Optimizing Joint Erasure- and Error-Correction Coding for Wireless Packet Transmissions

Christian R. Berger, *Student Member, IEEE,* Shengli Zhou, *Member, IEEE,*
Yonggang Wen, *Student Member, IEEE,* Peter Willett, *Fellow, IEEE,* and Krishna Pattipati, *Fellow, IEEE*

***Abstract*—To achieve reliable packet transmission over a wireless link without feedback, we propose a layered coding approach that uses error-correction coding within each packet and erasure-correction coding across the packets. This layered approach is also applicable to an end-to-end data transport over a network where a wireless link is the performance bottleneck. We investigate how to optimally combine the strengths of error- and erasure-correction coding to optimize the system performance with a given resource constraint, or to maximize the resource utilization efficiency subject to a prescribed performance. Our results determine the optimum tradeoff in splitting redundancy between error-correction coding and erasure-correction codes, which depends on the fading statistics and the average signal to noise ratio (SNR) of the wireless channel. For severe fading channels, such as Rayleigh fading channels, the tradeoff leans towards more redundancy on erasure-correction coding across packets, and less so on error-correction coding within each packet. For channels with better fading conditions, more redundancy can be spent on error-correction coding. The analysis has been extended to a limiting case with a large number of packets, and a scenario where only discrete rates are available via a finite number of transmission modes.**

I. INTRODUCTION

WIRELESS access to the Internet is by now very popular thanks to the success of wireless local area networks (WLANs). Also, in wireless sensor networks, sensor data is transmitted first via wireless links and then via wired links to the data fusion center. Wireless links are usually the bottleneck in such systems, due to the multipath fading effects and interference-rich environments.

In this paper, we analyze a wireless link in terms of both throughput and error rate. The considered system setup has the following characteristics:

- The channel of the wireless link is slowly fading. Due to channel fading, there is a non-zero outage probability of the channel not being able to support any prescribed rate.
- The feedback delay is very large; this prevents instant channel state feedback and the usage of any automatic repeat-request (ARQ) based solutions, e.g., [1], [2], [3].

While the fading channel assumption applies to most wireless scenarios, the assumption of large feedback delays is typical for satellite channels and also for acoustic underwater communication channels (see [4] and references therein). For such a system setup, we propose the following layered coding strategy:

1) Error-correction coding on a per packet basis, where the data stream is partitioned into packets with each packet encoded separately. Any packet with decoding errors will be discarded.
2) Erasure-correction coding (such as digital fountain codes [5], [6], [7]) across the data packets, viewing the underlying wireless channel as an erasure channel.

For this setup, we term the error-correction coding as intra-packet coding, or inner layer coding, and the erasure-correction coding as inter-packet coding, or outer layer coding. The inter-packet code treats the decoding errors of intra-packet coding as erasures, and can facilitate data recovery without requesting retransmission of lost data packets.

We would like to point out that our results can be applicable to a network scenario. Consider an end-to-end connection that consists of both wired and wireless links, and assume that the wireless link is the performance bottleneck. Error-correction coding is used to improve the error performance on the wireless link, while erasure-correction coding can be used to handle lost packets on the end-to-end connection level. The layered coding is well motivated because: i) When transmitting over a series of links, the inner layer coding is applied on a single link (our attention here is focused on the wireless link), while the outer layer coding is applied on the end-to-end connection. From the viewpoint of a single link, each packet is a separate transmission and is therefore processed separately. This reduces necessary buffer space at each link and avoids additional delays as decoding across packets is only done at the final destination. ii) The assumption of a large feedback delay could also be motivated in view of an end-to-end connection in a large network across many links.

We are mostly concerned with the problem of judiciously combining the strengths of intra-packet and inter-packet coding in this layered coding setup. In particular, we would like to know: given an overall efficiency (resource constraint), how should we split it into inter-packet coding and intra-packet

Manuscript received on the June 1, 2007, revised November 9, 2007 and March 10, 2008, accepted May 25, 2008. The editor responsible for the handling of this paper was Prof. B. Sundar Rajan. C. R. Berger and P. Willett are supported by the Office of Naval Research (ONR) grant N00014-07-1-0429. S. Zhou is supported by the ONR YIP grant N00014-07-1-0805. This work was presented in part at the 2007 IEEE Communication Theory Workshop, Sedona, Arizona.

C. R. Berger, S. Zhou, P. Willett and K. Pattipati are with the Department of Electrical and Computer Engineering, University of Connecticut, 371 Fairfield Way U-2157, Storrs, Connecticut 06269, USA (email: {crberger, shengli, willett, krishna}@engr.uconn.edu).

Y. Wen is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 (email: eewyg@mit.edu).

Digital Object Indentifier 10.1109/TW.2008.070581

1536-1276/08$25.00 © 2008 IEEE

coding for best system performance? Alternately, to meet a prescribed performance, how should we optimally combine inter-packet coding and intra-packet coding to minimize the total resources needed? We study three different scenarios. In the first, we consider capacity-achieving error-correcting coding with continuously variable coding rate; we investigate the optimal solutions in terms of the efficiency and transmission outage probability. In the second scenario, we investigate the limiting case of the first scenario where the number of data packets approaches infinity. We characterize the optimal solution in analytical form for some special cases. Interestingly, this scenario has an equivalent formulation as the case of maximizing throughput over an instant feedback channel. In the third scenario, we consider practical modulation and coding schemes where only discrete transmission rates are available.

In the presence of erasure-correction coding across packets, it is beneficial to maximize the number of successfully received packets across the wireless link, and it is irrelevant which packets are received as the message can be reconstructed from any combination of received packets[1]. For adverse channel conditions such as Rayleigh fading, heavier redundancy should be placed on the outer erasure-correction coding, transmitting many redundant packets. Instead of traditional schemes targeting packet error rates (PER) around or below 10^{-2}, the layered approach leads to a PER at the wireless link of above 10^{-1}. In Nakagami-m channels with improved fading conditions, more redundancy can be placed on the inner error-correction coding, with PER of above 10^{-2} for the practical ranges of signal to noise ratio (SNR). For practical modulation schemes with a limited set of transmission rates available, the results are approximations of the findings based on continuous rates, and there is only a small penalty on the achieved throughput.

Note that erasure-correction codes can be also used in noisy channels directly, see e.g., [8]. However, this requires maximum-likelihood decoding, e.g., belief propagation algorithms, which are highly complex and require processing all the noisy packets received. Given our context, it is also not desirable to forward quantized noisy packets when the wireless transmission could be just one hop in an end-to-end communication.

The rest of the paper is organized as follows. In Section II we introduce the general problem setup, in Section III we investigate the optimal tradeoff, then in Section IV we study the limiting case with infinite number of packets, and in Section V we work on systems with practical modulation and coding schemes. Lastly we conclude in Section VI.

II. System Model and Problem Statement

We consider the transmission of a finite-size data block over a wireless link. For simplicity, we consider a flat-fading wireless link with channel input-output relationship as

$$y = hs + w \tag{1}$$

where s is the transmitted signal, y is the received signal, w is additive white Gaussian noise (AWGN) with variance N_0 and h is a channel gain which is constant over each packet, but drawn independently between packets, as will be formalized soon. We assume that the feedback delay is large and hence any feedback mechanism such as ARQ cannot be used, nor is instant channel state information available at the sender.

