Cryptography and Communications
https://doi.org/10.1007/s12095-024-00712-3

RESEARCH

Integer syndrome decoding in the presence of noise

Vlad-Florin Drăgoi[1,2] · Brice Colombier[3] · Pierre-Louis Cayrel[3] · Vincent Grosso[3]

Received: 30 October 2023 / Accepted: 23 March 2024

Abstract
Code-based cryptography received attention after the NIST started the post-quantum cryptography standardization process in 2016. A central NP-hard problem is the binary syndrome decoding problem, on which the security of many code-based cryptosystems lies. The best known methods to solve this problem all stem from the information-set decoding strategy, first introduced by Prange in 1962. A recent line of work considers augmented versions of this strategy, with hints typically provided by side-channel information. In this work, we consider the integer syndrome decoding problem, where the integer syndrome is available but might be noisy. We study how the performance of the decoder is affected by the noise. First we identify the noise model as being close to a centered in zero binomial distribution. Second we model the probability of success of the ISD-score decoder in presence of a binomial noise. Third, we demonstrate that with high probability our algorithm finds the solution as long as the noise parameter d is linear in t (the Hamming weight of the solution) and t is sub-linear in the code-length. We provide experimental results on cryptographic parameters for the BIKE and *Classic McEliece* cryptosystems, which are both candidates for the fourth round of the NIST standardization process.

Keywords Code-based cryptography · Syndrome decoding problem · Information-set decoding

✉ Vlad-Florin Drăgoi
vlad.dragoi@uav.ro

Brice Colombier
b.colombier@univ-st-etienne.fr

Pierre-Louis Cayrel
pierre.louis.cayrel@univ-st-etienne.fr

Vincent Grosso
vincent.grosso@univ-st-etienne.fr

[1] Faculty of Exact Sciences, Aurel Vlaicu University of Arad, Arad, Romania

[2] LITIS, University of Rouen Normandy, Av. de l'Université, 76800 Saint-Étienne-du-Rouvray, France

[3] UJM-Saint-Etienne, CNRS Laboratoire Hubert Curien UMR, University of Lyon, 5516 F-42023 Saint-Etienne, France

Published online: 10 May 2024

1 Introduction

With the practical feasibility of a quantum computer of sufficient capacity getting more and more probable by the day, the threat posed by Shor's algorithm [47] on number theory based cryptosystems grows as well. To address this threat, NIST began a standardization process in 2016 for post-quantum cryptography. The fourth round of this process started in July 2022 when, in the Key Encapsulation Mechanism category, four candidates were submitted. Among them, the *Classic McEliece* [2] and the BIKE [3] cryptosystems are two solutions based on error-correcting codes. Their security relies on the intractability of the binary syndrome decoding problem (SDP) [6]. The SDP is the core hard problem of several cryptographic constructions, e.g., the FSB hash function [4], the SYND stream cipher [28] or the Stern identification scheme [51]. Given a parity-check matrix $\boldsymbol{H}$ of a binary linear code, a binary syndrome vector $\boldsymbol{s}^*$ and an integer t, the SDP consists in asking for fixed Hamming weight ($\mathrm{HW}(x) = t$) solution to the linear system $\boldsymbol{Hx} = \boldsymbol{s}^*$. There are three main techniques for solving the SDP: statistical decoding [14, 18, 27, 34, 43], information set decoding (ISD) [5, 9, 11, 12, 22, 23, 26, 36, 37, 39, 40, 46, 50] and generalized inverse based decoding [52]. Information Set Decoding was originally proposed by Prange in 1962 [46], and it has been incrementally refined since by Lee and Brickell [36], Stern [30, 50] and, more recently, by May, Meurer and Thomae [39] and by Becker, Joux, May and Meurer [5]. The complexity of the ISD method has been used to better tune the parameters of the cryptosystems [24] according to the required security levels.

1.1 Integer syndrome decoding

One recent line of work considers modified versions of the SDP, for which additional information is available, for instance via side-channel analysis on implementations of the aforementioned cryptosystems. In [33], authors study the case where parts of the error are known, or only their Hamming weight. The case where the *integer* syndrome $\boldsymbol{s}$ is available, instead of the binary one, as if the matrix-vector multiplication had been performed in the integer ring instead of the binary finite field, is considered in [20]. One method to obtain the integer syndrome is by laser fault injection attack, as presented in [15]. The problem one has to solve in this case is the integer syndrome decoding, referred to as $\mathbb{N} - \mathsf{SDP}$, where the input is the parity-check matrix $\boldsymbol{H}$, the integer syndrome vector $\boldsymbol{s}$ and the weight of the solution t. The same question is raised, whether $\boldsymbol{Hx} = \boldsymbol{s}$ admits a solution of weight t. This problem can be tackled down by means of Integer Linear Programming [15] or probabilistic methods [25]. Another method of obtaining an integer syndrome, much more feasible and realistic than laser fault injection, is by side-channel analysis [16].

Due to physical factors, the integer entries of the syndrome might not be perfectly accurate. Hence, in the resulting problem, the $\mathbb{N} - \mathsf{SDP}$ in the presence of noise, we are given a noisy integer syndrome $\widetilde{\boldsymbol{s}} = \boldsymbol{s} + \epsilon$, where ϵ models the noise as a vector of random variables. The solution proposed in [16] uses a combination of ISD techniques and the score decoder from [25]. In [16] the performance of the algorithm was evaluated by simulations. Still, no theoretical evidence exists for the performance of the ISD-score decoder. On top of that, the performance certainly depends on the distribution of the noise, and for that one needs to carry a deeper investigation into the side-channel part.

Recently, the leakage model (Hamming weight or distance) as well as various noise distributions were analyzed in [29]. It was shown that if an attacker has more that the final result (or approximation of the final result) then ISD-score decoders can be enhances or even outperformed. In particular, in [29] the attacker model assumes intermediate estimates of the syndrome computations are leaked. Hence, the extra information is used to locally correlate the information with the solution, making the resulting attacks more powerful.

1.2 Related work

Learning with errors and hints Not only code-based cryptosystems are vulnerable to such attacks. Similar results were obtained in the context of lattice-based cryptosystems by Bootle et al. [10]. The BLISS cryptosystem was cryptanalysed by means of similar hybrid attacks, where side-channel attacks revealed an Integer version of the Learning With Errors (ILWE). The ILWE problem is the lattice-based equivalent of the $\mathbb{N}-\mathsf{SDP}$. However, ILWE was solved with another technique that does not seem to work for $\mathbb{N}-\mathsf{SDP}$. Nevertheless, it points out that such scenarios extend broader than code-based cryptography.

Quantitative group testing Quantitative Group Testing (QGT) is an active field of research, lately boosted by the COVID-19 epidemic. In the QGT we are given a large population out of which some individuals suffer from a disease, and the goal is to identify the infected individuals. Possible applications of QGT go from bio-informatics [13], traffic monitoring [54] and confidential data transfer [1, 19] to machine learning [38, 55]. The $\mathbb{N}-\mathsf{SDP}$ can be also seen as a QGT in presence of noise. As we shall demonstrate, the algorithm we propose here, solves a noisy QGT instance, by adapting and improving (using coding theory tools, such as ISD techniques) a recent solution to the classical QGT [25]. We compare our findings with the results from [25] in two ways.

- In the noiseless setting (the algorithm we propose can be applied to a zero noise distribution) we obtain less restrictive conditions on the parameters for attaining a high success probability. Also, smaller syndrome entries are required in our case for finding the solution of $\mathbb{N}-\mathsf{SDP}$.
- In the noisy setting, by working on the proofs in [25], we derive a condition on the parameters to successfully retrieve a solution. However, we show that these conditions are more restrictive than ours, hence restraining even more the set of parameters.

1.3 Contributions

In this article, we analyze in detail the ISD-score decoder algorithm for the noisy $\mathbb{N}-\mathsf{SDP}$ problem and provide the following contributions.

Noise model We will focus on a Binomial noise model, more precisely, the vector $\boldsymbol{\epsilon}$ is such that each $\epsilon_i \sim -d + \mathcal{B}(2d, \frac{1}{2})$ (binomial centered in zero). One of the arguments for our choice is based on then noise description from [16, 29]. Due to implementation restrictions (width of the representation) there are differences in the observed noise for different widths (parameter w in [29]). The first type of errors in the estimation of the integer syndrome come from the accuracy of the side-channel distinguisher. Since, the accuracy corresponds to the probability of a correct guess, any wrong guess of the side-channel distinguisher will lead to an overestimation of the exact values with high probability. The second type of errors mentioned in [16, 29] is the double-cancellation and refers to the errors made in the approximation of the hamming distance between two vectors using the Hamming weight of these vectors. Thirdly, during computations the errors become dependent, inducing a propagation phenomenon.

Here, we will experimentally consider all these factors and show that when the size of the words increases the binomial noise model is a proper theoretical model for all the considered instances.

We simulated realistic side-channel noise on all cryptographic parameters and noticed the following. Our theoretical model, *i.e.* the binomial noise, fits all real cryptographic scenarios with d being from linear in t in the worst case (high σ values for extreme real cases) down to constant (for small and more realistic σ values). We show that all real scenarios lead to $d < \frac{t}{4}$ which, as we shall see later, is an acceptable noise level for practical purposes.

Performance of the ISD-score decoder We demonstrate that the ISD-score decoder finds a solution to the $\mathbb{N}-\mathsf{SDP}$ in the presence of noise with high probability, as long as the weight is sub-linear in n. Letting n, k, t, d be the $\mathbb{N}-\mathsf{SDP}$ parameters, δ a small constant (usually less than 3) and $W(x)$ the Lambert W function, we demonstrate the following.

Theorem Let $I = \left[\sqrt{\frac{t+2d}{n-k} W\left(\frac{n-t}{n-k-t+\delta+1}\frac{e\sqrt{2}}{\pi}\right)^2}, 1-\sqrt{\frac{t+2d-1}{n-k} W\left(\frac{t}{\delta+1}\frac{2e}{\pi}\right)^2}\right]$ and $\epsilon_i \sim -d + \mathcal{B}(2d, \frac{1}{2})$. If I is non-empty, then there is a value $\beta \in I$ such that the ISD-score decoder succeeds in finding a solution with probability at least

$$\left(1-\frac{e(n-t)}{\sqrt{2}\pi\beta(n-k-t+\delta+1)}\sqrt{\frac{t+2d}{n-k}}e^{-\frac{(n-k)\beta^2}{2(t+2d)}}\right)\left(1-\frac{et}{\pi(1-\beta)(\delta+1)}\sqrt{\frac{t+2d-1}{n-k}}e^{-\frac{(n-k)(1-\beta)^2}{2(t+2d-1)}}\right).$$

To reach our goal we partially build our demonstration on the techniques used in [25]. We incorporate the noise models into these techniques and, by using sharper inequalities, determine a much clearer condition for having a higher probability of success. For a more readable variant of our result we propose a slightly weaker version. We thus demonstrate the following.

Proposition Let $I_\beta = \left[\sqrt{\frac{2(t+2d)}{n-k}\ln\frac{n-t}{n-k-t+\delta+1}}, 1-\sqrt{\frac{2(t+2d-1)}{n-k}\ln\frac{t}{\delta+1}}\right]$ and $\epsilon_i \sim -d + \mathcal{B}(2d, \frac{1}{2})$. If $I_\beta \neq \emptyset$ then the probability of success of the ISD-score decoder is at least

$$\left(1-\frac{e}{2\pi}\frac{1}{\sqrt{\ln\frac{n-t}{n-k-t+\delta+1}}}\right)\left(1-\frac{e}{\sqrt{2}\pi}\frac{1}{\sqrt{\ln\frac{t}{\delta+1}}}\right).$$

The technical details of our proofs also provide theoretical and numerical evidence of the gain compared to [25]. In particular, for all the cryptographic parameters of BIKE and *Classic McEliece*, our analysis shows theoretical evidence of high success probability while, when using the results from [25], some parameters are outside of this scenario.

Information theoretic bounds Next, we demonstrate that our algorithm can retrieve solutions of weight $t \leq O\left(\frac{n-k}{\ln(n-k)}\right)$, where n is the length of the code and k the dimension. We also analyze the noise level tolerated by the $\mathbb{N}-\mathsf{SDP}$. We prove that the ISD-score decoder can tolerate noise levels that are linear in the weight of the solution t.

Another consequence of our approach is that when the noise is null and the ISD part is ignored, equivalently the ISD-score decoder boils down to the algorithm proposed in [25], the conditions we propose on the range of parameters, namely on t, are larger than those from [25]. In addition the techniques used in our demonstration allowed us to obtain sharper lower bound on the number of syndrome entries, or the number of rows in the parity-check matrix, required to find a solution, known as the information theoretic bound.

