JOINT SOURCE–CHANNEL MAPPINGS FOR THE RELAY CHANNEL

Johannes Karlsson and Mikael Skoglund

School of Electrical Engineering and the ACCESS Linnaeus Center,
Royal Institute of Technology (KTH), SE-100 44 Stockholm, Sweden. Email: johk@ee.kth.se

ABSTRACT

The three-node relay channel with a Gaussian source is studied for transmission subject to a low-delay constraint. A joint source–channel coding design algorithm is proposed and numerically evaluated. The designed system is compared to a reference system, based on modular source and channel coding, and the distortion-rate function for the Gaussian source, using known achievable rates on the relay channel. The structure of the source encoder and the relay mapping is visualized and discussed in order to gain understanding of how the system works. The relay mapping gets a structure that resembles a Wyner-Ziv code.

Index Terms— Estimation, joint source–channel coding, relay channel, quantization, sensor networks

1. INTRODUCTION

The relay channel has been studied extensively since its introduction [1]. With the increasing popularity of wireless sensor networks cooperative transmission is more relevant than ever.

In this paper we focus on the relay channel in the context of wireless sensor networks, where a possible application could be to implement the feedback link in a control system. We therefore study low-delay and energy efficient communication with a fidelity criterion on the source. Existing work on source and channel coding for the relay channel includes [2, 3]. However, whereas [2] looks at asymptotic high-SNR properties the present work is design oriented. Also, although [3] includes some practical results it relies on powerful channel codes. Because of this, the decoding is not instantaneous but a significant delay is needed for the message to be decoded.

In what follows a *low-delay* joint source–channel coding scheme for the relay channel is proposed and evaluated. To the authors' knowledge there are no similar existing results in this direction. The approach used here is related to the one being used for bandwidth compression–expansion in [4, 5, 6] and distributed source coding in [7].

2. PROBLEM FORMULATION

We will study the three node system depicted in Figure 1. Our goal is to transmit information about the Gaussian random variable X from the source node to the destination node so that it can be reconstructed with the smallest possible distortion. Besides the direct link we also have a path from the source to the destination via the relay node. The rules for the communication are the following. For each source sample X we have T channel uses at hand. The source and the relay do not transmit at the same time but must share these channel uses, we therefore use K channel uses for the transmission from the source and the remaining $L = T - K$ channel uses for the transmission from the relay. The scenario is in other words that of a half-duplex orthogonal relay channel. All transmissions are disturbed by additive white Gaussian noise, the received symbols on each channel can therefore be expressed as

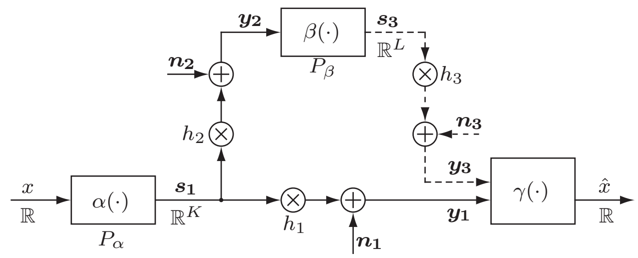

Fig. 1: Structure of the system.

$$y_i = h_i s_i + n_i \quad \forall i \in \{1, 2, 3\}, \tag{1}$$

where s_i is the transmitted symbol, h_i is a deterministic channel gain and n_i is independent white Gaussian noise with $E[n_i n_i^T] = I$ $\forall i \in \{1, 2, 3\}$. The transmitted symbols are given by the functions α and β according to

$$s_1 = s_2 = \alpha(x) \in \mathbb{R}^K, \tag{2}$$

$$s_3 = \beta(y_2) \in \mathbb{R}^L. \tag{3}$$

The equality $s_1 = s_2$ is due to the broadcast nature of a wireless channel. Without loss of generality we put the following constraint on the average transmit power per channel use for the source and the relay nodes

$$P_\alpha = \frac{1}{K}\mathrm{E}[\|S_1\|^2] \leq 1, \tag{4}$$

$$P_\beta = \frac{1}{L}\mathrm{E}[\|S_3\|^2] \leq 1. \tag{5}$$

At the destination node the received symbols are used to form an estimate of the transmitted value

$$\hat{x} = \gamma(y_1, y_3). \tag{6}$$

Given this system we want to find the optimal source encoder, relay mapping, and receiver — denoted α, β, and γ. To have a low-delay system we want the source and the relay nodes to work on a sample-by-sample basis restricting K and L to be integers. If $K > 1$, α will in general be a nonlinear mapping from the one-dimensional source space to the K-dimensional channel space. In a similar way β will be a nonlinear mapping from the K-dimensional input of the relay to the L-dimensional output. As distortion measure we use the mean squared error (MSE), $\mathrm{E}[(X - \hat{X})^2]$, "optimal" therefore refers to optimal in the minimum MSE sense.

