# Analysis of common attacks in public-key cryptosystems based on low-density parity-check codes

N. S. Skantzos,[1,2,*] D. Saad,[1,†] and Y. Kabashima[3,‡]
[1]*Neural Computing Research Group, Aston University, Aston Triangle, Birmingham B4 7ET, United Kingdom*
[2]*Institut for Theoretical Physics, Celestijnenlaan 200D, KULeuven, Leuven B-3001, Belgium*
[3]*Department of Computational Intelligence & Systems Science, Tokyo Institut of Technology, Yokohama 2268502, Japan*
(Received 8 May 2003; published 25 November 2003)

We analyze the security and reliability of a recently proposed class of public-key cryptosystems against attacks by unauthorized parties who have acquired partial knowledge of one or more of the private key components and/or of the plaintext. Phase diagrams are presented, showing critical partial knowledge levels required for unauthorized decryption.

DOI: 10.1103/PhysRevE.68.056125 PACS number(s): 89.70.+c, 03.67.Dd, 05.50.+q

## I. INTRODUCTION

An important aspect in many modern communication systems is the ability to exclude unauthorized parties from gaining access to confidential material. Although cryptosystems in general have an extensive history, until fairly recently they have been based on simple variations of the same theme: information security among authorized parties relies on sharing a secret key which is to be used for encryption and decryption of transmitted messages. While in this way confidentiality of the sent message may be secured, such systems suffer from the (obvious) drawback of nonsecure key distribution.

In 1978 Rivest, Shamir, and Adleman first devised a way to resolve this problem which led to the celebrated RSA *public-key* cryptosystem [1] (for historical accuracy, a similar system was suggested years earlier in the British GCHQ but was kept secret). The idea behind public-key cryptosystems is to differentiate between the encryption and decryption keys; private key(s) are assigned to authorized users, for decryption purposes, while transmitting parties only need to know the matching encryption (public) key [2]. The two keys are related by a function which generates the encryption mechanism from the decryption key with low computational costs, while the opposite operation (evaluating the decryption key from the encryption mechanism) is computationally infeasible. Such functions are called ''one-way'' or trapdoor functions; the RSA algorithm, for instance, is based on the intractability of factorizing large integers generated by taking the product of two large prime numbers.

The proliferation of digital communication in the last few decades has brought in a demand for secure communication leading to the invention of several other public-key cryptosystems, most notable of which are the El-Gammal cryptosystem (based on the discrete logarithm problem), systems based on elliptic curves, and the McEliece cryptosystem (based on linear error-correcting codes) [3]. A common denominator of all public-key algorithms is the high computational complexity of the task facing the unauthorized user; this is typically related to hard computational problems that cannot be solved in practical time scales.

A new public-key cryptosystem based on a diluted Ising spin-glass system has been recently proposed in Ref. [4]. The suggested cryptosystem is similar in spirit to that of McEliece and relies on exploiting physical properties of the MacKay-Neal (MN) low-density parity-check (LDPC) error-correcting codes. In particular, in the context of MN codes it has been shown [4–6] that for certain parameter values successful decoding is highly likely, while for others (particularly when the number of parity checks per bit and the number of bits per check tend to infinity) the ''perfect'' solution, describing full retrieval of the sent message, admits only a very narrow basin of attraction; iterative algorithmic solutions lead in this case, almost certainly, to a decryption failure. One can use these properties to devise a LDPC based cryptosystem [4]. The narrow basin of attraction ensures that a random initialization of the decryption equations will fail to converge to the plaintext solution while the naive approach of trying all possible initializations is clearly doomed for a sufficiently large plaintext size. The ''one-way'' function relies on the hard computational task of decomposing a dense matrix (the public key) into a combination of sparse and dense matrices (private keys) [7].

In this paper we examine the suggested cryptosystem from an adversary's viewpoint. We consider an unauthorized party that has acquired partial or full knowledge of one or more of the private keys, and/or of the message, and we evaluate the critical knowledge levels required for unauthorized decryption. In addition, we examine the decryption reliability by authorized users due to the probabilistic nature of the cryptosystem.

The paper is organized as follows. In the following section we give an outline of the suggested cryptosystem. In Sec. III we formulate unauthorized-decryption scenarios with partial knowledge based on a statistical mechanical framework. In Sec. IV we derive the observable quantity that measures decryption success of the unauthorized user as a function of the attack parameters and in Sec. V we examine various cases and present numerical results as well as the related phase diagrams. In Secs. VI and VII we briefly study the basin of attraction of the ferromagnetic solution, and the reliability of the decryption mechanism (for authorized us-

*Email address: skantzon@aston.ac.uk
†Email address: saadd@aston.ac.uk
‡Email address: kaba@dis.titech.ac.jp

 

ers), respectively. The implications of the analysis are discussed in Sec. VIII.

## II. DESCRIPTION OF THE CRYPTOSYSTEM

The cryptosystem suggested in Ref. [4] is based on the framework of MN error-correcting codes [5]. An outline of the encryption/decryption process is as follows.

A plaintext represented by $\boldsymbol{\xi} \in \{0,1\}^N$ is encrypted to the ciphertext $\boldsymbol{r} \in \{0,1\}^M$ (with $M > N$) using a predetermined generator matrix $G \in \{0,1\}$ and a corrupting vector $\boldsymbol{\zeta} \in \{0,1\}^M$ with $P(\zeta_i) = p\,\delta_{\zeta_i,1} + (1-p)\,\delta_{\zeta_i,0}$ for each component $1 \leqslant i \leqslant M$; the Kronecker tensor $\delta_{ab}$ returns 1 when the arguments are equal ($a = b$) and zero otherwise. The generated ciphertext is of the form

$$\boldsymbol{r} = G\boldsymbol{\xi} + \boldsymbol{\zeta} \pmod 2. \tag{1}$$

The $(M \times N)$ matrix $G$ together with the corruption rate $p \in [0,1]$ constitutes the *public* key.

The encryption matrix $G$ is constructed by choosing a dense matrix $D$ (of dimensionality $N \times N$) and two randomly selected sparse matrices $A$ (of dimensionality $M \times N$) and $B$ (of dimensionality $M \times M$) through $G = B^{-1}AD \pmod 2$. The matrices $A$ and $B$ are characterized by $K$ and $L$ nonzero elements per row and $C$ and $L$ nonzero elements per column, respectively (*irregular* constructions with values that vary from column to column or row to row may also be considered). The resulting dense matrix $G$ is *modeled* as being characterized by $K'$ and $C'$ nonzero elements per row and per column, respectively, with $K', C' \to \infty$ (while $K'/C' = N/M$ is finite). In fact, the dense matrix $G$ is of an irregular form due to the inverse of the sparse matrix $B$ as well as the product taken with the dense matrix $D$; we will model the matrix $G$ by a regular dense matrix to simplify the analysis. The parameters $K$, $C$, and $L$ define a particular cryptosystem while the matrices $A$, $B$, and $D$ constitute the *private* key.

The authorized user may obtain the plaintext from the received ciphertext $\boldsymbol{r}$ by taking the (mod 2) product $B\boldsymbol{r} = A\hat{\boldsymbol{\xi}} + B\boldsymbol{\zeta}$, where $\hat{\boldsymbol{\xi}} = D\boldsymbol{\xi}$. Finding a set of solutions $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ such that the equation

$$A\boldsymbol{\sigma} + B\boldsymbol{\tau} = A\hat{\boldsymbol{\xi}} + B\boldsymbol{\zeta} \pmod 2 \tag{2}$$

is true will lead to candidate solutions of the decryption problem (of which the most probable one will be detected according to a further selection criterion). This will be followed by a product with $D^{-1}$ to obtain the original plaintext. For particular choices of $K$ and $L$, solving the above equation can be achieved via iterative methods which have common roots in both graphical models and physics of disordered systems such as belief propagation [5], belief revision [8], and more recently survey propagation [9]; where state probabilities for the decrypted message bits $P(\boldsymbol{\sigma}, \boldsymbol{\tau} | \boldsymbol{r})$ are calculated by solving iteratively a set of coupled equations, describing conditional probabilities of the ciphertext bits given the plaintext and vice versa. This problem is identical to the decoding problem of a regular MN error-correcting code; for the explicit iterative decoding equations see Eqs. (54) and (55) as well as Refs. [5,10].

The unauthorized user, on the other hand, faces the task of finding the most probable solutions to the equation

$$G\boldsymbol{\xi} + \boldsymbol{\zeta} = G\boldsymbol{\sigma} + \boldsymbol{\tau} \pmod 2. \tag{3}$$

The above decryption equation is effectively identical to the decoding problem of Sourlas error-correcting codes [11], with the public matrix $G$ being dense. Most notably, in the context of Sourlas codes, finding solutions to Eq. (3) is strongly dependent on initial conditions: for all initial conditions other than the plaintext itself, the iterative equations of belief propagation will fail to converge to the plaintext solution [4–6,12] such that obtaining the correct solution for Eq. (3) without knowledge of the private key will become infeasible. Obtaining the private keys by decomposing $G$ into $A$, $B$, and $D$ is known to be a hard computational problem even if the values of $K$, $C$, and $L$ are known [7].

