

Statistical Thermodynamics of Natural Images

Greg J. Stephens,^{*} Thierry Mora,[†] Gašper Tkačik,[‡] and William Bialek

*Joseph Henry Laboratories of Physics and Lewis-Sigler Institute for Integrative Genomics, Princeton University,
Princeton, New Jersey 08544, USA*

(Received 31 January 2012; published 2 January 2013)

The scale invariance of natural images suggests an analogy to the statistical mechanics of physical systems at a critical point. Here we examine the distribution of pixels in small image patches and show how to construct the corresponding thermodynamics. We find evidence for criticality in a diverging specific heat, which corresponds to large fluctuations in how “surprising” we find individual images, and in the quantitative form of the entropy vs energy. We identify special image configurations as local energy minima and show that average patches within each basin are interpretable as lines and edges in all orientations.

DOI: [10.1103/PhysRevLett.110.018701](https://doi.org/10.1103/PhysRevLett.110.018701)

PACS numbers: 89.75.Da, 89.20.-a

Providing a precise mathematical description of the structure in the visual world is a challenging problem. One approach is to seek a generative model of the probability distribution of images such that randomly drawn samples resemble images from the natural environment. Such a model would provide a rigorous basis for practical algorithms in image coding, processing, and recognition [1]. Previous work, for example, has applied techniques drawn from Ising models and statistical mechanics to problems such as image restoration [2]. It is also plausible that our brains have learned an approximation to this probabilistic model, allowing for efficient representations and effective solutions of many seemingly difficult computational problems. In this view, aspects of vision ranging from the responses of individual neurons to gestalt perceptual rules would be seen not as artifacts of the brain’s circuitry, but rather as matched to the statistical structure of the physical world [3–6].

One statistical feature of natural images that hints about the nature of the underlying probability distribution is scale invariance. In particular, Field observed that the spatial patterns of image intensity from natural environments have power spectra that approximate $S_I \propto 1/k^2$, which is consistent with the hypothesis of scale invariance and simple dimensional analysis, and he suggested a direct connection to the distribution of receptive field parameters across neurons in visual cortex [7]. The intuition that scale invariance is a strong constraint on the form of the probability distribution comes from statistical mechanics. For most physical systems at generic values of the temperature and other parameters, correlations and power spectra exhibit a characteristic length ξ beyond which structures

approach statistical independence. Scale invariance emerges only when we tune the temperature to a special value T_c , a critical point which marks a second order phase transition between two different phases such as ferromagnet and paramagnet [8]. In the modern theory of critical phenomena such scale invariance can occur while violating the naive expectations of dimensional analysis, so that power spectra can acquire “anomalous dimensions,” $S \propto 1/k^{2-\eta}$. Further, scaling extends beyond low order statistics, so that the full probability distributions are predicted to be invariant (but non-Gaussian) under appropriate scaling transformations. Both anomalous scaling and invariant non-Gaussian distributions for local features have been observed in an ensemble of natural scenes [9,10].

The analogy between scaling in natural images and physical systems at their critical point raises the question of whether there are analogs to the *thermodynamic* features of criticality. For example, can we generalize a given natural image ensemble to a family of ensembles indexed by a “temperature,” and show that there is something special (i.e., critical) about the temperature of the real ensemble? If there is an analog of the diverging specific heat at T_c , what does this say about the nature of images? What are the order parameters that characterize the underlying phase transition?

To address these questions we analyzed the image ensemble of Ref. [9] and focused on the 45 images with 256×256 pixel regions covering $\sim 15^\circ \times 15^\circ$ scenes in the woods of Hacklebarney State Park in New Jersey; an example is shown in Fig. 1(a). We sampled the distribution of images in small patches and we quantized the gray-scale images into two levels, with the quantization threshold chosen so that the numbers of black and white pixels are equal over the ensemble.