1-4244-1484-9/08/$25.00 ©2008 IEEE

3. DESIGN

The expected distortion for a given system can be written as

$$D = \mathrm{E}[(X-\hat{X})^2] = \iiiint p(x)p(\boldsymbol{y_1}|\alpha(x))p(\boldsymbol{y_2}|\alpha(x)) p(\boldsymbol{y_3}|\beta(\boldsymbol{y_2}))(x-\gamma(\boldsymbol{y_1},\boldsymbol{y_3}))^2 \mathrm{d}x\mathrm{d}\boldsymbol{y_1}\mathrm{d}\boldsymbol{y_2}\mathrm{d}\boldsymbol{y_3}, \quad (7)$$

where $p(\cdot)$ and $p(\cdot|\cdot)$ denote probability density functions (pdf:s) and conditional pdf:s, respectively. What we would like is to find α, β, and γ such that D is minimized (given the power constraint $P_\alpha \leq 1, P_\beta \leq 1$). There are two problems with this direct approach. First, it is very hard to optimize all parts of the system simultaneously; second, the optimal mappings could be arbitrary nonlinear mappings with no closed form expressions. To make the problem feasible we take the following suboptimal approach. Instead of optimizing all parts of the system simultaneously we use the common strategy of optimizing one part at a time keeping the others fixed. The second problem is solved by discretizing each dimension of the channel space into M equally spaced points according to

$$\mathcal{S} = \{-\Delta\frac{M}{2}, -\Delta(\frac{M}{2}-1), \ldots, \Delta(\frac{M}{2}-1), \Delta\frac{M}{2}\} \quad (8)$$

and restricting the outputs of the source and the relay node to satisfy $\boldsymbol{s_1} \in \mathcal{S}^K$ and $\boldsymbol{s_3} \in \mathcal{S}^L$, respectively. At the receiving side the same approximation is made using a hard decision decoding rule — for instance, $\boldsymbol{y_1}$ is decoded according to

$$\hat{\boldsymbol{y}}_{\boldsymbol{1}} = \arg\min_{\boldsymbol{y'_1} \in \mathcal{S}^K} \|\boldsymbol{y_1} - h_1\boldsymbol{y'_1}\|^2 \quad (9)$$

where the hat will be used to indicate that the value has been discretized. This approximation is expected to be good as long as $h_i\Delta$ is small in relation to the channel noise $\boldsymbol{n_i}$. In the following analysis $P(\cdot|\cdot)$ will be used for conditional probabilities — for example, $P(\hat{\boldsymbol{y}}_{\boldsymbol{3}}|\boldsymbol{s_1})$ denotes the probability that the relay receives $\hat{\boldsymbol{y}}_{\boldsymbol{3}}$ given that $\boldsymbol{s_1}$ is transmitted from the source.

3.1. Optimal Source Encoder

With the above approximation and assuming β and γ are kept fixed we can write the optimal source encoder α as

$$\alpha(x) = \arg\min_{\boldsymbol{s_1}\in\mathcal{S}^K} \left(D_\alpha(x,\boldsymbol{s_1}) + \lambda\|\boldsymbol{s_1}\|^2\right) \quad (10)$$

