

Informed detection revisited

Jeffrey A. Bloom¹ and Matt L. Miller²

¹ Sarnoff Corporation, Princeton NJ 08540, USA,

² NEC Labs America, Princeton NJ 08540, USA,

Abstract. Watermarking systems can employ either *informed detection*, where the original cover work is required, or *blind detection*, where it is not required. While early systems used informed detection, recent work has focused on blind detection, because it is considered more challenging and general. Further, recent work on “dirty-paper watermarking” has suggested that informed detection provides no benefits over blind detection.

This paper discusses the dirty-paper assumptions and questions whether they apply to real-world watermarking. We discuss three basic ways in which an informed video-watermark detector, developed at Sarnoff, uses the original work: canceling interference between the cover work and the watermark, canceling subsequent distortions, and tailoring the watermark to the perceptual characteristics of the source. Of these, only the first is addressed by theoretical work on dirty-paper watermarking. Whether the other two can be accomplished equally well with blind watermarking is an open question.

1 Introduction

A watermarking system embeds a message into a cover work, such as an image, video clip, or audio stream, without perceptibly changing that work. Systems can be divided into two major categories according to whether detecting and decoding the message requires knowledge of the original, unwatermarked cover work. In a system that employs *blind detection*, only the watermarked work is required during the detection process, while in *informed detection*, the detection process requires both the watermarked work and the unwatermarked original.

At first glance, it may appear that informed detection should yield better performance than blind detection. An informed watermark detector can subtract the unwatermarked work from the watermarked one, easily canceling out any interference in the message that results from the cover work. Nevertheless, the past decade of watermarking research has seen a steady decline in interest in informed-detection systems, culminating in a belief that informed detection should *not* offer any fundamental advantage over blind detection. The main purpose of the present paper is to discuss whether that belief is justified.

Judging from our experience, the declining interest in informed watermarking can be attributed to several causes. From the point of view of watermark researchers, blind detection is attractive because it is more generally applicable

than informed detection. An informed detector can only be used in applications, such as traitor tracing or proof of ownership (see [9]), where those who need to detect the watermark have access to the original. In other applications, such as copyright notification and copy prevention (again, see [9]), the general public must be able to detect watermarks, so we cannot employ informed detection without publicly distributing the unmarked originals (which, presumably, would defeat the purpose of the watermark). On the other hand, a blind system can be used for any application, provided its performance is sufficiently good.

Furthermore, there is a consensus that developing blind watermarking systems is more challenging, and therefore more interesting, than developing informed systems. The problem of canceling out interference from the cover work without knowledge of the original at the detector cannot be solved without inventing clever ideas that make for good research publications and patents. Informed detection, by contrast, seems like a solved problem.

At the same time, from the point of view of watermark users, the possible applications of informed detection were not as attractive as applications requiring blind detection. A few companies tested the market for proof-of-ownership systems (e.g., [20]), but found that market to be weak. Other potential applications of informed detection, such as traitor tracing, were overshadowed by a quest for a silver bullet: a system that could *prevent* illegal copying. It was the quest for such copy prevention systems that led to projects like SDMI [14] and watermarking for DVD [1], which inspired much watermarking research in the late 90's.

Finally, in 2000, Chen & Wornell [2] introduced Costa's *dirty paper* result [8] into watermarking research. This result strongly suggests that, in theory, the performance of a properly designed blind-detection system should be just as good as any informed-detection system. Application of Costa's basic principles led directly to dramatic improvements in the performance of blind-detection watermarking systems [3, 11, 6, 17, 18], and it appeared that there should no longer be any reason to work on informed detection.

However, the quest for a silver bullet against illegal copying became bogged down in real-world problems. For a variety of reasons – mostly non-technical – no watermarking system has yet been deployed for DVD, and SDMI-compliant devices have yet to find a market. There is, therefore, a growing interest in less-complete aids to copyright enforcement, most notably the application of watermarks for traitor tracing. This interest has been encouraged by the recent success of a video watermark in identifying two men responsible for a substantial movie-pirating operation [13]. With rising interest in a potential application of informed detection, it is worth revisiting the question of whether blind and informed systems are, in fact, equally capable.

