Optimal Relay Functionality for SNR Maximization in Memoryless Relay Networks

Krishna Srikanth Gomadam, Syed Ali Jafar
Center for Pervasive Communications and Computing
Electrical Engineering and Computer Science
University of California, Irvine, CA 92697-2625
Email: kgomadam@uci.edu, syed@uci.edu

Abstract

We explore the SNR-optimal relay functionality in a memoryless relay network, i.e. a network where, during each channel use, the signal transmitted by a relay depends only on the last received symbol at that relay. We develop a generalized notion of SNR for the class of memoryless relay functions. The solution to the generalized SNR optimization problem leads to the novel concept of minimum mean squared uncorrelated error (MMSUE) estimation. For the elemental case of a single relay, we show that MMSUE estimate is a scaled version of the MMSE estimate. This scheme, that we call estimate and forward (EF), performs better than the best of amplify and forward (AF) and demodulate and forward (DF) in both parallel and serial relay networks. We determine that AF is near-optimal at low transmit power in a parallel network, while DF is near-optimal at high transmit power in a serial network. For hybrid networks that contain both serial and parallel elements, the advantage of EF over the best of AF and DF is found to be significant. Error probabilities are provided to substantiate the performance gain obtained through SNR optimality. We also show that, for Gaussian inputs, AF, DF and EF are identical.

Index Terms

Memoryless relay networks, estimate and forward, relay function, MMSUE, parallel relay networks, serial relay networks, hybrid relay networks
I. INTRODUCTION

The traditional wireless communication problem is to design effective coding and decoding techniques to enable reliable communication at data rates approaching the capacity of a channel. The channel is defined by a set of given assumptions regarding the physical signal propagation environment between the transmitter and the receiver. However, recent focus on cooperative communications presents a remarkable change of paradigm where in addition to the physical environment, the network is the channel [1]. In other words, with cooperative communications the effective channel between the original source and the final destination of a message depends not only on the given physical signal propagation conditions but also the signal processing at the cooperating nodes. The change of paradigm is quite significant. With cooperative communications, not only is there a need to optimally design the encoder and decoder at the source and destination, but also to design the channel itself by optimally choosing the functionality of the intermediate relay nodes. The choice of relay function is especially important as it directly affects the potential capacity benefits of cooperation which have been shown to be quite significant [2]–[6].

A number of relay strategies have been studied in literature. These strategies include amplify-and-forward [7] [8], where the relay sends a scaled version of its received signal to the destination, demodulate-and-forward [8] in which the relay demodulates individual symbols and retransmits, decode-and-forward [9] in which the relay decodes the entire message, re-encodes it and re-transmits it to the destination, and compress-and-forward [10] [4] where the relay sends a quantized version of its received signal. In [11], gains are determined for AF relays to minimize the MMSE of the source signal at the destination. It is shown that significant saving in power is achieved if there is no power constraint on the relays. Similarly in [12], gains for AF relays in a multiuser parallel network are determined to achieve a joint minimization of the mean squared error of the source signals at the destination.

From a practical standpoint, the benefits of cooperation are offset by the cost of cooperation in terms of the required processing complexity and transmit power at the relay nodes. The complexity of the signal processing at the relay could range from highly sophisticated decode-and-forward or compress-and-forward techniques [13] that require joint processing of a long sequence of received symbols to memoryless schemes such as amplify-and-forward or demodulate-and-forward that process only one symbol at a time. Clearly, the most desirable schemes are those that approach the limits of cooperative capacity with minimal processing complexity at the relays. Memoryless relay functions are highly relevant for this objective. In addition to their simplicity, memoryless relays are quite powerful in their capacity benefits. For example, the memoryless scheme of amplify-and-forward is known to be the capacity-optimal relay scheme for many interesting cases [14]–[18]. The effect of finite block-length processing at the relay on the capacity of serial networks is analyzed in [19], [20]. In [21], the memoryless MMSE estimate and
forward scheme has been shown numerically to be capacity optimal for a single relay system. For a single relay
and with BPSK modulation, the BER-optimal memoryless scheme is found by Faycal and Medard [22]. The BER-
optimal relay function turns out to be a Lambert W function normalized by the signal and noise power. In this paper
we explore the SNR-optimal signal processing function for memoryless networks with possibly multiple relays.
While SNR optimality does not always guarantee capacity or BER optimality, it is a practically useful performance
metric. SNR-optimization is especially interesting for its greater tractability that allows analytical results where
capacity and BER optimizations may be intractable, e.g. with multiple relays.

A. Notations

Throughout the paper, * represents the conjugation operation. \( \mathcal{E}[x] \), \(|x|\) and \( \Re(x) \) denote the standard expectation
operator, the absolute and the real part of \( x \) respectively.

II. Shaping the Relay Channel: Amplification, Demodulation and Estimation

In this section, we discuss the relay functions of common memoryless forwarding strategies and provide new
perspectives that lead us to a novel and superior memoryless forwarding technique.

A. Soft and Hard Information: Amplify and Demodulate

Within the class of memoryless relay strategies, amplification and demodulation are the most basic forwarding
techniques [8]. An AF relay simply forwards the received signal \( r \) after scaling it down to satisfy its power constraint.
The relay function for AF can be written as

\[
f_{AF}(r) = \sqrt{\frac{P_R}{P + 1}} r,
\]

where \( P \) and \( P_R \) are the source and relay power respectively. Evidently with AF, the relay tries to provide soft
information to the destination. A disadvantage with this technique is that significant power is expended at the relay
when \(|r|\) is high. In DF schemes, demodulation of the received symbol at the relay is followed by modulation with
its own power constraint \( P_R \). For BPSK modulation, the relay function for DF can be expressed as

\[
f_{DF}(r) = \sqrt{P_R} \text{sign}(r),
\]

where \( \text{sign}(r) \) outputs the sign of \( r \). Due to demodulation, the relay transmitted signal carries no information about
the degree of uncertainty in the relay’s choice of the optimal demodulated symbol. Demodulation at the relays
can lead to severe performance degradation in some scenarios. For example, in a parallel relay network, reliability information can be utilized to achieve better performance over DF.