where

$$D_\alpha(x,\boldsymbol{s_1}) = \mathrm{E}[(x-\hat{X})^2|x,\boldsymbol{s_1}] = \sum_{\hat{\boldsymbol{y}}_{\boldsymbol{3}}}\sum_{\hat{\boldsymbol{y}}_{\boldsymbol{2}}}\sum_{\hat{\boldsymbol{y}}_{\boldsymbol{1}}} P(\hat{\boldsymbol{y}}_{\boldsymbol{1}}|\boldsymbol{s_1})P(\hat{\boldsymbol{y}}_{\boldsymbol{2}}|\boldsymbol{s_1})P(\hat{\boldsymbol{y}}_{\boldsymbol{3}}|\beta(\hat{\boldsymbol{y}}_{\boldsymbol{2}}))(x-\gamma(\hat{\boldsymbol{y}}_{\boldsymbol{1}},\hat{\boldsymbol{y}}_{\boldsymbol{3}}))^2 \quad (11)$$

and $\lambda\|\boldsymbol{s_1}\|^2$ is a Lagrange term that is included for the following reason: $\|\boldsymbol{s_1}\|^2$ is a measure of the power that is needed to transmit the signal $\boldsymbol{s_1}$, the term $\lambda\|\boldsymbol{s_1}\|^2$ can therefore be used to control the transmit power of the source node by penalizing signals that would use too much power. When λ is set to the "correct" value, the source encoder will not encode x to the signal that gives the lowest distortion but rather to the signal that gives the lowest distortion conditioned that the power constraint in (4) is fulfilled.

3.2. Optimal Relay Mapping

In a similar way the optimal relay mapping β is given by

$$\beta(\hat{\boldsymbol{y}}_{\boldsymbol{2}}) = \arg\min_{\boldsymbol{s_3}\in\mathcal{S}^L} \left(D_\beta(\hat{\boldsymbol{y}}_{\boldsymbol{2}},\boldsymbol{s_3}) + \eta\|\boldsymbol{s_3}\|^2\right) \quad (12)$$

where

$$D_\beta(\hat{\boldsymbol{y}}_{\boldsymbol{2}},\boldsymbol{s_3}) = \mathrm{E}[(X-\hat{X})^2|\hat{\boldsymbol{y}}_{\boldsymbol{2}},\boldsymbol{s_3}] = \sum_{\hat{\boldsymbol{y}}_{\boldsymbol{1}}}\sum_{\hat{\boldsymbol{y}}_{\boldsymbol{3}}} \int_x P(\alpha(x)|\hat{\boldsymbol{y}}_{\boldsymbol{2}})P(\hat{\boldsymbol{y}}_{\boldsymbol{1}}|\alpha(x))P(\hat{\boldsymbol{y}}_{\boldsymbol{3}}|\boldsymbol{s_3})(x-\gamma(\hat{\boldsymbol{y}}_{\boldsymbol{1}},\hat{\boldsymbol{y}}_{\boldsymbol{3}}))^2\mathrm{d}x. \quad (13)$$

In (12) η is the Lagrange multiplier which makes sure that the power constraint (5) is satisfied.

3.3. Optimal Receiver

Since we use the MSE as a distortion measure, it is a well known fact from estimation theory that the optimal receiver is the expected value of X given the received symbols,

$$\hat{x} = \gamma(\hat{\boldsymbol{y}}_{\boldsymbol{1}},\hat{\boldsymbol{y}}_{\boldsymbol{3}}) = \mathrm{E}[X|\hat{\boldsymbol{y}}_{\boldsymbol{1}},\hat{\boldsymbol{y}}_{\boldsymbol{3}}]. \quad (14)$$

3.4. Design Algorithm

Given the above expressions for the source encoder, the relay mapping, and the receiver it will be possible to optimize the system iteratively. We do this by keeping two parts of the system fixed while we optimize the third part. One common problem with an iterative technique like the one suggested here is that the final solution will depend on the initialization of the algorithm, if the initialization is bad we are likely to end up in a poor local minimum. One method that has proven to be helpful in counteracting this is channel relaxation [8, 5] which works in the following way. A system is first designed for a noisy channel, the solution obtained is then used as an initialization when designing a system for a less noisy channel. The noise is reduced and the process is repeated until the desired noise level is reached. The intuition behind this method is that an optimal system for a noisy channel has a simple structure and is easy to find, as the channel noise is decreased more structure is gradually added to form the final system. The design algorithm is formally stated below.

