EXPLORING DISCRETE APPROACHES TO LOSSY COMPRESSION SCHEMES FOR NATURAL IMAGE PATCHES

Ram Mehta[1], Sarah Marzen[2], Christopher Hillar[1]

[1]Redwood Center for Theoretical Neuroscience, [2]Department of Physics
University of California, Berkeley

ABSTRACT

Optimal compressions in a rate-distortion sense are usually discrete random variables, so clever discretizations of natural images might be key to developing better compression schemes. A new image compression method achieved good perceptual coding performance by using as primitives memories of a Hopfield network trained on discretized natural images. Here we explore why Hopfield network fixed-points are good lossy perceptual features even though the implied generative model (a second-order Lenz-Ising model) does not provide a state-of-the-art match to the true probability distribution of discretized natural images. Even so, we demonstrate that this deterministic coding scheme can achieve near-optimality by comparing with the rate-distortion function for discretized natural image patches.

Index Terms— natural images, Hopfield network, recurrent neural network, image compression, Lenz-Ising model

1. INTRODUCTION

For decades, researchers have tried to find image compression schemes that can achieve rate-distortion bounds [1]. To reduce the file size of digital images many lossy compression methods, including those based on linear coding, require quantization of real-valued variables computed (typically linearly) from the intensity values of the image. However, optimal codings in a rate-distortion sense are usually discrete [2, 3]. Thus, clever discretizations of natural images might be key to optimal lossy compression.

Recently, a new efficient image compression method [4] that uses a recurrent network of McCulloch-Pitts [5] abstract ON/OFF neurons was found to be highly competitive with widely-used JPEG [6] in a high-fidelity regime, according to a standard perceptual distortion measure [7, 8]. The method uses as discrete primitives all of the memories of a Hopfield network [9] trained with minimum probability flow estimation [10, 11] on a collection of 3 million discretized image patches from the van Hateren natural image database [12].

Although Hopfield networks have been used for various image processing tasks by many previous researchers [13], we could find only a few papers using such networks to find primitives of natural images for lossy compression [14, 15].

At first glance, the performance of this discrete lossy compression method is surprising. Our perceptual responses are highly sensitive to the higher-order statistics of natural images [16, 17], but second-order Lenz-Ising models [18] – the underlying probabilistic model of data for the Hopfield network – are sensitive only to correlation structure. Along those lines, high-quality coding usually requires a good fit between the model and data, but as we show here, the probability distribution of discretized image patches is generally not well-described by this particular statistical model.

We address aspects of these issues through a series of experiments on a range of input patches used for training. First, we show that these networks detect nontrivial local correlation structure in ternarized natural image patches; moreover, this structure is still present in networks trained on whitened data. Then, we verify that the statistical patch model underlying the learned Hopfield networks indeed does not predict the frequencies of the actual image ensemble, but does have similar local extrema. We suspect that this qualitative match underlies the quality of the network's lossy encoding. To test this hypothesis, we used the Blahut-Arimoto [19, 20] algorithm to find the rate-distortion function for ternarized ON/OFF natural image patches. Using this tool, we demonstrate that this deterministic network coding is nearly-optimal.

2. BACKGROUND

In [4] it was found that high perceptual quality digital image compression was possible by leveraging discrete local structure in natural images. One important component of that approach was to train a Hopfield network over discretized natural image patches. We briefly review the scheme.

Given a grayscale $L \times L$ digital image patch, we remove its mean and then set the patch's variance to be 1. We call such a patch *normalized*. Next, we discretize each pixel x in such a patch to a pair of abstract ON/OFF neurons $(ON, OFF) \in \{0, 1\}^2$ according to a parameter $\alpha \geq 0$. When $x > \alpha$, the discretized pixel is assigned $(ON, OFF) = (1, 0)$; similarly,
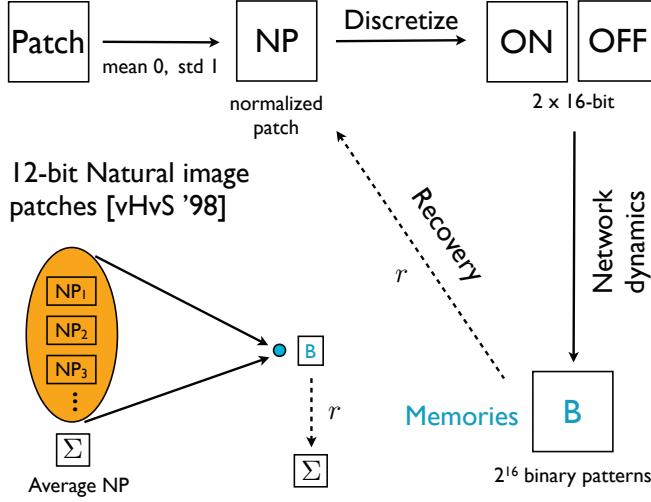
Fig. 1. **Image compression using a discrete network [4]**. Image patches are normalized, discretized, and then used to train Hopfield networks as described in the main text. We examine different discretization procedures (binarization instead of ternarization) and different training ensembles (natural images versus whitened versions).

