# Compressed Sensing over Galois Fields: Sensing Bounds and Recovery Algorithms

Heung-No Lee, JinTaek Seong, and Suje Lee

Gwangju Inst. of Sci. and Tech.(GIST), Korea

heungno@gist.ac.kr

**Abstract: In this paper, we aim to consider compressed sensing (CS) system over the Galois fields (GF). Sensing bounds based on the classical Gilbert Varshamov bounds are discussed. We obtain the `spark` of sensing matrices randomly generated in GF which can be used to give another perspective on the sensing bounds. A signal recovery algorithm based on posterior distribution updates is introduced. We anticipate this CS framework over GF will be useful for digital compressed sensing systems.**

**Keywords:** Compressed Sensing, Gilbert Varshamov bounds, Syndrome Decoding, Minimum distance, MAP algorithms.

## I. Introduction

Compressed sensing has provided a new signal acquisition framework with which one can take samples of a given signal of interest while compressing it simultaneously. This compressed sample taking is done via linear projection of the given signal against a prescribed set of kernels, i.e., one linearly projected sample per kernel. In its standard form, this compressed sensing operation is developed over the field of real numbers. In this presentation, we are interested in the development of compressed sensing over the finite fields. Fundamental limits on sensing measurement requirements as we vary the size of the finite field will be discussed. When compressed sensing is put to work in digital systems, the signals and the kernels should be represented in a finite precision manner anyhow; thus, the study of compressed sensing over finite fields should be of interest for implementation point of view. We aim to present our understanding on compressive sensing via Gilbert Varshamov (GV) bounds, new results on average `spark` calculation results, and proposal of the a posteriori (AP) signal recovery algorithm, and provide discussion on how they are related with each other.

We make note of existence of a few prior studies relevant to the content of this paper. Draper and Malekpour [2] have studied compressed sensing over finite fields and obtained fundamental bounds on sensing requirements using the error exponent analysis techniques of the channel coding theory. Ardestanizadeh, Cheraghchi, and Shokrollahi [6] have studied the question how much bit precision on the compressive measurements will be needed for good recovery of sparse signals of a finite size alphabet, say $q$. They assumed the use of Vandermonde frames [5] and obtained that the precision requirement is $O(K\log_2 q + K^2 \log \frac{N}{K})$. Zhang and Pfister [4] discussed the connection between compressed sensing and error correction codes, and proposed the use of low density parity check matrices over GF($q$) and a verification based iterative decoding schemes.

## II. Compressed Sensing via Syndrome Decoding

In this section, we aim to draw analogy between parity checking in coding theory and the under-determined equation in compressed sensing by recasting the basic compressed sensing equation

$$y = Fx \tag{1}$$

as a coding theoretic parity-checking equation. Treat $y$ as an $M \times 1$ syndrome vector, $F$ as an $M \times N$ parity check matrix, $M < N$, and $x$ as an $N \times 1$ *K-sparse* error vector. Note that this model is valid for real, complex, and finite fields GF($q$). Finite fields can be useful for implementing the CS system in digital forms, with a finite precision representation, say $\log_2(q)$ bit precision, done to the coefficients of the elements of the sensing matrices and signals.

We assume that $K \le t$ where $t$ means the number of errors a given code defined by an $F$ can correct. Let $U = N - M$. The rate $R$ of the code is $U/N$. We can then find the $N \times U$ generator matrix $G$ from $F$ using the relationship that $FG = 0$ (e.g. using Gaussian elimination on $F$) where 0 denotes the $M \times U$ all zero matrix. Let $\mathcal{C}$ be the codebook—collection of all codewords. Each $N \times 1$ codeword $c$ can be generated by multiplying an arbitrary $U \times 1$ message vector $m$ to the generator matrix, i.e., $c = Gm$. We assume $c$ is sent over a noisy channel where the noisy channel introduces an additive random error pattern $x$ to $c$, and the output of the channel is $z = c + x$.