1. Choose some initial mappings for β and γ.
2. Let $\boldsymbol{A} = (h_1^2, h_2^2, h_3^2)$ be the channel gains for which the system should be optimized. Create $\boldsymbol{A'} \leq \boldsymbol{A}$.
3. Design a system for $\boldsymbol{A'}$ according to:
   (a) Set the iteration index $k = 0$ and $D^{(0)} = \infty$.
   (b) Set $k = k + 1$.
   (c) Find the optimal source encoder α by using (10).
   (d) Find the optimal receiver γ by using (14).
   (e) Find the optimal relay mapping β by using (12).
   (f) Find the optimal receiver γ by using (14).
   (g) Evaluate the distortion $D^{(k)}$ for the system. If the relative improvement of $D^{(k)}$ compared to $D^{(k-1)}$ is less than some threshold $\delta > 0$ go to Step 4. Otherwise go to Step b.
4. If $\boldsymbol{A'} = \boldsymbol{A}$ stop the iteration. Otherwise increase $\boldsymbol{A'}$ according to some scheme (e.g., linearly) and go to Step 3 using the current system as initialization when designing the new system.

4. SIMULATION RESULTS

To evaluate the algorithm we have designed a system for the case $K = 2$, $L = 1$, and channel gains $h_1^2 = 5$ dB, $h_2^2 = 15$ dB, and $h_3^2 = 10$ dB. The values are chosen to reflect a scenario in which

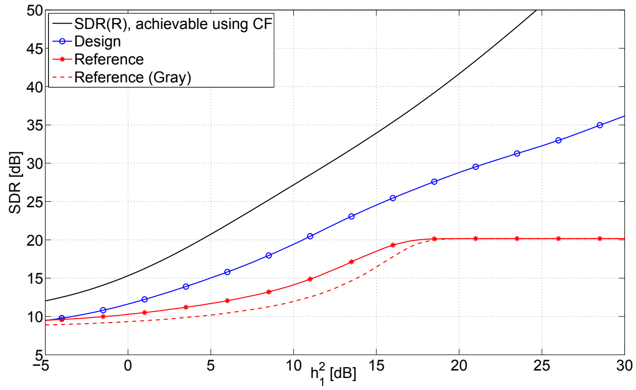

Fig. 2: Simulation results when the quality of the direct link is varied while $h_2^2 = 15$ dB and $h_3^2 = 10$ dB.

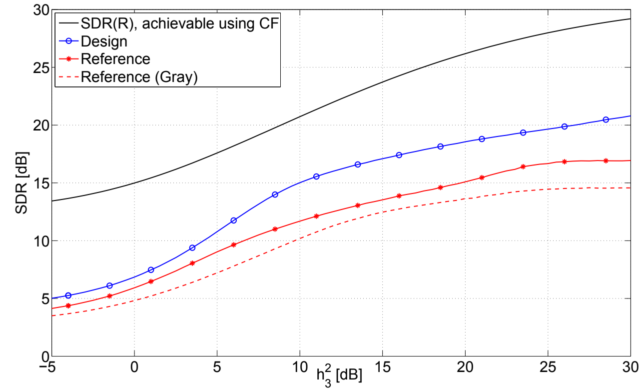

Fig. 3: Simulation results when the quality of the relay link is varied while $h_1^2 = 5$ dB and $h_2^2 = 15$ dB.

it is beneficial to make use of the relay node in the communication. We will compare the performance against a reference system and the distortion-rate function for a memoryless Gaussian source [9] using the achievable rate of the compress and forward (CF) scheme [10] with the previous mentioned assumptions, that is, $K/T = 2/3$ and orthogonal transmissions. The performance will be evaluated using the signal-to-distortion ratio (SDR) defined as

$$\text{SDR} = 10\log_{10}\left(\frac{\text{E}[X^2]}{\text{E}[(X-\hat{X})^2]}\right). \tag{15}$$

4.1. Reference System

As a reference system we design a system in a more modular fashion where we take off-the-shelf components and put them together. Instead of the source encoder $\alpha(\cdot)$ we use a 16-level Lloyd-Max quantizer followed by a 16-QAM mapping to the channel space. The relay node makes a hard decision on the received signal and re-encodes the decoded symbol with 16-PAM. At the destination node the received signals are once again decoded with a hard decision and finally x is reconstructed as the expected value of x given the decoded symbols. The reference system is optimized in the sense that we use a source optimized quantizer, a good choice of the QAM mapping (for comparison, we have also included a Gray mapping — where the indices of adjacent QAM symbols differ in only one bit.), and an optimal receiver (given the hard decoded received symbols).