when $x < -\alpha$, we have $(ON, OFF) = (0, 1)$; and finally, when $x \in [-\alpha, \alpha]$, we have $(ON, OFF) = (0, 0)$. We choose α so that 2α is the smallest pixel intensity difference that can occur. We can thus convert any grayscale $L \times L$ image patch into a binary vector of length $2L^2$. When no pixel maps to $(ON, OFF) = (0, 0)$, we call the patch *binary*.

A Hopfield network is then trained using minimum probability flow estimation on millions of discretized normalized patches obtained from a natural image database. After learning, the network acts deterministically on ternary patches by converging its linear threshold recurrent dynamics (each ON/OFF pair is viewed as a single unit for the asynchonous dynamics). It was found that such networks have all binary patches as fixed-points / memories, and that these primitives can serve as good feature labels for image compression. Specifically, a converged discretized normalized patch can be remapped to a continuous patch by averaging over natural images those normalized patches giving the same binary output. The mechanics are illustrated in Fig. 1 and Fig. 2.

To perform compression, a grayscale digital image is broken up into non-overlapping $L \times L$ patches, and then each patch is independently decomposed into a mean, a variance, and L^2 (ON, OFF) neurons. Means and variances are saved losslessly as PNG files while converged discretized patches are encoded with an entropy coder (e.g. Huffman) to get bytes on disk. To reconstruct the image, each ON/OFF patch is replaced by its continuous normalized representative and then means and variances are restored. Such a scheme is rate-competitive with JPEG for the same perceptual error.

We also explore here another, simpler discretization scheme that was used by previous workers, in which im-

age pixels are set to 0 if they are below the patch mean (or median), and set to 1 if they are above it [21, 22]. The workflow post-discretization was the same as for the ON/OFF networks except that we use single binary Hopfield units to represent data as opposed to pairs of ON/OFF neurons.

3. RESULTS

3.1. Ternary ON/OFF image patch correlations reflect higher-order image statistics

Weights of inferred Hopfield networks (second-order Lenz-Ising models) reflect the pairwise statistics of discretized image patches. However, higher-order statistics, i.e. greater than second-order, are crucial in explaining our perceptual response to natural images. We examine the role of higher-order statistics in learned parameters by estimating networks over whitened images as well as normal ones. As whitened images have no pairwise correlations, only higher-order image statistics can lead to structure in the inferred weights of discretized whitened images.

Fig. 3 shows learned weights for discretized $4{\times}4$ ON/OFF patches from the van Hateren natural image database (top
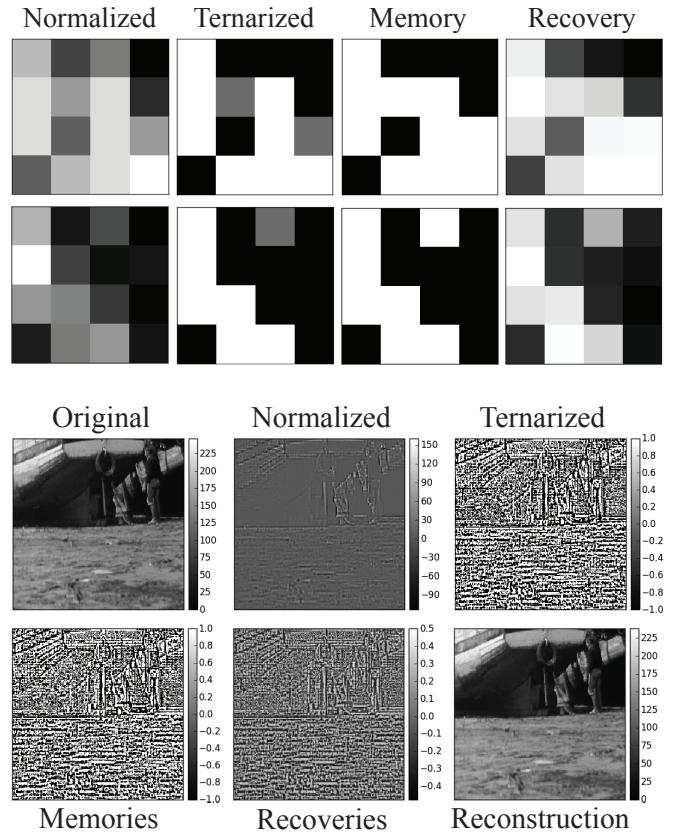

