Potential Limits on Power–Line Communication over Impulsive Noise Channels

Vladimir B. Balakirsky and A. J. Han Vinck

Institute for Experimental Mathematics,
University of Essen
45326 Essen, Germany
Phone: +49-201-1837650, Fax: +49-201-1837636
E-mail: vbal, vinck@exp-math.uni-essen.de

Abstract

We investigate the influence of side information of the sender and the receiver on the performance of data transmission systems over impulsive noise channels. The "hidden" side information is the note that the in–phase and the quadrature components of the transmitted signal at a fixed time instant meet the same transmission conditions. The "exposed" side information is received when the impulsive noise channel is tested before the transmission starts. We present optimum encoding and decoding algorithms maximizing the gain in the performance reached by making use of side information. We measure the gain in the performance in the maximum transmission rate.

1. Introduction

Middleton's ("Class A") model for the impulsive noise [1], [2] serves as a basic model for many power–line communication channels. It is usually understood in such a way that the in–phase and the quadrature components of the transmitted signal pass through the Gaussian channel whose variance is a random variable. The particular value of the variance at each time instant is determined by some source of impulsive noise. The conventional decoding procedures ([3], [4], and other papers) independently treat the results of transmission of the in–phase and the quadrature components without making use of the fact that the variances of the noise, which corrupted transmission of the components at any fixed time instant, were the same. *We represent data transmission over Middleton's impulsive noise channel as transmission over two parallel channels and show that if the coincidence of variances is taken into account, then the gain in the performance can be rather big, especially for small signal–to–noise ratios.* Such a representation also assumes a possible generalization of data transmission scheme to the case when there are T parallel channels and the variance of the noise at each time instant is the same. We will show that this "side information" becomes more and more valuable when T increases.

Another possibility that can be used to improve the performance of a communication scheme is inspection of the channel, which brings the knowledge about the actual variance of the noise at each time instant. We will distinguish between the cases $(-,-),(-,+),(+,-),(+,+)$, where the first sign (the second sign) shows whether the modulator (the decoder) knows the variance of the noise or not; this is a more general setup than the one, which was proposed in [5]. The adaptation of the modulator to the state of the channel is realized by transmitting the components of the codeword by signals of varying amplitudes $\pm\sqrt{E_1},\ldots,\pm\sqrt{E_n}$ in the cases $(+,-)$ and $(+,+)$ in such a way that the energy constraint,

$$E_1 + \ldots + E_n = nE \qquad (1)$$

is satisfied, where E is a given constant. The adaptation of the decoder to the state of the channel is realized by modifying the decoding rule. We will measure the performance, which is attained by making use of the side information, by the maximum transmission rate and compute these rates for four different cases.

The paper is organized as follows. We begin with the presentation of Middleton's model as a model for data transmission over time–varying parallel channels. Then we analyze the maximum likelihood decoding algorithm for parallel channels. Finally, we present results on the maximum transmission rates when there is side information about the state of the channel.

2. Description of Middleton's model of impulsive noise channels

A conventional presentation of Middleton's ("Class A") model of impulsive noise channels is as follows. At each time instant t, the variance of an additive white Gaussian noise, σ_t is independently chosen according to a Poisson

This work was partially supported by the DFG.

distribution,

$$\Pr\Big\{\sigma = \sigma(k)\Big\} = Q(k), \quad k = 0, 1, \ldots,$$

where

$$\sigma(k) \stackrel{\triangle}{=} \Big(\frac{k\sigma_I^2/A + \sigma_G^2}{\sigma_I^2 + \sigma_G^2}\Big)^{1/2} \tag{2}$$

and

$$Q(k) \stackrel{\triangle}{=} \frac{e^{-A}A^k}{k!} \tag{3}$$

The parameter A is called the impulsive index and σ_I^2 and σ_G^2 represent the impulsive noise power and the Gaussian noise power, respectively. The transmitted signal, the noise, and the received signal are complex numbers, denoted by $x = x_1 + ix_2$, $Z = Z_1 + iZ_2$, and $Y = Y_1 + iY_2$, where $i \stackrel{\triangle}{=} \sqrt{-1}$. The transmission is described as $Y = x + Z$. The conditional probability that the noise takes some value $z = z_1 + iz_2$ in the state k is defined as

$$\Pr\{\, Z = z \,|\, k \,\} \stackrel{\triangle}{=} \mathrm{G}(z_1; 0, \sigma(k)^2)\mathrm{G}(z_2; 0, \sigma(k)^2),$$

where

$$\mathrm{G}(z; m, \sigma^2) \stackrel{\triangle}{=} \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(z-m)^2}{2\sigma^2}\right\}$$

for all m, σ, and z.

