

Research Memorandum No. 1077

10/28/2008

Channel Estimation and Code Word Inference for Mobile Digital Satellite Broadcasting  
Reception

Hamada Masatoshi (Japan Broadcasting Corporation (NHK))  
Ikeda Shiro (The Institute of Statistical Mathematics)

The Institute of Statistical Mathematics

4-6-7 Minami-Azabu, Minato-ku,  
Tokyo, 106-8569, Japan

# Channel Estimation and Code Word Inference for Mobile Digital Satellite Broadcasting Reception

Masatoshi Hamada  
The Japan Broadcasting Corporation (NHK)  
hamada@ism.ac.jp

Shiro Ikeda  
The Institute of Statistical Mathematics  
shiro@ism.ac.jp

October 28, 2008

## Abstract

This paper proposes a method of improving reception of digital satellite broadcasting in a moving vehicle. According to some studies, the antennas used for mobile reception will be smaller in the next generation and reception will be more difficult because of a fading multipath channel with delays in a low carrier-to-noise ratio. Commonly used approaches to reduce the inter symbol interference caused by a fading multipath channel with delays are pilot sequences and diversity reception. Digital satellite broadcasting, however, does not transmit pilot sequences for channel estimation and it is not possible to install multiple antennas in a vehicle. This paper does not propose any change to the broadcasting standards but discusses how to process currently available digital satellite signals to obtain better results. Our method does not rely on the pilot sequences or diversity reception, but consists of channel estimation and stochastic inference methods. For each task, two methods are proposed. The maximum likelihood estimation and higher order statistics matching methods are proposed for the estimation, and the marginal with the joint probability inference methods are proposed for the stochastic inference. The improvements were confirmed through experiments with numerical simulations and real data. The computational costs are also discussed for future implementation.

## 1 Introduction

Since the 1990s, it has become popular to install television reception systems in vehicles and more people are viewing television programs in the mobile environment [13].

Mobile reception is generally difficult. Among the many causes of the difficulties, we discuss two: the Doppler shift and the fading multipath channel. The influence of the Doppler shift can be removed with the phase locked loop (PLL). When a fast PLL is used, the influence of the Doppler shift is negligible.

Commonly used approaches to reducing the influence of fading are the pilot sequence and diversity reception [3]. If the code includes a predetermined pilot sequence, the fading channel can be estimated. However, the current digital satellite broadcasting has only a one-byte known sequence, which is not enough for channel estimation. On the other hand, it is not possible to install multiple satellite antennas for diversity reception in vehicles, either. These are the reasons mobile reception of the satellite broadcasting more difficult than that of terrestrial broadcasting.

STRADA (Panasonic) is one of the commercial mobile reception products for the digital satellite broadcasting. It recommends that users should install an automatic machine tracking parabolic antenna to enhance the direct reception and reduce the reflected paths. Although this effectively reduces the influence of the fading, the antenna is rather large and expensive. Some studies of next-generation mobile reception of the digital satellite broadcasting have indicated that smaller antennas are preferable [16]. The carrier-to-noise (CNR) of such a compact antenna will be much lower than that for the standard parabolic antenna for static reception of digital satellite broadcasting [16]. What is worse is that the directivity is poor and it is not possible to eliminate the reflected waves that cause a fading multipath channel with delays[16].

What we propose in this paper is not a new standard for digital satellite broadcasting. No pilot sequence is assumed, and the carrier frequency is not changed. Our proposal is to enhance the received signals that are currently available from the satellite while still using an antenna with poor directivity and low CNR. Our method consists of channel estimation and stochastic inference of the codes.

Blind channel estimation is a popular idea [9] and we discuss two estimation methods for it. One is the commonly used maximum likelihood estimation (MLE) and the other is the higher order statistics (HOS) matching method. For MLE, we need to utilize the expectation maximization (EM) algorithm [4]. The EM algorithm is an iterative algorithm and its computational cost is relatively high. The HOS matching method may not be as stable as MLE, but its computational cost is very low.

For the stochastic inference, we propose two methods. One is to utilize the marginal distribution of the fading channel model and the other is to utilize the joint distribution. The idea is equivalent to inference based on the graphical model [8], which is commonly used [19, 18, 11]. Our proposal is to find the trade-off between the computational cost and the improvement.

The combinations of the proposed estimation and the stochastic inference methods were verified with simulated data and measured real channel data of the digital satellite broadcasting. The results show that our method improved the reception. The computational cost is reasonable enough for implementation.

Although the blind channel estimation and the stochastic inference are shown in some former studies [9, 19, 18, 11], this paper newly shows two items. One is the practical channel model for the mobile digital satellite broadcasting reception. The reference [16] gives suggestions about a multipath channel in Ku-band land mobile satellite broadcasting. The multipath of twelve symbol time delay is considered in the reference. We consider the direct path and one symbol delay path for the channel model and deal the multipath of more than one symbol time delay as the noise. Our consideration gives a simple model which is enough for the practical use. The other is the specific computational expression of the HOS under the channel model. Although an idea of the HOS is shown in the reference [9], the specific computational expression is not explained. We show the constrained conditions to obtain the unique solution under the channel model.

This paper is organized as follows. Section 2 describes the digital satellite broadcasting system and the problem of mobile reception. Section 3 outlines the proposed method. Section 4 gives details of the channel description and two channel estimation methods. Section 5 shows two approaches for code inference. Section 6 presents experimental results for the numerical simulation and real mobile reception. Finally, section 7 concludes the paper with a summary and some discussion.

## 2 Digital Satellite Broadcasting

### 2.1 Encoding and Modulation

First, we describe the encoding and modulation processes of digital satellite broadcasting. In this paper, we focus on two standards for digital satellite broadcasting (ARIB STD-B20 and DVB-S), which are the standards for NHK BS digital.

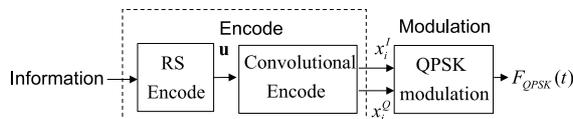


Figure 1: Encoding and Modulation.

The encoding and modulation processes are schematically shown in Fig. 1. First, information bits are encoded with the Reed-Solomon (RS) code [6] and the convolutional code. The RS code is based on Galois field  $GF(2^8)$ , whose element is 1 byte = 8 bits. The RS(204, 188) code is used, where each 188 bytes is encoded into 204 bytes. Basically, errors of 8 bytes can be corrected with this RS code. One block of RS(204, 188) is 204 bytes =  $8 \times 204 = 1632$  bits. The outputs of the RS code are interleaved and concatenated with a convolutional code, whose constraint length is 7. The code rate of the convolutional code is  $7/8$ . Finally, the output of the convolutional code is modulated with quadrature phase shift keying (QPSK). In this paper, we focus on the channel properties and the convolutional code and do not discuss the RS code in detail.

Let  $\mathbf{u} = (u_0, u_1, \dots, u_{m-1}) \in \{1, 0\}^m$  be the interleaved output of the RS code, where  $m > 7$ . Let us define the  $i$ -th output of the convolutional code as  $x_i = (x_i^I, x_i^Q) \in \{-1, +1\}^2$ , which is defined as follows:

$$\begin{aligned} x_i^I &= 1 - 2(u_i + u_{i-1} + u_{i-2} + u_{i-3} + u_{i-6}) \\ x_i^Q &= 1 - 2(u_i + u_{i-2} + u_{i-3} + u_{i-5} + u_{i-6}). \end{aligned} \quad (1)$$

Neglecting the further puncturing process, we define eq. (1) as the encoding of the convolutional code. Finally,  $x_i^I \in \{-1, +1\}$  and  $x_i^Q \in \{-1, +1\}$  are modulated by QPSK defined as

$$F_{QPSK}(t) = b[x_i^I \cos(2\pi f_c t) + x_i^Q \sin(2\pi f_c t)], \quad iT_s \leq t < (i+1)T_s, \quad (2)$$

where  $T_s$  is the duration of each bit (43.0404 ns),  $f_c$  is the carrier frequency, and  $b$  denotes the amplitude. In the case of NHK BS digital,  $f_s \equiv 1/T_s$ , which we call the ‘‘symbol rate’’ is 23.234 MHz,  $f_c$  is around 12 GHz, and  $f_c$  is divisible by  $f_s$ .