Our discussion of this question begins by describing, in Section 2, a watermark developed at Sarnoff Corporation between 2001 and 2003. This system is designed for traitor tracing in digital cinema, and employs informed detection. Though the ways its detector uses the unwatermarked original are not unique, some of them exploit aspects of real-world watermarking that differ dramatically

from Costa's theoretical formulation, and thus call into question the applicability of his result to watermarking. That is, we believe Costa's result *does not* indicate that watermarking with blind detection is necessarily equivalent to watermarking with informed detection. This point is argued and elaborated in Section 3. Finally, Section 4 concludes the paper with a summary of our argument and the challenges it poses.

2 Sarnoff's video watermark

Sarnoff's video watermark [16] was developed to help identify the source of unauthorized distributions of movies exhibited in digital cinemas³. Each authorized copy of a motion picture, including those distributed prior to the theatrical release, is watermarked with a unique serial number. In addition, each projector will embed a watermark indicating the identity of the screen onto which it is projecting, the show time, and perhaps some additional information about the operator or the showing. Insiders with direct access to the movie who make unauthorized copies can then be identified. Theaters that allow audience members to videotape the movie with a camcorder can be identified if a copy of that tape is distributed. Such a watermark provides the motion picture studies with a tool to help them understand and manage the piracy problem.

The requirements for such a watermark are quite strict. Though the data payload can be fairly small (about 0.5 bits per second is sufficient), the fidelity and robustness must be extremely high. Part of the point of digital cinema is to improve picture quality, so any degradation caused by watermarking will be unacceptable. For example, the watermark must be completely invisible in images with the quality shown in Figure 1 (a). At the same time, pirated copies can have very poor quality, so the watermark must be detectable in severely degraded video, such as that shown in Figure 1 (b).

To meet these requirements, the Sarnoff watermark uses very low frequency spatio-temporal patterns referred to as *carriers*. Each carrier is a spatially-limited pattern, one third to two thirds of the screen height in size, that appears and disappears slowly over time. Carriers are modulated (e.g., by sign or phase) to encode 1's and 0's and added to the frame sequence. This design takes advantage of the human visual system's low sensitivity to extremely low frequencies. At the same time, it makes for extremely robust watermarks, because virtually all processing to which pirated video might be subjected preserves the low frequencies.

Unfortunately, in spite of the eye's insensitivity to low frequencies, randomly-placed carriers are occasionally visible. To counter this problem, the embedder employs some automated perceptual analysis to identify locations where carriers can be invisibly placed. The watermark is then embedded at a subset of these locations only.

³ Original work was funded in part under the U.S. Department of Commerce, National Institute of Standards and Technology, Advanced Technology Program, Cooperative Agreement Number 70NANB1H3036.

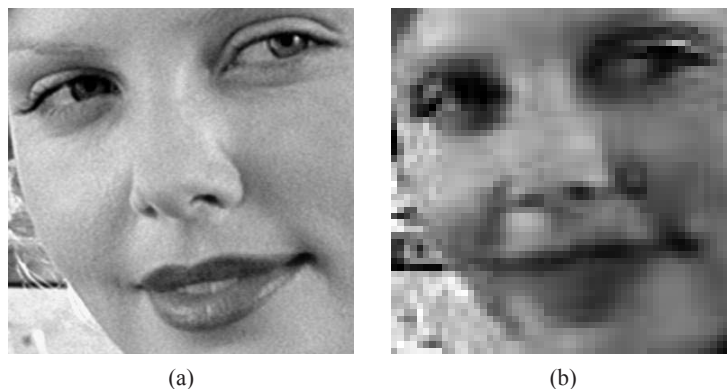


Fig. 1. A digital cinema application places extreme requirements on fidelity and robustness. The images shown are close-ups of a small region of a movie frame (courtesy of Miramax Film Corporation). The fidelity of the high resolution detail of the original content, shown in (a), must be maintained while the watermark must be detectable in video with quality like that in (b).