An alternative description can be presented in the following way. Suppose that $\boldsymbol{x} = (x_1, x_2)$ is a vector of length 2, called a subblock. Components of this vector are transmitted over 2 parallel additive white Gaussian noise channels with the same variance of the noise. Let $\boldsymbol{y} = (y_1, y_2)$ stand for the received subblock.

In foregoing discussions we will assume that the transmitted components at time instant t are chosen from the set $\{-\sqrt{E_t}, +\sqrt{E_t}\}$ and that the demodulator makes a hard decision about the received signal. Thus, we are dealing with data transmission over 2 parallel time varying binary symmetric channels whose crossover probability is a random variable independently chosen from the set $\{p_t(0), p_t(1), \ldots\}$, where

$$p_t(k) \stackrel{\triangle}{=} \frac{1}{\sigma(k)\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{(z-\sqrt{E_t})^2}{2\sigma(k)^2}} dz, \quad k = 0, 1, \ldots. \tag{4}$$

If the energy is a constant for all time instants t, we will omit the argument t in the specifications of E and $p(0), p(1), \ldots$, i.e.,

$$E_t = E \;\Rightarrow\; p_t(k) = p(k), \; k = 0, 1, \ldots$$

We can also generalize the setup above by introducing $T \geq 2$ parallel channels. Namely, let $\mathbf{x}$ and $\mathbf{y}$ be binary vectors of length nT. Express them as results of concatenation of n subblocks of length T and write

$$\mathbf{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n), \quad \mathbf{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n), \tag{5}$$

where $\boldsymbol{x}_t, \boldsymbol{y}_t \in \{0,1\}^T$ for all $t = 1, \ldots, n$. Introduce the conditional probability of receiving the vector $\mathbf{y}$ if the vector $\mathbf{x}$ was sent and $\varepsilon_1, \ldots, \varepsilon_n$ were the crossover probabilities of the channels as

$$W(\mathbf{y}|\mathbf{x}, \varepsilon_1, \ldots, \varepsilon_n) \stackrel{\triangle}{=} \prod_{t=1}^{n} \omega(\boldsymbol{y}_t|\boldsymbol{x}_t, \varepsilon_t), \tag{6}$$

$$\omega(\boldsymbol{y}_t|\boldsymbol{x}_t, \varepsilon_t) \stackrel{\triangle}{=} \varepsilon_t^{d_t}(1-\varepsilon_t)^{T-d_t}, \tag{7}$$

$$d_t \stackrel{\triangle}{=} \mathrm{wt}(\boldsymbol{x}_t \oplus \boldsymbol{y}_t), \tag{8}$$

where wt stands for the Hamming weight of a binary vector. *The data transmission problem can be formulated as constructing block codes and decoding algorithms for the channel defined by* (6)–(8).

3. Maximum likelihood decoding for Middleton's model of impulsive noise channels when the state of the channel is unknown

Suppose that a block code of length nT is used to transmit data over *a binary symmetric channel whose crossover probability is an i.i.d. random variable taking value $p(k)$ with probability $Q(k)$, $k = 0, 1, \ldots$* Let

$$\overline{p} \stackrel{\triangle}{=} \sum_{k=0}^{\infty} Q(k)p(k)$$

denote the average value of the crossover probability. If $\overline{p} \leq 1/2$, then the maximum likelihood decoding rule is exactly the same as for the binary symmetric channel having a fixed crossover probability smaller than $1/2$; namely, the decoder has to minimize the Hamming distance between the received vector and all possible codewords. It is easy to see that the channel capacity is equal to $1 - h(\overline{p})$, where

$$h(a) = -a\log a - (1-a)\log(1-a), \quad a \in (0,1),$$

is the binary entropy function. However, our conclusions become different if the codeword is split in n subblocks of length T, and components of each subblocks are transmitted over T parallel channels with the same crossover probability in each channel at each time instant. *The maximum likelihood decoding in this case can be presented as the minimization of the sum $\beta(d_1) + \ldots + \beta(d_n)$, where the possible values of the parameters, $\beta(0), \ldots, \beta(T)$, are computed as a function of $Q(k), p(k), k = 0, 1, \ldots$* The arguments for such a conclusion are presented below.