We would like to point to the fact that there may exist more than one triplet of matrices $\{A,B,D\}$ such that $G = B^{-1}AD$. With $D$ being a dense matrix, finding a set of matrices $A'$, $B'$, and $D'$ such that their combination produces $G = (B')^{-1}A'D'$ requires an exponentially diverging number of operations, with respect to the system size, making the decomposition computationally infeasible. For $D = \boldsymbol{1}$ (as was the original formulation in Ref. [4]) finding a pair of sparse matrices $A'$ and $B'$ such that $G = (B')^{-1}A'$ requires only a number of operations that is polynomial in $N$, and the cryptosystem is therefore not secure.

Other advantages and drawbacks of the new cryptosystem appear in Ref. [4].

## III. FORMULATION OF THE ATTACK

An essential ingredient of any cryptosystem is a certain level of robustness against attacks. The robustness of the current cryptosystem against attacks with no additional secret information has already been reported in Ref. [4]. In this section we study the vulnerability of the new cryptosystem to various attacks, characterized by partial knowledge of the secret keys and/or the plaintext itself; the additional information manifests itself in a set of decryption equations similar to Eq. (2) in which partial information of the secret keys (and plaintext) is used in conjunction with the publicly available information of Eq. (3). The explicit knowledge of the matrix $D$ required for the final step (an additional hurdle for a potential attacker) has been all but ignored in the analysis, as we focus on the feasibility of the decryption operation itself.

The cumulative information provided by the different sets of equations will potentially allow for a successful decryption. To this extent, knowledge of the matrix $B$ is of utmost importance since obtaining partial knowledge of the syndrome vector and Eq. (2) is only accessible through decryption using the matrix $B$. Let us consider that an unauthorized user has acquired knowledge of a number of rows $\gamma_A M$, $\gamma_B M$, and $\gamma_D M$ of the secret matrices $A$, $B$, and $D^{-1}$ (with $\gamma_\star \in [0,1]$). Relation (2) then provides $\gamma M$

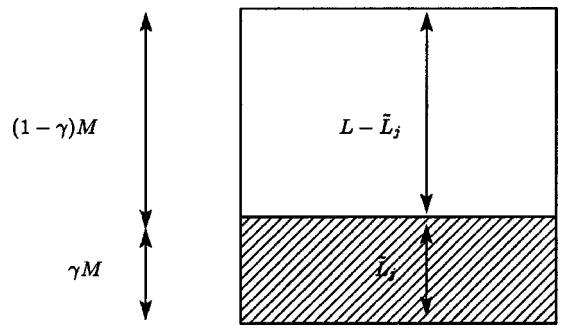


FIG. 1. The matrix $B$ of dimensionality $M\times M$ used as a private key in decryption. The scenario we consider here is that unauthorized users have acquired knowledge of $\gamma M$ rows of the matrix. The $(\gamma M\times M)$ block may have $\tilde{L}_j=0,\ldots,L$ nonzero elements per column for all $j$.

$\equiv\min\{\gamma_A,\gamma_B,\gamma_D\}M$ decryption equations (4) based on sparse matrices. To analyze the attack we will thus from now on assume that a block $(\gamma M\times M)$ of all matrices is known to the unauthorized user with $\gamma\in[0,1]$ (Fig. 1). In this case, the products $\Sigma_{j=1}^{M}B_{ij}r_j$ for $i=1,\ldots,\gamma M$ can be taken and the unauthorized user will arrive at the following decryption problem:

$$\text{private: } (\hat{A}\boldsymbol{\sigma})_i+(\hat{B}\boldsymbol{\tau})_i=(\hat{A}\boldsymbol{\xi})_i+(\hat{B}\boldsymbol{\zeta})_i \quad \text{for rows } i=1,\ldots,\gamma M, \tag{4}$$

,

$$\text{public: } (G\boldsymbol{\sigma})_i+(\boldsymbol{I}\boldsymbol{\tau})_i=(G\boldsymbol{\xi})_i+(\boldsymbol{I}\boldsymbol{\zeta})_i \quad \text{for rows } i=1,\ldots,M, \tag{5}$$

where we absorbed the matrix $D$ using $\boldsymbol{\sigma}\to D\boldsymbol{\sigma}$ and $\boldsymbol{\xi}\to D\boldsymbol{\xi}$; in practice, after decryption, one will have to use the inverted matrix $D^{-1}$, or part of it, to obtain the original plaintext itself (rather than its rotated version $D\boldsymbol{\xi}$). All solutions $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ will have to simultaneously satisfy Eqs. (4) and (5). The matrices $\hat{A}$ and $\hat{B}$ will be described by $K$ and $L$ nonzero elements per row. The average number of known nonzero elements per column in $\hat{A}$ and $\hat{B}$ will be denoted $\overline{C}$ and $\overline{L}$, respectively. Since $\gamma$ is the probability of selecting a nonzero element in the known part of the private key it follows that $\overline{C}=\gamma C$ and $\overline{L}=\gamma L$. For all columns $j=1,\ldots,M$ we will denote the number of nonzero elements in $\hat{A}$ and $\hat{B}$ by the random variables $\tilde{C}_j(=\Sigma_{i=1}^{\gamma M}\hat{A}_{ij})$ and $\tilde{L}_j(=\Sigma_{j=1}^{\gamma M}\hat{B}_{ij})$ which are described by the distributions

$$P(\tilde{C}_j;C)=\binom{C}{\tilde{C}_j}\gamma^{\tilde{C}_j}(1-\gamma)^{C-\tilde{C}_j},\quad \tilde{C}_j=0,\ldots,C, \tag{6}$$

$$P(\tilde{L}_j;L)=\binom{L}{\tilde{L}_j}\gamma^{\tilde{L}_j}(1-\gamma)^{L-\tilde{L}_j},\quad \tilde{L}_j=0,\ldots,L. \tag{7}$$

To facilitate the statistical mechanical description we will now replace the field $\{0,1;+(\mathrm{mod}\ 2)\}$ by the more familiar Ising spin representation [11] $\{-1,1;\times\}$. Equations (4) and (5) will also be modified. From the matrices $\hat{A},\hat{B}$, and $G,\boldsymbol{I}$ we construct the binary tensors $\mathcal{A}=\{\mathcal{A}_{\langle i_1,\ldots,i_K;j_1,\ldots,j_L\rangle};1\leqslant i_1<\cdots<i_K\leqslant N,1\leqslant j_1<\cdots<j_L\leqslant M\}$ and $\mathcal{G}=\{\mathcal{G}_{\langle i_1,\ldots,i_{K'};j\rangle};1\leqslant i_1<\cdots<i_{K'}\leqslant N,1\leqslant j\leqslant M\}$. The elements of these tensors are $\mathcal{A}_{\langle i_1\ldots i_K;j_1\ldots j_L\rangle}=1$ if $\hat{A}$ and $\hat{B}$ have, respectively, a row in which the elements $\{i_1,\ldots,i_K\}$ and $\{j_1,\ldots,j_L\}$ are all 1 and 0 otherwise. Similarly, $\mathcal{G}_{\langle i_1,\ldots,i_{K'};j\rangle}=1$ if $G$ and $\boldsymbol{I}$ have, respectively, a row in which the elements $\{i_1,\ldots,i_{K'}\}$ and $\{j\}$ are all 1 and 0 otherwise. The notation we used to indicate tensor elements, $\langle i_1\ldots i_K\rangle$, denotes that the sites $i_1,\ldots,i_K$ are ordered and different.

The fact that the number of nonzero elements per column in $\hat{A},\hat{B}$ and $G,\boldsymbol{I}$, respectively, are $\tilde{C}_i,\tilde{L}_i$ and $C',1$, for all columns, will be imposed by the constraints

$$\sum_{i_2,\ldots,i_K;j_1,\ldots,j_L}\mathcal{A}_{\langle i_1,\ldots,i_K;j_1,\ldots,j_L\rangle}=\tilde{C}_{i_1}\quad\forall\ i_1=1,\ldots,M, \tag{8}$$

$$\sum_{i_1,\ldots,i_K;j_2,\ldots,j_L}\mathcal{A}_{\langle i_1,\ldots,i_K;j_1,\ldots,j_L\rangle}=\tilde{L}_{j_1}\quad\forall\ j_1=1,\ldots,M, \tag{9}$$

$$\sum_{i_2,\ldots,i_{K'};j}\mathcal{G}_{\langle i_1,\ldots,i_{K'};j\rangle}=C'\quad\forall\ i_1=1,\ldots,M, \tag{10}$$

$$\sum_{i_1,\ldots,i_{K'}}\mathcal{G}_{\langle i_1,\ldots,i_{K'};j\rangle}=1\quad\forall\ j=1,\ldots,M. \tag{11}$$

To compress notation in what follows we will denote the set of indices involved in the tensors $\mathcal{A}$ and $\mathcal{G}$ by $\Lambda_K=\langle i_1,\ldots,i_K\rangle$ and $\Omega_L=\langle j_1,\ldots,j_L\rangle$.

