Distance verification for LDPC codes

Ilya Dumer*, Alexey A. Kovalev†, and Leonid P. Pryadko‡
* ECE Department, University of California at Riverside, USA (e-mail: dumer@ee.ucr.edu)
† Department of Physics & Astronomy, University of Nebraska-Lincoln, USA (e-mail: alexey.kovalev@unl.edu)
‡ Department of Physics & Astronomy, University of California at Riverside, USA (e-mail: leonid@ucr.edu)

***Abstract*—The problem of finding code distance has been long studied for the generic ensembles of linear codes and led to several algorithms that substantially reduce exponential complexity of this task. However, no asymptotic complexity bounds are known for distance verification in other ensembles of linear codes. Our goal is to re-design the existing generic algorithms of distance verification and derive their complexity for LDPC codes. We obtain new complexity bounds with provable performance expressed in terms of the erasure-correcting thresholds of long LDPC codes. These bounds exponentially reduce complexity estimates known for linear codes.**

Index Terms – Distance verification, complexity bounds, LDPC codes, erasure correction, covering sets

I. Introduction

This paper addresses the problem of finding code distances of LDPC codes with provable complexity estimates. Note that finding code distance d of a generic code is an NP-hard problem. This is valid for both the exact setting [1] and the evaluation problem [2], [3], where we only verify if d belongs to some interval $[\delta, c\delta]$ given some constant $c \in (1, 2)$. To this end, all algorithms of distance verification discussed in this paper have exponential complexity 2^{Fn} in blocklength n and our goal is to reduce the complexity exponent F.

Below we address generic algorithms of *distance verification* - known for linear codes - and re-design these algorithms for LDPC codes. The main problem is that such algorithms heavily rely on the properties of the randomly chosen generator (or parity-check) matrices. These properties have not been proved (or do not hold) for the smaller ensembles of codes, such as cyclic codes, LDPC codes, and others. Therefore, we will use a different technique and derive complexity estimates for LDPC codes using a single parameter, which is the erasure-correcting threshold of a specific code ensemble. We then define this threshold via the average weight spectra of LDPC codes. This technique is different from the generic approach. In particular, we calculate the average complexity of distance verification and then discard a vanishing fraction of LDPC codes that have atypically high complexity. Our main result is the new complexity bounds for distance verification for ensembles of LDPC codes or other ensembles with a given erasure-correcting threshold. These algorithms perform with an arbitrarily high level of accuracy.

Here, however, we leave out some efficient algorithms that require more specific estimates to perform distance verification with provable complexity. In particular, we do not address belief propagation (BP) algorithms, which can end at the stopping sets and therefore fail to furnish distance verification with an arbitrarily high likelihood. Some other algorithms also include impulse techniques [4] that apply list decoding BP algorithms to the randomly induced errors. Simulation results presented in [4] show that impulse techniques can also be effective in distance verification albeit with a lesser fidelity.

II. Background

Let $C[n, k]$ be a linear binary code of length n and dimension k. The problem of verifying distance d of a linear code (finding a minimum-weight codeword) is related to the decoding problem: find an error of minimum weight that gives the same syndrome as the received codeword. The number of operations N required for distance verification can usually be defined by some positive exponent $F = \overline{\lim}\,(\log_2 N)/n$ as $n \to \infty$. For example, for any code $C[n, k]$, inspection of all 2^k distinct codewords has (time) complexity exponent $F = R$, where $R = k/n$ is the code rate. Given substantially large memory, one can instead consider the syndrome table that stores the list of all q^r syndromes and coset leaders, where $r = n - k$. This setting gives (space) complexity $F = 1 - R$.

To proceed with the more efficient algorithms, we also need to consider some parameters of the shortened and punctured codes. Let G and H denote a generator and parity check matrices of a code $C[n, k]$. Let I be some subset of $g \geq k$ positions and J be the complementary subset of $\eta = n - g$, $\eta \leq r$, positions. Consider the punctured code $\widehat{C}_I = \{c_I : c \in C\}$ generated by submatrix G_I of size $k \times g$. The complementary shortened code $C_J = \{c_J : c_I = 0\}$ has parity-check matrix H_J of size $r \times \eta$. These matrices include at most k and η linearly independent rows, respectively. Let $b(G_I) = k - \operatorname{rank} G_I$ and $b(H_J) = \eta - \operatorname{rank} H_J$ denote the co-ranks of these two matrices. Throughout the paper, we use the following simple statement.

