Statistical-Physics-Based Reconstruction in Compressed Sensing

F. Krzakala,[1,*] M. Mézard,[2] F. Sausset,[2] Y. F. Sun,[1,3] and L. Zdeborová[4]

[1]*CNRS and ESPCI ParisTech, 10 rue Vauquelin, UMR 7083 Gulliver, Paris 75005, France*
[2]*Université Paris-Sud and CNRS, LPTMS, UMR8626, Bâtiment 100, 91405 Orsay, France*
[3]*LMIB and School of Mathematics and Systems Science, Beihang University, 100191 Beijing, China*
[4]*Institut de Physique Théorique (IPhT), CEA Saclay, and URA 2306, CNRS, 91191 Gif-sur-Yvette, France*

Compressed sensing has triggered a major evolution in signal acquisition. It consists of sampling a sparse signal at low rate and later using computational power for the exact reconstruction of the signal, so that only the necessary information is measured. Current reconstruction techniques are limited, however, to acquisition rates larger than the true density of the signal. We design a new procedure that is able to reconstruct the signal exactly with a number of measurements that approaches the theoretical limit, i.e., the number of nonzero components of the signal, in the limit of large systems. The design is based on the joint use of three essential ingredients: a probabilistic approach to signal reconstruction, a message-passing algorithm adapted from belief propagation, and a careful design of the measurement matrix inspired by the theory of crystal nucleation. The performance of this new algorithm is analyzed by statistical-physics methods. The obtained improvement is confirmed by numerical studies of several cases.

Subject Areas: Complex Systems, Computational Physics, Statistical Physics

I. INTRODUCTION

The ability to recover high-dimensional signals using only a limited number of measurements is crucial in many fields, ranging from image processing to astronomy or systems biology. Examples of direct applications include speeding up magnetic resonance imaging without the loss of resolution, sensing and compressing data simultaneously [1], and the single-pixel camera [2]. Compressed sensing is designed to directly acquire *only* the necessary information about the signal. This is possible whenever the signal is compressible, that is, sparse in some known basis. In a *second step* one uses computational power to reconstruct the signal exactly [1,3,4]. Currently, the best known generic method for exact reconstruction is based on converting the reconstruction problem into one of convex optimization, which can be solved efficiently using linear programming techniques [3,4]. This so-called ℓ_1 reconstruction is able to reconstruct accurately, provided the system size is large and the ratio of the number of measurements M to the number of nonzero elements K exceeds a specific limit which can be proven by careful analysis [4,5]. However, the limiting ratio is significantly larger than 1. In this paper, we improve on the performance of ℓ_1 reconstruction, and in the best possible way: We introduce a new procedure that is able to reach the optimal limit $M/K \to 1$.

*Corresponding author;
fk@espci.fr

This procedure, which we call *seeded belief propagation* (s-BP), is based on a new, carefully designed, measurement matrix. It is very powerful, as illustrated in Fig. 1. Its performance will be studied here with a joint use of numerical and analytic studies using methods from statistical physics [7].

II. RECONSTRUCTION IN COMPRESSED SENSING

The mathematical problem posed in compressed-sensing reconstruction is easily stated. Given an unknown signal which is an N-dimensional vector $\mathbf{s}$, we make M measurements, where each measurement amounts to a projection of $\mathbf{s}$ onto some known vector. The measurements are grouped into an M-component vector $\mathbf{y}$, which is obtained from $\mathbf{s}$ by a linear transformation $\mathbf{y} = \mathbf{Fs}$. Depending on the application, this linear transformation can be, for instance, associated with measurements of Fourier modes or wavelet coefficients. The observer knows the $M \times N$ matrix $\mathbf{F}$ and the M measurements $\mathbf{y}$, with $M < N$. His aim is to reconstruct $\mathbf{s}$. This is impossible in general, but compressed sensing deals with the case where the signal $\mathbf{s}$ is sparse, in the sense that only $K < N$ of its components are nonzero. We shall study the case where the nonzero components are real numbers and the measurements are linearly independent. In this case, exact signal reconstruction is possible in principle whenever $M \geq K + 1$, using an exhaustive enumeration method which tries to solve $\mathbf{y} = \mathbf{Fx}$ for all $\binom{N}{K}$ possible choices of locations of nonzero components of $\mathbf{x}$: Only one such choice gives a consistent linear system, which can then be inverted. However, one is typically interested in instances of large-size signals where $N \gg 1$, with
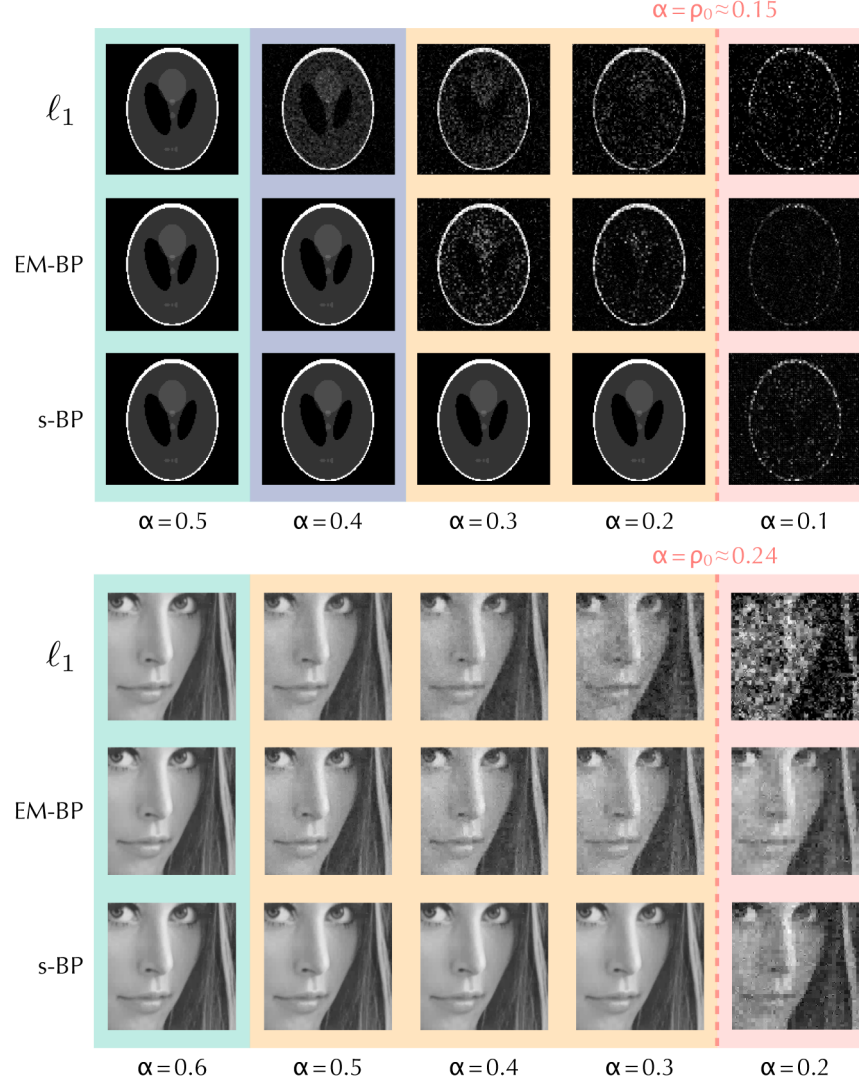
FIG. 1.   Two illustrative examples of compressed sensing in image processing. Top: The original image, known as the Shepp-Logan phantom, of size $N = 128^2$, is transformed via one step of Haar wavelets into a signal of density $\rho_0 \approx 0.15$. With compressed sensing, one is thus in principle able to reconstruct the image exactly with $M \geq \rho_0 N$ measurements, but practical reconstruction algorithms generally need M larger than $\rho_0 N$. The five columns show the reconstructed figure obtained from $M = \alpha N$ measurements, with decreasing acquisition rate α. The first row is obtained with the ℓ_1-reconstruction algorithm [3,4] for a measurement matrix with independent identically distributed (iid) elements of zero mean and variance $1/N$. The second row is obtained with belief propagation, using exactly the same measurements as used in the ℓ_1 reconstruction. The third row is the result of the seeded belief propagation, introduced in this work, which uses a measurement matrix based on a chain of coupled blocks (see Fig. 4). The running time of all three algorithms is comparable. (Asymptotically, they are all quadratic in the limit of the large signal size.) In the lower set of images, we took as the sparse signal the relevant coefficients after two-step Haar transform of the picture of Lena. Again for this signal, with density $\rho_0 = 0.24$, the s-BP procedure reconstructs exactly down to very low measurement rates. (Details are in Appendix G; the data is available online [6].)

$M = \alpha N$ and $K = \rho_0 N$. The enumeration method solves the compressed-sensing problem in the regime where measurement rates, α are at least as large as the signal density, $\alpha \geq \rho_0$, but in a time which grows exponentially with N, making it totally impractical. Therefore, $\alpha = \rho_0$ is the fundamental reconstruction limit for perfect reconstruction in the noiseless case, when the nonzero components of the signal are real numbers drawn from a continuous distribution. A general and detailed discussion of information-theoretically optimal reconstruction has been developed recently in [8–10].

In order to design practical "low-complexity" reconstruction algorithms, it has been proposed [3,4] to find a vector **x** which has the smallest ℓ_1 norm, $\sum_{i=1}^{N} |x_i|$, within the subspace of vectors which satisfy the constraints $\mathbf{y} = \mathbf{Fx}$, using efficient linear-programming techniques.

In order to measure the performance of this strategy, one can focus on the measurement matrix $\mathbf{F}$ generated randomly, with independent Gaussian-distributed matrix elements of mean zero and variance $1/N$, and the signal vector $\mathbf{s}$ having density $0 < \rho_0 < 1$. The analytic study of the ℓ_1 reconstruction in the thermodynamic limit $N \to \infty$ can be done using either geometric or probabilistic methods [5,11], or with the replica method [12–14]. The analysis shows the existence of a sharp phase transition at a value $\alpha_{\ell_1}(\rho_0)$. When α is larger than this value, the ℓ_1 reconstruction gives the exact result $\mathbf{x} = \mathbf{s}$ with probability going to 1 in the large N limit; when $\alpha < \alpha_{\ell_1}(\rho_0)$, the probability that it gives the exact result goes to zero as N grows. As shown in Fig. 2, $\alpha_{\ell_1}(\rho_0) > \rho_0$ and therefore the ℓ_1 reconstruction is suboptimal: It requires more measurements than would be absolutely necessary, in the sense that, if one were willing to do brute-force combinatorial optimization, no more than $\rho_0 N$ measurements would be necessary.

We introduce a new measurement and reconstruction approach, designated the seeded belief propagation (s-BP), that allows one to reconstruct the signal by a practical method that needs only $\approx \rho_0 N$ measurements. We shall now discuss s-BP's three ingredients: (1) a probabilistic approach to signal reconstruction, (2) a message-passing algorithm adapted from belief propagation [15], which is a procedure known to be efficient in various hard computational problems [16,17], and (3) an

innovative design of the measurement matrix inspired by the theory of crystal nucleation in statistical physics and from recent developments in coding theory [18–21]. Some previous works on compressed sensing have used these ingredients separately. In particular, adaptations of belief propagation have been developed for the compressed-sensing reconstruction, both in the context of ℓ_1 reconstruction [11,22,23] and in a probabilistic approach [24]. It is the combined use of these three ingredients that allows us to reach the $\alpha = \rho_0$ limit.

III. A PROBABILISTIC APPROACH

For the purpose of our analysis, we consider the case where the signal $\mathbf{s}$ has independent identically distributed (iid) components: $P_0(\mathbf{s}) = \prod_{i=1}^{N}[(1 - \rho_0)\delta(s_i) + \rho_0\phi_0(s_i)]$, with $0 < \rho_0 < 1$. In the large-N limit, the number of nonzero components is $\rho_0 N$. Our approach handles general distributions $\phi_0(s_i)$.