For the system described in Eqs. (4) and (5) the microscopic state probability $P(\boldsymbol{\sigma},\boldsymbol{\tau})$ can be written as

$$P(\boldsymbol{\sigma},\boldsymbol{\tau}|\boldsymbol{\xi},\boldsymbol{\zeta},\mathcal{A},\mathcal{G})=\frac{1}{Z}[\Delta(\boldsymbol{\sigma},\boldsymbol{\tau};\boldsymbol{\xi},\boldsymbol{\zeta},\mathcal{A})\Delta(\boldsymbol{\sigma},\boldsymbol{\tau};\boldsymbol{\xi},\boldsymbol{\zeta},\mathcal{G})\times\Phi(\boldsymbol{\sigma};\boldsymbol{\xi})\Phi(\boldsymbol{\tau};\boldsymbol{\zeta})]e^{-\beta H(\boldsymbol{\sigma},\boldsymbol{\tau})} \tag{12}$$

(notice that the dependence on $\boldsymbol{\xi},\boldsymbol{\zeta}$ is not explicit, but through the received vector $\boldsymbol{r}$) where $Z$ is the partition function and $H(\boldsymbol{\sigma},\boldsymbol{\tau})$ the energy

$$H(\boldsymbol{\sigma},\boldsymbol{\tau})=-F_\sigma\sum_{i=1}^{N}\sigma_i-F_\tau\sum_{j=1}^{M}\tau_j \tag{13}$$

with $F_\sigma=\frac{1}{2}\ln(1-p_\sigma)/p_\sigma$ and $F_\tau=\frac{1}{2}\ln(1-p_\tau)/p_\tau$. The fields $F_\sigma$ and $F_\tau$ represent prior knowledge of the statistics from which the plaintext and the corrupting vector are drawn, such that

$$P(\xi_i)=(1-p_\sigma)\,\delta_{\xi_i,1}+p_\sigma\delta_{\xi_i,-1},\quad p_\sigma\in[0,1], \tag{14}$$

$$P(\zeta_j)=(1-p_\tau)\,\delta_{\zeta_j,1}+p_\tau\delta_{\zeta_j,-1},\quad p_\tau\in[0,1]. \tag{15}$$

The indicator functions $\Delta(\boldsymbol{\sigma},\boldsymbol{\tau};\boldsymbol{\xi},\boldsymbol{\zeta},\mathcal{A})$ and $\Delta(\boldsymbol{\sigma},\boldsymbol{\tau};\boldsymbol{\xi},\boldsymbol{\zeta},\mathcal{G})$ restrict the space of solutions $\boldsymbol{\sigma}\in\{-1,1\}^N$ and $\boldsymbol{\tau}\in\{-1,1\}^M$ to those that obey Eqs. (4) and (5):

$$\Delta(\boldsymbol{\sigma},\boldsymbol{\tau};\boldsymbol{\xi},\boldsymbol{\zeta},\mathcal{A})=\prod_{\Lambda_K\Omega_L}\left[1+\frac{1}{2}\mathcal{A}_{\Lambda_K\Omega_L}\times\left(\prod_{i\in\Lambda_K}\sigma_i\xi_i\prod_{j\in\Omega_L}\tau_j\zeta_j-1\right)\right], \tag{16}$$

$$\Delta(\boldsymbol{\sigma},\boldsymbol{\tau};\boldsymbol{\xi},\boldsymbol{\zeta},\mathcal{G})=\prod_{\Lambda_{K'}\Omega_{L'}}\left[1+\frac{1}{2}\mathcal{G}_{\Lambda_{K'}\Omega_{L'}}\times\left(\prod_{i\in\Lambda_{K'}}\sigma_i\xi_i\prod_{j\in\Omega_{L'}}\tau_j\zeta_j-1\right)\right], \tag{17}$$

and finally the terms $\Phi(\cdots)\in\{0,1\}$ correspond to

$$\Phi(\boldsymbol{\sigma};\boldsymbol{\xi})=\prod_{i=1}^{N}\left[(1-c_i)+c_i\delta_{\sigma_i,\xi_i}\right], \tag{18}$$

$$\Phi(\boldsymbol{\tau};\boldsymbol{\zeta})=\prod_{i=1}^{M}\left[(1-d_i)+d_i\delta_{\tau_i,\zeta_i}\right], \tag{19}$$

where the quenched variables $c_i,d_j\in\{0,1\}$ model prior knowledge of bits of the plaintext and the corrupting vector such that if for some $i$ the plaintext bit $\xi_i$ is known then the thermal variable $\sigma_i$ takes the quenched plaintext value (and similarly for the corruption vector $\zeta_j$ and $\tau_j$). For the distribution of $c_i$ and $d_j$ we will consider

$$P(c_i)=w_\sigma\,\delta_{c_i,1}+(1-w_\sigma)\,\delta_{c_i,0}\quad w_\sigma\in[0,1], \tag{20}$$

$$P(d_j)=w_\tau\delta_{d_j,1}+(1-w_\tau)\,\delta_{d_j,0}\quad w_\tau\in[0,1]. \tag{21}$$

The system described by Eq. (12) represents a set of variables interacting via multispin ferromagnetic couplings of finite connectivity, represented by a combination of matrices, in the presence of the random fields $\xi_iF_\sigma$ and $\zeta_jF_\tau$. At $\beta=1$ (which corresponds to the Nishimori temperature [13]) we will evaluate the free energy per plaintext bit

$$f=-\lim_{N\to\infty}\frac{1}{\beta N}\langle\ln Z\rangle_\Gamma. \tag{22}$$

The macroscopic observable we are interested in calculating is the overlap $m=\lim_{N\to\infty}(1/N)\Sigma_i\xi_i\hat{\xi}_i$ between the plaintext and the bayes marginal posterior maximizer (MPM) estimate of the plaintext $\hat{\xi}_i\equiv\mathrm{sgn}\Sigma_{\sigma_i=\pm}\sigma_i\,p(\sigma_i|\boldsymbol{r})$ where $p(\sigma_i|\boldsymbol{r})$ is the microscopic state probability (12). Disorder averages $\langle\rangle_\Gamma$ are taken over the probability distributions (14), (15), (20), (21) and over the distribution of the tensors $\mathcal{A}$ and $\mathcal{G}$ obeying constrains (8)–(11):

$$\langle\mathcal{F}(\mathcal{A})\rangle_{\mathcal{A},\{\tilde{C}_i,\tilde{L}_i\}}=\frac{1}{\mathcal{N}}\sum_{\{\mathcal{A}_{\Lambda_K\Omega_L}\}}\prod_{i=1}^{N}\left\langle\delta\left[\sum_{\Lambda_K\Omega_{L/i\in\Lambda_K}}\mathcal{A}_{\Lambda_K\Omega_L}-\tilde{C}_i\right]\right\rangle_{P(\tilde{C}_i)}\times\prod_{j=1}^{M}\left\langle\delta\left[\sum_{\Lambda_K\Omega_{L/j\in\Omega_L}}\mathcal{A}_{\Lambda_K\Omega_L}-\tilde{L}_j\right]\right\rangle_{P(\tilde{L}_j)}\mathcal{F}(\mathcal{A}), \tag{23}$$

$$\langle\mathcal{F}(\mathcal{G})\rangle_{\mathcal{G}}=\frac{1}{\mathcal{N}'}\sum_{\{\mathcal{G}_{\Lambda_{K'}\Omega_{L'}}\}}\prod_{i=1}^{N}\delta\left[\sum_{\Lambda_{K'}\Omega_{L'/i\in\Lambda_{K'}}}\mathcal{A}_{\Lambda_{K'}\Omega_{L'}}-C'\right]\times\prod_{j=1}^{M}\delta\left[\sum_{\Lambda_{K'}\Omega_{L'/j_1\in\Omega_{L'}}}\mathcal{G}_{\Lambda_{K'}\Omega_{L'}}-1\right]\mathcal{F}(\mathcal{G}), \tag{24}$$

where $\mathcal{N}$ and $\mathcal{N}'$ are the corresponding normalization constants.

The parameters $w_\sigma,w_\tau,F_\sigma,F_\tau$, and $\gamma$ describe the attack characteristics.

## IV. THE FREE ENERGY AND DECRYPTION OBSERVABLES

The calculation generally follows that of Refs. [6,10]. To perform the various disorder averages we begin by invoking the replica identity $\langle\ln Z\rangle=\lim_{n\to0}1/n\ln\langle Z^n\rangle$ and making the gauge transformations $\sigma_i\to\sigma_i\xi_i$, $\tau_i\to\tau_i\zeta_i$, $\mathcal{A}_{\Lambda_K\Omega_L}\to\mathcal{A}_{\Lambda_K\Omega_L}\Pi_{i\in\Lambda_K}\xi_i\Pi_{j\in\Omega_L}\zeta_j$, and $\mathcal{G}_{\Lambda_{K'}\Omega_{L'}}\to\mathcal{G}_{\Lambda_{K'}\Omega_{L'}}\Pi_{i\in\Lambda_{K'}}\xi_i\Pi_{j\in\Omega_{L'}}\zeta_j$. This will allow us to disentangle the variables $\{\xi,\zeta\}$ from expressions involving the tensors $\mathcal{A}$ and $\mathcal{G}$ in Eqs. (16) and (17). Replacing the $\delta$ functions in Eqs. (23) and (24) by their integral representations allows us to perform the tensor summations. This leads to site-factorized expressions with (an infinite number) $m$ of replica indices. They take the form

$$q_{\alpha_1,\ldots,\alpha_m}=\sum_{i=1}^{N}Z_i\sigma_i^{\alpha_1},\ldots,\sigma_i^{\alpha_m},$$

$$r_{\alpha_1,\ldots,\alpha_m}=\sum_{i=1}^{N}X_i\sigma_i^{\alpha_1},\ldots,\sigma_i^{\alpha_m}, \tag{25}$$

$$t_{\alpha_1,\ldots,\alpha_m}=\sum_{j=1}^{M}Y_j\tau_j^{\alpha_1},\ldots,\tau_j^{\alpha_m},$$

