^j} \right\rangle\right| = O\left(\frac{\|X\|_2^2}{\sqrt{k_c}}\right) \qquad i \neq j \quad i,\ j = 1, \ldots, k_c \tag{15}$$

The channel $\tilde{H}$ is a k_c long vector with iid L zero mean entries with variance $1/L$, with indices that are chosen uniformly from the first Wt_d positions and

$$P\left(\left|\sum_{j=1,\ j\neq i}^{k_c} \tilde{H}_j \left\langle \overline{X^i}, \overline{X^j} \right\rangle\right| > \beta\sqrt{k_c}\right) \xrightarrow[W\to\infty]{} 0$$
$$i = 1, 2, \ldots, k_c \tag{16}$$

The noise vector Z contains standard Gaussian noise.

Claim: if the signal to noise ratio during active communication periods satisfies $\text{snr} = o\left(\frac{\log(W/L)}{W/L}\right)$, then the datarate diminishes in the limit of large bandwidth:

$$\frac{I(Y;\mathbf{X})}{\frac{1}{2}k_c\text{snr}} \xrightarrow[W\to\infty]{} 0$$

with probability one.

IV. Discussion of the Proof

We prove the theorem in Section V by showing that

$$\lim_{W\to\infty} \frac{I\left(Y;\tilde{H}|\mathbf{X}\right)}{\frac{1}{2}k_c\text{snr}} = 1 \tag{17}$$

and applying

$$\begin{aligned} I\left(Y;\mathbf{X}\right) &= I\left(Y;\tilde{H},\mathbf{X}\right) - I\left(Y;\tilde{H}|\mathbf{X}\right) &(18)\\ &\le \frac{1}{2}k_c\text{snr} - I\left(Y;\tilde{H}|\mathbf{X}\right) &(19)\end{aligned}$$

The last term in (19) is the datarate penalty due to channel uncertainty.

The proof of Theorem 1 is based on calculating the mmse estimate of the channel response $\tilde{H}$, given the transmitted and the received signals. We show that this mmse estimate is a vector with an $o(1)$ norm if the snr is low, essentially because the noise overwhelms the information carrying signal.

The channel uncertainty penalty is upper bounded by the channel entropy, and the bottom graph of Figure 1 thus saturates at

$$\begin{aligned} \frac{1}{2}k_c\text{snr} &= L\log\frac{Wt_d}{L} + \theta\left(L\right) &(20)\\ \text{snr} &= \theta\left(\frac{\log W/L}{W/L}\right) + \theta\left(\frac{L}{W}\right) &(21)\end{aligned}$$

where the first part in the RHS of (20) is a large bandwidth approximation of the entropy of the paths' delays and the second part corresponds to their gains.

The channel uncertainty penalty (last term in (19) is calculated by applying the I-mmse connection, in particular Theorem 2 of [7]. This theorem gives a simple formula to the achievable rate of communications over a known vector channel in terms of the error of the mmse estimate of the transmitted signal. We reverse the roles of H and $\mathbf{X}$ in our usage of Theorem 2 of [7], i.e. consider $\mathbf{X}$ as known and H as the estimated party.

The proof continues by explicitly calculating an estimate of the channel in positions where the actual channel gains are nonzero. The estimate diminishes in the large bandwidth limit so the error converges to the channel gains themselves. The minimal mean square error (23) is $k_c\left(1-o(1)\right)$, the mutual information (22) is $\frac{1}{2}k_c\text{snr}\left(1-o(1)\right)$ and (19) is $\frac{1}{2}\text{snr}\, o\left(k_c\right)$.

V. Sketch of the Proof of Theorem 1

We start by re-writing the last term of (19). Using our notation, the I-mmse connection says that as long as the vector H satisfies $\mathrm{E}\|H\|_2^2 < \infty$ we have

$$\begin{aligned} &I\left(H;\sqrt{\text{snr}}\,\mathbf{X}H + Z\middle|\,\mathbf{X}\right) = \\ &\qquad \frac{1}{2}\int_0^{\text{snr}} E_X\left[\text{mmse}\left(\text{snr}\right)\right]d\text{snr} \end{aligned} \tag{22}$$

where the expected mmse is given by

$$\begin{aligned} &E_X\left[\text{mmse}\left(\text{snr}\right)\right] = \\ &\qquad E_{X\tilde{H}Z}\left[\left\|\mathbf{X}\tilde{H} - \mathbf{X}\hat{H}\left(Y;\text{snr}\right)\right\|_2^2\right] \end{aligned} \tag{23}$$

and $\hat{H}$ is the mmse estimate of $\tilde{H}$ given both $\mathbf{X}$ and Y. We will show that the mmse estimate is a vector with an $o(1)$ norm in low snr conditions.