For reliable transfer, we consider a layered coding approach, where an erasure-correction-code is applied across data packets (inter-packet coding) and an error-correction code is applied within each data packet (intra-packet coding). Our objective is to study how to complement the strengths of erasure-correction coding and error-correction coding for optimal performance given some fixed resource, or for maximal resource utilization efficiency subject to a prescribed performance.

We now specify the details of the system model and the problem formulation. The data block of N_d bits is partitioned into k packets with N_b bits per packet:

$$N_d = kN_b. \tag{2}$$

An erasure-correction code (e.g., digital fountain code [5], [6], [7]) is used to code across these k packets to generate a stream of K encoded packets.

Intra-packet error-correction coding is applied to improve the error performance. Suppose that each packet consists of N_s information symbols after error-correction coding and modulation. The physical data rate is accordingly

$$R_p = \frac{N_b}{N_s} \ \left[\text{bits/symbol}\right]. \tag{3}$$

In this paper, we assume that

Assumption 1 *The channel is constant for the duration of one packet and independent across packets (a.k.a. block fading), and the channel amplitude $|h|$ has a Nakagami-m distribution with a probability density function (pdf) of $\lambda = |h|^2$ as*

$$f(\lambda) = \frac{m^m}{\Gamma(m)}\lambda^{m-1}e^{-m\lambda}.$$

Let $\gamma = E\left[|s|^2\right]/N_0$ denote the average signal to noise ratio (SNR). Due to the block fading assumption, packets will experience packet errors no matter what error-correction codes are used inside each packet. Assuming capacity-achieving Gaussian codebooks[2] with flexible encoding rate (the second assumption will be relaxed later), correct inner layer decoding is achieved iff the mutual information $I = \log_2(1 + \gamma|h|^2)$ is above the transmission rate. Hence, the packet error rate (PER) can be well approximated by

$$\text{PER} = \Pr\left(I < R_p\right) = \Pr\left(|h|^2 < \alpha\right) \tag{4}$$

where $\alpha = \left(2^{R_p} - 1\right)/\gamma$. So the PER is simply the probability distribution function of $|h|^2$; for Nakagami-m fading channels

[1]This is different from the ARQ principle, where lost packets have to be re-transmitted until correct reception.

[2]This assumption is not very limiting, as capacity achieving codes have been reported, i.e., Turbo and low density parity check (LDPC) codes, see e.g., [9].

this is [10]

$$\text{PER} = \int_0^\alpha \frac{m^m}{\Gamma(m)} \lambda^{m-1} e^{-m\lambda}\, d\lambda \tag{5a}$$

$$= 1 - \sum_{k=0}^{m-1} \frac{1}{k!} (m\alpha)^k\, e^{-m\alpha}. \tag{5b}$$

The probability of correct transmission of a packet across the wireless channel is

$$q = 1 - \text{PER} = \sum_{k=0}^{m-1} \frac{1}{k!} (m\alpha)^k\, e^{-m\alpha}. \tag{6}$$

The ergodic capacity of a Nakagami fading channel is defined as the average mutual information [11, Eq. (20)]:

$$C(\gamma, m) = E\left[\log_2(1+\gamma|h|^2)\right] \tag{7a}$$

$$= \int_0^\infty \log_2(1+\gamma\lambda)\, \frac{m^m}{\Gamma(m)} \lambda^{m-1} e^{-m\lambda}\, d\lambda \tag{7b}$$

$$= \log_2(e)\, e^{m/\gamma} \sum_{k=0}^{m-1} \left(\frac{m}{\gamma}\right)^k \Gamma\left(-k, \frac{m}{\gamma}\right) \tag{7c}$$

where $\Gamma(a,x) = \int_x^\infty t^{a-1} e^{-t}\, dt$ is the "upper tail" of the incomplete Gamma function. When $m = 1$, the Nakagami channel reduces to a Rayleigh fading channel.

Since we are dealing with a fading wireless channel, we define the coding rate of the intra-packet code, the inner rate, as a non-vanishing fraction of the channel ergodic capacity, similar to the approach used in [12],

$$r_i = \frac{R_p}{C}, \tag{8}$$

where we dropped the parameterization of C on γ and m in (7a) to shorten notation. By the definition of ergodic capacity, error free transmission is possible for $r_i < 1$ if channel coding is done across infinitely many packets to experience all channel realizations. On the contrary, for coding done inside one packet, there is always a non-zero probability of decoding error for any $r_i > 0$. The ergodic capacity is used in (8), as a means of normalization on the physical data rate, as in [12], since C varies depending on the operating SNR and R_p shall change accordingly.

We further assume that

Assumption 2 *Error detection based cyclic redundancy checking (CRC) is perfect, with a sufficiently reliable CRC code on each packet. The packets with CRC errors are discarded.*

With digital fountain codes, the original data of N_d bits can be reconstructed with high probability if ρk packets are received correctly, where $\rho \geq 1$ reflects the decoding overhead incurred by the erasure coding scheme. For reasonable N_d, it is assumed that the overhead is about five percent, i.e., $\rho \approx 1.05$ [9, Chapter 50]. With this we define the code rate of the erasure-correction code as the outer rate:

$$r_o = \frac{\rho k}{K}, \tag{9}$$

which is a quantity between zero and one, since we will obviously need to send at least as many packets as we will need to decode. An outage (or failure) happens when fewer than ρk correct packets are received. The probability of outage is then

$$P_{\text{outage}} = \sum_{i=0}^{\rho k-1} \binom{K}{i} q^i \cdot (1-q)^{K-i}. \tag{10}$$

In summary, since the data block of N_d bits travels through the wireless link via KN_s symbols and each symbol could carry a maximum of C bits on average, the overall efficiency for the data transfer is

$$\eta = \frac{N_d}{KN_sC} = \frac{k}{K} \cdot \frac{N_b}{N_sC} = \frac{1}{\rho}\ r_o \cdot r_i, \tag{11}$$

while the data transfer has an outage (or failure) probability specified in (10).

The definition of efficiency in (11) suggests two basic approaches to improving the overall efficiency. On the one hand, one can spend a small amount of redundancy coding each packet, ($r_i \approx 1$) and rely heavily on generating a lot of redundant data packets. In this case, the PER will be high and many packets will get lost when passing through the wireless channel. On the other hand, one can rely on strong FEC ($r_i \ll 1$) to improve the packet error rate, but that translates into a smaller number of data packets generated in total; hence the data transfer becomes more vulnerable to packet loss. It is unclear which strategy is better, and clearly there exists a tradeoff between these two different coding techniques.

We are motivated to investigate the following two dual problems:

1) Given a specified overall efficiency (resource constraint), how should we split it into inter-packet coding (r_o) and intra-packet coding (r_i) for the lowest outage probability?
2) To meet a prescribed outage probability (performance constraint), how should we optimally combine the strengths of inter-packet coding (r_o) and intra-packet coding (r_i) to maximize the resource utilization efficiency? (Maximal efficiency means minimal resources needed for a finite-size data transfer.)