Lemma 1. *For any linear code $C[n, k]$, matrices G_I and H_J have equal co-ranks $b(G_I) = b(H_J)$ on the complementary subsets I and J.*

Proof: Code C_J has size $2^{\dim C_J} = 2^{\eta - \operatorname{rank} H_J} = 2^{b(H_J)}$ and contains all (shortened) code vectors c with $c_I = 0$. On the other hand, for a given matrix G_I, there are $2^{b(G_I)}$ vectors c with $c_I = 0$. Thus, $b(G_I) = b(H_J)$. ■

Now consider the shortened codes C_J of length $\eta = \theta n$ taken over over different sets J. Let codes C_J have the average size $N_\theta = 2^{\dim C_J}$. Then Markov's inequality gives another useful estimate.

978-1-5090-1806-2/16/$31.00 ©2016 IEEE

Corollary 1. *For any subset J and any $t > 0$, at most a fraction $\frac{1}{t}$ of the shortened codes C_J have size exceeding tN_θ.*

We will now consider two ensembles of regular LDPC codes. Ensemble $\mathbb{A}(\ell, m)$ is defined by the equiprobable $r \times n$ matrices H that have all columns of weight ℓ and all rows of weight $m = \ell n/r$. Below we take $m \geq \ell \geq 3$. This ensemble also includes a smaller LDPC ensemble $\mathbb{B}(\ell, m)$ originally proposed by Gallager [5]. For each code in $\mathbb{B}(\ell, m)$, its parity-check matrix H consists of ℓ horizontal blocks $H_1, ..., H_\ell$ of size $\frac{r}{\ell} \times n$. The first block H_1 consists of m consecutive unit matrices of size $\frac{r}{\ell} \times \frac{r}{\ell}$. Any other block H_i is obtained by some random permutation $\pi_i(n)$ of n columns of H_1. Ensembles $\mathbb{A}(\ell, m)$ and $\mathbb{B}(\ell, m)$ have similar spectra and achieve the best asymptotic distance for a given code rate $1 - \ell/m$ among various LDPC ensembles studied to date [6].

Note that LDPC codes are defined by non-generic, sparse parity check matrices H_J. Below we will relate the co-ranks $b_J = b(H_J)$ of these matrices H_J to the erasure-correcting thresholds of LDPC codes. In doing so, we extensively use the average weight spectra derived for the ensemble $\mathbb{B}(\ell, m)$ in [5] and for ensemble $\mathbb{A}(\ell, m)$ in [6]. We note, however, that this analysis can readily be extended to other ensembles with the known average weight spectra.

Let $\alpha = \ell/m = 1 - R$. For any parameter $\beta \in [0, 1]$, consider the equation

$$\frac{(1+t)^{m-1} + (1-t)^{m-1}}{(1+t)^m + (1-t)^m} = 1 - \beta \tag{1}$$

that has a single positive root t. Also, let $h(\beta)$ be the binary entropy function. Below we extensively use the parameter

$$q(\alpha, \beta) = \alpha \log_2 \frac{(1+t)^m + (1-t)^m}{2t^{\beta m}} - \alpha m h(\beta), \tag{2}$$

and also take $q(\alpha, \beta) = -\infty$ if m is odd and $\beta \geq 1 - \frac{1}{m}$. Then Theorem 4 of [6] shows that a given codeword of weight βn belongs to some code in ensemble $\mathbb{A}(\ell, m)$ with probability $P(\alpha, \beta)$ such that

$$\lim_{n\to\infty} \tfrac{1}{n} \log_2 P(\alpha, \beta) = q(\alpha, \beta) \tag{3}$$

Lemma 2. *For any given subset J of size θn, where $\theta \leq 1$, codes $C_J(\ell, m)$ of the shortened ensemble $\mathbb{A}_J(\ell, m)$ have the average number N_θ of nonzero codewords such that*

$$\lim_{n\to\infty} \tfrac{1}{n} \log_2 N_\theta = f(\theta) \tag{4}$$

$$f(\theta) := \max_{0<\beta\leq 1} \{q(\alpha, \beta\theta) + \theta h(\beta)\} \tag{5}$$