From the relay functions of AF and DF, one can argue that an optimal relay function should provide soft information when there is an uncertainty in the received symbol, and at the same time should not expend a lot of power when the cost of power out-weighs the value of soft information.

**B. Estimate and Forward: A Novel Memoryless Forwarding Strategy**

The forwarding schemes can also be related to the fundamental signal processing operations: detection and estimation. In DF, the relay demodulates the received symbol employing MAP detection rule. So a DF function can be viewed as a MAP detector followed by a modulator. In a similar vein, AF can be viewed as a linear MMSE\(^1\) estimation scheme followed by normalization to satisfy the relay power constraint.

\[
f_{AF}(r) = \beta \hat{X}_{\text{linear}}(r),
\]

where the linear estimate \(\hat{X}_{\text{linear}}(r)\) obtained at the relay is given by

\[
\hat{X}_{\text{linear}}(r) = \frac{P}{P + 1} r,
\]

and

\[
\beta = \sqrt{\frac{P_R(P + 1)}{P^2}}.
\]

\(^1\)Because of the normalization associated with the relay power constraint all linear estimates are equivalent.
Viewing AF as linear MMSE leads naturally to the forwarding scheme of EF where the unconstrained MMSE estimate is forwarded. The unconstrained MMSE estimator is given by

$$\hat{X}(r) = \mathcal{E}(x|r).$$

1) **Relay Function for EF with BPSK modulation:** When the source employs BPSK modulation, the MMSE estimate at the relay is given by

$$\hat{X}(r) = \sqrt{P} \tanh(\sqrt{Pr}),$$

where $\tanh(z)$ returns the hyperbolic tangent of $z$. The relay function is therefore,

$$f_{EF}(r) = \sqrt{P} \frac{P_R}{\mathcal{E}[\tanh^2(\sqrt{Pr})]} \tanh(\sqrt{Pr}).$$

Note that all the memoryless schemes operate at sampled output of the matched filter. In this regard all the schemes have similar processing complexity. It is worth noting that while EF and AF require amplitude digitization, DF does not. Fig. 1 shows the relay functions for AF, EF and DF for $P = 1$. It can be seen that the relay function $f_{EF}$ is linear for small values of $|r|$. Its slope reduces gradually and ultimately becomes flat similar to $f_{DF}$. The function $f_{EF} = \sqrt{P} \tanh(\sqrt{Pr})$ is intuitively appealing for the following characteristics.

- Soft information in region of uncertainty.
- Limited power in region of high power cost.

The insights obtained in this section will be useful in understanding optimum relay functionalities in a relay network. In the next section, we determine the optimal memoryless strategy for a single relay network.

### III. Single Relay Channel

**A. Problem Statement**

Consider an elemental relay channel model as shown in the figure below, in which a single relay R assists the communication between the source S and the destination D. Both S-R and R-D links are assumed to be non-fading. Without loss of generality, channel gains for the source-relay and the relay-destination link can be incorporated into the model by modifying the source and relay power appropriately. There is no direct link between the source

![Fig. 2. Elementary Relay Channel](image)

and the destination, which may be due to the half duplex constraint of the nodes, where in the first slot D serves a
different set of nodes. The transmit power at the source and the relay is \( P \) and \( P_R \) respectively. At both the relay and the destination, the received symbol is corrupted by real additive white Gaussian noise of unit power. Relay \( R \) observes \( r \), a noisy version of the transmitted symbol \( x \). Based on the observation \( r \), the relay transmits a symbol \( f(r) \) which is received at the destination along with its noise \( n_2 \).

\[
\begin{align*}
    r &= x + n_1 \\
    y &= f(r) + n_2
\end{align*}
\]

The relay function \( f \) satisfies the average power constraint, \textit{i.e.} \( \mathcal{E}[|f(r)|^2] = P_R \). We seek to determine the memoryless relay function \( f(.) \) that maximizes SNR at the destination.

**B. What is the definition of SNR?**

Consider the MMSE estimate at the relay as a sum of the desired signal \( x \) and distortion \( e \).

\[ \mathcal{E}[x|r] = x + e \]

As MMSE minimizes \( \mathcal{E}[|e|^2] \), it may appear that MMSE maximizes SNR at the output of the relay. However, it is not meaningful to view SNR in terms of MMSE even though MMSE and SNR are closely related in terms of performance. This is because the MMSE error \( e \), in general, is correlated to the desired signal \( x \), \textit{i.e.}, part of the signal is contained in distortion \( e \).

Given an observation \( y \) that contains a desired signal \( x \) as well as some distortion (noise), SNR is traditionally defined as the power of the signal \( P_x \) divided by the power in the noise component \( P_n \). For observations of the form \( y = x + n \) where the observed power \( P_y = P_x + P_n \) (\textit{i.e.} signal and noise are uncorrelated) it is easy to separately identify the contribution of the signal power and the noise power to the observed power. However, what is the definition of SNR if the observation \( y \) is not already explicitly presented in the standard form \( y = x + n \) with uncorrelated signal and noise components? In general, the observation \( y \) may have an arbitrary and possibly non-linear dependence on the desired signal \( x \). For example, consider the signal at the destination: \( y = f(x + n_1) + n_2 \) with an arbitrary function \( f(.) \) describing the memoryless relay functionality. In order to define SNR one needs to separately identify the power contributions of the signal and noise components to the observed signal \( y \). If we can identify \( P_y = P_x + P_n \) then the definition of SNR readily follows as \( \frac{P_x}{P_n} \). In other words, the definition of SNR follows from a representation of the observation \( y \) in the form \( y = \alpha(x + e_u) \) where \( x \) and \( e_u \) are \textit{uncorrelated}, and \( \alpha \) is a scaling factor that does not affect the SNR. Then the SNR is given by \( \frac{\mathcal{E}[|x|^2]}{\mathcal{E}[|e_u|^2]} \).
Any signal $y$ that contains $x$ can be expressed in the form
\[ y = \frac{\mathcal{E}[x^*y]}{\mathcal{E}[|x|^2]} (x + e_u), \] (5)

Rearranging (5),
\[ e_u = \frac{P}{\mathcal{E}[x^*y]} y - x. \] (6)

