WIRELESS VIDEO CONFERENCING USING MULTIPLE DESCRIPTION CODING

Anshul Sehgal, Ashish Jagmohan, Narendra Ahuja

Beckman Institute, University of Illinois, Urbana - 60801
{asehgal, jagmohan, n-ahuja}@uiuc.edu

ABSTRACT

The transmission of video over wireless is a particularly involved problem - requiring as it does the transmission of a variable length highly compressed, delay-sensitive stream over an error prone medium. In this paper, this problem is addressed within the multiple description coding paradigm.

In particular the problem of wireless video conferencing over a constant bit-rate channel is addressed. A low-complexity, low-delay and adaptive coder that uses multiple description coding to combat channel errors is proposed. Error propagation is curtailed by using a leaky prediction mechanism. A low-complexity rate control scheme based on the LMS algorithm is also proposed.

1. INTRODUCTION

Wireless channels are inherently unreliable with high bit error rates (BERs) occurring during periods of deep fading - these errors can be modeled as erasure [1]. Therefore, any wireless transmission scheme requires the addition of a controlled amount of redundancy to combat the effect of these erasures. However, this sacrifices coding efficiency. Thus a mechanism is required to adapt the amount of redundancy added to the channel condition.

Multiple description coding (MDC) is a joint source-channel coding technique in which the source is encoded into multiple descriptions, such that there is redundancy among the descriptions. These descriptions are transmitted over separate channels (separated in time/frequency). In the wireless scenario, since the fading experienced over the channels is independent (and thus the failure probabilities of the channels are independent), the degradation in quality is graceful with increase in the subset of lost coefficients.

The core issue in designing MD coders is the trade-off between coding efficiency and redundancy between the descriptions, such that the degradation in quality in the event of failures is graceful. In this paper, we design an MD coder for a wireless video conferencing system. Thus, in addition to the above constraints, we desire our scheme to have low-complexity and low end-to-end delay.

Over the last few years, MD transmission of images has received significant attention. Details of previous work on MD transmission of images can be found in [2],[3],[4]. These schemes work well for images, however they cannot be directly employed for transmission of motion compensated video due to the problem of error propagation. The problem of MD video transmission is relatively new with little past work. Details of work on the MD transmission of video can be found in [5],[6], [7]. We present a scheme that adapts the redundancy incorporated into the transmitted video data to the wireless channel conditions. The problem of error propagation is combatted using a leaky prediction mechanism. The proposed scheme uses the correlating transform [8] to incorporate redundancy between the channels, and transmission over two channels. Extensions to a larger number of channels can be made.

2. THE PROPOSED ALGORITHM

Orchard et. al [4] proposed the correlating transform, which was later analyzed in detail and put in a rate-distortion framework by Goyal [9]. The parameter θ of the correlating transform controls the amount of *spatial redundancy* between the two descriptions. Due to space constraints we do not present any details of the correlating transform (these can be found in [4]). The correlating transform can be used for robust transmission of compressed images (assuming the DCT coefficients to be zero-mean, Gaussian, independent random variables). The transmission of compressed video, however, involves an additional difficulty. Video is differentially encoded, and hence errors tend to propagate. Consider the scenario where the video stream is transmitted along two wireless channels, separated in time/frequency. In particular, consider the case where video frames *1* through *i-1* have been reconstructed without errors, and frames *i, i+1, ..* are to be transmitted in the presence of channel errors. The residual obtained by differentially encoding frame *i* (with respect to frame *i-1*) is split between the two channels using the correlating transform. (The motion vectors for each frame are transmitted over both channels.) In this case, even if one of the descriptions of frame *i* is lost, an acceptable reconstruction of frame *i* is possible (because of the redundancy incorporated by the correlating transform). However, the errors in frame *i* will propagate to the future frames *i+1, i+2* .. since frames *i+1, i+2* .. will be encoded with respect to the original (undistorted) frame *i* . Over a duration of time, if further errors occur during the transmission of frames *i+1, i+2, ...* , the quality of the reconstructed video gets progressively worse until the next intra-coded frame. Thus an MD image coding scheme like the correlating transform cannot be directly used for the transmission of differentially encoded video.