In the quantized images, an $L \times L$ pixel region can take on 2^{L^2} possible states, and our data set provides $\sim 3 \times 10^6$ samples of these states. Thus the distribution is well sampled for $L = \{2, 3, 4\}$. A direct estimate of the entropy

Published by the American Physical Society under the terms of the [Creative Commons Attribution 3.0 License](https://creativecommons.org/licenses/by/3.0/). Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

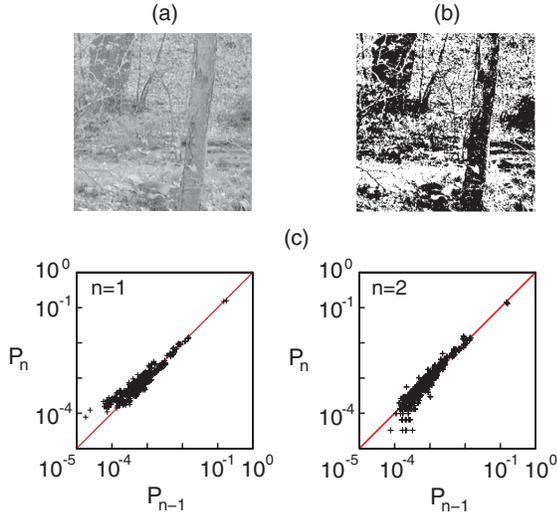


FIG. 1 (color). Natural images and patch scaling. (a) An example from the ensemble [9] and after quantizing into two equally-populated levels (b). Though most intensity information is lost, the image retains substantial structure. (c) The distribution of black and white pixels in 3×3 patches is invariant to block scaling. We quantify the difference in distributions through the Jensen-Shannon divergence D_{JS} and find $D_{JS}[P_1, P_0] = 8.5 \pm 0.1 \times 10^{-3}$ bits while $D_{JS}[P_2, P_1] = 3.3 \pm 0.2 \times 10^{-3}$ bits.

gives $S(4 \times 4) = 11.154 \pm 0.002$ bits, much less than the 16 bits from random pixels; similarly $S(3 \times 3) = 6.580 \pm 0.003 < 9$ bits. Thus, substantial local structure remains in the discretized patches.

We test for scaling by analyzing how the distribution of states in $L \times L$ patches evolves when we coarse grain the images. Following the approach for spin systems [11], we take the quantized image $\sigma(\vec{x})$ and create new quantized images by applying majority rule to the pixels in 3×3 blocks. This can be iterated and we denote the resulting distributions of states P_n , where P_0 is the distribution without blocking. Scale invariance implies that the P_n are the same, independent of n , because the distribution of states is at a fixed point of this renormalization transformation [8]. In Fig. 1(c) we test this prediction, showing that scaling is accurately obeyed over four decades in probability; small deviations are partly due to aliasing and diminish as we iterate away from single-pixel resolution. Importantly, this test of scale invariance involves the full patch distribution, and thus goes beyond the power law behavior of the spectrum (a second order moment) or the invariance of distributions of features (e.g., the outputs of local filters) evaluated at a single point. Similar results hold for 4×4 patches.

Each patch of our discrete images is described by a set of binary variables $\vec{\sigma}$. We imagine that the distribution $P(\vec{\sigma})$ is the Boltzmann distribution for a physical system at temperature $T = 1$, with an “energy” function $E(\vec{\sigma})$ describing each possible patch, $P(\vec{\sigma}) = \frac{1}{Z} e^{-E(\vec{\sigma})}$.

Following similar methods from dynamical systems [12,13], we define the distribution at any temperature T ,

$$P_T(\vec{\sigma}) \equiv \frac{1}{Z(T)} e^{-E(\vec{\sigma})/T} = \frac{1}{Z(T)} [P(\vec{\sigma})]^{1/T}, \quad (1)$$

where $Z(T) = \sum_{\vec{\sigma}} [P(\vec{\sigma})]^{1/T}$. The entropy is $S(T) = -\sum_{\vec{\sigma}} P_T(\vec{\sigma}) \log P_T(\vec{\sigma})$ and the heat capacity is given by the usual thermodynamic relation, $C(T) = T \partial S(T) / \partial T$. We also note that the heat capacity is proportional to the variance in energy or log probability, $C(T) = \langle [\delta E(\vec{\sigma})]^2 \rangle_T / T^2$, where $\langle \dots \rangle_T$ denotes an average in the distribution $P_T(\vec{\sigma})$.