$$u_{\alpha_1,\ldots,\alpha_m}=\sum_{j=1}^{M}V_j\tau_j^{\alpha_1},\ldots,\tau_j^{\alpha_m}, \tag{26}$$

which we insert in the expression for the free energy via suitably defined $\delta$ functions (giving rise to the Lagrange multipliers $\hat{q}_{\alpha_1,\ldots,\alpha_m}$, $\hat{r}_{\alpha_1,\ldots,\alpha_m}$, $\hat{t}_{\alpha_1,\ldots,\alpha_m}$, and $\hat{u}_{\alpha_1,\ldots,\alpha_m}$). To proceed with the calculation one needs to assume a certain order parameter symmetry for the above

quantities and their conjugates for all $m>1$. The simplest such assumption renders all replica $m$-tuples equivalent and all order parameters within this replica symmetric scheme need only depend on the number $m$. This effect can be described by the introduction of suitably defined distributions, the moments of which completely define the $m$-index order parameters

$$q_{\alpha_1,\ldots,\alpha_m}=q\int dx\ \pi(x)\ x^m,\quad \hat{q}_{\alpha_1,\ldots,\alpha_m}=\hat{q}\int dx\ \hat{\pi}(x)x^m, \tag{27}$$

$$r_{\alpha_1,\ldots,\alpha_m}=r\int dy\rho(y)y^m,\quad \hat{r}_{\alpha_1,\ldots,\alpha_m}=\hat{r}\int dy\hat{\rho}(y)y^m, \tag{28}$$

$$t_{\alpha_1,\ldots,\alpha_m}=t\int dx\phi(x)x^m,\quad \hat{t}_{\alpha_1,\ldots,\alpha_m}=\hat{t}\int dx\hat{\phi}(x)x^m, \tag{29}$$

$$u_{\alpha_1,\ldots,\alpha_m}=u\int dy\psi(y)y^m,\quad \hat{u}_{\alpha_1,\ldots,\alpha_m}=\hat{u}\int dy\hat{\psi}(y)y^m, \tag{30}$$

where all integrals are over the interval $[-1,1]$. The Nishimori condition ($\beta=1$), which corresponds to MPM decoding [14], also ensures that this simplest replica-symmetric scheme is sufficient to describe the thermodynamically dominant state [13,15]. Furthermore, it is worthwhile mentioning that extending the replica symmetric calculation to include the one-step replica symmetry breaking ansatz is unlikely to modify the location of the transition points identified under the replica-symmetric ansatz, as has been recently shown in a similar system [16]. Using the above ansatz we perform the trace over the spin variables, and in the limit $n\to 0$ we obtain:

$$\begin{aligned}-\beta f=\mathrm{Extr}\Big\{&-\overline{C}J_{1a}[\pi,\hat{\pi}]-\frac{\overline{C}L}{K}J_{1b}[\rho,\hat{\rho}]-C'J_{1c}[\phi,\hat{\phi}]\\&-\frac{C'}{K'}J_{1d}[\psi,\hat{\psi}]+\frac{\overline{C}}{K}J_{2a}[\pi,\rho]+\frac{C'}{K'}J_{2b}[\phi,\psi]\\&+J_{3a}[\hat{\pi},\hat{\phi}]+\frac{\overline{C}}{K}\frac{L}{=L}J_{3b}[\hat{\rho},\hat{\psi}]\Big\}-\left(\frac{\overline{C}}{K}+\frac{C'}{K'}\right)\ln 2,\end{aligned} \tag{31}$$

where the extremization is taken over the distributions defined in Eqs. (27)–(30) and the various integrals $J_{\star\star}$ are given by

$$J_{1a}[\pi,\hat{\pi}]=\int dxd\hat{x}\pi(x)\hat{\pi}(\hat{x})\ln(1+x\hat{x}),$$

$$J_{1b}[\rho,\hat{\rho}]=\int dyd\hat{y}\rho(y)\hat{\rho}(\hat{y})\ln(1+y\hat{y}), \tag{32}$$

$$J_{1c}[\phi,\hat{\phi}]=\int dxd\hat{x}\phi(x)\hat{\phi}(\hat{x})\ \ln(1+x\hat{x}),$$

$$J_{1d}[\psi,\hat{\psi}]=\int dyd\hat{y}\psi(y)\hat{\psi}(\hat{y})\ln(1+y\hat{y}), \tag{33}$$

$$\begin{aligned}J_{2a}[\pi,\rho]=\int\left[\prod_{k=1}^{K}dx_k\pi(x_k)\prod_{\ell=1}^{L}dy_\ell\rho(y_\ell)\right]\\\times\ln\left(1+\prod_k x_k\prod_\ell y_\ell\right),\end{aligned} \tag{34}$$

$$J_{2b}[\phi,\psi]=\int dy\psi(y)\left[\prod_{k=1}^{K'}dx_k\ \phi(x_k)\right]\ln\left(1+y\prod_k x_k\right), \tag{35}$$

$$\begin{aligned}J_{3a}[\hat{\pi},\hat{\phi}]=\int\prod_{c'=1}^{C'}d\hat{\phi}(y_{c'})\Bigg\{&(1-\gamma)^C\Big\langle\ln\sum_{\lambda=\pm}[(1-c)\\&+c\delta_{\lambda,1}]e^{\beta F_\sigma\xi\lambda}\prod_{c'}(1+y_{c'}\lambda)\Big\rangle_{c,\xi}\\&+\Big\langle\int\left[\prod_{c=1}^{\tilde{C}}d\hat{\pi}(x_c)\right]\Big\langle\ln\sum_{\lambda=\pm}[(1-c)+c\delta_{\lambda,1}]\\&\times e^{\beta F_\sigma\xi\sigma}\prod_c(1+x_c\lambda)\\&\times\prod_{c'}(1+y_{c'}\lambda)\Big\rangle_{c,\xi}\Big\rangle_{\tilde{C}}\Bigg\},\end{aligned} \tag{36}$$

$$\begin{aligned}J_{3b}[\hat{\rho},\hat{\psi}]=\int dy\ \ \hat{\psi}(y)\Bigg\{&(1-\gamma)^L\Big\langle\ln\sum_{\lambda=\pm}[(1-d)\\&+d\delta_{\lambda,1}]e^{\beta F_\tau\zeta\lambda}(1+y\lambda)\Big\rangle_{d,\zeta}\\&+\Big\langle\int\left[\prod_{\ell=1}^{\tilde{L}}d\hat{\rho}(x_\ell)\right]\Big\langle\ln\sum_{\lambda=\pm}[(1-d)+d\delta_{\lambda,1}]\\&\times e^{\beta F_\tau\zeta\lambda}\prod_\ell(1+x_\ell\lambda)(1+y\lambda)\Big\rangle_{d,\zeta}\Big\rangle_{\tilde{L}}\Bigg\},\end{aligned} \tag{37}$$

where

$$\overline{C}=\sum_{\tilde{C}=0}^{C}P(\tilde{C};C)\tilde{C},\quad \overline{L}=\sum_{\tilde{L}=0}^{L}P(\tilde{L};L)\tilde{L} \tag{38}$$

Averages denoted $\langle\cdots\rangle_{\tilde{C}}$ and $\langle\cdots\rangle_{\tilde{L}}$ are over the densities (6) and (7) with $\tilde{C}=1,\ldots,C$ and $\tilde{L}=1,\ldots,L$. Functional

differentiation of Eq. (31) with respect to the densities of Eqs. (27)–(30) results in the following saddle point equations:

$$\hat{\pi}(\hat{x}) = \int \left[ \prod_{k=1}^{K-1} dx_k \pi(x_k) \prod_{l=1}^{L} dy_l \rho(y_l) \right] \times \delta\left[ \hat{x} - \prod_{k=1}^{K-1} x_k \prod_{l=1}^{L} y_l \right], \tag{39}$$

$$\hat{\rho}(\hat{y}) = \int \left[ \prod_{k=1}^{K} dx_k \pi(x_k) \prod_{l=1}^{L-1} dy_l \rho(y_l) \right] \times \delta\left[ \hat{y} - \prod_{k=1}^{K} x_k \prod_{l=1}^{L-1} y_l \right], \tag{40}$$

$$\hat{\phi}(\hat{x}) = \int dy \, \psi(y) \left[ \prod_{k=1}^{K'-1} dx_k \phi(x_k) \right] \delta\left[ \hat{x} - y \prod_{k=1}^{K'-1} x_k \right], \tag{41}$$

$$\hat{\psi}(\hat{y}) = \int \left[ \prod_{k=1}^{K'} dx_k \phi(x_k) \right] \delta\left[ \hat{y} - \prod_{k=1}^{K'} x_k \right] \tag{42}$$

and

$$\pi(x) = w_\sigma \delta[x-1] + \frac{(1-w_\sigma)}{\overline{C}} \left\langle \tilde{C} \int \left[ \prod_{c'=1}^{C'} d\hat{\phi}(\hat{y}_{c'}) \prod_{c=1}^{\tilde{C}-1} d\hat{\pi}(\hat{x}_c) \right] \times \left\langle \delta\left( x - \tanh\left[ \beta F_\sigma \xi + \sum_{c=1}^{\tilde{C}-1} \operatorname{arctanh}(\hat{x}_c) + \sum_{c'=1}^{C'} \operatorname{arctanh}(\hat{y}_{c'}) \right] \right) \right\rangle_\xi \right\rangle_{\tilde{C}}, \tag{43}$$