It is easy to verify that $e_u$ is uncorrelated to $x$.
\[ \mathcal{E}[x^*e_u] = \mathcal{E} \left[ x^* \left( \frac{\mathcal{E}[|x|^2]}{\mathcal{E}[x^*y]} y - x \right) \right] \]
\[ = \frac{\mathcal{E}[|x|^2]}{\mathcal{E}[x^*y]} \mathcal{E}[x^*y] - \mathcal{E}[x^*x] = 0 \] (7)

To calculate the SNR of the received signal $y$, we need to identify the error term in the received signal $y$, that is uncorrelated to the signal $x$. The scaling factor $\frac{\mathcal{E}[x^*y]}{\mathcal{E}[|x|^2]}$ in (5) is common to both the signal and error terms. Therefore, the generalized SNR is defined as follows:
\[ \text{GSNR} = \frac{\mathcal{E}[|x|^2]}{\mathcal{E}[|e_u|^2]} = \frac{\mathcal{E}[|x|^2]}{\mathcal{E}[|\alpha - x|^2]}, \] (9)
where $\alpha = \frac{\mathcal{E}[x^*y]}{\mathcal{E}[|x|^2]}$. The advantage of the generalized definition lies in its applicability to both linear and nonlinear relay functions. Note that the conventional definition of SNR for point to point links is a special case of the generalized SNR. For example, consider a received signal $y = hx + n$. The conventional SNR is $|h|^2 P$. To obtain the generalized SNR, we need to express $y$ as a sum of the signal $x$ and uncorrelated error $e_u$ in the following form.
\[ y = \frac{\mathcal{E}[x^*y]}{\mathcal{E}[|x|^2]} (x + e_u), \]
where $\frac{\mathcal{E}[x^*y]}{\mathcal{E}[|x|^2]} = h$, in this case. Therefore $e_u = \frac{n}{h}$. The generalized SNR from (9) is,
\[ \text{GSNR} = \frac{P}{|h|^2} = |h|^2 P, \]
which is also the conventional definition of SNR.

The GSNR concept can be viewed as a decomposition of an observation into a component along the desired signal space and its orthogonal signal (uncorrelated noise) space. The orthogonal projections are evident in the second moment constraint $P_y = P_x + P_n$. GSNR is therefore as natural and meaningful a metric as the orthogonal projections themselves. While GSNR optimization does not guarantee capacity or BER optimality it is interesting to note that all three metrics (BER, capacity, GSNR) lead to very similar optimal relay functions for BPSK. The
BER optimal Lambert-W function is very similar to the GSNR optimal tan-hyperbolic function (EF). Moreover, in a separate work we have shown that, numerically, the GSNR optimal EF function is also capacity optimal for BPSK modulation [21].

C. Optimal Relay Function

We first derive the optimal estimation method that maximizes GSNR. Based on this result, we determine the optimal relay function.

Theorem 1: Given an observation $r$ that contains (is correlated with) the signal $x$, the MMSUE estimate of $x$ is

$$\hat{X}(r) = \frac{\mathcal{E} \left[ |x|^2 \right]}{\mathcal{E} \left[ x^* \mathcal{E} [x|r] \right]} \mathcal{E} [x|r],$$

regardless of the joint distribution of $x$ and $r$.

Proof: Given the observed signal $r$, we wish to find the MMSUE estimate $\hat{X}(r)$, i.e. the estimate that minimizes the mean squared error uncorrelated to $x$. The optimization problem is

$$\min_{\hat{X}(r)} \mathcal{E} \left[ |\hat{X}(r) - x|^2 \right] \quad \text{s.t.} \quad \mathcal{E} \left[ (\hat{X}(r) - x)^* x \right] = 0.$$  

Employing Lagrange multipliers$^2$, we write the constrained minimization as the minimization of

$$\text{MSUE} = \mathcal{E} \left[ |\hat{X}(r) - x|^2 \right] - \lambda \mathcal{E} \left[ x^* (\hat{X}(r) - x) \right] - \lambda^* \mathcal{E} \left[ x \left( \hat{X}^*(r) - x^* \right) \right]$$

$$= \mathcal{E} \left[ |\hat{X}(r)|^2 \right] + (-\lambda + 1) \mathcal{E} \left[ x^* \hat{X}(r) \right] + (-\lambda^* + 1) \mathcal{E} \left[ x \hat{X}^*(r) \right] - P (-\lambda - \lambda^* - 1)$$

$$= \mathcal{E}_r \left[ |\hat{X}(r)|^2 - (\lambda - 1) \hat{X}(r) \mathcal{E} [x^*|r] - (\lambda^* - 1) \hat{X}^*(r) \mathcal{E} [x|r] \right] + P (\lambda + \lambda^* + 1)$$

From the above equation, it is clear that MSUE is minimized when

$$\hat{X}(r) = (\lambda - 1) \mathcal{E} [x|r],$$

and the minimum MSUE is

$$\text{MMSUE} = P (2\Re (\lambda) + 1) - (\lambda - 1)^2 \mathcal{E}_r \left[ |\mathcal{E} [x|r]|^2 \right].$$  

$^2$As the constraint $\mathcal{E}[x e_u^{*}]$ is a complex quantity, the Lagrange multipliers are $\lambda$ and $\lambda^*$ corresponding to $\mathcal{E}[x e_u^{*}]$ and $\mathcal{E}[x^* e_u]$ respectively.
From the constraint that
\[ \mathcal{E}[x^*e_u] = 0 \Rightarrow \mathcal{E}[x^*\hat{X}(r)] = P, \]
we have
\[ \lambda - 1 = \frac{P}{\mathcal{E}[x^*\mathcal{E}[x|r]]}. \]
Therefore \( \hat{X}(r) = \frac{P}{\mathcal{E}[x^*\mathcal{E}[x|r]]}\mathcal{E}[x|r] \) is the SNR maximizing estimate. It should be noted that the above result is completely general and is valid for any input distribution at the source, and noise distribution at the relay and destination.

**Theorem 2:** For a network with a single relay that has a power constraint \( P_R \), the relay function that maximizes GSNR at the destination is
\[ f(r) = \sqrt{\frac{P_R}{\mathcal{E}[[x^2 + e^2]^{\frac{1}{2}}]}} \mathcal{E}[x|r], \]
regardless of the input and noise distributions.