The mmse estimate is given by

$$\hat{H} = E\left[H|Y,\mathbf{X}\right] \tag{24}$$

We lower bound the mmse by calculating it in a system that is given additional information on $\tilde{H}$, namely which of its positions satisfy (16).

$$\hat{H}' = E\left[H|Y,\mathbf{X},I_{(16)}\right] \tag{25}$$

where $I_{(16)}$ is a list of indices $\{i\}$ where $\tilde{H}_i$ satisfies (16). The additional information can only reduce the mmse.

$$\hat{H}' = \int\int\ldots\int Hf\left(H|Y,\mathbf{X}\right)dH_1\;dH_2\ldots dH_L \tag{26}$$

The conditional probability density in (26) is manipulated using the independence of the transmitted signal from the channel.

$$\hat{H}' = \frac{\int\int\ldots\int Hf\left(Y|\mathbf{X},H\right)f\left(H\right)dH_1\;dH_2\ldots dH_L}{\int\int\ldots\int f\left(Y|\mathbf{X},H\right)f\left(H\right)dH_1\;dH_2\ldots dH_L} \tag{27}$$

The conditional probability density in (27) is Gaussian. We denote by $f_s(\;)$ the probability density of a k_c long vector of iid standard Gaussian variables

$$f_s(S) = \frac{1}{\left(2\pi\right)^{k_c/2}}\exp\left(-\frac{1}{2}\left|S\right|^2\right) \tag{28}$$

$$\hat{H}' = \int\int\ldots\int \quad H f_s\left(Y-\sqrt{\text{snr}}\mathbf{X}H\right) f\left(H\right)dH_1\ dH_2\ldots dH_L \Bigg/ \int\int\ldots\int \quad f_s\left(Y-\sqrt{\text{snr}}\mathbf{X}H\right) f\left(H\right)dH_1\ dH_2\ldots dH_L \tag{29}$$

and proceed to examine the components of the vector $\hat{H}'$. Consider first positions (indices) j where $\tilde{H}_j = 0$. At these positions, any non-zero value of $\hat{H}'_j$ increases the estimation error and can be disregarded in the calculation of a lower bound on the mmse. We now examine the positions where $\tilde{H}_i \neq 0$ and (16) holds, and look at the estimates of each such value:

$$\hat{H}'_i = \int\int\ldots\int \quad H_i f_s\left(Y-\sqrt{\text{snr}}\mathbf{X}H\right) f\left(H\right)dH_1\ dH_2\ldots dH_L \Bigg/ \int\int\ldots\int \quad f_s\left(Y-\sqrt{\text{snr}}\mathbf{X}H\right) f\left(H\right)dH_1\ dH_2\ldots dH_L \tag{30}$$

Assuming (16) we show that the nominator of (30) is negligible when compared to its denominator, and the mmse estimate is small.

We first approximate both integrals in (30) by Riemann sums over sampled groups of values of each position in the vector H with a sampling step ΔH, and define $p\left(H\right) = f\left(H\right)\Delta H^L$. The sampling is done over a tight enough grid that the resulting errors are small.

An upper bound to (30) is calculated by breaking the sum in the denominator to a series of sums over disjoint groups of values of H, where each group corresponds to a single assignment of H in the nominator. The set $\mathcal{B}$ consists of assignments with a non-zero value $H_i \neq 0$. Rewriting the discrete approximation of (30) we get

$$\hat{H}'_i = \frac{\sum_{H\in\mathcal{B}} H_i f_s\left(Y-\sqrt{\text{snr}}\mathbf{X}H\right)p\left(H\right)}{\sum_{G\in\mathcal{B}+\mathcal{B}^c} f_s\left(Y-\sqrt{\text{snr}}\mathbf{X}G\right)p\left(G\right)} \tag{31}$$

the notation G was introduced to improve clarity.