Fig. 2. Sample patches (top) and image (bottom) discretized and recovered using scheme from [4].
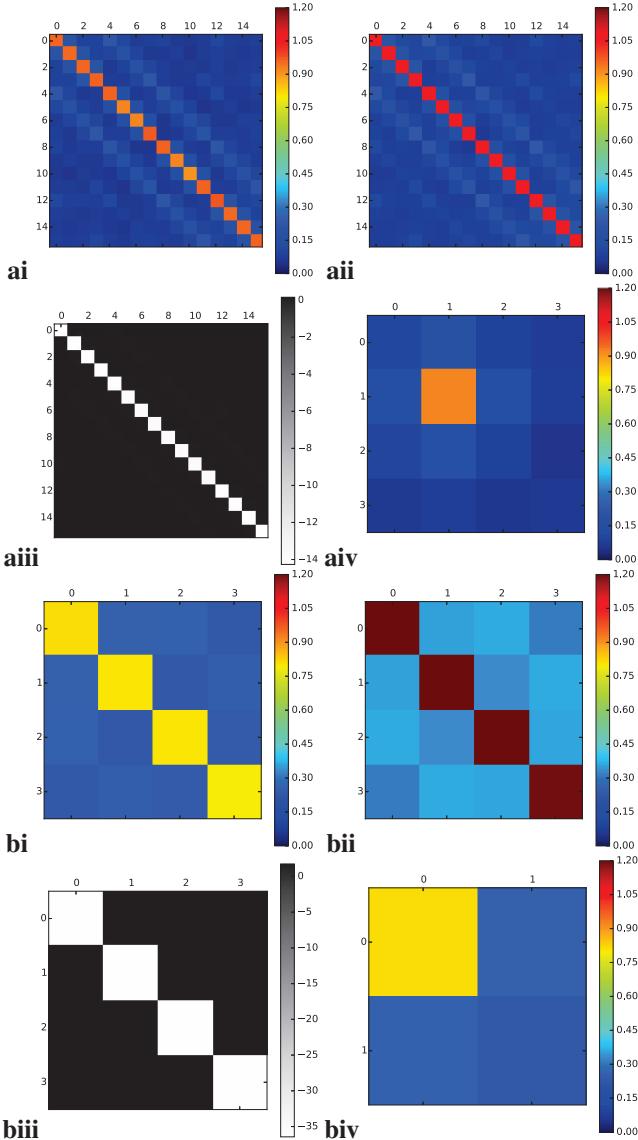
ai

aii

aiii

aiv

bi

bii

biii

biv

Fig. 3. **ON/OFF weights and thresholds after training on 3 million digital image patches**. **a)** 4×4 ON/OFF network trained on natural images. **i)** ON-ON weights, **ii)** OFF-OFF, **iii)** ON-OFF, and **iv)** single ON neuron's weights to neighboring ON neurons. For **ai** and **aii**, diagonal squares represent thresholds (parameters have been rescaled so that the average threshold is 1); for **aiv**, threshold is at neuron location (second from top and left in patch). **b)** As **a** using 2×2 ON/OFF network trained using whitened natural images.

four) and discretized 2×2 ON/OFF patches from whitened versions of these images (bottom four). In both cases, non-trivial correlation structure is being detected by the networks. The nearest neighbor structure of the weight matrix in Fig. 3**a** suggests that comparing binarized image patches to the ferromagnetic Lenz-Ising model makes for a surprisingly good analogy [22]. Notice also that the weight matrices are sparsely-connected, as are effective spacial connectivities between ganglion neurons in retina [23].

3.2. Good lossy encodings not matching input statistics

Typically, natural image (and most machine learning) models are evaluated on their likelihood [21]. High likelihood implies a successful match between the model's output statistics and the statistics of the data. Intuitively, it seems that good lossy encodings must implicitly require high likelihood models. Here, we describe how that view is an oversimplification.