In a system with a critical point, the specific heat diverges at T_c in the thermodynamic limit, corresponding here to patches with many pixels. Are there precursors of this divergence in the small patches that we can accurately sample? Figure 2(a) shows the specific heat for 2×2 , 3×3 , and 4×4 patches in our image ensemble, calculated directly from the distributions $P(\vec{\sigma})$. Even when normalized by the number of pixels $N = L^2$ in each patch (since the heat capacity is extensive), larger patches reveal a larger specific heat with a clear peak as a function of temperature, and this peak is shifting toward $T = 1$.

To calibrate our intuition about the specific heat estimated from small patches, we show analogous computations on the nearest neighbor ferromagnetic Ising model in two dimensions, defined by $E(\vec{\sigma}) = -J \sum_{\langle ij \rangle} \sigma_i \sigma_j$, where $\sum_{\langle ij \rangle}$ denotes a sum over neighboring pairs of pixels. Monte Carlo simulations of this model generate binary “images,” and so many of the practical sampling questions are very similar. In Fig. 2(b) the specific heat again shows a peak which grows and moves toward the true critical temperature $T_c = 1$ for larger patches [14]. Quantitatively the behavior is less dramatic than in the images, perhaps because the divergence of the specific heat in the

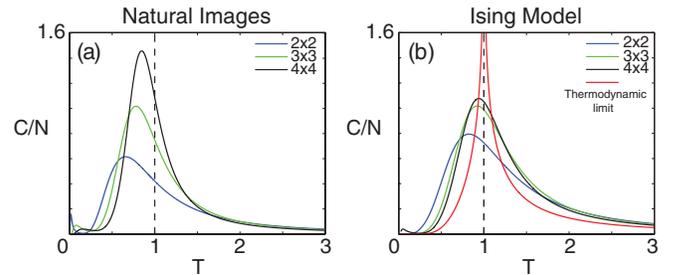


FIG. 2 (color). Diverging specific heats of natural images. (a) The specific heat C/N constructed from natural images for $L \times L$ patches of linear dimension $L = 2, 3, 4$ pixels. Away from the natural operating temperature ($T = 1$) the distribution is defined by Eq. (1). The peak in the specific heat near $T = 1$ suggests that natural images are drawn from a critical ensemble. (b) As in (a) but constructed from Monte Carlo simulations of an Ising model with $T_c = 1$; also shown is the exact behavior in the thermodynamic limit [28]. Even small patches display hints of underlying critical behavior.

thermodynamic limit is very gentle (logarithmic). Although this Ising system is much simpler than the ensemble of images, the similarities of Figs. 2(a) and 2(b) are consistent with an underlying divergence of the specific heat at a critical temperature near $T = 1$.

A complementary perspective on thermodynamics is the microcanonical ensemble, corresponding to fixed energy rather than fixed temperature. Recall that all thermodynamic quantities can be recovered from the partition function, $Z(T) = \sum_{\vec{\sigma}} e^{-E(\vec{\sigma})/T}$. We rewrite this sum by grouping together all states that have the same energy,

$$\begin{aligned} Z(T) &= \sum_{\vec{\sigma}} e^{-E(\vec{\sigma})/T} = \int dE \left[\sum_{\vec{\sigma}} \delta(E - E(\vec{\sigma})) \right] e^{-E/T} \\ &= \int dE \rho(E) e^{-E/T}, \end{aligned} \quad (2)$$

which defines the density of states $\rho(E)$. For a large system, the density of states becomes a smooth function, and we can define an entropy $S(E)$ at fixed energy as the log of the number of states in a narrow range of energies Δ , so that $\rho(E) = (1/\Delta)e^{S(E)}$ and $Z(T) = \frac{1}{\Delta} \int dE e^{S(E)-E/T}$. Further, both the energy and entropy are extensive variables proportional to the size of the system N , taken here to be the number of pixels in a patch. We define $\epsilon = E/N$ and $s(\epsilon) = S(E = N\epsilon)/N$, and the partition function becomes

$$Z(T) = \frac{N}{\Delta} \int d\epsilon e^{N[s(\epsilon) - \epsilon/T]}. \quad (3)$$

For large N , the integral is dominated by the point where the exponent is maximal, an energy such that $ds(\epsilon)/d\epsilon = dS(E)/dE = 1/T$. This connects the (microcanonical) description at fixed energy with the (canonical) description at fixed temperature. In this language, the specific heat, $C = \frac{N}{T^2} \left[-\frac{d^2s(\epsilon)}{d\epsilon^2} \right]^{-1}$ diverges where the second derivative of the entropy vs energy vanishes, the hallmark of a second order phase transition.