An obvious solution to this problem is to intra-code each frame, thereby removing the inter-frame dependencies (i.e. treat video as a sequence of unrelated images) . However this would result in poor compression efficiency. An interesting way of viewing this problem is as one of finding the 'best' frame with respect to which the current frame can be differentially encoded. Thus at one extreme we have intra-coding, which can equivalently be looked upon as differential encoding with respect to a blank frame. This results in no propagation of error but also in poor compression efficiency. At the other extreme is differential encoding of the current frame with respect to the previous frame - which results in good compression efficiency but also in propagation of errors. The tradeoff involved, across the two extremes, is between good com-

0-7803-6685-9/01/$10.00©2001 IEEE

pression efficiency and low propagation of error. To maximize the PSNR of the reconstructed stream for a given compression efficiency, it may be best to encode the current frame with respect to a frame which is 'intermediate' between the above two extremes (i.e. a blank frame and the previous frame). In this paper, we use leaky prediction to generate such an intermediate frame.

Consider two consecutive uncompressed video frames F_1, F_2. In a standard differentially encoded stream, we differentially encode frame F_2 with respected to motion compensated and quantized F_1. In the proposed scheme, we differentially encode F_2 with respect to an intermediate frame F_1' that is derived from F_1. Next, we see how F_1' is obtained from F_1.

Frame F_2 is divided into 8×8 blocks, the DCT of the luminance and chrominance samples of each of these blocks is then taken. (for simplicity we consider only the luminance samples in the following discussion, chrominance samples are treated in exactly the same manner). Let $\mu^{\mathbf{F_2}} = \{\mu_1^{F_2}, \mu_2^{F_2}, ..., \mu_{64}^{F_2}\}$ be a vector of the means of the 64 DCT coefficients averaged over all the blocks of frame F_2. Similarly, let $\sigma^{\mathbf{F_2}} = \{\sigma_1^{F_2}, \sigma_2^{F_2}, ..., \sigma_{64}^{F_2}\}$ be the vector of the variances of the 64 DCT coefficients averaged over all the blocks of frame F_2 ($\mu^{\mathbf{F_1}}$ and $\sigma^{\mathbf{F_1}}$ denote the corresponding quantities for frame F_1). Consider any one of the 8×8 blocks from frame F_2. This block is motion compensated with respect to a quantized version of some block of frame F_1. [1] Let $\mathbf{A}^{\mathbf{F_1}} = \{\mathrm{A}_1^{F_1}, \mathrm{A}_2^{F_1}, ..., \mathrm{A}_{64}^{F_1}\}$ be the vector of the 64 DCT coefficients of this block. We consider $\{A_1^{F_1}, A_2^{F_1}, ..., A_{64}^{F_1}\}$ to be random variables independent of each other[2] with mean vector $\mu^{\mathbf{F_1}}$ and variance vector $\sigma^{\mathbf{F_1}}$. Let $\mathbf{a}^{\mathbf{F_1}} = \mathbf{A}^{\mathbf{F_1}} - \mu^{\mathbf{F_1}}$ denote corresponding zero mean independent random variables. The motion compensated block of F_2 corresponding to $\mathbf{a}^{\mathbf{F_1}}$ is differentially encoded with respect to $\mathbf{a}'^{\mathbf{F_1}}$ (derived from $\mathbf{a}^{\mathbf{F_1}}$). Next, we proceed to derive $\mathbf{a}'^{\mathbf{F_1}}$ from $\mathbf{a}^{\mathbf{F_1}}$. Consider the case of $a_i^{F_1}$, $1 < i < 64$. Let

$$E\{(a_i^{F_1} - a_i'^{F_1})^2\} = \alpha_i^2 \tag{1}$$

where α_i is the redundancy parameter for DCT coefficient i in the block $\mathbf{a}^{\mathbf{F_1}}$. α_i equals zero corresponds to ordinary differential encoding (high coding efficiency, high error propagation) and α_i equals $\sigma_i^{F_1}$ corresponds to intra-coding of F_2 (low coding efficiency, no error propagation). As described earlier, we use leaky prediction to mitigate the effects of error propagation. Thus, $a_i'^{F_1}$ is a scaled version of $a_i^{F_1}$. i.e.

