Image Data-hiding Based on Capacity-approaching Dirty-paper Coding
Yang Yang, Yong Sun, Vladimir Stanković, and Zixiang Xiong
Dept. of Electrical and Computer Eng., Texas A&M University, College Station, TX 77843

ABSTRACT
We present an image data-hiding scheme based on near-capacity dirty-paper codes. The scheme achieves high embedding rates by "hiding" information into mid-frequency DCT coefficients among each DCT block of the host image. To reduce the perceptual distortion due to data-hiding, the mid-frequency DCT coefficients are first perceptually scaled according to Watson's model. Then a rate-1/3 projection matrix in conjunction with a rate-1/5 capacity-approaching dirty-paper code is applied. We are able to embed 1500 information bits into $256 \times 256$ images, outperforming, under a Gaussian noise attack, currently the best known data-hiding scheme by 33%. Robustness tests against different attacks, such as low-pass filtering, image scaling, and lossy compression, show that our scheme is a good candidate for high-rate image data-hiding applications.

Keywords: Data-hiding, watermarking, dirty-paper coding, perceptual shaping, and robustness.

1. INTRODUCTION
Digital data-hiding, or more specifically digital watermarking, is a technique of unnoticeable altering a digital media (such as images, audio, or video sequences) to embed information. It has applications in broadcast monitoring, owner identification, transaction tracking, and copy control. In most of these applications, digital watermarking techniques must possess three important properties: imperceptibility, robustness, and security. The design challenge is to develop a scheme that can embed as many information bits as possible while respecting these properties.

Since usually the receiver has no access to the original data, we consider an informed watermarking scheme that incorporates informed embedding (i.e., the encoder knows the original data) and blind detection. It was recognized recently that informed watermarking can be modelled as communication with side information at the encoder, where the encoder has some non-causal knowledge about the transmission channel (e.g., channel interference) in the form of side information, which is not available at the decoder. When the channel is additive white Gaussian noise (AWGN) we have the celebrated dirty-paper coding (DPC) problem, in which the decoder can completely cancel out the effect of the interference caused by the side information.

Chen and Wornell first pointed out the equivalence between DPC and digital watermarking. Indeed, DPC can be viewed as a special case of informed watermarking with a Gaussian host signal, an AWGN attack, and mean-square error (MSE) distortion measure. Note that the capacity of a DPC channel is the same as that of a pure AWGN channel, and hence equals to the watermarking capacity when the original data is known to the decoder.

Recently, it was shown that the DPC capacity can be asymptotically achieved using nested lattice codes. Erez et al. constructed a scheme that combines dithered lattice quantization and minimum mean-square error (MMSE) scaling to transform the power-constrained AWGN channel into a so-called modulo-lattice additive noise channel; the scheme achieves the DPC capacity as the lattice dimension goes to infinity. Following this approach, Erez and ten Brink proposed close-to-capacity practical dirty paper codes, which employ vector quantization (VQ) and iterative decoding of capacity-approaching channel codes. The best performance reported with a memory-6 VQ is within 1.3 dB of the DPC capacity at an embedding rate of 0.25 bit/sample. Moreover, another 0.5 dB gain at the same rate is achieved by a DPC scheme introduced using memory-10 trellis coded quantization (TCQ) and systematic irregular repeat accumulate (IRA) codes, with lower complexity compared to the scheme of.

Based on recent successful dirty-paper code designs, practical schemes for real-world applications have been developed. Miller et al. employed trellis-based dirty-paper codes for image watermarking. At an embedding
rate of 0.0156 bit/pixel, the scheme of\(^6\) shows robustness against important valumetric attacks including AWGN, valumetric scaling, low-pass filtering, and lossy compression. Recently, Comesaña \textit{et al.}\(^3\) achieved higher embedding rates by replacing the trellis-based dirty-paper code by the DPC design of.\(^8\) The scheme of\(^6\) embeds 1122 information bits into the 256 \(\times\) 256 Lena image (corresponding to an embedding rate of 0.017 bit/pixel), by applying DPC on the perceptually shaped mid-frequency Discrete Cosine Transform (DCT) coefficients as encoder side information. The robustness of this scheme against the AWGN attack (in the perceptually shaped DCT domain) was evaluated, showing a bit-error rate (BER) of \(10^{-3}\) at a watermark-to-noise ratio (WNR) of -7.9 dB, which is 3.64 dB away from the DPC capacity. However, the scheme\(^3\) exploits only low-memory VQ (memory-2) with the IRA code profile used in,\(^8\) which was not specifically designed for this application. Hence there are still room for improvements.