## 2.2 Static Reception

Digital satellite broadcasting is generally received fairly well from a fixed place with a BS/CS (broadcasting satellite and communication satellite) receiving dish antenna having a rather large bore diameter (40 to 50 cm). The received CNR range of static reception is very high (more than 14 dB) because of the parabolic antenna. A conventional method of demodulation and decoding is shown in Figure 2. The

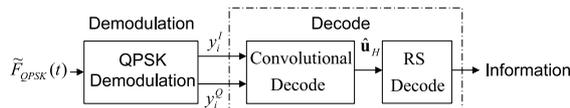


Figure 2: Conventional Demodulation and Decoding.

coherent QPSK demodulation is defined as the time average of the products of  $\cos(2\pi f_c t)$  or  $\sin(2\pi f_c t)$  and the received signal  $\tilde{F}_{QPSK}(t)$ . Let us define  $\mathbf{y}_i = (y_i^I, y_i^Q)^T$  as follows:

$$y_i^I = B \int_{iT_s}^{(i+1)T_s} \tilde{F}_{QPSK}(t) \cos(2\pi f_c t) dt, \quad y_i^Q = B \int_{iT_s}^{(i+1)T_s} \tilde{F}_{QPSK}(t) \sin(2\pi f_c t) dt, \quad (3)$$

where  $B$  is a factor for making  $y_i^{I^2} + y_i^{Q^2} = 2$ . Each component of  $\mathbf{y}_i$  is not binary in practice because of the noise. However, in the case of static reception, the difference between  $\mathbf{y}_i$  and  $\mathbf{x}_i$  is small. In the following, we define  $\mathbf{x}$  as the sequence of  $x_q, \dots, x_{q+N}$ . In a similar manner,  $\mathbf{y}$  and  $\mathbf{u}$  show the sequence. The basic idea for decoding  $\mathbf{y}$  to  $\hat{\mathbf{u}}_H$  [15] is summarized in the following equation:

$$\hat{\mathbf{u}}_H = \underset{\mathbf{u}}{\operatorname{argmin}} d_H(\operatorname{sign}(\mathbf{y}), \mathbf{x}(\mathbf{u})), \quad (4)$$

where  $d_H(\cdot, \cdot)$  is the Hamming distance,  $\mathbf{x}(\mathbf{u})$  is the function defined in eq. (1), and  $\operatorname{argmin}_{\mathbf{u}}$  defines the  $\mathbf{u}$  that minimizes the quantity. In eq. (4), the  $\mathbf{u}$  that minimizes the total Hamming distance is selected. In the following, we denote the decoded result of  $\mathbf{y}$  as  $\hat{\mathbf{u}}_H$ . Some errors in  $\hat{\mathbf{u}}_H$  are corrected by the RS decoding, and the final error rate is very low in static reception.

Since the reception antennas for static reception are large enough to focus only on the direct path, it is possible to assume that the channel is stationary and memoryless. Under this assumption, the received information about each bit can be regarded as an independent observation, and the decoding with Hamming distance works well.

## 2.3 Mobile Reception

Compared with static reception, mobile reception of digital satellite broadcasting is difficult. We can see this from the fact that there are very few commercial products for mobile reception.

The difficulty arises from two main reasons: the Doppler shift and the fading multipath channel [15]. We explain these first.

In the mobile reception, the relative velocity between the satellite and the vehicle is not 0, so it produces a Doppler shift. If a vehicle moves at 100 km/h toward the satellite, the Doppler shift is around 1 kHz because the carrier frequency is 12 GHz. This is much smaller than the symbol rate, which is  $f_s = 23$  MHz, but the Doppler shift influences the QPSK demodulation. One of the general approaches to reducing the influence of the Doppler shift is to use a phase locked loop (PLL).

In the case of static reception, dish antennas are used. Since they have narrow directivities, the influence of the multipath channel can be neglected. However, in mobile reception, antennas are generally small because they must be installed in a vehicle and their directivities are poor. We assume that the direct path between the satellite and the vehicle is not occluded, but that the influence of the multipath channel is not negligible. Moreover, it is natural to assume that the characteristics of the channel change as the vehicle moves.

Next, let us discuss how these problems are solved in commercial products. Among the few commercial digital satellite broadcasting receivers for vehicles, STRADA (Panasonic) is the one with a reasonable price. Although details of the receiver are not available, it seems (from personal communication) the system uses a fast PLL. A PLL resets the influence of the Doppler shift at every acquisition time and if the PLL is fast enough, the influence of the Doppler shift can be neglected.

To reduce the influence of the fading channel, Panasonic recommends that customers should use an antenna with a tracking system that has narrow directivity, which makes the main path is dominant and reflections small.

The strategy of using a PLL and an antenna with a tracking system for the Doppler shift and fading multipath channel, respectively, is effective.

## 3 Proposed Method

### 3.1 Our Strategy

The ultimate goal of our research is to make a compact receiving system with a reasonable price. We do not propose any change to the broadcasting standard. We receive the signal that is currently available from the satellite and improve the quality of the received broadcasting.

Our strategy is different from the available commercial products. To reduce the influence of the Doppler shift, we also use a fast PLL, but we do not use a tracking antenna because an antenna with a tracking system is rather large and expensive.

According to discussions of the ideal next-generation mobile reception antennas [16], a three-beam antenna that covers 360 degree (each beam is 120 degree) in azimuth angle has good characteristics [7]. Antennas of this type are compact enough to be installed to vehicles. However, the CNR range of such antennas will be relatively low [16]. Another study indicates that the next-generation antenna gain for mobile reception will be around 25 dBi to 37 dBi [17]. What is worse is that since the directivities of such antennas are poor, it is not possible to ignore the reflected waves even if the carrier frequency is high (over 10 GHz) and the channel becomes a fading multipath channel with delays [16, 13, 15].

One of the commonly used approaches for channels with memory is to install multiple antennas for diversity reception. However, it is not possible to install multiple digital satellite broadcasting antennas in a vehicle. Another commonly used approach is to utilize pilot sequences. These enable the characteristics of the fading channel to be estimated. However, the NHK digital satellite broadcasting has only one fixed word (8 bits) in each block of RS(204, 188), and this is not long enough as the pilot sequence.

The method that we propose does not rely on diversity reception or pilot sequences, but uses the stochastic channel model and recovers the information bits by using stochastic inference.

### 3.2 Channel Model and Stochastic Inference

The channel model is generally described as a stochastic model  $p(\mathbf{y}|\mathbf{x}(\mathbf{u}))$ , which is the conditional probability of  $\mathbf{y}$  when  $\mathbf{x}(\mathbf{u})$  is given. When  $\mathbf{y}$  is observed, a natural inference on  $\mathbf{u}$  is to take  $\hat{\mathbf{u}} = \operatorname{argmax}_{\mathbf{u}} p(\mathbf{x}(\mathbf{u})|\mathbf{y})$  by assuming that the prior distribution of  $\mathbf{x}$  is uniform as  $p(\mathbf{x}) = 1/2^{2N}$  ( $N$  is the length of  $\mathbf{x}$ ) and using the Bayes theorem:

$$p(\mathbf{x}(\mathbf{u})|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x}(\mathbf{u}))p(\mathbf{x})}{\sum_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}(\mathbf{u}))p(\mathbf{x})} \propto p(\mathbf{y}|\mathbf{x}(\mathbf{u})).$$

Therefore,

$$\hat{\mathbf{u}} = \underset{\mathbf{u}}{\operatorname{argmax}} p(\mathbf{x}(\mathbf{u})|\mathbf{y}) = \underset{\mathbf{u}}{\operatorname{argmax}} p(\mathbf{y}|\mathbf{x}(\mathbf{u})). \quad (5)$$

Moreover, if the channel is a binary symmetric channel (BSC) with flipping probability  $\alpha < 1/2$ , which is a typical memoryless channel, then

$$p(\mathbf{y}|\mathbf{x}(\mathbf{u})) = \prod_i p(y_i|x_i(\mathbf{u})) = \prod_i \alpha^{d_H(y_i, x_i(\mathbf{u}))} (1-\alpha)^{2-d_H(y_i, x_i(\mathbf{u}))} \propto \left(\frac{1-\alpha}{\alpha}\right)^{-d_H(\mathbf{y}, \mathbf{x}(\mathbf{u}))}.$$

Since  $\alpha < 1/2$ ,  $((1-\alpha)/\alpha) > 1$ , the maximization of  $p(\mathbf{y}|\mathbf{x}(\mathbf{u}))$  is equivalent to the minimization of  $d_H(\mathbf{y}, \mathbf{x}(\mathbf{u}))$  regardless of  $\alpha$ . Therefore eq. (5) becomes equivalent to eq. (4).

When the channel has memory, the bits of  $\mathbf{y}$  are not independent and eq. (4) does not work. We still use eq.(5), based on the stochastic channel model with memory. What we need to discuss below is how to build the channel model and how to utilize it to solve eq. (5).

In the case of a BSC, parameter  $\alpha$  does not have any influence on the decoding algorithm, but in the case of fading channels, the parameters must be estimated and used to solve eq. (5). The channel model and parameter estimation methods are discussed in detail in section 4. We use the stochastic model with adjacent inter symbol interference (ISI). We explain why we chose this model in section 4. The parameter estimation of the channel is not simple because the channel input sequence  $\mathbf{u}$  is hidden from the observed signal  $\mathbf{y}$ . We discuss two methods [19, 9]: MLE and HOS matching. For MLE, we use the EM algorithm[4]. Although the EM algorithm is simple, iterative computation is inevitable and this takes time. The estimation based on HOS may not be as stable as that of MLE, but the computational cost is very low, which is better for practical implementation. We compare them in section 6.