Using (6)–(8) and the memoryless property of Middleton's model we have

$$\begin{aligned} W(\mathbf{y}|\mathbf{x}) &\stackrel{\triangle}{=} \sum_{\gamma_1, \ldots, \gamma_n} \Pr\{\, (\varepsilon_1, \ldots, \varepsilon_n) = (\gamma_1, \ldots, \gamma_n) \,\} \\ &\qquad \cdot W(\mathbf{y}|\mathbf{x}, \varepsilon_1, \ldots, \varepsilon_n) \\ &= \prod_{t=1}^{n} \sum_{\gamma \in \{p(0), p(1), \ldots\}} \Pr\{\, \varepsilon = \gamma \,\}\omega(\boldsymbol{y}_t|\boldsymbol{x}_t, \gamma), \end{aligned}$$

Table 1: The parameters of a time–varying binary symmetric channel received from Middleton's model for $A = 1$, $\Gamma = 10^{-3}$, $E = \sigma = 1$.

k	$Q(k)$	$p(k)$
0	0.369	0.000
1	0.369	0.158
2	0.185	0.239
3	0.062	0.281
4	0.015	0.308
	$\overline{p} =$	0.124
	$1 - h(\overline{p}) =$	0.457

where

$$\Pr\{\,\varepsilon = p(k)\,\} = Q(k), \quad k = 0, 1, \ldots$$

Equivalently, using (8) we have

$$-\log W(\mathbf{y}|\mathbf{x}) = \sum_{t=1}^{n} F(d_t),$$

where

$$F(d) \stackrel{\triangle}{=} -\log \sum_{k \geq 0} Q(k) p^d(k)(1 - p(k))^{T-d}$$

for all $d = 0, \ldots, T$. If

$$\beta(d) \stackrel{\triangle}{=} \frac{F(d) - F(0)}{F(1) - F(0)},$$

then $\beta(0) = 0$, $\beta(1) = 1$ and

$$\begin{aligned} -\log W(\mathbf{y}|\mathbf{x}) &= \sum_{t=1}^{n} \Big[F(0) + (F(1) - F(0))\beta(d_t) \Big] \\ &= nF(0) + (F(1) - F(0)) \sum_{t=1}^{n} \beta(d_t). \end{aligned}$$

One can easily see that $F(1) > F(0) > 0$. Therefore, the maximization of the conditional probability of receiving $\mathbf{y}$ over all possible codewords that could be sent over the channel is equivalent to the minimization of the sum $\beta(d_1) + \ldots + \beta(d_n)$.

Some parameters of channel received from Middleton's model for $E_1 = \ldots = E_n = E$, where $E = \sigma = 1, A = 1, \Gamma = 10^{-3}$, and

$$\sigma^2 \stackrel{\triangle}{=} \sigma_G^2 + \sigma_I^2, \quad \Gamma \stackrel{\triangle}{=} \sigma_G^2/\sigma_I^2$$

are given in Table 1 and corresponding values of $\beta(0), \ldots, \beta(T)$ are given in Table 2. In particular, for this

Table 2: Coefficients $\beta(d)$ of the optimum decoding function for T parallel time–varying binary symmetric channels whose parameters are specified in Table 1.

d	$T = 1$	$T = 2$	$T = 4$	$T = 8$
0	0	0	0	0
1	1	1	1	1
2		1.64	1.58	1.49
3			2.13	1.95
4			2.64	2.38
5				2.78
6				3.16
7				3.52
8				3.87
$C =$	0.457	0.460	0.466	0.477

example we conclude that two subblocks with one error in each of them are better meaning the maximization of the conditional probability of $\mathbf{y}$ given $\mathbf{x}$ than one subblock with 3 errors if $T = 4$. If $T = 8$, then the conclusion is opposite. The formula for the capacity of the resulting channel will be given below as capacity for the case $(-, -)$.