Simulations We have demonstrated the performance of the $\mathbb{N} - \mathsf{SDP}$ in the presence of noise for different cryptographic parameters. To be more exact, we have chosen two code-based candidates at the NIST standardization process, *Classic McEliece* and BIKE. For both candidates we have considered increasing noise levels from $\mathcal{B}(\frac{t}{4}, \frac{1}{2})$ to $\mathcal{B}(t, \frac{1}{2})$. For *Classic McEliece* the parameter t exceeds a bit the maximum theoretic limit ($t = n/\log_2(n)$). Still, in simulations the ISD-score decoder finds the solution to the $\mathbb{N} - \mathsf{SDP}$ with noise using the optimization parameter $\delta = 3$ and all the syndrome entries. In the case of BIKE, where $t = \mathcal{O}(\sqrt{n})$ with $\delta \leq 3$ only a fraction of 0.15 syndrome entries are sufficient for the ISD-score decoder to find the solution.

On another track, we compare the ISD-score decoder with the ILP solution proposed in [15, 20]. In the noiseless setting there is a significant gap between the two solutions in terms of efficiency, and here we refer to timing and ratio of syndrome required. On both aspects our algorithm outperforms the ILP solutions (interior points or simplex). In the noisy setting the difference is even more significant, the ILP fails to find a solution even for the smallest noise considered. While our algorithm benefits from the advantage of the ISD part, when $\delta \geq 3$ it can be modified to continue with a generic ISD, the ILP on the other side does not posses such a feature.

We insist on the cryptographic context since it represents the origin of the underlying problem. The algorithm presented here was applied on practical instances. Although the public code of both BIKE and *Classic McEliece* are not random codes, they are indistinguishable from random codes. Also, here we do not necessarily insist on the fact that these instances are or not distinguishable, but on the fact that the public matrix entries follow a Bernoulli distribution. There are many constructions where the public codes can be distinguished from random codes, however, the their parity-check matrix is statistically "close" to a Bernoulli matrix (*e.g.* Niederreiter variant based on Reed-Muller codes [49], or on polar codes [48]). To support our models we have tested two hypothesis: i) the public parity-check matrix of BIKE and *Classic Mceliece* is distributed as a Bernouilli matrix, ii) its entries are independent; using a statistical test. We have noticed no significant difference, our hypothesis being validated with a p-value greater than or equal to 0.999 for all cryptographic parameter sets.

Summary of results compared to other models Let us briefly state how our results compare to other models such as those from Table 1.

- **Integer and noisy integer syndrome:** the noisy scenario is much more realistic than the perfect integer syndrome entries as shown in several articles [16, 29]. Hence compared

Table 1 Attacker model and algorithms for variants of SDP

Extra information	Algorithms	Performance	Noise	Refs.
Integer syndrome	ILP	Simulation	–	[15, 20]
	ISD-score decoder	Simulation	–	[16]
Noisy Integer syndrome	ISD-score decoder	Simulation	Double Cancellation	[16]
		Theoretical + Simulation	Side-channel Double Cancellation	This article
Integer syndrome and intermediate values	ISD-score decoder T-test Decoder	Simulation Simulation	Side-channel Double Cancellation	[29]
Parts of the solution	ISD with hints	Theoretical + Simulation	–	[33]

Performance refers to success probability and complexity of the algorithms to [15, 16, 20] our results feature a more practical applicability. Also, the noise model analyzed here comes from more realistic scenarios compared to other works such as [16]. Indeed, by considering the side-channel leakage model in addition to the double cancellation, we converge towards a real life scenario. In the noiseless scenario, our method outperforms both the Quantitative Group Testing analysis [25] and the ILP decoder [15, 20].

- **Parts of solution leakage model:** This model is not yet realistic and the assumption of having the exact value of some entries in the solution seems for the moment a bit unrealistic. Even-though the algorithm proposed here [33] are exponential in the Hamming weight of the solution and do not tolerate any noise. Let us emphasize that our method has the feature to incorporate this leakage model as well. Indeed, we can easily reduce the dimension of our problem by simply updating the syndrome with the solution entries equal to 1 and puncture the parity-check matrix on the positions of the solutions equal to 0.
- **Intermediate values:** This particular scenario requires one not only to obtain the noisy integer entries but also to have access to additional information such as the intermediate values involved in the computations. The T-test score decoder from [29] has better performances in simulation compared to our method, however, no theoretical evidence was presented.

Short note A 5 page short version of this article was presented at the Information Theory Workshop (ITW) 2022 [21]. We are extending on this short version by providing the following contributions.

- We give full proofs of the results in [21] with additional comments. We extend the results by providing sharper statement, quantifying exactly the error probability, e.g., Corollary 2 and Theorem 2 from [21]. On top of that the technical details from the proofs reflect the gain of our method compared to [25].
- We provide a detailed numerical simulated evidence of the approximated real noise model. We do illustrate that our theoretical model of noise provides a good approximation of the real noise model for the cryptographic parameters considered in the NIST standardization process.
- Compared to the conference version [21], here, we compare other similar methods for solving the $\mathbb{N}-\mathsf{SDP}$ with the ISD-score decoder. Indeed, we complete the analysis by including the ILP sovers in the comparison, from both a success rate point of view as well as computation time point of view.

Outline of the article

In Section 2, we introduce the SDP and its variants, $\mathbb{N}-\mathsf{SDP}$ and $\mathbb{N}-\mathsf{SDP}$ in the presence of noise. We also recall the cryptographic context where these problems occur. Section 3 begins by recalling the score decoder proposed in [16]. Then, it analyzes the distribution of the discriminant function for the $\mathbb{N}-\mathsf{SDP}$ in the presence of noise. The section ends with the description of the ISD-score decoder. Next, we analyze the success probability of the

ISD-score decoder in Section 4. The theoretical results from this part are being compared with numerical values from our implementation of the algorithm in Section 5. The section also makes a parallel between the efficiency of the ISD-score decoder and other methods such as ILP. Finally, we conclude the article in Section 6.

2 Preliminaries

2.1 Definitions and notations

Let us begin by fixing the necessary notations. A finite field is denoted by $\mathbb{F}$, and the ring of integers by $\mathbb{Z}$. The basis of the natural logarithm is denoted as by e, and ln denotes natural logarithm. We write $\mathbb{N}_n^* = \{1, \dots, n\}$ and $\mathbb{Z}_{-n,n} = \{-n, \dots, 0, \dots, n\}$. Matrices and vectors are written in bold capital, respectively small letters. We also use $\mathrm{HW}(\boldsymbol{c})$ to denote the Hamming weight of the vector $\boldsymbol{c}$, i.e., the number of non-zero positions of $\boldsymbol{c}$.

For $p \in [0, 1]$ and $n \in \mathbb{N}^*$ a random variable X that follows a distribution depending on p and n will be marked as $X \sim \mathcal{D}$, in particular, $X \sim \mathcal{B}er(p)$ for the Bernoulli distribution and by $X \sim \mathcal{B}(n, p)$ for the Binomial distribution.

We denote by $W(x)$ the Lambert W function [32], which is the converse of the function $x = ye^y$. In other words for any positive real number x the solution to the previous equation is $y = W(x)$ (to be more precise we only use the first real branch of W which is usually denoted $W_0(x)$ [17]). We will also require the asymptotic expansion near $x = \infty$ which is

$$W(x) = \ln x - \ln \ln x + O\left(\frac{\ln \ln x}{\ln x}\right).$$

Error correcting codes Let n and k be two positive integers such that $k \leq n$. An $[n, k]$ linear code can be defined as a sub-vector space of dimension k of the vector space $\mathbb{F}^n$. A code can be specified either by its generator matrix $\boldsymbol{G} \in \mathbb{F}^{k \times n}$ (a basis for the code), or by its parity-check matrix $\boldsymbol{H} \in \mathbb{F}^{(n-k)\times n}$ (a basis for the dual code). A code $\mathcal{C}$ is in standard form if its generator matrix is $\boldsymbol{G} = (\boldsymbol{I}_k \mid \boldsymbol{T})$. The minimum distance, or the Hamming distance of a code $\mathcal{C}$, is the minimum of all $\mathrm{HW}(v)$ for $\boldsymbol{v} \in \mathcal{C}$, $\boldsymbol{v} \neq \boldsymbol{0}$.

One of the main features of linear codes is their ability to decode noisy information/data. Several general decoding strategies exist, the syndrome decoding problem being one of them.

Definition 1 (Binary syndrome decoding problem SDP)

Inputs: $\boldsymbol{H} \in \mathbb{F}_2^{(n-k)\times n}$, $\boldsymbol{s}^* \in \mathbb{F}_2^{n-k}$, $t \in \mathbb{N}^*$.
Output: $\boldsymbol{x} \in \mathbb{F}_2^n$ such that $\boldsymbol{Hx} = \boldsymbol{s}^*$, and $\mathrm{HW}(\boldsymbol{x}) = t$.

This problem is NP-Complete [6] and, as we shall quickly see, it constitutes the building block of code-based solution for post-quantum cryptography.

2.2 The Niederreiter encryption framework

Both, *Classic McEliece* [2] and BIKE [3], are based on the Niederreiter encryption scheme [42]. The key generation, encryption and decryption functions of the Niederreiter cryptosystem are given in Algorithms 1, 2 and 3 respectively.

Algorithm 1 Niederreiter key generation

```
function KEYGEN($n$, $k$, $t$)
    $\mathcal{C}$ an $[n, k]$ code that corrects $t$ errors
    A parity-check matrix of $\mathcal{C}$: $\boldsymbol{H}$
    An $n \times n$ permutation matrix $\boldsymbol{P}$
    An $(n - k) \times (n - k)$ invertible matrix $\boldsymbol{S}$
    Compute $\boldsymbol{H}_{\text{pub}} = \boldsymbol{SHP}$
    pk $= (\boldsymbol{H}_{\text{pub}}, t)$
    sk $= (\boldsymbol{S}, \boldsymbol{H}, \boldsymbol{P})$
    return (pk, sk)
```

Algorithm 2 Niederreiter encryption

```
function ENCRYPT($\boldsymbol{m}$, pk)
    Encode $\boldsymbol{m} \rightarrow \boldsymbol{x}$ with HW() $= t$
    Compute $\boldsymbol{s}^* = \boldsymbol{H}_{\text{pub}}\boldsymbol{x}$
    return $\boldsymbol{s}^*$
```

To practically instantiate the schemes one must choose a family of error correcting codes, e.g., binary Goppa codes for *Classic McEliece* that posses strong security arguments. One of the required feature is to be indistinguishable from random codes, which is the case of all submitted proposals. Such a requirement has a theoretical implication (semantic security arguments) and a practical implication, we can set parameters as if we were dealing with random codes. In this is the case then breaking the *confidentiality* of the Niederreiter-like schemes resumes to solving the SDP for a random-like code. Hence, the sets of (n, k, t) parameters defined in [2] and [3] (see Table 2) are given with respect to the working factor of the best algorithm for solving the SDP. There are two different type of algorithms for solving SDP, statistical decoding [18, 34, 41, 44] and Information Set Decoding (ISD) [7, 8, 11, 22, 36, 37, 39, 40, 45, 50].

Algorithm 3 Niederreiter decryption

```
function DECRYPT($\boldsymbol{s}^*$, sk)
    Compute $\boldsymbol{x}^{'} =$ Decode$(\boldsymbol{S}^{-1}\boldsymbol{s}^*, \boldsymbol{H})$
    Compute $\boldsymbol{m}$ from $\boldsymbol{P}^{-1}\boldsymbol{x}^{'}$
    return $\boldsymbol{m}$
```

Table 2 (n, k, t) parameters for *Classic McEliece* and BIKE

	n	k	t
Classic McEliece	3488	2720	64
	4608	3360	96
	6688	5024	128
	8192	6528	128
BIKE	24646	12323	134
	49318	24659	199
	81946	40973	264

Let us shortly recall the ideas behind the ISD techniques, e.g., the Prange variant.

1. Randomly permute the columns of $\boldsymbol{H}$ (let $\boldsymbol{P}$ be the permutation matrix)
2. Compute the standard form of $\boldsymbol{H}^* = \boldsymbol{H}\boldsymbol{P}$, i.e.,

$$\boldsymbol{Q}\boldsymbol{H}^* = \boldsymbol{Q}\boldsymbol{H}\boldsymbol{P} = \left(\boldsymbol{T}\ \boldsymbol{I}_{n-k}\right) \tag{1}$$

3. If $\mathrm{HW}(\boldsymbol{Q}\boldsymbol{s}^*) \leq t$ then return $\boldsymbol{P}\begin{pmatrix}\boldsymbol{0}_k \\ \boldsymbol{Q}\boldsymbol{s}^*\end{pmatrix}$;
   else go to Step 1

Since $\boldsymbol{H}\boldsymbol{x} = \boldsymbol{s}$ we can see that

$$\boldsymbol{Q}\boldsymbol{H}\boldsymbol{P}\,\underbrace{\boldsymbol{P}^{-1}\boldsymbol{x}}_{\boldsymbol{x}^*} = \left(\boldsymbol{T}\ \boldsymbol{I}_{n-k}\right)\begin{pmatrix}\boldsymbol{x}_1^* \\ \boldsymbol{x}_2^*\end{pmatrix} = \boldsymbol{Q}\boldsymbol{s}^*,$$

which yields

$$\boldsymbol{T}\boldsymbol{x}_1^* + \boldsymbol{x}_2^* = \boldsymbol{Q}\boldsymbol{s}^*.$$

Now, if $\boldsymbol{x}_1^* = \boldsymbol{0}$ we deduce $\boldsymbol{x}_2^* = \boldsymbol{Q}\boldsymbol{s}^*$ and thus $\boldsymbol{x}^* = \begin{pmatrix}\boldsymbol{0} \\ \boldsymbol{Q}\boldsymbol{s}^*\end{pmatrix}$ is a valid solution to the SDP. Prange's algorithm samples permutations until the vector $\boldsymbol{x}_1^*$ equals zero, or equivalently until an information set is found. Variants of ISD offer time optimizations by allowing different relaxations of the weight condition on $\boldsymbol{x}_1^*$.