Proof: For any set J of size θn, consider codewords c of weight $\beta\theta n$ that have support on J. For any $\beta \in (0, 1]$, codes in $\mathbb{A}_J(\ell, m)$ contain the average number $N_\theta(\beta) = P(\alpha, \beta\theta) \binom{\theta n}{\beta\theta n}$ of such codewords c. Then

$$\max_{0<\beta\leq 1} \frac{\log_2 N_\theta(\beta)}{n} \sim \max_{0<\beta\leq 1} \{q(\alpha, \beta\theta) + \theta h(\beta)\} \tag{6}$$

which gives asymptotic equalities (4) and (5). ■

III. Distance verification for LDPC codes

A. Two main parameters for complexity estimates.

Two essential differences separate LDPC ensembles from random codes in regards to complexity estimates. These differences are closely related to two parameters, δ_* and θ_*, which are the roots of the equations

$$\begin{aligned} &\delta_* : h(\delta_*) + q(\alpha, \delta_*) = 0 \\ &\theta_* : f(\theta_*) = 0. \end{aligned} \tag{7}$$

Note that δ_* is the average relative code distance in the ensemble $\mathbb{A}(\ell, m)$. Indeed, for $\theta = 1$, equality (6) shows that the average number of codewords $N_\theta(\beta)$ of length n and weight βn has asymptotic order

$$\tfrac{1}{n} \log_2 N(\beta) \sim h(\beta) + q(\alpha, \beta) \tag{8}$$

For any code rate $R = 1 - \ell/m$, δ_* falls below the GV distance $h^{-1}(1-R)$ of random codes (see [5] and [6] for more details). For example, $\delta_* \sim 0.02$ for the $\mathbb{A}(3, 6)$ ensemble of rate $R = 1/2$, whereas $h^{-1}(0.5) \sim 0.11$. The smaller distances δ_* will reduce the complexity of distance verification.

Parameter θ_* also plays a significant role in distance verification. Namely, consider a code ensemble $\mathbb{C}$ of growing length $n \to \infty$. Let N_θ be the number of nonzero codewords in the shortened codes C_J averaged over all codes $C \in \mathbb{C}$ and all subsets J of size θn. Then we use the following statement.

Lemma 3. *Let the ensemble $\mathbb{C}$ have a vanishing average number $N_\theta \to 0$ of nonzero codewords in the shortened codes C_J of length θn. Then most codes $C \in \mathbb{C}$ correct most erasure subsets J, with the exception of a vanishing fraction of codes C and subsets J.*

Proof: A code $C \in \mathbb{C}$ fails to correct some erasure set J of weight θn iff code C_J has $N_J(C) \geq 1$ nonzero codewords. Let M_θ be the average fraction of such codes C_J taken over all codes C and all subsets J. Note that $M_\theta \leq N_\theta$. Per Markov's inequality, no more than a fraction $\sqrt{M_\theta}$ of codes C may leave a fraction $\sqrt{M_\theta}$ of sets J uncorrected. ■

More generally, we say that an ensemble of codes $\mathbb{C}$ has the erasure-correcting threshold θ_* if $N_\theta \to 0$ for any $\theta < \theta_*$ and $N_\theta \geq 1$ for any $\theta > \theta_*$ on the sets J of size θn. Here ensembles $\mathbb{A}(\ell, m)$ and $\mathbb{B}(\ell, m)$ satisfy Lemma 3 for any $\theta < \theta_*$ of (7). Thus, θ_* serves as a lower bound on the erasure-correcting capacity of LDPC codes under ML decoding. Alternatively, one can use other thresholds, such as the threshold for message-passing algorithms. Note also that ensembles $\mathbb{A}(\ell, m)$ and $\mathbb{B}(\ell, m)$ are permutation-invariant and therefore yield the same fraction of uncorrected codes C for each erasure subset J. Then for any $\varepsilon > 0$, the bound $N_J(C) \leq 2^{\varepsilon n}$ holds on *all* subsets J (except for a fraction of $2^{-\varepsilon n}$ of codes C) as long as $N_\theta \leq 1$.