III. Optimal Combining of Inter- and Intra-Packet Coding

When ρk and K are large, it becomes difficult to evaluate the outage probability in (10). According to the DeMoivre-Laplace Theorem [10], one can approximate the binomial distribution used in (10) as a Gaussian distribution with mean Kq and variance $Kq(1-q)$. The outage probability in (10) is then approximated as

$$P_{\text{outage}} \approx \int_{-\infty}^{\rho k} \frac{1}{\sqrt{2\pi Kq(1-q)}} \exp\left(-\frac{(x-Kq)^2}{2Kq(1-q)}\right)\, dx$$

$$= Q\left(\frac{Kq - \rho k}{\sqrt{Kq(1-q)}}\right). \tag{12}$$

According to [10, pg. 105-109], this approximation is accurate if *i)* $K \gg 1$, *ii)* $Kq \gg 1$, and *iii)* $|Kq - \rho k|$ is on the order of a few standard deviations, $\sqrt{Kq(1-q)}$. For a reasonable coding scenario we can assume $k \gg 1$ and with (9) we have

$K > k \gg 1$, since $r_o \leq 1$ and $\rho \geq 1$. Also we are only interested in small P_{outage}, e.g., $10^{-1} \ldots 10^{-3}$, which means that $Kq - \rho k$ is between two to four standard deviations, and $Kq > k \gg 1$.

After a few substitutions, we simplify (12) as

$$P_{\text{outage}} = Q\left(\sqrt{\frac{\rho N}{N_s C}} \frac{q - r_o}{\sqrt{r_o r_i q(1-q)}}\right). \tag{13}$$

Note that q depends on r_i via (8) and (6).

A. Optimizing Performance under Resource Constraint

We will first look into the problem of performance optimization subject to a resource constraint, as formulated as

$$\begin{aligned} &\text{minimize } P_{\text{outage}}, \\ &\text{subject to } \eta \geq \eta_0, \end{aligned} \tag{14}$$

where η_0 is prescribed. Since $Q(\cdot)$ is a monotonically decreasing function of its argument, the problem in (14) is equivalent to

$$\begin{aligned} &\text{maximize } J(r_i, r_o) := \frac{q - r_o}{\sqrt{r_i r_o (1-q) q}}, \\ &\text{subject to } r_i r_o \geq \rho \eta_0. \end{aligned} \tag{15}$$

Using the standard Lagrangian method, we formulate the objective function as:

$$L(r_i, r_o, \lambda) = J(r_i, r_o) - \lambda\left(r_i r_o - \rho\eta_0\right), \tag{16}$$

where λ is the Lagrangian multiplier. Setting $\partial L/\partial r_i = \partial L/\partial r_o = 0$ leads to

$$\lambda = \frac{1}{r_o}\frac{\partial J}{\partial r_i} = \frac{1}{r_i}\frac{\partial J}{\partial r_o}. \tag{17}$$

The requisite partial derivatives are

$$\frac{\partial J}{\partial r_i} = r_o \frac{r_i q(1-q)\dot{q} - \frac{1}{2}(q - r_o)\left(q(1-q) + r_i[\dot{q} - 2\dot{q}q]\right)}{[r_i r_o q(1-q)]^{\frac{3}{2}}} \tag{18}$$

$$\frac{\partial J}{\partial r_o} = r_i \frac{r_o q(1-q)(-1) - \frac{1}{2}(q - r_o) q(1-q)}{[r_i r_o q(1-q)]^{\frac{3}{2}}}, \tag{19}$$

where $\dot{q}$ is the derivative of q with respect to r_i. Based on (6), we have

$$\dot{q} = \frac{\partial q}{\partial \alpha}\frac{\partial \alpha}{\partial r_i} = -\frac{m^m}{\Gamma(m)}\alpha^{m-1}e^{-m\alpha}\frac{\ln(2)C}{\gamma}2^{Cr_i}. \tag{20}$$

Substituting (18) and (19) into (17), we obtain

$$r_o = \frac{-r_i q \dot{q}}{2q(1-q) - r_i(2q-1)\dot{q}} \tag{21}$$

The value for r_i can be solved numerically from the constraint

$$r_i r_o = \frac{-r_i^2 q\dot{q}}{r_i(2q-1)\dot{q} - 2q(1-q)} = \rho\eta_0. \tag{22}$$

In short, we find the optimal (r_i, r_o) based on (21) and (22).

Remark 1 Multiplying (17) by $r_i r_o$ leads to

$$r_i \frac{\partial J}{\partial r_i} = r_o \frac{\partial J}{\partial r_o}, \tag{23}$$

which is equivalent to

$$\frac{\partial J}{\partial \ln r_i} = \frac{\partial J}{\partial \ln r_o}. \tag{24}$$

Hence, the optimal (r_i, r_o) operates at the point where the increments of $\ln r_i$ and $\ln r_o$ lead to the same growth in the objective function.

B. Maximizing Efficiency under Performance Constraint

We now look into the problem of maximizing the resource utilization efficiency subject to a prescribed performance:

$$\max_{\{r_i, r_o\}} r_i r_o \quad \text{subject to } P_{\text{outage}} \leq \text{constant} \tag{25}$$

The problem in (25) is the dual problem of (14). Since adding the coding redundancy can only improve the system performance, P_{outage} shall be monotonically decreasing when $r_i r_o$ decreases. As such, we can solve (25) based on a bisectional search on $r_i r_o$, where for each tentative point $r_i r_o$ we obtain (r_i, r_o) via the solution of (14). The final solution of (r_i, r_o) is found when the performance constraint is met.

C. Numerical Examples

1) Resource Constrained Optimization: To give some numerical insight, we elaborate on an example. For the purpose of simulation we define

$$K_0 := \frac{\rho N_d}{C N_s} = 2^8 \tag{26}$$

which is the number of packets needed when transmitting at capacity. Keeping this ratio constant, the data size scales with the channel capacity and we do not need to explicitly consider the coding overhead ρ. Further, we define the numbers of symbol per packet and the prescribed efficiency:

$$N_s = 2^8 \tag{27}$$

$$\rho\eta_0 = 0.5 \tag{28}$$

The fading channel experiences Rayleigh fading $(m = 1)$ which is constant for the duration of one packet as described in Section II, where the capacity C for a given SNR is calculated via (7c). In the simulation, for each level of SNR, we iterate through r_i and calculate the following parameters:

$$N_b = \lfloor r_i C \rfloor \tag{29a}$$

$$\rho k = \left\lceil \frac{\rho N_d}{N_b} \right\rceil \tag{29b}$$

$$K = \left\lfloor \frac{\rho k}{\rho\eta_0} r_i \right\rfloor \tag{29c}$$

If the number of transmitted packets K is smaller than the number of correctly received packets necessary to decode ρk, the outage probability is naturally set to one.