In this paper, we present a high-rate image data-hiding scheme by applying the best known dirty-paper code designs of.\(^16,17\) Similar to,\(^3\) our scheme operates in the DCT domain of the host image and modifies only 22 mid-frequency DCT coefficients among each 8 \(\times\) 8 DCT block. To reduce the perceptual distortion caused by the watermark signal, the extracted DCT coefficients are first perceptually scaled according to Watson’s model.\(^20\) The resulting scaled coefficients are the actual data-hiding host and can be viewed as independent samples with the same perceptual sensitivities. Note that, since the DPC capacity does not depend on the distribution of the host signal, our scheme is expected to perform equally well for all real images.

Information bits are embedded into the scaled DCT coefficients at a rate of one bit per 15 samples. However, to accomplish such a small embedding rate, time-sharing is necessary for a good performance.\(^8\) Hence, to implement time-sharing and achieve the desired embedding rate, we employ a three-to-one projection function and a rate-1/5 near-capacity dirty-paper code of.\(^16\) The rate-1/5 DPC scheme in\(^16\) serially concatenates a rate-1/5 IRA code designed using EXIT-chart-based techniques,\(^19\) and a 256-state TCQ with 16-PAM constellation. At the block length of 2 \(\times\) \(10^5\) bits, this IRA code introduces 0.34 dB loss to the capacity on AWGN channels, while TCQ provides 1.33 dB granular gain. Hence the overall performance is 1.07 dB away from the dirty-paper limit.\(^4\) However, because of the limited size of practical images, in image data-hiding the block length has to be reduced, which somewhat worsens performance.

We simulate the proposed data-hiding scheme by modelling the attack as a Gaussian channel with independent identically distributed (i.i.d.) noise components in the perceptually shaped DCT domain. In our experiments, 1500 information bits are embedded into the 256 \(\times\) 256 Lena image, which is equivalent to an embedding rate of 0.0229 bit/pixel. The turbo cliff of this scheme appears at the signal-to-noise ratio (SNR) of -3.13 dB (with a BER less than 10\(^{-3}\)), which is 1.83 dB away from the DPC capacity (thus, the loss due to the reduced block length is 1.83-1.07=0.76 dB). To compare our scheme with the data-hiding scheme of,\(^3\) we convert our results to a function of the WNR; then the turbo cliff shows up at WNR=7.9 dB, which is the same as in.\(^3\) However, we are able to embed 378 more bits (or 33 \% more) than the scheme in\(^3\) under the identical simulation conditions.

We also perform robustness tests on our scheme against different valumetric attacks.\(^6\) The results show that our scheme has comparable robustness to the scheme in,\(^6\) while embedding 46\% more information bits. Hence our scheme is a good candidate for high-rate image data-hiding applications.

The rest of this paper is organized as follows. Section 2 provides a general framework of DPC-based image data-hiding. Section 3 reviews the best known dirty-paper code designs.\(^16,17\) The proposed image watermarking scheme is then detailed in Section 4. Simulation results against the AWGN attack are shown in Section 5. Finally, the performance of our scheme in other robustness tests is given in Section 6.

### 2. IMAGE DATA-HIDING BASED ON DIRTY-PAPER CODING

Throughout the paper, random variables are denoted by capital letters, e.g., \(X\). An \(n\)-length vector of samples drawn from a random variable \(X\) is denoted by \(X^n\), and the \(i\)-th sample of a random vector \(X^n\) is \(x_i\).