For LDPC codes, $\theta_* < \alpha$, where $\alpha = 1 - R$ is the erasure-correcting threshold for random linear codes. For example, $\theta_* = 0.483$ for the ensemble $\mathbb{A}(3, 6)$ of LDPC codes. See also papers [7]–[10], where parameter θ_* is discussed in a greater detail for both ML decoding and message-passing decoder.

The reduced erasure-correcting threshold θ_* will increase complexity estimates for LDPC codes. In the sequel, we will show that the first factor (the smaller distance δ_*) outweighs the second factor (the smaller threshold θ_*) and reduces complexity of distance verification for LDPC codes.

B. Sliding window (SW) technique for LDPC codes

This technique of [11] decodes generic linear codes $C[n,k,d]$ generated by the randomly chosen $(Rn\times n)$ matrices G. Note that most such codes have full dimension $k=Rn$ and meet the asymptotic GV bound $d/n \to h^{-1}(1-R)$. It is shown in [11] that nearly full decoding (that has error probability similar to that of ML decoding) can be performed for most codes $C[n,k,d]$ with complexity of order $2^{nR(1-R)}$. Below we modify this algorithm for other ensembles of codes, such as $\mathbb{A}(\ell,m)$ or $\mathbb{B}(\ell,m)$.

Proposition 1. *Consider any ensemble of codes $\mathbb{C}$ with an average relative distance δ_* and an erasure-correcting bound θ_*. For most codes $C\in\mathbb{C}$, SW technique performs distance verification with complexity of exponential order 2^{Fn} or less, where*

$$F=(1-\theta_*)h(\delta_*) \tag{9}$$

Proof: Consider a sliding window $I(i,s)$, which is the set of s cyclically consecutive positions for some $i=0,\dots,n-1$. We choose $s=(1-\theta_*+\varepsilon)n$, where $\varepsilon>0$ is a parameter such that $\varepsilon\to 0$ as $n\to\infty$. A window $I(i,s)$ can change its Hamming weight only by one when it moves from position i to $i+1$; thus any codeword c of weight $d=\delta_* n$ has at least one window $I(i,s)$ with the average Hamming weight $v=\lfloor \delta_* s\rfloor$. For each window I, we inspect all $L=\binom{s}{v}$ vectors c_I of weight v. Here

$$\tfrac{1}{n}\log_2 L\sim(1-\theta_*+\varepsilon)h(\delta_*)$$

We then encode each vector c_I performing erasure correction on the complementary sets $J=\overline{I}$ of size $(\theta_*-\varepsilon)n$. Thus, a typical vector c_I generates the average number N_θ of nonzero codewords c_J. Given L vectors c_I and n sets $I=I(i,s)$, we obtain the average encoding complexity of $n^3N_\theta L$. Here we take the average over different codes $C\in\mathbb{C}$. Thus, at most a vanishing fraction n^{-1} of such codes have complexity above $n^4N_\theta L$ for all n subsets I. This gives (9) as $\varepsilon\to 0$. ■

C. Matching Bipartition (MB) technique for LDPC codes

Below we briefly discuss MB-technique of [12], [13]. It works for any linear code and yields the lowest asymptotic complexity for very high code rates $R\to 1$.

Proposition 2. *MB technique performs distance verification for a linear code of distance $\delta_* n$ with complexity of exponential order 2^{Fn}, where*

$$F=h(\delta_*)/2 \tag{10}$$

Proof: To find an (unknown) vector e of weight $d=\delta_* n$, we use the "left" window I_ℓ of length $s_\ell=\lfloor n/2\rfloor$ starting in any position i and the complementary "right" window I_r of length $s_r=\lceil n/2\rceil$. At least one choice of i gives the average weights $v_\ell=\lfloor d/2\rfloor$ and $v_r=\lceil d/2\rceil$ for truncated vectors e_ℓ and e_r in windows I_ℓ and I_r. The number L of vectors e_ℓ and e_r has the order of

$$\tfrac{1}{n}\log_2 L\sim\tfrac{1}{n}\log_2\tbinom{s_r}{v_r}\sim h(\delta_*)/2$$

We calculate the syndromes of all vectors e_ℓ and e_r and try to match two vectors with equal syndromes. This matching is performed by sorting the elements of the combined set with complexity of order $Ln\log_2 L$, which gives exponent (10). ■

Exponents (9) and (10) give the combined estimate

$$F=\min\{(1-\theta_*)h(\delta_*),h(\delta_*)/2\} \tag{11}$$

Here parameters δ_* and θ_* are defined for LDPC codes in (7).