Considerations above show that *partitioning of the transmitted codeword in subblocks brings a knowledge, which is similar to "soft information" at the decoder, while interleaving the components of the transmitted codeword, which is usually used in applications to generate a model of a memoryless channel, makes this information not available.*

4. Maximum transmission rates under the energy constraint and side information of the participants

Let us consider the data transmission over impulsive noise channels in a slightly different way. Suppose, as before, that the message to be delivered to the receiver is encoded by a codeword $\mathbf{x}$ of a binary block code of length nT. This codeword and the received vector $\mathbf{y}$ are represented as concatenations of subblocks of length T (see (5)). Components of the same subblock of the vector $\mathbf{x}$ are simultaneously transmitted over T parallel channels. There is also a participant, called a Controller, who generates a vector of states of the channels. Let us denote this vector by $\mathbf{k} = (k_1, \ldots, k_n)$. The values of components of the vector $\mathbf{k}$ are independently chosen according to a Poisson distribution $Q(k), k = 0, 1, \ldots$ defined in (3). We assume that the sender or/and the receiver may know the realization of the vector $\mathbf{k}$ in advance, which corresponds to testing the channel before the transmission starts and requires a capability of predicting the future. Components of the vector $\mathbf{k}$ specify the variance of the noise corrupting the transmitted subblocks at time instants $1, \ldots, n$ according to (2).

Suppose that if the sender knows the vector $\mathbf{k}$, then he uses this knowledge in the modulator by sending signals of different amplitudes at different time instants. The encoding remains unchanged, i.e., the fixed codeword always corresponds to the same message. If the receiver knows $\mathbf{k}$, then he puts this information to the decoder by changing the decoding rule. The demodulation procedure stays unchanged at each time instant, i.e., if the correlation of the received signal with the pulse of a positive amplitude is greater than 0, then the demodulator outputs bit 1, otherwise he outputs 0.

We distinguish between cases $(-,-)$, $(-,+)$, $(+,-)$, $(+,+)$, where the first sign (the second sign) shows whether the sender (the receiver) knows $\mathbf{k}$ or not. The maximum transmission rate in every case is defined as the ratio of the logarithm of the number of messages that can be delivered to the receiver with "a small" decoding error probability. The small decoding error probability is understood in information theory sense : if the rate is less than the maximum transmission rate, then there exists a sequence of codes having that rate and increasing lengths such that the corresponding sequence of decoding error probabilities converges to 0; conversely, if there is a sequence of codes having the fixed rate and increasing lengths such that the corresponding sequence of decoding error probabilities converges to 0, then the rate is less than the maximum transmission rate.

Notice that the situation we meet differs from the side information direction, as it is usually understood in information theory. The difference is due to the fact that our encoder is ignorant about the channel in all cases and cannot adopt codewords to transmission conditions. From mathematical point of view these assumptions essentially simplify the problem of computing the maximum transmission rate, which allows us to derive analytical results by exporting known methods of analysis to the scheme under consideration. Nevertheless, conclusions can be of interest from a conceptual point of view and they can modify conventional technical tools used to build practical systems.

Probably the closest reference to directions considered in communication theory should be given to systems of data transmission with cochannel interference [6], [7] and other papers. The main difference in the setup in our considerations is that we assume that T independent data streams should be transmitted over parallel channels or cochannels with interference. This puts the system to a multi–user framework, but then we meet the point that the description of the channel essentially differs from the existing basic models.