The block diagram of our image data-hiding scheme based on DPC is shown in Figure 1. A message \(B\) is to be embedded into the original image host \(X_i\). Since the goal is to provide a scheme that is robust to various types of attacks (including, for example, low-pass filtering and compression), it is more efficient to embed information into transformed coefficients rather than the image itself. Therefore, \(X_i\) is first transformed to \(X_i = \mathcal{F}(X_i)\) using certain image transformation \(\mathcal{F}\) (e.g., DCT). Prior to information embedding, all the transformed coefficients are
perceptually shaped (i.e., scaled) such that they all bear the same perceptual sensitivities. This step is necessary since introducing unequal perceptual distortion between coefficients will lead to more visual quality degradation compared to the case when all coefficients are equally distorted in the perceptual sense.\textsuperscript{14} The resulting scaled coefficients $X_s$ are then randomly permuted with a secret key that is known to both the encoder and the decoder. This permutation operation not only increases the security of the watermark, but also breaks the dependence between consecutive coefficients. Instead of embedding information in all transformed coefficients, we modify only some of them (the mid-frequency coefficients in each $8 \times 8$ DCT block) to ensure the robustness against certain attacks (e.g., the low-pass filtering attack). We denote the permuted and “extracted” coefficients (the mid-frequency once in our case) as $X_h$, which is the actual data-hiding host and can be viewed as the independent samples with the same perceptual sensitivities.

![Block diagram of the image data-hiding scheme based on dirty-paper coding.](image)

Dirty-paper encoder should embed a binary message $B$ into the data-hiding host $X_h$ at a target rate of $R_{WM}$ bit/sample. The DPC schemes of\textsuperscript{8,16} operate at the lowest embedding rates of 0.2-0.25 bit/sample. Since the practical image data-hiding rates under the constraint of unnoticeable embedding are much lower (e.g., 0.0156 bit/pixel in\textsuperscript{8}), we have to provide an interface when applying dirty-paper codes of\textsuperscript{8,16}. An intuitive solution for such an interface proposed in\textsuperscript{8} is to:

- linearly combine each $k$ samples of $X_h$ into one, where $k = R_{DPC}/R_{WM}$, resulting in $X_p$,
- apply dirty-paper code of rate $R_{DPC}$ on $X_p$, and
- after embedding, equally distribute the watermarking power to each of $k$ samples.

This three-step procedure, referred to as “projection”, “dirty-paper encoding”, and “inverse-projection” in Figure 1, is the key structure of this image data-hiding framework that bridges the gap between existing dirty-paper schemes and this specific application. Finally, the watermarked signal $Y_h$ sequentially passes through the inverse operations, yielding $Y_s$, $Y_t$, and $Y_i$, which correspond to $X_s$, $X_t$, and $X_i$, respectively. The watermarking power $P_w$ is defined as the MSE between the DCT coefficients of the watermarked image $Y_i$ and the original one $X_i$, i.e., $P_w = \mathbb{E}[(Y_i - X_i)^2]$, where $\mathbb{E}[]$ is the expectation over all DCT coefficients.

In data-hiding applications, the watermarked image $Y_i$ has to be robust to different types of alternations, called attacks. Common attacks include additive Gaussian noise attack, low-pass filtering attack, lossy compression...
attack, valumetric scaling attack, and rotation attack. An efficient data-hiding scheme is capable of preserving the visual quality of the image after these attacks. Quantitatively, we can evaluate the MSE distortion between the altered transform coefficients \( Z_i \) (after the attack) and the watermarked coefficients \( Y_i \) as \( P_n = \mathcal{E}[(Z_i - Y_i)^2] \). Then the constraint \( P_n \leq P \) for some \( P > 0 \) limits the perceptual degradation of the image at the receiver. Note that other distortion measures (e.g., Watson’s distortion) can be used instead of MSE, but there are no closed-form expressions for the data-hiding capacity under general measures and attacks. One exception is the case with the AWGN attack and MSE distortion measure, which is equivalent to the DPC problem.

Under the constraint \( P_n \leq P \), the watermark decoder needs to make the decision on each embedded information bit based on the received image \( Z_i \). The decoder preprocesses \( Z_i \) following the preprocessing steps for \( X_i \) at the encoder, resulting in \( Z_k \), the altered version of \( Y_k \). After projection, \( Z_p \) is fed to the DPC decoder which outputs the reconstructed information message \( \hat{B} \). The bit error probability between \( B \) and \( \hat{B} \) gauges the overall performance of an image data-hiding scheme as a function of the WNR \( P_w/P_n \).