When  $\mathbf{y}$  is observed, the channel model gives us  $p(\mathbf{x}(\mathbf{u})|\mathbf{y})$ . In section 5, we discuss two methods of stochastic inference ([18, 11]) based on  $p(\mathbf{x}(\mathbf{u})|\mathbf{y})$ .

One is to use the marginal distribution  $p(x_i(\mathbf{u})|\mathbf{y})$  and compute  $\mathbf{u}$  based on it (marginal inference). The other is to use the joint probability of  $p(\mathbf{x}(\mathbf{u})|\mathbf{y})$  to compute  $\mathbf{u}$  (joint inference) (Fig. 3). In each case, the decoded codeword of  $\mathbf{u}$  is further decoded by the RS decoder (Fig. 3). Since the computational cost is slightly different, we compare the costs and results of the methods.

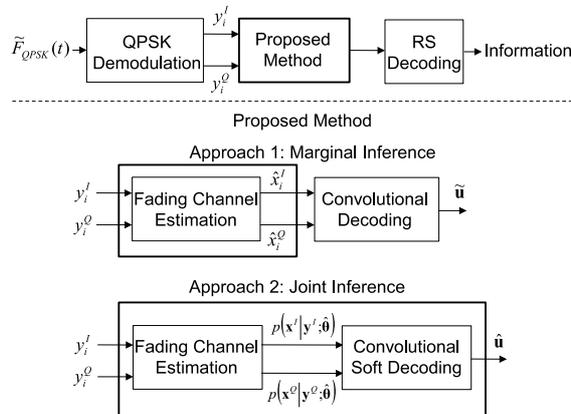


Figure 3: Schematic Diagram of the Proposed Method.

For the following discussion, we define variables  $\hat{\mathbf{x}}$ ,  $\tilde{\mathbf{u}}$ , and  $\hat{\mathbf{u}}$  as follows.

$\hat{\mathbf{x}}$  : The stochastic inference of the hard bits based on the fading channel estimation.

$\tilde{\mathbf{u}}$  : The stochastic inference on  $\mathbf{u}$  based on  $p(x_i(\mathbf{u})|\mathbf{y})$ .

$\hat{\mathbf{u}}$  : The stochastic inference on  $\mathbf{u}$  based on  $p(\mathbf{x}(\mathbf{u})|\mathbf{y})$ .

## 4 Fading Channel and Estimation

### 4.1 Model of Fading Channel

Let us assume that  $\mathbf{y}$  does not have a Doppler shift because a fast PLL is used, but that the influence of the fading channel remains. Our problem is to remove the influence of the fading channel. More precisely, we would like to infer the probability of the code sequence  $\mathbf{u}$  from observations  $\mathbf{y}$ . First, we build the channel model. The autocorrelations of  $\mathbf{y}$ , which were measured in twelve real routes with a moving vehicle, are shown in Fig. 4 (details are given in section 6). Since  $\mathbf{x}$  is the output of an interleaver, we expect its autocorrelation to be 0 for  $\tau \neq 0$ . However, the result shows that  $\mathbf{y}$  has correlations for  $\tau \neq 0$ . This indicates that the channel with memory has ISI. Although the channel can have longer correlations, Fig. 4 shows that the main contribution to the fading channel is the adjacent interfering path. On the

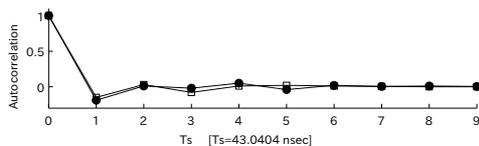


Figure 4: Autocorrelations of  $\{y_i\}$  with the Measured Data.

basis of this result, we assume that the number of multiple paths is two, that is, the main path and one adjacent interfering path. We also assume additive Gaussian noise. The mathematical formulation of the channel is given by

$$y_i \simeq b_0 x_i + b_1 x_{i-1} + w_i, \quad w_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_4), \quad (6)$$

where  $x_i = (x_i^I, x_i^Q)$ ,  $y_i = (y_i^I, y_i^Q)$ . The parameters  $b_0$  and  $b_1$  are coefficients,  $\mathcal{N}$  is the density function of the normal distribution,  $\sigma^2$  is the variance of the noise,  $I_4$  is the 4 dimensional identity matrix and  $w_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_4)$  means  $w_i$  is generated from the density function of the normal distribution. Even though this model is simple, the experiments described in section 6 show that it greatly improves the results. Our ultimate goal is real-world application, where realtime processing is necessary. The computational cost of estimation and inference with this simple model is low, so it can easily be implemented in a realtime system.

Equation (6) gives the distribution of  $y_i$  when  $x_i$  and  $x_{i-1}$  are given. In other words, let  $\theta = (b_0, b_1, \sigma^2)$  and

$$y_i \sim p(y_i | x_i, x_{i-1}; \theta) = \mathcal{N}(b_0 x_i + b_1 x_{i-1}, \sigma^2 I_4). \quad (7)$$

Where  $y_i \sim p(y_i | x_i, x_{i-1}; \theta)$  means  $y_i$  is generated from the probability function  $p(y_i | x_i, x_{i-1}; \theta)$ .

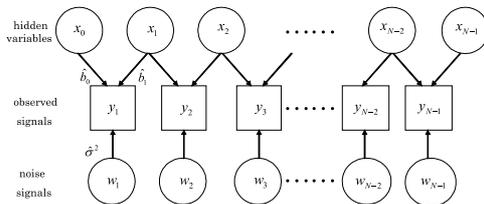


Figure 5: Graphical Model of the Channel.

This is the channel model and Fig. 5 shows it as a graphical model. On the basis of the discussion in section 4.1, we define the channel model as in eq. (7), and the parameter  $\theta$  must be estimated from data in order to perform further inferences. For the estimation of the channel parameters  $\theta = (b_0, b_1, \sigma^2)$ , we compare two methods: MLE and HOS matching.

### 4.2 Channel estimation by EM algorithm

Let us start with MLE. Since the channel input data  $\mathbf{x}$  is hidden from the observed signal  $\mathbf{y}$ , we use the EM algorithm[4] to compute the MLE. We split the data into blocks and estimate the parameters for

each block. The block length is set to  $N = 2000$ . Let block  $\mathbf{y} = \{y_1, \dots, y_N\}$ . For the estimation, we use MLE, which is defined as

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} p(\mathbf{y}; \boldsymbol{\theta}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \log p(\mathbf{y}; \boldsymbol{\theta}).$$

The likelihood  $p(\mathbf{y}; \boldsymbol{\theta})$  is rewritten as

$$p(\mathbf{y}; \boldsymbol{\theta}) = \sum_{\mathbf{x}} p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}) = \sum_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta})p(\mathbf{x}).$$

From eq. (7),  $p(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta})$  is easily rewritten as

$$p(\mathbf{y}|\mathbf{x}; \boldsymbol{\theta}) = \prod_{i=1}^N p(y_i | x_i, x_{i-1}; \boldsymbol{\theta}),$$

and the natural choice of  $p(\mathbf{x})$ , which is the prior of  $\mathbf{x}$ , is a uniform distribution  $p(\mathbf{x}) = 1/2^{2(N+1)}$ . Now the distributions  $p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta})$  and  $p(\mathbf{y}; \boldsymbol{\theta})$  become

$$p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}) = \frac{1}{2^{2(N+1)}} \prod_{i=1}^N p(y_i | x_i, x_{i-1}; \boldsymbol{\theta}) \quad (8)$$

$$p(\mathbf{y}; \boldsymbol{\theta}) = \frac{1}{2^{2(N+1)}} \sum_{\mathbf{x}} \prod_{i=1}^N p(y_i | x_i, x_{i-1}; \boldsymbol{\theta}). \quad (9)$$

This is a mixture of normal distributions, and  $\mathbf{x}$  is the hidden stochastic variable. In this situation, one widely used method for MLE is the EM algorithm. We use the EM algorithm to compute the MLE of  $\boldsymbol{\theta} = (b_0, b_1, \sigma^2)$ . The EM algorithm is a simple iterative method that consists of E- and M-steps for maximizing the log-likelihood function [4]. The algorithm is as follows.