The PER and outage probability are evaluated via (5b),(10) and we plot the outage probability in Fig. 1. First we notice that for r_i below 0.5 there is no possible r_o, since r_o can't be larger than unity and $\rho\eta_0 = r_o r_i = 0.5$, leading to an outage probability of one. Otherwise, the tradeoff between the redundancy of error-correction coding and the redundant

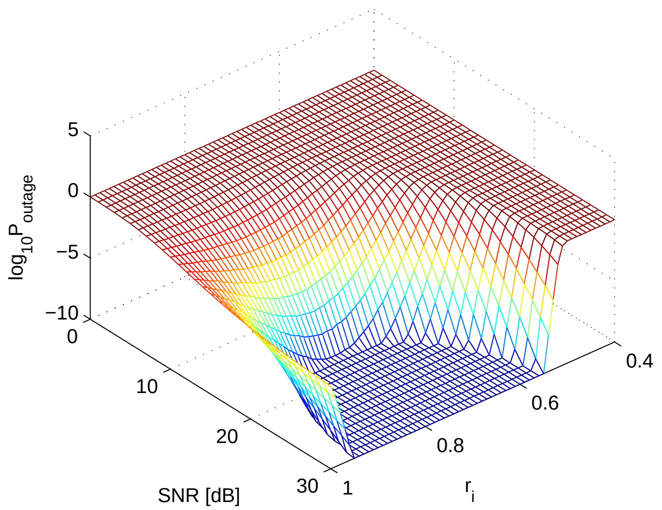

Fig. 1. For a prescribed efficiency $\rho\eta_0 = 0.5$, the outage probability can be determined for each SNR and pair of (r_i, r_o); values below 10^{-10} are displayed as 10^{-10}, there are 256 symbols per packet and the data size scales with the capacity as in (26).

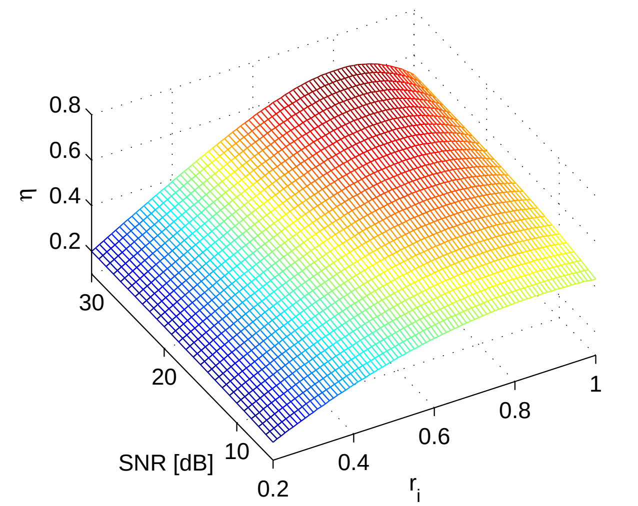

Fig. 3. The maximum achievable efficiency for a given outage probability of 10^{-2} shows a clear maximum for each SNR; there are 256 symbols per packet and the data size scales with the capacity as in (26).

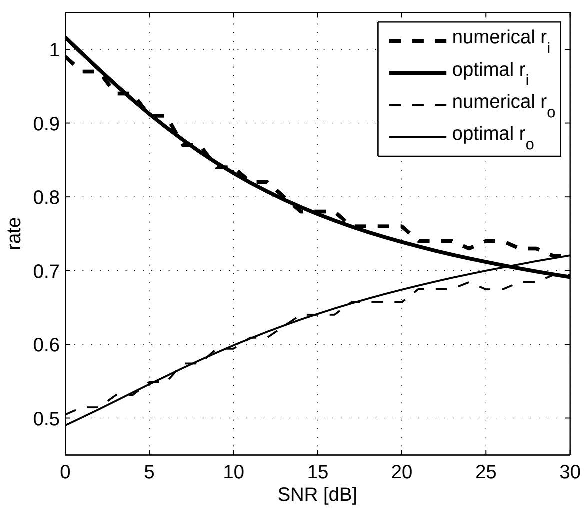

Fig. 2. Minimizing P_{outage} over r_i gives the following pairs of (r_i, r_o) for each value of SNR.

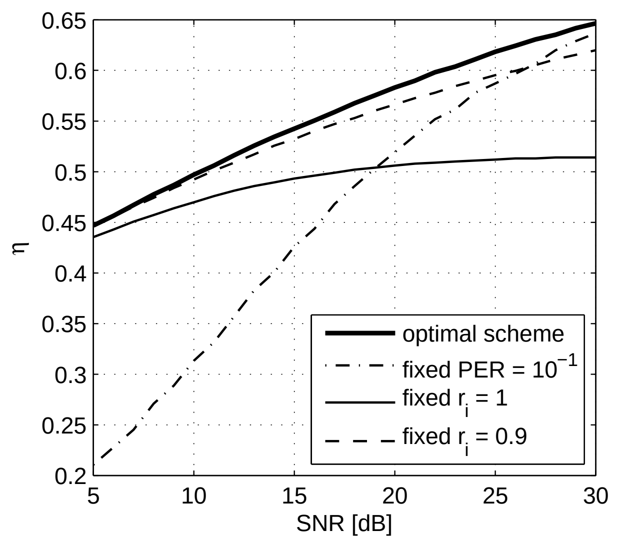

Fig. 4. Comparison of the optimum efficiency with the efficiency of sub-optimum schemes, configuring the inner layer coding to either achieve a constant PER or to transmit at a fixed fraction of capacity.

packets for erasure coding leads to a clear global maximum for each fixed SNR.

We pick the r_i which returns the smallest P_{outage} for each SNR, and show the result in Fig. 2, together with the corresponding r_o. Fig. 2 also shows the results derived via (21) based on the Gaussian approximation in (12). The curves based on the Gaussian approximation are smooth, since no quantization in K, k and ρk is taken into account. The values obtained via Gaussian approximation agree very well with the optimal values obtained via brute-force search.

2) Performance Constrained Optimization: We look at a similar numerical example, but fix

$$P_{\text{outage}} = 10^{-2}. \tag{30}$$

The other parameters stay unchanged as defined in (26),(27). We start with Rayleigh fading $(m = 1)$, but will extend to Nakagami fading. We evaluate N_b, ρk, K as in (29a)-(29c) and the PER from (5b).

a) Optimal Efficiency: To determine the achievable η while satisfying the given outage probability, we use numerical search. We evaluate P_{outage} via (10) for increasing K; then the efficiency can be calculated as, c.f. (26),

$$\eta = K_0/K. \tag{31}$$

The results are shown in Fig. 3, where the plot appears concave and shows a clear global maximum. We compare this to some sub-optimum schemes in Fig. 4, where the physical link is fixed to either having a certain packet error rate, e.g., PER$=10^{-1}$ or to transmit at a certain fraction of capacity, e.g., $r_i = 1$ or $r_i = 0.9$. We see that keeping the PER fixed is not reasonable for low SNR, as on a Rayleigh channel this leads to

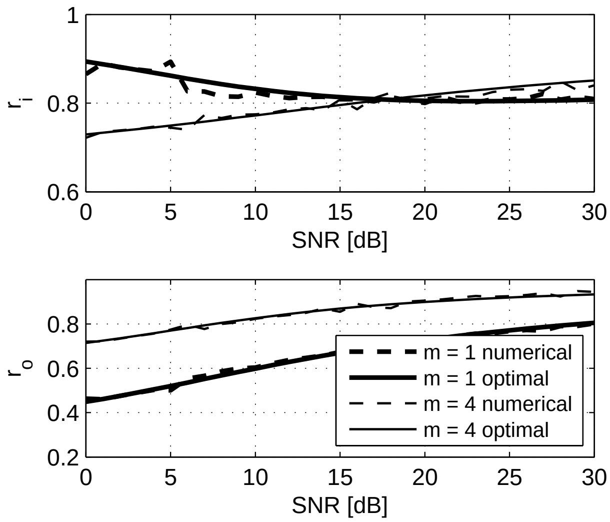

Fig. 5. Displayed are the r_i and r_o which lead to the highest efficiency for a given SNR, to achieve an outage probability of 10^{-2}, for Rayleigh ($m = 1$) and Nakagami-4 fading channels.