The quality of a lossy encoding of X using codes $\tilde{X}$ can be evaluated by its proximity to the rate-distortion curve [1]:

$$R(D) = \min_{p(\tilde{\mathbf{x}}|\mathbf{x}):\, \mathbb{E}[d(\tilde{\mathbf{x}},\mathbf{x})]\leq D} I[\tilde{X}; X]. \tag{1}$$

This function gives the minimum coding rate of a codebook $p(\tilde{\mathbf{x}}|\mathbf{x})$ achieving an expected distortion of no more than D given a fixed distortion measure d. Formally, the codebook $p(\tilde{\mathbf{x}}|\mathbf{x})$ is a nonnegative matrix representing the probability of coding a state $\mathbf{x}$ drawn from X with the new state $\tilde{\mathbf{x}}$ from $\tilde{X}$.

Optimal lossy encodings $p(\tilde{\mathbf{x}}|\mathbf{x})$ satisfy, for some nonnegative parameter β, the following set of equations:

$$p(\tilde{\mathbf{x}}|\mathbf{x}) = \frac{\sum_{\mathbf{x}'} p(\mathbf{x}')p(\tilde{\mathbf{x}}|\mathbf{x}')}{Z(\mathbf{x}, \beta)} e^{-\beta d(\tilde{\mathbf{x}},\mathbf{x})}, \tag{2}$$

in which $Z(\mathbf{x}, \beta)$ is a normalizing factor and $p(\mathbf{x})$ is the distribution over inputs $\mathbf{x}$ (e.g., discretized image patches X). As β increases, we achieve less lossy encodings of the data. For instance, the case $D = 0$ corresponds to $\beta \to \infty$, giving the entropy of X as the optimal coding cost, whereas $\beta \to 0$ determines the trivial coding $R = 0$. Viewed as a mapping, expression (2) can be iterated to find an optimal point (R_β, D_β) on the rate-distortion function for any fixed β, and by sweeping out β, the rate-distortion function (1) is traced out parametrically (this is called the Blahut-Arimoto algorithm).

Here, we use the Hamming distance between $2L^2$-bit vectors (L^2-bit in the binary case) as our distortion measure d. The rate of the Hopfield networks is taken to be the entropy of the distribution of memories after converging dynamics over discretized natural image patches.

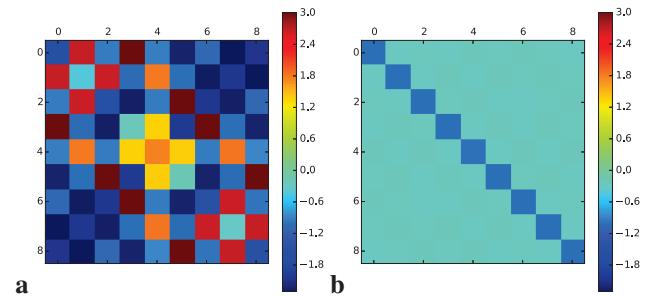In Fig. 5, we show both the statistical model fit and coding quality of Hopfield networks for 3×3 binarized image

a

b

Fig. 4. **Networks trained on 3 million 3×3 binarized patches**. **a)** Learned 3×3 binary network and weights; thresholds on diagonal (parameters scaled so that mean absolute value of thresholds is 1). **b)** Same as in **a** but network trained on whitened natural images.

patches. The Lenz-Ising model is not a strong statistical model of the input probability distribution, and the network compression does not yield a very effective lossy code.