Instead of using a minimization procedure, we shall adopt a probabilistic approach. We introduce a probability measure $\hat{P}(\mathbf{x})$ over vectors $\mathbf{x} \in \mathbb{R}^N$ which is the restriction of the Gauss-Bernoulli measure $P(\mathbf{x}) = \prod_{i=1}^{N}[(1 - \rho) \times \delta(x_i) + \rho\phi(x_i)]$ to the subspace $|\mathbf{y} - \mathbf{Fx}| = 0$ [25]. In this paper, we use a distribution $\phi(x)$, which is a Gaussian with mean $\bar{x}$ and variance σ^2, but other choices for $\phi(x)$ are possible. It is crucial to note that we do not require *a priori* knowledge of the statistical properties of
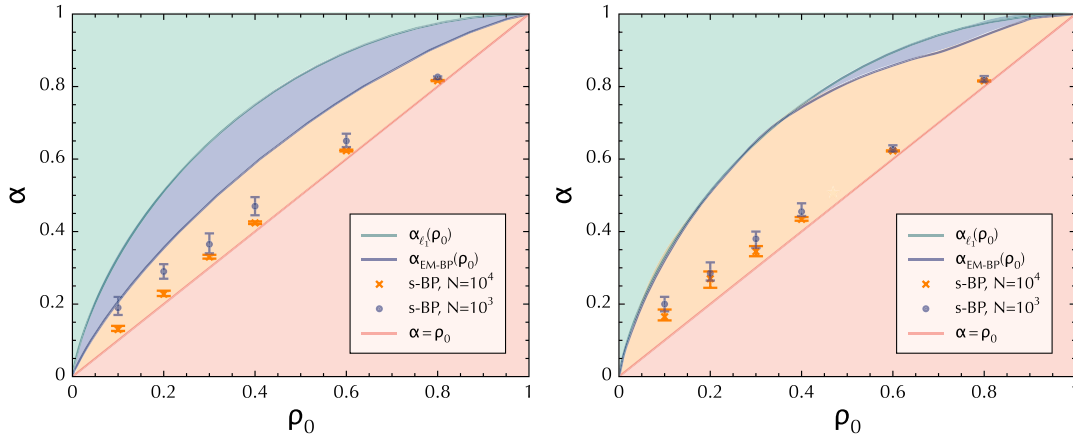

FIG. 2. Phase diagrams for compressed-sensing reconstruction for two different signal distributions. On the left, the $\rho_0 N$ nonzero components of the signal are independent Gaussian random variables with zero mean and unit variance. On the right, they are independent ± 1 variables. The measurement rate is $\alpha = M/N$. On both sides, we show, from top to bottom: (a) The phase transition α_{ℓ_1} for ℓ_1 reconstruction [5,11,12] (which does not depend on the signal distribution). (b) The phase transition $\alpha_{\text{EM-BP}}$ for EM-BP reconstruction based for both sides on the probabilistic model with Gaussian ϕ. (c) The data points which are numerical-reconstruction thresholds obtained with the s-BP procedure with $L = 20$. The point gives the value of α where exact reconstruction was obtained in 50% of the tested samples; the top of the error bar corresponds to a success rate of 90%; the bottom of the bar to a success of 10%. The shrinking of the error bar with increasing N gives numerical support to the existence of the phase transition that we have studied analytically. These empirical reconstruction thresholds of s-BP are quite close to the $\alpha = \rho_0$ optimal line, and they move closer to it with increasing N. The parameters used in these numerical experiments are detailed in Appendix E. (d) The line $\alpha = \rho_0$ that is the theoretical reconstruction limit for signals with continuous ϕ_0. An alternative presentation of the same data using the convention of Donoho and Tanner [5] is shown in Appendix F.

the signal: We use a value of ρ not necessarily equal to ρ_0, and the ϕ that we use is not necessarily equal to ϕ_0. The important point is to use $\rho < 1$ (which reflects the fact that one searches for a sparse signal).

Assuming that **F** is a random matrix, either where all the elements are drawn as independent Gaussian random variables with zero mean and the same variance or are of the carefully designed type of "seeding matrices" described below, we demonstrate in Appendix A that, for any ρ_0-dense original signal **s** and any $\alpha > \rho_0$, the probability $\hat{P}(\mathbf{s})$ of the original signal goes to 1 when $N \to \infty$. This result holds independent of the distribution ϕ_0 of the original signal, which does not need to be known. In practice, we see that **s** also dominates the measure when N is not very large. In principle, sampling configurations **x** proportionally to the restricted Gauss-Bernoulli measure $\hat{P}(\mathbf{x})$ thus gives asymptotically the exact reconstruction in the whole region $\alpha > \rho_0$. This idea stands at the root of our approach and is at the origin of the connection with statistical physics (where one samples with the Boltzmann measure).

IV. SAMPLING WITH EXPECTATION-MAXIMIZATION BELIEF PROPAGATION

The exact sampling from a distribution such as $\hat{P}(\mathbf{x})$, Eq. (A2), is known to be computationally intractable [26]. However, an efficient approximate sampling can be performed using a message-passing procedure that we now describe [24,27–29]. We start from the general belief-propagation formalism [17,30,31]: For each measurement $\mu = 1, \ldots, M$ and each signal component $i = 1, \ldots, N$, one introduces a "message," $m_{i \to \mu}(x_i)$, which is the probability of x_i in a modified measure where measurement μ has been erased. In the present case, the canonical belief-propagation equations relating these messages can be simplified [11,22–24,32] into a closed form that uses only the expectation $a_{i \to \mu}^{(t)}$ and the variance $v_{i \to \mu}^{(t)}$ of the distribution $m_{i \to \mu}^{(t)}(x_i)$ (see Appendix B). An important ingredient that we add to this approach is the learning of the parameters in $P(\mathbf{x})$: the density ρ, and the mean $\bar{x}$ and variance σ^2 of the Gaussian distribution $\phi(x)$. These are three parameters to be learned using update equations based on the gradient of the so-called Bethe free entropy, in a way analogous to the expectation maximization [33–35]. This leads to the expectation-maximization-belief-propagation (EM-BP) algorithm that we will use in the following discussion for reconstruction in compressed sensing. It consists of iterating the messages and the three parameters, starting from random messages $a_{i \to \mu}^{(0)}$ and $v_{i \to \mu}^{(0)}$, until a fixed point is obtained. Perfect reconstruction is found when the messages converge to the fixed point where $a_{i \to \mu} = s_i$ and $v_{i \to \mu} = 0$.

Like the ℓ_1 reconstruction, the EM-BP reconstruction also has a phase transition. Perfect reconstruction is

achieved with probability 1 in the large-N limit if and only if $\alpha > \alpha_{\text{EM-BP}}$. Using the asymptotic replica analysis, as explained below, we have computed the line $\alpha_{\text{EM-BP}}(\rho_0)$ when the elements of the $M \times N$ measurement matrix **F** are independent Gaussian random variables with zero mean and variance $1/N$ and the signal components are iid. The location of this transition line does depend on the signal distribution (see Fig. 2), contrary to the location of the ℓ_1 phase transition.

Notice that our analysis is fully based on the case when the probabilistic model has a Gaussian ϕ. Not surprisingly, EM-BP performs better when $\phi = \phi_0$ (see the left panel of Fig. 2), where EM-BP provides a sizable improvement over ℓ_1. In contrast, the right panel of Fig. 2 shows an adversary case when we use a Gaussian ϕ to reconstruct a binary ϕ_0; in this case, there is nearly no improvement over the ℓ_1 reconstruction.

V. DESIGNING SEEDING MATRICES

In order for the EM-BP message-passing algorithm to be able to reconstruct the signal down to the theoretically optimal number of measurements $\alpha = \rho_0$, one needs to use a special family of measurement matrices **F** that we call "seeding matrices." If one uses an unstructured **F**, for instance, a matrix with independent Gaussian-distributed random elements, EM-BP samples correctly at large α. At small enough α, however, a metastable state appears in the measure $\hat{P}(\mathbf{x})$. The EM-BP algorithm is trapped in this state and is therefore unable to find the original signal (see Fig. 3), just as a supercooled liquid gets trapped in a glassy state instead of crystallizing. It is well known in crystallization theory that the crucial step is to nucleate a large enough seed of crystal. This is the purpose of the following design of **F**.

We divide the N variables into L groups of N/L variables, and the M measurements into L groups. The number of measurements in the p-th group is $M_p = \alpha_p N/L$, so that $M = [(1/L) \sum_{p=1}^{L} \alpha_p] N = \alpha N$. We then choose the matrix elements $F_{\mu i}$ independently in such a way that, if i belongs to group p and μ to group q, then $F_{\mu i}$ is a random number chosen from the normal distribution with mean zero and variance $J_{q,p}/N$ (see Fig. 4). The matrix $J_{q,p}$ is an $L \times L$ coupling matrix (and the standard compressed-sensing matrices are obtained using $L = 1$ and $\alpha_1 = \alpha$). Using these new matrices, one can shift the BP phase transition very close to the theoretical limit. In order to get an efficient reconstruction with message passing, one should use a large enough α_1. With a good choice of the coupling matrix $J_{p,q}$, the reconstruction first takes place in the first block and then propagates as a wave in the following blocks
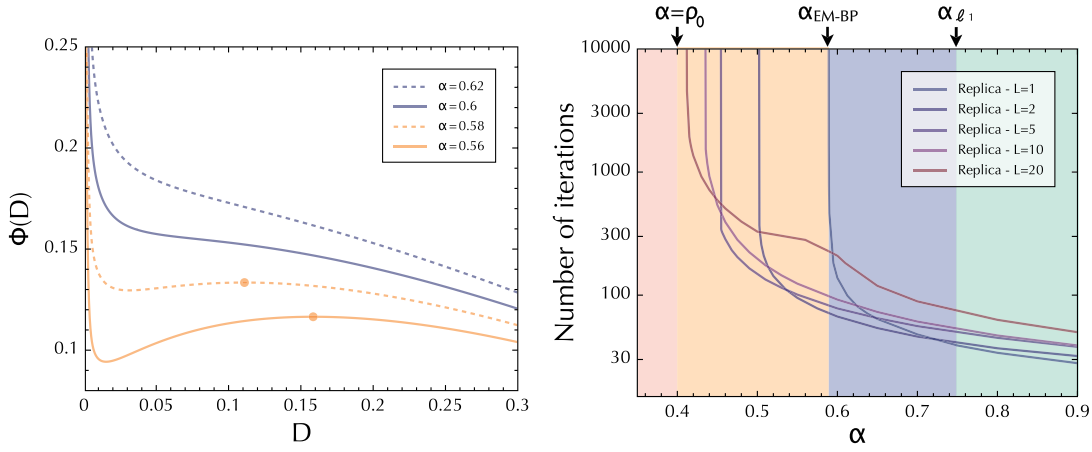
FIG. 3. When sampling from the probability $\hat{P}(\mathbf{x})$, in the limit of large N, the probability that the reconstructed signal x is at a squared distance $D = \sum_i (x_i - s_i)^2/N$ from the original signal s is written as $ce^{N\Phi(D)}$, where c is a constant and $\Phi(D)$ is the free entropy. Left: $\Phi(D)$ for a Gauss-Bernoulli signal with $\rho_0 = \rho = 0.4$ in the case of an unstructured measurement matrix $\mathbf{F}$ with independent random Gaussian-distributed elements. The evolution of the EM-BP algorithm is basically a steepest ascent in $\Phi(D)$ starting from large values of D. It goes to the correct maximum at $D = 0$ for large values of α, but it is blocked in the local maximum that appears for $\alpha < \alpha_{\text{EM-BP}}(\rho_0 = 0.4) \approx 0.59$. For $\alpha < \rho_0$, the maximum is not at $D = 0$, and exact inference is impossible. The seeding matrix $\mathbf{F}$, leading to the s-BP algorithm, succeeds in eliminating this local maximum. Right: Convergence time of the EM-BP and s-BP algorithms obtained through the replica analysis for $\rho_0 = 0.4$. The EM-BP convergence time diverges as $\alpha \rightarrow \alpha_{\text{EM-BP}}$ with the standard $L = 1$ matrices. The s-BP strategy allows one to go beyond the threshold: Using $\alpha_1 = 0.7$ and increasing the structure of the seeding matrix (here, $L = 2, 5, 10, 20$), we approach the limit $\alpha = \rho_0$. (Details of the parameters are given in Appendix E.)

$p = 2, 3, \ldots$, even if their measurement rate α_p is small. In practice, we use $\alpha_2 = \ldots = \alpha_L = \alpha'$, so that the total measurement rate is $\alpha = [\alpha_1 + (L - 1)\alpha']/L$. The whole reconstruction process is then analogous to crystal nucleation, where a crystal is growing from its seed (see Fig. 5). Similar ideas have been used recently in the design of sparse coding matrices for error-correcting codes [18–21].
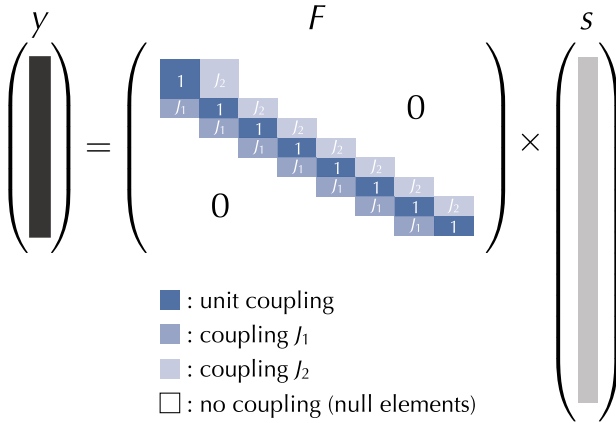

FIG. 4. Construction of the measurement matrix $\mathbf{F}$ for seeded compressed sensing. The elements of the signal vector are split into L (here, $L = 8$) equal-sized blocks; the number of measurements in each block is $M_p = \alpha_p N/L$ (here, $\alpha_1 = 1$, $\alpha_p = 0.5$ for $p = 2, \ldots, 8$). The matrix elements $F_{\mu i}$ are chosen as random Gaussian variables with variance $J_{q,p}/N$ if variable i is in the block p and measurement μ is in the block q. In the s-BP algorithm, we use $J_{p,q} = 0$, except for $J_{p,p} = 1$, $J_{p,p-1} = J_1$, and $J_{p-1,p} = J_2$. Good performance is typically obtained with relatively large J_1 and small J_2.
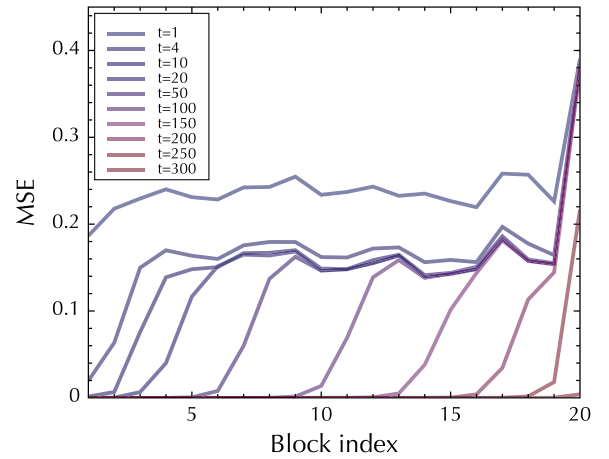

FIG. 5. Evolution of the mean-squared error at different times as a function of the block index for the s-BP algorithm. Exact reconstruction first appears in the left block whose rate $\alpha_1 > \alpha_{\text{EM-BP}}$ allows for seeded nucleation. It then propagates gradually block by block, driven by the free-entropy difference between the metastable and equilibrium states. After little more than 300 iterations, the whole signal of density $\rho_0 = 0.4$ and size $N = 50\,000$ is exactly reconstructed well inside the zone forbidden for BP (see Fig. 2). Here, we used $L = 20$, $J_1 = 20$, $J_2 = 0.2$, $\alpha_1 = 1.0$, $\alpha = 0.5$.