**E-step** In the  $r$ -th round of the E-step, a  $Q$  function, which is defined below, is computed.

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^r) = \frac{1}{N} \sum_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}; \boldsymbol{\theta}^r) \log p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta})$$

From eqs. (7) and (8),  $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^r)$  is rewritten as

$$\begin{aligned} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^r) &= -\frac{1}{N} \sum_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}; \boldsymbol{\theta}^r) \frac{|y_i - (b_1 x_{i-1} + b_0 x_i)|^2}{2\sigma^2} - \log 2\pi\sigma^2 \\ &= -\frac{1}{N} \sum_{i=1}^N \frac{y_i^2}{2\sigma^2} - \frac{b_1^2 + b_0^2}{\sigma^2} - \log 2\pi\sigma^2 + \frac{2b_1}{\sigma^2} C_{\mathbf{y}\mathbf{x}_{-1}}^r + \frac{2b_0}{\sigma^2} C_{\mathbf{y}\mathbf{x}}^r - \frac{2b_0 b_1}{\sigma^2} C_{\mathbf{x}\mathbf{x}_{-1}}^r, \end{aligned} \quad (10)$$

where the fact  $|x_i|^2 = 2$  is used and  $C_{\mathbf{y}\mathbf{x}_{-1}}^r$ ,  $C_{\mathbf{y}\mathbf{x}}^r$ , and  $C_{\mathbf{x}\mathbf{x}_{-1}}^r$  are defined as follows:

$$\begin{aligned} C_{\mathbf{y}\mathbf{x}_{-1}}^r &= \frac{1}{2N} \sum_{i=1}^N \sum_{x_{i-1}} p(\mathbf{x}|\mathbf{y}; \boldsymbol{\theta}^r) y_i \cdot x_{i-1}, \quad C_{\mathbf{y}\mathbf{x}}^r = \frac{1}{2N} \sum_{i=1}^N \sum_{x_i} p(\mathbf{x}|\mathbf{y}; \boldsymbol{\theta}^r) y_i \cdot x_i \\ C_{\mathbf{x}\mathbf{x}_{-1}}^r &= \frac{1}{2N} \sum_{i=1}^N \sum_{x_{i-1}, x_i} p(\mathbf{x}|\mathbf{y}; \boldsymbol{\theta}^r) x_{i-1} \cdot x_i. \end{aligned} \quad (11)$$

It is necessary to compute  $C_{\mathbf{y}^N \mathbf{x}_{-1}}^r$ ,  $C_{\mathbf{y}^N \mathbf{x}_{-1}}^r$ , and  $C_{\mathbf{x}\mathbf{x}_{-1}}^r$  at every iteration. For the computation, we can utilize the belief propagation (BP) algorithm. Details are given in section A.

**M-step** At each iteration, the M-step updates  $\boldsymbol{\theta}$  as

$$\boldsymbol{\theta}^{r+1} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^r).$$

If we equate  $\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}^r)/\partial \boldsymbol{\theta}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{r+1}} = \mathbf{0}$ , it becomes

$$b_0^{r+1} = \frac{C_{\mathbf{y}\mathbf{x}}^r - C_{\mathbf{y}\mathbf{x}_{-1}}^r C_{\mathbf{x}\mathbf{x}_{-1}}^r}{1 - (C_{\mathbf{x}\mathbf{x}_{-1}}^r)^2}, \quad b_1^{r+1} = \frac{C_{\mathbf{y}\mathbf{x}_{-1}}^r - C_{\mathbf{y}\mathbf{x}}^r C_{\mathbf{x}\mathbf{x}_{-1}}^r}{1 - (C_{\mathbf{x}\mathbf{x}_{-1}}^r)^2}$$

$$\sigma^{r+1^2} = \frac{1}{2N} \sum_{i=1}^N |y_i|^2 + b_1^{r+1^2} + b_0^{r+1^2} - 2b_1^{r+1} C_{\mathbf{y}\mathbf{x}_{-1}}^r - 2b_0^{r+1} C_{\mathbf{y}\mathbf{x}}^r + 2b_0^{r+1} b_1^{r+1} C_{\mathbf{x}\mathbf{x}_{-1}}^r.$$

At  $r = 0$ ,  $\boldsymbol{\theta}^0$  is initialized as  $\hat{b}_0^0 = 0.8(2 - \hat{\sigma}^2)$ ,  $\hat{b}_1^0 = 0.2(2 - \hat{\sigma}^2)$  and  $\hat{\sigma}^2{}^0 = 0.3$ . In practice, the E- and M-steps are repeated and the algorithm with the initial condition converges well with a few to 20 iterations. Finally, the parameter is estimated as  $\hat{\boldsymbol{\theta}} = (\hat{b}_0, \hat{b}_1, \hat{\sigma}^2)$ .

### 4.3 Channel estimation by Higher Order Statistics

Next, we explain the HOS matching. From the model in eq. (6) and the assumptions that  $x_i$  is independent, its mean is 0, and noise is independent additive Gaussian noise, we have the following relations:

$$C_{\mathbf{y}\mathbf{y}} = 2(b_0^2 + b_1^2 + \sigma^2), \quad C_{\mathbf{y}\mathbf{y}_{-1}} = 2b_0 b_1,$$

$$C_{\mathbf{y}^4} = 4\left\{(b_0^2 + b_1^2 + \sigma^2)^2 + 4\sigma^2(b_0^2 + b_1^2 + \sigma^2) - 2\sigma^4 + 4b_0^2 b_1^2\right\} = 4\left\{2\sigma^2 C_{\mathbf{y}\mathbf{y}} + \frac{1}{2}C_{\mathbf{y}\mathbf{y}}^2 + C_{\mathbf{y}\mathbf{y}_{-1}}^2 - \sigma^4\right\} \quad (12)$$

where we used the fact that  $E_{p(z)}[z^4] = 3\sigma^4$  if  $z \sim \mathcal{N}(0, \sigma^2)$  and the following quantity

$$C_{\mathbf{y}\mathbf{y}} = E_{p(\mathbf{y})}[y_i^{Q^2} + y_i^{I^2}], \quad C_{\mathbf{y}^4} = E_{p(\mathbf{y})}[y_i^{Q^4} + y_i^{I^4}], \quad C_{\mathbf{y}\mathbf{y}_{-1}} = E_{p(\mathbf{y})}[y_i^I y_{i-1}^I + y_i^Q y_{i-1}^Q].$$

In practice, we replace these with the following example means.

$$\hat{C}_{\mathbf{y}\mathbf{y}} = \frac{1}{N} \sum_i (y_i^{Q^2} + y_i^{I^2}), \quad \hat{C}_{\mathbf{y}^4} = \frac{1}{N} \sum_i (y_i^{Q^4} + y_i^{I^4}), \quad \hat{C}_{\mathbf{y}\mathbf{y}_{-1}} = \frac{1}{N} \sum_i (y_i^I y_{i-1}^I + y_i^Q y_{i-1}^Q).$$

The unknown parameters  $b_0$ ,  $b_1$ , and  $\sigma^2$  can be estimated by solving the set of equations in eq. (12). It is easy to check that the solutions are not unique from the fact that the equations in eq. (12) are symmetric for  $b_0$  and  $b_1$ .

Since  $b_0$  shows the signal strength from the direct path, we impose two constraints:  $b_0 > 0$  and  $b_0 > |b_1|$ .

$$\hat{\sigma}^2 = \hat{C}_{\mathbf{y}\mathbf{y}} - \sqrt{\frac{3}{2}\hat{C}_{\mathbf{y}\mathbf{y}}^2 + 2\hat{C}_{\mathbf{y}\mathbf{y}_{-1}}^2 - \hat{C}_{\mathbf{y}^4}},$$

$$\hat{b}_0 = \frac{1}{\sqrt{2}} \left( B^2 + \sqrt{B^2 - \hat{C}_{\mathbf{y}\mathbf{y}_{-1}}^2} \right)^{1/2}, \quad \hat{b}_1 = \frac{\hat{C}_{\mathbf{y}\mathbf{y}_{-1}}}{2\hat{b}_0}, \quad \text{where } B = \frac{1}{2}\hat{C}_{\mathbf{y}\mathbf{y}} - \hat{\sigma}^2.$$

This method strongly depends on the assumption that the noise is Gaussian, and the solution may not be as stable as that of MLE. However, the computation is straightforward and its cost is low. We compare these two methods through experiments.

## 5 Inference of Convolutional Code

When the parameters of the channel are estimated with the MLE or HOS matching methods, the channel model  $p(\mathbf{y}|\mathbf{x}(\mathbf{u}); \boldsymbol{\theta})$  is given. We would like to make the stochastic inference on  $\mathbf{u}$ , which is the inputs of the convolutional code, based on  $p(\mathbf{x}(\mathbf{u})|\mathbf{y}; \boldsymbol{\theta})$ .

In the following, we propose two approaches for stochastic inference. The first approach is to apply stochastic inference to the hard bits of  $\mathbf{x}$  and decode them by the conventional convolutional decoding. We call this approach the marginal inference approach. The other approach is to apply stochastic inference directly to the input sequence of the convolutional encoding  $\mathbf{u}$ . We show the graphical model of the transmitting convolutional code words and the channel model. And the stochastic inference on the convolutional code words is based on the graphical model. We call this method the joint inference approach. Both approaches are discussed below.