The detector receives the suspect video (i.e. the pirated copy) along with two pieces of side information: the original unwatermarked video and a description of the carrier locations used by the embedder, as determined by perceptual analysis. Note that this second piece of side information could just as well be obtained by re-running the perceptual analysis on the original video, so, in principle, only the first piece of side information is truly needed.

The detector first adjusts the original video's geometry, timing, and color histogram so that it matches the suspect video (see [4, 5] for details on how this is done) and subtracts the two, obtaining a difference video. It then adjusts the pattern of carriers the same way that it adjusted the original video, and correlates it with the difference video. This yields either a positive or negative correlation for each carrier, which is then decoded to obtain the watermark information.

The embedding and detection processes are described in more detail in [16]. For our purposes here, however, we need only focus on the three basic ways in which the original video is used in the detector:

1. The original is subtracted from the suspect video to cancel out interference between the video and the watermark.
2. The pattern of carriers is adjusted according to the registration of the original and suspect videos, thus canceling out the effects of temporal, geometric, and histogram distortions.
3. The pattern of carriers that the detector looks for, and thus the appearance of the watermark, depends on the perceptual properties of the original video.

The question we face is whether these functions – canceling interference, canceling distortions, and perceptual shaping – could, in theory, be performed equally

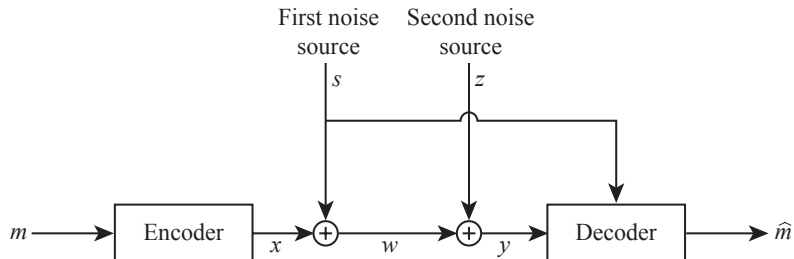


Fig. 2. Channel with informed decoding.

well with a blind detector. To answer this question, we must take a closer look at the theoretical reasons for believing that they might be.

3 Informed detection vs. blind detection

The reason for believing that blind and informed detection are equivalent comes from Costa's dirty-paper result of 1983 [8]. This says that the two *are* equivalent, for a specific type of channel that is highly analogous to watermarking. After briefly reviewing the result here, we go on to discuss the differences between Costa's channel and the reality of watermarking, and how those differences are exploited in Sarnoff's watermark.

3.1 Costa's dirty-paper result

The channels Costa discussed are illustrated in Figure 2 and Figure 3. A message, m , is mapped to a signal vector, x , by the encoder. This signal is limited by a power constraint:

$$\frac{1}{n} \sum_i x_i^2 \leq P \quad (1)$$

where n is the number of elements in the vector. x is then corrupted by two additive, Gaussian noise vectors, s and z . s is drawn from the *first noise source*, and z is drawn from the *second noise source*. We refer here to $x + s$ as w , and to $w + z$ as y . y is received by the decoder, and decoded to the received message, \hat{m} . Figure 2 shows the case of *informed decoding*, in which s is provided to the decoder as side information. Figure 3 shows the case of *informed encoding*, in which s is provided to the encoder, before it must select x . The question is what are the capacities of these two channels?