VI. ANALYSIS OF THE PERFORMANCE OF THE SEEDED BELIEF PROPAGATION PROCEDURE

The s-BP procedure is based on the joint use of seeding-measurement matrices and of the EM-BP message-passing reconstruction. We have studied it with two methods: direct numerical simulations and analysis of the performance in the large-N limit. The analytical result is obtained by a combination of the "replica" method and the "cavity" method (also known as "density evolution" or "state evolution"). The replica method is a standard method in statistical physics [7], which has been applied successfully to several problems of information theory [17,28,36,37], including compressed sensing [12–14]. It can be used to compute the free-entropy function Φ associated with the probability $\hat{P}(\mathbf{x})$ (see Appendix D). The cavity method shows that the dynamics of the message-passing algorithm is a gradient dynamics leading to a maximum of this free entropy.

When applied to the usual case of the full $\mathbf{F}$ matrix with independent Gaussian-distributed elements (case $L = 1$), the replica computation shows that the free-entropy function $\Phi(D)$ for configurations constrained to be at a mean-squared distance D has a global maximum at $D = 0$ when $\alpha > \rho_0$, which confirms that the Gauss-Bernoulli probabilistic reconstruction is in principle able to reach the optimal compression limit $\alpha = \rho_0$. However, for $\alpha_{\text{EM-BP}} > \alpha > \rho_0$, where $\alpha_{\text{EM-BP}}$ is a threshold that depends on the signal and on the distribution $P(\mathbf{x})$, a secondary local maximum of $\Phi(D)$ appears at $D > 0$ (see Fig. 3). In this case, the EM-BP algorithm converges instead to this secondary maximum and does not reach exact reconstruction. The threshold $\alpha_{\text{EM-BP}}$ is obtained analytically as the smallest value of α such that $\Phi(D)$ is decreasing (Fig. 2). This theoretical study has been confirmed by numerical measurements of the number of iterations needed for EM-BP to reach its fixed point (within a given accuracy). This convergence time of EM-BP to the exact reconstruction of the signal diverges when $\alpha \to \alpha_{\text{EM-BP}}$ (see Fig. 3). For $\alpha < \alpha_{\text{EM-BP}}$ the EM-BP algorithm converges to a fixed point with strictly positive mean-squared error (MSE). This "dynamical" transition is similar to the one found in the cooling of liquids which go into a supercooled glassy state instead of crystallizing, and it appears in the decoding of error-correcting codes [16,17] as well.

We have applied the same technique to the case of seeding-measurement matrices ($L > 1$). The cavity method allows us to analytically locate the dynamical phase transition of s-BP. In the limit of large N, the MSE E_p and the variance messages V_p in each block $p = 1, \ldots L$, the density ρ, the mean $\bar{x}$, and the variance σ^2 of $P(\mathbf{x})$ all evolve according to a dynamical system which can be computed exactly (see Appendix E). One can see numerically if this dynamical system converges to the fixed point corresponding to exact reconstruction ($E_p = 0$ for all p). This study can be used to optimize the design of

the seeding matrix $\mathbf{F}$ by choosing α_1, L, and $J_{p,q}$ in such a way that the convergence to exact reconstruction is as fast as possible. In Fig. 3, we show the convergence time of s-BP predicted by the replica theory for different sets of parameters. For optimized values of the parameters, in the limit of a large number of blocks L and large system sizes N/L, s-BP is capable of exact reconstruction close to the smallest possible number of measurements, $\alpha \to \rho_0$. In practice, finite-size effects slightly degrade this asymptotic threshold saturation, but the s-BP algorithm nevertheless reconstructs signals at rates close to the optimal one regardless of the signal distribution, as illustrated in Fig. 2.

We illustrate the evolution of the s-BP algorithm in Fig. 5. The nonzero signal elements are Gaussian with zero mean, unit variance, density $\rho_0 = 0.4$, and measurement rate $\alpha = 0.5$, which is deep in the glassy region where all other known algorithms fail. The s-BP algorithm first nucleates the native state in the first block and then propagates it through the system. We have also tested the s-BP algorithm on real images where the nonzero components of the signal are far from Gaussian, and the results are nevertheless very good, as shown in Fig. 1. This illustration shows that the quality of the result is not due to a good guess of $P(\mathbf{x})$. It is also important to mention that the gain in performance in using seeding-measurement matrices is really specific to the probabilistic approach: We have computed the phase diagram of ℓ_1 minimization with these matrices and found that, in general, the performance is slightly degraded with respect to the phase diagram of the full measurement matrices, in the large N limit. This result demonstrates that it is the combination of the probabilistic approach, the message-passing reconstruction with parameter learning, and the seeding design of the measurement matrix that is able to reach the best possible performance.

VII. PERSPECTIVES

The seeded compressed-sensing approach introduced here is versatile enough to allow for various extensions. One aspect worth mentioning is the possibility to write the EM-BP equations in terms of N messages instead of the $M \times N$ parameters as described in Appendix C. This is basically the step that goes from the relaxed-BP [24] to the approximate-message-passing (AMP) [11] algorithm. It could be particularly useful when the measurement matrix has some special structure, so that the measurements $\mathbf{y}$ can be obtained in many fewer than $M \times N$ operations (typically in $N \log N$ operations). We have also checked that the approach is robust enough for the introduction of a small amount of noise in the measurements (see Appendix H). Finally, let us mention that, in the case where *a priori* information on the signal is available, it can be incorporated in this approach through a better choice of ϕ, which can considerably improve the performance of the algorithm. For signal with density ρ_0, the worst case, which

we addressed here, is when the nonzero components of the signal are drawn from a continuous distribution. Better performance can be obtained with our method if these nonzero components come from a discrete distribution and one uses this distribution in the choice of ϕ. Another interesting direction in which our formalism can be extended naturally is the use of nonlinear measurements and different types of noises. Altogether, this approach turns out to be very efficient for both random and structured data, as illustrated in Fig. 1, and offers an interesting perspective for practical compressed-sensing applications. Data and code are available online at [6].

ACKNOWLEDGMENTS

Note added.—During the review process for our paper, we became aware of the work [38], in which the authors give a rigorous proof of our result that the threshold $\alpha = \rho_0$ can be reached asymptotically by the s-BP procedure.

APPENDIX A: PROOF OF THE OPTIMALITY OF THE PROBABILISTIC APPROACH

Here we give the main lines of the proof that our probabilistic approach is asymptotically optimal. We consider the case where the signal **s** has iid components

$$P_0(\mathbf{s}) = \prod_{i=1}^{N}[(1 - \rho_0)\delta(s_i) + \rho_0\phi_0(s_i)], \quad (A1)$$

with $0 < \rho_0 < 1$. And we study the probability distribution

$$\hat{P}(\mathbf{x}) = \frac{1}{Z}\prod_{i=1}^{N}\{dx_i[(1 - \rho)\delta(x_i) + \rho\phi(x_i)]\}$$
$$\times \prod_{\mu=1}^{M}\delta_\epsilon\left[\sum_i F_{\mu i}(x_i - s_i)\right], \quad (A2)$$

with a Gaussian $\phi(x)$ of mean zero and unit variance. We stress here that we consider general ϕ_0, i.e., ϕ_0 is not necessarily equal to $\phi(x)$ (and ρ_0 is not necessarily equal to ρ). The measurement matrix **F** is composed of iid elements $F_{\mu i}$ such that, if μ belongs to block q and i belongs to block p, then $F_{\mu i}$ is a random number generated from the Gaussian distribution with zero mean and variance $J_{q,p}/N$. The function $\delta_\epsilon(x)$ is a centered Gaussian distribution with variance ϵ^2.

We show that, with probability going to 1 in the large N limit (at fixed $\alpha = M/N$), the measure $\hat{P}$ (obtained with a generic seeding matrix **F** as described in the main text) is dominated by the signal if $\alpha > \rho_0$, $\alpha' > \rho_0$ [as long as $\phi_0(0)$ is finite].

We introduce the constrained partition function:

$$Y(D, \epsilon) = \int \prod_{i=1}^{N}(dx_i[(1 - \rho)\delta(x_i) + \rho\phi(x_i)])$$
$$\times \prod_{\mu=1}^{M}\delta_\epsilon\left[\sum_i F_{\mu i}(x_i - s_i)\right]\mathbf{1}\left[\sum_{i=1}^{N}(x_i - s_i)^2 > ND\right], \quad (A3)$$

and the corresponding "free-entropy density": $\mathcal{Y}(D, \epsilon) = \lim_{N\to\infty}\mathbb{E}_\mathbf{F}\mathbb{E}_\mathbf{s}\log Y(D, \epsilon)/N$. The notations $\mathbb{E}_\mathbf{F}$ and $\mathbb{E}_\mathbf{s}$ denote, respectively, the expectation value with respect to **F** and to **s**. **1** denotes an indicator function, equal to one if its argument is true, and equal to zero otherwise.

The proof of optimality is obtained by showing that, under the conditions above, $\lim_{\epsilon\to0}\mathcal{Y}(D, \epsilon)/[(\alpha - \rho_0) \times \log(1/\epsilon)]$ is finite if $D = 0$ (statement 1), and it vanishes if $D > 0$ (statement 2). This proves that the measure $\hat{P}$ is dominated by $D = 0$, i.e., by the neighborhood of the signal $x_i = s_i$. The standard "self-averageness" property, which states that the distribution (with respect to the choice of **F** and **s**) of $\log Y(D, \epsilon)/N$ concentrates around $\mathcal{Y}(D, \epsilon)$ when $N \to \infty$, completes the proof. We give here the main lines of the first two steps of the proof.

We first sketch the proof of statement 2. The fact that $\lim_{\epsilon\to0}\mathcal{Y}(D, \epsilon)/[(\alpha - \rho_0)\log(1/\epsilon)] = 0$ when $D > 0$ can be derived by a first moment bound:

$$\mathcal{Y}(D, \epsilon) \leq \lim_{N\to\infty}\frac{1}{N}\mathbb{E}_\mathbf{s}\log Y_{\text{ann}}(D, \epsilon), \quad (A4)$$

where $Y_{\text{ann}}(D, \epsilon)$ is the "annealed partition function" defined as

$$Y_{\text{ann}}(D, \epsilon) = \mathbb{E}_\mathbf{F}Y(D, \epsilon). \quad (A5)$$

In order to evaluate $Y_{\text{ann}}(D, \epsilon)$, one can first compute the annealed partition function in which the distances between **x** and the signal are fixed in each block. More precisely, we define

$$Z(r_1, \cdots, r_L, \epsilon) = \mathbb{E}_\mathbf{F}\int\prod_{i=1}^{N}\{dx_i[(1 - \rho)\delta(x_i) + \rho\phi(x_i)]\}$$
$$\times \prod_{\mu=1}^{M}\delta_\epsilon\left[\sum_i F_{\mu i}(x_i - s_i)\right]$$
$$\times \prod_{p=1}^{L}\delta\left[r_p - \frac{L}{N}\sum_{i\in B_p}(x_i - s_i)^2\right].$$

By noticing that the M random variables $a_\mu = \sum_i F_{\mu i}(x_i - s_i)$ are independent Gaussian random variables, one obtains

$$Z(r_1, \cdots, r_L, \epsilon) = \prod_{p=1}^{L}\left[2\pi\left(\epsilon^2 + \frac{1}{L}\sum_{q=1}^{L}J_{pq}r_q\right)\right]^{-N\alpha_p/2}$$
$$\times \prod_{p=1}^{L}e^{(N/L)\psi(r_p)}, \quad (A6)$$

where

$$\psi(r) = \lim_{n\to\infty} \frac{1}{n} \log\Big\{\int \prod_{i=1}^{n} (dx_i[(1-\rho)\delta(x_i)$$

$$+ \rho\phi(x_i)])\delta\Big[r - \frac{1}{n}\sum_{i=1}^{n}(x_i - s_i)^2\Big]\Big\}. \quad (A7)$$

The behavior of $\psi(r)$ is easily obtained by standard saddle-point methods. In particular, when $r \to 0$, one has $\psi(r) \simeq \frac{1}{2}\rho_0 \log r$.