$$\rho(x) = w_\tau \delta[x-1] + \frac{(1-w_\tau)}{\overline{L}} \left\langle \tilde{L} \int d\hat{\psi}(\hat{y}) \left[ \prod_{l=1}^{\tilde{L}-1} d\hat{\rho}(\hat{y}_l) \right] \times \left\langle \delta\left( x - \tanh\left[ \beta F_\tau \zeta + \sum_{l=1}^{\tilde{L}-1} \operatorname{arctanh}(\hat{x}_l) + \operatorname{arctanh}(\hat{y}) \right] \right) \right\rangle_\zeta \right\rangle_{\tilde{L}}, \tag{44}$$

$$\phi(x) = w_\sigma \delta[x-1] + (1-w_\sigma) \int \prod_{c'=1}^{C'-1} d\hat{\phi}(y_{c'}) \left\{ (1-\gamma)^C \times \left\langle \delta\left( x - \tanh\left[ \beta F_\sigma \xi + \sum_{c'=1}^{C'-1} \operatorname{arctanh}(\hat{y}_{c'}) \right] \right) \right\rangle_\xi + \left\langle \int \left[ \prod_{c=1}^{\tilde{C}} d\hat{\pi}(\hat{x}_c) \right] \left\langle \delta\left( x - \tanh\left[ \beta F_\sigma \xi + \sum_{c=1}^{\tilde{C}} \operatorname{arctanh}(\hat{x}_c) + \sum_{c'=1}^{C'-1} \operatorname{arctanh}(\hat{y}_{c'}) \right] \right) \right\rangle_\xi \right\rangle_{\tilde{C}} \right\}, \tag{45}$$

$$\psi(x) = w_\tau \delta[x-1] + (1-w_\tau) \left\{ (1-\gamma)^L \times \langle \delta[x - \tanh(\beta F_\tau \zeta)] \rangle_\zeta + \left\langle \int \left[ \prod_{l=1}^{\tilde{L}} d\hat{\rho}(\hat{x}_l) \right] \times \left\langle \delta\left( x - \tanh\left[ \beta F_\tau \zeta + \sum_{l=1}^{\tilde{L}} \operatorname{arctanh}(\hat{x}_l) \right] \right) \right\rangle_\zeta \right\rangle_{\tilde{L}} \right\}. \tag{46}$$

In general, the coupled set of equations (39)–(46) are to be solved numerically. Among the set of $\boldsymbol{\sigma}$ that satisfy Eqs. (4) and (5) we choose the MPM estimate of the plaintext $\hat{\xi}_i = \operatorname{sgn} \Sigma_{\sigma_i = \pm} \sigma_i \, p(\sigma_i | \boldsymbol{r}) = \operatorname{sgn} \langle \sigma_i \rangle$ (thermal average) by using Nishimori's condition (or $\beta = 1$) [13]. Then, the overlap $m = \lim_{N \to \infty} 1/N \Sigma_i \xi_i \hat{\xi}_i$ becomes

$$m = w_\sigma + (1-w_\sigma) \int dh \; P(h) \operatorname{sign}(h), \tag{47}$$

$$P(h) = \int \left[ \prod_{c'=1}^{C'} d\hat{\phi}(\hat{y}_{c'}) \right] \left\{ (1-\gamma)^C \left\langle \delta\left( h - \tanh\left[ \beta F_\sigma \xi + \sum_{c'=1}^{C'} \operatorname{arctanh}(\hat{y}_{c'}) \right] \right) \right\rangle_\xi + \left\langle \int \left[ \prod_{c=1}^{\tilde{C}} d\hat{\pi}(\hat{x}_c) \right] \times \left\langle \delta\left( h - \tanh\left[ \beta F_\sigma \xi + \sum_{c=1}^{\tilde{C}} \operatorname{arctanh}(\hat{x}_c) + \sum_{c'=1}^{C'} \operatorname{arctanh}(\hat{y}_{y'}) \right] \right) \right\rangle_\xi \right\rangle_{\tilde{C}} \right\}, \tag{48}$$

from which it can be seen that the perfect (ferromagnetic) solution $m = 1$ is achieved when $w_\sigma = 1$ (complete knowledge of the solution) or when $\hat{\phi}(x) = \delta[x-1]$. This also

implies that all densities involved in Eq. (31) $\lambda(x)=\{\pi(x),\ldots,\hat{\psi}(x)\}$ acquire the form $\lambda(x)=\delta[x-1]$ giving a free energy of the form

$$f_{FM}=\left(\frac{C'}{K'}-\frac{C}{K}\right)\ln 2-\frac{C}{K}\beta F_\tau\langle\zeta\rangle_\zeta. \tag{49}$$

The physical meaning of the terms $w_\star\,\delta[x-1]$ in Eqs. (43)–(46) is that the acquired microscopic knowledge gives a probabilistic weight at the ferromagnetic state. The state $m=0$ is obtained if $w_\sigma=F_\sigma=0$ and $\hat{\pi}(x)=\hat{\phi}(x)=\delta[x]$ (paramagnetic solution).

## V. PHASE DIAGRAMS

In this section we obtain numerical solutions for various attack scenarios. In all cases studied we assume an unbiased plaintext ($p_\sigma=1/2$, $F_\sigma=0$); for brevity we refer to the remaining bias parameter, the corruption level denoted $p_\tau$ in previous sections, simply as $p$. All experiments have been carried out using a regular cryptosystem with $K=L=2$, being the original cryptosystem suggested in Ref. [4]. In principle, one can use any set of regular or irregular matrices, provided one identifies the corresponding dynamical transition point. However, having been thoroughly studied previously, the current construction serves as a particularly suited benchmark.

Solving the coupled equations (39)–(46) we typically observe that for sufficiently small values of $p$ the ferromagnetic state $m=1$ is the only stable solution whereas at a corruption value that marks the dynamical (spinodal) transition $p_s$, an exponential number of solutions with $m\neq 1$ are created [either suboptimal ferromagnetic or paramagnetic, depending on the values of $(K,C,L)$]. For all $p>p_s$ perfect decryption will be difficult to obtain. This transition also defines the corruption level below which an unauthorized attacker, who has acquired partial information of the secret keys, will be successful.

We will concentrate on two main attacks: (i) the attacker has partial knowledge of the keys (primarily the matrix $B$); (ii) the attacker has partial microscopic knowledge of the plaintext and/or corruption vector.

In Fig. 2 we present a phase diagram describing regions with perfect ($m=1$) or partial/null ($|m|<1$) decryption success as evaluated from solving Eqs. (31) and (47). We plot the dynamical transition corruption level $p_s$ as a function of the private key fractional knowledge $\gamma$ for different values of $w_\sigma$ and $w_\tau$ (we have set $p_\sigma=1/2$ which corresponds to an 'unbiased' plaintext). In the limit $\gamma=0$ (i.e., no knowledge of the matrices), while $m=1$ may be a stable solution, the decryption dynamics is fully dominated by $|m|<1$ states. For $\gamma=1$ the cryptosystem describes a specific MN code and perfect decryption can occur below $p_s$.

The interaction between sparsely (4) and densely (5) connected decryption components is nonlinear and nontrivial; however, as a first approximation one can view the fractional matrix knowledge $\gamma$ as changing the effective sparse component, which is the main contributor in the decryption process.

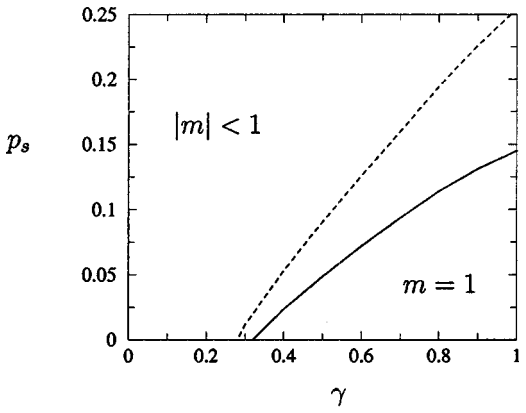


FIG. 2. Phase diagram of the spinodal corruption rate against the fractional knowledge of the private key $\gamma$ for a $(K,C,L)=(2,6,2)$ cryptosystem for $(w_\sigma,w_\tau)=(0,0)$ (solid line) and (0.2,0.2) (dashed line). Microscopic knowledge of the plaintext and the corrupting vector enlarges the perfect decryption area, as expected.

To that end $\gamma$ will have a direct impact on the effective code rate $N/(M\gamma)$, the average connectivity $\gamma C$, and the connectivity distribution. It is clear that at an effective code rate 1 ($\gamma=N/M=1/3$ in the case of the parameters used in Fig. 2) decryption is even not theoretically feasible. The reason Fig. 2 points to a possibility of decryption below this value is due to additional information brought in by the dense components we ignored in this simplistic description.

We also examined the effect of prior microscopic knowledge of the plaintext/corrupting vector $(w_\sigma,w_\tau>0)$ on the area of perfect decryption; which clearly increases with the knowledge provided, as expected. Also this can be viewed as a change to the effective code rate. This time, the partial microscopic knowledge of either plaintext or corrupting vector (or both) serves to reduce the effective number of variables and hence the code rate itself; lower code rate will typically allow for perfect decryption in worse corruption conditions as can be seen in Fig. 2

To understand the implication of these results let us assume using the cryptosystem described in Fig. 2 at a chosen corruption level of $p=0.1$ (which is chosen much smaller that $p_s$ to increase the decryption reliability). In this case knowing about 70% of the matrices (secret keys) will be sufficient for decrypting the ciphertext. True, there is still a need to know the dense matrix $D^{-1}$ for extracting the plaintext itself and the exposed fraction of the secret key is significant; but still there is a weakness that may be exploited by a skillful attacker.