Clearly, in the case of informed decoding (Figure 2), the capacity can be computed as if the first noise source does not exist. The decoder can simply subtract s from y , obtaining $x + z$. Thus, the capacity is given by the standard formula for capacity of an additive Gaussian channel:

$$C = \frac{1}{2} \log \left(1 + \frac{P}{\sigma_z} \right) \quad (2)$$

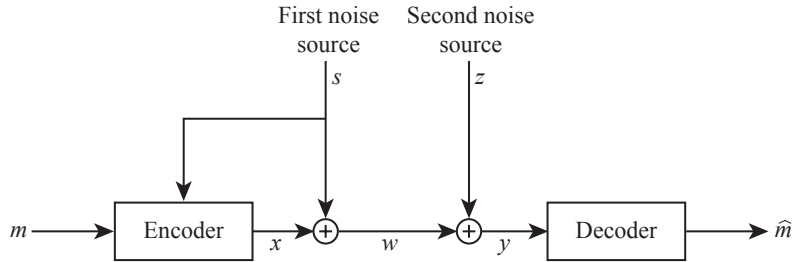


Fig. 3. Channel with informed encoding.

where C is the capacity, and σ_z is the standard deviation of the second noise source.

In the case of informed encoding (Figure 3), it is tempting to try a similar approach, by encoding the message with some vector, u , and letting $x = u - s$. This would mean that $w = u$ and $y = u + z$, so the resulting channel looks like one that has a single noise source. The problem is that $\sum (u_i - s_i)^2$ will generally be larger than $\sum u_i^2$, so u is more limited by the power constraint than x . In general, u must be limited by $\sum u_i^2 \leq P - \sigma_s^2$, where σ_s is the standard deviation of the first noise source. Thus, the code rate achieved by this approach must be lower than the capacity of Equation 2.

Costa showed, however, that higher code rates can be achieved using

$$x = u - \alpha s \quad (3)$$

where α is a carefully chosen constant. When α is computed as

$$\alpha = \frac{P}{P + \sigma_z} \quad (4)$$

the highest possible code rate is actually *equal* to the capacity of Equation 2. Thus, knowledge of the first noise source at *either* the encoder or the decoder allows its complete cancellation for purposes of computing channel capacity.

The implications of this for watermarking become clear when we see that the channels of Figures 2 and 3 are essentially the same as informed and blind watermarking, respectively. The first noise source is analogous to the original cover work. w is analogous to the watermarked work. The second noise source is analogous to subsequent distortions applied to the watermarked work. And the power constraint is analogous to a limit on perceptibility of the watermark.

If these analogies are perfect, then Costa's result means that the cover works have no effect on watermark capacity, and blind and informed watermarking should have equivalent performance. We now, therefore, examine each of these analogies in turn.

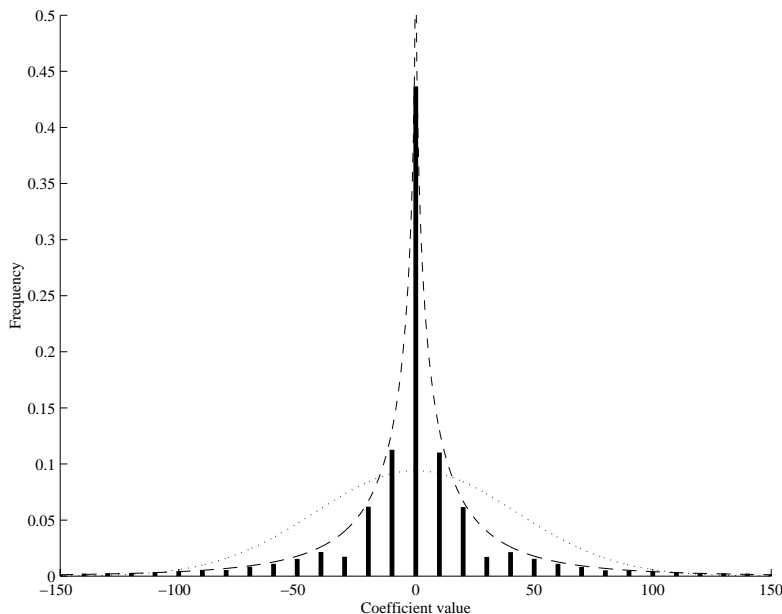


Fig. 4. Example of the distribution of values for an image transform coefficient. The bars show the distribution of one coefficient in the 8x8 block DCT of 100 images. The dotted line shows the best approximation of this distribution as a Gaussian. The dashed line shows the best approximation of the distribution as a generalized Gaussian, with a shape parameter of 0.5.