D. Covering set (CS) technique for LDPC codes

This probabilistic technique was proposed in [14] and has become a benchmark in cryptography since the classical paper [15]. It lowers complexity estimate (11) for all but very high code rates $R\to 1$. CS technique has also been studied for distance verification of specific code families (see [16] and [17]); however, provable results [18], [19] are only known for generic random codes.

Below we choose any LDPC ensemble and describe CS technique in the following proposition.

Proposition 3. *Consider any code ensemble $\mathbb{C}$ with an average relative distance δ_* and an erasure-correcting bound θ_*. For most codes $C\in\mathbb{C}$, CS technique performs distance verification or corrects up to $\delta_* n$ errors with complexity of order 2^{Fn} or less, where*

$$F=h(\delta_*)-\theta_* h(\delta_*/\theta_*) \tag{12}$$

Proof: Let e be some unknown codeword of weight d in a given code $C\in\mathbb{C}$. Alternatively, we can consider an error vector e of weight d. To find e, we repeatedly try to cover all d nonzero positions of e with some subsets $J=\{i_1,\dots,i_s\}$ of $s=\theta n$ positions, where $\theta=\theta_*-\varepsilon$ and $\varepsilon\to 0$ as $n\to\infty$. To cover every possible d-set, we need no less than

$$T(n,s,d)=\binom{n}{d}\Big/\binom{s}{d}$$

sets J. Below we randomly choose a larger number of

$$T=T(n,s,d)n\ln n \tag{13}$$

sets J. Following Theorem 13.4 of [20] it is easy to see that T trials fail to yield such an (n,s,d)-covering with a probability less than $e^{-n\ln n}$.

Recall that $N_\theta\to 0$ for the shortened codes C_J. Let $C_J(b)$ be a code that contains 2^b-1 nonzero codewords for some $b=0,\dots,\theta n$. Also, let $\alpha_\theta(b)$ be the fraction of codes $C_J(b)$ in the ensemble $\mathbb{C}_J$. Then

$$N_\theta=\sum_{b=0}^{\theta n}\left(2^b-1\right)\alpha_\theta(b) \tag{14}$$

A parity-check matrix H_J of any code $C_J(b)$ has rank $s-b$ and size $r\times s$, where $r=n-k$ is the number of parity

checks. By Gaussian elimination, matrix H_J can be modified into a new $r \times s$ matrix $\mathcal{H}_J$ that includes b zero rows. We will also place $s-b$ unit columns $u_i = (0...01_i0...0)$ in the first positions $i \in [1, s-b]$ of $\mathcal{H}_J$, and b other columns g_j in the last positions $j \in [s-b+1, s]$. Let $v = \mathcal{H}e^T$ denote the syndrome of vector e (possibly modified by the Gaussian elimination procedure).

A. First, consider general error correction given a syndrome $v \neq 0$. If $b = 0$ in a given trial J, then matrix $\mathcal{H}_J$ has full rank and we obtain vector e of weight $wt(v)$. If $b > 0$, we assume that v contains only zero symbols in the last b positions. Then CS algorithm inspects all 2^b linear combinations (LC) of the last columns g_j. Let $LC(p)$ denote some LC that includes p columns. If $LC(p) + v$ has weight w, we obtain vector e of weight $w + p$ by adding w unit columns u_i.