We proceed to divide the entire range of vectors G into non-overlapping subgroups, such that for each $H\in\mathcal{B}$ we have a corresponding subgroup $\mathcal{A}\left(H\right)$, such that

$$J(H) = \frac{H_i f_s\left(Y-\sqrt{\text{snr}}\mathbf{X}H\right)p\left(H\right)}{\sum_{G\in\mathcal{A}(H)} f_s\left(Y-\sqrt{\text{snr}}\mathbf{X}G\right)p\left(G\right)} \tag{32}$$

diminishes in the limit of large bandwidth. The convergence of (31) to zero follows directly, as the nominator of (31) is a sum of the nominators of $J(H)$ for all $H\in\mathcal{B}$, and the denominator of (31) is a sum of the denominators of the different $J(H)$. $J(H)$ can be uniformly upper-bounded as shown below, and the same bound applies to (31).

The subgroups $\mathcal{A}(H)$ are created randomly. For each assignment of G that has a non-zero value in the i^{th} position, it is put in the subgroup $\mathcal{A}(G)$. For a vector G with $G_i = 0$ we (uniformly) choose one of its non-zero taps and replace it to the i^{th} position. To clarify the process, let us say that the j^{th} position of the vector G was chosen. We calculate a new vector H by

$$H_i = G_j\ ;\quad H_j = 0\ ;\quad H_k = G_k \text{ for } k\neq i,\ j \tag{33}$$

and assign the vector G to the subgroup $\mathcal{A}(H)$.

Each group $\mathcal{A}(H)$ contains H and about $\left(k_c - L\right)/L$ other members, each different from H in exactly two positions. We ensure that groups' sizes do not deviate significantly from $\left(k_c - L\right)/L$ by relocating members from large groups into suitable smaller ones.

Denote by $H^{i\to k}$ a member of $\mathcal{A}(H)$ that differs from H by exchanging the values in its i^{th} and k^{th} positions, and define $H^{i\to i} = H$. The set $\mathcal{K}(H)$ holds the values of k such that $H^{i\to k}\in\mathcal{A}(H)$.

The terms $p\left(H\right)$ in the nominator of (32) and $p\left(G\right)$ in the denominator are identical for all members of the group $\mathcal{A}(H)$ because of our assumptions on iid gains and a uniform spread of the path delays.

$$J(H) = \frac{H_i \exp\left\{-\frac{1}{2}\left\|Y-\sqrt{\text{snr}}\mathbf{X}H^{i\to i}\right\|_2^2\right\}}{\sum_{k\in\mathcal{K}(H)} \exp\left\{-\frac{1}{2}\left\|Y-\sqrt{\text{snr}}\mathbf{X}H^{i\to k}\right\|_2^2\right\}} \tag{34}$$

The denominator of (34) contains a sum over about k_c/L exponential factors with different values of k, including $k = i$ and the nominator holds a signal such factor with $k = i$. We take a close look at their exponent and introduce the notation $V\left(H_i, k\right)$ for a k_c-long vector with the value H_i at the k^{th} positions and zeros elsewhere.

$$
\begin{aligned}
- & \quad \frac{1}{2}\left\|Y-\sqrt{\text{snr}}\mathbf{X}H^{i\to k}\right\|_2^2 = \\
- & \quad \frac{1}{2}\|Y\|_2^2-\frac{1}{2}\left\|\sqrt{\text{snr}}\mathbf{X}\left(H^{i\to k}-V\left(H_i,k\right)\right)\right\|_2^2 && (35)\\
- & \quad \frac{1}{2}\left\|\sqrt{\text{snr}}\mathbf{X}V\left(H_i,k\right)\right\|_2^2 && (36)\\
- & \quad \left\langle\sqrt{\text{snr}}\mathbf{X}\left(H^{i\to k}-V\left(H_i,k\right)\right),\right. \\
& \quad \left.\sqrt{\text{snr}}\mathbf{X}V\left(H_i,k\right)\right\rangle && (37)\\
+ & \quad \left\langle\sqrt{\text{snr}}\mathbf{X}\tilde{H},\sqrt{\text{snr}}\mathbf{X}\left(H^{i\to k}-V\left(H_i,k\right)\right)\right\rangle && (38)\\
+ & \quad \left\langle\sqrt{\text{snr}}\mathbf{X}\tilde{H},\sqrt{\text{snr}}\mathbf{X}V\left(H_i,k\right)\right\rangle && (39)\\
+ & \quad \left\langle Z,\sqrt{\text{snr}}\mathbf{X}\left(H^{i\to k}-V\left(H_i,k\right)\right)\right\rangle && (40)\\
+ & \quad \left\langle Z,\sqrt{\text{snr}}\mathbf{X}V\left(H_i,k\right)\right\rangle && (41)
\end{aligned}
$$

The proof proceeds by analyzing each line (35)-(41), to show that the nominator of (34) is much smaller than the denominator. The terms (35), (36), (37) and (40) are equal across the entire group $\mathcal{A}(H)$ and thus do not have any effect on $J(H)$. We omit much of the detail due to limited space, and discuss only the analysis of the significant parts in the nominator and denominator of (34). The term (39) is the dominant term in the nominator of (34), and we show that its exponent is much smaller than (41) in the denominator of (34).