2.3 Integer version of the syndrome decoding problem

Recent message recovery attacks are pointing the encryption step, where the cipher-text is obtained from the multiplication of the public parity-check matrix $\boldsymbol{H}_{\mathrm{pub}}$ and the secret error vector $\boldsymbol{x}$. Hence, in [15, 16, 29] the matrix-vector multiplication is targeted as leakage point (line 3 in Algorithm 2). The physical scenario reveals the possibility of retrieving extra information during the multiplications. More exactly, it was shown that it is possible to either change the instruction code in the Flash memory and thus set it to ADD instead of XOR [15] or to recover by side-channel measurements an approximation of the real/natural value of $\boldsymbol{s}^*$ [16]. Both scenarios lead to a modified version of the binary SDP.

Definition 2 ($\mathbb{N} - \mathsf{SDP}$)

> **Inputs:** $\boldsymbol{H} \in \{0,1\}^{(n-k)\times n}$, $\boldsymbol{s} \in \mathbb{N}^{n-k}$, $t \in \mathbb{N}^*$.
> **Output:** $\boldsymbol{x} \in \{0,1\}^n$, such that $\boldsymbol{H}\boldsymbol{x} = \boldsymbol{s}$, and $\mathrm{HW}(\boldsymbol{x}) = t$.

As pointed out in [29] the integer value $\boldsymbol{s}$ is often difficult to obtain. More exactly, in [29] in was shown that there are different type of noise that interfere with the intermediate estimations, fact that leads to a noisy integer syndrome. To define $\mathbb{N} - \mathsf{SDP}$ in the presence of noise as generally as possible, we model the noise $\epsilon = (\epsilon_1, \ldots, \epsilon_{n-k})$ as a vector of random variables $\epsilon_i \sim \mathcal{D}$, where $\mathcal{D}$ is a discrete probability distribution. In the $\mathbb{N} - \mathsf{SDP}$ in the presence of noise, instead of having access to an instance of the $\mathbb{N} - \mathsf{SDP}$, *i.e.*, $(\boldsymbol{H}, \boldsymbol{s}, t)$, we are given a noisy syndrome $\widetilde{\boldsymbol{s}} = \boldsymbol{s} + \epsilon$ and the value $\boldsymbol{s}^* = \boldsymbol{s} \pmod 2$ (component-wise).

Definition 3 $\mathbb{N} - \mathsf{SDP}$ in the presence of noise ϵ)

> **Inputs:** $\boldsymbol{H} \in \{0,1\}^{(n-k)\times n}$, $\widetilde{\boldsymbol{s}} \in \mathbb{Z}^{n-k}$
> $\boldsymbol{s}^* \in \{0,1\}^{n-k}$, $t \in \mathbb{N}^*$

Output: $\boldsymbol{x} \in \{0,1\}^n$, such that $\boldsymbol{Hx} = \boldsymbol{s}^*$ with $\mathrm{HW}(\boldsymbol{x}) = t$
$\boldsymbol{s}^* = \boldsymbol{s} \mod 2$, and $\widetilde{\boldsymbol{s}} = \boldsymbol{s} + \epsilon$.

Remark that $\mathbb{N}-$ SDP in presence of noise is the SDP with additional information. Under certain conditions, we hope that, given $(\boldsymbol{H}, \boldsymbol{s}^*, t, \widetilde{\boldsymbol{s}})$, we can find $\boldsymbol{x}$, solution to the SDP. Also, when the noise is zero we face the classic $\mathbb{N}-$ SDP.

3 ISD-score decoder

The idea of assigning a score to each column was already used for the $\mathbb{N}-$ SDP in [16]. The objective is to distinguish columns of $\boldsymbol{H}$ in the support of the solution vector from columns which are outside the support. We shall begin by defining a score decoder, as introduced in [25], that proved to be particularly discriminant in the context of $\mathbb{N}-$ SDP. For a better illustration of the nice features of the decoder in the presence of noise, we will express it in function of the noiseless decoder. As we shall see, this method allows not only to derive a particularly simple relation between those two, but also to deduce conditions on the tolerated noise level.

Definition 4 Let $\boldsymbol{H} \in \{0,1\}^{(n-k)\times n}$, $\boldsymbol{s} \in \mathbb{N}^{n-k}$ and $t \in \mathbb{Z}^*$ be the input of $\mathbb{N}-$ SDP. Then define the score of a column:

$$\forall i \in \mathbb{N}_n^* \quad \psi_i(\boldsymbol{s}) = \sum_{\ell=1}^{n-k} \left(h_{\ell,i} \boldsymbol{s}_\ell + (1 - h_{\ell,i})(t - \boldsymbol{s}_\ell) \right). \tag{2}$$

For the $\mathbb{N}-$ SDP in the presence of noise we shall use $\psi_i(\widetilde{\boldsymbol{s}})$. The next result, rephrased from [25], expresses the capability of the score decoder to distinguish between columns in the support of the solution vector from columns which are outside the support.

Theorem 1 *Let $\boldsymbol{H} \in \{0,1\}^{(n-k)\times n}$ be a random matrix, with distribution given by $h_{j,i} \sim \mathcal{B}er(\frac{1}{2})$ and $\boldsymbol{s} \in \mathbb{N}^{n-k}$ such that $\exists\ \boldsymbol{x} \in \{0,1\}^n$ with $HW(\boldsymbol{x}) = t$ satisfying $\boldsymbol{Hx} = \boldsymbol{s}$. Then*

$$\psi_i(\boldsymbol{s}) \sim \begin{cases} \mathcal{B}((n-k)t, \frac{1}{2}) & , i \notin \mathsf{Supp}(\boldsymbol{x}) \\ \mathcal{B}((n-k)(t-1), \frac{1}{2}) + n - k & , i \in \mathsf{Supp}(\boldsymbol{x}) \end{cases}$$

Straightforward from Theorem 1 we have $\mathbb{E}(\psi_i(\boldsymbol{s})) = (n-k)t/2$ for $i \notin \mathsf{Supp}(\boldsymbol{x})$ and $\mathbb{E}(\psi_i(\boldsymbol{s})) = (n-k)t/2 + (n-k)/2$ for $i \in \mathsf{Supp}(\boldsymbol{x})$.

The difference in the average value points out that ψ can be a distinguisher between positions in the support and outside the support of the vector $\boldsymbol{x}$. In addition, the variance also differs, fact that will be used in the tail bounds. Moving forward, we will consider the noisy version of this problem in the next section.

3.1 Score decoder in the presence of noise

As in [16], we make some assumptions on the noise considered here, *i.e.*, ϵ_i are independent and identically distributed random variables, the noise does not depend on the distribution of the entries in $\boldsymbol{H}$ and the distribution $\mathcal{D}$ is symmetric.

Proposition 1 ([16]) *For $j \in \mathbb{Z}_{n-k}^*$ let ϵ_j be independent and identically distributed discrete random variables following a symmetric distribution over the set $\mathbb{Z}_{-d,d}$, such that ϵ_j and $h_{i,j}$ are independent.*

Then

$$\mathsf{Prob}\,(\psi_i(\widetilde{\boldsymbol{s}}) - \psi_i(\boldsymbol{s}) = \alpha) = \mathsf{Prob}\left(\sum_{j=1}^{n-k} \epsilon_j = \alpha\right).$$

Proof Let $Y_{\ell,i} = (2h_{\ell,i} - 1)\epsilon_\ell$. Then we have,

$$\begin{aligned}\psi_i(\widetilde{\boldsymbol{s}}) &= \sum_{\ell=1}^{n-k} \big(h_{\ell,i}(\widetilde{s_\ell} + (1 - h_{\ell,i})(t - \widetilde{s_\ell}))\big)\\ &= \sum_{\ell=1}^{n-k} \big(h_{\ell,i}(s_\ell + \epsilon_\ell + (1 - h_{\ell,i})(t - s_\ell - \epsilon_\ell))\big)\\ \psi_i(\widetilde{\boldsymbol{s}}) &= \psi_i(\boldsymbol{s}) + \sum_{\ell=1}^{n-k} \underbrace{\big(h_{\ell,i}\epsilon_\ell - (1 - h_{\ell,i})\epsilon_\ell\big)}_{Y_{l,i}}\end{aligned}$$

For any fixed value of $\ell \in \mathbb{Z}^*_{n-k}$ we have $\mathsf{Prob}(Y_{\ell,i} = \alpha_\ell) = \mathsf{Prob}(\epsilon_\ell = \alpha_\ell)$ for any $\alpha_\ell \in \mathbb{Z}_{-d,d}$ (using the symmetry property and the independence of $h_{\ell,i}$ and ϵ_ℓ). Hence $Y_{\ell,i}$ follows the same distribution as ϵ_ℓ. Thus, $\psi_i(\widetilde{\boldsymbol{s}}) - \psi_i(\boldsymbol{s}) \in \mathbb{Z}_{-(n-k)d,(n-k)d}$ with probability distribution $\mathsf{Prob}(\psi_i(\widetilde{\boldsymbol{s}}) - \psi_i(\boldsymbol{s}) = \alpha) = \mathsf{Prob}\left(\sum_{j=1}^{n-k} \epsilon_j = \alpha\right)$.

Keeping the difference $\psi_i(\widetilde{\boldsymbol{s}}) - \psi_i(\boldsymbol{s})$ as small as possible resumes to controlling the sum of ϵ_j. The variance of ϵ_j plays a crucial role in the distinguishing capacity of ψ.

Proposition 2 *For any $j \in \mathbb{Z}^*_{n-k}$ let ϵ_j be a discrete random variable satisfying the conditions from Proposition 1 and let $\sigma^2 = Var(\epsilon_j)$. Let $g(n, k, t)$ be a function in the parameters of* $\mathbb{N} - \mathsf{SDP}$. *Then for any $\alpha > \sigma\sqrt{(n-k)g(n, k, t)}$*

$$\mathsf{Prob}(\psi_i(\widetilde{\boldsymbol{s}}) - \psi_i(\boldsymbol{s}) \geq \alpha) \leq \frac{1}{g(n, k, t)}. \tag{3}$$

Proof Use Chebyshev's inequality for the sum of ϵ_j and the linearity of the variance.

The case of centered binomial noise

Corollary 1 *Let $d \in \mathbb{N}$ and $\epsilon_i \sim -d + \mathcal{B}(2d, \frac{1}{2})$. Then*

- *for $i \notin$* $\mathsf{Supp}(\boldsymbol{x})$

$$\psi_i(\widetilde{\boldsymbol{s}}) \sim -d(n-k) + \mathcal{B}\left((n-k)(t + 2d), \frac{1}{2}\right);$$

- *for $i \in$* $\mathsf{Supp}(\boldsymbol{x})$

$$\psi_i(\widetilde{\boldsymbol{s}}) \sim -(d-1)(n-k) + \mathcal{B}\left((n-k)(t - 1 + 2d), \frac{1}{2}\right).$$

Moreover, $\mathbb{E}(\psi_i(\widetilde{\boldsymbol{s}})) = \mathbb{E}(\psi_i(\boldsymbol{s}))$ and $Var(\psi_i(\widetilde{\boldsymbol{s}})) = Var(\psi_i(\boldsymbol{s})) + (n-k)d/2$.

To maintain the capability to distinguish between positions inside the support and positions outside the support, the noise parameter d from $\mathcal{B}(2d, \frac{1}{2})$ should be restricted.

Corollary 2 *Let $\epsilon_i \sim -d + \mathcal{B}(2d, \frac{1}{2})$ and $g(n, k, t)$ an unbounded function in t, n, k. Then we have*

$$\mathsf{Prob}\left(|\psi_i(\widetilde{\boldsymbol{s}}) - \psi_i(\boldsymbol{s})| \leq \sqrt{\frac{d(n-k)g(n,k,t)}{2}}\right) \geq 1 - \frac{2}{g(n,t,k)}.$$

Moreover, for any $d \leq \frac{n-k}{8g(n,k,t)}$, the function $\psi(\widetilde{\boldsymbol{s}})$ distinguishes positions in $\mathsf{Supp}(\boldsymbol{x})$ *from positions outside* $\mathsf{Supp}(\boldsymbol{x})$.