The overall decoding algorithm successively tries to find a vector e of weight $d = 1, 2....$ For any given d, it runs over all subsets J and ends once we find a vector e of weight $w+p = d$. For any given code $C_J(b)$, this procedure includes one Gaussian elimination and up to b vector additions, which gives complexity $\mathcal{D}_\theta(b) \leq n^3 + rb2^b \leq n^3 2^b$. For a given set J, different codes $C_J(b)$ yield the average complexity

$$\mathcal{D}_\theta(J) \leq \sum_{b=0}^{\theta n} n^3 2^b \alpha_\theta(b) = n^3 N_\theta + n^3 \qquad (15)$$

Thus, CS algorithm has the total average complexity $\mathcal{D}_{ave} \sim n^3 T$ for all T sets J. Then at most a vanishing fraction $1/n$ of codes C have complexity $\mathcal{D} \geq n^4 T$, which gives the exponent (12) for the remaining codes in ensemble $\mathbb{C}$ as $n \to \infty$.

B. Vector e forms a codeword with syndrome $v = 0$. Then any code $C_J(0)$ has no nonzero codewords, and CS algorithm skips the above case $b = 0$. Also, we consider only $2^b - 1$ nonzero combinations LC(p) for the last b columns in any $C_J(b)$. Thus, we replace (15) with a similar inequality

$$\mathcal{D}_\theta(J) \leq \sum_{b=1}^{\theta n} n^3 \left(2^b - 1\right) \alpha_\theta(b) \leq n^3 N_\theta + n^3 \qquad (16)$$

that satisfies complexity bound (12). ■

Remark. The existing CS algorithms employ some stringent properties of random ensembles of linear codes. For example, the algorithm of [18] uses the fact that most random binary $r \times n$ matrices H, except an exponentially small fraction $\binom{n}{r}^{1-c}$ for $c > 1$, have all $r \times r$ submatrices H_J with nearly-full rank $r - b$, where

$$0 \leq b \leq b_{\max} = \sqrt{c \log_2 \binom{n}{r}} \qquad (17)$$

Thus, all shortened codes C_J have limited size 2^b for most linear codes C. For LDPC codes, we use a slightly weaker condition. Our technique discards codes C_J of large size 2^b that form an exponentially small fraction of all codes C_J.

Fig. 1 summarizes complexity estimates for LDPC codes. For comparison, we also plot two generic exponents valid for most linear codes. Note that these codes meet the GV bound and have parameters $h(\delta_*) = \theta_* = 1 - R$. Then the combination (11) of SW and MB algorithms gives exponent $F = \min\{R(1-R), (1-R)/2\}$, whereas exponent (12) of CS algorithm reads as $F = (1-R)\left[1 - h\left(\delta/(1-R)\right)\right]$. For LDPC codes, we similarly consider the exponents (11) and (12). Here we consider ensembles $\mathbb{A}(\ell, m)$ or $\mathbb{B}(\ell, m)$ for various LDPC (ℓ, m) codes with code rates ranging from 0.125 to 0.8. With the exception of low-rate codes, all LDPC codes of Fig. 1 achieve a substantial reduction in complexity exponent for distance verification compared to the generic linear codes.

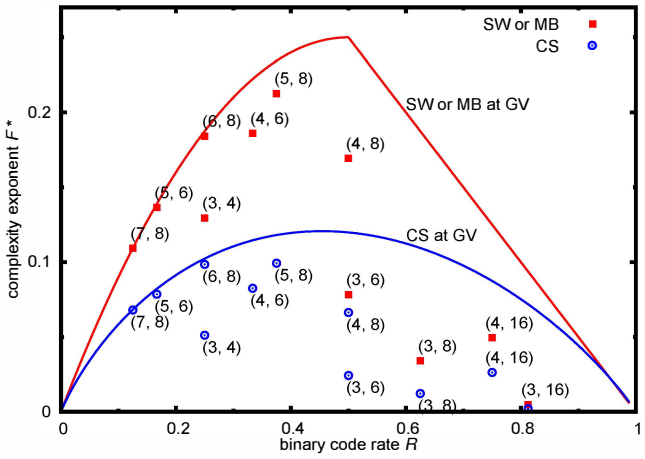

Fig. 1. Complexity exponents for the binary codes meeting the GV bound and for some (ℓ, m)-regular LDPC codes as indicated. Abbreviation "SW or MB" stands for the Sliding Window or Matching Bipartition techniques (marked with filled boxes), and "CS" stands for the Covering-Set technique (marked with empty circles).