Fig. 6 shows the coding and statistical model quality of the Hopfield networks for 2×2 ternarized ON/OFF image patches. These networks are far better lossy encoders of the input probability distribution (Fig. 6**a**); however, they clearly provide a suboptimal statistical model (Fig. 6**b**). Although the frequencies of the model do not match the data very well, the two distributions have similar local maxima – and this, we speculate, is the key to the high quality of the coding scheme.
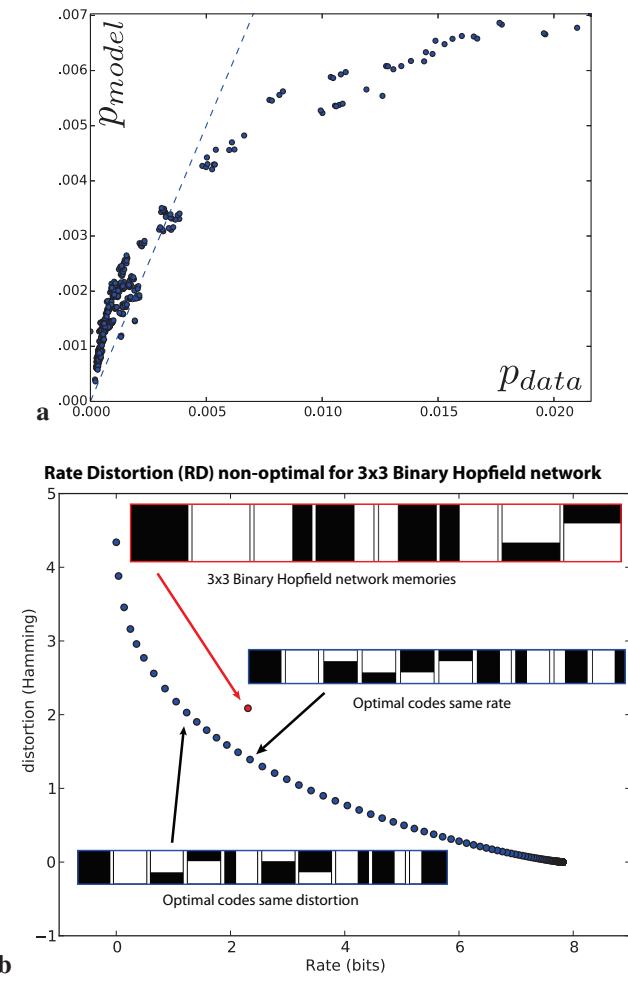

Fig. 6. **Modeling / coding** 2×2 **ON/OFF image patches**. **a)** Comparison of p_{data} and p_{model} by plotting both as a function of the particular binary vector. Bottom plot reveals that p_{data} and p_{model} have matching local maxima, but not matching frequencies. **b)** Expected distortion and rate for Hopfield networks trained on 2×2 ON/OFF natural image patches, relative to the true rate-distortion function.

Fig. 5. **Model fit and coding quality of Hopfield networks trained on** 3×3 **binarized natural image patches**. **a)** Scatterplot of p_{data} on the x-axis versus p_{model} on the y-axis, where p_{model} is taken to be the Lenz-Ising model underlying the Hopfield network in Fig. 4**a**. Notice that the estimation method underestimates the high probability states, while overestimating those will small probability. **b)** The expected distortion (Hamming distance) and rate (entropy of Hopfield memories) for Hopfield networks trained on 3×3 binarized natural image patches, relative to rate-distortion function.
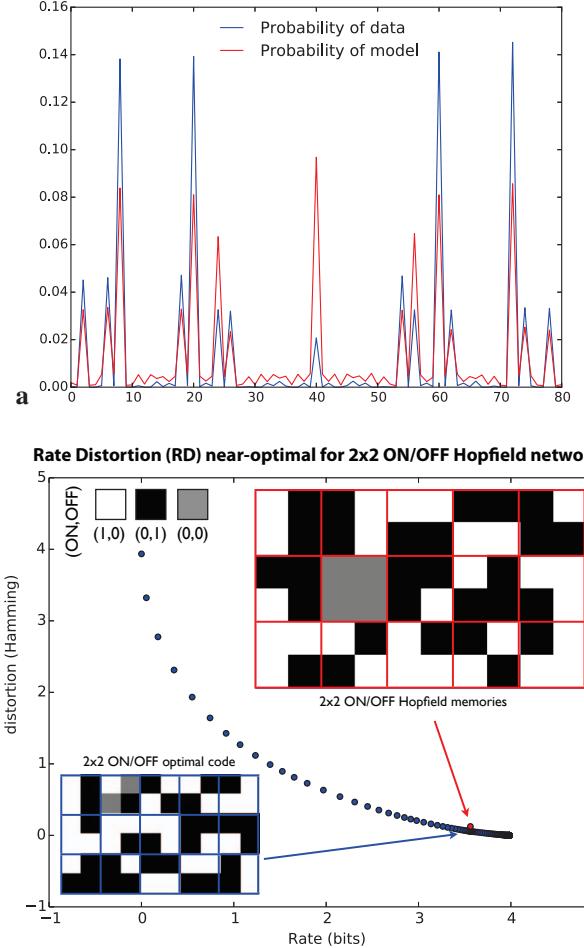
4. CONCLUSION