4.2. Numerical Results

Before presenting the results there are some implementation aspects that are worth mentioning. In Step 2 of the algorithm $\boldsymbol{A'}$ was initially set to $(-5,-5,-5)$ dB and in Step 4 all components of $\boldsymbol{A'}$ were linearly increased until the final channel gains were reached. β was initialized as a linear mapping and γ was randomly initialized. The parameter Δ was set to $\Delta = 8/M$ and as earlier stated the approximation is expected to be good as long as $h_i\Delta$ is small in relation to the channel noise, therefore the parameter M was gradually increased along with $\boldsymbol{A'}$. In the final system we use $M = 64$ points per dimension in the channel space.

We will show the results of two different simulation scenarios. In the first scenario (Figure 2) we investigate how the system performs when the signal-to-noise ratio (SNR) of the direct link is changed while keeping the other links fixed. We assume that the source encoder and the relay mapping are fixed but that the receiver has perfect channel state information and therefore adapts to the current channel state using (14). At $h_1^2 = 5$ dB, which is the design SNR, the system is about 3.3 dB better than the reference system. When the quality of the direct link gets better we can see that the performance of the designed system also increases in an almost linear fashion. If the system would have been truly analog this scaling with the SNR would have continued as the SNR approaches infinity, but now since we have discretized the channel space, at some point the performance will be saturated. For the reference system this effect takes place at lower SNRs, we can see that already at $h_1^2 \approx 17$ dB the performance is saturated at an SDR of 20 dB which is due to the 4 bits used by the quantizer in the source node. The gap to the achievable SDR using CF is significant, however it should be pointed out that in order to achieve this performance infinite block lengths are required whereas our maximum block length is 2. It is also clearly seen that a Gray mapping is far from ideal for this problem. In the second scenario (Figure 3) we instead change the SNR of the link from the relay to the destination keeping the other links fixed. At the design SNR, $h_3^2 = 10$ dB, we of course have the same performance gain over the reference system as in the previous scenario. When the SNR increases the gain is more or less the same over the entire SNR region shown. In this scenario even the achievable SDR using CF becomes saturated at some point, this is because the bottle necks are the links from the source node.

4.3. Structure of α and β

The source encoder α is a mapping from the one-dimensional source space to the two-dimensional channel space. One input value gives rise to two output values and because of this α performs a bandwidth expansion. In its simplest form the two output values would be the same and we would have a repetition code. One way to visualize the mapping is to mark the points in $\mathcal{S}^2$ which are most likely to be transmitted. This is done in Figure 4(a) where the probabilities of the marked points sum up to 0.995. The mapping is such that small negative values of x are mapped to one end of the curve and as x is increased the mapping follows the curve to the other end. Values around zero — which are the most likely values for a Gaussian source — are mapped to the center of the curve which lies close to the origin where $\|\boldsymbol{s_1}\|^2$ is small. The transmission power for these values is hence minimized. In contrast, values that are less probable are instead mapped to points in the channel space that use more energy. This structure is due to the Lagrange term in (12), similar