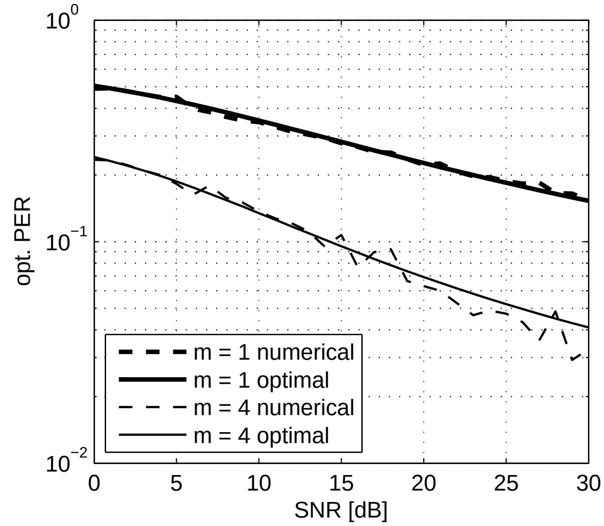

Fig. 6. Comparison of packet error rates corresponding to the optimal r_i for Rayleigh ($m = 1$) and Nakagami-4 fading channels.

a very low physical throughput. On the other hand transmitting at a fixed fraction of capacity does better than fixing the PER; and the question is then to decide what fraction to use.

Next we plot the optimal coding rates r_i and r_o as the "$m = 1$" plots in Fig. 5. The brute-force search results match very well with results based on (21) from Gaussian approximation. The plots confirm that the optimal r_i for Rayleigh fading channels is rather constant for medium to high SNR. Generally, we see that the outer rate r_o starts very low, then increases, while the inner rate r_i is slowly decreasing.

b) Rayleigh vs. Nakagami Fading Channels: Next, we simulate the same setup using a Nakagami-m fading channel with $m = 4$, which implies less severe channel fading conditions. Comparison of the optimal and numerical rates to those of the Rayleigh case are plotted in Fig. 5. We see that the Nakagami fading channel leads to a much higher outer rate, as fewer packets are lost. The inner rate varies only slightly across the considered SNR range, but compared to the Rayleigh case starts with a lower value and shows a monotonically increasing behavior. This reveals that for a less severe fading environment it pays to go with a lower rate at the inner layer to achieve a reliable transmission, especially for low SNR.

Further, we evaluate the PERs corresponding to the optimal r_i (see Fig. 6). While for Rayleigh fading ($m = 1$) the PER decreases very slowly, reaching barely 10^{-1} at 30 dB, for Nakagami-m fading ($m = 4$) the PER is smaller. In general, the "raw" PER at the wireless link is quite high, demonstrating the utilization of a second level of redundancy across the packets.

IV. Rate Optimization in a Special Case

We now consider a special case where the data size approaches infinity: $N_d \to \infty$. We shall find an explicit solution for this special case, and show that the optimal solution in Section III of finite N_d has the same trend as that of infinite N_d.

For a given SNR and a finite packet size N_s, the outage probability will go to zero as long as $r_o < q$, since

$$\lim_{N_d \to \infty} Q\left(\sqrt{\frac{\rho N_d}{N_s C}} \frac{q - r_o}{\sqrt{r_o r_i q(1-q)}}\right) = 0. \tag{32}$$

Intuitively, as N_d goes to infinity, we will have $K \to \infty$ packets transmitted through the wireless channel. By the law of large numbers, the receiver will have Kq correct packets *almost surely*. As long as $Kq \geq \rho k$, the receiver decodes the whole message back. In this case, the maximum network rate shall be equal to the transmission success probability, i.e.,

$$r_o = q. \tag{33}$$

The optimization problem (34) becomes

$$\max_{r_i} \; r_i q. \tag{34}$$

Remark 2 Let us consider a different setup where an immediate ARQ is available on top of error-correction coding for each packet. Let Cr_i denote the transmission rate for each packet after coding, and $1 - q$ denotes the packet error rate. The system throughput is then $Cr_i q$, which can be maximized by adjusting the rate of the error-correcting code. This is mathematically equivalent to the problem in (34). Our results in this section are hence applicable to such an system with combined ARQ and error-correcting coding.

We have the following analytical solution to the optimization problem in (34).

Result 1 *On a Rayleigh fading channel, the inner layer rate maximizing* (34) *is given by:*

$$r_i = \frac{W(\gamma)}{\ln(2) C}, \tag{35}$$

	$m=1$	$m=2$	$m=3$	$m=4$	$m=6$	$m=8$	$m=12$	$m=16$
r_i	1	$\frac{1+\sqrt{5}}{4}\approx 0.809$	≈ 0.757	≈ 0.736	≈ 0.725	≈ 0.726	≈ 0.735	≈ 0.746
r_o	$e^{-1}\approx 0.368$	≈ 0.519	≈ 0.604	≈ 0.660	≈ 0.729	≈ 0.771	≈ 0.820	≈ 0.849

TABLE I
OPTIMAL RATES FOR VANISHING SNR.

where $W(\gamma)$ is the Lambert-W function [13]. This leads to the network rate as:

$$r_o = \exp\left[-\frac{1}{W(\gamma)} + \frac{1}{\gamma}\right]. \tag{36}$$

At high SNR, we have

$$\lim_{\gamma\to\infty} r_i = 1, \quad \lim_{\gamma\to\infty} r_o = 1. \tag{37}$$

Proof: Setting the derivative of $r_i q = 0$ leads to

$$q + r_i \dot{q} = 0. \tag{38}$$

Substituting (6) and (20) into (38), we obtain

$$\exp\left(-\frac{2^{r_i C}-1}{\gamma}\right)\left[1 - r_i \ln(2) C \frac{2^{r_i C}}{\gamma}\right] = 0. \tag{39}$$

Since the exponential function is not equal to zero, the optimal r_i satisfies

$$\gamma = \ln(2) C r_i \, e^{\ln(2) C r_i}. \tag{40}$$

Using the Lambert-W function $W(y)$ to denote the solution to $y = xe^x$, we obtain (35). Using (5b), we obtain r_o in (36).