3. NEAR-CAPACITY DIRTY-PAPER CODING

In recent years several close-to-capacity dirty-paper code designs have been proposed. Erez and ten Brink designed dirty-paper codes based on lattice precoding and trellis shaping. This scheme with non-systematic IRA codes and 64-state trellis based VQ performs 1.3 dB away from the DPC capacity at embedding rate of 0.25 bit/sample. Bennatan et al. developed a code design based on superposition coding and successive cancellation decoding; it employs 512-state TCQ and low-density parity-check (LDPC) codes, and is also 1.3 dB away from the DPC capacity.

The best performed DPC code design is proposed based on TCQ and IRA codes. The block diagram of this scheme is shown in Figure 2, where, in analogy to Figure 1, \( B \) denotes the transmitted message and \( X_p \) is the encoder side information. The key structure of this scheme is the combination of trellis-based lattice coding, lattice decoding, and MMSE scaling; hence, it is a practical realization of the lattice precoding approach for DPC proposed in.

![Figure 2. Near-capacity dirty-paper code design of.](image)

3.1. Lattice Precoding

Let \( \Lambda \) be an \( n \)-dimensional lattice quantizer with basic Voronoi cell \( \mathcal{V} \). Associated with \( \mathcal{V} \) are several important quantities: the cell volume \( |\mathcal{V}| \), the second moment \( P(\Lambda) \), and the normalized second moment \( G(\Lambda) \), defined by \( |\mathcal{V}| = \int_\mathcal{V} dx, P(\Lambda) = \frac{1}{n|\mathcal{V}|} \int_\mathcal{V} |x|^2 dx, \) and \( G(\Lambda) = P(\Lambda)/|\mathcal{V}|^{\frac{3}{2}} \), respectively. The minimum of \( G(\Lambda) \) over all \( n \)-dimensional lattices is denoted as \( G_n \), where \( G_n > \frac{1}{27\pi^2} \), \( \forall n \) and \( \lim_{n\to\infty} G_n = \frac{1}{27\pi^2} \). The granular gain of \( \Lambda \) is \( g(\Lambda) = -10\log_{10} 12G(\Lambda) \), which is maximally 1.53 dB. Let \( D \) be a random dither uniformly distributed over \( \mathcal{V} \). For any source codewords (or constellation points) \( v \in \mathcal{V} \), the encoder transmits \( U = [v - \alpha X_p - D] \mod \Lambda \), while the decoder receives \( Z_p = U + X_p + N \) and computes \( Z_p' = [\alpha Z_p + D] \mod \Lambda = [v + N'] \mod \Lambda \), where \( \alpha \) is a scaling factor and \( N' = [(1 - \alpha)D + \alpha N] \mod \Lambda \) is the equivalent modulo lattice channel noise.

The maximum achievable rate of the modulo lattice channel is \( \frac{1}{2}I(V; Z_0') \), achieved by a uniformly distributed input \( V \) over \( \mathcal{V} \). Due to the dither \( D \), \( U \) is independent of \( v \) and uniformly distributed over \( \mathcal{V} \) with \( E[||U||^2] = \)
\( P_U = P(\Lambda) \). Then, for \( n > 1 \), 
\[
\frac{1}{n} I(V; Z'_p) \text{ can be lower bounded by assuming } D \text{ has i.i.d. Gaussian components}
\]
and using the MSE-optimal \( \alpha = \frac{P_U}{P_U+P_N} = \frac{\text{SNR}}{\text{SNR}+1} \), yielding
\[
\frac{1}{n} I(V; Z'_p) \geq \frac{1}{2} \log_2 (1 + \text{SNR}) - \frac{1}{2} \log_2 2\pi e G(\Lambda).
\]

(1)

Note that for any finite \( n \), the components of \( D \) (and \( U \)) are not Gaussian or independent. In practice, \( n \) has to be high for the i.i.d. Gaussian assumption to be approximately true and for the lower bound in (1) to be tight.

### 3.2. Dirty-paper Coding Based on TCQ and Nonsystematic IRA Code

TCQ is efficient means of implementing an equivalent high-dimensional lattice quantizer \( \Lambda \). A dither \( D \) is generated by keeping random inputs that are only quantized to zero by TCQ. At the encoder, side information \( X_p \) is first linearly scaled by \( \alpha \), then \( \alpha X_p + D \) is quantized to \( v \) by a coset of the TCQ selected by the message \( B \) so that the obtained quantization error \( U = v - \alpha X_p - D \) satisfies the power constraint \( E[U^2] = P_U \). \( U \) is transmitted over the side-information channel. The decoder receives \( Z_p = U + X_p + N \) and finds the codeword \( \hat{v} \) closest to \( Z'_p = \alpha Z_p + D = v + N' \), where \( N' = (1 - \alpha)(-U) + \alpha N \) is the equivalent channel noise. Note that the “mod \( \Lambda \)” operation is implicitly implemented in TCQ. Finally, the index of the bin containing \( \hat{v} \) is identified as the decoded message.