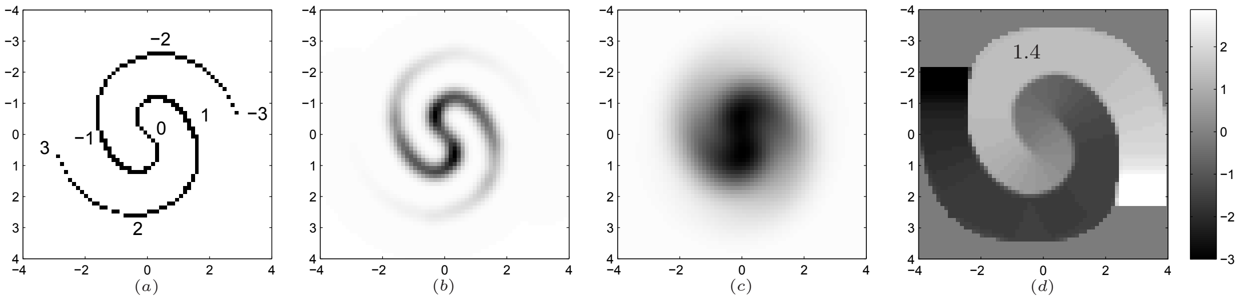

Fig. 4: Structure of α and β when $h_1^2 = 5$ dB, $h_2^2 = 15$ dB, and $h_3^2 = 10$ dB. (a) The points in $\mathcal{S}^2$ that are most likely to be transmitted, the x-axis shows the first channel use and the y-axis shows the second channel use. (For ease of comparison the axes of the subsequent figures have been divided by the corresponding channel gain.) (b) Receive probabilities at the relay node, i.e., $P(\hat{\boldsymbol{y}}_2)$. (c) Receive probabilities from the direct link at the receiver, i.e., $P(\hat{\boldsymbol{y}}_1)$ (d) Relay mapping, the color in the figure together with the colorbar shows how the 2-dimensional input is mapped to the 1-dimensional output.

results have been been obtained in [5, 6, 7].

Moving on to the relay, Figure 4(b) shows the receive probabilities for different points at the relay. We can see that the curve is somewhat smoothed but still distinguishable as opposed to the received points from the direct link at the receiver (Figure 4c) which reminds of a Gaussian distribution. Clearly the relay node needs to help the receiver distinguish which point, or at least which region, of the curve that was transmitted. Looking at Figure 4(d), which shows the relay mapping, we can see that this is exactly what the relay does. Something that is interesting to notice is that the relay is *not* the inverse of the source encoder which it would be if the relay tried to estimate x and send the estimate to the receiver. This is easiest seen by the fact that for some of the outer parts of the curve, the relay uses the same output symbol for large regions (e.g., $s_3 \approx 1.4$ for the upper part of the curve) which means that the relay does not send an estimate of what was received but rather just tells the receiver that the transmitted point was on the upper part of the curve. Using this information the receiver estimates x based on the value received from the direct link conditioned that the transmitted point was on the upper part of the curve. This could be seen as a kind of Wyner-Ziv coding which has also been proposed as a possible coding scheme for the relay [10].

5. CONCLUSIONS

We have proposed a low-delay joint source–channel coding design for the relay channel. A reference system based on modular source and channel coding has also been implemented for comparison. The numerical results show that the joint design scales well with the SNR of the system and does not saturate as quickly as the reference system, this is especially true when the SNR of the direct link is increased. More interesting however is the structure of the source encoder and the relay mapping that together make it possible for the receiver to output a good estimate of the source.

References

[1] E. C. van der Meulen, “Three-terminal communication channels,” *Adv. Appl. Prob.*, vol. 3, no. 1, pp. 120–154, 1971.

[2] D. Gündüz and E. Erkip, “Source and channel coding for cooperative relaying,” *IEEE Trans. on Information Theory*, vol. 53, no. 10, pp. 3454–3475, October 2007.

[3] H. Y. Shutoy, D. Gündüz, E. Erkip, and Y. Wang, “Cooperative source and channel coding for wireless multimedia communications,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 2, pp. 295–307, August 2007.

[4] V. Vaishampayan, *Combined source–channel coding for bandlimited waveform channels*, Ph.D. thesis, University of Maryland, 1989.

[5] A. Fuldseth and T. A. Ramstad, “Bandwidth compression for continuous amplitude channels based on vector approximation to a continuous subset of the source signal space,” *Proc. of IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, pp. 3093–3096, April 1997.

[6] P. A. Floor, T. A. Ramstad, and N. Wernersson, “Power constrained channel optimized vector quantizers used for bandwidth expansion,” *Proc. of International Symposium on Wireless Communication Systems*, pp. 667–671, October 2007.

[7] N. Wernersson, J. Karlsson, and M. Skoglund, “Distributed scalar quantizers for Gaussian channels,” *Proc. of IEEE Int. Symposium on Information Theory*, pp. 1741–1745, June 2007.

[8] S. Gadkari and K Rose, “Noisy channel relaxation for VQ design,” *Proc. of IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, pp. 2048–2051, May 1996.

[9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscience, 2 edition, 2006.

[10] A. Høst-Madsen and J. Zhang, “Capacity bounds and power allocation for wireless relay channels,” *IEEE Trans. on Information Theory*, vol. 51, no. 6, pp. 2020–2040, June 2005.