$$a_i'^{F_1} = k_i a_i^{F_1} \tag{2}$$

where $k_i \geq 0$ is the leakage factor. Comparing equations 1 and 2,

$$k_i = \left(1 - \frac{\alpha_i}{\sigma_{a_i}}\right) \tag{3}$$

In a sense, the parameter α_i is a measure of the amount of information of F_1 that is contained in F_1'. Since frame F_2 is differentially encoded with respect to F_1' the information of F_1 that is not contained in F_1' is effectively retransmitted in the residual of frame F_2. Therefore, the vector $\vec{\alpha} = \{\alpha_1, \alpha_2, ..., \alpha_{64}\}$ can be used to control the inter-frame (temporal) redundancy and the propagation of errors. When $\vec{\alpha}$ is close to $\sigma^{\mathbf{F_1}}$, the amount of temporal redundancy is very large, since most of the information is retransmitted. Hence the propagation of error is curtailed. For a smaller value of $\vec{\alpha}$ the amount of retransmission is lower, hence the rate of recovery from an error is smaller.

[1] We note here that we are assuming that the blocks of frame F_1 are ergodic in space. Therefore the statistics (mean, variance) of this block is assumed to be the same as that of any of the encoded blocks of F_1

[2] This is a valid assumption since the DCT is a close approximation to the KLT for natural images.

3. SYSTEM DESCRIPTION

Consider a sequence of uncompressed frames $F_1, F_2, ..., F_n$ that need to be encoded and transmitted. Frame F_1 is encoded in the intra-mode (as in MPEG) using a quantization scale [3] parameter Δ^1 (Δ^1 denotes the quantization parameter for frame 1). Next, the correlating transform is applied to a pair of DCT coefficients in each DCT block [4] of F_1 to generate two descriptions of F_1. The parameter θ (spatial redundancy) of the correlating transform controls the redundancy between the two descriptions, let θ^1 denote the choice of the θ for frame F_1. Next, frame F_1' is generated with respect to which frame F_2 is encoded, let $\vec{\alpha}^1$ be the choice of the temporal redundancy parameter for F_1 for differentially encoding F_2. The residual obtained by motion compensating F_2 with respect to F_1' is then intra-coded in the same manner as F_1, let Δ^2 and θ^2 be the choice of the quantization scale and the spatial redundancy parameter for the residual of frame F_2. Let α^2 denote the temporal redundancy parameter for F_2 while encoding F_3. Hence, the choice of the vectors $\theta = \{\theta^1, \theta^2, ..., \theta^n\}$, $\mathbf{\Delta} = \{\Delta^1, \Delta^2, ..., \Delta^n\}$, $\alpha = \{\tilde{\alpha}^1, \tilde{\alpha}^2, ..., \tilde{\alpha}^n\}$ governs the trade-off between compression efficiency and robustness to error and error propagation.

Next, we see how the parameters Δ^i, $\tilde{\alpha^i}$ and θ^i (the quantization and the spatial and temporal redundancy parameters for frame i) need to be adjusted so as to facilitate real-time good quality reconstruction of the video sequence after transmitting it over a constant bandwidth, error prone wireless channel. The aim is to maximize the quality of the received video (in the presence of channel errors) while staying within the bandwidth constraints imposed by the channel. When the channel is unreliable, θ and α are used to increase the amount of intra-frame and inter-frame redundancy. To stay within the bandwidth constraints, the quantization step size has to be reduced. When the channel is reliable, we would like a fine scale parameter Δ so that the reconstruction is of high quality and correspondingly adjust θ and α such that the amount of redundancy is low. This choice can be made at the encoder on the basis of feedback received from the decoder or the channel.

Another concern is the packetization of the data in the two streams, assuming that the two channels available have equal bandwidth, it would be advisable to generate the two streams such that they have equal rates. This is accomplished by splitting the *correlated coefficients* between the two streams in the manner described next. For each DCT block of frame F_i, the DCT coefficients are scanned into two arrays along the two curves shown in Figure 3. Let the DCT coefficients scanned along curve 1 be denoted by $\vec{a} = \{a_1, a_2, ..., a_{32}\}$ and those scanned along curve 2 by $\vec{b} = \{b_1, b_2, ..., b_{32}\}$. Taking the correlating transform of the pairs of coefficients, we get $[c_1, d_1]^T = \mathbf{T}[\mathrm{a}_1, \mathrm{b}_1]^{\mathrm{T}}$, $[c_2, d_2]^T = \mathbf{T}[\mathrm{a}_2, \mathrm{b}_2]^{\mathrm{T}}$,...,$[c_{32}, d_{32}]^T = \mathbf{T}[\mathrm{a}_{32}, \mathrm{b}_{32}]^{\mathrm{T}}$, where $\mathbf{T}$ is the correlating transform matrix. The coefficients $\{c_1, d_1, ..., c_{32}, d_{32}\}$ are then split into two sets as follows: $Description\ 1 = \{c_1, d_2$

[3] The coefficients are quantized after applying the correlating transform.