To compare the importance of prior microscopic knowledge of plaintext versus that of the corrupting vector we plotted in Fig. 3 the phase diagrams for $(w_\sigma,w_\tau)=\{(0.1,0),(0.2,0)\}$ and $(w_\sigma,w_\tau)=\{(0,0.1),(0,0.2)\}$ which describe two complementary scenarios. The effect is quite similar, taking into account the information provided by the two vectors (the plaintext is unbiased but of length $N$ while the corruption vector is biased but of length $M$). For high $\gamma$ values microscopic knowledge of the corrupting vector becomes more informative than that of the plaintext, an effect

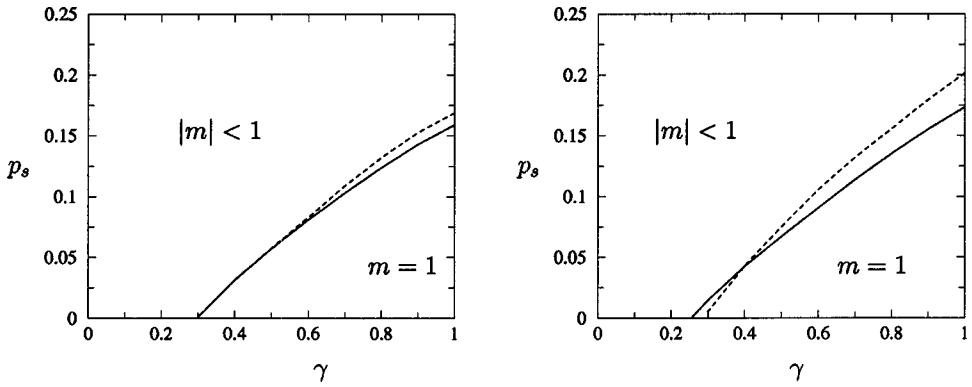


FIG. 3. Phase diagrams of the spinodal corruption rates against the fractional knowledge of the private key $\gamma$ for a $(K,C,L)=(2,6,2)$ cryptosystem. Left picture: $(w_\sigma,w_\tau)=(0.1,0)$ (solid line) and $(0,0.1)$ (dashed line). Right picture: $(w_\sigma,w_\tau)=(0.2,0)$ (solid line) and $(0,0.2)$ (dashed line). For sufficiently large $\gamma$ values microscopic knowledge of the corrupting vector becomes more important to the unauthorized user than that of the plaintext; this effect becomes more emphasized as the fraction of known bits increases.

which becomes more emphasized as the fraction of known bits increases.

In Fig. 4 we compare two cryptosystems with $(K,C,L)=(2,4,2)$ and $(K,C,L)=(2,3,2)$ for $(w_\sigma,w_\tau)=(0,0)$. We see that smaller $C$ values (i.e., higher code rates) will reduce the area of perfect decryption. On the one hand, this will increase the secret information required for perfect decryption at each corruption level; on the other hand, it will reduce the corruption level that can be used and will expose the cryptosystem to attacks based on an exhaustive search of corruption vectors.

The security of a cryptosystem may be compromised without a full recovery of the plaintext; also partial recovery of the plaintext may pose a significant threat. To study the effect of partial knowledge of the matrices and plaintext on the ability to obtain high overlap between the decrypted ciphertext and plaintext, we conducted several experiments, an example of which appears in Fig. 4. Here we show the overlap obtained $m$ as function of the corruption rate $p$ for a specific cryptosystem $(K,C,L)=(2,6,2)$ along the line $\gamma=0.8$ and for two different choices of $w_\sigma$. Prior to the dynamical transition points both ciphertexts are decrypted perfectly; this corresponds to corruption and partial knowledge levels below the solid and dashed lines of Fig. 2.

Above the dynamical transition point, new suboptimal solutions are created and the overlap value obtained deteriorates with the corruption level. However, the two different choices of $w_\sigma$ values lead to two different deterioration patterns: while overlap in the system with no microscopic knowledge of the plaintext deteriorates very rapidly, the system with $w_\sigma=0.2$ provides solutions with high overlap values even if the corruption is high. As a consequence, we see that the effect of microscopic knowledge goes beyond a shift in the dynamical transition point; it also influences decryption beyond that point (in fact, it goes even beyond Shannon's limit).

## VI. BASIN OF ATTRACTION

The increasingly narrowing basin of attraction for the ferromagnetic solution, as the connectivity values $K$, $C$, and $L\to\infty$, is central to the security level offered by the cryptosystem. The effect has been reported in a number of papers in the statistical physics [4,12] and information-theory [5] literature; in this section we will show that the basin of at-

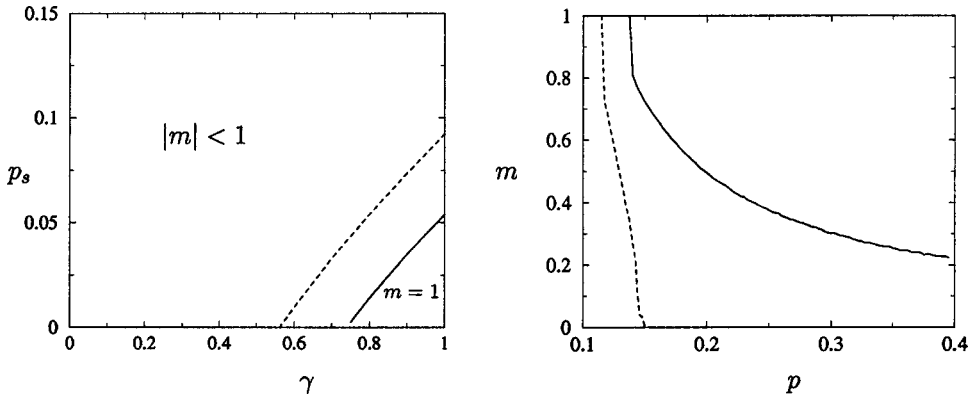


FIG. 4. Left: Comparison between two different cryptosystems with $(K,C,L)=(2,3,2)$ (solid line) and $(K,C,L)=(2,4,2)$ (dashed line). Smaller $C$ values correspond to higher rate codes and lead to smaller regions in parameter space where perfect decryption is possible. Right: Overlap $m$ as function of the corrupting rate $p$ obtained from Eq. (47) for a $(K,C,L)=(2,6,2)$ cryptosystem and along the line $\gamma=0.8$ for $(w_\sigma,w_\tau)=(0.2,0)$ (solid line) and $(w_\sigma,w_\tau)=(0,0)$ (dashed line).

traction shrinks as the connectivity increases, to a value of $O(1/K)$ as $K,C\to\infty$.

To provide a rough evaluation of the basin of attraction (BOA) for obtaining the ferromagnetic solution we focus on Eq. (2) in the limit $K,C\to\infty$. BOA clearly depends on the algorithm used; here we focus on the belief propagation (BP) algorithm, which is empirically known to be the best practical algorithm for solving problems of the current type. As far as we explored, no other schemes such as the naive mean field and the belief revision algorithms exhibit better performance than BP, which implies that our consideration on BP is at least of a certain practical significance (survey propagation [9] has not yet been tested for these systems).

Let us represent prior knowledge on plaintext $\boldsymbol{\xi}$ and noise $\boldsymbol{\zeta}$ (in Ising spin representation) as the *prior probabilities*

$$P_i^o(\sigma_i)=\frac{\exp(F_{\sigma i}\sigma_i)}{2\cosh(F_{\sigma i})}, \tag{50}$$

$$P_j^o(\tau_j)=\frac{\exp(F_{\tau j}\tau_j)}{2\cosh(F_{\tau j})}, \tag{51}$$

respectively. Here, the parameters $F_{\sigma i}$ and $F_{\tau j}$ express confidence of the prior knowledge per variable, which is a generalization of the global prior terms $F_\sigma, F_\tau$ used earlier. Notice that this representation includes the case that certain bits are completely determined by setting $|F_{\sigma i}|$ (or $|F_{\tau j}|$) $\to\infty$, enabling us to cover various scenarios. In the following, we assume that the fraction of completely determined bits is less than 1 when $N,M\to\infty$. Given prior probabilities (50) and (51), and the indicator function $\Delta(\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\xi},\boldsymbol{\zeta},\mathcal{A})$ which is the alternative to parity check Eq. (2), the Bayesian framework provides the *posterior probability*

$$P^{post}(\boldsymbol{\sigma},\boldsymbol{\tau})=\frac{\Delta(\boldsymbol{\sigma},\boldsymbol{\tau},\boldsymbol{\xi},\boldsymbol{\zeta},\mathcal{A})\prod_{i=1}^{N}P_i^o(\sigma_i)\prod_{j=1}^{M}P_j^o(\tau_i)}{Z}, \tag{52}$$

where $Z$ is the normalization constant. Using Eq. (52), one can determine the best possible action for minimizing the expected value of a given cost function [14]. As a cost function, we select here the Hamming distance between the correct plain text $\boldsymbol{\xi}$ and its estimates $\hat{\boldsymbol{\xi}}$, $L(\hat{\boldsymbol{\xi}},\boldsymbol{\xi})=N-\Sigma_{i=1}^{N}\hat{\xi}_i\xi_i$; this selection naturally offers the MPM decoding $\widehat{\xi}_i=\mathrm{sgn}(m_i^\sigma)$ as the optimal estimation strategy, where

$$m_i^\sigma=\sum_{\boldsymbol{\sigma},\boldsymbol{\tau}}\sigma_i\ P^{post}(\boldsymbol{\sigma},\boldsymbol{\tau}), \tag{53}$$

is the average of spin $\sigma_i$ over the posterior probability and $\mathrm{sgn}(x)=1$ for $x>0$ and $-1$, otherwise.