The key idea of the distinguisher is that for $i \notin \mathsf{Supp}(\boldsymbol{x})$ the upper limit of the confidence interval will be smaller that the lower limit of the confidence interval for $i \in \mathsf{Supp}(\boldsymbol{x})$. More exactly, to distinguish with probability at least $1 - \frac{1}{g(n,t,k)}$ one needs to have

$$\frac{(n-k)t}{2} + \sqrt{\frac{d(n-k)g(n,k,t)}{2}} \leq \frac{(n-k)t}{2} + \frac{n-k}{2} - \sqrt{\frac{d(n-k)g(n,k,t)}{2}}, \tag{4}$$

which yields $d \leq \frac{n-k}{8g(n,k,t)}$. In particular, we can put $g(n, k, t) = \ln\ln t$ or $g(n, k, t) = \ln\ln n$ depending on the wanted speed of convergence. Figure 1 shows the distribution of ψ_i values for different levels of noise, ranging from $d = 0$, *i.e.* the noiseless setting, to a very high noise of $\mathcal{B}(2t, \frac{1}{2})$. Notice that the distinguishing capability is much higher for the BIKE parameters, as shown in Fig. 1a, than for the *Classic McEliece* parameters, as shown in Fig. 1b.
Bernoulli noise

Proposition 3 *Let $\epsilon_i \sim \mathcal{B}er(\{0, 1\}, 1/2)$. Then $\psi_i(\widetilde{\boldsymbol{s}})$ is a random variable that follows the distribution*

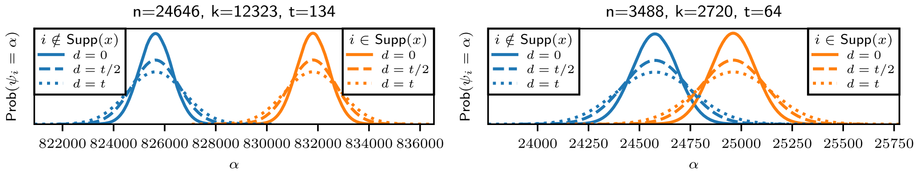
$$\psi_i(\widetilde{\boldsymbol{s}}) \sim \begin{cases} \mathcal{B}((n-k)(t+2), \frac{1}{2}) - (n-k) & , i \notin \mathsf{Supp}(\boldsymbol{x}) \\ \mathcal{B}((n-k)(t+1), \frac{1}{2}) & , i \in \mathsf{Supp}(\boldsymbol{x}) \end{cases}$$

Moreover, $\mathbb{E}(\psi_i(\widetilde{\boldsymbol{s}})) = \mathbb{E}(\psi_i(\boldsymbol{s}))$ *and* $Var(\psi_i(\widetilde{\boldsymbol{s}})) = Var(\psi_i(\boldsymbol{s})) + (n-k)/2$.

Notice that, in the case of a Bernoulli type of noise, the behavior is equivalent to the case of a centered binomial noise. (equivalent to $d = 1$ in Corollary 1). Indeed, the result in Proposition 3 is equivalent to the one given in Corollary 1 with $d = 1$.

3.2 Combining ISD and score decoder

The idea in [16] was to boost the distinguishing capability of the score decoder with ISD-like techniques. To this end, the score decoder is integrated in the "permutation" step of the ISD method. Indeed, this method starts by performing a permutation on the columns of $\boldsymbol{H}$ that will hopefully rearrange the solution in a useful way. In the original ISD methods, permutations are sampled randomly until a "good" one is obtained. Thanks to the extra-information provided



(a) Example parameters set of BIKE

(b) Example parameters set of *Classic McEliece*

Fig. 1 Distribution of ψ_i for $\epsilon \sim -d + \mathcal{B}(2d, \frac{1}{2})$

by $\boldsymbol{s}$ or $\widetilde{\boldsymbol{s}}$, the function ψ allows to construct a permutation which by no means is random. Indeed, we have seen that ψ, by its nature, allows one to distinguish between positions in the support of $\boldsymbol{x}$ and positions outside. Hence, the underlying permutation, hopefully is a "good" permutation. As pointed out in [16], sorting the list of values $\psi_i(\widetilde{\boldsymbol{s}})$ in descending order is equivalent to generating a permutation $\boldsymbol{\Pi}$. Algorithm 4 finds a solution to the $\mathbb{N}-\mathsf{SDP}$ in the presence of noise as long as $\boldsymbol{\Pi}$ is "good" enough.

Algorithm 4 Prange Score Decoder

```
function PRANGE SCORE DECODER($\boldsymbol{H}, \widetilde{\boldsymbol{s}}, \boldsymbol{s}^*, t$)
    Compute $\boldsymbol{\Pi}$ from the list $\psi_i(\widetilde{\boldsymbol{s}})$
    Compute $\boldsymbol{A}^*, \boldsymbol{H}^* \leftarrow \texttt{rref}(\boldsymbol{H}\boldsymbol{\Pi})$
    if HW$(A^*s^*) = t$ then
        return $\boldsymbol{x} = \boldsymbol{\Pi}\begin{pmatrix}\boldsymbol{A}^*\boldsymbol{s}^*\\ \boldsymbol{0}_{n-r}\end{pmatrix}$        ▷ $r = \texttt{rank}(\boldsymbol{A})$
```

The procedure $\texttt{rref}(\boldsymbol{H}\boldsymbol{\Pi})$, which stands for "reduced row echelon form", is equivalent to performing a partial Gaussian elimination over $\mathbb{F}_2$. Indeed, there is an $(n-k)\times(n-k)$ non-singular matrix $\boldsymbol{A}^*$ such that, $\boldsymbol{A}^*\boldsymbol{H}\boldsymbol{\Pi} = \left[\begin{bmatrix}\boldsymbol{I}_r\\ \boldsymbol{0}_{n-k-r,r}\end{bmatrix} \parallel \boldsymbol{B}^*\right]$ where $\boldsymbol{H}\boldsymbol{\Pi} = [\boldsymbol{A} \parallel \boldsymbol{B}]$ with $\boldsymbol{A}$ a $(n-k)\times r$ matrix satisfying $\boldsymbol{A}^*\boldsymbol{A} = \begin{bmatrix}\boldsymbol{I}_r\\ \boldsymbol{0}_{n-k-r,r}\end{bmatrix}$, and $\boldsymbol{B}^* = \boldsymbol{A}^*\boldsymbol{B}$.

In the case of a full rank matrix $\boldsymbol{A}$ we have $\boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{I}_{n-k}$. From the description of the algorithm above, the following result can be deduced.

Proposition 4 ([16]) PRANGE SCORE DECODER *outputs a valid solution as long as there exists at least one set* $L \subset \mathbb{N}_n^* \setminus \mathsf{Supp}(\boldsymbol{x})$ *with* $\#L \geq n-r$ *such that* $\min\{\psi_i(\widetilde{\boldsymbol{s}}), i \in \mathsf{Supp}(\boldsymbol{x})\} > \max\{\psi_i(\widetilde{\boldsymbol{x}}), i \in L\}$.

The overall time complexity of PRANGE SCORE DECODER is $\mathcal{O}((n-k)^3)$, since it is dominated by the partial Gaussian elimination, *i.e.* the computation of $\boldsymbol{A}^*$.

Since the permutation $\boldsymbol{\Pi}$ might not move all the positions in the support of $\boldsymbol{x}$ in the first $n-k$ positions, more powerful ISD methods may be used, *e.g.* Lee-Brickell [36], Stern [50] or Dumer [22]. The idea is to allow a number of δ positions from $\mathsf{Supp}(\boldsymbol{x})$ outside the first $n-k$ positions. This is equivalent to extending PRANGE SCORE DECODER so that it covers error vectors with a more general pattern. The LEE- BRICKELL SCORE DECODER, where δ positions are searched exhaustively, is thus proposed in [16] as a possible solution.

When the Lee-Brickell variant is used and $\delta = \mathcal{O}(1), k = \mathcal{O}(n)$, the work factor of the resulting algorithm becomes polynomial in n.

Proposition 5 *The* δ*-ISD-score decoder outputs a valid solution as long as there are at most* δ *indices* $i \in \mathsf{Supp}(\boldsymbol{x})$ *with values* $\psi_i(\widetilde{\boldsymbol{s}}) < \psi_j(\widetilde{\boldsymbol{s}})$ *with* j *in a set* $J \subset \mathbb{N}_n$ *of cardinality* $n-k$.

4 Success probability of the ISD-score decoder

The following result gives a condition on the parameters for having a high probability of success for the ISD score decoder on the $\mathbb{N}-\mathsf{SDP}$ in presence of noise.

Algorithm 5 Lee-Brickell Score Decoder ([16])

```
function LEE- BRICKELL SCORE DECODER($\boldsymbol{H}, \widetilde{\boldsymbol{s}}, \boldsymbol{s}^*, t$)
    Compute $\boldsymbol{\Pi}$ from the list $\psi_i(\widetilde{\boldsymbol{s}})$
    Set $\boldsymbol{H\Pi} = [\boldsymbol{A} \parallel \boldsymbol{B}]$
    Compute $\boldsymbol{A}^*, \boldsymbol{H}^* \leftarrow$rref$(\boldsymbol{H\Pi})$ and $\boldsymbol{B}^* = \boldsymbol{A}^*\boldsymbol{B}$
    Compute $\boldsymbol{s}^{'} = \boldsymbol{A}^*\boldsymbol{s}^*$
    if HW$(\boldsymbol{s}^{'}) == t$ then
        return $\boldsymbol{x} = \boldsymbol{\Pi}(\boldsymbol{s}^{'} \parallel \boldsymbol{0}_k)^t$
    else
        for $i \leftarrow 1, \delta$ do
            $S$ = Gener-Subsets$(\{1, \ldots, k\}, i)$
            for $E$ in $S$ do
                $\boldsymbol{x}^{''} \leftarrow$ Vector$(\{0, 1\}, k, E)$
                $\boldsymbol{x}^{'} \leftarrow \boldsymbol{s}^{'} - \boldsymbol{B}^*\boldsymbol{x}^{''}$
                if HW$(\boldsymbol{x}^{'}) == t - i$ then
                    return $\left(\boldsymbol{\Pi}(\boldsymbol{x}^{'} \parallel \boldsymbol{x}^{''})^t, \boldsymbol{\Pi}\right)$
```

Theorem 2 *Let* $I = \left[\sqrt{\frac{t+2d}{n-k} W\left(\frac{n-t}{n-k-t+\delta+1}\frac{e\sqrt{2}}{\pi}\right)^2}, 1 - \sqrt{\frac{t+2d-1}{n-k} W\left(\frac{t}{\delta+1}\frac{2e}{\pi}\right)^2}\right]$ *and* $\epsilon_i \sim -d + \mathcal{B}(2d, \frac{1}{2})$. *If* I *is non-empty, then there is a value* $\beta \in I$ *such that the ISD-score decoder succeeds in finding a valid solution with probability at least*

$$\left(1 - \frac{e(n-t)}{\sqrt{2}\pi\beta(n-k-t+\delta+1)}\sqrt{\frac{t+2d}{n-k}}e^{-\frac{(n-k)\beta^2}{2(t+2d)}}\right)\left(1 - \frac{et}{\pi(1-\beta)(\delta+1)}\sqrt{\frac{t+2d-1}{n-k}}e^{-\frac{(n-k)(1-\beta)^2}{2(t+2d-1)}}\right).$$

4.1 Technicalities of theorem 2

To prove this theorem we shall use 3 steps. More precisely, we first give an estimation on the tails of the distributions $\psi_i(\widetilde{\boldsymbol{s}})$, then we insert these results into a generic upper bound on the probability of success of the ISD-score decoder, and finally we study the range of parameters for which our conditions are valid.

4.1.1 Tail bounds on the distribution

Firstly we have the following result on the distribution of ψ in the noiseless scenario.

Theorem 3 *Let* $\beta \in (0, 1)$ *and* $B_\beta = \frac{(n-k)t}{2} + \frac{\beta(n-k)}{2}$. *Then we have for* $i \notin \mathsf{Supp}(\boldsymbol{x})$

$$\mathsf{Prob}\left(\psi_i(\boldsymbol{s}) \geq B_\beta\right) \leq \frac{e}{\sqrt{2}\pi\beta}\sqrt{\frac{t}{n-k}}e^{-\frac{n-k}{2t}\beta^2}, \tag{5}$$

for $i \in \mathsf{Supp}(\boldsymbol{x})$

$$\mathsf{Prob}\left(\psi_i(\boldsymbol{s}) \leq B_\beta\right) \leq \frac{e}{\pi(1-\beta)}\sqrt{\frac{t-1}{n-k}}e^{-\frac{n-k}{2(t-1)}(1-\beta)^2}. \tag{6}$$

Moving forward, in the case of a binomial noise we have

Theorem 4 *Let* $\epsilon_i \sim -d + \mathcal{B}(2d, \frac{1}{2})$, $\beta \in (0, 1)$ *and* B_β *as previously defined. Then we have for* $i \notin \mathsf{Supp}(\boldsymbol{x})$

$$\mathsf{Prob}\left(\psi_i(\widetilde{\boldsymbol{s}}) \geq B_\beta\right) \leq \frac{e}{\sqrt{2}\pi\beta}\sqrt{\frac{t+2d}{n-k}}e^{-\frac{(n-k)\beta^2}{2(t+2d)}}, \tag{7}$$

for $i \in \mathsf{Supp}(\boldsymbol{x})$

$$\mathsf{Prob}\left(\psi_i(\widetilde{\boldsymbol{s}}) \leq B_\beta\right) \leq \frac{e}{\pi(1-\beta)}\sqrt{\frac{t+2d-1}{n-k}}e^{-\frac{(n-k)(1-\beta)^2}{2(t+2d-1)}}. \tag{8}$$

The proof of the two theorems above is given in the Appendix. Let us denote the two upper bounds in Theorem 4 by $\mathrm{Ub}_{\mathsf{Supp}(\boldsymbol{x})}(n,k,t,\beta)$ and $\mathrm{Ub}_{\mathsf{Supp}(\boldsymbol{x})^c}(n,k,t,\beta)$.