**Proof:** Consider the received signal at the destination,
\[ y = f(r) + n_2, \quad (11) \]
where \( \mathcal{E}[[f(r)]^2] = P_R \). The relay function can be viewed as an estimation followed by a scaling at the relay to satisfy the power constraint. Thus the relay function can be expressed as a scaled version of \( x + e_u \).
\[ f(r) = \beta(x + e_u), \]
where \( \beta \) is given by
\[ \beta^2 = \frac{P_R}{\mathcal{E}[[x + e_u]^2]} = \frac{P_R}{P + \text{MSUE}}. \]
The received signal at the destination is
\[ y = f(r) + n_2 = \sqrt{\frac{P_R}{P + \text{MSUE}}} (x + e_u) + n_2. \]
The generalized SNR at the destination is then given by,
\[ \text{GSNR}_D = \frac{P}{\frac{P}{P + \text{MSUE}}} \cdot \quad (12) \]
Clearly minimizing MSUE amounts to maximizing GSNR at the destination. Due to power normalization at the relay, the scaling factor associated with MMSUE estimate does not affect GSNR. Therefore, the optimal relay function is a scaled MMSE estimator and the scaling factor is determined by the relay power constraint. It is given by

\[ f(r) = \sqrt{\frac{P_R}{E_r[|E(x|r)|^2]}} E[x|r]. \]

Theorem 2 implies that any scaled version of MMSE estimator is GSNR optimal. Thus, regardless of the power constraint at the relay, EF maximizes GSNR at the destination. The relation between MMSE and MMSUE is discussed in Appendix A. In the next section, we compare the performance of the memoryless schemes for a dual hop relay network.

IV. COMPARATIVE ANALYSIS

For any relay function, GSNR at the destination can be obtained from (12). Therefore calculation of MSUE of DF and AF allows a direct comparison of these schemes with the GSNR optimal EF. To determine the MSUE estimate from any relay function, we only need to obtain the scaling factor that allows the relay function to be expressed as in (5). In this section, we compare the schemes for BPSK modulation and illustrate the concept of generalized SNR.

A. BPSK Modulation

We express the relay function of a demodulating relay for BPSK as

\[ f_{DF}(x + n) = \sqrt{P_R} \text{sign}(x + n) = \sqrt{\frac{P_R}{P}} (x + d). \]

where \( d \) is the Euclidean distance between the input symbol \( x \) and the demodulated symbol. The distribution of \( d \) conditioned on \( x \) is given by

\[
\begin{cases}
0 & 1 - \epsilon \\
-2x & \epsilon
\end{cases}
\]

where \( \epsilon = Q\left(\sqrt{P}\right) \), the probability of symbol error. As seen from the error distribution, the demodulation error \( d \) is correlated with \( x \). The correlation between the input and the error is given by

\[ \mathcal{E}(xd) = -2P\epsilon. \]
The uncorrelated error can be calculated from (5) according to which
\[ e_u = P \frac{P}{E(x f_{DF}(x+n))} f_{DF}(x+n) - x \]
\[ = P(x+d) - x \frac{2cx + d}{P - 2P\epsilon} - x = \frac{2cx + d}{1 - 2\epsilon}. \] (16)

From (14), the power of the uncorrelated error in (16) can be calculated and is given by
\[ MSUE_{DF} = \frac{4P\epsilon(1-\epsilon)}{(1-2\epsilon)^2} \] (17)

To characterize the MSUE associated with the DF relay function, we first consider \( \epsilon \), the probability of decision error at the relay.
\[ \epsilon = Q\left(\sqrt{P}\right) = \frac{1}{\sqrt{2\pi}} \int_{\frac{-\sqrt{P}}{2}}^{\infty} \exp(-\frac{x^2}{2})dx \]
\[ = \frac{1}{2} - \frac{1}{\sqrt{2\pi}} \int_{0}^{\sqrt{P}} \exp(-\frac{x^2}{2})dx \] (19)

For small values of \( P \),
\[ \epsilon = \frac{1}{2} - \sqrt{\frac{P}{2\pi}}. \]

Therefore at low source transmit power, the mean squared uncorrelated error can be expressed as a function of \( P \).
\[ MSUE_{DF}(P) = \frac{4P \left(\frac{1}{2} - \sqrt{\frac{P}{2\pi}}\right) \left(\frac{1}{2} + \sqrt{\frac{P}{2\pi}}\right)}{\left(1 - \left(1 - \frac{2P}{\pi}\right)\right)^2} \]
\[ = 2\pi \left(\frac{1}{4} - \frac{P}{2\pi}\right) \] (20)

As \( P \to 0 \), the uncorrelated error power shoots up to \( \frac{\pi}{2} \). It should be noted that the noise variance at the relay is 1. This suggests that DF is not preferable at low \( P \).

As the relay function of an AF relay is a scaled version of the received signal \( r \), it is simple to determine the MSUE associated with the relay function. From (5), we have
\[ e_u = \frac{P}{E(x f_{AF}(x+n))} f_{AF}(x+n) - x = n. \]

The MSUE of AF is therefore the same as the noise variance, \( MSUE_{AF} = 1 \), interestingly independent of the source transmit power.

Fig. 3 plots the MSUE as a function of transmit power for all the three schemes. Several interesting observations
can be made. It can be seen that AF is close to optimum (EF) at low $P$ while DF is near optimal at high $P$. In the intermediate range, both AF and DF are far from optimal. It is well known that AF suffers from noise amplification at low SNR [23], which is in contrast to the results here. When we view the relay operation as an estimation, it is only natural that the estimation error is high at low $P$, which results in noise amplification. In fact AF is very close to optimum among all memoryless relay functions at low $P$. Rather it is DF that suffers the most from noise/error amplification. However, AF is inefficient at high $P$ as $\text{MSUE}_{AF} = 1$ does not decrease with $P$, while uncorrelated error in DF and EF vanishes at high $P$. The MSUE of the schemes for extreme values of $P$ is listed in Table I.

<table>
<thead>
<tr>
<th>Relay Function</th>
<th>$P \to 0$</th>
<th>$P \to \infty$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplify</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Demodulate</td>
<td>$\frac{\pi}{2}$</td>
<td>0</td>
</tr>
<tr>
<td>Estimate</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**TABLE I**

MSUE OF THE RELAY FUNCTIONS FOR BPSK MODULATION.