Nominator: The term from (39)

$$
\begin{aligned}
b_i &= \left\langle\sqrt{\text{snr}}\mathbf{X}\tilde{H},\sqrt{\text{snr}}\mathbf{X}V\left(H_i,k\right)\right\rangle \\
&= \text{snr}\,H_i\sum_{j=1}^{k_c}\tilde{H}_j\left\langle\overline{X^{j-1}},\overline{X^{i-1}}\right\rangle && (42)
\end{aligned}
$$

is the dominant term in the nominator of (34). We use (5) and (16) show that with a high probability

$$
\begin{aligned}
|b_i| &\leq \text{snr}\,H_i^2\left(1+o(1)\right)k_c+\text{snr}\,H_i\beta\sqrt{k_c} \\
&= O\left(\frac{\text{snr}\,k_c}{L}\right) && (43)
\end{aligned}
$$

Denominator: The term

$$
c_k=\left\langle Z,\sqrt{\text{snr}}\mathbf{X}V\left(H_i,k\right)\right\rangle=\sqrt{\text{snr}}\,H_i\left\langle Z,\overline{X_k}\right\rangle \quad (44)
$$

from (41) is the dominant term in the denominator of (34). The sum of exponents of form $\exp\left\{c_k\right\}$ is lower bounded by a single factor with $k^\star\in\mathcal{K}(H)$. We use asymptotic order statistics to show that there is $k^\star\in\mathcal{K}(H)$ such that $c_{k^\star}=\sqrt{\frac{k_c\text{snr}}{L}}\sqrt{2\log\frac{k_c}{L}}$ in the limit.

The terms $\{c_k\}$ are mutually Gaussian zero mean random variables, with

$$
\begin{aligned}
\mathbf{var}\left(c_k\right) &= \text{snr}\,H_i^2\left\|\overline{X^k}\right\|_2^2=\text{snr}\,H_i^2\left\|X\right\|_2^2 && (45)\\
\text{cov}\left(c_k,c_m\right) &= \text{snr}\,H_i^2\left\langle\overline{X^k},\overline{X^m}\right\rangle \\
&= \text{snr}\,H_i^2O\left(\frac{\|X\|_2^2}{\sqrt{k_c}}\right) && (46)
\end{aligned}
$$

where the last equality is from (6).

We collect $\{c_k\}$ into the vector C of length $m=|\mathcal{K}(H)|$ and mark its correlation matrix by $\mathbf{R}$, a positive definite matrix, with a constant value $b=\text{snr}\,H_i^2\left\|X\right\|_2^2$ on its diagonal and significantly smaller values off-diagonal.

The mean and variance of the maximal of m iid $\sim N\left(0,\sigma^2\right)$ random variables are given by [18], where the mean is $\sqrt{2\ln m}-o(1)$ and the variance is $O\left(1/\ln m\right)$ These results cannot be directly applied to the maximal $\{c_k\}$ because these variables are correlated. We show below that the correlations among $\{c_k\}$ are insignificant in the limit of large bandwidth in the sense that there is a $c_{k^\star}$ that is very similar to the maximal of iid Gaussians. In order to proceed we calculate the trace of the matrix $\mathbf{R}^2$, using the Frobenius norm of $\mathbf{R}$:

$$
\begin{aligned}
\text{trace}\left(\mathbf{R}^2\right) &= \left\|\mathbf{R}\right\|_F^2 && (47)\\
&= \text{snr}^2H_i^4\sum_{k\in\mathcal{K}(H)}\sum_{j\in\mathcal{K}(H)}\left\langle\overline{X^k},\overline{X^j}\right\rangle^2 && (48)\\
&= \text{snr}^2H_i^4\sum_{k\in\mathcal{K}(H)}\left\langle\overline{X^k},\overline{X^k}\right\rangle^2 \\
&\quad +\text{snr}^2H_i^4\sum_{k\in\mathcal{K}(H)}\sum_{j\in\mathcal{K}(H)\ j\neq k}\left\langle\overline{X^k},\overline{X^j}\right\rangle^2 && (49)\\
&= \text{snr}^2H_i^4\left\|X\right\|_2^4\left(m+m(m-1)O\left(\frac{1}{k_c}\right)\right) \\
&= \text{snr}^2H_i^4\left\|X\right\|_2^4m\left(1+O\left(\frac{1}{L}\right)\right) && (50)\\
&= \text{snr}^2H_i^4\left\|X\right\|_2^4m\left(1+o\left(1\right)\right) && (51)
\end{aligned}
$$

Lemma 1: Consider a real matrix $\mathbf{R}$ of size $m\times m$ that satisfies the following three conditions:

- It is positive definite
- Its diagonal entries equal b
- The trace of the matrix $\mathbf{R^2}$ equals $(1+o(1))mb^2$, as m increases

Then there is a positive definite matrix $\mathbf{R^{1/2}}$ such that

- $\mathbf{R^{1/2}}\left(\mathbf{R^{1/2}}\right)^T=\mathbf{R}$

- The diagonal elements of $\mathbf{R^{1/2}}$ are no bigger than $\sqrt{b}$.
- Almost all the diagonal elements of $\mathbf{R^{1/2}}$ almost equal $\sqrt{b}$, in the sense that no more than $o(1)m$ elements deviate from $\sqrt{b}$ by more than $o(1)\sqrt{b}$.
- The sum of squares of the off-diagonal elements in each row of $(1-o(1))\,m$ rows of $\mathbf{R^{1/2}}$, i.e. the sum $\sum_{j=1,\ j\neq i}^{m}\left(\left(\mathbf{R^{1/2}}\right)_{ij}\right)^2$ for almost all values of i, is $o(1)b$.

The proof is given in the appendix. Using Lemma 1, the vector C is represented with a standard Gaussian vector $Z_c \sim (0, I)$ and a square matrix $\mathbf{R^{1/2}}$

$$C = \mathbf{R^{1/2}} Z_c \tag{52}$$

We now look at $c_{k^\star}$ that corresponds to the maximal element of the vector Z_c. The expectations below are taken with respect to the transmitted signal X.

$$\begin{aligned} c_{k^\star} &= \mathbf{R^{1/2}}_{k^\star k^\star} Z_{ck^\star} && (53)\\ E\left[c_{k^\star}\right] &= \mathbf{R^{1/2}}_{k^\star k^\star}\left(\sqrt{2\ln m}-o(1)\right) && (54)\\ &= \sqrt{\text{snr}\, k_c} H_i \sqrt{2\ln\frac{k_c}{L}}\,(1+o(1)) && (55)\\ \mathbf{var}(c_{k^\star}) &= {\mathbf{R^{1/2}}^2}_{k^\star k^\star} O\left(\frac{1}{\ln m}\right) \\ &\quad + \sum_{m=1,\ m\neq k^\star}^{m} {\mathbf{R^{1/2}}^2}_{mk} && (56)\\ &= O\left(\frac{1}{\ln m}\right) o(1)\text{snr}\, H_i^2 k_c && (57) \end{aligned}$$

and conclude that in the limit

$$c_{k^\star} = \sqrt{\text{snr}\, k_c} H_i \sqrt{2\ln\frac{k_c}{L}}\,(1+o(1)) \tag{58}$$

Conclusion of the Proof: We upper bound (34) in the limit of large bandwidth using the significant terms in the nominator and denominator. Using (43):

$$\begin{aligned} J(H) &\leq \frac{H_i \exp\left\{-\frac{1}{2}\left\|Y-\sqrt{\text{snr}}\mathbf{X}H^{i\to i}\right\|_2^2\right\}}{\max_{k\in\mathcal{K}(H)}\exp\left\{-\frac{1}{2}\left\|Y-\sqrt{\text{snr}}\mathbf{X}H^{i\to k}\right\|_2^2\right\}} \\ &\leq H_i \exp\left\{\theta\left(\frac{k_c\text{snr}}{L}\right)-\sqrt{\frac{k_c\text{snr}}{L}}\sqrt{2\log\frac{k_c}{L}}\right\} \end{aligned} \tag{59}$$

For snr $= o\left(\frac{\log k_c/L}{k_c/L}\right) = o\left(\frac{\log W/L}{W/L}\right)$ the expression (59) diverges to $-\infty$ as the bandwidth increases, and $J(H) \xrightarrow[W\to\infty]{} 0$. The mmse estimate of the channel (31) thus also diminishes. The proof of Theorem 1 is completed by noting that (23) is $k_c\,(1-o(1))$, thus (22) is $\frac{1}{2}k_c\text{snr}\,(1-o(1))$. The ratio (17) is thus shown to approach unity and the proof is complete.