4.1.2 A general bound on the success probability using tail estimations

A general theorem regarding the success probability of ISD-score decoder can be stated. For that we suppose that the distribution $\psi_i(\widetilde{\boldsymbol{s}})$ when $i \in \mathsf{Supp}(\boldsymbol{x})$ has to be different from $\psi_i(\widetilde{\boldsymbol{s}})$ when $i \notin \mathsf{Supp}(\boldsymbol{x})$, e.g., it is at least shifted. If not it is obvious that ISD-score decoder can not retrieve a valid solution with high probability.

Theorem 5 *Let* $\psi_i(\widetilde{\boldsymbol{s}})$ *be random variables and* $f(n,k,t,d,B)$, $g(n,k,t,d,B)$ *be two functions such that*

$$\mathsf{Prob}(\psi_i(\widetilde{\boldsymbol{s}}) \leq B) \leq e^{-f(n,k,t,d,B)} \quad , i \in \mathsf{Supp}(\boldsymbol{x}) \tag{9}$$

$$\mathsf{Prob}(\psi_i(\widetilde{\boldsymbol{s}}) \geq B) \leq e^{-g(n,k,t,d,B)} \quad , i \notin \mathsf{Supp}(\boldsymbol{x}) \tag{10}$$

The ISD-score decoder finds the solution if $\exists B^*$ *such that*

- $0 \leq 1 - \frac{t}{\delta+1}e^{-f(n,k,t,d,B^*)} \leq 1,$
- $0 \leq 1 - \frac{n-t}{n-k-t+\delta+1}e^{-g(n,k,t,d,B^*)} \leq 1,$
- $\frac{t}{\delta+1}e^{-f(n,k,t,d,B^*)} + \frac{n-t}{n-k-t+\delta+1}e^{-g(n,k,t,d,B^*)}$ *is close to zero,*

Typically, the theorem gives a sufficient condition for having a high probability of success. Indeed, if one finds a value B_β for which the lower bound tends to 1 then the Score function achieves its goal, namely to distinguish positions in the support of $\boldsymbol{x}$ from those outside it. The proof of this result is given in the Appendix.

Combining the tail bounds on the distribution of $\psi_i(\widetilde{\boldsymbol{s}})$ with the condition on β^* for having a high probability of success enables the following result. Denote

$$\mathrm{Lb}_{\mathsf{Supp}(\boldsymbol{x})^c} = 1 - \frac{e(n-t)}{\sqrt{2}\pi\beta(n-k-t+\delta+1)}\sqrt{\frac{t+2d}{n-k}}e^{-\frac{(n-k)\beta^2}{2(t+2d)}},$$

$$\mathrm{Lb}_{\mathsf{Supp}(\boldsymbol{x})} = 1 - \frac{et}{\pi(1-\beta)(\delta+1)}\sqrt{\frac{t+2d-1}{n-k}}e^{-\frac{(n-k)(1-\beta)^2}{2(t+2d-1)}}.$$

Proposition 6 *Let* $\epsilon_i \sim -d + \mathcal{B}(2d, \frac{1}{2})$. *If* $\exists \beta^* \in (0,1)$ *such that* $\mathrm{Lb}_{\mathsf{Supp}(\boldsymbol{x})}, \mathrm{Lb}_{\mathsf{Supp}(\boldsymbol{x})^c} \in [0,1]$. *The probability that ISD-score decoder succeeds in finding a valid solution is at least* $\mathrm{Lb}_{\mathsf{Supp}(\boldsymbol{x})}\mathrm{Lb}_{\mathsf{Supp}(\boldsymbol{x})^c}$.

Corollary 3 *When* $d = 0$ *and* $\delta = 0$ *the condition on* β^* *simplifies to*

- $0 \leq \frac{et}{\pi(1-\beta)}\sqrt{\frac{t}{n-k}}e^{-\frac{(n-k)(1-\beta)^2}{2t}} \leq 1,$
- $0 \leq \frac{e(n-t)}{(\sqrt{2}\pi\beta)(n-k-t)}\sqrt{\frac{t}{n-k}}e^{-\frac{(n-k)\beta^2}{2t}} \leq 1,$
- $\frac{et}{\pi(1-\beta)}\sqrt{\frac{t}{n-k}}e^{-\frac{(n-k)(1-\beta)^2}{2t}} + \frac{e(n-t)}{(\sqrt{2}\pi\beta)(n-k-t)}\sqrt{\frac{t}{n-k}}e^{-\frac{(n-k)\beta^2}{2t}}$ *is close to zero,*

To fairly compare with state-of-the-art techniques such as the algorithm in [25], which is only valid for the noiseless scenario, we adapted the conditions from [25] to the noise model considered here. This gives two similar functions in β, namely $1 - \frac{n-t}{n-k-t} e^{-\frac{(n-k)\beta^2}{2(t+2d)}}$, and $1 - te^{-\frac{(n-k)(1-\beta)^2}{2(t+2d-1)}}$. In Fig. 2, we plot the modified functions from [25] (dashed lines) and $\mathrm{Lb}_{\mathsf{Supp}(\boldsymbol{x})}$, $\mathrm{Lb}_{\mathsf{Supp}(\boldsymbol{x})^c}$ (solid lines).

In dark green and light green, the valid interval/region for the adapted functions from [25], and our functions, respectively, are represented. Notice that for all parameter sets and all noise levels considered here, our function offers a larger interval. Hence, this implies that for some sets of parameters, *e.g.,* in Fig. 2d, the interval is empty w.r.t. conditions in [25], while w.r.t. our conditions the interval exists.

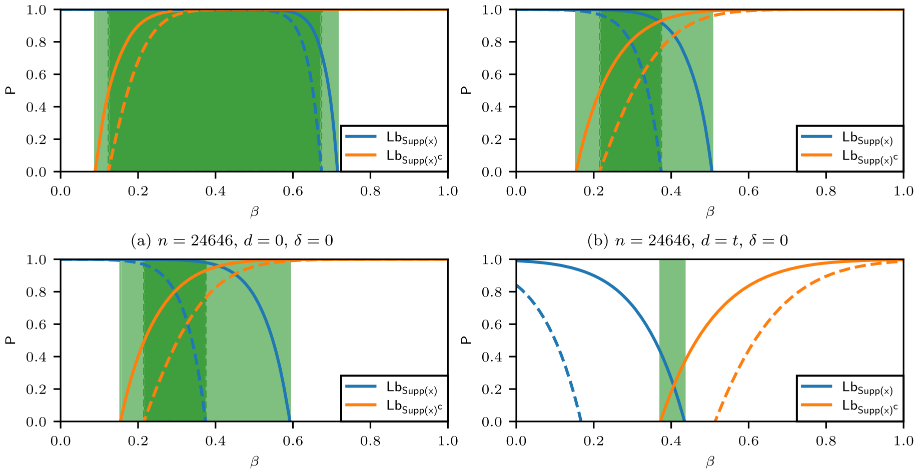
4.1.3 Range of valid parameters

Here, we shall determine the conditions on the parameters such that the conditions in Proposition 6 are satisfied. We will begin by determining the existence of β^*. We will need to denote by $W(x)$ the Lambert W function.

Proposition 7 *For any* $\beta \geq \sqrt{\frac{t+2d}{n-k} W\left(\frac{n-t}{n-k-t+\delta+1} \frac{e}{\sqrt{2\pi}}\right)^2}$ *we have that* $\frac{n-t}{n-k-t+\delta+1} \mathrm{Ub}_{\mathsf{Supp}(\boldsymbol{x})^c}(n, k, t, d, \beta) \leq 1$, *and for any* $\beta \leq 1 - \sqrt{\frac{t+2d-1}{n-k} W\left(\frac{t}{\delta+1} \frac{e}{\pi}\right)^2}$ *we have that* $\frac{t}{\delta+1} \mathrm{Ub}_{\mathsf{Supp}(\boldsymbol{x})}(n, k, t, d, \beta) \leq 1$.

Having both functions positive and strictly smaller than 1, at the same time, can be achieved as long the interval defined by the two extreme points, in the previous proposition is non-empty, i.e., $\sqrt{\frac{t+2d}{n-k} W\left(\frac{n-t}{n-k-t+\delta+1} \frac{e}{\sqrt{2\pi}}\right)^2} \leq 1 - \sqrt{\frac{t+2d-1}{n-k} W\left(\frac{t}{\delta+1} \frac{e}{\pi}\right)^2}$.

To give a more sensitive meaning of our result, we could approximate the value of the Lambert W function by $W(m) = \ln m - \ln\ln m + \frac{\ln\ln m}{\ln m}$ as m tends to infinity. Using only



(a) $n = 24646,\ d = 0,\ \delta = 0$

(b) $n = 24646,\ d = t,\ \delta = 0$

(c) $n = 24646,\ d = t,\ \delta = 3$

(d) $n = 3488,\ d = t,\ \delta = 3$

Fig. 2 Valid β interval from the bounds in [25] (dashed lines) and the proposed ones (solid lines)

the first term we define $I_\beta = \left[\sqrt{\frac{2(t+2d)}{n-k}\ln\frac{n-t}{n-k-t+\delta+1}}, 1-\sqrt{\frac{2(t+2d-1)}{n-k}\ln\frac{t}{\delta+1}}\right]$. Hence, we deduce the following result.

Proposition 8 *If $I_\beta \neq \emptyset$ then the probability of success of the ISD-score decoder is at least*

$$\left(1-\frac{e}{2\pi}\frac{1}{\sqrt{\ln\frac{n-t}{n-k-t+\delta+1}}}\right)\left(1-\frac{e}{\sqrt{2}\pi}\frac{1}{\sqrt{\ln\frac{t}{\delta+1}}}\right).$$

Typically, our result gives a sub-interval where the conditions are safely satisfied. When simulations are to be performed, one could solve the inequalities in order to determine a more accurate interval. However, using more terms in the expansion of $W(m)$ yields the following.

Corollary 4 *Let $f_{n,k,t,\delta} = \frac{n-t}{n-k-t+\delta+1}$ and $f^*_{t,\delta} = \frac{t}{\delta+1}$. The extreme points of the interval where the first two conditions in Theorem 6 are satisfied, converges to*
$\sqrt{\frac{t+2d}{n-k}\left(2\ln f_{n,k,t,\delta} - \ln 2\ln f_{n,k,t,\delta} + \frac{\ln 2\ln f_{n,k,t,\delta}}{2\ln f_{n,k,t,\delta}}\right)}$,
and $1-\sqrt{\frac{t+2d-1}{n-k}\left(2\ln f^*_{t,\delta} - \ln 2\ln f^*_{t,\delta} + \frac{\ln 2\ln f^*_{t,\delta}}{2\ln f^*_{t,\delta}}\right)}$.

4.2 Information-theoretic bounds

4.2.1 Bounding the value of *t*

To see how large the weight of the error t must be to in order to still posses a non-empty interval, the following rough estimate can be used.

Theorem 6 **(Upper bound on** t **)** *Let $k \le n-t+\delta+1-(n-t)(\delta+1)/t$ and $d = ct/2$. Then $I_\beta \neq \emptyset$ as long as we have*

$$t \le \frac{n-k}{8(1+c)W\left(\frac{n-k}{8(1+c)(\delta+1)}\right)} \tag{11}$$

Moreover, when $n \to \infty$, we have that $t \le \mathcal{O}\left(\frac{n-k}{\ln(n-k)}\right)$.

Using a first term approximation for the Lambert W function near infinity, we obtain a threshold on t. More exactly this value can be approximated by $\frac{n-k}{8(1+c)\ln\frac{n-k}{8(1+c)(\delta+1)}}$.

Now, recall that we have determined a preliminary condition on d, such that the ψ function can distinguish between positions in the support of the solution and outside it. This condition was $d \le \frac{n-k}{8\ln\ln(n-k)}$. Taking a slightly smaller noise level, *e.g.* $d = \frac{n-k}{8\ln(n-k)} \le \frac{n-k}{8\ln\ln(n-k)}$ validates the choice in the hypothesis $d = ct/2$, as per Theorem 6 $t \le \mathcal{O}\left(\frac{n-k}{\ln(n-k)}\right)$. Taking into account this condition and the hypothesis of Theorem 6, *i.e.* $d = ct/2$, we deduce the following upper bound on t

$$d = \frac{ct}{2} \le \frac{n-k}{8\ln t} \;\Rightarrow\; t\ln t \le \frac{n-k}{4c}. \tag{12}$$

This improves the constant term by $t \le \dfrac{n-k}{4cW(\frac{n-k}{4c})}$.

Remark 1 When we consider the conditions from Proposition 7 we can deduce a similar, but stronger condition on t. Indeed under the same assumption on k we have

$$t \leq \frac{n-k}{(12(1+c)W\left(\frac{1}{3}\left(\frac{n-k}{4(1+c)(\delta+1)}\right)^{\frac{2}{3}}\right)} \tag{13}$$

Asymptotically we obtain the same behavior, however, the constant factors matter when numerical simulations are performed (Table 3).