**B. Higher Order Constellations**

We know from Theorem 2 that EF is GSNR optimal for all modulation schemes. For fixed input power $P$, increasing the number of constellation points $M$ will result in an increased MSUE for EF. This is rather intuitive from the fact that increasing the number of constellation points for fixed power increases the estimation error. Fig. 4 shows the relay functions for 4-PAM constellation set. Interestingly, the relay functions of the schemes become more and more similar with increase in constellation points.
For Gaussian inputs, the unconstrained MMSE estimate and the linear MMSE estimate are equivalent.

\[ \mathcal{E}[x|r] = \frac{P}{P+1} r \]

Thus AF and EF strategies are the same for a Gaussian source. In this context, it can also be shown that DF and AF are equivalent for Gaussian inputs. The notion of demodulation of symbols from a Gaussian source is explained through the following. A Gaussian distribution is quantized into a number of states with the probability of the \( i^{th} \) state given by,

\[ \Pr(x_i) = \frac{1}{\sqrt{2\pi P}} \int_{(i-1)\Delta x}^{i\Delta x} \exp\left(\frac{-x^2}{2P}\right) dx. \]

Suppose the source transmits symbols \( x_i \) according to the probability distribution above, then the MAP detection rule at the relay is given by

\[ \hat{X}(r) = \arg \max_{x_i} \Pr(x|r) \]

In the limit \( \Delta x \to 0 \), \( x \) and \( r \) become jointly Gaussian. It is well known that the conditional mean \( \mathcal{E}(x|r) \) maximizes the joint probability. Therefore \( \mathcal{E}(x|r) \) which is also the MMSE estimate is the output of the ML detector. Thus for Gaussian inputs AF, EF and DF are identical, and the relay function is linear.

V. PARALLEL RELAY NETWORK

A Gaussian parallel relay channel [14] is shown in Fig. 5. It consists of a single source destination pair with \( L \) relays that assist in the communication. All the links are assumed to be non-fading with unequal channel gains and information is transferred in two time slots. The relays observe \( \{r_i\}_{i=1}^L \), the noisy version of the transmitted signal.
Fig. 5. Parallel Relay Channel

\[ r_i = g_i x + n_i \]  \hspace{1cm} (22)

where \( g_i \) is the gain of the link between the source and the \( i^{th} \) relay. \( n_i \) denotes an additive white Gaussian noise with unit power. Since the relays are assumed to be memoryless, each relay transmits a signal that is a function of its observation \( r_i \). For ease of notation, we denote the transmit power at the source as \( P \) and the relay transmit power as \( \{P_i\}_{i=1}^L \). Without loss of generality, the channel gain for the relay-destination links can be introduced through the relay transmit power. The destination receives the sum of all relay transmitted symbols along with its own noise.

\[ y = \sum_{i=1}^L f_i(r_i) + n \]

By viewing relay operation as an estimation we have,

\[ f_i(r_i) = f_i(g_i x + n_i) = \alpha_i(x + e_i) \]

where \( e_i \) is the uncorrelated estimation error at the \( i^{th} \) relay, \( \alpha_i = \sqrt{\frac{P}{P + \text{MSUE}_i}} \) and \( \text{MSUE}_i = \mathcal{E}[|e_i|^2] \), the mean square uncorrelated error associated with the \( i^{th} \) relay function. The received signal at the destination can be expressed as

\[ y = \sum_{i=1}^L \alpha_i(x + e_i) + n \]  \hspace{1cm} (23)

For any forwarding scheme, the GSNR at the destination is given by

\[ \text{GSNR} = \frac{(\sum_{i=1}^L \alpha_i)^2 P}{\sum_{i=1}^L \alpha_i^2 \text{MSUE}_i + \sum_{i=1}^L \sum_{j=1, j \neq i}^L \alpha_i \alpha_j C_{ij} + 1} \]  \hspace{1cm} (24)

where \( C_{ij} = \mathcal{E}(e_i^* e_j) \) is the correlation between errors \( e_i \) and \( e_j \) at relays \( i \) and \( j, i \neq j \).

For AF, the correlation \( C \) is always zero as the error terms represent independent AWGN noise. For both DF and
EF, the correlation depends on the modulation scheme and is not always zero. Although each MSUE\(_i\) is minimized by EF, due to the possibility of error correlation maximum GSNR is not always guaranteed. However for most constellation sets, the error correlation can be shown to be either very close or exactly equal to zero. Theorem 6 in Appendix C shows that the correlation between errors from different relays is zero for any MPSK constellation.

A. Effect of Error Correlation on EF

The correlation between errors at the output of EF relays, in general, is not zero for all constellations. However it is negligible for many constellation sets like M-QAM and does not result in significant GSNR loss. For example, the correlation is in the order of \(10^{-3}\) for 16-QAM modulation. In fact, the correlation can be expected to decrease for large QAM constellations where the ‘edge effects’ become insignificant. However due to the combination of \(L(L-1)\) terms for the correlation expression in (24), the performance of the system may degrade for very large values of \(L\). In a symmetrical relay network with unit channel gains and with transmit power at the relays equal to the source power \(P\), the GSNR for any relay function is

\[
\text{GSNR} = \frac{L^2P}{L \text{MMSUE} + L(L-1)C + 1 + \frac{\text{MMSUE}}{P}}. \tag{25}
\]

Suppose the source employs a modulation scheme that results in nonzero correlation between errors with EF in a parallel network with unit channel gains, GSNR\(_{AF}\) > GSNR\(_{EF}\) only if

\[
C > \frac{1 - \text{MMSUE}}{L(L-1)} \left( L + \frac{1}{P} \right) \tag{26}
\]

As the scaling associated with the correlation is \(L(L-1)\), its effect is prominent for large \(L\). For a given error correlation \(C\), GSNR\(_{AF}\) > GSNR\(_{EF}\) if

\[
L \gg \frac{1 - \text{MMSUE}}{C} + 1 \tag{27}
\]

The above relation suggests that even if the correlation with EF is nonzero, the number of relays has to be very large for AF to outperform EF, with modulation schemes like M-QAM.