When $\gamma \to \infty$, the capacity is $C \approx \log_2(\gamma)$. Hence, we have

$$\lim_{\gamma\to\infty} r_i = \lim_{\gamma\to\infty} \frac{W(\gamma)}{\ln(\gamma)}. \tag{41}$$

Since both numerator and denominator diverge, we use l'Hôpital's rule,

$$\lim_{\gamma\to\infty} r_i = \lim_{\gamma\to\infty} \frac{W(\gamma)}{\gamma\,[W(\gamma)+1]} \Big/ \frac{1}{\gamma} = 1, \tag{42}$$

where the derivative of the Lambert-W function is given in [13]. The limit of r_o for high SNR can be directly found from (36), as $W(\gamma) \to \infty$ for $\gamma \to \infty$. ∎

Result 2 *For general Nakagami-m fading channels, the optimal rates of the problem in* (34) *at a vanishing SNR are given in Table I.*

Proof: For vanishing SNR, we use the approximation $\ln(1+x) \approx x$ to calculate the capacity:

$$C = \int_0^\infty \frac{\ln(1+\gamma\lambda)}{\ln(2)} f_m(\lambda)\, d\lambda \approx \frac{\gamma}{\ln(2)} \int_0^\infty \lambda f_m(\lambda)\, d\lambda = \frac{\gamma}{\ln(2)} \tag{43}$$

since the mean of a Nakagami-m variable is one. We then have

$$\lim_{\gamma\to 0} \alpha = \lim_{\gamma\to 0} \frac{e^{r_i\gamma}-1}{\gamma} = \lim_{\gamma\to 0} \frac{r_i e^{r_i \gamma}}{1} = r_i \tag{44}$$

via l'Hôpital's rule. Substituting (6) and (20) into (38), we obtain

$$e^{-m r_i}\left[\sum_{k=0}^{m-1} \frac{1}{k!}(m r_i)^k - \frac{1}{(m-1)!}(m r_i)^m\right] = 0. \tag{45}$$

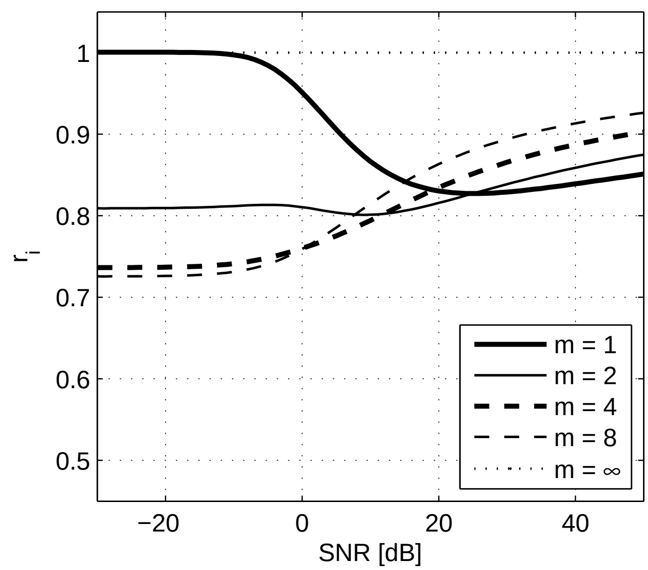

Fig. 7. Optimal r_i as a function of SNR for different Nakagami-m channel models, in the case of infinite data block length $N_d \to \infty$.

For different m, we solve (45) numerically to get r_i as shown in Table I. We then obtain $r_o = q$ using (6). ∎

The inner layer rates maximizing (34) for general Nakagami-m fading can be seen in Fig. 7, where we used numerical maximization except for the analytic results in the Rayleigh case. It is interesting to observe that the general shape of the curves changes for $m > 2$, similar to what we observed in Fig. 5 (finite data case). For high SNR all curves seem to rise monotonically, while for vanishing SNR they approach a constant value as predicted by Result 2.

Numerically solving (45), the values are given in the Table of Result 2.

the values are given in the Table of Result 2. We find the optimal inner rate r_i is decreasing for $m < 8$ and slowly increases again afterwards, reaching a theoretic limit of $r_i = 1$ for $m \to \infty$; the optimal outer rate r_o is monotonically increasing. Therefore for severely fading channels, i.e., $m \approx 1$, it is better to invest less redundancy on the inner layer coding (or transmit close to capacity) for vanishing SNR. As the gain on these channels in terms of packet error rate per redundancy increase in error-correction is rather low, more redundant packets in the outer layer, erasure-correction, are preferable.

Fig. 8 compares the optimal rates of finite data lengths to those of infinite data length. In general, both r_i and r_o are smaller in the finite data length case, since $r_o = q$ is equivalent to an outage probability of 0.5, c.f., (13) – accordingly we need to increase redundancy to lower the outage probability. The

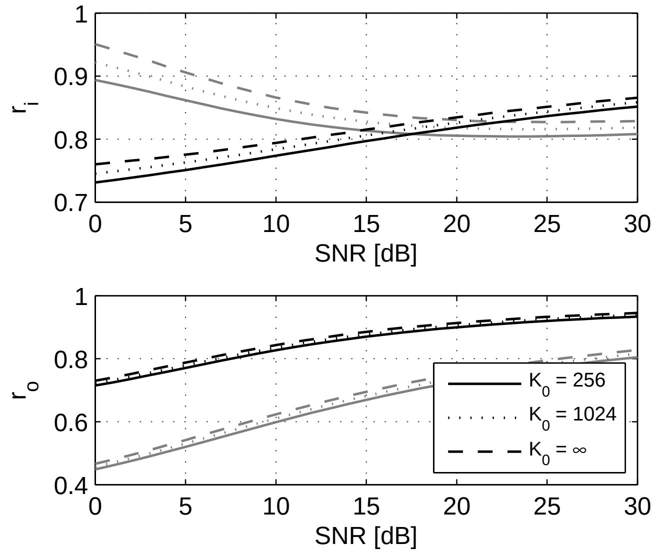

Fig. 8. Comparison of optimal r_i, r_o for different data lengths, for Rayleigh (gray) and Nakagami-4 (black) fading channels.

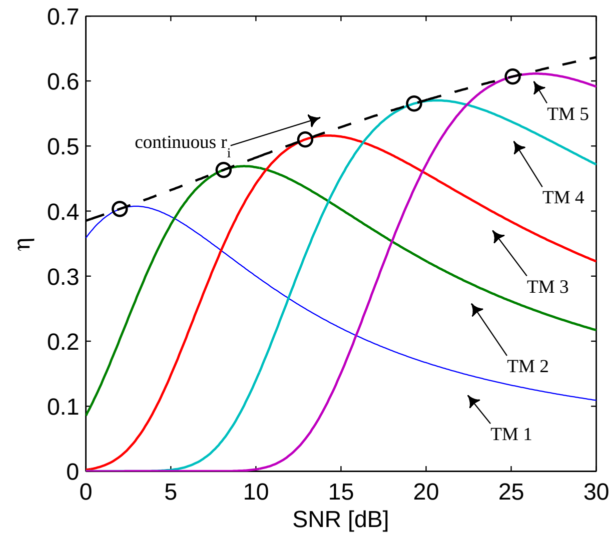

Fig. 9. For each discrete transmission mode (TM), the optimum performance is achieved for a limited SNR region ($P_{\text{outage}} = 10^{-2}$); modulation mode and coding rate of each TM are defined in Table II.

curves with finite K_0 approach the limiting case of infinite K_0 uniformly across the SNR range, showing similar behaviors.

V. Optimization with Discrete Transmission Rates

In Sections III and IV we have assumed continuously available inner layer coding rates. In practice, however, only discrete rates are available depending on the chosen modulation and coding pairs. In this section, we consider how the performance is affected when only a finite number of transmission modes are available.