4.2.2 Bounding the required ratio of syndrome entries

The existence of a value such that the ISD-score decoder succeeds in finding a solution using fewer syndrome entries could be deduced. It suffices to replace $(n-k)$ with $\gamma(n-k)$, where $\gamma \in (0, 1]$ represents the percentage of syndrome entries required to achieve a high probability. This value can be deduced from Theorem 6. Typically, given a number of rows $n-k$, the maximum value of t for which the success probability is close enough to 1 also determines the minimum number of required rows. More exactly, for a fixed value of t and $n-k$, we can compute $\gamma(n-k)$, the value for which t satisfies $8t(1+c)\ln\frac{t}{\delta+1} = \gamma(n-k)$. By Theorem 6, with only $\gamma(n-k)$ rows, one can recover a solution of weight at most t with high probability. Formally, the following holds.

Corollary 5 *Let $d = ct/2$ where c is a constant. Then the minimum quantity of information required by the ISD-score decoder to find a valid solution is $4(1+c)t\ln\frac{t}{\delta+1}$. Moreover, in the noiseless scenario, the minimum quantity of information becomes $4t\ln\frac{t}{\delta+1}$.*

Consequently, we deduce that one could improve the constant term, however, not lower than $2(1+c)t\ln\frac{t}{\delta+1}$.

On a parallel track, when we analyze the condition on $n-k$ from Theorem 6 we find that $n-k \geq t + (\delta+1)n/t - 2(\delta+1)$. Equivalently this leads to a minimum dimension of order $\mathcal{O}\left(\max(t, \frac{n}{t})\right)$. Hence, there is a universal condition on the fraction on syndrome

Table 3 Theoretical upper bounds on t

$n-k$	1000	5000	10000	15000	20000	25000	30000	35000	40000	45000	50000
$\delta = 0, c = 0$	35	129	230	324	415	502	588	672	755	836	916
	43	156	276	388	495	598	699	798	895	990	1084
$\delta = 3, c = 0$	50	167	291	406	515	620	722	822	920	1016	1111
	62	206	356	493	624	749	871	990	1106	1220	1333
$\delta = 0, c = 0.25$	30	107	190	268	343	415	485	554	622	689	755
	36	130	230	322	410	495	578	659	739	817	895
$\delta = 3, c = 0.25$	42	140	243	338	428	515	599	681	762	841	920
	52	173	298	412	520	624	724	822	919	1013	1106
$\delta = 0, c = 0.5$	26	92	164	230	293	355	415	473	531	588	644
	32	112	198	276	352	424	495	564	632	699	765
$\delta = 3, c = 0.5$	37	122	210	291	368	443	515	585	654	722	789
	46	150	258	356	448	537	624	708	790	871	950

entries required to solve the problem regardless of the relation between d and t. Indeed, choosing the particular relation $d = ct/2$ allows to determine the maximum value of t, but this value depends on d. Typically, all these variables are linked together. That is why we can express the tolerated noise level in function of the syndrome entries and weight t, or the minimum syndrome entries in function of the maximum decodable weight t which depends on the noise level.

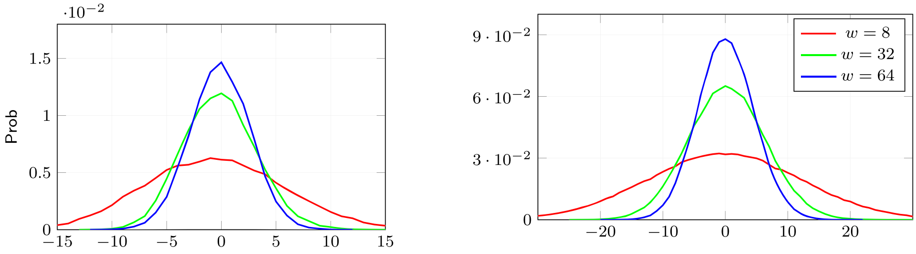
5 Experimental results

The following experiments have been carried out on a standard laptop embedding an 8-core processor running at 1.6 GHz and 32 GB of RAM. The ILP solver we used is provided by the Scipy Python package [53] under the `scipy.optimize.linprog` function. The score decoder is implemented using the Numpy Python package [31] to perform matrix computations.

5.1 Noise model

We have simulated the noise model as per [29]. More exactly, we will handle here two type of errors, coming from the accuracy of the side-channel distinguisher and double-cancellation. Thus, the noise will consist of two parts $\mathrm{HW}(\boldsymbol{b}_{i,j}) + \mathcal{N}(0, \sigma^2)$, where the Hamming weight HW depends on the width of the representation (8, 32 or 64-bit values) and the noise variance σ^2 affects the side-channel distinguisher accuracy. The accuracy of distinguisher (a) is approximated using the 3-σ rule $a \simeq \mathrm{erf}\left(\frac{1}{2\sqrt{2}\sigma}\right)$, where erf is the Gauss error function [56]. We have also taken into account the parity of $\boldsymbol{s}^*$ to correct wrong estimated values of $\widetilde{\boldsymbol{s}}$. The noise model for $\sigma = 0.25$, $\sigma = 0, 5$ and two parameter sets for the *Classic McEliece* KEM are illustrated in Fig. 3. The plotted distribution is a truncated shifted distribution since, i) one out of two values are equal to 0 (correction with respect to the parity of the binary syndrome) and ii) shifted depending on the width of the representation and the values of σ.

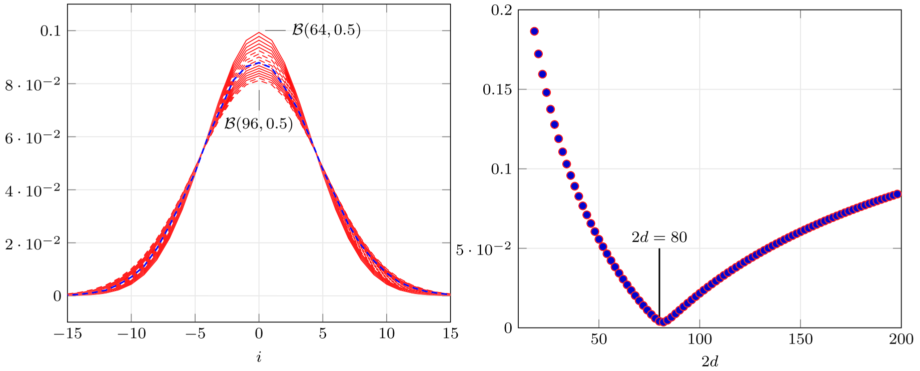
To determine the closest distribution to the simulated noise, we have first plotted in Fig. 4a a possible interval of values d where the most probable $\mathcal{B}(d, 0.5)$ could be. Then we have computed the Euclidean distance between the simulated noise distribution and $\mathcal{B}(2d, 0.5) - d$ for $d = 5..100$. The sequence of distances is decreasing from $d = 5$ to $d = 40$ where the minimum is reached (the distance for $d = 40$ equals 0.004), and then increasing.



(a) $n = 8192, t = 128, \sigma = 0.25$

(b) $n = 8192, t = 128, \sigma = 0.5$

Fig. 3 Simulated noise distribution



(a) Simulate noise vs. $\mathcal{B}(2d, 0.5) - d$ for $d = 32..48$

(b) Euclidean distance between $\mathcal{B}(d, 0.5)$ and simulated noise.

Fig. 4 Binomial model for simulated noise ($\sigma = 0.5$) in the case of the *Classic McEliece* of parameters $n = 8192, t = 128$

In Table 4, we have computed the "best" (the closest w.r.t. the Euclidean distance) binomial distributions for the simulated noise. Light green signifies values of d close to $t/8$, while dark colors indicates large values of d, typically larger than $3t$. As expected, the value of d is increasing with σ and decreasing with w. The first parameter (σ) induces errors in the estimation of the intermediate weights which obviously affects negatively the parameter d. We can see that in Table 4a where we have computed the values for values of σ up to 0.75

Table 4 Closest binomial distributions to simulated noise. Each value represents the best value for the parameter d in $\mathcal{B}(2d, 0.5) - d$

(a) Noise $\sigma \in \{0.25, 0.50, 0.75\}$

Classic McEliece $n = 3488, k = 2720, t = 64$

	$w = 8$	$w = 32$	$w = 64$
$\sigma = 0.25$	35	9	6
$\sigma = 0.50$	132	33	17
$\sigma = 0.75$	208	50	27

Classic McEliece $n = 8192, k = 6528, t = 128$

	$w = 8$	$w = 32$	$w = 64$
$\sigma = 0.25$	82	22	15
$\sigma = 0.50$	309	76	41
$\sigma = 0.75$	491	115	63

BIKE $n = 24646, k = 12323, t = 134$

	$w = 8$	$w = 32$	$w = 64$
$\sigma = 0.25$	247	59	30
$\sigma = 0.50$	≥ 500	224	113
$\sigma = 0.75$	≥ 500	342	172

(b) Noise $\sigma \in \{0.1, 0.2, 0.3, 0.4\}$

Classic McEliece $n = 3488, k = 2720, t = 64$

	$w = 8$	$w = 32$	$w = 64$
$\sigma = 0.1$	5	5	5
$\sigma = 0.2$	10	5	5
$\sigma = 0.3$	63	10	8
$\sigma = 0.4$	123	26	13

Classic McEliece $n = 8192, k = 6528, t = 128$

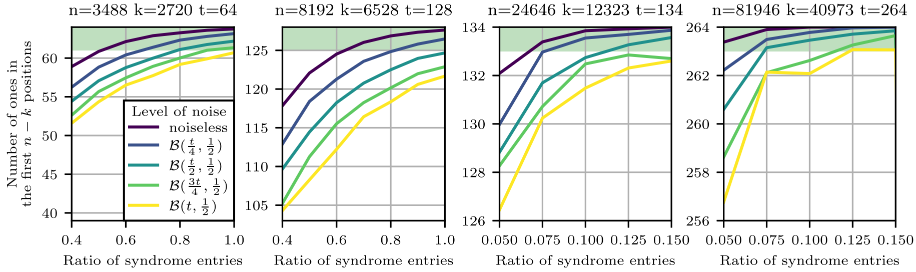
	$w = 8$	$w = 32$	$w = 64$
$\sigma = 0.1$	5	5	5
$\sigma = 0.2$	25	10	10
$\sigma = 0.3$	154	39	22
$\sigma = 0.4$	251	62	34

(extremely noisy setting). In real situations [16] the largest values do not exceed $\sigma = 0.2$. The second parameter (w) has a converse influence on d, since the larger the width of the registers the lower the noise level. This mainly comes from the fact that when the registers are large there are fewer blocks n/w on which the noise gets accumulated. Hence, there is a smaller influence in the intermediate estimated values that comes from this side. Keeping all these in mind, we see that light green colors are predominating in the small σ and large w region. For realistic scenarios in Table 4b the first two lines in each parameter set shows that $d < t/4$ for all w.

5.2 Success probability and ratio of syndrome entries

The following experiments look at the number of syndrome entries required to bring $t - \delta$ ones in the first $n - k$ positions, as dictated by the ISD method. Results are shown in Fig. 5, for both the *Classic McEliece* and the BIKE cryptosystems. Let us explain the meaning of the plots, when these are read horizontally. One way this could be read is as the weight of solutions retrieved by the ISD-score decoder with probability 1. The green stripe represents the region corresponding to possible values of δ. The value of δ for the $[t - \delta; t]$ interval is lower for the BIKE cryptosystem since it comes with much larger values of n, making the exhaustive search for the correct permutation much more costly. Conversely, we allow for $\delta = 3$ in the case of *Classic McEliece* since the n values are smaller. For example, when $n = 8192$ and noise level equal to t we can hope to retrieve solutions of weight at most 122 (which is smaller than the proposed parameters), while for the same length and noise smaller than $t/2$ we can retrieve any solution of weight at most 128 using the ISD-score decoder using $\delta = 3$, or equivalently solutions of weight 125 using the Prange-score decoder. To summarize, except for the case $n = 8192$ with noise levels strictly greater than $t/2$, all the plots suggest that the ISD-score decoder is able to retrieve with high probability a valid solution of weight t in presence of noise.

We can also read the plots vertically. This gives us the ratio of syndrome entries required to find a solution of given weight with high probability. The abscissa of the points of intersection between the curves and the green stripe gives minimum percentage of syndrome entries required in the ISD-score decoder to successfully retrieve a valid solution of weight t. For the BIKE cryptosystem, the ratio of syndrome entries required to bring at least $t - 1$ ones in the first $n - k$ positions ranges from 4.75% to 6.5%. For the *Classic McEliece* cryptosystem, the ratio of syndrome entries required to bring at least $t - 3$ ones in the first $n - k$ positions



(a) *Classic McEliece* parameters set and the $[t-3, t]$ interval (b) BIKE parameters set and the $[t-1, t]$ interval

Fig. 5 Number of ones in the first $n-k$ positions for some of the *Classic McEliece* and BIKE sets of parameters and different levels of a centered binomial noise

ranges from 48% to 62%. We have also computed the best theoretical lower bound we could hope for, *i.e.*, the percentage of syndrome entries should be at least $\frac{2(1+c)t}{n-k} \ln \frac{t}{\delta+1}$. When comparing the experimental results shown in Fig. 5 and Table 5, we observe that theoretical values are around 10% smaller than the experimental values.