B. Error Correlation in DF

Similar to EF, the error correlation in DF depends on the modulation scheme at the source. From (16), the uncorrelated error associated with the DF relay function for BPSK is

\[
e_i = \frac{d_i + 2\epsilon_i x}{1 - 2\epsilon_i}, \tag{28}
\]
where $\epsilon_i$ depends on the transmit power and the source-relay channel. From the distribution of $d_i$ in (13), we can calculate the correlation between errors $e_i$ and $e_j$ at relay $i$ and $j$.

$$
\mathcal{E}(e_i e_j) = \frac{\mathcal{E}[d_i d_j] + 2\epsilon_i \mathcal{E}[x d_i] + 2\epsilon_j \mathcal{E}[x d_j] + 4\epsilon_i \epsilon_j \mathcal{E}[|x|^2]}{(1 - 2\epsilon_i)(1 - 2\epsilon_j)} \\
= \frac{4\epsilon_i \epsilon_j P - 8\epsilon_i \epsilon_j P + 4\epsilon_i \epsilon_j P}{(1 - 2\epsilon_i)(1 - 2\epsilon_j)} = 0
$$

(29)

For BPSK modulation and with unit channel gain for all the links, the GSNR at the destination is

$$
\text{GSNR}_{DF} = \frac{P L^2 (1 - 2\epsilon)^2}{4 P L \epsilon (1 - \epsilon) + 1}
$$

(30)

For large M-QAM constellation, ignoring edge effects, the demodulation error can be assumed to be independent of the transmitted symbol with the error distribution given by

$$
d = \begin{cases} 
0 & 1 - \epsilon \\
d_{\text{min}} & \epsilon/4 \\
-d_{\text{min}} & \epsilon/4 \\
j d_{\text{min}} & \epsilon/4 \\
-j d_{\text{min}} & \epsilon/4 
\end{cases}
$$

(31)

where, from [24] we have

$$
d_{\text{min}} = \sqrt{\frac{6P}{M - 1}} \\
\epsilon \leq 4 \left( 1 - \frac{1}{\sqrt{M}} \right) Q \left( \sqrt{\frac{3P}{M - 1}} \right)
$$

(32)

Then the mean squared uncorrelated error is given by

$$
\text{MSUE}_{DF} \approx d_{\text{min}}^2 \epsilon
$$

Here we assume that decision errors occur only among the nearest neighbors. Although this is an optimistic assumption, it closely predicts the performance of the system at medium and high $P$ where the assumption is justified. The effective SNR at the destination is,

$$
\text{GSNR}_{DF} \approx \frac{P L^2}{L d_{\text{min}}^2 \epsilon + 1}.
$$

(33)
C. Asymptotic GSNR Comparison

While EF is superior to AF and DF at all SNR regardless of the number of relays, it will be interesting to characterize the asymptotic gain of EF as a function of $L$ and $P$. For ease of analysis, we restrict the channel gains to be the same and the individual relay power to be equal to the source power. From the GSNR expressions of the schemes, we have the following ratios.

$$\frac{\text{GSNR}_{EF}}{\text{GSNR}_{AF}} = \frac{L P + P + 1}{L P \text{MSUE}_{EF}(P) + P + \text{MSUE}_{EF}(P)}$$

$$\frac{\text{GSNR}_{EF}}{\text{GSNR}_{DF}} = \frac{L P \text{MSUE}_{DF}(P) + P + \text{MSUE}_{DF}}{L P \text{MSUE}_{EF}(P) + P + \text{MSUE}_{EF}(P)}$$

Notice that the above expressions do not include the correlation term that may be associated with EF. Therefore, it is valid only for the zero correlation case, for example MPSK modulation. For any input distribution $p_X(x)$, i.e., for all modulation schemes,

$$\text{MSUE}_{EF}(P) \leq 1, \ \forall P. \quad (34)$$

1) Fixed $P$: With a large number of relays,

$$L \rightarrow \infty, \quad \frac{\text{GSNR}_{EF}(P)}{\text{GSNR}_{AF}(P)} = \frac{\text{MSUE}_{AF}}{\text{MSUE}_{EF}} \quad (34)$$

$$L \rightarrow \infty, \quad \frac{\text{GSNR}_{EF}(P)}{\text{GSNR}_{DF}(P)} = \frac{\text{MSUE}_{DF}}{\text{MSUE}_{EF}} \quad (35)$$

It is clear from the above expressions that the MSUE of the schemes determine the GSNR gain. We know from Section IV that MSUE$_{EF}(P)$ decreases with $P$ and ultimately becomes zero as $P \rightarrow \infty$. This implies that in a large relay network, maximum gain over AF is obtained for high transmit power $P$. Similarly maximum gain over DF is obtained at low $P$. This is due to the fact that DF is inefficient at low $P$ as indicated by its mean squared uncorrelated error.

2) Fixed $L$: For a fixed number of relays, the GSNR gain of EF over AF for BPSK modulation at high $P$ is approximately $L + 1$. Similarly the GSNR gain of EF over DF at low $P$ is $\frac{\pi}{2}$ as indicated by the following expressions.

$$P \rightarrow \infty, \quad \frac{\text{GSNR}_{EF}(L)}{\text{GSNR}_{AF}(L)} = L + 1 \quad \frac{\text{GSNR}_{EF}(L)}{\text{GSNR}_{DF}(L)} = 1 \quad (36)$$

$$P \rightarrow 0, \quad \frac{\text{GSNR}_{EF}(L)}{\text{GSNR}_{AF}(L)} = 1 \quad \frac{\text{GSNR}_{EF}(L)}{\text{GSNR}_{DF}(L)} = \frac{\pi}{2} \quad (37)$$
The above expressions clearly demonstrate the inefficiency of DF and AF at low and high SNR respectively.