The rate in a practical transmission mode is determined by the modulation scheme, i.e., the number of bits per symbol, and the coding rate which makes some fraction of the transmitted bits redundant to protect against errors. We will consider square quadrature amplitude modulation (QAM) with capacity-achieving channel coding (e.g., Turbo codes or LDPC codes [9]). Each symbol carries an even integer number (say b) of bits. For example, QPSK, 16-QAM,..., and 256-QAM have 2, 4,..., and 8 bits per symbol, respectively. With a code of rate R_c, the actual inner rate is $r_i = b \cdot R_c/C$. We consider the following rates listed in Table II.

When using discrete transmission modes, we cannot optimize over the continuous variable r_i. In fact, discrete transmission modes lead to a constant number of bits per symbol, $R_p = r_i C$, which defines curves on the (SNR, r_i)-plane among which we can choose. Since C is an increasing function of SNR and R_p is constant, the r_i on these curves are proportional to $1/C$ going from infinity to zero with increasing SNR.

Fig. 9 shows the efficiency η for all transmission modes evaluated as in (31). The outage probability is $P_{\text{outage}} = 10^{-2}$ as previously; other parameters stay also unchanged as in (26), (27) and evaluated further via (29a)-(29c), (5b). As comparison we included the optimal η for continuous inner coding rates from Fig. 4. We notice that each transmission mode achieves a local maximum for a certain SNR and is

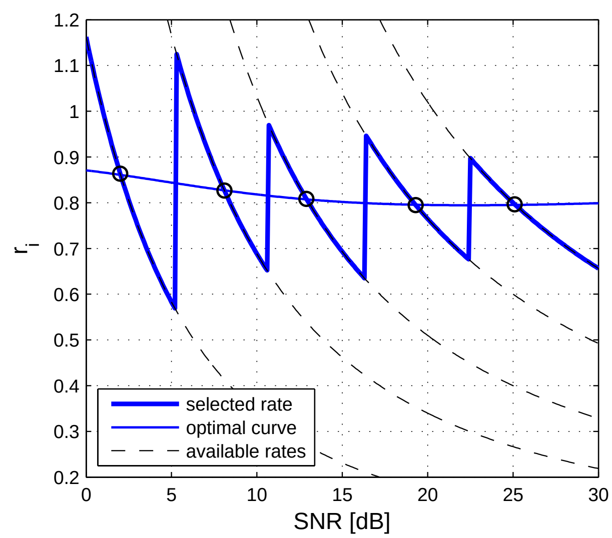

Fig. 10. Optimal inner coding rate r_i with discrete transmission modes, Rayleigh fading ($m = 1$) and $P_{\text{outage}} = 10^{-2}$.

suboptimum otherwise, since ultimately each r_i connected to a certain transmission mode will go to zero, since it is a decreasing function of SNR as explained before. The optimal efficiency from Fig. 4 is the convex hull of the discrete curves, as it stems from the maximization over all r_i.

Optimizing over the discrete set of transmission modes, the resulting inner rates is shown in Fig. 10; also included are the lines corresponding to constant R_p for each available transmission mode (dashed lines) and the optimal curves from Fig. 5. As the optimal inner coding rate is relatively constant, we observe that the actual inner rate jumps between the discrete available values to stay close for increasing SNR.

Although the actual inner coding rate is "oscillating" around the optimal rate due to the discrete rate constraint, the effective

TABLE II
DISCRETE TRANSMISSION MODES WITH CODING

	TM 1	TM 2	TM 3	TM 4	TM 5
Modulation	QPSK	16-QAM	16-QAM	64-QAM	256-QAM
Code Rate R_c	1/2	1/2	3/4	3/4	3/4
R_p [bits/sym]	1.0	2.0	3.0	4.5	6.0

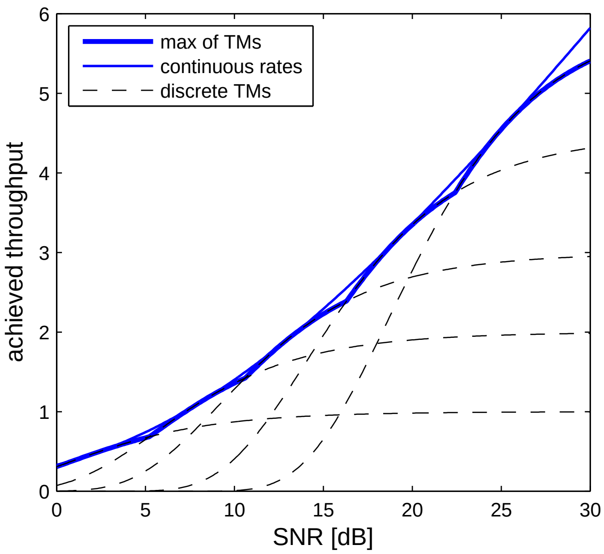

Fig. 11. Achieved throughput with discrete transmission modes, Rayleigh fading ($m = 1$) and $P_{\text{outage}} = 10^{-2}$.

system throughput Cr_ir_o stays fairly smooth (see Fig. 11). This, because the changes in the inner rate are compensated by the changes of the outer rate. Due to the finite number of transmission rates, the optimal throughput is not always achieved. However, with practical transmission modes the throughput is strictly increasing, and not far away from the optimal throughput for continuous rates.

VI. CONCLUSION

We investigated the tradeoff between erasure- and error-correction coding for packet transmissions over a wireless link. We presented solutions to optimally splitting the redundancy between these two codes. Our results show that on very unreliable channels, e.g., Rayleigh fading, the tradeoff leans towards more redundancy at the outer layer coding across packets, and less so on the inner layer coding within each packet. For channels with better fading conditions, more redundancy can be spent on the inner layer coding. Thanks to erasure-correction coding, less reliable transmission due to error-correction coding are needed to maximize the throughput over the wireless link. We have also extended our analysis to practical modulation and coding schemes with discrete rates.

REFERENCES

[1] S. Sesia, G. Caire, and G. Vivier, “Incremental redundancy hybrid ARQ schemes based on low-density parity-check codes,” *IEEE Trans. Commun.*, vol. 52, no. 8, pp. 1311–1321, Aug. 2004.

[2] E. Soljanin, N. Varnica, and P. Whiting, “Incremental redundancy hybrid ARQ with LDPC and Raptor codes,” *IEEE Trans. Inform. Theory*, Sept. 2005, submitted for publication; available at http://netlib.bell-labs.com/who/emina/papers/hybridarq_final.pdf.

[3] J.-F. Cheng, “Coding performance of hybrid ARQ schemes,” *IEEE Trans. Commun.*, vol. 54, no. 6, pp. 1017–1029, June 2006.

[4] J.-H. Cui, J. Kong, M. Gerla, and S. Zhou, “The challenges of building mobile underwater wireless networks for aquatic applications,” *IEEE Network, Special Issue on Wireless Sensor Networking*, vol. 20, no. 3, pp. 12–18, May/June 2006.

[5] M. Luby, M. Mitzenmacher, M. Shokrollahi, and D. Spielman, “Efficient erasure correcting codes,” *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 569–584, Feb. 2001.