The aim of the code design in\(^{16}\) is to use the strongest possible TCQ to achieve most of the 1.53 dB granular gain so that the distribution of \( U \) (and the dither \( D \)) approaches Gaussian, while employing IRA codes to approach the capacity. Then the practical performance loss \( \Delta \text{SNR} \) (in dB) of this design can be separated into two parts: packing loss \( \Delta \text{SNR}_p \) due to IRA codes and the modulo loss \( \Delta \text{SNR}_m \) due to TCQ, i.e., \( \Delta \text{SNR} = \Delta \text{SNR}_p + \Delta \text{SNR}_m \). Assuming that \( \Delta \text{SNR}_p = 0.34 \) dB, Table 1 lists the predicted total performance loss \( \Delta \text{SNR} \) when the target rate is 0.25 bit/sample. These predictions are confirmed by simulation results based on 100 blocks of transmission. For example, a BER of \( 4.76 \times 10^{-5} \) is obtained when \( \text{SNR}=2.993 \) dB, which is 0.83 dB away from the DPC capacity at the rate of 0.25 bit/sample. This performance matches the predicted 0.49+0.34=0.83 dB gap in Table 1 when the granular gain of TCQ is 1.38 dB.

**Table 1. Performance of the dirty-paper code design of.\(^{16}\)**

<table>
<thead>
<tr>
<th># of TCQ States</th>
<th>Granular Gain (dB)</th>
<th>Predicted ( \Delta \text{SNR} ) (dB)</th>
<th>Practical ( \Delta \text{SNR} ) (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>1.33</td>
<td>0.99</td>
<td>0.984</td>
</tr>
<tr>
<td>512</td>
<td>1.36</td>
<td>0.90</td>
<td>0.902</td>
</tr>
<tr>
<td>1024</td>
<td>1.38</td>
<td>0.83</td>
<td>0.835</td>
</tr>
</tbody>
</table>

### 4. PROPOSED IMAGE DATA-HIDING SCHEME

The near-capacity dirty-paper code design based on TCQ and IRA codes\(^{16}\) provides powerful tools for construction of a high-embedding-rate image watermarking scheme under an AWGN attack. One can simply incorporate the dirty-paper code designs\(^{16}\) into the general image data-hiding framework in Figure 1. However, several special properties of image data-hiding applications bound the performance of such a straightforward implementation away from the capacity. The most important limitation is the codeword length, which is often in the range between 1000 and 10000 bits due to the limited size of practical image sources (for comparison, the codeword length in\(^{16}\) is \( N = 240000 \) bits). This urges us to design dirty-paper codes that operate well for lower codeword lengths. In this section, we first describe each component of our image data-hiding scheme in detail, and then provide new dirty-paper code designs that outperform those in\(^{16}\) for the image data-hiding application.

#### 4.1. Dirty-paper Coding Based Image Data-hiding Scheme

Our goal is to embed 1500 information bits into the 256 × 256 gray-scale Lena image with imperceptible modifications, while minimizing the minimum tolerable WNR for almost error-free (with the BER less than \( 10^{-3} \)) information reconstruction. The corresponding image-wise information embedding rate is 0.0226 bit/pixel.
Our scheme operates in the DCT domain of the host image $X_i$. DCT coefficients $X_t$ are perceptually shaped according to Watson’s model, which is a widely used DCT-based visual model that tries to estimate “the just noticeable difference” for each DCT coefficient of an image. The model includes two main parts: luminance masking and contrast masking, but in our scheme, we consider only the former. In short, the luminance mask for each DCT coefficient is a sensitivity threshold as a function of the magnitude and the local brightness (DC coefficient) of the DCT block. This function is defined as:

$$t_{i,j,k}^L = t_{i,j}(C_{0,0,k}^0/C_{0,0})^{\alpha T},$$

where $t_{i,j,k}^L$ is the luminance-masked threshold for the $(i, j)$-th DCT coefficient of the $k$-th block. $C_{0,0,k}^0$ represents the DC coefficient of the $k$-th block, which is normalized by the average DC value $C_{0,0}$ of the image. The exponent $\alpha T$ is a constant equal to 0.649. The multiplier $t_{i,j}$ is the intrinsic frequency sensitivity for the $(i, j)$-th DCT coefficient, which is determined by the background luminance and horizontal and vertical spatial frequencies. An example of $t_{i,j}$ values is shown in Table 2. Using this perceptual model, all the DCT coefficients are scaled by the corresponding thresholds to generate $X_s$; this means $X_s = X_t/T^L$, where “/” denotes component-wise division, and $T^L$ is the vector form of $t_{i,j,k}^L$.

**Table 2.** DCT frequency sensitivity table.

<table>
<thead>
<tr>
<th>$i$ \ $j$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.5265</td>
<td>1.5265</td>
<td>1.1608</td>
<td>1.8157</td>
<td>2.4446</td>
<td>3.2950</td>
<td>4.4093</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.5265</td>
<td>1.7377</td>
<td>1.4706</td>
<td>1.6133</td>
<td>2.0162</td>
<td>2.6353</td>
<td>3.4865</td>
<td>4.6082</td>
</tr>
<tr>
<td>2</td>
<td>1.1608</td>
<td>1.4706</td>
<td>1.8952</td>
<td>2.1787</td>
<td>2.3832</td>
<td>3.2066</td>
<td>4.0643</td>
<td>5.2122</td>
</tr>
<tr>
<td>3</td>
<td>1.3816</td>
<td>1.6133</td>
<td>2.1787</td>
<td>2.8532</td>
<td>3.3788</td>
<td>4.0891</td>
<td>5.0172</td>
<td>6.2294</td>
</tr>
<tr>
<td>4</td>
<td>1.8157</td>
<td>2.0162</td>
<td>2.3832</td>
<td>3.3788</td>
<td>4.2512</td>
<td>5.1923</td>
<td>6.2913</td>
<td>7.6440</td>
</tr>
</tbody>
</table>

Note that there are 1024 $8 \times 8$ DCT blocks in a $256 \times 256$ image. From each such block, we extract only 22 mid-frequency AC coefficients (see Figure 3), which are going to be altered (i.e., used as a host signal) later. Hence the extracted source consists of $22 \times 1024 = 22528$ DCT coefficients in total. To ensure security of the embedded information, it is pseudo-randomly permuted with a seed (private key) known to both the encoder and decoder. In this way the actual host signal $X_h$ is formed. Besides providing security, the benefit of the permutation lies in breaking correlation between neighboring DCT coefficients, which in turn results in an i.i.d. host signal. Particularly, the host signal $X_h$, i.e., mid-frequency DCT coefficients, can be well modelled by a generalized Gaussian distribution.

![Figure 3. Extraction of 22 mid-frequency DCT coefficients from a $8 \times 8$ DCT block.](image)

The watermark encoder wants to hide as many information bits as possible into the host signal. However, it
has to keep the resulting watermarked image perceptually the same as the original one, which limits the maximum amount of information that can be embedded. In our scheme, a rate-1/15 watermark encoder is implemented, which is equivalent to hiding $22528/15 \approx 1500$ bits into $X_h$ (the remaining $22528-1500 \times 15=28$ samples of $X_h$ are kept unchanged) with a total image data-hiding rate of $0.0226 \text{ bit/pixel}$. However, Erez et al.\cite{erez2004} showed that to achieve the best performance under a very low embedding rate, a lattice precoding-based dirty-paper scheme must combine time-sharing and a higher rate dirty-paper code. Hence in our design, the embedding rate of 1/15 is broken into two parts: a three-to-one projection (for simulating time-sharing) followed by a rate-1/5 dirty-paper code based on TCQ and IRA codes.