3.2 First noise vs. cover works

Is Costa's first noise source a good analogy for the original cover work? In general, no. The problem is that most types of media are not Gaussian distributed.

The simplest models that provide reasonable matches to real distributions of images and audio clips are generalized Gaussians in transform domains. For example, Figure 4 shows the distribution of values for one block DCT coefficient in a corpus of 100 images. The dotted line shows the best fit obtainable with a normal Gaussian. The dashed line shows a fit obtainable with a generalized Gaussian, computed as

$$P(x) = \frac{\lambda\beta}{2\Gamma\left(\frac{1}{\beta}\right)} e^{-|\lambda x|^\beta} \quad (5)$$

with the *shape parameter*, β , equal to 0.5, and the scaling constant, λ , equal to 0.2583.

In truth, the distribution of cover works is more complex than that shown in Figure 4. Different image textures, different musical instruments, different voices, etc. all have different statistical characteristics, and real works of media are drawn from a complex mixture of these distributions.

Fortunately, it has been shown that none of this matters. Cohen & Lapidoth [7], and Erez, Shamai & Zamir [12] have proven independently (and in different ways) that Costa’s result holds regardless of the distribution of the first noise source. As long as this noise is additive, its effect on channel capacity can be cancelled at the encoder.

This means that, in principle, Sarnoff’s watermark detector should not need to subtract the original from the suspect video. Instead, the embedder could use coding techniques based on Costa’s proof (*dirty paper* coding [17]) to cancel interference from the cover work.

3.3 Second noise vs. digital media distortions

The analogy between Costa’s second noise source and actual media distortions is weaker than that between the first noise source and original cover works. Digital media is rarely corrupted by the addition of Gaussian noise. Instead, more common distortions include operations like

- volumetric scaling (multiplying each pixel or audio sample by some scaling factor),
- spectral filtering,
- non-linear filtering,
- quantization for lossy compression, and
- geometric distortions.

This distortions are completely different from additive noise.

Unlike the first noise source, furthermore, deviation from a Gaussian distribution in the second noise source is likely to invalidate Costa’s result. In the same paper where they proved that the first noise source can be arbitrary, Cohen & Lapidoth put forth a conjecture that Costa’s result holds *only* if the second noise is additive Gaussian. Though they did not quite prove this conjecture, their argument for its validity is very strong.

One possible escape from Cohen & Lapidoth’s conjecture is that it only holds for systems that compute x in the manner of Equation 3. It is likely that, with different equations for x , Costa’s result can be obtained for some larger class of second noise sources. However, it is far from clear that this should include every conceivable noise source, or even the common types of distortion listed above.

A major issue with digital media distortions is that they are *not* additive. The change made in a work by each of the distortions listed above is highly dependent on the work itself. That is, there is significant mutual information between w and z . This is a complete departure from Costa’s basic dirty-paper channel, and allows an informed detector to do something Costa’s decoder has no opportunity to do: obtain information about the distortion by looking at the cover work.

In Sarnoff’s system, the detector obtains information about geometric and histogram distortions by comparing the unmarked original with the suspect video. It uses this information to cancel out the effect of these distortions, thus

effectively increasing the channel capacity beyond what could be achieved in a channel that had no first noise source. There is, as yet, no reason to believe that a similar increase in channel capacity can be obtained with a blind detector.

3.4 Power constraint vs. fidelity limit

The third part of the analogy between Costa’s channel and watermarking is the mapping of watermarking’s fidelity constraint with the channel’s power constraint. This, too, is not a perfect analogy.