[6] M. Luby, “LT codes,” in *Proc. 43rd Annual IEEE Symposium on Foundations of Computer Science*, Nov. 2002, pp. 271–280.

[7] M. Shokrollahi, “Raptor codes,” *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2551–2567, June 2006.

[8] O. Etesami and A. Shokrollahi, “Raptor codes on binary memoryless symmetric channels,” *IEEE Trans. Inform. Theory*, vol. 52, no. 6, pp. 2033–2051, May 2006.

[9] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*, 1st ed. Cambridge, UK: Cambridge University Press, 2003.

[10] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. New York: McGraw-Hill, 2002.

[11] M.-S. Alouini and A. J. Goldsmith, “Adaptive modulation over Nakagami fading channels,” *Wireless Personal Communications*, vol. 13, no. 1-2, pp. 119–143, Nov. 2004.

[12] L. Zheng and D. Tse, “Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels,” *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[13] E. W. Weisstein, “Lambert W-Function,” *MathWorld–A Wolfram Web Resource*, http://mathworld.wolfram.com/LambertW-Function.html.

PLACE
PHOTO
HERE

Christian R. Berger (S05) was born in Heidelberg, Germany, on September 12th, 1979. He received the Dipl.-Ing. degree in electrical engineering from the Universität Karlsruhe (TH), Karlsruhe, Germany in 2005. During this degree he also spent a semester at the National University of Singapore (NUS), Singapore, where he took both undergraduate and graduate courses in electrical engineering. He is currently working towards his Ph.D. degree in electrical engineering at the University of Connecticut (UCONN), Storrs.

In the summer of 2006, he was as a visiting scientist at the Sensor Networks and Data Fusion Department of the FGAN Research Institute, Wachtberg, Germany. His research interests lie in the areas of communications and signal processing, including distributed estimation in wireless sensor networks, wireless positioning and synchronization, underwater acoustic communications and networking.

PLACE PHOTO HERE

Shengli Zhou (M'03) received the B.S. degree in 1995 and the M.Sc. degree in 1998, from the University of Science and Technology of China (USTC), Hefei, both in electrical engineering and information science. He received his Ph.D. degree in electrical engineering from the University of Minnesota (UMN), Minneapolis, in 2002. He has been an assistant professor with the Department of Electrical and Computer Engineering at the University of Connecticut (UCONN), Storrs, since 2003.

His general research interests lie in the areas of wireless communications and signal processing. His recent focus has been on underwater acoustic communications and networking. He served as an Associate Editor for the IEEE Transactions on Wireless Communications from February 2005 to January 2007. He received the ONR Young Investigator award in 2007.

PLACE PHOTO HERE

Yonggang Wen (M'08) was born in Nanchang, Jiangxi, China in 1977. He received his BEng degree from the University of Science and Technology of China (USTC), Hefei, Anhui, China, in 1999, his MPhil degree from the Chinese University of Hong Kong (CUHK), Hong Kong, in 2001, and his PhD from Massachusetts Institute of Technology, in 2007.

He worked from 1997 to 1999 as a Research Assistant in the Personal Communication Network and Spread Spectrum Lab at USTC, with much attention to the G3 mobile communication. From 1999 to 2001, he was a Research Assistant in the Lightwave Communication Lab at CUHK, working on optical code-division multiplexing systems and high-rate optical routing networks. From 2001 to 2007, He was a Research Assistant at the Laboratory for Information and Decision Systems of MIT, working on fault detection/isolation/prognosis for optical networks. He is now working at Cisco. His research interests focus on mesh-based network architecture including ultra-reliable all-optical network architecture, efficient network management algorithms and schedule over packet-based networks.

PLACE PHOTO HERE

Peter Willett (F'03) received his BASc (Engineering Science) from the University of Toronto in 1982, and his PhD degree from Princeton University in 1986. He has been a faculty member at the University of Connecticut ever since, and since 1998 has been a Professor.

His primary areas of research have been statistical signal processing, detection, machine learning, data fusion and tracking. He has interests in and has published in the areas of change/abnormality detection, optical pattern recognition, communications and industrial/security condition monitoring.

He is editor-in-chief for IEEE Transactions on Aerospace and Electronic Systems, and until recently was associate editor for three active journals: IEEE Transactions on Aerospace and Electronic Systems (for Data Fusion and Target Tracking) and IEEE Transactions on Systems, Man, and Cybernetics, parts A and B. He is also associate editor for the IEEE AES Magazine, editor of the AES Magazines periodic Tutorial issues, associate editor for ISIFs electronic Journal of Advances in Information Fusion, and is a member of the editorial board of IEEEs Signal Processing Magazine. He has been a member of the IEEE AESS Board of Governors since 2003. He was General Co-Chair (with Stefano Coraluppi) for the 2006 ISIF/IEEE Fusion Conference in Florence, Italy, Program Co-Chair (with Eugene Santos) for the 2003 IEEE Conference on Systems, Man, and Cybernetics in Washington DC, and Program Co-Chair (with Pramod Varshney) for the 1999 Fusion Conference in Sunnyvale.

PLACE PHOTO HERE

Krishna R. Pattipati (S77-M80-SM91-F95) received the B.Tech degree in Electrical Engineering with highest honors from the Indian Institute of Technology, Kharagpur, in 1975, and the MS and Ph.D. degrees in Systems Engineering from the University of Connecticut in 1977 and 1980, respectively. From 1980-86 he was employed by ALPHATECH, Inc., Burlington, MA. Since 1986, he has been with the University of Connecticut, where he is a Professor of Electrical and Computer Engineering. His current research interests are in the areas of adaptive organizations for dynamic and uncertain environments, multi-user detection in wireless communications, signal processing and diagnosis techniques for complex system monitoring, and multi-object tracking. Dr. Pattipati has published over 370 articles, primarily in the application of systems theory and optimization (continuous and discrete) techniques to large-scale systems. He has served as a consultant to Alphatech, Inc. and IBM Research and Development, and is a cofounder of Qualtech Systems, Inc., a small business specializing in intelligent diagnostic software tools.

Dr. Pattipati was selected by the IEEE Systems, Man, and Cybernetics Society as the Outstanding Young Engineer of 1984, and received the Centennial Key to the Future award. He was elected a Fellow of the IEEE in 1995 for his contributions to discrete-optimization algorithms for large-scale systems and team decision making. Dr. Pattipati has served as the Editor-in-Chief of the IEEE Transactions on SMC: Part B- Cybernetics during 1998-2001, Vice-President for Technical Activities of the IEEE SMC Society (1998-1999), and as Vice-President for Conferences and Meetings of the IEEE SMC Society (2000-2001). He was co-recipient of the Andrew P. Sage award for the Best SMC Transactions Paper for 1999, Barry Carlton award for the Best AES Transactions Paper for 2000, the 2002 and 2008 NASA Space Act Award for "A Comprehensive Toolset for Model-based Health Monitoring and Diagnosis", the 2003 AAUP Research Excellence Award and the 2005 School of Engineering Teaching Excellence Award at the University of Connecticut. He also won the best technical paper awards at the 1985, 1990, 1994, 2002, 2004 and 2005 IEEE AUTOTEST Conferences, and at the 1997 and 2004 Command and Control Conferences.