**D. Numerical Results**

Fig. 6 provides the GSNR performance of the schemes in a parallel relay network with equal channel gains. The corresponding error probabilities are provided in Fig. 7. The error probabilities closely follow the trend exhibited in GSNR. It can be seen that EF achieves substantial error probability gains over AF and DF for all values of $P$. As seen in GSNR plot, AF is superior to DF at low $P$ while DF performs better than AF at high $P$. This is also indicated by (36) and (37).
VI. SERIAL RELAY NETWORK

A serial relay network with Gaussian noise at all receivers is shown in Fig. 8. All the relays are memoryless and employ a relay function to transmit a symbol based on its received symbol. Throughout the section, we assume unit channel gain for the links and equal transmit power at all the nodes. It should be noted that the relay functions, in general, need not be the same for all the relays. This is due to the fact that the noise distribution gets altered at every hop depending on the relay function of the preceding relay.

A. AF

The AF relay function for the $k^{th}$ relay is

$$f_k(r_k) = \beta r_k,$$

where $\beta^2 = \frac{P}{P+n}$. The received symbol at the destination can be expressed as

$$y = \beta L x + \sum_{i=1}^{L} \beta^n n_i + n. \quad (38)$$

Therefore the GSNR at the destination is given by

$$\text{GSNR}_{AF} = \frac{\beta^2 L P}{1 + \sum_{i=1}^{L} \beta^{2i}}. \quad (39)$$

The MSUE associated with the $k^{th}$ relay function is given by

$$\text{MSUE}_{AF}^k = 1 + \sum_{j=1}^{k-1} \beta^{-2j}. \quad (40)$$

B. DF

The DF relay function for the $k^{th}$ relay is given by

$$f_k(r_k) = f_{k-1}(r_{k-1}) + d_k, \quad k = 1...L,$$

where $f_0(r_0) = x$ and $d_k$ represents the decision error at the $k^{th}$ relay. We express the received signal at the destination as

$$y = x + \sum_{i=1}^{L} d_i + n. \quad (40)$$
From (31), the effective GSNR at the destination for a large QAM modulation can be approximated as

$$\text{GSNR}_{DF} \approx \frac{P}{Ld_{min}^2 + 1},$$

and the MSUE associated the $k^{th}$ relay function is given by

$$\text{MSUE}_{DF}^k \approx kd_{min}^2 \epsilon.$$ 

C. EF

With estimate and forward at all the relays, the relay functions are different. This is because the distribution of the noise in the received signal at the relay gets altered due to nonlinear operations performed at the preceding relay. The relay function for the $k^{th}$ relay is given by

$$f_k(r_k) = \alpha_k \mathbb{E}[x|r_k = f_{k-1}(r_{k-1}) + n_k],$$

where $\alpha_k$ is such that $\mathbb{E}[|f_k(r_k)|^2] = P$ is satisfied.

**Theorem 3:** In a serial relay network, the last stage relay functionality that maximizes GSNR is estimate and forward, regardless of the relay function in the preceding relays and the modulation scheme at the source.

**Proof:** Regardless of the relay functions at the preceding relays, the received signal at the last relay can be expressed in the same form as (5). From Theorems 1 and 2 which are valid for all input and noise distributions, it is straightforward that EF at the last stage relay maximizes GSNR at the destination.
Fig. 10. BER of schemes in a serial network ($L = 2$) for BPSK modulation.

D. Performance Comparison

Fig. 9 compares the destination GSNR of the schemes for two serial relays. Here the relay functions for DF and AF remain the same for both the relays. For EF, $f_1(r_1) = \alpha_1 \tanh(\sqrt{P} r_1)$ and $f_2(r_2) = \alpha_2 \mathbb{E}[x| r_2 = \alpha_1 \tanh(\sqrt{P} r_1) + n_2]$. As expected, EF is the best performing scheme and DF closely follows it. It can be easily noticed that in a serial network, the GSNR decreases with each stage. AF, being power inefficient at high $P$, suffers the most due to multi-hop communication.

$$\text{GSNR}_{AF} = \frac{P}{1 + \sum_{k=1}^{L} \frac{1}{\beta_k}} < \frac{P}{L + 1}$$  \hspace{1cm} (41)

For large M-QAM modulation, GSNR at the destination with DF scheme can be approximated as

$$\text{GSNR}_{DF} = \frac{P}{Ld_{\text{min}, \epsilon}^2 + 1}. \hspace{1cm} (42)$$

Clearly when $d_{\text{min}, \epsilon}^2 < 1$ (at high $P$ regime),

$$\text{GSNR}_{DF} \geq \frac{P}{L + 1}. \hspace{1cm} (43)$$

indicates that DF is superior to AF at high $P$. We can also observe the case where $\text{GSNR}_{AF} > \text{GSNR}_{DF}$ at low $P$ (when $d_{\text{min}, \epsilon}^2 > 1$). Note that the variance of the error components associated with DF ($d_{\text{min}, \epsilon}^2$) decreases exponentially with $P$, while those in AF ($\frac{1}{\beta_k}$ for the $k^{th}$ relay) decreases linearly with $P$. These observations can be clearly seen in Fig. 9 where AF performs slightly better than DF at very low $P$. Gradually with increase in $P$, DF outperforms AF and the performance gap widens with further increase in $P$. 
E. Hybrid Relay Networks

From the previous sections, it is clear that EF is well suited to both parallel and serial relay networks, and substantial performance gain can be obtained over AF and DF in many scenarios. We also observe that DF is close to optimal in a serial relay network at high $P$ whereas AF is near-optimal in parallel relay networks at low $P$. In these regimes, the performance gain of EF over the best of AF and DF is limited. Now, consider a network consisting of both parallel and serial subnetworks as shown in Fig. 11. Due to the presence of parallel and serial elements together in the network, we find a significant performance degradation in both AF and DF at all $P$. This is a scenario where EF obtains a large gain over the best of DF and AF. Fig. 12 compares the performance of schemes for the hybrid network in Fig. 11. It can be noticed that EF performs significantly better than the best of DF and AF at all $P$. This can also be observed in Fig. 13 that displays the error probability of the schemes for the hybrid network. We expect the performance gain to increase for a large network with both parallel and serial elements.
VII. CONCLUSION

In this work, we explored optimum relay functions in a memoryless relay network that maximize SNR at the destination. We developed a notion of generalized SNR as an extension to conventional SNR definition for the relay channel problem. From an estimation perspective, a general framework for determining SNR for any memoryless relay processing is developed. For the single relay case, the generalized SNR is maximized by the MMSUE estimate which is related to the traditional MMSE estimate by a constant scaling factor. We also considered parallel and serial networks with multiple relays and established the superiority of EF over DF and AF. We found that AF is near-optimal at low transmit power in a parallel network while DF is near-optimal at high transmit power in a serial network. For hybrid networks that consist of serial and parallel subnetworks, the advantage of EF over the best of AF and DF is found to be significant. By viewing AF as linear estimation and DF as detection, we showed that AF, DF, and EF are identical for Gaussian sources.