[4] We use the pairing suggested in [4] since this provides the maximum variation in the amount of spatial redundancy added

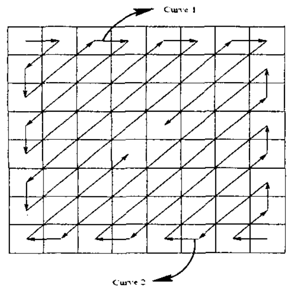

Fig. 1. Scanning of the DCT coefficients into two vectors along curve 1 and curve 2.

$, c_3, d_4, ..., d_{32}\}$ and $Description\ 2 = \{d_1, c_2, d_3, c_4, ..., c_{32}\}$. Assuming that the variances of the DCT coefficients decay smoothly with frequency, the above splitting ensures that the bit rate of the two descriptions is approximately the same. A simpler alternative would be to transmit the set $\vec{c}$ on one channel and $\vec{d}$ on the other for a block and interchange the descriptions for the following block. However, for small values of θ, the vector $\vec{c}$ would have much more energy as compared to $\vec{d}$, hence in case the description with $\vec{c}$ is lost, the reconstruction of this block would be poor. On the other hand, in the proposed scheme, the degradation in quality would be more graceful.

4. ADAPTIVE PARAMETER SELECTION

As mentioned earlier, any practical MD coding scheme for video has to be channel adaptive to efficiently utilize the bandwidth. For this purpose, it is assumed that the decoder is able to provide ACK/NACK feedback to the encoder via a reliable feedback channel. The channel being modeled is a constant bit-rate wireless channel (like CBR in WATM). It is further assumed that errors occur in bursts and typically last up to 200 ms (2 video frames at 10 fps). The feedback received at the encoder is inevitably delayed leading to inaccuracies in channel state estimation. The smaller the delay, the better is the channel state estimation.[5]

Standard techniques for optimally deciding the encoder parameters based on the channel state make use of dynamic programming techniques or the method of Lagrange multipliers [1]. These operations often require large computation which is not suitable for low-complexity systems. Motivated by this concern, a low-complexity ad-hoc parameter adaptation scheme is proposed that controls the spatial (θ) and the temporal (α) redundancy added.

The redundancy parameters are adapted to the channel state depending on the feedback from the decoder as briefly explained next. The transmitted streams have a nominal value of $\theta = \theta_0$. This compensates for the delay in feedback - i.e. it combats errors which occur in frames before the encoder receives the decoder error feedback and can incorporate further redundancy. Upon receiving NACKs from the decoder, the encoder increases the value of θ gradually, in increments of θ_i (so as to avoid addition of excessive redundancy). When an ACK is received from the decoder, the value of θ is immediately reduced back to $\theta = \theta_0$.

The interframe redundancy parameter α combats the effect of error propagation. The nominal value of α is zero. When a NACK is received the encoder directly increases the value of α to α_{shoot} after a delay corresponding to the difference between the average error burst length and the delay in feedback (assumed to be lesser than the burst length). Once ACKs begin to arrive, the value of α is decreased gradually (in decrements of α_d) - the gradual decrease is because error propagation manifests itself over a relatively long-term period.

Deciding the spatial and temporal redundancy parameters in the above manner combats errors, however, since these redundancy parameters are dynamically adjusted depending upon the feedback from the decoder, the rate of transmission over the channel also fluctuates with the redundancy added. This is not acceptable since the channels are constant bandwidth channels. Hence, to maintain constant rate of transmission, a mechanisim which adjusts the quantization parameter with dynamically changing redundancy is required. The choice of such a quantization parameter is explained next.