Using (A6), we obtain, in the small ϵ limit:

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{\mathbf{s}} \log Y_{\text{ann}}(D, \epsilon)$$

$$= \max_{r_1,\cdots,r_L}\Big[\frac{\rho_0}{2}\frac{1}{L}\sum_{p=1}^{L}\log r_p - \frac{1}{L}\sum_{p=1}^{L}\frac{\alpha_p}{2}\log\Big(\epsilon^2 + \frac{1}{L}\sum_{q=1}^{L}J_{pq}r_q\Big)\Big],$$

$$(A8)$$

where the maximum over $r_1, \ldots, r_L$ is to be taken under the constraint $r_1 + \ldots + r_L > LD$. Taking the limit of $\epsilon \to 0$ with a finite D, at least one of the distance r_p must remain finite. It is then easy to show that

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{\mathbf{s}} \log Y_{\text{ann}}(D, \epsilon) = \log(1/\epsilon)\Big[\alpha - \rho_0 - \frac{1}{L}(2\alpha' - \rho_0)\Big],$$

$$(A9)$$

where α' is the fraction of measurements in blocks 2 to L. As $\alpha' > \rho_0$, this is less singular than $\log(1/\epsilon)(\alpha - \rho_0)$, which proves statement 2.

On the contrary, when $D = 0$, we obtain from the same analysis

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{\mathbf{s}} \log Y_{\text{ann}}(D, \epsilon) = \log(1/\epsilon)(\alpha - \rho_0). \quad (A10)$$

This annealed estimate actually gives the correct scaling at small ϵ, as can be shown by the following lower bound. When $D = 0$, we define $\mathcal{V}_0$ as the subset of indices i where $s_i = 0$, $|\mathcal{V}_0| = N(1 - \rho_0)$, and $\mathcal{V}_1$ as the subset of indices i where $s_i \neq 0$, $|\mathcal{V}_1| = N\rho_0$. We obtain a lower bound on $Y(0, \epsilon)$ by substituting $P(\mathbf{x})$ with the factors $(1 - \rho)\delta(x_i)$ when $i \in \mathcal{V}_0$ and with $\rho\phi(x_i)$ when $i \in \mathcal{V}_1$. This gives

$$Y(0, \epsilon) > \exp\{N[(1 - \rho_0)\log(1 - \rho) + \rho_0\log(\rho)$$

$$- (\alpha/2)\log(2\pi) + (\rho_0 - \alpha)\log\epsilon)]\}$$

$$\times \int \prod_{i\in\mathcal{V}_1} du_i \phi(s_i + \epsilon u_i)\exp\Big(-\frac{1}{2}\sum_{i,j\in\mathcal{V}_1} M_{ij}u_iu_j\Big),$$

$$(A11)$$

where $M_{ij} = \sum_{\mu=1}^{\alpha N} F_{\mu i}F_{\mu j}$. The matrix M, of size $\rho_0 N \times \rho_0 N$, is a Wishart-like random matrix. For $\alpha > \rho_0$, generically, its eigenvalues are strictly positive, as we show below. Using this property, one can show that, if

$\prod_{i\in\gamma_1}\phi(s_i) > 0$, the integral over the variables u_i in (A11) is strictly positive in the limit $\epsilon \to 0$. The divergence of $\mathcal{Y}(0, \epsilon)$ in the limit $\epsilon \to 0$ is due to the explicit term $\exp[N(\rho_0 - \alpha)\log\epsilon]$ in Eq. (A11).

The fact that all eigenvalues of M are strictly positive is well known in the case of $L = 1$ where the spectrum has been obtained by Marčenko and Pastur [39]. In general, the fact that all the eigenvalues of M are strictly positive is equivalent to saying that all the lines of the $\alpha N \times \rho_0 N$ matrix F (which is the restriction of the measurement matrix to columns with nonzero signal components) are linearly independent. In the case of seeding matrices with general L, this statement is basically obvious by construction of the matrices, in the regime where, in each block q, $\alpha_q > \rho_0$ and $J_{qq} > 0$. A more formal proof can be obtained as follows. We consider the Gaussian integral

$$Z(v) = \int \prod_{i=1}^{n} d\phi_i \exp\Big[-\frac{1}{2}\sum_{i,j,\mu} F_{\mu i}F_{\mu j}\phi_i\phi_j + \frac{v}{2}\sum_i \phi_i^2\Big].$$

$$(A12)$$

This quantity is finite if and only if v is smaller than the smallest eigenvalue, $\lambda_{\min}$, of M. We now compute the annealed average $Z_{\text{ann}}(v) = \mathbb{E}_{\mathbf{F}}Z(v)$. If $Z_{\text{ann}}(v)$ is finite, then the probability that $\lambda_{\min} \leq v$ goes to zero in the large N limit. Using methods similar to the one above, one can show that

$$\frac{2L}{n}\log\mathbb{E}_{\mathbf{F}}Z(v) = \max_{r_1,\ldots,r_p}\Big[\sum_{p=1}^{L}(\log r_p + vr_p)$$

$$- \sum_{p=1}^{L}\frac{\alpha_p}{\rho_0}\log\Big(1 + \frac{1}{L}\sum_q J_{pq}r_q\Big)\Big]. \quad (A13)$$

The saddle-point equations

$$\frac{1}{r_p} + v = \frac{1}{L}\sum_{q=1}^{L}\frac{\alpha_q}{\rho_0}\frac{J_{qp}}{1 + \frac{1}{L}\sum_s J_{qs}r_s} \quad (A14)$$

have a solution at $v = 0$ (and by continuity also at $v > 0$ small enough), when $\frac{1}{L}\sum_{q=1}^{L}\frac{\alpha_q}{\rho_0} = \frac{\alpha}{\rho_0} > 1$. (It can be found, for instance, by iteration.) Therefore, $\frac{2L}{n}\log\mathbb{E}_{\mathbf{F}}Z(v)$ is finite for some $v > 0$ small enough, and therefore $\lambda_{\min} > 0$.

APPENDIX B: DERIVATION OF EXPECTATION-MAXIMIZATION BELIEF PROPAGATION

In this and the next sections, we present the message-passing algorithm that we used for reconstruction in compressed sensing. In this section, we derive its message-passing form, where $O(NM)$ messages are being sent between each signal component i and each measurement μ. This algorithm was used in [24], where it was called the *relaxed belief propagation*, as an approximate algorithm for the case of a sparse measurement matrix $\mathbf{F}$. In the case that we use here of a measurement matrix which is not

sparse (that is, a finite fraction of the elements of **F** is nonzero, and all the nonzero elements scale as $1/\sqrt{N}$), the algorithm is asymptotically exact. We show here for completeness how to derive it. In the next section, we then derive asymptotically equivalent equations that depend only on $O(N)$ messages. In statistical physics terms, this corresponds to the TAP equations [27] with the Onsager reaction term, which are asymptotically equivalent to the BP on fully connected models. In the context of compressed sensing, this form of equations has been used previously [11], and it is called *approximate message passing* (AMP). In cases when the matrix **F** can be computed recursively (e.g., via fast Fourier transform), the running time of the AMP-type message passing is $O(N \log N)$ (compared to the $O(NM)$ for the non-AMP form). Apart from this speeding up, both classes of message passing give the same performance.

We derive here the message-passing algorithm in the case where measurements have additive Gaussian noise; the noiseless case limit is easily obtained at the end. The posterior probability of **x** after the measurement of **y** is given by

$$\hat{P}(\mathbf{x}) = \frac{1}{Z} \prod_{i=1}^{N} [(1 - \rho)\delta(x_i) + \rho\phi(x_i)]$$
$$\times \prod_{\mu=1}^{M} \frac{1}{\sqrt{2\pi\Delta_\mu}} e^{-(1/2\Delta_\mu)(y_\mu - \sum_{i=1}^{N} F_{\mu i}x_i)^2}, \quad \text{(B1)}$$

where Z is a normalization constant (the partition function) and Δ_μ is the variance of the noise in measurement μ. The noiseless case is recovered in the limit $\Delta_\mu \to 0$. The optimal estimate, which minimizes the MSE with respect to the original signal **s**, is obtained from averages of x_i with respect to the probability measure $\hat{P}(\mathbf{x})$. Exact computation of these averages would require exponential time; belief propagation provides a standard approximation. The canonical BP equations for probability measure $\hat{P}(\mathbf{x})$ read

$$m_{\mu \to i}(x_i) = \frac{1}{Z^{\mu \to i}} \int \left[\prod_{j(\neq i)} m_{j \to \mu}(x_j) dx_j \right] e^{-(1/2\Delta_\mu)(\sum_{j \neq i} F_{\mu j}x_j + F_{\mu i}x_i - y_\mu)^2}, \quad \text{(B2)}$$

$$m_{i \to \mu}(x_i) = \frac{1}{Z^{i \to \mu}} [(1 - \rho)\delta(x_i) + \rho\phi(x_i)] \prod_{\gamma \neq \mu} m_{\gamma \to i}(x_i), \quad \text{(B3)}$$

where $Z^{\mu \to i}$ and $Z^{i \to \mu}$ are normalization factors ensuring that $\int dx_i m_{\mu \to i}(x_i) = \int dx_i m_{i \to \mu}(x_i) = 1$. These are integral equations for probability distributions that are still practically intractable in this form. We can, however, take advantage of the fact that, after we properly rescale the linear system $\mathbf{y} = \mathbf{Fx}$ in such a way that elements of **y** and **x** are of $O(1)$, the matrix $F_{\mu i}$ has random elements with variance of $O(1/N)$. Using the Hubbard-Stratonovich transformation [40]

$$e^{-(\omega^2/2\Delta)} = \frac{1}{\sqrt{2\pi\Delta}} \int d\lambda \, e^{-(\lambda^2/2\Delta) + (i\lambda\omega/\Delta)} \quad \text{(B4)}$$

for $\omega = (\sum_{j \neq i} F_{\mu j}x_j)$, we can simplify Eq. (B2) as follows:

$$m_{\mu \to i}(x_i) = \frac{1}{Z^{\mu \to i}\sqrt{2\pi\Delta_\mu}} e^{-(1/2\Delta_\mu)(F_{\mu i}x_i - y_\mu)^2} \int d\lambda \, e^{-(\lambda^2/2\Delta_\mu)} \prod_{j \neq i} \left[\int dx_j m_{j \to \mu}(x_j) e^{F_{\mu j}x_j/\Delta_\mu(y_\mu - F_{\mu i}x_i + i\lambda)} \right]. \quad \text{(B5)}$$

Now we expand the last exponential around zero; because the term $F_{\mu j}$ is small in N, we keep all terms that are of $O(1/N)$. Introducing means and variances as new messages

$$a_{i \to \mu} \equiv \int dx_i x_i m_{i \to \mu}(x_i), \quad \text{(B6)}$$

$$v_{i \to \mu} \equiv \int dx_i x_i^2 m_{i \to \mu}(x_i) - a_{i \to \mu}^2, \quad \text{(B7)}$$

we obtain

$$m_{\mu \to i}(x_i) = \frac{1}{Z^{\mu \to i}\sqrt{2\pi\Delta_\mu}} e^{-(1/2\Delta_\mu)(F_{\mu i}x_i - y_\mu)^2} \int d\lambda \, e^{-(\lambda^2/2\Delta_\mu)} \prod_{j \neq i} [e^{(F_{\mu j}a_{j \to \mu}/\Delta_\mu)(y_\mu - F_{\mu i}x_i + i\lambda) + (F_{\mu j}^2 v_{j \to \mu}/2\Delta_\mu^2)(y_\mu - F_{\mu i}x_i + i\lambda)^2}]. \quad \text{(B8)}$$

Performing the Gaussian integral over λ, we obtain

$$m_{\mu \to i}(x_i) = \frac{1}{\tilde{Z}^{\mu \to i}} e^{-(x_i^2/2)A_{\mu \to i} + B_{\mu \to i} x_i},$$

$$\tilde{Z}^{\mu \to i} = \sqrt{\frac{2\pi}{A_{\mu \to i}}} e^{B_{\mu \to i}^2/2A_{\mu \to i}}, \quad \text{(B9)}$$

where we introduce

$$A_{\mu \to i} = \frac{F_{\mu i}^2}{\Delta_\mu + \sum_{j \neq i} F_{\mu j}^2 v_{j \to \mu}}, \quad \text{(B10)}$$

$$B_{\mu \to i} = \frac{F_{\mu i}(y_\mu - \sum_{j \neq i} F_{\mu j} a_{j \to \mu})}{\Delta_\mu + \sum_{j \neq i} F_{\mu j}^2 v_{j \to \mu}}, \quad \text{(B11)}$$

and the normalization $\tilde{Z}^{\mu \to i}$ contains all the x_i-independent factors. The noiseless case corresponds to $\Delta_\mu = 0$.

To close the equations on messages $a_{i \to \mu}$ and $v_{i \to \mu}$, we notice that

$$m_{i \to \mu}(x_i) = \frac{1}{\tilde{Z}^{i \to \mu}} [(1-\rho)\delta(x_i) + \rho\phi(x_i)]$$

$$\times e^{-(x_i^2/2)\sum_{\gamma \neq \mu} A_{\gamma \to i} + x_i \sum_{\gamma \neq \mu} B_{\gamma \to i}}. \quad \text{(B12)}$$

Messages $a_{i \to \mu}$ and $v_{i \to \mu}$ are, respectively, the mean and variance of the probability distribution $m_{i \to \mu}(x_i)$. For general $\phi(x_i)$ the mean and variance (B6) and (B7) will be computed using numerical integration over x_i. Equations (B6) and (B7) together with (B10)–(B12) then lead to closed iterative message-passing equations.