To formulate the results let us introduce some notations. For all $k = 0, 1, \ldots$ and $\boldsymbol{x} \in \{0,1\}^T$, denote

$$P(k, \boldsymbol{x}) \stackrel{\triangle}{=} p(k)^{\mathrm{wt}(\boldsymbol{x})} (1 - p(k))^{T - \mathrm{wt}(\boldsymbol{x})}. \tag{9}$$

We also write $\mathbf{P}(k) = (P(k, \boldsymbol{x}), \boldsymbol{x} \in \{0,1\}^T)$ and denote the entropy of the probability distribution specified by the vector $\mathbf{P}(k)$ by

$$H(\mathbf{P}(k)) \stackrel{\triangle}{=} \sum_{\boldsymbol{x} \in \{0,1\}^T} P(k, \boldsymbol{x}) \log \frac{1}{P(k, \boldsymbol{x})}.$$

If E_t is replaced in (4) with $E(k)$, we write $p'(k)$ instead of $p(k)$ and using this parameter in (9) compute components of the distribution $\mathbf{P}'(k) = (P'(k, \boldsymbol{x}), \boldsymbol{x} \in \{0,1\}^T)$.

Theorem. *Consider a binary symmetric channel defined by* (6)–(8). *Let* $R^{(-,-)}$, $R^{(-,+)}$, $R^{(+,-)}$, $R^{(+,+)}$ *denote the maximum transmission rates for the cases* $(-,-), (-,+), (+,-), (+,+)$, *respectively. Then*

$$R^{(-,-)} = 1 - \frac{1}{T} H\Big(\sum_{k=0}^{\infty} Q(k) \mathbf{P}(k) \Big), \tag{10}$$

$$\begin{aligned} R^{(-,+)} &= 1 - \frac{1}{T} \sum_{k=0}^{\infty} Q(k) H(\mathbf{P}(k)) \qquad (11) \\ &= 1 - \sum_{k=0}^{\infty} Q(k) h(p(k)), \\ R^{(+,-)} &= 1 - \frac{1}{T} \min_{E(0), E(1), \ldots} H\Big(\sum_{k=0}^{\infty} Q(k) \mathbf{P}'(k) \Big), \\ R^{(+,+)} &= 1 - \frac{1}{T} \min_{E(0), E(1), \ldots} \sum_{k=0}^{\infty} Q(k) H(\mathbf{P}'(k)) \\ &= 1 - \min_{E(0), E(1), \ldots} \sum_{k=0}^{\infty} Q(k) h(p'(k)), \end{aligned}$$

where the minimization is carried out on all $E(0), E(1), \ldots \geq 0$ *satisfying the constraint*

$$\sum_{k=0}^{\infty} Q(k) E(k) = E.$$

Formula (10) follows from the representation of our data transmission scheme as transmission of a codeword consisting of symbols $\boldsymbol{x}_t \in \{0,1\}^T$, $t = \overline{1,n}$, over a discrete memoryless channel with the output alphabet $\{0,1\}^T$ and crossover probabilities

$$W(\boldsymbol{y}_t | \boldsymbol{x}_t) = \sum_{k \geq 0} Q(k) p(k) \omega(\boldsymbol{y}_t | \boldsymbol{x}_t, p(k))$$

where the conditional probability distribution $\omega(\boldsymbol{y}_t | \boldsymbol{x}_t, p(k))$ is defined in (7), $p(k)$ is defined in (4)

Table 3: The values of the code rates that can be attained in binary Middleton's impulsive noise model for $A = 1, \Gamma = 10^{-3}$, $\sigma = 1$.

T	E	$R^{(-,-)}$	$R^{(-,+)}$	$\hat{R}^{(+,-)}$	$\hat{R}^{(+,+)}$	E	$R^{(-,-)}$	$R^{(-,+)}$	$\hat{R}^{(+,-)}$	$\hat{R}^{(+,+)}$
1	0.1	0.192	0.391	0.219	0.404	1.0	0.457	0.554	0.557	0.638
2		0.205		0.230			0.460		0.559	
4		0.233		0.253			0.466		0.563	
8		0.281		0.295			0.477		0.571	
1	0.2	0.238	0.412	0.280	0.439	1.5	0.546	0.622	0.660	0.717
2		0.247		0.288			0.548		0.661	
4		0.268		0.305			0.552		0.663	
8		0.307		0.336			0.559		0.667	
1	0.3	0.275	0.432	0.330	0.471	2.0	0.617	0.679	0.736	0.777
2		0.283		0.336			0.618		0.736	
4		0.299		0.349			0.621		0.738	
8		0.331		0.374			0.626		0.740	
1	0.4	0.308	0.452	0.373	0.501	2.5	0.674	0.726	0.794	0.824
2		0.314		0.378			0.675		0.794	
4		0.328		0.389			0.677		0.794	
8		0.354		0.408			0.681		0.796	
1	0.5	0.338	0.471	0.411	0.528	3.0	0.720	0.766	0.838	0.860
2		0.343		0.415			0.721		0.838	
4		0.354		0.424			0.723		0.838	
8		0.377		0.441			0.726		0.839	

when E is substituted for E_t, and $Q(k), k = 0, 1, \ldots$ is the Poisson distribution (see (3)).