## 5.1 Marginal Inference Approach

In the marginal inference approach, we first estimate  $\mathbf{x}$  based on the marginal probability  $p(x_i(\mathbf{u})|\mathbf{y})$ . Let  $\hat{\mathbf{x}}$  be the inference of the sequence  $\mathbf{x}$ . With an abuse of the notation, the inference  $\hat{\mathbf{x}}$  is defined as follows:

$$\hat{x}_i = \text{sign } \bar{x}_i, \quad \bar{x}_i = \sum_{x_i} x_i p(x_i|\mathbf{y}; \hat{\boldsymbol{\theta}}) \quad (13)$$

This is the maximization of the posterior marginals; that is,

$$\hat{x}_i = \underset{x_i}{\text{argmax}} p(x_i|\mathbf{y}; \hat{\boldsymbol{\theta}}).$$

The BP algorithm (A) is an efficient method for computing  $\bar{x}_i$ . To decode  $\hat{\mathbf{x}}$  to  $\tilde{\mathbf{u}}$ , we use eq. (4):

$$\tilde{\mathbf{u}} = \underset{\mathbf{u}}{\text{argmin}} d_H(\text{sign}(\hat{\mathbf{x}}), \mathbf{x}(\mathbf{u})). \quad (14)$$

The standard decoding algorithm of convolutional codes solves this problem. We define  $\tilde{\mathbf{u}}$  as the decoded word of this method.

## 5.2 Joint Inference Approach

In the joint inference approach, the stochastic inference on  $\mathbf{u}$  is done based on  $p(\mathbf{x}(\mathbf{u})|\mathbf{y}; \hat{\boldsymbol{\theta}})$ . The graphical model representation is shown in Fig. 6, where code words  $\mathbf{u}$  as well as  $\mathbf{x}$  are the hidden stochastic variables. We infer code  $\hat{\mathbf{u}}$  from the posterior distribution given observed signal  $\mathbf{y}$ .

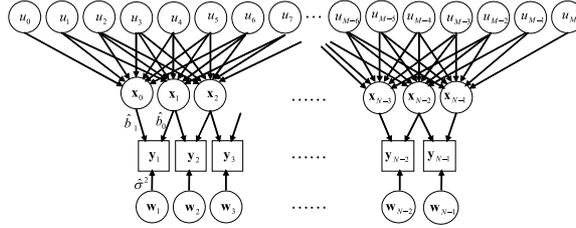


Figure 6: Graphical Model of Convolutional Code.

The idea of the inference is to find the path in the trellis that maximizes  $p(\mathbf{u}|\mathbf{y}; \hat{\boldsymbol{\theta}})$  as

$$\hat{\mathbf{u}} = \underset{\mathbf{u}}{\text{argmax}} p(\mathbf{u}|\mathbf{y}; \hat{\boldsymbol{\theta}}) = \underset{\mathbf{u}}{\text{argmax}} \sum_{\mathbf{x}} p(\mathbf{u}|\mathbf{x}) p(\mathbf{x}|\mathbf{y}; \hat{\boldsymbol{\theta}}). \quad (15)$$

From eq. (8), we have  $p(\mathbf{x}|\mathbf{y}; \boldsymbol{\theta})$  as

$$p(\mathbf{x}|\mathbf{y}; \hat{\boldsymbol{\theta}}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x}; \hat{\boldsymbol{\theta}})}{\sum_{\mathbf{x}} p(\mathbf{x})p(\mathbf{y}|\mathbf{x}; \hat{\boldsymbol{\theta}})} = \frac{1}{Z} \prod_{i=1}^{N-1} p(y_i|x_i, x_{i-1}; \hat{\boldsymbol{\theta}}) = \frac{1}{Z'} \prod_{i=1}^{N-1} \psi_i(x_i, x_{i-1}),$$

where

$$Z = 2^{2N} \sum_{\mathbf{x}} p(\mathbf{y}, \mathbf{x}; \hat{\boldsymbol{\theta}}), \quad \psi_i(x_i, x_{i-1}) = \exp \left[ -\frac{|y_i - (b_0 x_i + b_1 x_{i-1})|^2}{2\sigma^2} \right]$$

$$Z' = (2\pi\sigma^2)^{N-1} Z = \sum_{\mathbf{x}} \prod_{i=1}^{N-1} \psi_i(x_i, x_{i-1}).$$

We use the BP algorithm[14] to compute  $p(\mathbf{x}|\mathbf{y}; \hat{\boldsymbol{\theta}})$  efficiently. In the following, we rewrite  $p(\mathbf{u}|\mathbf{y}; \hat{\boldsymbol{\theta}})$  as  $p(\mathbf{u})$  and denote  $\mathbf{u}_r^q$  for  $u_r, \dots, u_q$ , ( $q > r$ ) for simplicity because the joint distribution  $p(\mathbf{u}_0^{M-1})$  is defined as

$$p(\mathbf{u}_0^{M-1}) = p(\mathbf{u}_0^6) p(u_7|\mathbf{u}_0^6) \cdot \dots \cdot p(u_{M-1}|\mathbf{u}_{M-8}^{M-2})$$

$$\log p(\mathbf{u}_0^{M-1}) = \log p(\mathbf{u}_0^6) + \log p(u_7|\mathbf{u}_0^6) + \dots + \log p(u_{M-1}|\mathbf{u}_{M-8}^{M-2}).$$

The individual conditional probabilities are defined as

$$p(u_i|\mathbf{u}_{i-7}^{i-1}) = \frac{p(\mathbf{u}_{i-7}^i)}{p(\mathbf{u}_{i-7}^{i-1})}.$$

The trellis for the computation of  $\hat{\mathbf{u}}$  is shown in Fig. 7. Figure 7 starts from the initial status  $\log p(\mathbf{u}_0^6)$  and then moves to the next status  $\log p(u_7|\mathbf{u}_0^6)$ . Final status is  $\log p(u_{M-1}|\mathbf{u}_{M-8}^{M-2})$ . The sequence  $\mathbf{u}_0^{M-1}$  is searched for in the trellis as the path that maximizes  $p(\mathbf{u}_0^{M-1})$ .

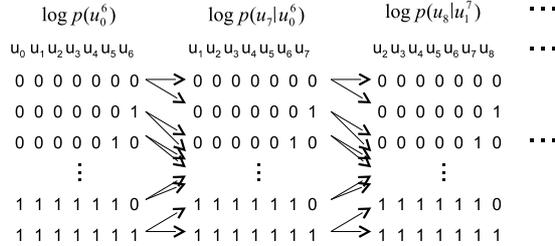


Figure 7: Convolutional Code Inference.

## 6 Numerical Experiments

### 6.1 Simulated Data

We have proposed two channel estimation methods (MLE and HOS matching) and two stochastic inference methods (marginal and joint inference). The combinations of each estimation and inference methods were tested in two numerical experiments: one with simulated data and the other with real mobile channel data.

First, we explain the experiment with simulated data. We generated data sets according to eq. (6). Each data set consists of 2000 data points, where  $b_0$  is fixed to 1.0,  $b_1$  is a sample taken from a uniform distribution of  $[-1/K, 1/K]$ , and the additive noise follows the Gaussian distribution as

$$\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_4).$$

Note that  $b_0$  and  $b_1$  are fixed for each data set, and we repeated the experiments 100 times for each  $\sigma$  by drawing  $b_1$  from a uniform distribution.

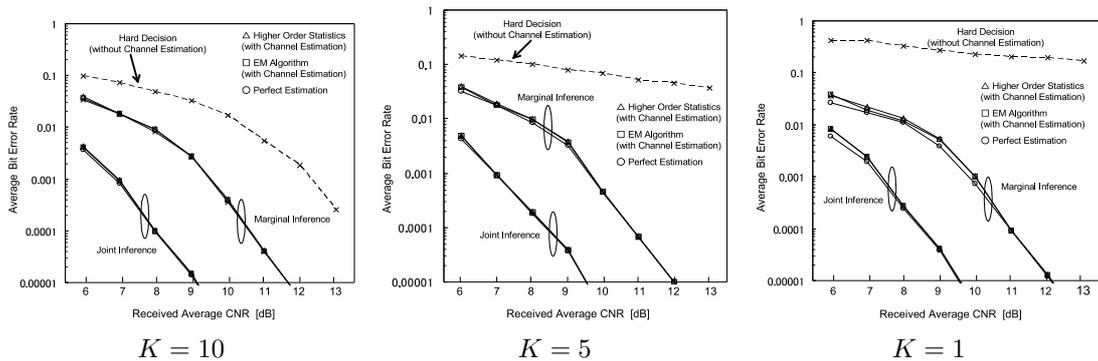


Figure 8: Error Performances of the Outputs of Convolutional Code.

The error performances of the convolutional decoding are shown in Fig. 8. Each graph shows the results for a different  $K$ . Each figure shows 7 results: combinations of the two channel estimation methods and two inference methods, two inference methods with true parameters (since we know the

channel model parameters in the simulation), and standard decoding methods without the channel model or joint inference (hard decision + Viterbi decoding).

As shown in Fig. 8, the method (MLE estimation or HOS matching) did not influence the results, but the joint inference worked better than the marginal inference. In each case, the results were close to those obtained using the true parameters.