In all the specific examples shown here and in the main part of the paper, we have used a Gaussian $\phi(x_i)$ with mean $\bar{x}$ and variance σ^2. We define two functions

$$f_a(X, Y) = \left[\frac{\rho(Y + \bar{x}/\sigma^2)}{\sigma(1/\sigma^2 + X)^{3/2}} \right]$$

$$\times \left[(1-\rho)e^{-[(Y+\bar{x}/\sigma^2)^2/2(1/\sigma^2+X)]+(\bar{x}^2/2\sigma^2)} \right.$$

$$\left. + \frac{\rho}{\sigma(1/\sigma^2 + X)^{1/2}} \right]^{-1}, \quad \text{(B13)}$$

$$f_c(X, Y) = \left[\frac{\rho}{\sigma(1/\sigma^2 + X)^{3/2}} \left(1 + \frac{(Y + \bar{x}/\sigma^2)^2}{1/\sigma^2 + X} \right) \right]$$

$$\times \left[(1-\rho)e^{-[(Y+\bar{x}/\sigma^2)^2/2(1/\sigma^2+X)]+(\bar{x}^2/2\sigma^2)} \right.$$

$$\left. + \frac{\rho}{\sigma(1/\sigma^2 + X)^{1/2}} \right]^{-1} - f_a^2(X, Y). \quad \text{(B14)}$$

Then the closed form of the BP update is

$$a_{i \to \mu} = f_a\left(\sum_{\gamma \neq \mu} A_{\gamma \to i}, \sum_{\gamma \neq \mu} B_{\gamma \to i} \right),$$

$$a_i = f_a\left(\sum_\gamma A_{\gamma \to i}, \sum_\gamma B_{\gamma \to i} \right), \quad \text{(B15)}$$

$$v_{i \to \mu} = f_c\left(\sum_{\gamma \neq \mu} A_{\gamma \to i}, \sum_{\gamma \neq \mu} B_{\gamma \to i} \right),$$

$$v_i = f_c\left(\sum_\gamma A_{\gamma \to i}, \sum_\gamma B_{\gamma \to i} \right), \quad \text{(B16)}$$

where the a_i and v_i are the mean and variance of the marginal probabilities of variable x_i.

As we discussed in the main text, the parameters ρ, $\bar{x}$, and σ are usually not known in advance. However, their values can be learned within the probabilistic approach. A standard way to do so is called expectation maximization [33]. One realizes that the partition function

$$Z(\rho, \bar{x}, \sigma) = \int \prod_{i=1}^N dx_i \prod_{i=1}^N \left[(1-\rho)\delta(x_i) \right.$$

$$\left. + \frac{\rho}{\sqrt{2\pi}\sigma} e^{-[(x_i-\bar{x})^2/2\sigma^2]} \right]$$

$$\times \prod_{\mu=1}^M \frac{1}{\sqrt{2\pi\Delta_\mu}} e^{-(1/2\Delta_\mu)(y_\mu - \sum_{i=1}^N F_{\mu i} x_i)^2}, \quad \text{(B17)}$$

is proportional to the probability of the true parameters ρ_0, $\bar{s}$, σ_0, given the measurement **y**. Hence, to compute the most probable values of parameters, one searches for the maximum of this partition function. Within the BP approach, the logarithm of the partition function is the Bethe free entropy expressed as [17]

$$F(\rho, \bar{x}, \sigma) = \sum_\mu \log Z^\mu + \sum_i \log Z^i - \sum_{(\mu i)} \log Z^{\mu i}, \quad \text{(B18)}$$

where

$$Z^i = \int dx_i \prod_\mu m_{\mu \to i}(x_i) [(1-\rho)\delta(x_i) + \frac{\rho}{\sqrt{2\pi}\sigma} e^{-[(x_i-\bar{x})^2/2\sigma^2]}], \quad \text{(B19)}$$

$$Z^\mu = \int \prod_i dx_i \prod_i m_{i \to \mu}(x_i) \frac{1}{\sqrt{2\pi\Delta_\mu}} e^{-[(y_\mu - \sum_i F_{\mu i} x_i)^2/2\Delta_\mu]}, \quad \text{(B20)}$$

$$Z^{\mu i} = \int dx_i m_{\mu \to i}(x_i) m_{i \to \mu}(x_i). \quad \text{(B21)}$$

The stationarity condition of Bethe free entropy (B18) with respect to ρ leads to

$$\rho = \frac{\sum_i \frac{1/\sigma^2 + U_i}{V_i + \bar{x}/\sigma^2} a_i}{\sum_i [1 - \rho + \frac{\rho}{\sigma(1/\sigma^2 + U_i)^{1/2}} e^{[(V_i + \bar{x}/\sigma^2)^2/2(1/\sigma^2 + U_i)] - (\bar{x}^2/2\sigma^2)}]^{-1}}, \tag{B22}$$

where $U_i = \sum_\gamma A_{\gamma \to i}$, and $V_i = \sum_\gamma B_{\gamma \to i}$. Stationarity with respect to $\bar{x}$ and σ gives

$$\bar{x} = \frac{\sum_i a_i}{\rho \sum_i [\rho + (1 - \rho)\sigma(1/\sigma^2 + U_i)^{1/2} e^{-[(V_i + \bar{x}/\sigma^2)^2/2(1/\sigma^2 + U_i)] + (\bar{x}^2/2\sigma^2)}]^{-1}}, \tag{B23}$$

$$\sigma^2 = \frac{\sum_i (v_i + a_i^2)}{\rho \sum_i [\rho + (1 - \rho)\sigma(1/\sigma^2 + U_i)^{1/2} e^{-[(V_i + \bar{x}/\sigma^2)^2/2(1/\sigma^2 + U_i)] + (\bar{x}^2/2\sigma^2)}]^{-1}} - \bar{x}^2. \tag{B24}$$

In statistical physics, conditions (B22) are known as Nishimori conditions [34,36]. In the expectation maximization equations (B22)–(B24), these conditions are used iteratively for the update of the current guess of parameters. A reasonable initial guess is $\rho_{\text{init}} = \alpha$. The value of $\rho_0 \bar{s}$ can also be obtained with a special line of measurement consisting of a unit vector; hence, we assume that, given an estimate of ρ, the $\bar{x} = \rho_0 \bar{s}/\rho$. In the case where the matrix $\mathbf{F}$ is random, with Gaussian elements of zero mean and variance $1/N$, we can also use the following equation for learning the variance: $\sum_{\mu=1}^M y_\mu^2/N = \alpha \rho_0 \langle s^2 \rangle = \alpha \rho(\sigma^2 + \bar{x}^2)$.

APPENDIX C: AMP FORM OF THE MESSAGE PASSING

In the large N limit, the messages $a_{i \to \mu}$ and $v_{i \to \mu}$ are nearly independent of μ, but one must be careful to keep the correcting Onsager reaction terms. Let us define

$$\omega_\mu = \sum_i F_{\mu i} a_{i \to \mu}, \qquad \gamma_\mu = \sum_i F_{\mu i}^2 v_{i \to \mu}, \tag{C1}$$

$$U_i = \sum_\mu A_{\mu \to i}, \qquad V_i = \sum_\mu B_{\mu \to i}. \tag{C2}$$

Then we have

$$U_i = \sum_\mu \frac{F_{\mu i}^2}{\Delta_\mu + \gamma_\mu - F_{\mu i}^2 v_{i \to \mu}} \simeq \sum_\mu \frac{F_{\mu i}^2}{\Delta_\mu + \gamma_\mu}, \tag{C3}$$

$$\begin{aligned} V_i &= \sum_\mu \frac{F_{\mu i}(y_\mu - \omega_\mu + F_{\mu i} a_{i \to \mu})}{\Delta_\mu + \gamma_\mu - F_{\mu i}^2 v_{i \to \mu}} \\ &\simeq \sum_\mu F_{\mu i} \frac{(y_\mu - \omega_\mu)}{\Delta_\mu + \gamma_\mu} + f_a(U_i, V_i) \sum_\mu F_{\mu i}^2 \frac{1}{\Delta_\mu + \gamma_\mu}. \end{aligned} \tag{C4}$$

We now compute ω_μ:

$$\begin{aligned} a_{i \to \mu} &= f_a(U_i - A_{\mu \to i}, V_i - B_{\mu \to i}) \\ &\simeq a_i - A_{\mu \to i} \frac{\partial f_a}{\partial X}(U_i, V_i) - B_{\mu \to i} \frac{\partial f_a}{\partial Y}(U_i, V_i). \end{aligned} \tag{C5}$$

To express $\omega_\mu = \sum_i F_{\mu i} a_{i \to \mu}$, we see that the first correction term has a contribution in $F_{\mu i}^3$ and can be safely neglected. On the contrary, the second term has a contribution in $F_{\mu i}^2$, which one should keep. Therefore,

$$\omega_\mu = \sum_i F_{\mu i} f_a(U_i, V_i) - \frac{(y_\mu - \omega_\mu)}{\Delta_\mu + \gamma_\mu} \sum_i F_{\mu i}^2 \frac{\partial f_a}{\partial Y}(U_i, V_i). \tag{C6}$$

The computation of γ_μ is similar; it gives:

$$\begin{aligned} \gamma_\mu &= \sum_i F_{\mu i}^2 v_i - \sum_i F_{\mu i}^3 \frac{(y_\mu - \omega_\mu)}{\Delta_\mu + \gamma_\mu} \frac{\partial f_c}{\partial Y}(U_i, V_i) \\ &\simeq \sum_i F_{\mu i}^2 f_c(U_i, V_i). \end{aligned} \tag{C7}$$

For a known form of matrix $\mathbf{F}$, these equations can be slightly simplified further by using the assumptions of the BP approach about independence of $F_{\mu i}$ and BP messages. This plus a law of large number implies that, for matrix $\mathbf{F}$ with Gaussian entries of zero mean and unit variance, one can effectively "replace" every $F_{\mu i}^2$ by $1/N$ in Eqs. (C3), (C4), (C6), and (C7). This leads, for homogeneous or bloc matrices, to even simpler equations and a slightly faster algorithm.

Equations (C3), (C4), (C6), and (C7), with or without the later simplification, give a system of closed equations. They are a special form [$P(\mathbf{x})$ and hence functions f_a, f_c are different in our case] of the approximate message passing of [11].

The final reconstruction algorithm for the general measurement matrix and with the learning of the $P(\mathbf{x})$ parameters can hence be summarized in a schematic way (see Algorithm 1).

Note that in practice we use "damping." At each update, the new message is obtained as u times the old value plus $1 - u$ times the newly computed value, with a damping

Algorithm 1. EM-BP(y_μ, $F_{\mu i}$, criterium, t_{max})

Initialize randomly messages U_i from interval [0,1] for every component;
Initialize randomly messages V_i from interval $[-1, 1]$ for every component;
Initialize messages $\omega_\mu \leftarrow y_\mu$;
Initialize randomly messages γ_μ from interval [0,1] for every measurement;
Initialize the parameters $\rho \leftarrow \alpha$, $\bar{x} \leftarrow 0$, $\sigma^2 \leftarrow 1$.
conv $\leftarrow$ criterium + 1; $t \leftarrow 0$;
while conv > criterium and $t < t_{max}$:
    do $t \leftarrow t + 1$;
        for each component i:
            do $x_i^{old} \leftarrow f_a(U_i, V_i)$;
                Update U_i according to Eq. (C3).
                Update V_i according to Eq. (C4).
        for each measurement μ:
            do Update ω_μ according to Eq. (C6).
                Update γ_μ according to Eq. (C7).
        Update ρ according to Eq. (B22).
        Update $\bar{x}$ and σ^2 according to Eq. (B24).
        for each component i:
            do $x_i \leftarrow f_a(U_i, V_i)$;
        conv $\leftarrow$ mean($|x_i - x_i^{old}|$);
return signal components **x**

$0 < u < 1$, typically $u = 0.5$. For both the update of messages and learning of parameters, empirically this damping speeds up the convergence. Note also that the algorithm is relatively robust with respect to the initialization of

messages. The reported initialization was used to obtain the results in Fig. 1. However, other initializations are possible. Note also that, for specific classes of signals or measurement matrices, the initial conditions may be adjusted to take into account the magnitude of the values $F_{\mu i}$ and y_μ.

For a general matrix **F**, one iteration takes $O(NM)$ steps. We have observed the number of iterations needed for convergence to be basically independent of N; however, the constant depends on the parameters and the signal (see Fig. 3). For matrices that can be computed recursively (i.e., without storing all their NM elements), a speeding-up is possible, as the message-passing loop takes only $O(M + N)$ steps.

APPENDIX D: REPLICA ANALYSIS AND DENSITY EVOLUTION—FULL MEASUREMENT MATRIX

Averaging over disorder leads to replica equations that describe the $N \to \infty$ behavior of the partition function as well as the density evolution of the belief-propagation algorithm. The replica trick evaluates $\mathbb{E}_{\mathbf{F},\mathbf{s}}(\log Z)$ via

$$\Phi = \frac{1}{N}\mathbb{E}(\log Z) = \frac{1}{N}\lim_{n \to 0}\frac{\mathbb{E}(Z^n) - 1}{n}. \tag{D1}$$

In the case where the matrix **F** is the full measurement with all elements iid from a normal distribution with zero mean and variance unity, one finds that Φ is obtained as the saddle-point value of the function:

$$\Phi(Q, q, m, \hat{Q}, \hat{q}, \hat{m}) = -\frac{\alpha}{2}\frac{q - 2m + \rho_0\langle s^2\rangle + \Delta}{\Delta + Q - q} - \frac{\alpha}{2}\log(\Delta + Q - q) + \frac{Q\hat{Q}}{2} - m\hat{m} + \frac{q\hat{q}}{2}$$
$$+ \int \mathcal{D}z \int ds[(1 - \rho_0)\delta(s) + \rho_0\phi_0(s)]\log\left\{\int dx e^{-(\hat{Q}+\hat{q}/2)x^2+\hat{m}xs+z\sqrt{\hat{q}}x}[(1 - \rho)\delta(x) + \rho\phi(x)]\right\}. \tag{D2}$$

Here, $\mathcal{D}z$ is a Gaussian integration measure with zero mean and variance equal to one; ρ_0 is the density of the signal; and $\phi_0(s)$ is the distribution of the signal components and $\langle s^2\rangle = \int ds s^2 \phi_0(s)$ is its second moment. Δ is the variance of the measurement noise. The noiseless case is recovered by using $\Delta = 0$.