Since each sample of the host signal $X_h$ is assumed to bear the same perceptual sensitivity (due to perceptual shaping), the three-to-one projection operation is simply the addition of each three consecutive host samples. If we write $X_h$ and the projected samples $X_p$ as column vectors, the projection can be written as $X_p = TX_h$, where

$$
T_{7500 \times 22500} = \begin{bmatrix}
1 & 1 & 1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & 1 & 1 \\
\end{bmatrix}.
$$

The employed rate-1/5 DPC scheme exploits memory-8 TCQ and non-systematic IRA codes designed using EXIT chart technique.\cite{kostakis2006} (The dirty-paper code design details are given in the next subsection.) It embeds 1500 bits into $X_p$ and outputs $Y_p$ as the watermarked signal. Let $W_p = Y_p - X_p$ be a watermark, equally distributed among $X_h$ samples by an “inverse projection” operation. This means $Y_h = X_h + SW_p$, where $S = \frac{1}{3}T^t$ satisfies $TS = I$, with $I$ being the identity matrix. Then, $Y_h$ passes an inverse permutation, and the resulting signal replaces the corresponding samples in $X_s$. Finally, the watermarked image $Y_i$ is obtained after the inverse shaping and inverse DCT transform. An example of the watermarked image together with the original Lena image is shown in Figure 4.

![Figure 4](image.png)

**Figure 4.** (a) The original $256 \times 256$ Lena image; (b) the same image after watermarking. The embedding rate is 0.0226 bit/pixel.

### 4.2. Dirty-paper Code Design for Image Data-hiding

The dirty-paper codes proposed in\cite{erez2004,farsad2005} are designed for large codeword lengths (e.g., for the code rate $k/n = 1/4$, $k = 60000$ and $n = 240000$ bits). The used IRA profiles are carefully designed using EXIT chart technique.\cite{kostakis2006} However, in our scheme with rate-$k/n = 1/5$ dirty-paper code, the size of the image imposes much shorter codeword lengths of $k = 1500$ and $n = 7500$ bits. Hence it is necessary to design new IRA profiles for this
specific codeword length to achieve better performance. In our code design, we fix the check node degree profile to degree 1: 20% and degree 2: 80%. Starting with the variable node profile in, we run a number of simulations to get the EXIT chart, and manually adjust variable node degrees and percentages such that the variable node EXIT curve and the check node EXIT curve are separated from each other. We can do this because the check node EXIT curve can be assumed to be independent of the variable node profile. The resulting variable node profile is: degree 2: 57.67%, degree 3: 28.53%, degree 10: 13.07%, and degree 366: 0.73%.

A problem with small codeword lengths is the high probability of short cycles and small-weight error events, which arise in the construction of the IRA code graph (or equivalently, the interleaver) and may cause severe performance degradation. Figure 5 (a) and (b) shows that when short-length cycles exist among degree-2 variable nodes, the IRA code becomes singular, i.e., flipping the value of these degree-2 nodes does not change the parity bits. Then the decoder can only “guess” the bits and hence the error floor is high. In general, short cycles also harm the iterative decoding algorithm since they introduce positive feedback in the associated log-likelihood ratios. Figure 5 (c) is an example of a weight-2 error event caused by one degree-2 variable node. When its value changes, only two parity bits are flipped. This will give us a minimum distance of 2, while it is obviously unacceptable. One can think about many other cases that may cause small-weight error events. This is due to the nature of IRA codes and cannot be completely eliminated. Our solution is to use the progressive edge growth algorithm to construct the IRA code graph edge by edge. This algorithm is devised for general Tanner graphs to maximize the girth in a best effort sense by progressively establishing edges between variable and check nodes in an edge-by-edge manner. We modified the algorithm to take into account the problem of small-weight errors. The simulation shows that this method is very effective in the sense of lowering the error floors.

![Examples of short cycles and small-weight error events.](image)

**5. SIMULATION RESULTS**

In this section we report our simulation results when the attacker adds i.i.d. Gaussian noise to the signal $Y_h$ (and hence to $Y_p$). Thus, $Z_h = Y_h + N_h$, where $N_h \sim \mathcal{N}(0, \sigma^2_h)$. Note that this setting is the same as in the DPC problem, and the WNR value corresponds to the SNR value in DPC, because both $X_h$ and $Y_h$ are equally scaled version of $X_s$ and $Y_s$, respectively. The obtained simulation results are plotted in Figure 6.