In the analogy, the vector x is the pattern added to the cover work s . By limiting the magnitude of x with a power constraint, we hope to limit the perceptibility of the change caused by adding it to s . In reality, however, the perceptibility of changes in a work is highly dependent on that work. For example, an audio signal of pure white noise can be dramatically changed without changing its perceptual effect, while a recording of a single, clear musical note is much less forgiving. This means that, at a minimum, the power limit P should be viewed as dependent on s .

Furthermore, the limit should also depend on x itself. Figure 5 shows an example. This shows two different patterns for x , x_0 and x_1 , and the perceptual effect of adding each pattern, with the same power, to the same image. Because x_1 is shaped to have energy only in areas where the image has texture, its impact is much less perceptible than the impact of x_0 . Thus, the power constraint for x_0 should be tighter than for x_1 . Clearly, however, this is image-dependent, so to capture the true fidelity limit, we would need a power constraint based on both s and x .

It is not clear what effect this complication has on channel capacity. Certainly, an informed embedder can shape x to take advantage of perceptual phenomena, improving the performance of a system with blind detection. This has been demonstrated in numerous practical watermarking systems. The question, however, is whether better performance can be obtained when the detector is informed.

In the Sarnoff watermark, both the embedder and detector are informed, so the channel matches that shown in Figure 6. This allows the embedder and the detector to agree on the placement of the carriers, effectively designing a unique code for each cover video, according to that video’s perceptual characteristics. Whether this yields a fundamental improvement in channel capacity is an open question – a question that is not addressed by the analogy with Costa’s channel.

3.5 A possible solution: perceptually-uniform space

One possible way to make the analogy between Costa’s channel and watermarking exact would be to use a perceptually-uniform space. In principle, we could define some non-linear transform that, when applied to an image, maps it into a space where Euclidian distance corresponds to perceptual distance. This should simplify the two main areas where actual watermarking is more complex than Costa’s channel.

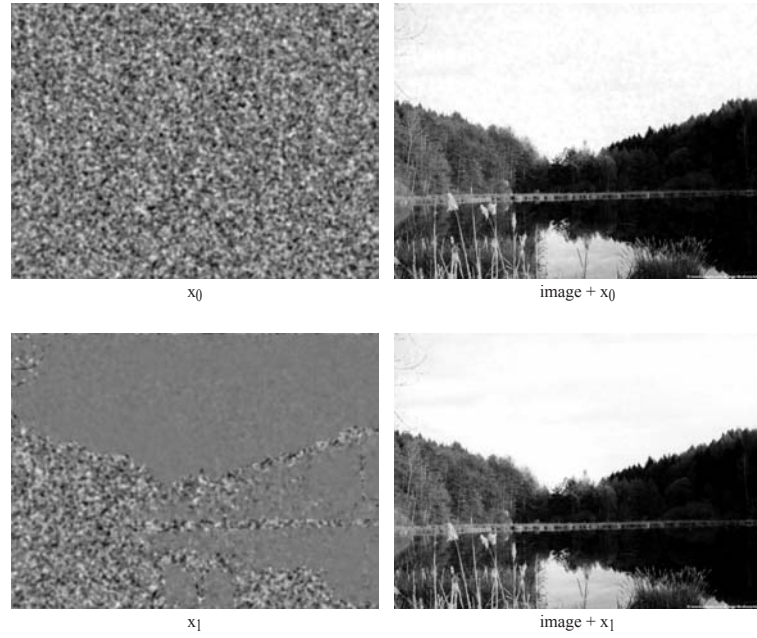


Fig. 5. Effect of adding two different patterns, x_0 and x_1 , to the same image with the same power.