APPENDIX

PROOF OF LEMMA 1

Since $\mathbf{R}$ is a positive definite matrix, we can represent it by

$$\mathbf{R} = \mathbf{SVS}^T \tag{60}$$

where $\mathbf{S}$ is real and unitary, and $\mathbf{V}$ is a diagonal matrix with the positive eigenvalues $\{\sigma_i\}$ on its diagonal. Define $\mathbf{V^{1/2}}$ as a diagonal matrix with the positive square roots $\sqrt{\sigma_i}$ on its diagonal, and and define

$$\mathbf{R^{1/2}} = \mathbf{SV^{1/2}S}^T \tag{61}$$

Obviously $\mathbf{R} = \mathbf{R^{1/2}}\left(\mathbf{R^{1/2}}\right)^T$

We calculate an upper bound on the diagonal elements of $\mathbf{R^{1/2}}$ from

$$\begin{aligned} &(\mathbf{R})_{ii} = b = \sum_{j=1}^{M}\left(\left(\mathbf{R^{1/2}}\right)_{ij}\right)^2 \geq \left(\left(\mathbf{R^{1/2}}\right)_{ii}\right)^2 \\ \Longrightarrow\quad &\left|\left(\mathbf{R^{1/2}}\right)_{ii}\right| \leq b \end{aligned} \tag{62}$$

Divide the indices of $\{\sigma_i\}_{i=1}^{M}$ into two sets, $\mathcal{G}$ and its complements $\mathcal{G}^c$:

$$\mathcal{G} = \{i|\sigma_i \geq b\} \tag{63}$$

and

$$\mathcal{G}^c = \{i|\sigma_i < b\} \tag{64}$$

define $\epsilon_1 \in [0,1]$, $\epsilon_2 > 0$ such that $|\mathcal{G}| = \epsilon_1 M$ and

$$\sum_{i\in\mathcal{G}}\sigma_i = b\epsilon_1 M\,(1+\epsilon_2) \tag{65}$$

Obviously

$$|\mathcal{G}^c| = (1-\epsilon_1)\,M \tag{66}$$

and

$$\sum_{i\in\mathcal{G}^c}\sigma_i = \sum_{i=1}^{M}\sigma_i - \sum_{i\in\mathcal{G}}\sigma_i = Mb\,(1-\epsilon_1\,(1+\epsilon_2)) \tag{67}$$

The arithmetic mean of the eigenvalues in $\mathcal{G}$ is

$$\frac{\sum_{i\in\mathcal{G}}\sigma_i}{|\mathcal{G}|}=(1+\epsilon_2)b \tag{68}$$

and the mean of the eigenvalues in $\mathcal{G}^c$ is

$$\begin{aligned}\frac{\sum_{i\in\mathcal{G}^c}\sigma_i}{|\mathcal{G}^c|} &= \frac{Mb\left(1-\epsilon_1\left(1+\epsilon_2\right)\right)}{M\left(1-\epsilon_1\right)} \\ &= b\left(1-\frac{\epsilon_1\epsilon_2}{1-\epsilon_1}\right) \end{aligned}\tag{69}$$

Using the convexity of the quadratic function:

$$\sum_{i\in\mathcal{G}}\sigma_i^2\geq|\mathcal{G}|\left(\frac{\sum_{i\in\mathcal{G}}\sigma_i}{|\mathcal{G}|}\right)^2=\epsilon_1M\left(1+\epsilon_2\right)^2b^2 \tag{70}$$

and similarly

$$\begin{aligned}\sum_{i\in\mathcal{G}^c}\sigma_i^2\geq|\mathcal{G}^c|\left(\frac{\sum_{i\in\mathcal{G}^c}\sigma_i}{|\mathcal{G}^c|}\right)^2 &= \\ \left(1-\epsilon_1\right)Mb^2\left(1-\frac{\epsilon_1\epsilon_2}{1-\epsilon_1}\right)^2 & \end{aligned}\tag{71}$$