IV. Further Extensions

In this paper, we study provable algorithms of distance verification for LDPC codes and derive complexity estimates using only the relative distance δ_* and the erasure-correcting threshold θ_* averaged over a given ensemble of codes. For LDPC codes, these algorithms exponentially reduce generic complexity estimates known for random linear codes. More generally, this approach can be used for any ensemble of codes with a given erasure-correcting threshold.

One particular extension is any ensemble of irregular LDPC codes with the known parameters δ_* and θ_*. Note that parameter θ_* has been studied for both ML decoding and message-passing decoding of irregular codes [7], [8], [10]. For ML decoding, this parameter can also be derived using the weight spectra obtained for irregular codes in papers [21], [22].

Another direction is to design more advanced algorithms of distance verification for LDPC codes. Most of such algorithms known to date for linear $[n, k]$ codes combine Matching Bipartition (MB) techniques with the Covering Set (CS) algorithms. In particular, the algorithm of [23] first applies CS technique seeking some slightly corrupted information set of k bits. It also tries to select some small subset of Δ parity bits, every time assuming that these bits are error-free. Then MB technique is applied to correct information bits in the

$[k+\Delta, k]$-code with Δ correct parity bits. This algorithm reduces the maximum complexity exponent $\max_R F(R)$ of CS technique from 0.1208 to 0.1167. A slightly more efficient algorithm of [24] (see also [25]) reduces this exponent to 0.1163 using a lightly corrupted block of length greater than k. Later, this algorithm has been re-established for cryptographic setting in [26], [27] with many applications related to the McEliece cryptosystem. More recently, the maximum complexity exponent $F(R)$ has been further reduced to 0.1019 using some robust MB techniques that allow randomly overlapping partitions [28]. An important observation is that both MB and CS techniques can be applied to LDPC codes; therefore our conjecture is that provable complexity bounds for distance verification also carry over to the above techniques. These more advanced algorithms can again slightly reduce the exponent of CS complexity for LDPC codes; however, their detailed description is beyond the scope of this paper.

Finally, one more approach is to combine LDPC-specific message-passing algorithms with the subsequent erasure correction. Such an approach has been used in [29] for quantum LDPC codes that require stringent self-orthogonality conditions. The corresponding complexity exponent closely approaches exponent $F(R)$ for self-orthogonal LDPC codes that have high code rate and low distance. For all other instances, this approach requires substantial improvements as complexity exponents exceed the exponent $F(R)$ obtained in the current paper.

Acknowledgment. The work of L.P. Pryadko was supported in part by ARO Grant W911NF-14-1-0272 and NSF Grant PHY-1416578. The work of A.A. Kovalev was supported in part by NSF Grant PHY-1415600.

REFERENCES

[1] A. Vardy, "The intractability of computing the minimum distance of a code," *IEEE Trans. Inform. Theory*, vol. 43, no. 6, p. 17571766, 1997.

[2] I. Dumer, D. Micciancio, and M. Sudan, "Hardness of approximating the minimum distance of a linear code," *IEEE Trans. Inform. Theory*, vol. 49, no. 1, pp. 22–37, 2003.

[3] Q. Cheng and D. Wan, "A deterministic reduction for the gap minimum distance problem," in *STOC 2009*, 2009, pp. 33–38.

[4] D. Declercq and M. Fossorier, "Improved impulse method to evaluate the low weight profile of sparse binary linear codes," in *2008 IEEE Intern. Symposium on Info. Theory*, July 2008, pp. 1963–1967.

[5] R. G. Gallager, *Low Density Parity Check Codes*. Cambridge, MA: M.I.T Press, 1963.

[6] S. Litsyn and V. Shevelev, "On ensembles of low-density parity-check codes: asymptotic distance distributions," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 887–908, Apr 2002.

[7] M. Luby, M. Mitzenmacher, A. Shokrollahi, and D. Spielmani, "Efficient erasure correcting codesl," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 569–584, February 2001.

[8] T. Richardson and R. Urbanke, "The capacity of low-density parity check codes under message passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, February 2001.

[9] T. Richardson, A. Shokrollahi, and R. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 619–637, February 2001.

[10] H. Pishro-Nik and F. Fekri, "On decoding of low-density parity-check codes over the binary erasure channel," *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 439–454, March 2004.