In this setting, parity checking on $z$ shall return the zero syndrome, i.e., $y = Fz = F(c+x) = Fx$, unless there is zero errors or the error pattern $x$ is a codeword, i.e. $x \in \mathcal{C}$; otherwise it will give a non-zero syndrome vector. The code is linear and hence it contains the all-zero codeword. The error correction capability of this code $\mathcal{C}$ can be parameterized by its minimum distance $d_{min}$. The minimum distance $d_{min}$ is the minimum

Hamming weight (the number of non-zero coefficients) of any codeword, since the code is linear, i.e.,

$$d_{\min} \triangleq \min_{c\neq 0, c\in\mathcal{C}} w_H(c). \quad (2)$$

But a codeword is a word that satisfies the parity check equation, i.e., $Fx=0$. From this observation, we may also write that $d_{\min}$ is also the smallest number $d$ that there exists a set of $d$ columns of matrix $F$ that are linearly dependent; this definition is the same as that of the `spark` in compressed sensing. This discussion will continue further in Section III. From the coding theory, we note that, a code defined by its parity check matrix $F$ with $d_{min}$ can correct *all error patterns with weight smaller or equal to t*, and $t$ is given by

$$t = \left\lfloor \frac{d_{\min}-1}{2} \right\rfloor \approx \frac{d_{\min}}{2}. \quad (3)$$

Our discussion up to (3) implies that all $K$-sparse error vectors $x$ can be uniquely determined from the syndrome equation $y = F x$ as long as $K \le t$. Notice that this is a deterministic guarantee, rather than probabilistic, on the recovery of the sparse vectors. Such a code or a matrix $F$ with $d_{min}$ can be constructed. Namely, we can construct an $F$ so that any collection of less than or equal to $d_{min} - 1$ columns of $F$ is linearly independent. This means that $d_{min}$ - 1 can be as large as the rank of $F$ which is further upper bounded by $M$ since $M < N$. Hence, we have the Singleton bound,

$$d_{\min} - 1 \le M. \quad (4)$$

Those codes that achieve the Singleton bound with equality are called *maximum distance separable* codes. They include the *repetition* codes and *Reed Solomon* (RS) codes. Real- or complex-valued RS code like sensing matrix $F$ with $d_{min}$ can also be found. The examples are given in [5],[7]. From (4), we can obtain $\frac{M}{2N} \ge \frac{d_{\min}}{2N} + \frac{1}{2N}$, by dividing both sides by $2N$. By defining the compression ratio $\rho_{comp} \triangleq \frac{M}{N}$ and the error correction ratio (ECR) $\rho_t \triangleq \frac{t}{N}$, we have

$$\rho_{comp} \ge 2\rho_t. \quad (5)$$

We call (5) the *CS Singleton bound*. Any $x$ whose sparsity ratio $\rho_{sp} \triangleq \frac{K}{N}$ is smaller than or equal to ECR (i.e., $K \le t$), can be uniquely determined from syndrome $y$. Fig. 1 shows the CS Singleton bound.

On the other hand, the Gilbert-Varshamov bound tells us the existence of a $t$ error correcting linear block code. The rate $R$ of such a code is given by,

$$R(\delta) \ge 1 - H_q(\delta) \quad (6)$$

where $\delta \triangleq \frac{d_{\min}}{N}$, $R = \frac{N-M}{N}$ and $H_q(\delta)$ is the $q$-ary entropy function. It is the lower bound on the rate required to have the relative minimum distance $\delta$. Eq. (6) can then be written as,

$$\rho_{comp} \le H_q(2\rho_{sp}) \quad (7)$$

for $\rho_{sp} \in [0, 0.5]$.

It is interesting to note that $H_q(2\rho_{sp})$ approaches the line with slope 2 as $q$ increases. The required code rate can be as large as what this lower bound predicts for a long block length. It is then an upper bound on the redundancy. The number of check equations required for a sensing matrix to have the relative minimum distance is at most what this bound can tell us. One needs at most this much redundancy to be able to find at least a single sensing matrix with the relative minimum distance $\delta$. It can be shown that an ensemble of parity check (PC) codes, say $\mathcal{C}(N, d_s, d_c, q)$ block codes of length $N$, check degree $d_c$, signal element degree $d_s$, and GF($q$), approach the GV bound from *above* as the degrees are increased. This is from our observation. Thus, GV bounds in fact work as a benchmark, instead of as an upper bound. The check degree and the signal element degree indicate the number of non-zero entries in any row of a sensing matrix and the number of non-zero entries in any column respectively. We focus on the cases in this paper that the degrees are fixed for each row and column. Thus, for a compressed sensing system with a large field GF($q$), the sufficient condition is close to

$$\rho_{comp} \gtrsim 2\rho_{sp}. \quad (8)$$

This means that if $\rho_{comp} \gtrsim 2\rho_{sp}$, a good sensing matrix exists and can be found randomly. As the dimension of the system approaches infinity, a randomly selected code out of an ensemble will behave as good as what these bounds can predict, with probability getting close to 1.