Formula (11) is derived from the observation that computation of the maximum transmission rate can be done for a system of successive data transmission over channels having crossover probabilities $p(k), k = 0, 1, \ldots$ when the 0–th channel is used $nQ(0)$ times, the 1–st channel is used $nQ(1)$ times, etc. (we assume that n tends to infinity and does not pay attention to the fact that $nQ(k)$ should be an integer for all $k = 0, 1, \ldots$).

In both cases above we apply the well–known Shannon's result on the capacity of a discrete memoryless channel.

Computations of $R^{(+,-)}$ and $R^{(+,+)}$ are based on the arguments above for the cases $(-,-)$ and $(-,+)$, respectively with the only difference that the energy of the transmitted signal at time instant t depends on the value of k. As a result, we have to substitute $p_t(k)$ for $p(k)$ and proceed as before.

Numerical illustration of the results stated in the theorem is given in Table 3, where $\hat{R}^{(+,-)}$ and $\hat{R}^{(+,+)}$ are estimates of $R^{(+,-)}$ and $R^{(+,+)}$ obtained with the local optimization over $E(0), E(1), \ldots$ These data illustrate the point that the knowledge of the state of the channel can essentially affect the maximum transmission rate. We can also observe that if the signal–to–noise ratio is small, then the knowledge of the decoder is more important than the knowledge of the modulator. For large signal–to–noise ratios the conclusion is opposite.

5. Conclusion

Channels for power–line communication are known as very unstable transmission media. Therefore a designer of data transmission systems has to use all possibilities for increasing the reliability. Probably the most important point is a proper understanding of the influence of side information on the performance of the system and knowing the methods that allow one to achieve a gain in the performance. We addressed these issues from information theory point of view.

References

[1] D. Middleton, "Statistical–physical model of electromagnetic interference," *IEEE Trans. Electromagn. Compat.*, vol. 19, no. 3, pp. 106–126, 1977.

[2] A. D. Spaulding and D. Middleton, "Optimum reception in an impulsive interference environment – Part 1 : Coherent detection," *IEEE Trans. Commun.*, vol. 25, no. 9, pp. 910–923, 1977.

[3] D. Umehara, M. Kawai, and Y. Morihiro, "An iterative detection for M–ary SS system over impulsive noise channel," *Proc. 6–th Int. Symposium on Power–Line Communications and Its Applications (ISPLC'2002)*, F–N. Pavlidou (ed.), Athens, Greece, pp. 203–207, 2002.

[4] H. Matsuo, D. Umehara, M. Kawai, and Y. Morihiro, "An iterative detection for OFDM over impulsive noise channel," *Proc. 6–th Int. Symposium on Power–Line Communications and Its Applications (ISPLC'2002)*, F–N. Pavlidou (ed.), Athens, Greece, pp. 213–217, 2002.

[5] V. B. Balakirsky and A. J. Han Vinck, "Communication over time–varying channels under power constraints and side information of the participants," *Proc. 6–th Int. Symposium on Power–Line Communications and Its Applications (ISPLC'2002)*, F–N. Pavlidou (ed.), Athens, Greece, pp. 106–110, 2002.

[6] R. Krishnamurthi and S. C. Gupta, "The error performance of Gray encoded QPSK and 8–PSK schemes in a fading channel with cochannel interference," *IEEE Trans. Commun.*, vol. 41, no. 12, pp. 1773–1776, 1993.

[7] V. V. Veeravalli, "On performance analysis for signaling on correlated fading channels," *IEEE Trans. Commun.*, vol. 49, no. 11, pp. 1879–1883, 2001.