Computational cost for an exact evaluation of the spin average (53) increases as $O(2^{N+M})$, which implies that MPM decoding is practically difficult. An alternative approach is to resort to an approximation such as BP. In the current case, this means to iteratively solving the coupled equations (for details of the derivation see Refs. [5,10])

$$\hat{m}_{\mu i}^\sigma=J_\mu\prod_{l\in\mathcal{L}^\sigma(\mu)\setminus i}m_{\mu l}^\sigma\prod_{j\in\mathcal{L}^\tau(\mu)}m_{\mu j}^\tau,$$

$$\hat{m}_{\mu j}^\tau=J_\mu\prod_{l\in\mathcal{L}^\sigma(\mu)}m_{\mu l}^\sigma\prod_{k\in\mathcal{L}^\tau(\mu)\setminus j}m_{\mu k}^\tau, \tag{54}$$

$$m_{\mu i}^\sigma=\tanh\left(F_{\sigma i}+\sum_{\nu\in\mathcal{M}^\sigma(i)\setminus\mu}\mathrm{arctanh}(\hat{m}_{\nu i}^\sigma)\right),$$

$$m_{\mu j}^\tau=\tanh\left(F_{\tau j}+\sum_{\nu\in\mathcal{M}^\tau(j)\setminus\mu}\mathrm{arctanh}(\hat{m}_{\nu j}^\tau)\right), \tag{55}$$

where $J_\mu\equiv(\Pi_{l\in\mathcal{L}^\sigma(\mu)}\xi_l\Pi_{j\in\mathcal{L}^\tau(\mu)}\zeta_j)$, $\mathcal{L}^\sigma(\mu)$ and $\mathcal{L}^\tau(\mu)$ are the sets of indices of nonzero elements in $\mu$th row of $A$ and $B$, respectively, and $\mathcal{M}^\sigma(i)$ and $\mathcal{M}^\tau(j)$ are similarly defined for columns of $A$ and $B$, respectively. $\mathcal{L}^\sigma(\mu)\setminus i$ denotes a set of indices in $\mathcal{L}^\sigma$ other than $i$, and similarly for other symbols. The variables $m_{\mu i}^{\sigma/\tau}$ and $\hat{m}_{\mu i}^{\sigma/\tau}$ represent pseudo-posterior-averages of $\sigma_i$ (or $\tau_j$) when the $\mu$th check $J_\mu$ is left out, and the influence of a newly added $J_\mu$ on $\sigma_i$ (or $\tau_j$), respectively (see Refs. [5,10] for details). Using $\hat{m}_{\mu i}^\sigma$, the posterior average $m_i^\sigma$ is obtained as

$$m_i^\sigma=\tanh\left(F_{\sigma i}+\sum_{\mu\in\mathcal{M}^\sigma(i)}\mathrm{arctanh}(\hat{m}_{\mu i}^\sigma)\right). \tag{56}$$

Let us investigate the condition necessary for finding the correct solution by iterating Eqs. (54) and (55) in the limit $K,C\to\infty$. For this purpose, we first employ the gauge transformation $\xi_i m_{\mu i}^\sigma\to m_{\mu i}^\sigma$, $\xi_i\hat{m}_{\mu i}^\sigma\to\hat{m}_{\mu i}^\sigma$, $\zeta_j m_{\mu j}^\tau\to m_{\mu j}^\tau$, $\zeta_j\hat{m}_{\mu j}^\tau\to\hat{m}_{\mu j}^\tau$ and $J_\mu(\Pi_{l\in\mathcal{L}^\sigma(\mu)}\xi_l\Pi_{j\in\mathcal{L}^\tau(\mu)}\zeta_j)\to 1$. This decouples the quenched random variables $\xi_i$ and $\zeta_j$ from Eq. (54), as $J_\mu$ becomes independent of the quenched variables, and the BP equations can be expressed as

$$\hat{m}_{\mu i}^\sigma=\prod_{l\in\mathcal{L}^\sigma(\mu)\setminus i}m_{\mu l}^\sigma\prod_{j\in\mathcal{L}^\tau(\mu)}m_{\mu j}^\tau,$$

$$\hat{m}_{\mu j}^\tau=\prod_{l\in\mathcal{L}^\sigma(\mu)}m_{\mu l}^\sigma\prod_{k\in\mathcal{L}^\tau(\mu)\setminus j}m_{\mu k}^\tau, \tag{57}$$

$$m_{\mu i}^\sigma=\tanh\left(F_i^\sigma\xi_i+\sum_{\nu\in\mathcal{M}^\sigma(i)\setminus\mu}\mathrm{arctanh}(\hat{m}_{\nu i}^\sigma)\right),$$

$$m_{\mu j}^\tau=\tanh\left(F_j^\tau\zeta_j+\sum_{\nu\in\mathcal{M}^\tau(j)\setminus\mu}\mathrm{arctanh}(\hat{m}_{\nu j}^\tau)\right). \tag{58}$$

The expression of the correct solution is also converted to $m_{\mu i}^\sigma=1$ and $m_{\mu j}^\tau=1$. Notice that any state which is characterized by decreasing absolute values $|m_{\mu i}^\sigma|<1-\varepsilon$ and $|m_{\mu i}^\tau|<1-\varepsilon$ for an arbitrary fixed positive number $\varepsilon>0$ is attracted to a locally stable solution $\hat{m}_{\mu i}^\sigma\sim 0$, $\hat{m}_{\mu j}^\tau\sim 0$, $m_{\mu i}^\sigma=\tanh(F_i^\sigma\xi_i)$, and $m_{\mu j}^\tau=\tanh(F_j^\tau\zeta_j)$ for $K\to\infty$ in a single update since products on the right hand sides of Eq. (57) vanish. To provide a rough evaluation of the BOA for the correct

(ferromagnetic) solution $m^{\sigma}_{\mu i}=1$ and $m^{\tau}_{\mu j}=1$, let us assume that $m^{\sigma}_{\mu i}$ and $m^{\tau}_{\mu j}$ are randomly distributed at $1-\varepsilon(K)$ and $-[1-\varepsilon(K)]$ with probabilities $1-p(K)$ and $p(K)$, respectively, where $\varepsilon(K)$ and $p(K)$ are small parameters to characterize the BOA for a large $K$. Under this assumption, $\hat{m}^{\sigma}_{\mu i}$ and $\hat{m}^{\sigma}_{\mu j}$ are distributed at $\pm[1-\varepsilon(K)]^{K+L}\sim\pm[1-\varepsilon(K)]^{K}$ with probability $\{1\pm[1-2p(K)]^{K+L}\}/2\sim\{1\pm[1-2p(K)]^{K}\}/2$, respectively. If either $[1-\varepsilon(K)]^{K}$ or $[1-2p(K)]^{K}$ is negligible, the absolute values of $m^{\sigma}_{\mu i}$ and $m^{\tau}_{\mu j}$ become sufficiently smaller than 1, and therefore, the state is trapped in a locally stable solution in the second iteration [19]. This implies that the critical condition is given by $\varepsilon(K)\sim O(1/K)$ and $p(K)\sim O(1/K)$ for large $K$. In terms of the macroscopic overlap, this means $m^{0}_{cr}\approx 1-O(1/K)$.

## VII. RELIABILITY

Unlike most of the commonly used cryptosystems which are based on a deterministic decryption procedure, the current cryptosystem relies on a probabilistic decryption process. The evaluation of decryption success for an *authorized* user is therefore as important as assessing the level of robustness against attacks.

In practical scenarios, decryption success generally depends on the plaintext size. Analysis of finite size effects in the belief propagation based decryption procedure is difficult. A principled alternative that we pursue here is based on evaluating the *average error exponent* of the current cryptosystem; this provides the expected error level at any given corruption level when maximum likelihood decoding is employed, and therefore represents a lower bound to the expected error rate. Moreover, the corruption levels employed are far below the critical (thermodynamic) transition point, we therefore *assume* that belief propagation decryption will provide similar performance to maximum likelihood decoding; clearly, the lower bound will become looser as we get close to the dynamical transition point.