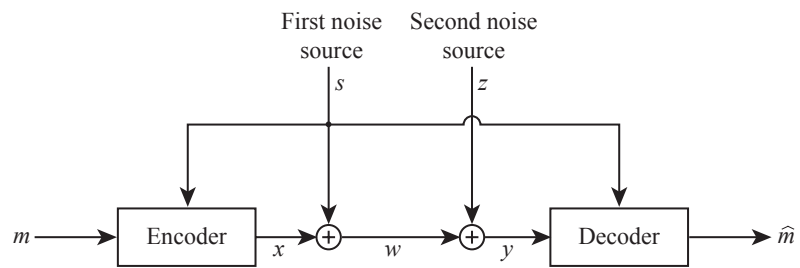


Fig. 6. Channel with both encoding and decoding informed.

Clearly, in a perceptually-uniform space, the power constraint will be a good analogy for watermarking's fidelity limit, because the perceptual difference between two images is, by definition, indicated entirely by the Euclidian distance between them. Perceptual distortion does not vary with location in the space, nor does it vary with direction of the distortion vector.

Less clearly, a perceptually-uniform space might also simplify the distribution of subsequent distortions, making it more analogous to the second noise source in Costa's channel. The reason for this is that all the distortions to which media will be subject are intended to preserve (to some degree) the perceptual quality of that media. Thus, if we are in a space where the perceptibility of a distortion is dependent only on the length of the distortion vector, and independent of the starting point, then we might be justified in assuming that the distribution of subsequent distortions is likewise independent of the work being distorted. Basically, the dependancy between the distortion and the work is subsumed in the non-linearities of the transform into perceptually-uniform space. Under such an assumption, we should be able to represent distortions with a simple, additive noise source.

Unfortunately, defining a perceptually-uniform space is extremely difficult, and may never be possible in practice. There exist several perceptual models for judging small differences between images [10, 15, 19], and these might serve as starting points for developing such a transform, but there are, as yet, no good models for measuring larger perceptual differences. A space that's uniform over small changes should typically be sufficient to make the fidelity limit analogous to a power constraint, because the fidelity limits typically only allow small changes. But subsequent distortions, especially if introduced by a pirate, can be much larger (see Figure 1 again). We have little hope, at present, of making a space that is uniform over a large enough area to allow such distortions to be modelled as simple additive noise.

Thus, even if a perceptually-uniform space is theoretically possible, it is likely to be impractical. In the absence of such a space, informed detection will probably provide additional channel capacity.

4 Conclusion

The application of dirty-paper coding to the watermarking channel has suggested that blind detection with side information available at the encoder is equivalent to informed detection, in which side information is made available to the detector. This extension relies on the validity of three important assumptions: the distribution of cover works can be modelled as an additive Gaussian noise source, the distribution of noise applied to a watermarked work prior to detection can be modelled as an additive Gaussian noise source, and the maximum allowable watermark power is independent of the cover work.

While the first of these assumptions does not directly hold, there is research that shows that Costa's result applies regardless of the distribution of cover works. The other two assumptions are not so easily reconciled. The distortions

applied to watermarked works are highly dependent on the cover work and are neither additive nor Gaussian. The fidelity limit places restrictions on the shape of the watermark as well as on its power. Further, that shape is dependent on the cover work. Thus, it appears that Costa's result *does not* indicate that watermarking with blind detection is necessarily equivalent to watermarking with informed detection.

Sarnoff's watermarking system for traitor tracing in digital cinema represents an example of informed detection in which the detector can infer information about the second noise source and the fidelity constraint applied during embedding. Costa's framework does not address such systems.

If the watermarking community is to continue to rely on Costa's dirty paper model to assess the capacity of the watermarking channel, there are a number of questions that must be addressed.

1. We need to better understand the role of the second noise source. It appears that providing the detector with information about real distortions, those that are non-additive and highly dependent on the cover work, can improve the channel capacity. Can we prove that such knowledge does not improve capacity?
2. We need to better understand the role of the fidelity constraint. Can we design a perceptually uniform space in which the fidelity constraint becomes a Euclidian power constraint? Or can we extend Costa's model to address content-dependent fidelity constraints rather than power?
3. An issue that has not been addressed here is the challenge of distinguishing between distortions that are introduced by the second noise source and the watermark pattern itself. Given the original content, how can a detector infer and compensate for distortions without damaging the watermark pattern?