APPENDIX

A. Relation between MMSUE and MMSE

Although the GSNR arising out of MMSUE and MMSE estimates are identical, they are fundamentally distinct as the objectives optimized by them are different. MMSUE is the minimum achievable uncorrelated error power while MMSE is the minimum achievable distortion.

Theorem 4: For any input distribution, the exact relation between MMSUE and MMSE is

\[ \text{MMSUE} = \frac{\text{MMSE} - \frac{\mu^2}{P}}{(1 + \frac{\mu^2}{P})^2}, \]
and the MMSUE is always greater than or equal to MMSE. Further, the correlation between the MMSE error $e$ and the input $x$ is always non-positive.

**Proof:** We express the MMSE estimate in the form of (5)

$$\hat{X}(r) = x + e = x + \frac{\mu}{P}x + e_u,$$

where $\mu = \mathcal{E}[x^*e]$ and $e_u$ is uncorrelated to $x$. Thus, the MMSE and MMSUE errors are related by the following.

$$e_u = e - \frac{\mu}{P}x$$

(44)

Evaluating the mean squared value of $e_u$, we have the following relation.

$$\text{MMSUE} = \frac{\text{MMSE} - \frac{\mu^2}{P}}{(1 + \frac{\mu}{P})^2}$$

Now, since MMSE estimation is the optimal mean squared error minimizing method, we have the relation

$$\mathcal{E}(|e|^2) \leq \frac{1}{(1 + \frac{\mu}{P})^2} \mathcal{E}(|e_u|^2)$$

$$\leq \frac{1}{(1 + \frac{\mu}{P})^2} \left( \mathcal{E}(|e|^2) - \frac{\mu^2}{P} \right)$$

$$\leq \frac{1}{(1 + \frac{\mu}{P})^2} \mathcal{E}(|e|^2),$$

which implies $\mu \leq 0$. □

For Gaussian inputs, a unique relationship between MMSE estimate and the correlation exists, which is $\mu = \mathcal{E}(xe) = -\text{MMSE} = -\frac{P}{P+1}$. A direct consequence of the negative correlation of the error with the signal $x$ leads to the following inequality.

$$\text{SNR} \leq \frac{P}{\text{MMSE}}.$$

**B. Rotational Property of EF**

**Theorem 5:** For all MPSK constellation inputs, the MMSE estimate has the property

$$\hat{X}(re^{j\theta_m}) = e^{j\theta_m} \hat{X}(r),$$

where $\theta_m = \frac{2\pi m}{M}$, $m = 0, 1, ..., M - 1$, the signal phases of MPSK.
Proof:

\[ E[x|r = re^{j\theta_m}] = \frac{\sqrt{P}}{M} \sum_{k=0}^{M-1} e^{j\theta_k} \Pr[x = \sqrt{Pe^{j\theta_k}}|re^{j\theta_m}] \]

\[ = \frac{\sqrt{P}}{M} \sum_{k=0}^{M-1} e^{j\theta_k} \Pr[x = \sqrt{Pe^{j\theta}}e^{-j\theta_n}|r] \]

\[ = \frac{\sqrt{P}}{M} e^{j\theta_m} \sum_{k=0}^{M-1} e^{j\theta(k-m)} \Pr[x = \sqrt{Pe^{j\theta(k-m)}}|r] \]

\[ = e^{j\theta_m} E[x|r] \]

C. Zero Error Correlation of EF

**Theorem 6:** Error at the relays that estimate and forward are uncorrelated with each other if MPSK modulation is employed at the source.

Proof: Expressing error as the difference of the estimate and the actual symbol, we have

\[ C = E[e_1e_2^*] \]

\[ = E[\hat{X}(r_1) - x)(\hat{X}^*(r_2) - x^*)] \]

\[ = \frac{P^2 (E[x|r_1]E[x^*|r_2])}{E[x^*E(x|r_1 = g_1x + n_1)E[x^*E(x^*|r_2 = g_2x + n_2)]} - P. \]

\[ E[x^*E(x|x_1)] = \frac{1}{M} \sum_{i=0}^{M-1} E_n \text{E}[x|r = g_1x_i + n_1]E_n \text{E}[x^*|r_2 = g_2x_i + n_2] \]

\[ = \frac{1}{M} \sum_{i=0}^{M-1} E_n \text{E}[x|r = g_1x_i + n]E_n^* \text{E}[x^*|r = g_2x_i + n] \]

\[ = E_n \text{E}[x|r = g_1x_0 + n]E_n^* \text{E}[x^*|r = g_2x_0 + n], \]

where (47) is obtained by applying Theorem 5 in (46).

\[ E[x^*E(x|x_1)] = \frac{1}{M} \sum_{i=0}^{M-1} x_i E_n \text{E}[x|r = x_i + n] \]

\[ = \sqrt{P} \frac{1}{M} \sum_{i=0}^{M-1} E_n \text{E}[x|r = x_0 + n] \]

\[ = \sqrt{P} E_n \text{E}[x|r = g_1x_0 + n], \]
where (48) is reduced to (49) using Theorem 5. Substituting (47) and (49) in (45), we have

\[
C = \mathcal{E}[e_1 e_2^*] \\
= \frac{P^2 \mathcal{E}_n \mathcal{E}[x|r = g_1 x_0 + n] \mathcal{E}_n^* \mathcal{E}[x|r = g_2 x_0 + n]}{P \mathcal{E}_n \mathcal{E}[x|r = g_1 x_0 + n] \mathcal{E}_n^* \mathcal{E}[x|r = g_2 x_0 + n]} - P \\
= 0.
\]

Hence proved. 

\[\Box\]

ACKNOWLEDGMENT

We would like to thank the anonymous reviewers for their helpful comments and suggestions that greatly improved the quality of the paper.

REFERENCES