The physical meaning of the order parameters is

$$Q = \frac{1}{N}\sum_i\langle x_i^2\rangle, \qquad q = \frac{1}{N}\sum_i\langle x_i\rangle^2, \qquad m = \frac{1}{N}\sum_i s_i\langle x_i\rangle, \tag{D3}$$

whereas the other three $\hat{m}$, $\hat{q}$, $\hat{Q}$ are auxiliary parameters. Performing the saddle-point derivative with respect to m, q, $Q - q$, $\hat{m}$, $\hat{q}$, and $\hat{Q} + \hat{q}$, we obtain the following six self-consistent equations [using the Gaussian form of $\phi(x)$, with mean $\bar{x}$ and variance σ^2]:

$$\hat{m} = \frac{\alpha}{\Delta + Q - q} = \hat{Q} + \hat{q}, \qquad \hat{Q} = \frac{\alpha}{\Delta + Q - q} - \alpha\frac{q - 2m + \rho_0\langle s^2\rangle + \Delta}{(\Delta + Q - q)^2}, \tag{D4}$$

$$m = \frac{\rho_0\rho}{\hat{Q} + \hat{q} + 1/\sigma^2} \int \mathcal{D}z \int ds s\phi_0(s) \frac{\hat{m}s + z\sqrt{\hat{q}} + \bar{x}/\sigma^2}{(1-\rho)\sigma\sqrt{\hat{Q} + \hat{q} + 1/\sigma^2}e^{(\bar{x}^2/2\sigma^2)-[(\hat{m}s+z\sqrt{\hat{q}}+\bar{x}/\sigma^2)^2/2(\hat{Q}+\hat{q}+1/\sigma^2)]} + \rho}, \tag{D5}$$

$$Q - q = \frac{(1-\rho_0)\rho}{(\hat{Q} + \hat{q} + 1/\sigma^2)\sqrt{\hat{q}}} \int \mathcal{D}z z \frac{z\sqrt{\hat{q}} + \bar{x}/\sigma^2}{(1-\rho)\sigma\sqrt{\hat{Q} + \hat{q} + 1/\sigma^2}e^{(\bar{x}^2/2\sigma^2)-[(z\sqrt{\hat{q}}+\bar{x}/\sigma^2)^2/2(\hat{Q}+\hat{q}+1/\sigma^2)]} + \rho}$$

$$+ \frac{\rho_0\rho}{(\hat{Q} + \hat{q} + 1/\sigma^2)\sqrt{\hat{q}}} \int \mathcal{D}z z \int ds\phi_0(s) \frac{\hat{m}s + z\sqrt{\hat{q}} + \bar{x}/\sigma^2}{(1-\rho)\sigma\sqrt{\hat{Q} + \hat{q} + 1/\sigma^2}e^{(\bar{x}^2/2\sigma^2)-[(\hat{m}s+z\sqrt{\hat{q}}+\bar{x}/\sigma^2)^2/2(\hat{Q}+\hat{q}+1/\sigma^2)]} + \rho}, \tag{D6}$$

$$Q = \frac{(1-\rho_0)\rho}{(\hat{Q} + \hat{q} + 1/\sigma^2)^2} \int \mathcal{D}z \frac{(z\sqrt{\hat{q}} + \bar{x}/\sigma^2)^2 + \hat{Q} + \hat{q} + 1/\sigma^2}{(1-\rho)\sigma\sqrt{\hat{Q} + \hat{q} + 1/\sigma^2}e^{(\bar{x}^2/2\sigma^2)-[(z\sqrt{\hat{q}}+\bar{x}/\sigma^2)^2/2(\hat{Q}+\hat{q}+1/\sigma^2)]} + \rho}$$

$$+ \frac{\rho_0\rho}{(\hat{Q} + \hat{q} + 1/\sigma^2)^2} \int \mathcal{D}z \int ds\phi_0(s) \frac{(\hat{m}s + z\sqrt{\hat{q}} + \bar{x}/\sigma^2)^2 + \hat{Q} + \hat{q} + 1/\sigma^2}{(1-\rho)\sigma\sqrt{\hat{Q} + \hat{q} + 1/\sigma^2}e^{(\bar{x}^2/2\sigma^2)-[(\hat{m}s+z\sqrt{\hat{q}}+\bar{x}/\sigma^2)^2/2(\hat{Q}+\hat{q}+1/\sigma^2)]} + \rho}. \tag{D7}$$

We now show the connection between this replica computation and the evolution of belief-propagation messages, studying first the case where one does not change the parameters ρ, $\bar{x}$, and σ. Let us introduce parameters m_{BP}, q_{BP}, Q_{BP}, defined via the belief-propagation messages as:

$$m_{BP}^{(t)} = \frac{1}{N}\sum_{i=1}^{N} a_i^{(t)}s_i, \qquad q_{BP}^{(t)} = \frac{1}{N}\sum_{i=1}^{N}(a_i^{(t)})^2,$$
$$Q_{BP}^{(t)} - q_{BP}^{(t)} = \frac{1}{N}\sum_{i=1}^{N} v_i^{(t)}. \tag{D8}$$

The density- (state-)evolution equations for these parameters can be derived in the same way as in [11,32], and this leads to the result that m_{BP}, q_{BP}, Q_{BP} evolve under the update of BP in exactly the same way as according to iterations of Eqs. (D5)–(D7). Hence, the analytical equations (D5)–(D7) allow us to study the performance of the BP algorithm. Note also that the density-evolution equations are the same for the message-passing and for the AMP equations. It turns out that the above equations close in terms of two parameters: the mean-squared error $E_{BP}^{(t)} = q_{BP}^{(t)} - 2m_{BP}^{(t)} + \rho_0\langle s^2\rangle$ and the variance $V_{BP}^{(t)} = Q_{BP}^{(t)} - q_{BP}^{(t)}$. From Eqs. (D4)–(D7) one easily obtains a closed mapping $(E_{BP}^{(t+1)}, V_{BP}^{(t+1)}) = f(E_{BP}^{(t)}, V_{BP}^{(t)})$.

In the main text, we defined the function $\Phi(D)$, which is the free entropy restricted to configurations $\mathbf{x}$ for which $D = \sum_{i=1}^{N}(x_i - s_i)^2/N$ is fixed. This is evaluated as the saddle point over Q, q, $\hat{Q}$, $\hat{q}$, $\hat{m}$ of the function $\Phi(Q, q, (Q - D + \rho_0\langle s^2\rangle)/2, \hat{Q}, \hat{q}, \hat{m})$. This function is plotted in Fig. 3(a).

In presence of expectation-maximization learning of the parameters, the density evolution for the conditions (B22) and (B24) are

$$\rho^{(t+1)} = \rho^{(t)}\left(\int \mathcal{D}z \int dx_0[(1-\rho_0)\delta(x_0) + \rho_0\phi_0(x_0)]\frac{g(\hat{Q} + \hat{q}, \hat{m}x_0 + z\sqrt{\hat{q}})}{1 - \rho + \rho g(\hat{Q} + \hat{q}, \hat{m}x_0 + z\sqrt{\hat{q}})}\right)$$
$$\times \left(\int \mathcal{D}z \int dx_0[(1-\rho_0)\delta(x_0) + \rho_0\phi_0(x_0)]\frac{1}{1 - \rho + \rho g(\hat{Q} + \hat{q}, \hat{m}x_0 + z\sqrt{\hat{q}})}\right)^{-1}, \tag{D9}$$

$$\bar{x}^{(t+1)} = \frac{1}{\rho}\left(\int \mathcal{D}z \int dx_0[(1-\rho_0)\delta(x_0) + \rho_0\phi_0(x_0)]f_a(\hat{Q} + \hat{q}, \hat{m}x_0 + z\sqrt{\hat{q}})\right)$$
$$\times \left(\int \mathcal{D}z \int dx_0[(1-\rho_0)\delta(x_0) + \rho_0\phi_0(x_0)]\frac{g(\hat{Q} + \hat{q}, \hat{m}x_0 + z\sqrt{\hat{q}})}{1 - \rho + \rho g(\hat{Q} + \hat{q}, \hat{m}x_0 + z\sqrt{\hat{q}})}\right)^{-1}, \tag{D10}$$

$$(\sigma^2)^{(t+1)} = \frac{1}{\rho}\left(\int \mathcal{D}z \int dx_0[(1-\rho_0)\delta(x_0) + \rho_0\phi_0(x_0)][f_a(\hat{Q} + \hat{q}, \hat{m}x_0 + z\sqrt{\hat{q}})^2 + f_c(\hat{Q} + \hat{q}, \hat{m}x_0 + z\sqrt{\hat{q}})]\right)$$
$$\times \left(\int \mathcal{D}z \int dx_0[(1-\rho_0)\delta(x_0) + \rho_0\phi_0(x_0)]\frac{g(\hat{Q} + \hat{q}, \hat{m}x_0 + z\sqrt{\hat{q}})}{1 - \rho + \rho g(\hat{Q} + \hat{q}, \hat{m}x_0 + z\sqrt{\hat{q}})}\right)^{-1} - [\bar{x}^{(t+1)}]^2. \tag{D11}$$

The density-evolution equations now provide a mapping,

$$(E_{\text{EM-BP}}^{(t+1)}, V_{\text{EM-BP}}^{(t+1)}, \rho^{(t+1)}, \bar{x}^{(t+1)}, \sigma^{(t+1)})$$
$$= f(E_{\text{EM-BP}}^{(t)}, V_{\text{EM-BP}}^{(t)}, \rho^{(t)}, \bar{x}^{(t)}, \sigma^{(t)}), \qquad \text{(D12)}$$

obtained by complementing the previous equations on $E_{\text{EM-BP}}^{(t)}, V_{\text{EM-BP}}^{(t)}$ with the update equations (D9)–(D11). The next section gives explicitly the full set of equations in the case of seeding matrices; the ones for the full matrices are obtained by taking $L = 1$. These are the equations that we study to describe analytically the evolution of the EM-BP algorithm and obtain the phase diagram for the reconstruction (see Fig. 2).

APPENDIX E: REPLICA ANALYSIS AND DENSITY EVOLUTION FOR SEEDING-MEASUREMENT MATRICES

Many choices of J_1 and J_2 actually work very well, and good performance for seeding-measurement matrices can be easily obtained. In fact, the form of the matrix that we have used is by no means the only one that can produce the seeding mechanism, and we expect that better choices, in terms of convergence time, finite-size effects, and sensibility to noise, could be unveiled in the near future.

With the matrix presented in this work, and in order to obtain the best performance (in terms of phase transition limit and of speed of convergence), one needs to optimize the value of J_1 and J_2 depending on the type of signal. Fortunately, this can be analyzed with the replica method. The analytic study in the case of seeding-measurement matrices is in fact done using the same techniques as for the full matrix. The order parameters are now the MSE $E_p = q_p - 2m_p + \rho_0 \langle s^2 \rangle$ and variance $V_p = Q_p - q_p$ in each block $p \in \{1, \ldots, L\}$. Consequently, we obtain the final dynamical system of $2L + 3$-order parameters describing the density evolution of the s-BP algorithm. The order parameters at iteration $t + 1$ of the message-passing algorithm are given by:

$$E_q^{(t+1)} = \int dx^0 [(1 - \rho_0)\delta(x^0) + \rho_0 \phi_0(x^0)] \int \mathcal{D}z (f_a(\hat{Q}_q + \hat{q}_q, \hat{m}_q x^0 + z\sqrt{\hat{q}_q}) - x^0)^2, \qquad \text{(E1)}$$

$$V_q^{(t+1)} = \int dx^0 [(1 - \rho_0)\delta(x^0) + \rho_0 \phi_0(x^0)] \int \mathcal{D}z f_c(\hat{Q}_q + \hat{q}_q, \hat{m}_q x^0 + z\sqrt{\hat{q}_q}), \qquad \text{(E2)}$$

$$\rho^{(t+1)} = \rho^{(t)} \left(\frac{1}{L} \sum_{p=1}^{L} \int \mathcal{D}z \int dx_0 [(1 - \rho_0)\delta(x_0) + \rho_0 \phi_0(x_0)] \frac{g(\hat{Q}_p + \hat{q}_p, \hat{m}_p x_0 + z\sqrt{\hat{q}_p})}{1 - \rho + \rho g(\hat{Q}_p + \hat{q}_p, \hat{m}_p x_0 + z\sqrt{\hat{q}_p})} \right)$$
$$\times \left(\frac{1}{L} \sum_{p=1}^{L} \int \mathcal{D}z \int dx_0 [(1 - \rho_0)\delta(x_0) + \rho_0 \phi_0(x_0)] \frac{1}{1 - \rho + \rho g(\hat{Q}_p + \hat{q}_p, \hat{m}_p x_0 + z\sqrt{\hat{q}_p})} \right)^{-1}, \qquad \text{(E3)}$$