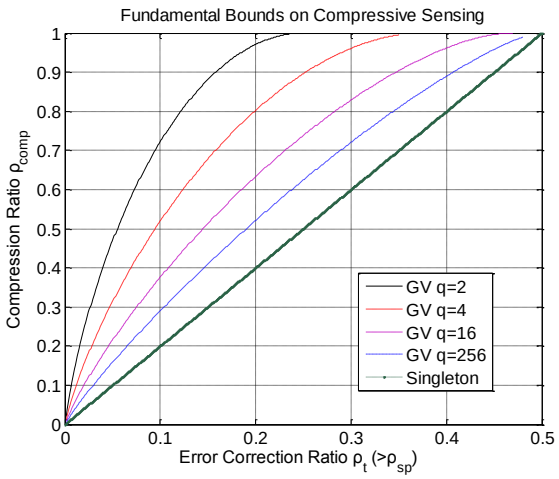


Figure 1 : Gilbert-Varshamov Compressed Sensing Bounds for Sensing Matrices over GF($q$) and The Singleton Bound

## III. The spark of Sensing Matrix over GF($q$)

In this section, we aim to find the ensemble average sparks of GF($q$) LDPC codes. The spark of an $M \times N$ matrix is the smallest number $S$ such that there exists a set of $S$ columns of the matrix that are linearly dependent. One should note that spark of a sensing matrix and $d_{min}$ for a parity check matrix are exactly the same. In fact, they are two different names for the same thing. From the Singleton bound, then, $S-1 \le M$. Finding the spark of a sensing matrix is of paramount importance in compressed sensing because it can provide a limit on how sparse a signal has to be for guaranteed unique recovery. For example, if the spark of a certain $M \times N$ sensing matrix $F$ is given to be $S$, then any signal $x$ with sparsity $K$ can be uniquely determined from the combinatorial $L_0$ minimization routine, as long as $K \le \frac{S}{2}$. That is, $K \le \frac{S}{2}$ is the sufficient condition for the $L_0$ norm minimization solution, subject to $y = Fx$ constraint, to return the unique solution. Otherwise, say $K = \frac{S}{2}+1$ for example, the solution is not unique, which can be easily proved. The $L_0$ minimization is known as an NP-hard problem since it is combinatorial. The sufficient condition, thus, provides a meaningful benchmark on the required sparsity.

Finding the spark of a matrix is thus desired, but it requires a combinatorial search and hence is an NP-hard problem by itself. In this paper, we find the average spark of an ensemble of sensing matrices. For a system with a large block length $N$, the average spark of an ensemble of sensing matrices is close to the spark of an individual sensing matrix randomly selected out of the ensemble. That is, it can be shown that the spark of an individual sensing matrix concentrates around the ensemble averaged spark.

**Theorem 1** : The average spark of an ensemble of GF($q$) random sensing matrices, i.e., $\mathcal{C}(N, d_s, d_c, q)$, is given by

$$\begin{aligned}\text{spark}(N, d_s, d_c, q) = \min\Big\{ & d \in \{2, \cdots, N\} \Big| \\ & d\left(1-d_v\right)\ln(q-1) + \left(d_s\right)\ln\tfrac{N_d}{\binom{N}{d}} \ge 0 \Big\}\end{aligned} \tag{9}$$

where the variable $N_d$ is a function of the *chec*k degree $d_c$ and it is given by

$$N_d = \text{Coeff}_d\left( p(x)^{\frac{N}{d_c}} \right) \tag{10}$$

where $p(x) = \sum_{i=0}^{d_c} p_i x^i$, $p_i = \binom{d_c}{i}$, for even $i$, $p_i = 0$ for odd $i$, $0 \le i \le d_c$, $\text{Coeff}_d(\cdot)$ denotes the coefficient of the term $x^d$ in the expansion of the argument polynomial, and we assume $\frac{N}{d_c}$ is integer.

## IV. Signal Detection Algorithms

In this section, we aim to discuss how to detect the sparse signal $x$ measured from a sensing matrix $F$ selected randomly out of an ensemble $\mathcal{C}(N, d_s, d_c, q)$ [3].

The sparse signal values can be obtained by solving the following problem:

$$\tilde{x}_t := \arg\max_{\tau_t \in \text{GF}(q)} P(x_t = \tau_t \mid y, C) \quad \text{s.t. } y = F\,x \tag{11}$$

where $t = 0, 1, \cdots, N-1$ and the symbol "$C$" in the conditioning compartment means satisfaction of all the $M$ "check" relations. The function $P(x_t = \tau_t \mid y, C)$ is the posterior distribution, given the observation, and after enforcing the check relations. This posterior distribution is updated for each element of the signal $x$.