First, we model the rate of transmission. Assuming the DCT coefficients are Gaussian distributed, the rate of transmission over the two channels for a given value of θ, α can be calculated. Denoting the variance of the residual for coefficient $a_{i,j}$ of frame j by $\sigma_{ea_i,j}$ and the total number of DCT blocks in each frame by M, the rate of transmission over channel 1 is given by:

$$R_{Ch_1} = M \times \log 2\pi e \prod_{i=1,i'=64}^{i=32,i'=33} \sigma_{c(i,i'),j} \tag{4}$$

where

$$\sigma_{c(i,i'),j} = \cos^2(\theta)k^2(\alpha_i, \sigma_{a_i}, \sigma_{ea_i,j}) + \sin^2(\theta)k^2(\alpha_{i'}, \sigma_{a_{i'}}, \sigma_{ea_{i'},j}) - \frac{\Delta_{c(i,i')}}{12} \tag{5}$$

where

$$k^2(\alpha_i, \sigma_{a_i}, \sigma_{ea_i,j}) = \sigma^2_{ea_i,j} + \alpha_i^2 + \frac{\alpha_i \sigma_{ea_i,j}}{\sigma_{a_i}} \tag{6}$$

However, the rate for compressed video data deviates from this. While the actual rate can be computed by compressing the data for multiple values of the parameters, this would be impractical for a low-complexity system. The need for accurate estimation of the rate can be seen as follows. Since the channel is constant bit-rate, in the presence of errors, the quantization has to become coarser to compensate for the increased redundancy. To avoid over-compensation (leading to bandwidth wastage) or under-compensation (leading to exceeding the allocated bandwidth), it is imperative to be able to compute the quantization step size required without having to actually encode the stream for a range of values of Δ. Therefore, the value of Δ is computed using the LMS algorithm, as follows[6] :

$$\Delta_{F_i} = \Delta_{F_{i-1}} + \lambda_1(R_q^{F_i} - R_q^{F_{i-1}})\frac{\partial \Delta}{\partial R_q}|_P + \lambda_2(\theta^{F_i} - \theta^{F_{i-1}})\frac{\partial \Delta}{\partial \theta}|_P + \lambda_3(\vec{\alpha}^{F_i} - \vec{\alpha}^{F_{i-1}})^T \nabla_{\vec{\alpha}} \Delta \tag{7}$$

where the set $P = \{\theta^{F_{i-1}}, \vec{\alpha}^{F_{i-1}}, \Delta^{F_{i-1}}\}$, $\{\lambda_1, \lambda_2, \lambda_3\}$ are the acceleration factors and $R_q^{F_i}$ is set to R_o, the channel rate. The partial derivatives are not known for the actual compressed data, hence equation (4) is used to compute the them.

[5] This model has been used before ([1]).

[6] For ease of implementation, we have assumed all the components of $\vec{\alpha}$ to be equal.

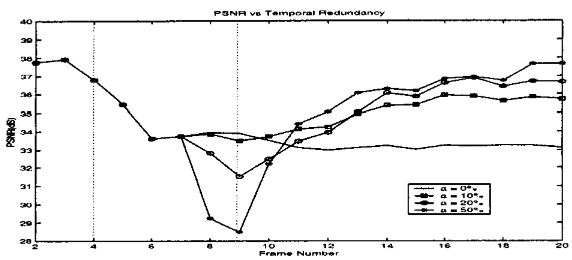

Fig. 2. Variation of PSNR with changing temporal redundancy ($\alpha_{max} = a, \alpha_d = a/6$)

5. RESULTS

Preliminary results are presented here, the final paper will contain more extensive simulation results.

Results are shown for the 'Miss America' stream with frame size 176×144 pixels. The stream was transmitted over two simulated constant bit-rate, 60 kbps channels at 10 fps. The correlated DCT coefficients were split between the two streams so as to have approximately equal rate streams. The streams were encoded using the MPEG quantization matrices and the set of scale parameters. The streams were then run-length and variable length coded. The motion vectors were transmitted over both channels.

Figure 2 shows the PSNR values with varying temporal redundancy and zero spatial redundancy in the presence of a channel burst erasure (between frames 4 and 9) for a constant bit-rate channel. The maximum value of $\alpha = \alpha_{shoot}$ is set to $a\%$ of the coefficient variance for each coefficient, α_d is set to $a/6$, θ_{avg} and θ_o are both set to zero. As can be seen from the Figure, for large values of α_{shoot}, the PSNR drops more during the error period, however the recovery from errors is also fast.