To apply the microcanonical framework to our image ensemble we define E for every observed state as the log of the probability,

$$E(\vec{\sigma}) = -\ln P(\vec{\sigma}) + c, \quad (4)$$

where c sets the (arbitrary) zero of energy and is chosen so that the most probable state has zero energy. We now define the cumulative distribution, $\mathcal{N}(E) = \int_0^E dE' \rho(E')$, which counts the number of possible image patches for which the observed log probability is greater than $-E + c$. If $S(E)$ is increasing, then this integral is dominated by the behavior near its upper limit, so that

$$\begin{aligned} \mathcal{N}(E) &= \frac{N}{\Delta} \int_0^{E/N} d\epsilon e^{Ns(\epsilon)} \approx \frac{N}{\Delta} \left[N \frac{ds(\epsilon)}{d\epsilon} \right]^{-1} e^{Ns(\epsilon=E/N)} \\ \Rightarrow s(\epsilon) &= \frac{1}{N} \ln \mathcal{N}(E = N\epsilon) + \frac{\ln(T/\Delta)}{N}. \end{aligned} \quad (5)$$

Note that the second term in this equation vanishes for large N , and so we approximate the entropy per pixel as a function of energy per pixel by the first term.

By counting the number of possible image patches with probability greater than a certain level, we can use Eq. (5) to construct the entropy vs energy and hence derive all other thermodynamic functions. We do this in Fig. 3(a) using rectangular $L \times L'$ patches of size from 8 pixels up to 50 pixels; in Fig. 3(b) we show the same analysis for the Monte Carlo simulations of the Ising model. We note that sampling problems are less serious at low energies (states with high probability), so we expect that even if we look at regions where we can't sample the whole distribution we will get the correct low energy behavior.

The results on different sized image patches are remarkably consistent, which is indicative of the thermodynamic limit despite the small region sizes. Since the real image ensemble is at $T = 1$, we seek the energy at which $dS/dE = 1$ which is difficult, since the plot is very nearly linear with unit slope. However, if the point where $dS/dE = 1$ is also a place where $d^2S/dE^2 = 0$, then $T = 1$ is a critical point. The fact that S/N vs E/N is linear with unit slope is direct evidence that the ensemble of natural images is at criticality.

Our approach to the thermodynamics of images is connected to Zipf's law [15] for which the probability of the rank order r th most common word drawn from English text is $p_r \propto 1/r$. Many other authors have considered generalized Zipf-like distributions [16], $p_r \propto 1/r^\alpha$, and there has been much discussion about the meaning of these relationships. Suppose we identify the Zipf-like distribution $p_r = A/r^\alpha$ with a Boltzmann distribution at $T = 1$, $p_r = (1/Z)e^{-E_r}$. Then the energy of the state at rank r is $E_r = \alpha \ln r - \ln(AZ)$. In the limit of a large system with many possible states, r has a uniform distribution, and hence $\rho(E) \approx |dE_r/dr|^{-1}$; this gives $\rho(E) = r/\alpha$. But we also have $r = (AZ)^{1/\alpha} e^{E_r/\alpha}$, so we find

$$\rho(E) = \frac{1}{\alpha} (AZ)^{1/\alpha} e^{E/\alpha} \Rightarrow S_{\text{Zipf}}(E) = E/\alpha + \text{const.}$$

Thus a generalized Zipf's law is equivalent to an entropy that is (exactly) linear in the energy. The original Zipf's

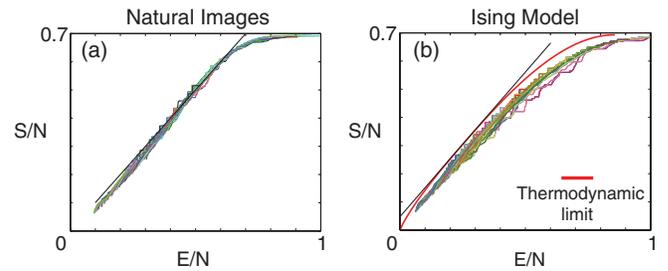


FIG. 3 (color). Indication of critical behavior in the plot of entropy vs energy. (a) The results for rectangular image patches of size 8 to 50 pixels (different colors). (b) Results for Monte Carlo simulations of the 2D Ising model.