**Theorem 2**: The *aposteriori* probability (AP) that the first value, $x_0 = \tau_0 \in \text{GF}(q)$, given the observation $y$ and enforcing the checks (checks should be satisfied), is given by

$$\begin{aligned}\text{P}(x_t = \tau \mid y, C) = & \frac{P\left(x_t = \tau \mid y\right)}{P\left(C \mid y\right)} \times \\ & \prod_{p=1}^{d_c} \left[ \sum_{x_{t,p}} P\left(C_{i_p} \mid x_t = \tau_0, x_{t,p}, y\right) P(x_{0,0} \mid y) \right]\end{aligned} \tag{12}$$

We apply the same procedure and obtain the AP result for each element of $x$. A single round of calculation of the posterior distribution $\{\text{P}(x_t = \tau_t \mid \mathbf{y}, C) : \tau_t \in \mathcal{X}\}$ for each and every element, $t = 0, 1, \cdots, N-1$, constitutes a single iteration. In a single iteration, therefore, all $N$ different posterior distributions are updated once. We repeat this iteration multiple times. Why do we need iterations? Why can't we be satisfied with a single iteration? This has to do with our choice on the density, controlled by the two degrees, of the sensing matrix. Choosing a sparse sensing matrix (small degrees) would be desired because when the matrix is sparse, the iterative algorithm works well, a lesson learned from the experience on the low density parity check codes. In a single iteration, only local information is gathered because of sparse connections. Through multiple iterations, it is hoped that and thus the algorithm is only sub-optimal, the entire information from observation $y$ available via enforcing checking relations prescribed in the sparse matrix can be gathered. An enough number of iterations should be repeated before convergence can be seen on the value of each signal element.

It can be shown that the *check* posterior results, i.e.,

$\sum_{x_{t,p}} P\left(C_{i_p} \mid x_t = \tau_0, x_{t,p}, y\right) P(x_{t,p} \mid y)$ in (12), can be obtained from a series of convolution operations of the probability distribution functions of the signal variables connected to the pertinent check $C_{i_p}$. For example, suppose $x_t$ is connected via its first check to $x_3$ and $x_6$, then it is the convolution of the two distributions, one for the signal element $x_3$ and the other for $x_6$. The convolution operations can be conveniently done in the frequency domain using FFT and IFFT.

In [3], a couple of ideas on iterations based on identifying the support set detection are also included. One of them is aiming to obtain the posterior distribution of the state $S_t$ of $t$-th signal element. A state value $S_t$ is binary, 1 for the non-zero value of $x_t$, and 0 for the zero values. Then, the state posterior either $\Pr\{S_t = 1 \mid y, C\}$ and $\Pr\{S_t = 0 \mid y, C\}$ can be updated in each iteration. The log ratio of the posterior probabilities on the state is maintained in each iteration. For the state posterior calculation, the prior information on signal sparseness is utilized. When the log ratio is greater than 0, then the pertinent state is more like to be 1; otherwise it is zero. At the end of each iteration, we can threshold the log ratios, determine the indices of non-zero states, and form an estimate of the support set. Once a support set estimate of size $K$ is given, one can then attempt to solve the over-determined problem (by collecting only those columns of matrix $F$ and those elements of $x$ corresponding to the non-zero indices) and find a solution $\tilde{x}$. If this one is found to satisfy the observation, i.e., $y = F\tilde{x}$, i.e., it is declared to be the solution; then the iteration can be put to stop.

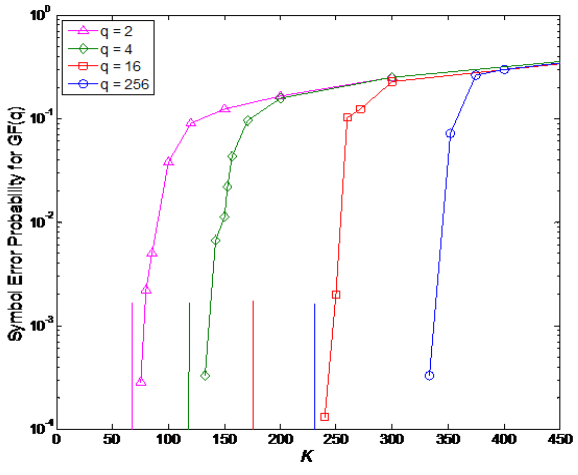


Figure 2 : Simulation results of a $\mathcal{C}(N$=1200, $d_s$ = 3, $d_c$ = 6) code with different field sizes, compared with the Gilbert-Varshamov bounds indicated by lines.