To verify that the public parity-check matrix is close to a Bernoulli matrix, we verify the bias of the coefficients. We use a χ^2 test with the null hypothesis that the distribution of 1 and 0's follows a uniform distribution. For all sets of parameters, we obtain a p-value higher than 0.999, allowing us to accept the null hypothesis. We also check that the distribution of transitions between the coefficients follow a uniform distribution, using a χ^2 test and uniform distribution as null hypothesis. We obtain a p-value higher than 0.999, allowing us to accept the null hypothesis.

5.3 ILP solver and ISD-score decoder

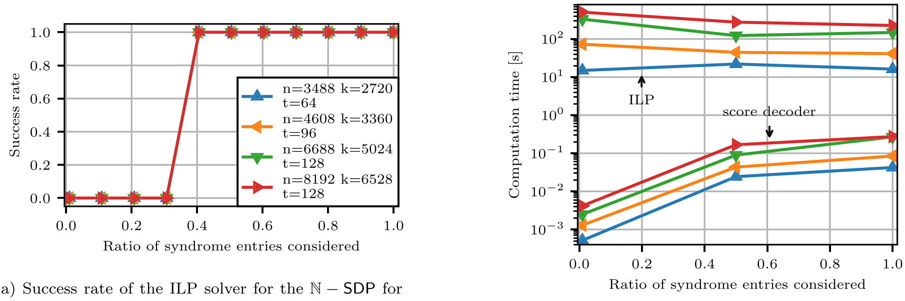
Percentage of required entries To compare the ILP solver with the ISD-score decoder we used the parameters for the *Classic McEliece* proposal. We decided to consider only the *Classic McEliece* because the execution time of the ILP solver for the smallest parameters of BIKE exceeded tens of minutes for a single instance of the $\mathbb{N}-\mathsf{SDP}$. Obtaining in a reasonable time a solid statistical evidence of the performance of the ILP solver for BIKE, would assume a much more optimized implementation of the solver, which is not the main purpose of this article. The results for the ILP solver in the noiseless scenario are given in Fig. 6a. The success rate is computed for ten evenly spaced ratios ranging from 1 to 100%.

We observe that the behavior is the same for all sets of parameters. When considering 30% of syndrome entries, the ILP solver failed at recovering the error vector ten times out of ten. Conversely, when considering 40% of syndrome entries, the ILP solver succeeded at recovering the error vector ten times out of ten. Hence, the main drawback of the ILP solver, when compared to the ISD-score decoder, is that the ILP cannot be used when only a small percentage of syndrome entries are known.

Noisy setting In a noisy setting, the differences between the ILP solver and the ISD-score decoder is even more dramatic. Indeed, the ILP solver either succeeds in finding a valid solution, with t ones in the first t positions, or it fails. Conversely, the ISD-score decoder succeeds if $t-\delta$ ones are in the first $(n-k)$ positions, providing a much larger margin in the noisy setting.

Table 5 Theoretical lower bound on the ratio of syndrome entries necessary for the ISD-score decoder

n	noiseless	$\mathcal{B}\left(\frac{t}{4},\frac{1}{2}\right)$	$\mathcal{B}\left(\frac{t}{2},\frac{1}{2}\right)$	$\mathcal{B}\left(\frac{3t}{4},\frac{1}{2}\right)$	$\mathcal{B}\left(t,\frac{1}{2}\right)$
Classic McEliece					
3488	0.46	0.58	0.69	0.81	0.92
4608	0.49	0.61	0.73	0.86	0.98
6688	0.53	0.67	0.80	0.93	1.00
8192	0.53	0.67	0.80	0.93	1.00
BIKE					
24646	0.09	0.11	0.14	0.16	0.18
49318	0.07	0.09	0.11	0.13	0.15
81946	0.06	0.08	0.09	0.11	0.13



(a) Success rate of the ILP solver for the $\mathbb{N}-\mathsf{SDP}$ for four sets of parameters and different ratios of syndrome entries considered

(b) Computation time of the ILP solver and the ISD-score decoder

Fig. 6 ILP and ISD-score decoder performance for $\mathbb{N}-\mathsf{SDP}$

Eventually, the permutation returned by the ISD-score decoder is always better than a random permutation. Therefore, one can always resort to exhaustive search afterwards.

Computation time When comparing the time required by the two algorithms for retrieving a valid solution, we notice a significant gap between the two algorithms. From Fig. 6b we can see that it takes less than 0.1 s for the ISD-score decoder, while for the ILP it takes at least 10 s for any of the parameters of the *Classic McEliece* scheme. Broadly speaking, the ILP solver is three orders of magnitude slower than the ISD-score decoder.

6 Conclusion

This article evaluated the efficiency of the score decoder for integer syndrome decoding in the presence of noise. We proved that, even in the presence of noise, this decoder is indeed able to successfully bring $t-\delta$ ones in the first $n-k$ positions, as required by the ISD-based methods. We then experimentally validate this capability considering the parameter sets of two post-quantum cryptosystems, *Classic McEliece* and BIKE. Future works could investigate other types of noise, improve the efficiency of the decoder, or consider other type of distributions. For example, LDPC or MDPC parity-check matrices offer interesting results in simulations, fact that opens the following question. What is the influence of the matrix sparsity on the success probability?

Appendix A proof of theorem 1

Proof By definition 4 we have that

$$\psi_i(\boldsymbol{s}) = \sum_{\ell=1}^{n-k} \big(h_{\ell,i}s_\ell + (1-h_{\ell,i})(t-s_\ell)\big)\,. \tag{14}$$

Let us denote $X_\ell = h_{\ell,i}s_\ell + (1-h_{\ell,i})(t-s_\ell))$. As $s_\ell = \sum_{j\in\mathsf{Supp}(\boldsymbol{x})} h_{\ell,j}$ we deduce that

$$X_\ell = h_{\ell,i}\sum_{j\in\mathsf{Supp}(\boldsymbol{x})} h_{\ell,j} + (1-h_{\ell,i})(t-\sum_{j\in\mathsf{Supp}(\boldsymbol{x})} h_{\ell,j})). \tag{15}$$

If $i \notin \mathsf{Supp}(\boldsymbol{x})$ then

$$X_\ell = \begin{cases} \sum_{j \in \mathsf{Supp}(\boldsymbol{x})} h_{\ell,j} \quad = s_\ell \quad , \text{if } h_{\ell,i} = 1 \\ t - \sum_{j \in \mathsf{Supp}(\boldsymbol{x})} h_{\ell,j} = t - s_\ell \text{ , if } h_{\ell,i} = 0 \end{cases}$$

As $s_\ell \sim \mathcal{B}(t, \frac{1}{2})$ we deduce that $X_\ell \sim \mathcal{B}(t, \frac{1}{2})$ for all $i \notin \mathsf{Supp}(\boldsymbol{x})$, and by independence we obtain $\psi_i(\boldsymbol{s}) \sim \mathcal{B}((n-k)t, \frac{1}{2})$, then $\mathbb{E}(\psi_i(\boldsymbol{s})) = \frac{(n-k)t}{2}$ and $Var(\psi_i(\boldsymbol{s})) = \frac{(n-k)t}{4}$.

If $i \in \mathsf{Supp}(\boldsymbol{x})$ we have that s_ℓ and $h_{\ell,i}$ are dependent random variables. Hence we obtain

$$X_\ell = \begin{cases} 1 + \sum_{j \in \mathsf{Supp}(\boldsymbol{x}) \setminus \{i\}} h_{\ell,j} \quad , \text{if } h_{\ell,i} = 1 \\ 1 + (t-1) - \sum_{j \in \mathsf{Supp}(\boldsymbol{x}) \setminus \{i\}} h_{\ell,j} \text{ , if } h_{\ell,i} = 0 \end{cases}$$

As $s_\ell - h_{\ell,i} \sim \mathcal{B}(t-1, \frac{1}{2})$ we deduce that $X_\ell \sim 1 + \mathcal{B}(t-1, \frac{1}{2})$ for all $i \in \mathsf{Supp}(\boldsymbol{x})$, and by independence of the variables X_ℓ we obtain $\psi_i(\boldsymbol{s}) \sim (n-k) + \mathcal{B}((n-k)(t-1), \frac{1}{2})$.

Appendix B proof of corollary 2

Proof Apply Proposition 2 and Corollary 1 to obtain the the results. In order to determine the upper bound on d, we start by computing the intervals of confidence for $\psi_i(\widetilde{\boldsymbol{s}})$ from Proposition 2. This yields an interval $I_{\widetilde{\boldsymbol{s}}}(i)$ defined by the two extremal points $\mathbb{E}(\psi_i(\widetilde{\boldsymbol{s}})) \pm \sqrt{\frac{d(n-k)g(n,k,t)}{2}}$, i.e.,

- $i \notin \mathsf{Supp}(\boldsymbol{x})$ the points $\frac{(n-k)t}{2} \pm \sqrt{\frac{d(n-k)g(n,k,t)}{2}}$
- $i \in \mathsf{Supp}(\boldsymbol{x})$ the points $\frac{(n-k)t}{2} + \frac{n-k}{2} \pm \sqrt{\frac{d(n-k)g(n,k,t)}{2}}$.

The two intervals are disjoint if we have

$$2\sqrt{\frac{d(n-k)g(n,k,t)}{2}} \leq \frac{n-k}{2} \tag{16}$$

Hence, we obtain $d \leq \frac{n-k}{8g(n,k,t)}$.

Appendix C proof of theorem 3 and theorem 4

Let us begin by a useful result.

Lemma 1 ([35]) *Let $X \sim \mathcal{B}(n, \frac{1}{2})$ and $\frac{n}{2} \leq \alpha \leq n$. Then*

$$\mathsf{Prob}(X \geq \alpha) \leq \frac{\alpha+1}{2\alpha - n + 1} \mathsf{Prob}(X = \alpha). \tag{17}$$

Lemma 2 ([25]) *Let $X \sim \mathcal{B}(n, \frac{1}{2})$ and $\alpha \leq \frac{n}{2}$. Then*

$$\mathsf{Prob}\left(X = \frac{n}{2} + \alpha\right) \leq \frac{e}{2\pi} \sqrt{\frac{n}{\frac{n^2}{4} - \alpha^2}} e^{-\frac{2\alpha^2}{n}}. \tag{18}$$

Proposition 9 *Let $X \sim \mathcal{B}(n, \frac{1}{2})$ and $\alpha < n$. Then*

$$\mathsf{Prob}\left(X \geq \frac{n}{2} + \frac{\alpha}{2}\right) \leq \frac{e}{2\pi}\left(1 + \frac{n+1}{\alpha+1}\right)\sqrt{\frac{n}{n^2-\alpha^2}}e^{-\frac{\alpha^2}{2n}}. \tag{19}$$

Proof Use Lemma 1 and 2.

We can now proceed to the proof of Theorem 3.

Proof Recall that

$$\psi_i(\boldsymbol{s}) \sim \begin{cases} \mathcal{B}((n-k)t, \frac{1}{2}) & \text{for } i \notin \mathsf{Supp}(\boldsymbol{x}) \\ n-k+\mathcal{B}((n-k)(t-1), \frac{1}{2}) & \text{for } i \in \mathsf{Supp}(\boldsymbol{x}) \end{cases}$$

By Proposition 9, for $i \notin \mathsf{Supp}(\boldsymbol{x})$ we have that

$$\begin{aligned} \mathsf{Prob}\left(\psi_i(\boldsymbol{s}) \geq B_\beta\right) &\leq \frac{\frac{e}{2\pi}\left(1+\frac{(n-k)t+1}{(n-k)\beta+1}\right)\sqrt{\frac{(n-k)t}{(n-k)^2t^2-(n-k)^2\beta^2}}}{e^{\frac{(n-k)^2\beta^2}{2(n-k)t}}} & (20) \\ &\leq \frac{e}{2\pi\beta}\sqrt{\frac{t+\beta}{t-\beta}}\sqrt{\frac{t}{(n-k)}}e^{-\frac{(n-k)\beta^2}{2t}} & (21) \\ &\leq \frac{e}{\sqrt{2}\pi\beta}\sqrt{\frac{t}{(n-k)}}e^{-\frac{(n-k)\beta^2}{2t}} & (22) \end{aligned}$$

For $i \in \mathsf{Supp}(\boldsymbol{x})$ we have that $\mathbb{E}(\psi_i(\boldsymbol{s})) = \frac{(n-k)t}{2} + \frac{n-k}{2}$. Hence, by Proposition 9 we obtain that $\mathsf{Prob}\left(\psi_i(\boldsymbol{s}) \leq \frac{(n-k)t}{2} + \frac{(n-k)\beta}{2}\right)$ is upper bounded by

$$\begin{aligned} &\leq \frac{\frac{e}{2\pi}\left(1+\frac{(n-k)(t-1)+1}{(n-k)(1-\beta)+1}\right)\sqrt{\frac{(n-k)(t-1)}{(n-k)^2(t-1)^2-(n-k)^2(1-\beta)^2}}}{e^{\frac{(n-k)^2(1-\beta)^2}{2(n-k)(t-1)}}} & (23) \\ &\leq \frac{e}{2\pi(1-\beta)}\sqrt{\frac{t-\beta}{t+\beta+2}}\sqrt{\frac{t-1}{(n-k)}}e^{-\frac{(n-k)(1-\beta)^2}{2(t-1)}} & (24) \\ &\leq \frac{e}{2\pi(1-\beta)}\sqrt{\frac{t-1}{(n-k)}}e^{-\frac{(n-k)(1-\beta)^2}{2(t-1)}}. & (25) \end{aligned}$$

As for Theorem 4 we have:

Proof Recall that we have

- for $i \notin \mathsf{Supp}(\boldsymbol{x})$

$$\psi_i(\widetilde{\boldsymbol{s}}) \sim -d(n-k) + \mathcal{B}\left((n-k)(t+2d), \frac{1}{2}\right);$$

- for $i \in \mathsf{Supp}(\boldsymbol{x})$

$$\psi_i(\widetilde{\boldsymbol{s}}) \sim -(d-1)(n-k) + \mathcal{B}\left((n-k)(t-1+2d), \frac{1}{2}\right).$$

The proof is thus identical with that of Theorem 3 by simply putting $t' = t + 2d$ when $i \notin \mathsf{Supp}(\boldsymbol{x})$ and $t' = t + 2d - 1$ when $i \in \mathsf{Supp}(\boldsymbol{x})$.