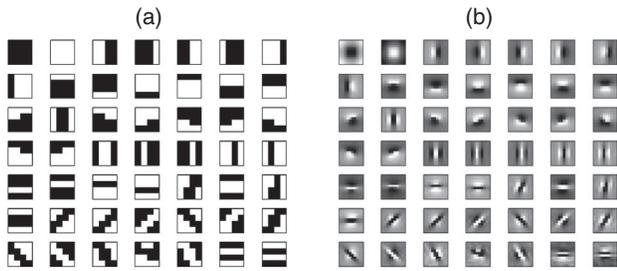


FIG. 4. Locally stable states. (a) The 49 most probable patches such that all single pixel inversions decrease their probability in the ensemble. Even in small (4×4) patches these locally stable states are interpretable as lines and edges at all positions and orientations. (b) The average surrounding 10×10 light-intensity images leading to these metastable states. These images resemble the average images that trigger responses of neurons in primary visual cortex.

law ($\alpha = 1$) corresponds to a unit slope, as we have found for image patches. Further, this simple linear relation corresponds exactly to a critical point [17,18].

Since the specific heat is the variance in the energy, criticality in the image ensemble means that the log probability has a broad distribution, with a formally divergent second moment. One consequence is that the approach toward typicality in the sense of information theory [19] is slow, which may be related to difficulties in compressing large natural images, or even in estimating their entropy (e.g., Ref. [20]). This large variance also implies large fluctuations in how surprised we should be by any given scene or segment of a scene, a relatively common experience.

Critical points mark the transition between phases characterized by different forms of order. What ordering would emerge if the distribution of natural images could be “cooled” from $T = 1$ toward $T = 0$? Ultimately, this is a question about the nature of the image patches that correspond to the low energy states. The lowest energy states of small patches are solid black or white blocks, as in a ferromagnet with aligned spins. But, searching through all 4×4 patches, we find ~ 100 states that are local minima, such that flipping any single pixel results in increased energy or reduced probability. In Fig. 4(a) we show 49 of these states, ordered in decreasing probability. We see that many of these states are interpretable, for example as edges between dark and light regions, and that much of the multiplicity arises from the different ways of realizing these patterns (e.g., the six possible cases of a single vertical edge). We can think of these local minima in the energy landscape [21] as being like the attractors in the Hopfield model of neural networks [22], or like the codewords in statistical mechanics approaches to error-correcting codes [23]. While we usually think of error-correcting coding as a construct, here it seems intrinsic to the natural signals (see Ref. [24] for a discussion of similar questions in neural systems).

We speculate that such image features acquire their importance due to their intrinsic properties of error correction. If this is true, then the visual system might build neurons which serve to identify the basins of attraction defined by these local minima in energy. If such cells respond only when the original gray-scale image corresponds to an image within a particular basin of attraction, then it is easy to compute the response-triggered average within our natural image ensemble, with the results shown in Fig. 4(b). These results have a strong resemblance to the spike-triggered average responses of neurons in visual cortex to natural scenes [25]. In more detail we find that perceptron-like models based on filtering through a single receptive field do rather poorly. Thus, if visual cortex builds a representation of the world based on the identification of these local minima, the computations necessarily involve nonlinear combinations of multiple filters [26], as observed [27].

G. J. S. and T. M. contributed equally. We thank D. Chigirev, S. E. Palmer, and E. Schneidman for helpful discussions, and D. L. Ruderman for recovering the original data from Ref. [9]. This work was supported in part by NSF Grants No. IIS-0613435, No. IBN-0344678, and No. PHY-0957573, by NIH Grant No. T32 MH065214, by the Human Frontier Science Program, and by the Swartz Foundation.