The maximum bit error rate (BER) to let RS decoding have an error ratio of less than  $10^{-5}$  is typically  $2 \times 10^{-3}$ . Although wide range of CNR is simulated in the reference [16], we consider 8 dB to 10 dB of CNR is the range of our experiments. This range is relatively low and the method shown in the reference [16] cannot improve the error performances enough. In this range, decoding without channel estimation cannot remove the influence of the fading channel. The marginal inference approach can improve the error performances, but cannot achieve sufficient quality. The joint inference approach can remove the influence of the fading channel.

## 6.2 Measuring Real Channels on Twelve Routes

Next, we explain the experiment with the data for real channels of satellite broadcasting mobile reception. The measurement conditions are listed in Table 1.

Table 1: Measurement Conditions.

Satellite	BSAT-2a
Reception frequency	11.84256 GHz
Places	Tokyo and Kanagawa, Japan
Elevation angle	38.1 degree
Antenna gain	25 dBi
Date	February 2, 2006
Weather	Fine

We used a multi-beam antenna [7] and a fast PLL whose acquisition time is equal to the symbol time. In our experiments, we installed a commercially produced Lunenburg lens antenna [7] (Sumitomo Electric Industries, LuneQ-40<sup>1</sup>) on the roof of a vehicle. The measuring system consisted of a low noise amplifier module and an orthogonal signal recorder with a PLL whose acquisition time was one QPSK symbol time.

The channel was measured on a fine day while the vehicle was moving along the K1-Yokohane highway from Haneda toward Yokohama, the urban road around Ohta Ichiba in Shinagawa, and the city road around Aomonoyokocho in Shinagawa. The routes were chosen so as not to have any serious shadowing. We measured at five places along the highway (H1–H5), three places along the urban road (U1–U3), and four places along the city road (C1–C4) (Fig. 9).

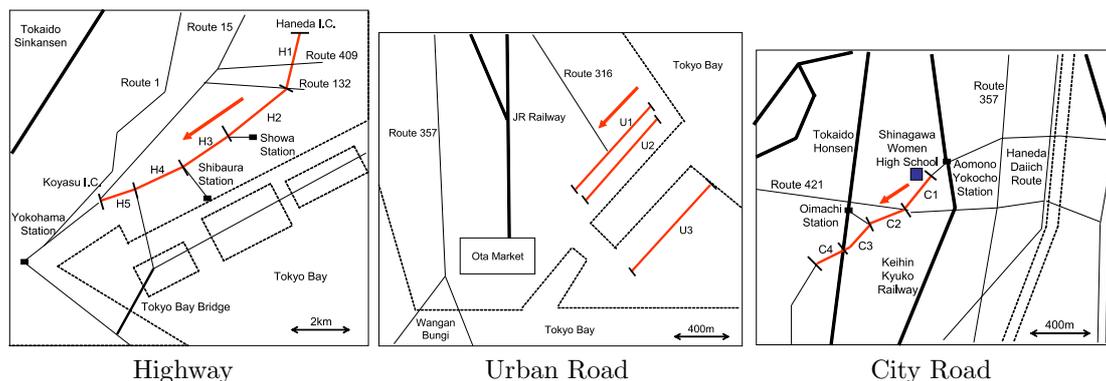


Figure 9: Measurement Routes (I.C.: interchange).

The autocorrelations for each route are shown in Fig. 10. The highway is an elevated expressway with

<sup>1</sup>Although this product is designed for static reception, the antenna was installed on a vehicle for the measurement.

few reflected waves. The urban road is surrounded by warehouses, and the city road is surrounded by office buildings, giving relatively many reflected waves and delays.

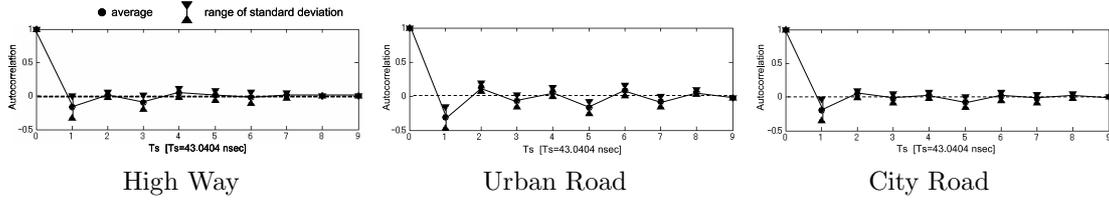


Figure 10: Examples of Autocorrelation.

In each route, the CNR was in the range of 8 to 10 dB. We recorded the frequency modulation (FM) signal broadcast from the satellite, where the original signal was known. The real channel was measured at the sampling rate 92.936 MHz and the channel was reconstructed in the form of autocorrelations, where the data rate was 23.234 MHz. Data for 120 s was collected. We used data for twelve real channel in the experiments (Fig. 11). In this case, the original data was known and the error rates were computed.

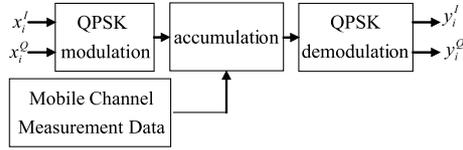


Figure 11: Block Diagram of the Measurement.

### 6.3 Reconstruction of the Mobile Channel

For the experiments, the measured mobile channels had to be reconstructed from the measured FM signals. An FM signal is generally expressed as

$$F_{FM}(t) = a \cos(2\pi(f_c + f_m(t))t),$$

where  $f_m(t)$  is the frequency of the transmission signal, which we know.

With the PLL, the mobile channel output signal of the FM signal, which is influenced by the Doppler shift and the fading channel, can be expressed as

$$\tilde{F}_{FM}(t) = \sum_{l=0}^{L_{FM}-1} a_l \cos(2\pi(f_c + f_m(t))(t - \tau_l) + \varpi_l(t)), \quad \varpi_l(t) = 2\pi f_{Dl}(t - \tau_l) + \phi_0,$$

where  $a_l$ ,  $\tau_l$ , and  $\varpi_l(t)$  are the amplitude, delay, and Doppler shift of the  $l$ th path, respectively. Since we know  $f_m(t)$ , the mobile channel characteristics are obtained by removing  $f_m(t)$  from the recorded signal. After that, the channel is approximated as a finite impulse response filter with 4096 taps, which is defined as

$$h_{FM}(t) \equiv \sum_{l=0}^{L_{FM}-1} a_l \cos(2\pi(f_c + f_{Dl})(t - \tau_l) + \phi_0) = \sum_{l=0}^{4095} a_l \cos(2\pi(f_c + f_{Dl})(t - l\hat{T}_S) + \phi_0).$$

Here,  $\hat{T}_S$  is given by the PLL. To get each multiple path amplitude and phase rotation, we multiply  $h_{FM}(t)$  by  $\cos(2\pi f_c(t - l\hat{T}_S))$  and obtain the signal observed through the low pass filter as

$$b_l \cos(2\pi f_{Dl}(t - l\hat{T}_S) + \phi_0).$$

The Hilbert transformation gives us

$$b_l \sin(2\pi f_{Dl}(t - l\hat{T}_S) + \phi_0). \quad (16)$$

In this way, the reproduced mobile channel is expressed as the filter. We can rewrite it as a filter from  $x_i$  to  $y_i$ , which is expressed as a matrix defined as follows:

$$H(t) \equiv \sum_{l=0}^{4095} \begin{pmatrix} b_l \cos \varpi_l(t) & b_l \sin \varpi_l(t) \\ -b_l \sin \varpi_l(t) & b_l \cos \varpi_l(t) \end{pmatrix}, \quad \varpi_l(t) = 2\pi f_{Dl}(t - l\hat{T}_s) + \phi_0.$$

The noise was also measured in the case of a different carrier, where no signal was sent. Letting the noise  $n(iT_s)$ , the received signal is reconstructed as

$$y_i = H(iT_s)x_i + n(iT_s).$$

## 6.4 Results

Here, we present the results obtained when our method was applied to the data for twelve mobile channels. First, to explain how the multiple paths are removed, we show the autocorrelations of  $\mathbf{y}$  and  $\hat{\mathbf{x}}$ , where  $\mathbf{y}$  is the reception data and  $\hat{\mathbf{x}}$  is the inference based on the fading channel, as expressed by eq. (13). The original codes were interleaved and the ideal autocorrelations were 0 except for  $\tau = 0$ . The autocorrelations for 3 out of 12 measured mobile channels (the results were similar for the other channels) are shown in Fig. 12. As  $\mathbf{y}$  was processed to  $\hat{\mathbf{x}}$ , the autocorrelations got closer to the expected function.

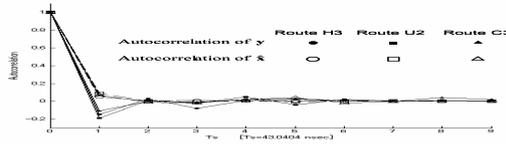


Figure 12: Autocorrelations.