$$\bar{x}^{(t+1)} = \frac{1}{\rho} \left(\frac{1}{L} \sum_{p=1}^{L} \int \mathcal{D}z \int dx_0 [(1 - \rho_0)\delta(x_0) + \rho_0 \phi_0(x_0)] f_a(\hat{Q}_p + \hat{q}_p, \hat{m}_p x_0 + z\sqrt{\hat{q}_p}) \right)$$
$$\times \left(\frac{1}{L} \sum_{p=1}^{L} \int \mathcal{D}z \int dx_0 [(1 - \rho_0)\delta(x_0) + \rho_0 \phi_0(x_0)] \frac{g(\hat{Q}_p + \hat{q}_p, \hat{m}_p x_0 + z\sqrt{\hat{q}_p})}{1 - \rho + \rho g(\hat{Q}_p + \hat{q}_p, \hat{m}_p x_0 + z\sqrt{\hat{q}_p})} \right)^{-1}, \qquad \text{(E4)}$$

$$(\sigma^2)^{(t+1)} = \frac{1}{\rho} \left(\frac{1}{L} \sum_{p=1}^{L} \int \mathcal{D}z \int dx_0 [(1 - \rho_0)\delta(x_0) + \rho_0 \phi_0(x_0)][f_a(\hat{Q}_p + \hat{q}_p, \hat{m}_p x_0 + z\sqrt{\hat{q}_p})^2 + f_c(\hat{Q}_p + \hat{q}_p, \hat{m}_p x_0 + z\sqrt{\hat{q}_p})] \right)$$
$$\times \left(\frac{1}{L} \sum_{p=1}^{L} \int \mathcal{D}z \int dx_0 [(1 - \rho_0)\delta(x_0) + \rho_0 \phi_0(x_0)] \frac{g(\hat{Q}_p + \hat{q}_p, \hat{m}_p x_0 + z\sqrt{\hat{q}_p})}{1 - \rho + \rho g(\hat{Q}_p + \hat{q}_p, \hat{m}_p x_0 + z\sqrt{\hat{q}_p})} \right)^{-1} - [\bar{x}^{(t+1)}]^2, \qquad \text{(E5)}$$

where:

$$\hat{m}_q = \frac{1}{L} \sum_p \frac{J_{pq}\alpha_p}{\Delta + (1/L)\sum_{r=1}^{L} J_{pr} V_r^{(t)}}, \qquad \text{(E6)}$$

$$\hat{q}_q = \frac{1}{L} \sum_p \frac{J_{pq}\alpha_p}{[\Delta + (1/L)\sum_{r=1}^{L} J_{pr} V_r^{(t)}]^2} \frac{1}{L} \sum_s J_{ps} E_s^{(t)}, \qquad \text{(E7)}$$

$$\hat{Q}_q = \frac{1}{L} \sum_p \frac{\alpha_p J_{pq}}{\Delta + (1/L) \sum_{r=1}^{L} J_{pr} V_r^{(t)}} - \hat{q}_q = \hat{m}_q - \hat{q}_q.$$

(E8)

The functions $f_a(X, Y)$, $f_c(X, Y)$ were defined in (B13) and (B14), and the function g is defined as

$$g(X, Y) = \frac{1}{\sqrt{1 + X\sigma^2}} \exp\left[\frac{(Y + \bar{x}/\sigma^2)^2}{2(X + 1/\sigma^2)} - \frac{\bar{x}^2}{2\sigma^2}\right]. \quad (E9)$$

This is the dynamical system that we use in the paper in the noiseless case ($\Delta = 0$) in order to optimize the values of α_1, J_1, and J_2. We can estimate the convergence time of the algorithm as the number of iterations needed in order to reach the successful fixed point (where all E_p and V_p vanish within some given accuracy). Figure 6 shows the convergence time of the algorithm as a function of J_1 and J_2 for Gauss-Bernoulli signals.

The numerical iteration of this dynamical system is fast. It allows one to obtain the theoretical performance that can be achieved in an infinite-N system. We have used it, in particular, to estimate the values of L, α_1, J_1, J_2 that have good performance. For Gauss-Bernoulli signals, using optimal choices of J_1, J_2, we have found that perfect reconstruction can be obtained down to the theoretical limit $\alpha = \rho_0$ by taking $L \to \infty$ (with corrections that scale as $1/L$). Recent rigorous work by Donoho, Javanmard, and

Montanari [38] extends our work and proves our claim that s-BP can reach the optimal threshold asymptotically.

Practical numerical implementation of s-BP matches this theoretical performance only when the size of every block is large enough (few hundreds of variables). In practice, for finite size of the signal, if we want to keep the block-size reasonable, we are hence limited to values of L of several dozens. Thus, in practice, we do not quite saturate the threshold $\alpha = \rho_0$, but exact reconstruction is possible very close to it, as illustrated in Fig. 2, for which the values that we used for the coupling parameters are listed in Table I.

In Fig. 3, we also presented the result of the s-BP reconstruction of the Gaussian signal of density $\rho = 0.4$ for different values of L. We have observed empirically that the result is rather robust for choices of J_1, J_2 and α_1. In this case, in order to demonstrate that very different choices give seeding matrices which are efficient, we used $\alpha_1 = 0.7$; and then $J_1 = 1043$ and $J_2 = 10^{-4}$ for $L = 2$; $J_1 = 631$ and $J_2 = 0.1$ for $L = 5$; $J_1 = 158$ and $J_2 = 4$ for $L = 10$; and $J_1 = 1000$ and $J_2 = 1$ for $L = 20$. One sees that a common aspect to all these choices is a large ratio J_1/J_2. Empirically, this seems to be important in order to ensure a short convergence time. A more detailed study of convergence time of the dynamical system will be necessary in order to give some more systematic rules for choosing the couplings. This study is left for future work.

Even though our theoretical study of the seeded BP was performed on an example of a specific signal distribution, the examples presented in Figs. 1 and 2 show that the performance of the algorithm is robust and also applies to images which are not drawn from that signal distribution.
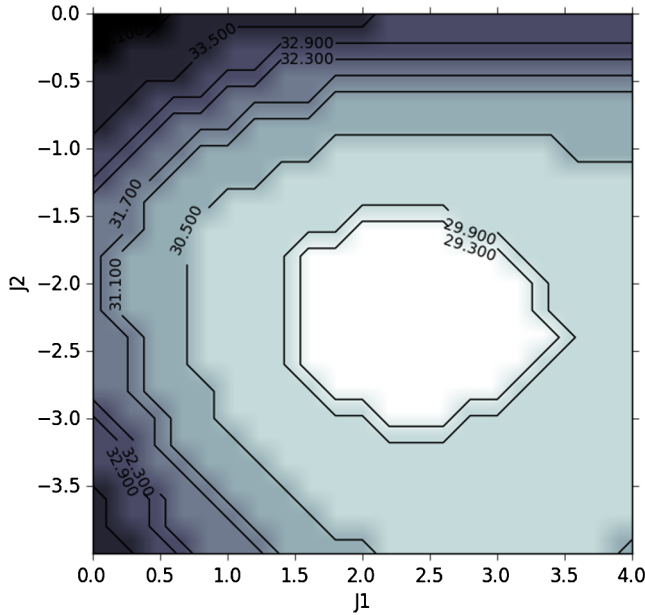

FIG. 6. Convergence time of the s-BP algorithm with $L = 2$ as a function of J_1 and J_2 (in log scale) for Gauss-Bernoulli signal with $\rho_0 = 0.1$. The color represents the number of iterations such that the MSE is smaller than 10^{-8}. The white region gives the fastest convergence. The measurement density is fixed to $\alpha = 0.25$. The parameters in s-BP are chosen as $L = 2$ and $\alpha_1 = 0.3$. The axes are in $\log_{10}$ scale.

TABLE I. Parameters used for the s-BP reconstruction of the Gaussian signal and the binary signal in Fig. 2.

ρ	α	α_1	α'	J_1	J_2	L
Gaussian signal						
0.1	0.130	0.3	0.121	1600	1.44	20
0.2	0.227	0.4	0.218	100	0.64	20
0.3	0.328	0.6	0.314	64	0.16	20
0.4	0.426	0.7	0.412	16	0.16	20
0.6	0.624	0.9	0.609	4	0.04	20
0.8	0.816	0.95	0.809	4	0.04	20
Binary signal						
0.1	0.150	0.4	0.137	64	0.16	20
0.2	0.250	0.6	0.232	64	0.16	20
0.3	0.349	0.7	0.331	16	0.16	20
0.4	0.441	0.8	0.422	16	0.16	20
0.6	0.630	0.95	0.613	16	0.16	20
0.8	0.820	1	0.811	4	0.04	20

APPENDIX F: PHASE DIAGRAM IN THE VARIABLES USED BY DONOHO AND TANNER

We show in Fig. 7 the phase diagram in the convention used by Donoho and Tanner [5], which might be more convenient for some readers.

APPENDIX G: DETAILS ON THE PHANTOM AND LENA EXAMPLES

In this section, we give a detailed description of the way we have produced the two examples of reconstruction in Fig. 1. It is important to stress that this figure is intended to be an illustration of the s-BP reconstruction algorithm. As such, we have used elementary protocols to produce true K-sparse signals and have not tried to optimize the sparsity nor to use the best possible compression algorithm. Instead, we have limited ourselves to the simplest Haar wavelet transform, to make the exact reconstruction and the comparison between the different approaches more transparent.

The Shepp-Logan example (top panels of Fig. 1) is a 128^2 picture that has been generated using the MATLAB implementation. The Lena picture (lower panels of Fig. 1 is a 128^2 crop of the 512^2 gray version of the standard test image. In the first case, we have worked with the sparse one-step Haar transform of the picture, while in the second case, we have worked with a modified picture where we have kept the 24% of largest (in absolute value) coefficients of the two-step Haar transform, while putting all others to zero. The data sets of the two images are available online [6]. Compressed sensing here is done as follows: The original of each image is a vector $\mathbf{o}$ of $N = L^2$ pixels. The unknown vector $\mathbf{x} = \mathbf{Wo}$ are the projections of the original image on a basis of one- or two-step Haar wavelets. It is sparse by construction. We generate a matrix $\mathbf{F}$ as described above, and construct $\mathbf{G} = \mathbf{FW}$. The measurements are obtained by $\mathbf{y} = \mathbf{Go}$, and the linear system for which one does reconstruction is $\mathbf{y} = \mathbf{Fx}$. Once $\mathbf{x}$ has been

TABLE II. Parameters used for the s-BP reconstruction of the Shepp-Logan phantom image and the Lena image of Fig. 1.

α	α_1	α'	J_1	J_2	L
Shepp-Logan phantom image					
0.5	0.6	0.495	20	0.2	45
0.4	0.6	0.395	20	0.2	45
0.3	0.6	0.295	20	0.2	45
0.2	0.6	0.195	20	0.2	45
0.1	0.3	0.095	1	1	30
Lena image					
0.6	0.8	0.585	20	0.1	30
0.5	0.8	0.485	20	0.1	30
0.4	0.8	0.385	20	0.1	30
0.3	0.8	0.285	20	0.1	30
0.2	0.5	0.195	1	1	30

found, the original image is obtained from $\mathbf{o} = \mathbf{W}^{-1}\mathbf{x}$. We used EM-BP and s-BP with a Gauss-Bernoulli $P(\mathbf{x})$.

On the algorithmic side, the EM-BP and ℓ_1 experiments were run with the same full Gaussian random-measurement matrices. The minimization of the ℓ_1 norm was done using the ℓ_1-MAGIC tool for MATLAB [41]. The minimization uses the lowest possible tolerance such that the algorithm outputs a solution. The coupling parameters of s-BP are given in Table II. A small damping was used in each iteration: We mixed the old and new messages, keeping 20% of the messages from the previous iteration. Moreover, because, for both Lena and the phantom, the components of the signal are correlated, we have permuted randomly the columns of the sensing matrix $\mathbf{F}$. This allows us to avoid the (dangerous) situation where a given block contains only zero signal components.

Note that the number of iterations needed by the s-BP procedure to find the solution is moderate. The s-BP algorithm coded in MATLAB finds the image in a few seconds.
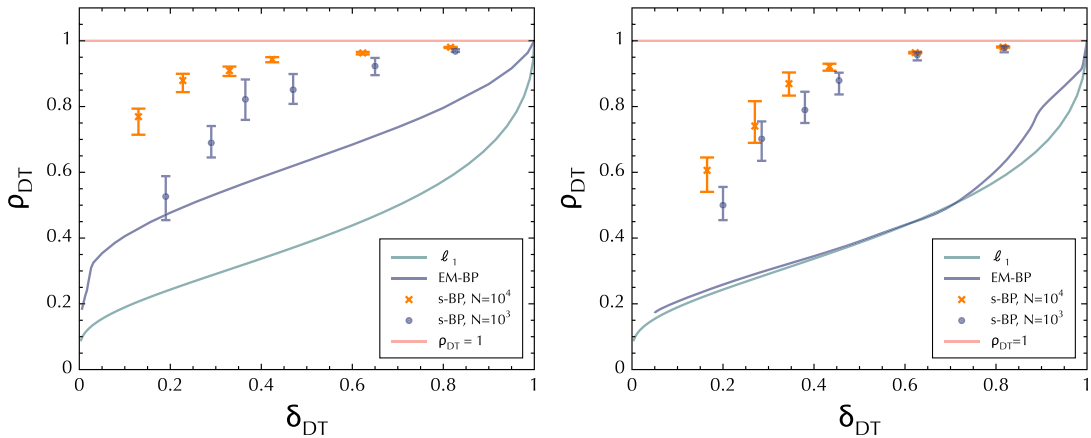

FIG. 7. Same data as in Fig. 2, but using the convention of [5]. The phase diagrams are now plotted as a function of the undersampling ratio $\rho_{DT} = K/M = \rho/\alpha$ and of the oversampling ratio $\delta_{DT} = M/N = \alpha$.