*Present address: Department of Physics and Astronomy, Vrije Universiteit Amsterdam, de Boelelaan 1081, 1081 HV Amsterdam, The Netherlands and Okinawa Institute of Science and Technology, 1919-1 Tancha, Onna-son, Okinawa, Japan.

[†]Present address: Laboratoire de physique statistique, UMR8550, CNRS and École normale supérieure, 24, rue Lhomond, 75005 Paris, France.

[‡]Present address: Institute of Science and Technology Austria, Am Campus 1, A-3400 Klosterneuburg, Austria.

- [1] E. R. Davies, *Machine Vision: Theory, Algorithms, Practicalities* (Morgan Kaufmann, San Francisco, 2005).
- [2] S. German and D. German, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6**, 721 (1984).
- [3] H. B. Barlow, in *Sensory Communication*, edited by W. Rosenblith (MIT Press, Cambridge, 1961), pp. 217–234.
- [4] J. J. Atick, *Network* **3**, 213 (1992).
- [5] E. Simoncelli and B. Olshausen, *Annu. Rev. Neurosci.* **24**, 1193 (2001).
- [6] W. Bialek, in *Physics of Biomolecules and Cells*, Proceedings of the Les Houches Summer School, Session LXXV, edited by H. Flyvbjerg, F. Jülicher, P. Ormos, and F. David (Springer-Verlag, Berlin, 2002), pp. 485–577; [arXiv:physics/0205030](https://arxiv.org/abs/physics/0205030).
- [7] D. J. Field, *J. Opt. Soc. Am. A* **4**, 2379 (1987).
- [8] J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, Cambridge, 1996).
- [9] D. L. Ruderman and W. Bialek, *Phys. Rev. Lett.* **73**, 814 (1994).

- [10] D.L. Ruderman, *Network* **5**, 517 (1994).
- [11] L.P. Kadanoff, *Physics* (Long Island City, N.Y.) **2**, 263 (1966).
- [12] M.J. Feigenbaum, M.H. Jensen, and I. Procaccia, *Phys. Rev. Lett.* **57**, 1503 (1986); M.J. Feigenbaum, *J. Stat. Phys.* **46**, 919 (1987).
- [13] C. Beck and F. Schlögl, *Thermodynamics of Chaotic Systems* (Cambridge University Press, New York, 1993).
- [14] A.E. Ferdinand and M.E. Fisher, *Phys. Rev.* **185**, 832 (1969).
- [15] G.K. Zipf, *Selected Studies of the Principle of Relative Frequency in Language* (Harvard University Press, Cambridge, 1932).
- [16] For a review, see M.E.J. Newman, *Contemp. Phys.* **46**, 323 (2005).
- [17] T. Mora and W. Bialek, *J. Stat. Phys.* **144**, 268 (2011).
- [18] Interestingly, if $\alpha \neq 1$, then there is no correspondence with thermodynamics.
- [19] C.E. Shannon, *Bell Syst. Tech. J.* **27**, 379 (1948); **27**, 623 (1948).
- [20] D.M. Chandler and D.J. Field, *J. Opt. Soc. Am. A* **24**, 922 (2007).
- [21] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
- [22] J.J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554 (1982).
- [23] N. Sourlas, *Nature (London)* **339**, 693 (1989).
- [24] G. Tkačik, E. Schneidman, M.J. Berry II, and W. Bialek, [arXiv:0912.5409](https://arxiv.org/abs/0912.5409).
- [25] D. Smyth, B. Willmore, G.E. Baker, I.D. Thompson, and D.J. Tolhurst, *J. Neurosci.* **23**, 4746 (2003); F.E. Theunissen, S.V. David, N.C. Singh, A. Hsu, W.E. Vinje, and J.L. Gallant, *Netw., Comput. Neural Syst.* **12**, 289 (2001).
- [26] W. Bialek and R. de Ruyter van Steveninck, [arXiv:q-bio/0505003](https://arxiv.org/abs/q-bio/0505003).
- [27] N.C. Rust, O. Schwartz, J.A. Movshon, and E.P. Simoncelli, *Neuron* **46**, 945 (2005).
- [28] L. Onsager, *Phys. Rev.* **65**, 117 (1944).