Figure 3 shows the output PSNR values for a working system with both temporal and spatial redundancy adaptively varied. θ_{avg} is set to $\pi/90$, $\theta_i = \pi/30$, $\alpha_{shoot} = 20\%$ and $\alpha_d = \alpha_{max}/5$. As can be seen, the incorporation of temporal redundancy facilitates speedy recovery of PSNR, from channel errors. Also, the LMS algorithm keeps the rate of transmission almost constant while the spatial and temporal redundnacy is being changed.

6. CONCLUSIONS

An algorithm for transmission of video over erasure channels is presented in this paper. A simple and effective rate control scheme based on the LMS algorithm is also presented. The leaky prediction mechanism helps recover from the channel errors fast, while spatial redundancy helps in controlling the amount of damage caused to the stream during periods of error. Future work includes optimizing the redundancy parameters based on source stream R-D characteristics.

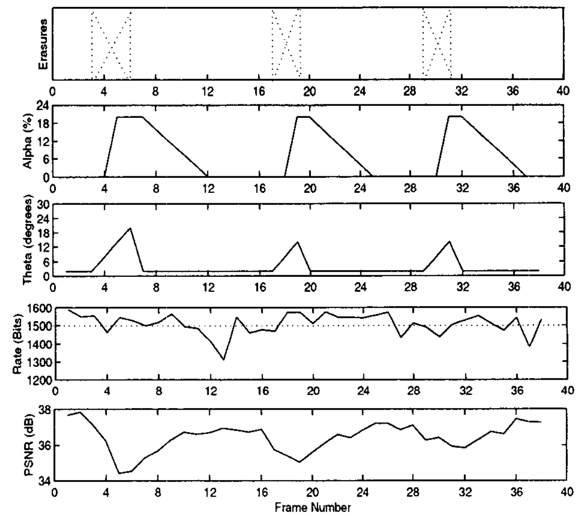

Fig. 3. Full system implementation with spatial and temporal redundancy

7. REFERENCES

[1] C. Hsu, A. Ortega and M. Khansari. Rate Control for Robust Video Transmission over Burst-Error Wireless Channels.*IEEE Journal on Selected Areas in Communications, Special Issue on Multimedia Network Radios*, Vol. 17, No. 5, pages 756-773, May 1999.

[2] S. Yang, V. Vaishampayan. Low-Delay Communication for Rayleigh Fading Channels: An Application of the Multiple Description Quantizer. *IEEE Transactions on Communications*, pages 2771-2783, vol. 43, no. 11, 1995.

[3] V. A. Vaishampayan. Design of multiple description scalar quantizers. *IEEE Trans. Inform. Theory*, vol. 39, pages 821-834, May 1993.

[4] M. Orchard, Y. Wang, V. Vaishampayan, A. Reibman. Redundancy Rate-Distortion Analysis of Multiple Description Coding Using Pair-wise Correlating Transform. *IEEE Signal Processing Society 1997 Workshop on Multimedia Signal Processing*, pages 821-834.

[5] A. C. Miguel, A. E. Mohr, and E. A. Riskin. SPIHT for Generalized Multiple Description Coding. *Proc. of the International Conference on Image Processing*, 1999.

[6] R. Puri, K. Ramchandran. Multiple Description Source Coding Through Forward Error Correction Codes.*Proc. of the 33rd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA*, October 1999.

[7] V. Vaishampayan. Application of multiple description codes to image and video transmission over lossy networks.*Proc. 7th International Workshop on Packet Video, Brisbane, Australia*. March 1996.

[8] Yao Wang, Orchard M. T., Reibman A. R. Multiple Description Image Coding for Noisy Channels by Pairing Transform Coefficients. *Proc. IEEE Worskshop on Multimedia Signal Processing*, pages 419-424, 1997.

[9] V. Goyal, J. Kovacevic. Optimal Multiple Description Transform Coding of Gaussian Vectors. *Proc. IEEE Data Compression Conference*, pages 388-397, 1998.

[10] L. Lin, A. Ortega. Bit-Rate Control Using Piecewise Approximated Rate-Distortion Characteristics. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 8, No. 4, pages 446-459 , Aug. 1998.