From the figure, we can observe that the turbo cliff appears at WNR=-7.9 dB (with a BER less than $10^{-3}$). The dirty-paper capacity at the rate of $R_{DPC}=0.2$ bit/sample is $C = 10 \log_{10}(2^{2R_{DPC}} - 1) = -4.96$ dB. However, since the three-to-one projection operation decreases the capacity by $10 \log_{10}(1/3) = -4.77$ dB, the WNR limit at the embedding rate of $R_{WM} = 1/15$ is $-4.96 - 4.77 = -9.73$ dB. Thus, our scheme performs 1.83 dB away from the theoretical bound. Compared to the 1.07 dB gap predicted for the large block length, we may conclude that the performance loss due to the reduced block length is 0.76 dB. Note that under the identical simulation conditions we achieve the same WNR of -7.9 dB as the best previously reported data-hiding scheme while being able to embed $1500 - 1122 = 378$ (or 33%) more bits.

**6. ROBUSTNESS TESTS**

In the previous section, we considered only the case of the AWGN attack on $Y_h$. In this section, we present simulation results for four different attacks: additive white Gaussian noise in the image domain (i.e., on $Y_I$), low-pass filtering, volumetric scaling, and JPEG lossy compression. Since the results are image-dependent, we
average the performance in terms of BER and message error rate (MER) over six standard gray-scale $256 \times 256$ images: Lena, Aer, House, Lady, Scene, and Space. All images are shown with the reduced size in Figure 7. After each attack, the altered pixel value of an image $Z_i$ is rounded to the nearest integer and bounded to the range of $[0, 255]$. Note that this introduces a small amount of distortion regardless of the attack.

6.1. Additive White Gaussian Noise

The attacker is an AWGN on the watermarked image $Y_i$; that is, $Z_i = Y_i + N_i$, where $N_h \sim \mathcal{N}(0, \sigma_N^2)$. The obtained results are shown in Figure 8 (a). Since the problem now is channel coding with side information over a colored Gaussian channel, the capacity limit is unknown. We can see that our scheme performs equally well for the range of $\sigma_N$ between 1 and 6 achieving MER below $2 \times 10^{-2}$. It is uniformly better than the scheme of. 

6.2. Low-pass Filtering

The watermarked images pass a time-domain low-pass Gaussian filter of width $\sigma_g$. The robustness curves are presented in Figure 8 (b). The turbo cliff (with MER less than 0.1) appears at around $\sigma_g = 0.48$, which is worse than the scheme of, where MER= 0.1 is reached at $\sigma_g = 0.60$.

6.3. Lossy Compression

If the attacker tries to compress the watermarked image $Y_i$ using JPEG compression standard, the watermark decoding performance will depend on the quality factor $QF \in [0, 100]$ used in the quantization of DCT coefficients. (The higher the quality factor $QF$ is, the less compressed the image is.) Comparing our results in Figure 9 (a) with those reported in, we observe that our scheme performs slightly worse (MER=0.1 is achieved at $QF= 70$). However, when $QF< 70$, the perceptual quality of the decompressed image $Z_i$ is quite poor due to a high compression noise. Thus in the range of $QF$’s where the image quality after the attack is still acceptable, both schemes succeed to achieve MER below 0.1.
6.4. Valumetric Scaling

The scaling attack changes the watermarked image linearly in value; that is, $Z_i = aY_i$, where $a$ is a scaling factor. The robustness performance in terms of $a$ is shown in Figure 9 (b). As opposed to the scheme of\textsuperscript{14}, we can see that the performance curve of our scheme is roughly symmetric about the line $a = 1$. Our scheme performs worse than the scheme of\textsuperscript{14} when $a < 1$ (scaling down), and slightly better when $a > 1$ (scaling up). This is reasonable because we use a lattice-based scheme which is in nature vulnerable under a scaling attack. Moreover, for images scaled by a factor $a$ outside the range (0.8, 1.2) (corresponding to the range of MER below 0.1), perceptual quality is usually unacceptable.

7. CONCLUSION

We have proposed a high-rate data-hiding scheme which embeds information bits into perceptually-shaped DCT coefficients of an image using near-capacity dirty-paper codes. At the same level of watermark-to-noise ratio under additive white Gaussian noise attack, our scheme can embed 33% more information bits than the technique
in. Other robustness tests on our scheme show comparable performance to the scheme of against additive white Gaussian noise, low-pass filtering, valumetric scaling, and lossy compression.

REFERENCES