## V. Simulation Results

Figure 2 shows the Monte Carlo simulation results of our MAP algorithms. The block length is $N$ = 1200. The number of observation is $M$ = 600; the maximum number of iterations is 20. For each sensing matrix, selected randomly out of $\mathcal{C}(N = 1200, d_s = 3, d_c = 6, q)$ ensemble, large enough signal vectors with sparsity $K$ are simulated with an aim to obtain at least 1000 errors for each simulation point. In addition, the same procedure is repeated over for 50 matrix selections, and thus a little bit of averaging is also done for matrix selections within a particular ensemble. Also, indicated in the figure are the sparsities obtained from the Gilbert-Varshamov bounds for $q$=2, 4, 16 and 256 and for (3, 6) codes. They are indicated as the lines, at 65, 120, 175, and 230, respectively in Figure 2. In addition, Table I shows the sparsity of various rate 1/2 matrices.

Table 1: The `spark` $S$ and relative `spark` $S/N$ (inside the parenthesis) obtained from Theorem 1 for ($N$=1200, $d_v$, $d_c$) ensembles and GF($q$). The rate $M/N$ is 1/2.

| ($d_v$, $d_c$) | $q$=2 | $q$=4 | $q$=16 | $q$=256 |
|---|---|---|---|---|
| (3,6) | 32 (0.027) | 64 (0.54) | 108 (0.09) | 121(0.10) |
| (4,8) | 78 (0.065) | 146(0.12) | 235 (0.20) | 284(0.24) |
| (5,10) | 102 (0.085) | 187(0.16) | 294 (0.25) | 370(0.30) |

We note that the sparsity limits obtained from simulation are much larger. Namely, they are 70, 130, 240, and 325 obtained from simulation. The Singleton bound at $N$=1200 gives a `spark` of 600 for rate 1/2 code.

## VI. Conclusion

We note that all three measures, the sparsity obtained from GV bounds, the ensemble average sparsity obtained from Theorem 1, and the simulation results of the iterative recovery algorithm, agree to the observation that as the field size $q$ increases, a given sensing matrix of rate 1/2 can have large `spark` and thus can be used to detect the signals with a larger sparsity $K$. Simulation results show that the iterative algorithm can far surpass the predictions made by the average `spark`s as well as by the GV bounds, which is very interesting, and calls for further study. We also note that as the field size is increased, the compressed sensing bound is $M \geq 2K$ for unique recovery under the Singleton bound. A sensing matrix that satisfies this can be found easily from the random construction, and the iterative recovery algorithm introduced here can be used to even surpass it.

**Acknowledgements**

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (Do-Yak Research Program, N0. 2011-0016496)

## References


[1] D. Baron, S. Sarvotham, R. G. Baraniuk, “Bayesian Sensing via Belief Propagation”, *IEEE Sig. Proc*., vol. 58, no. 1, pp. 269-280, Jan. 2010.

[2] Stark C. Draper and Sheida Malekpour, “Compressed Sensing over Finite Fields,” Proc. of *IEEE ISIT* , Seoul, Korea, 2009.

[3] Heung-No Lee, *Introduction to Compressed Sensing*, Lecture Note, 2011 Spring Semester, GIST, Korea.

[4] F. Zhang and H. D. Pfister, “Compressed Sensing and Linear Codes over Real Numbers,” arXiv:0806.3243. Accepted for publication in *IEEE Trans. Info. Theory*.

[5] M. Akcakaya and V. Tarokh, “A frame construction and a universal distortion bound for sparse representations,” *IEEE Trans. on Signal Proc.*, vol. 56, pp. 2443-2450, 2008.

[6] E. Ardestanizadeh, M. Cheraghchi, and A. Shokrollahi, “Bit Precision Analysis for Compressed Sensing,” Prof. of *IEEE ISIT*, Seoul, Korea, June 28-July 3, 2009.

[7] M. Vetterli, P. Marziliano, T. Blu, “Sampling signals with finite rate of innovation,” *IEEE Trans. Signal Proc*., vol.50, no. 6, pp. 1417-1428, June, 2002.