TABLE III. Mean-squared error obtained after reconstruction for the Shepp-Logan phantom image (top) and for the Lena image (bottom), with ℓ_1 minimization, BP, and s-BP.

	$\alpha = 0.5$	$\alpha = 0.4$	$\alpha = 0.3$	$\alpha = 0.2$	$\alpha = 0.1$
ℓ_1	0	0.0055	0.0189	0.0315	0.0537
BP	0	0	0.0213	0.025	0.0489
s-BP	0	0	0	0	0.0412

	$\alpha = 0.6$	$\alpha = 0.5$	$\alpha = 0.4$	$\alpha = 0.3$	$\alpha = 0.2$
ℓ_1	0	4.48×10^{-4}	0.0015	0.0059	0.0928
BP	0	4.94×10^{-4}	0.0014	0.0024	0.0038
s-BP	0	0	0	0	0.0038

For instance, the Lena picture at $\alpha = 0.5$ requires about 500 iterations and about 30 seconds on a standard laptop. Even for the most difficult case of Lena at $\alpha = 0.3$, we need around 2000 iterations, which took only about 2 minutes. In fact, s-BP (coded in MATLAB) is much faster than the MATLAB implementation of the ℓ_1-MAGIC tool on the same machine. We report the MSE for all these reconstruction protocols in Table III.

APPENDIX H: PERFORMANCE OF THE ALGORITHM IN THE PRESENCE OF MEASUREMENT NOISE

A systematic study of our algorithm for the case of noisy measurements can be performed using the replica analysis, but that study goes beyond the scope of the present work. In this section, we do, however, want to point out two important facts: (1) The modification of our algorithm to take the noise into account is straightforward, and (2) The results that we have obtained are robust to the presence of a

small amount of noise. As shown in Appendix B, the probability of **x** after the measurement of **y** is given by

$$
\hat{P}(\mathbf{x}) = \frac{1}{Z} \prod_{i=1}^{N} [(1 - \rho)\delta(x_i) + \rho\phi(x_i)]
$$
$$
\times \prod_{\mu=1}^{M} \frac{1}{\sqrt{2\pi\Delta_\mu}} e^{-(1/2\Delta_\mu)(y_\mu - \sum_{i=1}^{N} F_{\mu i} x_i)^2}. \quad \text{(H1)}
$$

Δ_μ is the variance of the Gaussian noise in measurement μ. For simplicity, we consider that the noise is homogeneous, i.e., $\Delta_\mu = \Delta$, for all μ, and discuss the result in the unit of standard deviation $\sqrt{\Delta}$. The AMP-form of the message passing including noise has already been given in Appendix C. The variance of the noise, Δ, can also be learned via the expectation-maximization approach, which reads:

$$
\Delta = \frac{\sum_\mu \frac{(y_\mu - \omega_\mu)^2}{(1 + \frac{1}{\Delta}\gamma_\mu)^2}}{\sum_\mu \frac{1}{1 + \frac{1}{\Delta}\gamma_\mu}}. \quad \text{(H2)}
$$

The dynamical system describing the density evolution has been written in Appendix E. It just needs to be complemented by the iteration on Δ according to Eq. (H2). We have used this dynamical system to study the evolution of MSE in the case where both $P(x)$ and signal distribution are Gauss-Bernoulli with density ρ, zero mean, and unit variance for the full measurement matrix and for the seeding one. Figure 8 shows the MSE as a function of α for a given Δ, using our algorithm, and the ℓ_1 minimization for comparison. The analytical results obtained by the study of the dynamical system are compared to numerical simulations, and they agree very well.
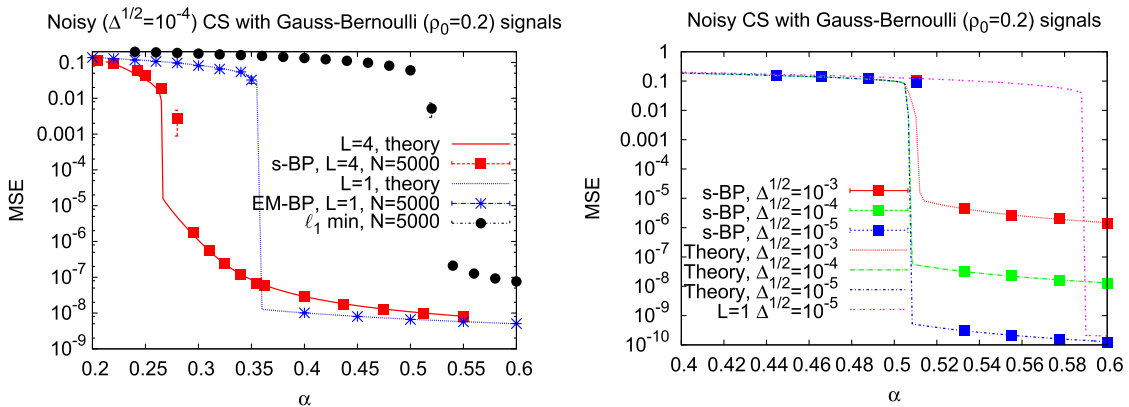

FIG. 8. Mean-squared error as a function of α for different values of the measurement noise. Left: Gauss-Bernoulli signal with $\rho_0 = 0.2$, and measurement noise with standard deviation $\sqrt{\Delta} = 10^{-4}$. The EM-BP ($L = 1$) and s-BP strategy ($L = 4$, $J_1 = 20$, $J = 0.1$, $\alpha_1 = 0.4$) are able to perform a very good reconstruction up to much lower value of α than the ℓ_1 procedure. Below a critical value of α, these algorithms show a first-order phase transition to a regime with much larger MSE. Right: Gauss-Bernoulli signal with $\rho_0 = 0.4$, with a noise with standard deviation $\sqrt{\Delta} = 10^{-3}$, 10^{-4}, 10^{-5}. s-BP, with $L = 9$, $J_1 = 30$, $J_2 = 8$, $\alpha_1 = 0.8$ decodes very well for all of these values of noises. In this case, ℓ_1 is unable to reconstruct for all $\alpha < 0.75$, well outside the range of this plot.

[1] E. J. Candès and M. B. Wakin, *An Introduction to Compressive Sampling*, IEEE Signal Process. Mag. **25**, 21 (2008).

[2] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, Ting Sun, K. F. Kelly, and R. G. Baraniuk, *Single-Pixel Imaging via Compressive Sampling*, IEEE Signal Process. Mag. **25**, 83 (2008).

[3] E. J. Candès and T. Tao, *Decoding by Linear Programming*, IEEE Trans. Inf. Theory **51**, 4203 (2005).

[4] D. L. Donoho, *Compressed Sensing*, IEEE Trans. Inf. Theory **52**, 1289 (2006).

[5] D. L. Donoho and J. Tanner, *Neighborliness of Randomly Projected Simplices in High Dimensions*, Proc. Natl. Acad. Sci. U.S.A. **102**, 9452 (2005).

[6] *Applying Statistical Physics to Inference in Compressed Sensing*, http://aspics.krzakala.org/.

[7] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin-Glass Theory and Beyond*, Lecture Notes in Physics, Vol. 9 (World Scientific, Singapore, 1987).

[8] Y. Wu and S. Verdu, *MMSE Dimension*, IEEE Trans. Inf. Theory **57**, 4857 (2011).

[9] Y. Wu and S. Verdu, *Optimal Phase Transitions in Compressed Sensing*, arXiv:1111.6822v1.

[10] D. Guo, D. Baron, and S. Shamai, in *47th Annual Allerton Conference on Communication, Control, and Computing, Monticello, Illinois* (IEEE, New York, 2009), pp. 52–59.

[11] D. L. Donoho, A. Maleki, and A. Montanari, *Message-Passing Algorithms for Compressed Sensing*, Proc. Natl. Acad. Sci. U.S.A. **106**, 18914 (2009).

[12] Y. Kabashima, T. Wadayama, and T. Tanaka, *A Typical Reconstruction Limit of Compressed Sensing Based on L_p-Norm Minimization*, J. Stat. Mech. (2009) L09003.

[13] S. Rangan, A. Fletcher, and V. Goyal, *Asymptotic Analysis of MAP Estimation via the Replica Method and Applications to Compressed Sensing*, arXiv:0906.3234v2.

[14] S. Ganguli and H. Sompolinsky, *Statistical Mechanics of Compressed Sensing*, Phys. Rev. Lett. **104**, 188701 (2010).

[15] J. Pearl, in *Proceedings of the American Association of Artificial Intelligence National Conference on AI* (Pittsburgh, 1982), ISBN 978-1-57735-192-4, pp. 133–136.

[16] T. Richardson and R. Urbanke, *Modern Coding Theory* (Cambridge University Press, Cambridge, England, 2008).

[17] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University, Oxford, England, 2009).

[18] A. Jimenez Felstrom and K. Zigangirov, *Time-Varying Periodic Convolutional Codes with Low-Density Parity-Check Matrix*, IEEE Trans. Inf. Theory **45**, 2181 (1999).

[19] S. Kudekar, T. Richardson, and R. Urbanke, in *Proceedings of the 2010 IEEE International Symposium on Information Theory Proceedings (ISIT), Austin, Texas* (IEEE, New York, 2010), pp. 684–688.

[20] M. Lentmaier and G. Fettweis, in *Proceedings of the 2010 IEEE International Symposium on Information Theory Proceedings (ISIT), Austin, Texas* (IEEE, New York, 2010), pp. 709–713.

[21] S. Hassani, N. Macris, and R. Urbanke, in *IEEE Information Theory Workshop (ITW) 2010*, doi:10.1109/CIG.2010.5592881 (IEEE, New York, 2010), pp. 1–5.

[22] D. Donoho, A. Maleki, and A. Montanari, in *IEEE Information Theory Workshop (ITW) 2010*, doi:10.1109/ITWKSPS.2010.5503228 (IEEE, New York, 2010), pp. 1–5.

[23] S. Rangan, in *Proceedings of the 2011 IEEE International Symposium on Information Theory Proceedings (ISIT), Austin, Texas* (IEEE, New York, 2011), pp. 2168–2172.

[24] S. Rangan, in *44th Annual Conference on Information Sciences and Systems (CISS), 2010* (IEEE, New York, 2010), pp. 1–6.

[25] For proper normalization, we regularize the linear constraint to $e^{-(\mathbf{y}-\mathbf{Fx})^2/(2\Delta)}/(2\pi\Delta)^{M/2}$, where Δ can be seen as a variance of noisy measurement, and consider the limit $\Delta \to 0$; see Appendix A.

[26] B. K. Natarajan, *Sparse Approximate Solutions to Linear Systems*, SIAM J. Comput. **24**, 227 (1995).

[27] D. J. Thouless, P. W. Anderson, and R. G. Palmer, *Solution of "Solvable Model of a Spin-Glass"*, Philos. Mag. **35**, 593 (1977).

[28] T. A. Tanaka, *A Statistical-Mechanics Approach to Large-System Analysis of CDMA Multiuser Detectors*, IEEE Trans. Inf. Theory **48**, 2888 (2002).

[29] D. Guo and C.-C. Wang, in *Information Theory Workshop (ITW) 2006, Chengdu* (IEEE, New York, 2006), pp. 194–198.

[30] F. R. Kschischang, B. Frey, and H.-A. Loeliger, *Factor Graphs and the Sum-Product Algorithm*, IEEE Trans. Inf. Theory **47**, 498 (2001).

[31] J. Yedidia, W. Freeman, and Y. Weiss, in *Exploring Artificial Intelligence in the New Millennium* (Morgan Kaufmann, San Francisco, 2003), pp. 239–236.

[32] A. Montanari and M. Bayati, IEEE Trans. Inf. Theory **57**, 764 (2011).

[33] A. Dempster, N. Laird, and D. Rubin, *Maximum Likelihood from Incomplete Data via the EM Algorithm*, J. R. Stat. Soc. Ser. B. Methodol. **39**, 1 (1977) [http://www.jstor.org/stable/2984875].

[34] Y. Iba, *The Nishimori Line and Bayesian Statistics*, J. Phys. A **32**, 3875 (1999).

[35] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, *Inference and Phase Transition in the Detection of Modules in Sparse Networks*, Phys. Rev. Lett. **107**, 065701 (2011).

[36] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing* (Oxford University Press, Oxford, England, 2001).

[37] D. Guo and S. Verdú, *Randomly Spread CDMA: Asymptotics via Statistical Physics*, IEEE Trans. Inf. Theory **51**, 1983 (2005).

[38] D. L. Donoho, A. Javanmard, and A. Montanari, *Information-Theoretically Optimal Compressed Sensing via Spatial Coupling and Approximate Message Passing*, arXiv:1112.0708v1.

[39] V. A. Marčenko and L. A. Pastur, *Distribution of Eigenvalues for Some Sets of Random Matrices*, Mathematics of the USSR-Sbornik **1**, 457 (1967) [http://iopscience.iop.org/0025-5734/1/4/A01/].

[40] J. Hubbard, *Calculation of Partition Functions*, Phys. Rev. Lett. **3**, 77 (1959); R. L. Stratonovich, *On a Method of Calculating Quantum Distribution Functions*, Sov. Phys. Dokl. **2**, 416 (1958).

[41] The ℓ_1-MAGIC tool is a free implementation that can be downloaded from http://users.ece.gatech.edu/~justin/l1magic/.