[11] G. S. Evseev, "Complexity of decoding for linear codes." *Probl. Peredachi Informacii*, vol. 19, no. 1, pp. 3–8, 1983, (In Russian). [Online]. Available: http://mi.mathnet.ru/eng/ppi1159

[12] I. I. Dumer, "On syndrome decoding of linear codes," in *Proc. Ninth All-Union Symp. Redundancy in Information Systems*. Nauka, May 1986, vol. 2, pp. 157–159, (In Russian).

[13] ——, "Two decoding algorithms for linear codes," *Probl. Peredachi Informacii*, vol. 25, no. 1, pp. 24–32, 1989, (In Russian). [Online]. Available: http://mi.mathnet.ru/eng/ppi635

[14] E. Prange, "The use of information sets in decoding cyclic codes," *Information Theory, IRE Transactions on*, vol. 8, no. 5, pp. 5–9, 1962.

[15] R. McEliece, "A public-key cryptosystem based on algebraic coding theory," JPL, Tech. Rep. DSN Progress Report 43-44, 1978.

[16] P. J. Lee and E. F. Brickell, "An observation on the security of mcelieces public-key cryptosystem," in *Advances in Cryptology - EUROCRYPT 1988*, 1988, pp. 275–280.

[17] J. S. Leon, "A probabilistic algorithm for computing minimum weights of large error-correcting codes," *IEEE Trans. Info. Theory*, vol. 34, no. 5, pp. 1354 –1359, Sep 1988.

[18] E. A. Kruk, "Decoding complexity bound for linear block codes," *Probl. Peredachi Inf.*, vol. 25, no. 3, pp. 103–107, 1989, (In Russian). [Online]. Available: http://mi.mathnet.ru/eng/ppi665

[19] J. T. Coffey and R. M. Goodman, "The complexity of information set decoding," *IEEE Trans. Info. Theory*, vol. 36, no. 5, pp. 1031 –1037, Sep 1990.

[20] P. Erdos and J. Spencer, *Probabilistic methods in combinatorics*. Budapest: Akademiai Kiado, 1974.

[21] C. Di, R. Urbanke, and T. Richardson, "Weight distributions: How deviant can you be?" in *Proc. Int. Symp. Information Theory (ISIT 2001), Washington, DC*, 2001, p. 50.

[22] S. Litsyn and V. Shevelev, "Distance distributions in ensembles of irregular low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3140–3159, Dec 2003.

[23] J. Stern, "A method for finding codewords of small weight," in *Coding Theory and Applications*, ser. LNCS, G. Cohen and J. Wolfmann, Eds. Heidelberg: Springer, 1989, vol. 388, pp. 106–113.

[24] I. Dumer, "On minimum distance decoding of linear codes," in *Fifth Soviet-Swedish intern. workshop Information theory*, G. Kabatianskii, Ed. Moscow: Nauka, Jan. 1991, pp. 50–52.

[25] A. Barg, "Complexity issues in coding theory," in *Handbook of Coding Theory*, V. Pless and W. C. Huffman, Eds. Amsterdam: Elsevier, 1998, pp. 649–754.

[26] M. Finiasz and N. Sendrier, "Security bounds for the design of code-based cryptosystems," in *Asiacrypt 2009*, ser. LNCS. Heidelberg: Springer, 2011, vol. 5912, pp. 88–105.

[27] D. J. Bernstein, T. Lange, and C. Peters, "Smaller decoding exponents: ball-collision decoding," in *CRYPTO 2011*, ser. LNCS. Heidelberg: Springer, 2011, vol. 6841, pp. 743–760.

[28] A. Becker, A. Joux, A. May, and A. Meurer, "Decoding random binary linear codes in 2 n/20 : How $1 + 1 = 0$ improves information set decoding," in *EUROCRYPT 2012*, ser. LNCS. Heidelberg: Springer, 2012, vol. 7237, pp. 520–536.

[29] I. Dumer, A. A. Kovalev, and L. P. Pryadko, "Numerical techniques for finding the distances of quantum codes," in *2014 IEEE Intern. Symposium on Info. Theory*, Honolulu, HI, June 2014, pp. 1086–1090.