The average block error rate $P_B(p)$ (i.e., erroneous decrypted plaintexts) takes the form

$$P_B(p)=e^{-ME(p)}, \tag{59}$$

where $E(p)$ is the average error exponent per noise level $p$ and $M$ the length of the ciphertext (in the particular case of LDPC codes we assume that short loops, which contribute polynomially to the block error probability [17], have been removed). The quantity $P_B(p)$ represents the probability by which candidate solutions $\{\boldsymbol{\sigma},\boldsymbol{\tau}\}$ are drawn from the set of those satisfying Eq. (4) (with $\gamma=1$; authorized decryption) other than the ones corresponding to the true plaintext and corrupting vector, $\boldsymbol{\sigma}=\boldsymbol{\xi}$ and $\boldsymbol{\tau}=\boldsymbol{\zeta}$, respectively. To evaluate this probability we introduce the indicator function

$$\Psi(\Gamma)=\lim_{\beta\to\infty}\lim_{\lambda_{1,2}\to\pm\lambda}[Z_1^{\lambda_1}(\Gamma;\beta_1)Z_2^{\lambda_2}(\Gamma;\beta_2)]_{\beta_1=\beta_2=\beta}, \tag{60}$$

where $\Gamma=\{\boldsymbol{\xi},\boldsymbol{\zeta},\mathcal{A}\}$ collectively denotes the set of quenched variables. The power $\lambda\in[0,1]$ is used in conjunction with the partition functions

$$Z_1(\Gamma;\beta_1)=\sum_{\boldsymbol{\sigma}\neq\boldsymbol{\xi}}\sum_{\boldsymbol{\tau}\neq\boldsymbol{\zeta}}e^{-\beta_1 H(\boldsymbol{\sigma},\boldsymbol{\tau})},$$

$$Z_2(\Gamma;\beta_2)=\sum_{\boldsymbol{\sigma}}\sum_{\boldsymbol{\tau}}e^{-\beta_2 H(\boldsymbol{\sigma},\boldsymbol{\tau})} \tag{61}$$

to provide an indicator function as explained below. The Hamiltonian $H(\boldsymbol{\sigma},\boldsymbol{\tau})$ is given by Eq. (13) and the trace over spin variables is restricted to those configurations satisfying Eq. (4). The above partition functions $Z_1$ and $Z_2$ differ only in the exclusion of the true plaintext and corrupting vector in the trace over variables; this enables us to identify instances where the maximum likelihood decoder chooses solutions that do not match the true (quenched variable) vectors. Hamiltonian (13) is proportional to the magnetizations $m_{\sigma}(\boldsymbol{\sigma})=(1/N)\Sigma_i\sigma_i$ and $m_{\tau}(\boldsymbol{\tau})=(1/M)\Sigma_i\tau_i$. Therefore, if the true plaintext and corrupting vectors have the highest magnetizations (decryption success), the Boltzmann factor $\exp[-\beta H(\boldsymbol{\sigma},\boldsymbol{\tau})]$ will dominate the sum over states in $Z_2$ in the limit $\beta\to\infty$ and $\Psi(\Gamma)=0$. Alternatively, if some other vectors $\boldsymbol{\sigma}\neq\boldsymbol{\xi}$ and $\boldsymbol{\tau}\neq\boldsymbol{\zeta}$ have the highest magnetizations of all candidates (decoding failure), its Boltzmann factor will dominate both $Z_1$ and $Z_2$ so that $\Psi(\Gamma)=1$. Separate temperatures $\beta_{1,2}$ and powers $\lambda_{1,2}$ have been introduced to determine whether obtained solutions are physical or not (values of these parameters will be obtained via the zero-entropy condition).

To derive the average error exponent $E(p)$ we take the logarithm of the above indicator function averaged with respect to the disorder variables $\Gamma=\{\boldsymbol{\xi},\boldsymbol{\zeta},\mathcal{A}\}$,

$$E(p)=\lim_{M\to\infty}\frac{1}{M}\ln\langle\Psi(\Gamma)\rangle_{\Gamma}. \tag{62}$$

The evaluation of Eq. (62) is similar in spirit to the analysis of Sec. IV. For details of this calculation we refer the reader to Ref. [18] where we also study and compare the reliability and average error exponents of various low-density parity-check codes.

Results describing $E(p)$ for authorized decryption of the cryptosystem [4] are presented in Fig. 5 where we plot $E(p)$ as function of the corruption level $p$ for $(K,C,L)=(2,8,2)$ (code rate 1/4) and $(K,C,L)=(2,4,2)$ (code rate 1/2) cryptosystems. It is clear that decryption errors decay very fast with the system size as we go away from the critical corruption level. For instance, in the case of $R=1/4$, using a corruption level of $p=0.13$ (Shannon's limit is at $p=0.20$) and a modest ciphertext size of $M=1000$ will result in a negligible block error probability $P_B=10^{-11}$.

## VIII. DISCUSSION

In this paper we have analyzed several security issues related to the recently suggested public-key cryptosystem of Ref. [4]. The suggested cryptosystem is based on the com-

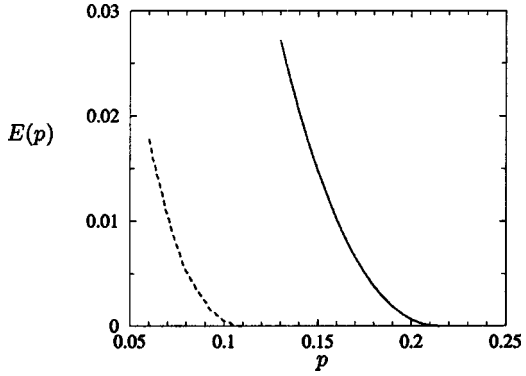


FIG. 5. Reliability exponent (62) as a function of the corruption level $p$ for the case $K=L=2$ and rates $R=1/2$ (dashed line) and $R=1/4$ (solid line).

putational difficulty of decomposing a dense matrix into a combination of dense and sparse matrices (obeying certain statistics) which is a known hard computational problem. We have considered several attack scenarios in which unauthorized parties have acquired partial knowledge of one or more of the private keys and/or microscopic knowledge of the plaintext and/or the ''corrupting vector.'' The analysis follows standard statistical mechanical methods of dealing with diluted spin systems within replica symmetric considerations. Of central importance to the unauthorized decryption is the dynamical transition which defines decryption success in practical situations. This has been calculated using a replica symmetric ansatz, which is sufficient for an accurate evaluation of the dynamic transition point; more involved one-step replica symmetry breaking schemes for similar systems predict the same dynamical transition point as that obtained by using the replica symmetric anzats [16]. Our phase diagrams show the dynamical threshold as a function of the partial acquired knowledge of the private key; they describe regions with perfect ($m=1$) or partial/null decryption success ($|m|<1$).

Public-key cryptosystems play an important role in modern communications. The increasing demand for secure transmission of information has lead to the invention of novel cryptosystems in recent years. To this extent and based on the insight gained by statistical physics analyses of error-correcting codes a new family of cryptosystems was suggested in Ref. [4]. This paper constitutes a first step in studying this class of cryptosystems by considering the potential success of possible attacks.

Several future research directions aimed at improving the security and reliability of this cryptosystem may include studying the efficacy of irregular code constructions and the use of novel decryption methods such as survey propagation [9] for pushing the dynamical transition point closer to the information theoretic limits.

## ACKNOWLEDGMENTS

We would like to thank Jort van Mourik for helpful discussions. Support from EPSRC research grant, Grant No. GR/N63178, the Royal Society (D.S., N.S.), and Grant-in-Aid, MEXT, Japan, No. 14084206 (Y.K.) are gratefully acknowledged. NS would also like to acknowledge support from the Fund for Scientific Research-Flanders, Belgium, for the final stages of this research. This work has been supported in part by the European Community's Human Potential Programme under Contract No. HPRN-CT-2002-00319, STIPCO.

[1] R.I. Rivest, A. Shamir, and L. Adleman, Commun. ACM **21**, 120 (1978).
[2] W. Diffie and M.E. Hellman IEEE Trans. Inf. Theory **22**, 644 (1976).
[3] D.R. Stinson *Cryptography, Theory and Practice* (Chapman and Hall, London, 1995).
[4] Y. Kabashima, T. Murayama, and D. Saad, Phys. Rev. Lett. **84**, 2030 (2000).
[5] D.J.C. MacKay, IEEE Trans. Inf. Theory **45**, 399 (2000).
[6] Y. Kabashima, T. Murayama and D. Saad, Phys. Rev. Lett. **84**, 1355 (2000).
[7] M.R. Garey and D.S. Johnson, *Computers and Intractability* (Freeman, San Francisco, 251).
[8] Y. Weiss Neural Comput. **12**, 1 (2000).
[9] M. Mézard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002).
[10] T. Murayama, Y. Kabashima, D. Saad, and R. Vicente, Phys. Rev. E **62**, 1577 (2000).
[11] N. Sourlas, Nature (London) **339**, 693 (1989).
[12] I. Kanter and D. Saad, Phys. Rev. Lett. **83**, 2660 (1999).
[13] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing* (Oxford University Press, Oxford, UK, 2001).
[14] Y. Iba J. Phys. A **32**, 3875 (1999).
[15] H. Nishimori and D. Sherrington, in *Disordered and Complex Systems*, edited by P. Sollich, A.C.C. Coolen, L.P. Hughston, and R.F. Streater, AIP Conf. Proc. 553 (AIP, Melville, New York, 2001), p. 67.
[16] S. Franz, M. Leone, A. Montanari, and F. Ricci-Tersenghi, Phys. Rev. E **66**, 046120 (2002).
[17] G. Miller and D. Burshtein, IEEE Trans. Inf. Theory **47**, 2696 (2001).
[18] N.S. Skantzos, J. van Mourik, D. Saad, and Y. Kabashima, J. Phys. A **36**, 11131 (2003).
[19] Although larger $C$ values would increase the absolute values of $m_{\mu i}$ and $m_{\mu j}$ in Eq. (58), this effect is relatively small and the critical condition is determined mainly by $K$ in Eq. (57).