While being valuable as an abstract functional model of memory, Hopfield networks have long been dismissed as serious candidates for real-world applications. The main shortcomings that have prevented the practical use of such networks are threefold: First and foremost, the number of generic robust memories is limited to a (small) number of fixed-points linear in the number n of neurons (e.g. $0.15n$ using the outer-product learning rule). Secondly, the application of Hopfield networks to noisy data was limited, since the original learning rules require supervised learning with clean patterns, while more advanced statistical training methods involve computationally intensive iterations. Thirdly, Hopfield networks are thought to be weak models of data because many datasets have higher-order correlations that might not be captured by a second-order model (i.e., determined by its means and

covariances). Nonetheless, new probabilistically motivated techniques to fit Hopfield networks to natural data produce networks that can be exploited to efficiently compress images at high perceptual quality. Here, we have explored in more detail how such an approach can well-encode images and demonstrate its near-optimality in the 2×2 case. Our work opens new avenues for processing and understanding natural data using discrete recurrent neural network architectures.

5. REFERENCES

[1] C.E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, 1948.

[2] R. Kenneth, "A mapping approach to rate-distortion computation and analysis," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 1939–1952, 1994.

[3] F. Matejka and C. A. Sims, "Discrete actions in information-constrained tracking problems," *Princeton University manuscript*, 2010.

[4] C. Hillar, R. Mehta, and K. Koepsell, "A Hopfield recurrent neural network trained on natural images performs state-of-the-art image compression," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 4092–4096.

[5] W.S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *Bulletin of mathematical biology*, vol. 5, no. 4, pp. 115–133, 1943.

[6] G.K. Wallace, "The JPEG still picture compression standard," *Communications of the ACM*, vol. 34, no. 4, pp. 30–44, 1991.

[7] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[8] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.

[9] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *PNAS*, vol. 79, no. 8, pp. 2554–2558, 1982.

[10] J. Sohl-Dickstein, P.B. Battaglino, and M.R. DeWeese, "New method for parameter estimation in probabilistic models: minimum probability flow," *Physical review letters*, vol. 107, no. 22, pp. 220601, 2011.

[11] C. Hillar, J. Sohl-Dickstein, and K. Koepsell, "Efficient and optimal Little-Hopfield auto-associative memory storage using minimum probability flow," *NIPS (DISCML Workshop)*, 2012.

[12] J.H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," *Proceedings of the Royal Society of London. Series B: Biological Sciences*, vol. 265, no. 1394, pp. 359–366, 1998.

[13] M. Egmont-Petersen, D. de Ridder, and H. Handels, "Image processing with neural networks – a review," *Pattern recognition*, vol. 35, no. 10, pp. 2279–2301, 2002.

[14] N. Martine and T. Jean-Bernard, "Neural approach for TV image compression using a Hopfield type network," in *Advances in Neural Information Processing Systems 1*, D.S. Touretzky, Ed., pp. 264–271. 1989.

[15] J.-S. Lin and S.-H. Liu, "A competitive continuous Hopfield neural network for vector quantization in image compression," *Engineering Applications of Artificial Intelligence*, vol. 12, no. 2, pp. 111–118, 1999.

[16] E. P. Simoncelli and B. A. Olshausen, "Natural image statistics and neural representation," *Annual review of neuroscience*, vol. 24, no. 1, pp. 1193–1216, 2001.

[17] H.E. Gerhard, F.A. Wichmann, and M. Bethge, "How sensitive is the human visual system to the local statistics of natural images?," *PLoS computational biology*, vol. 9, no. 1, pp. e1002873, 2013.

[18] E. Ising, "Beitrag zur Theorie des Ferromagnetismus," *Zeitschrift fur Physik*, vol. 31, pp. 253–258, 1925.

[19] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *Information Theory, IEEE Transactions on*, vol. 18, no. 1, pp. 14–20, 1972.

[20] R. E. Blahut, "Computation of channel capacity and rate-distortion functions," *Information Theory, IEEE Transactions on*, vol. 18, no. 4, pp. 460–473, 1972.

[21] M. Bethge and P. Berens, "Near-maximum entropy models for binary neural representations of natural images," *Advances in neural information processing systems*, vol. 20, pp. 97–104, 2008.

[22] G.J. Stephens, T. Mora, G. Tkacik, and W. Bialek, "Statistical thermodynamics of natural images," *Physical review letters*, vol. 110, no. 1, pp. 018701, 2013.

[23] J.W. Pillow, J. Shlens, L. Paninski, A. Sher, A.M. Litke, E.J. Chichilnisky, and E.P. Simoncelli, "Spatio-temporal correlations and visual signalling in a complete neuronal population," *Nature*, vol. 454, no. 7207, pp. 995–999, 2008.