The eigenvalues of the square matrix $(\mathbf{R})^2$ are $\{\sigma_i^2\}$ and we use the requirement on the trace of this matrix to conclude

$$\begin{aligned}\sum\sigma_i^2 &= \text{trace}\left((\mathbf{R})^2\right) &(72)\\ &= b^2M(1+o(1)) &(73)\\ &\geq |\mathcal{G}|\left(\frac{\sum_{i\in\mathcal{G}}\sigma_i}{|\mathcal{G}|}\right)^2+|\mathcal{G}^c|\left(\frac{\sum_{i\in\mathcal{G}^c}\sigma_i}{|\mathcal{G}^c|}\right)^2 \\ &= Mb^2\left(1+\epsilon_1\epsilon_2^2+\frac{\epsilon_1^2\epsilon_2^2}{1-\epsilon_1}\right) &(74)\end{aligned}$$

The inequality yeilds

$$\epsilon_1\epsilon_2^2+\frac{\epsilon_1^2\epsilon_2^2}{1-\epsilon_1}=o(1) \tag{75}$$

this is possible if and only if the product $\epsilon_1\epsilon_2$ is $o(1)$, i.e. either ϵ_1 or ϵ_2 (or both) are $o(1)$.

If ϵ_1 is $o(1)$ then the number of eigenvalues of $\mathbf{R}$ bigger than b is $o(M)$. If $\epsilon_2=o(1)$ and ϵ_1 is fixed, we upper bound the number of elements in $\mathcal{G}$ that are bigger than $\left(1+\sqrt{\epsilon_2}\right)b$. Denote the number of these elements by δM.

$$\begin{aligned}\sum_{i\in\mathcal{G}}\sigma_i &= Mb\epsilon_1(1+\epsilon_2) &(76)\\ &\geq Mb\left(\epsilon_1+\delta\sqrt{\epsilon_2}\right) &(77)\end{aligned}$$

so

$$\begin{aligned}\epsilon_1(1+\epsilon_2) &\geq \epsilon_1+\delta\sqrt{\epsilon_2} &(78)\\ \epsilon_1\epsilon_2 &\geq \delta\sqrt{\epsilon_2} &(79)\\ \delta &\leq \epsilon_1\sqrt{\epsilon_2}=o(1) &(80)\end{aligned}$$

One can similarly bound the number of eigenvalues of $\mathbf{R}$ significantly smaller than b. If $\epsilon_1=o(1)$, we upper bound the number of elements in $\mathcal{G}^c$ that are smaller than $\left(1-\sqrt{\epsilon_1}\right)b$. Denote the number of such elements by δM.

$$\begin{aligned}\sum_{i\in\mathcal{G}^c}\sigma_i &= Mb\left(1-\epsilon_1-\epsilon_1\epsilon_2\right) &(81)\\ &\leq Mb\left(1-\epsilon_1-\delta\sqrt{\epsilon_1}\right) &(82)\end{aligned}$$

so

$$\begin{aligned}-\epsilon_1\epsilon_2 &\leq -\delta\sqrt{\epsilon_1} &(83)\\ \epsilon_1\epsilon_2 &\geq \delta\sqrt{\epsilon_1} &(84)\\ \delta &\leq \epsilon_1\sqrt{\epsilon_1}=o(1) &(85)\end{aligned}$$

If $\epsilon_2=o(1)$, and ϵ_1 is fixed, we upper bound the number of elements in $\mathcal{G}^c$, that are smaller than $\left(1-\sqrt{\epsilon_2}\right)b$. Denote the number of the elements by δM.

$$\begin{aligned}\sum_{i\in\mathcal{G}^c}\sigma_i &= Mb\left(1-\epsilon_1-\epsilon_1\epsilon_2\right) &(86)\\ &\leq Mb\left(1-\epsilon_1-\delta\sqrt{\epsilon_2}\right) &(87)\end{aligned}$$

so

$$\begin{aligned}-\epsilon_1\epsilon_2 &\leq -\delta\sqrt{\epsilon_2} &(88)\\ \epsilon_1\epsilon_2 &\geq \delta\sqrt{\epsilon_2} &(89)\\ \delta &\leq \epsilon_1\sqrt{\epsilon_2}=o(1) &(90)\end{aligned}$$

and as a result, all but $o(M)$ eigenvalues of $\mathbf{R^{1/2}}$, i.e. $\left\{\sqrt{\sigma_i}\right\}$ are lower bounded by $(1-o(1))\sqrt{b}$. This is used to lower bound the trace:

$$\begin{aligned}\text{trace}\left(\mathbf{R^{1/2}}\right) &= \sum\sqrt{\sigma_i} \\ &\geq \left(1-o(1)\right)M\left(1-o(1)\right)\sqrt{b} \\ &= \left(1-o(1)\right)M\sqrt{b} &(91)\end{aligned}$$