References

1. J. A. Bloom, I. J. Cox, T. Kalker, J-P Linnartz, M. L. Miller, and B. Traw. Copy protection for DVD video. *Proc. IEEE*, 87(7):1267–1276, 1999.
2. B. Chen and G. W. Wornell. An information-theoretic approach to the design of robust digital watermarking systems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1999.
3. B. Chen and G. W. Wornell. Quantization index modulation: A class of provably good methods for digital watermarking and information embedding. In *Proc. Int. Symp. Inform. Theory (ISIT-2000)*, 2000.
4. H. Cheng. Temporal video registration. In *Proc. of IEEE Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP'03)*, volume 3, pages 489–92, Hong Kong, China, April 2003.
5. H. Cheng and M. Isnardi. Spatial, temporal and histogram video registration for digital watermark detection. In *Proc. of IEEE Int'l Conf. on Image Processing (ICIP'03)*, volume 2, pages 735–8, Barcelona, Spain, Sept. 2003.
6. Jim Chou, S. Sandeep Pradhan, and Kannan Ramchandran. On the duality between distributed source coding and data hiding. *Thirty-third Asilomar conference on signals, systems, and computers*, 2:1503–1507, 1999.

7. A. S. Cohen and A. Lapidoth. Generalized writing on dirty paper. In *International Symposium on Information Theory (ISIT)*, 2002.
8. M. Costa. Writing on dirty paper. *IEEE Trans. Inform. Theory*, 29:439–441, 1983.
9. I. J. Cox, M. L. Miller, and J. A. Bloom. *Digital Watermarking*. Morgan Kaufmann, 2001.
10. Scott Daly. The Visible Difference Predictor: An algorithm for the assessment of image fidelity. In A. B. Watson, editor, *Digital Images and Human Vision*, chapter 14, pages 179–206. MIT Press, 1993.
11. J. J. Eggers, J. K. Su, and B. Girod. A blind watermarking scheme based on structured codebooks. In *IEE Seminar on Secure Images and Image Authentication*, pages 4/1–4/21, 2000.
12. U. Erez, S. Shamai, and R. Zamir. Capacity and lattice-strategies for cancelling known interference. In *Proc. of the Cornell Summer Workshop on Inform. Theory*, Aug. 2000.
13. B. Fritz and T. M. Gray. Acad member tied to piracy bust. *Variety*, Jan. 23, 2004.
14. Secure Digital Music Initiative. SDMI portable device specification, 1999. Available at <http://www.sdmi.org>.
15. J. Lubin. The use of psychophysical data and models in the analysis of display system performance. In A. B. Watson, editor, *Digital Images and Human Vision*, chapter 14, pages 163–178. MIT Press, 1993.
16. J. Lubin, J. A. Bloom, and H. Cheng. Robust, content-dependent, high-fidelity watermark for tracking in digital cinema. *Security and Watermarking of Multimedia Contents V*, SPIE-5020:536–45, 2003.
17. M. L. Miller. Watermarking with dirty-paper codes. In *IEEE International Conference on Image Processing*, September 2001.
18. M. L. Miller, G. J. Doërr, and I. J. Cox. Applying informed coding and embedding to design a robust, high capacity watermark. *IEEE Transactions on Image Processing*, 13(6):792–807, 2004.
19. Christian J. van den Branden Lambrecht and Joyce E. Farrell. Perceptual quality metric for digitally coded color images. *Proc. EUSIPCO*, pages 1175–1178, 1996.
20. G. Voyatzis and I. Pitas. The use of watermark in the protection of digital multimedia products. *Proceedings of the IEEE, Special Issue on Identification and Protection of Multimedia Information*, 87(7):1197–1207, 1999.