Next we show how the proposed method improves the results. We show the experimental schematic diagrams in Fig.13. The results for  $\hat{\mathbf{u}}$  (the marginal inference approach) and  $\hat{\mathbf{u}}$  (the joint inference

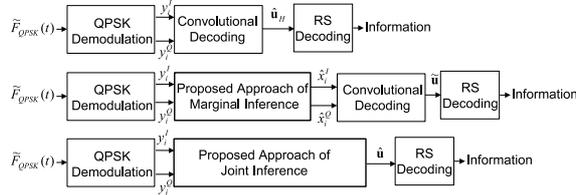


Figure 13: Schematic Diagram of Experiments.

approach) are shown in Tables 2 and 3. For both inference approaches, two channel estimation methods (HOS matching and the EM algorithm) were combined. The BER of the outputs of the convolutional decoding and the element errors of the RS decoding are shown in Tables 2 and 3, respectively. These tables also show the results with  $\hat{\mathbf{u}}_H$  (hard decision without channel estimation).

For NHK BS television reception, the quality of the received images was very good when the element error rate was less than  $2 \times 10^{-4}$  at the output of the RS decoding.

Comparing the marginal and joint inference approaches, the joint inference reached a good quality range in a practical CNR range (8 to 10 dB), while the marginal inference approach gave good results for some routes (e.g., H1, H2, and U1), but performed badly for others (e.g., H5, U3, and C4).

Comparing the HOS matching and the EM algorithm, although there were almost no differences for the numerical simulation, the EM algorithm worked slightly better than the HOS matching. The HOS matching strongly relies on the noise being Gaussian that  $E_{p(z)}[z^4] = 3\sigma^4$  if  $z \sim \mathcal{N}(0, \sigma^2)$ , each  $x$  is independent and the average of  $\mathbf{x}$  is zero. This may indicate that the noise distribution of a real channel is non-Gaussian. However, at the output of the RS decoder, there were almost no differences.

We give a quantitative comment of the computational cost by measuring the execution time of MATLAB program. Comparing the decoding time expressed in Eq.(4), although the method combined the MLE estimation with Joint Inference times from 2.2 to 20.2, the method combined the HOS estimation with Joint Inference times 1.3.

Table 2: BER of Convolutional Code.

Route	Average Received CNR [dB]	without Channel Estimation	with Channel Estimation				
		Hard Decision	Marginal Inference		Joint Inference		
			Higher Order	EM Algorithm	Higher Order	EM Algorithm	
High Way	H1	9.91	0.0034	0.0026	0.0022	<1E-04	<1E-04
	H2	9.47	0.0052	0.0038	0.0035	<1E-04	<1E-04
	H3	9.12	0.0052	0.0036	0.0035	<1E-04	<1E-04
	H4	8.68	0.0058	0.0045	0.0040	<1E-04	<1E-04
	H5	8.30	0.0065	0.0050	0.0043	<1E-04	<1E-04
Urban	U1	9.39	0.0054	0.0038	0.0038	<1E-04	<1E-04
	U2	9.19	0.0071	0.0061	0.0054	<1E-04	<1E-04
	U3	8.40	0.0073	0.0060	0.0051	<1E-04	<1E-04
City	C1	9.74	0.0054	0.0041	0.0037	<1E-04	<1E-04
	C2	9.38	0.0048	0.0036	0.0033	<1E-04	<1E-04
	C3	9.01	0.0088	0.0069	0.0067	<1E-04	<1E-04
	C4	8.10	0.0090	0.0072	0.0068	<1E-04	<1E-04

Table 3: Element Error Rate of RS Code.

Route	Average Received CNR [dB]	without Channel Estimation	with Channel Estimation				
		Hard Decision	Marginal Inference		Joint Inference		
			Higher Order	EM Algorithm	Higher Order	EM Algorithm	
High Way	H1	9.91	0.00058	<1E-04	<1E-04	<1E-04	<1E-04
	H2	9.47	0.0048	<1E-04	<1E-04	<1E-04	<1E-04
	H3	9.12	0.0063	0.00084	0.00084	<1E-04	<1E-04
	H4	8.68	0.0045	0.0016	0.00040	<1E-04	<1E-04
	H5	8.30	0.026	0.0078	0.0029	<1E-04	<1E-04
Urban	U1	9.39	0.0016	<1E-04	<1E-04	<1E-04	<1E-04
	U2	9.19	0.0058	0.0024	0.00061	<1E-04	<1E-04
	U3	8.40	0.027	0.0071	0.0036	<1E-04	<1E-04
City	C1	9.74	0.0023	0.00045	0.00011	<1E-04	<1E-04
	C2	9.38	0.0021	0.00030	<1E-04	<1E-04	<1E-04
	C3	9.01	0.038	0.0051	0.0051	<1E-04	<1E-04
	C4	8.10	0.016	0.0091	0.0064	<1E-04	<1E-04

For further reference, we present pictures some results in Fig. 14. These are the snapshots of the decoded images for route C4, whose received CNR was the lowest (around 8 dB). They show that the joint inference approach with both HOS matching and the EM algorithm removed the influence of the real digital satellite mobile channel.



Original picture.



Enlarged original picture.

## 7 Conclusion

We proposed mobile reception methods to improve the quality of the received information in digital satellite broadcasting. These methods do not use any diversity reception or pilot sequences [5], but use channel estimation and stochastic inference of code words based on a simple graphical model, which represents the fading channel and the convolutional code. We proposed two channel estimation methods and two stochastic inference methods. The estimation methods are MLE and HOS matching. The stochastic inference methods are marginal inference and joint inference.

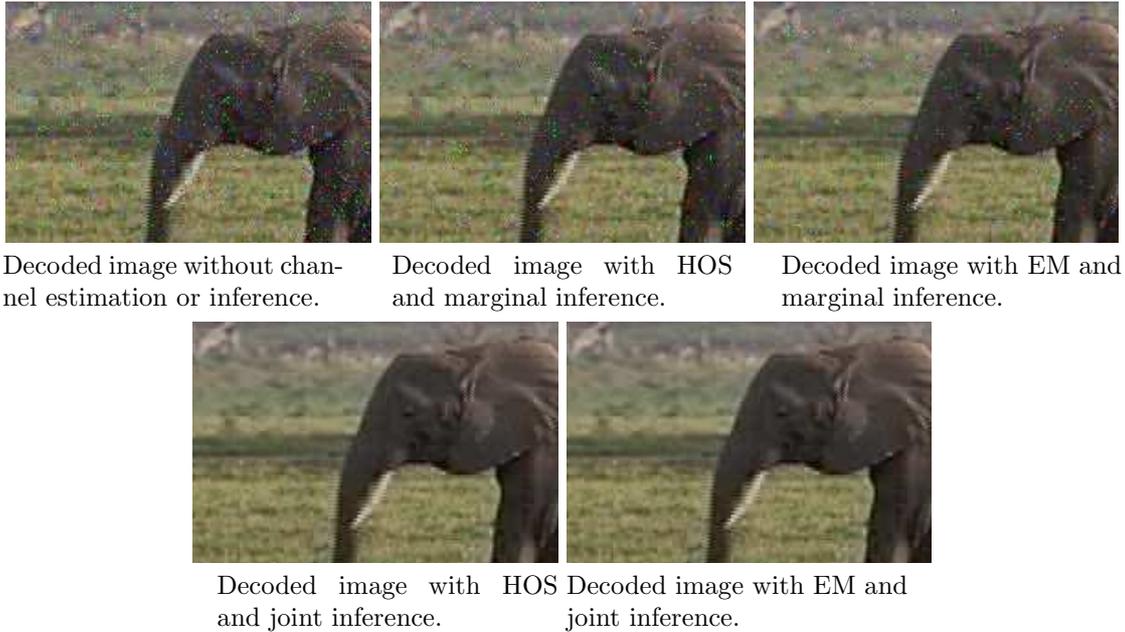


Figure 14: Snapshots of Picture Information for Route C4.

The error performances of the two estimation methods were very similar. For a real channel, the HOS matching method worked slightly worse than the MLE method. The reason for this might be related to the fact that HOS matching strongly relies on the assumption that the noise is Gaussian, whereas the distribution of the real channel might be non-Gaussian. However, the difference is small. It is also important to note that the computational cost of the HOS matching method is very small compared with the EM algorithm where iteration is inevitable.

We clarified that the inference of the code words combined with the channel estimation works well. The results of the numerical and real data experiments show that convolutional code inference by the marginal probability works well if the received CNR is high enough, but not in a low CNR range. On the other hand, inference by joint probability works well in a practical CNR range, from 8 to 10 dB.

For these reasons, we conclude that joint inference with HOS channel estimation is very effective for receiving BS digital satellite broadcasting in vehicles. Note that the CNR in the experiment with real data was lower than that of the currently available system, and the result was still good enough for a real application. It is also important to note that the computational cost is quite low and it is possible to implement it to a practical system that requires realtime processing.