This lower bound together with (62) leads to the conclusion that almost all diagonal elements $\left(\mathbf{R^{1/2}}\right)_{ii}$, that is at least $1-o(1)$ of them, must be $\left(1-o(1)\right)\sqrt{b}$.

How big is the energy of the off diagonal entries? for almost all of the rows, i.e. $(M - o(M))$ values of $i = 1, \ldots, M$ we have

$$\begin{aligned}\sum_j \left(\left(\mathbf{R^{1/2}} \right)_{ij, j \neq i} \right)^2 &= (\mathbf{R})_{ii} - \left(\left(\mathbf{R^{1/2}} \right)_{ii} \right)^2 \\ &= b - (1 - o(1))b \qquad (92) \\ &= o(1)b \qquad (93)\end{aligned}$$

The proof of Lemma 1 is thus complete.

References

[1] I. E. Telatar and D. N. C. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Transactions on Information Theory*, vol. 46, no. 4, pp. 1384–1400, Jul. 2000.

[2] M. Médard and R. G. Gallager, "Bandwidth scaling for fading multipath channels," *IEEE Transactions on Information Theory*, vol. 2002, no. 4, pp. 840–852, Apr. 2002.

[3] V. G. Subramanian and B. Hajek, "Broad-band fading channels: Signal burstiness and capacity," *IEEE Transactions on Information Theory*, vol. 48, no. 4, pp. 809–827, Apr. 2002.

[4] D. Porrat, D. Tse, and Şerban Nacu, "Channel uncertainty in ultra wideband communication systems," *IEEE Transactions on Information Theory*, vol. 53, no. 1, pp. 194–108, Jan. 2007.

[5] D. Porrat and U. Mitra, "Timing acquisition of wideband PPM systems over multipath," in *IEEE International Symposium on Information Theory*, Jul 2006, pp. 2583–2587.

[6] ——, "Multipath delay profile acquisition for ultra-wideband ppm systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 372–382, Oct. 2007, special Issue on Performance Limits of Ultra-Wideband Systems.

[7] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.

[8] G. Durisi, H. Bölcskei, and S. Shamai, "Capacity of underspread noncoherent WSSUS fading channels under peak signal constraints," in *IEEE International Symposium on Information Theory*, Jun. 2007, pp. 156–160.

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

[10] C. Luo and M. Médard, "Frequency-shift keying for ultrawideband - achieving rates of the order of capacity," in *40th Annual Allerton Conference on Communication, Control, and Computing*, 2002.

[11] C. Luo, M. Médard, and L. Zheng, "Error exponents for multitone frequency shift keying on wideband rayleigh fading channels," in *IEEE Globecom*, vol. 2, Dec. 2003, pp. 779–783.

[12] F. Saadane, A. Menouni, T. Knopp, and T. Aboutajdine, "Empirical eigenanalysis of indoor UWB propagation channels," in *IEEE Globecom*, vol. 5, Nov.-Dec. 2004, pp. 3215–3219.

[13] F. Saadane, T. Aboutajdine, A. Hayar, and R. Knopp, "On the estimation of the degrees of freedom of indoor UWB channel," in *IEEE VTC*, vol. 5, May-Jun. 2005-Spring, pp. 3147–3151.

[14] V. Raghavan, G. Hariharan, and A. M. Sayeed, "Capacity of sparse multipath channels in the ultra-wideband regime," *Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 357–371, Oct. 2007.

[15] A. Sayeed, V. Ragahvan, and J. Kotecha, "Capacity of space-time wireless channels: A physical perspective," in *IEEE Information Theory Workshop*, Oct. 2004.

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

[17] W. Kozek, "On the transfer function calculus for underspread LTV channels," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 219–223, Jan. 1997.

[18] H. Cramér, *Mathematical Methods of Statistics*. Princeton University Press, 1946.