Finally, we mention some of our plans for future work and conclude the paper. We would like to extend the idea to other modulations such as OFDM and CDMA [10, 1]. For the fading channel, we assumed a single interfering path, but we can extend it to multiple paths. The problem for such an extension is the computational cost for the estimation and inference. Roughly speaking, the computational cost for the inference increases exponentially with the number of paths. In this paper, we did not discuss the RS decoding algorithm. Although belief propagation is not used for decoding the RS codes, it is the standard algorithm for turbo codes[2] and LDPC (low-density parity check) codes [12]. If we implement the decoding of the RS code with a BP decoder, the whole process of the inference of the codes and the decoding can be unified with one BP algorithm. We intend to study this.

## References

- [1] Fumiyuki Adachi, Deepshika Garg, Shinsuke Takaoka, and Kazuaki takeda. Broadband CDMA techniques. *IEEE Wireless Communications*, 12(2):8–18, 2005.
- [2] Claude Berrou, Alain Glavieux, and Punya Thitimajshima. Near Shannon limit error-correcting coding and decoding: Turbo-codes. In *Proceedings of IEEE International Conference on Communications*, pages 1064–1070, Geneva, Switzerland, 1993.

- [3] James K. Cavers. An analysis of pilot symbol assisted modulation for Rayleigh fading channels. *IEEE Transactions on Vehicular Technology*, 40(4):686–693, 1991.
- [4] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39:1–38, 1977.
- [5] Rrank A. Dietrich and Wolfgang Utschick. Maximum ratio combining of correlated Rayleigh fading channels with imperfect channel knowledge. *IEEE Transactions on Communications*, 7(9):419–421, 2003.
- [6] G. David Forney Jr. *Concatenated Codes*. Research Monograph No.37. The MIT Pres, 1966.
- [7] Osamu Hashimoto, Katsuki Hirabayashi, Kouichi Ikeda, and Takahiko Yoshida. A study on multi-beam forming using sphere lens antenna. *IEICE Transactions on Communications (Japanese)*, J77-B-2(2):113–116, 1994.
- [8] Michael I. Jordan. *Learining in Graphical Models*. Kluwer Academic, 1998.
- [9] Karl-Dirk Kammeyer, Volker Kuhn, and Thorsten Petermann. Blind and nonblind turbo estimation for fast fading gsm channels. *IEEE Journal on Selected Areas in Communications*, 19(9):1718–1728, 2001.
- [10] Yungsoo Kim, Byung Jang Jeong, Jaehak Chung, Chan-Soo Hwang, Joon S. Ryu, Ki-Ho Kim, and Young Kyun Kim. Beyond 3G: Vision, requirements, and enabling technologies. *IEEE Communications Magazines*, 41:120–124, 2003.
- [11] Brian M. Kurkoski, Paul H. Siegel, and Jack Keil Wolf. Joint message-passing decoding of ldpc codes and partial-response channels. *IEEE Transactions on Information Theory*, 48(6):1410–1423, 2002.
- [12] David J. C. MacKay. Good error-correcting codes based on very sparse matrices. *IEEE Transactions on Information Theory*, 45:399–431, 1999.
- [13] Kunitoshi Nishikawa. Land vehicle antennas. *IEICE Transactions on Communications*, E86-B(3):993–1004, 2003.
- [14] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.
- [15] John G. Proakis. *Digital Communications*. McGraw-Hill, 1995.
- [16] Luka Rugini and Paolo Banelli. Block equalization for single-carrier satellite communications with high-mobility receivers. *IEEE GLOBECOM 2007 proceedings*, pages 5021–5025, 2007.
- [17] Yasuo Suzuki and Jiro Hirokawa. Development of planar antennas. *IEICE Transactions on Communications*, E86-B(3):909–924, 2003.
- [18] Andrew P. Worthen and Wayne E. Stark. Unified design of iterative receivers using factor. *IEEE Transactions on Information Theory*, 47(1):843–849, 2001.
- [19] Sau-Hsuan Wu, Urbashi Mitra, and C.-C. Jay Kuo. Graph representation for joint channel estimation and symbol detection. *IEEE Communications Society Globecom 2004*, pages 2329–2333, 2004.

## A Belief Propagation

The idea of belief propagation is explained in this section [14]. From eq. (8), we have the following relation:

$$p(\mathbf{x}|\mathbf{y}; \boldsymbol{\theta}) = \frac{p(\mathbf{x})p(\mathbf{y} | \mathbf{x}; \boldsymbol{\theta})}{\sum_{\mathbf{x}} p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta})} = \frac{1}{Z} \prod_{i=1}^{N-1} p(y_i | x_i, x_{i-1}; \boldsymbol{\theta}) = \frac{1}{Z'} \prod_{i=1}^{N-1} \psi_i(x_i, x_{i-1}),$$

where

$$Z = 2^{2N} \sum_{\mathbf{x}} p(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}), \quad \psi_i(x_i, x_{i-1}) = \exp \left[ -\frac{(y_i - (b_0 x_i + b_1 x_{i-1}))^2}{2\sigma^2} \right]$$

$$Z' = (2\pi\sigma^2)^{N-1} Z = \sum_{\mathbf{x}} \prod_{i=1}^{N-1} \psi_i(x_i, x_{i-1}).$$

It is important to note that  $p(\mathbf{x}|\mathbf{y}; \boldsymbol{\theta})$  is denoted only with local information  $\psi_i(x_i, x_{i-1})$ . Although  $\psi_i(x_i, x_{i-1})$  seems to have 16 states, we can simplify it to 8 because we assume that  $x_i^I, x_i^Q$ , and  $w_i^I, w_i^Q$  in (6) are independent of each other: that is,

$$\psi_i(x_i, x_{i-1}) = \psi_i^I(x_i^I, x_{i-1}^I) \psi_i^Q(x_i^Q, x_{i-1}^Q),$$

where

$$\psi_i^I(x_i^I, x_{i-1}^I) = \exp \left[ -\frac{(y_i^I - (b_0 x_i^I + b_1 x_{i-1}^I))^2}{2\sigma^2} \right], \quad \psi_i^Q(x_i^Q, x_{i-1}^Q) = \exp \left[ -\frac{(y_i^Q - (b_0 x_i^Q + b_1 x_{i-1}^Q))^2}{2\sigma^2} \right].$$

The values of eq. (11) can be efficiently computed by using the characteristics of the distribution. More precisely, the computation can be applied by passing messages, first from  $i = 1$  to  $i = N$  and then from  $i = N$  to  $i = 1$ . A practical BP algorithm for calculating the maximization of the marginal probability is as follows. On the graphical model shown in Fig. 5, the probability is calculated by multiplying two messages: the forward message and the backward message. The message passing in the graphical model is shown in Fig. 15. Since the messages from  $x_{i-1}$  to  $x_i$  and from  $x_i$  to  $x_{i-1}$  pass through  $y_i$ , the practical graph includes  $y_i$  in the equation and deletes  $y_i$  from the graph, as shown in Fig. 16.

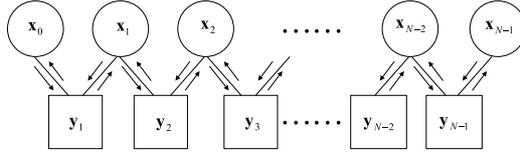


Figure 15: Message Passing in the Graphical Model.

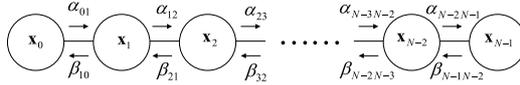


Figure 16: Message Passing of  $\mathbf{x}$ .

A practical algorithm for the graph in Fig. 16 is shown below.

**Set Initial Probability** To calculate the maximization of the marginal probability  $p_i$  of the  $i$ th node  $x_i$ , first the initial probability  $p_i^0$  is set. The initial probability is arbitrary within  $0 < p_i^0 < 1$ .

**Calculate Forward Message Passing** The message from the  $(i-1)$ th node to the  $i$ th node  $\alpha_{i-1, i}$  is calculated as

$$\alpha_{i-1, i} = \frac{1}{Z_{fi}} \sum_{x_i} \exp \left[ -\frac{(y_i - (b_0 x_i + b_1 x_{i-1}))^2}{2\sigma^2} \right] p_{i-1}^0, \quad (17)$$

where  $Z_{fi}$  is the normalization factor.

**Calculate Backward Message Passing** The message from the  $i$ th node to the  $(i-1)$ th node  $\beta_{i, i-1}$  is calculated as

$$\beta_{i, i-1} = \frac{1}{Z_{bi}} \sum_{x_{i-1}} \exp \left[ -\frac{(y_i - (b_0 x_i + b_1 x_{i-1}))^2}{2\sigma^2} \right] \beta_{i+1, i}, \quad (18)$$

where  $Z_{bi}$  is the normalization factor.

**Calculate Marginal Probability** The maximized marginal probability of the  $i$ th node  $p_i$  is computed as

$$p_i = \frac{1}{Z_{BP_i}} \alpha_{i-1} \beta_{i+1}, \quad (19)$$

where  $Z_{BP_i}$  is the normalization factor.

The practical algorithm is similar to the Forward-Backward algorithm of the hidden Markov models. Its computational cost is proportional to  $N$ , and it can easily be implemented in a realtime system.