Appendix D proof of theorem 5

Proof Let X_B denote the number of indices $j \in \mathsf{Supp}(\boldsymbol{x})$ for which $\psi_i(\widetilde{\boldsymbol{s}}) \le B$, and Y_B the number of indices $j \notin \mathsf{Supp}(\boldsymbol{x})$ for which $\psi_i(\widetilde{\boldsymbol{s}}) \ge B$. The probability of success of our algorithm equals

$$\begin{aligned}
&\sum_B \mathsf{Prob}(X_B \le \delta) \cdot \mathsf{Prob}\left(Y_B \le n-k-t+\delta\right) \\
&= \sum_B (1 - \mathsf{Prob}(X_B \ge \delta+1)) \cdot (1 - \mathsf{Prob}(Y_B \ge n-k-t+\delta+1)) \\
&\ge \sum_B \left(1 - \frac{t}{\delta+1} e^{-f(n,k,t,d,B)}\right) \cdot \left(1 - \frac{n-t}{n-k-t+\delta+1} e^{-g(n,k,t,d,B)}\right).
\end{aligned}$$

In the last equation we have used Markov's inequality. Also, the last sum is over those values B for which the two terms in the sum are both positive and smaller than 1. Now suppose that a B^* satisfying the required condition exists. Then the probability of success is

$$\begin{aligned}
&\ge \left(1 - \frac{t}{\delta+1} e^{-f(n,k,t,d,B^*)}\right) \cdot \left(1 - \frac{n-t}{n-k-t+\delta+1} e^{-g(n,k,t,d,B^*)}\right) \\
&\ge 1 - \frac{t}{\delta+1} e^{-f(n,k,t,d,B^*)} - \frac{n-t}{n-k-t+\delta+1} e^{-g(n,k,t,d,B^*)}.
\end{aligned}$$

Appendix E range of valid parameters: proofs and comments

Let us denote $f(n,k,t,d,\beta) = \frac{t}{\delta+1}\mathrm{Ub}_{\mathsf{Supp}(\boldsymbol{x})}(n,k,t,\beta)$ and $g(n,k,t,d,\beta) = \frac{n-t}{n-k-t+\delta+1}\mathrm{Ub}_{\mathsf{Supp}(\boldsymbol{x})^c}(n,k,t,d,\beta)$. The first useful results concerns the monotony of the two upper bounds.

Lemma 3 *The functions* $f(n,k,t,d,\beta)$, $g(n,k,t,d,\beta)$ *in* $\beta \in (0,1)$*, are positive increasing, and positive decreasing, respectively.*

Proof We have that both functions f, g are positive. We also have

$$\begin{aligned}
\frac{\partial f(n,k,t,d,\beta)}{\partial\beta} &= \frac{(n-k)(1-\beta)^2 + (t+2d-1)}{(1-\beta)(t+2d-1)} f(n,k,t,d,\beta) \\
\frac{\partial g(n,k,t,d,\beta)}{\partial\beta} &= -\frac{(n-k)\beta^2 + (t+2d)}{\beta(t+2d)} g(n,k,t,d,\beta).
\end{aligned}$$

Using the fact that f and g are positive we deduce the wanted result.

Now we can demonstrate Proposition 7.

Proof Let us consider the limit point β where the two functions equal 1. As the first function is decreasing we then obtain a lower bound on β.

$$g(n,k,t,d,\beta) = 1 \Leftrightarrow \tag{26}$$

$$\frac{n-t}{n-k-t+\delta+1}\frac{e}{\sqrt{2\pi}\beta}\sqrt{\frac{t+2d}{n-k}}\, e^{-\frac{(n-k)\beta^2}{2(t+2d)}} = 1 \Leftrightarrow \tag{27}$$

$$\left(\frac{n-t}{n-k-t+\delta+1}\frac{e}{\sqrt{2\pi}}\right)^2 \frac{t+2d}{(n-k)\beta^2} = e^{\frac{(n-k)\beta^2}{t+2d}}. \tag{28}$$

By letting $y = \frac{(n-k)\beta^2}{t+2d}$ we have

$$ye^y = \left(\frac{n-t}{n-k-t+\delta+1}\frac{e}{\sqrt{2\pi}}\right)^2, \tag{29}$$

admitting a real solution $y = W\left(\frac{n-t}{n-k-t+\delta+1}\frac{e}{\sqrt{2\pi}}\right)^2$, where W is the Lambert W function. From this we deduce $\beta = \sqrt{\frac{t+2d}{n-k}W\left(\frac{n-t}{n-k-t+\delta+1}\frac{e}{\sqrt{2\pi}}\right)^2}$. The second function is increasing hence, it gives an upper bound on β.

$$f(n,k,t,d,\beta) = 1 \Leftrightarrow \tag{30}$$

$$\frac{t}{\delta+1}\frac{e}{\pi(1-\beta)}\sqrt{\frac{t+2d-1}{n-k}}e^{-\frac{(n-k)(1-\beta)^2}{2(t+2d-1)}} = 1 \Leftrightarrow \tag{31}$$

$$\left(\frac{t}{\delta+1}\frac{e}{\pi}\right)^2\frac{t+2d-1}{(n-k)(1-\beta)^2} = e^{\frac{(n-k)(1-\beta)^2}{t+2d-1}}. \tag{32}$$

As in the first case we obtain $1-\beta = \sqrt{\frac{t+2d-1}{n-k}W\left(\frac{t}{\delta+1}\frac{e}{\pi}\right)^2}$.

Proposition 8 gives a slightly weaker condition, however, it helps understanding the order of magnitude of the parameters. Let us demonstrate the result.

Proof Let $\beta \geq \beta_1 = \sqrt{2\frac{t+2d}{n-k}\ln\frac{n-t}{n-k-t+\delta+1}}$. Then we have

$$g(n,k,t,d,\beta) = \frac{n-t}{n-k-t+\delta+1}\frac{e}{\sqrt{2\pi}\beta}\sqrt{\frac{t+2d}{n-k}}e^{-\frac{(n-k)\beta^2}{2(t+2d)}} \tag{33}$$

$$\leq \frac{n-t}{n-k-t+\delta+1}\frac{e}{2\pi\sqrt{\ln\frac{n-t}{n-k-t+\delta+1}}}e^{-\ln\frac{n-t}{n-k-t+\delta+1}} \tag{34}$$

$$= \frac{e}{2\pi}\frac{1}{\sqrt{\ln\frac{n-t}{n-k-t+\delta+1}}}. \tag{35}$$

Let $1-\beta \geq \beta_2 = \sqrt{2\frac{t+2d-1}{n-k}\ln\frac{t}{\delta+1}}$. Then

$$f(n,k,t,d,\beta) = \frac{t}{\delta+1}\frac{e}{\pi(1-\beta)}\sqrt{\frac{t+2d-1}{n-k}}e^{-\frac{(n-k)(1-\beta)^2}{2(t+2d-1)}} \tag{36}$$

$$\leq \frac{e}{\sqrt{2\pi}}\frac{1}{\sqrt{\ln\frac{t}{\delta+1}}}. \tag{37}$$

From this we deduce

$$\frac{\sqrt{2(t+2d)}}{\sqrt{n-k}}\sqrt{\ln\frac{n-t}{n-k-t+\delta+1}} \leq \beta \tag{38}$$

$$\beta \leq 1 - \frac{\sqrt{2(t+2d-1)}}{\sqrt{n-k}}\sqrt{\ln\frac{t}{\delta+1}}. \tag{39}$$

Now, suppose that $[\beta_1, \beta_2]$ is non-empty and take $\beta^* \in [\beta_1, \beta_2]$. Since $f(n, k, t, d, \beta)$ is increasing in β, this implies that

$$f(n, k, t, d, \beta^*) \leq f(n, k, t, d, \beta_2) = \frac{e}{\sqrt{2}\pi} \frac{1}{\sqrt{\ln \frac{t}{\delta+1}}} < 1, \tag{40}$$

as long as $t > 1.45(\delta + 1)$. Also, as $g(n, k, t, d, \beta)$ is decreasing in β we have that

$$g(n, k, t, d, \beta^*) \leq g(n, k, t, d, \beta_1) = \frac{e}{2\pi} \frac{1}{\sqrt{\ln \frac{n-t}{n-k-t+\delta+1}}} < 1, \tag{41}$$

as long as $k > 0.16(n - t) + \delta + 1$.

Equations (40) and (41) thus imply

$$\left(1 - \frac{e}{2\pi} \frac{1}{\sqrt{\ln \frac{n-t}{n-k-t+\delta+1}}}\right)\left(1 - \frac{e}{\sqrt{2}\pi} \frac{1}{\sqrt{\ln \frac{t}{\delta+1}}}\right). \tag{42}$$

The last result to demonstrate from this section is Theorem 6.

Proof Taking the simplified interval for β, the existence of this interval implies

$$\frac{\sqrt{2(t+2d)}}{\sqrt{n-k}} \sqrt{\ln \frac{n-t}{n-k-t+\delta+1}} \leq 1 - \frac{\sqrt{2(t+2d-1)}}{\sqrt{n-k}} \sqrt{\ln \frac{t}{\delta+1}} \tag{43}$$

$$\sqrt{\ln \frac{n-t}{n-k-t+\delta+1}} + \sqrt{\ln \frac{t}{\delta+1}} \leq \sqrt{\frac{n-k}{2(t+2d)}} \tag{44}$$

Using the condition on k we deduce

$$\sqrt{\ln \frac{n-t}{n-k-t+\delta+1}} \leq \sqrt{\ln \frac{n-t}{n-(n-t+\delta+1-\frac{(n-t)(\delta+1)}{t})-t+\delta+1}} \tag{45}$$

$$\sqrt{\ln \frac{n-t}{n-k-t+\delta+1}} \leq \sqrt{\ln \frac{t}{\delta+1}}. \tag{46}$$

Hence, the following should hold

$$2\sqrt{\ln \frac{t}{\delta+1}} \leq \sqrt{\frac{n-k}{2(1+c)t}} \tag{47}$$

$$\frac{t}{\delta+1} \ln \frac{t}{\delta+1} \leq \frac{n-k}{8(1+c)(\delta+1)}, \tag{48}$$

which is satisfied as long as $t \leq \frac{n-k}{8(1+c)W(\frac{n-k}{8(1+c)(\delta+1)})}$. The initial condition on k implies

$$t \leq \frac{n-k+2\delta+2-\sqrt{(n-k+2\delta+2)^2-4n(\delta+1)}}{2}$$

which is greater than or equal to $\frac{n-k}{8(1+c)W(\frac{n-k}{8(1+c)(\delta+1)})}$.

As for the asymptotic, use one term approximation for the LambertW function near infinity.

Proof of Remark 1.

$$\frac{\sqrt{t+2d}}{\sqrt{n-k}}\sqrt{W\left(\frac{n-t}{n-k-t+\delta+1}\right)^2} \leq 1-\frac{\sqrt{t+2d-1}}{\sqrt{n-k}}\sqrt{W\left(\frac{t}{\delta+1}\right)^2} \tag{49}$$

$W()$ being increasing we deduce

$$2\frac{\sqrt{t+2d-1}}{\sqrt{n-k}}\sqrt{W\left(\frac{t}{\delta+1}\right)^2} \leq 1 \tag{50}$$

$$\sqrt{W\left(\frac{t}{\delta+1}\right)^2} \leq \frac{1}{2}\sqrt{\frac{n-k}{(1+c)t}} \tag{51}$$

$$W\left(\frac{t}{\delta+1}\right)^2 \leq \frac{n-k}{4(1+c)t} \tag{52}$$

$$yW(y)^2 \leq \frac{n-k}{4(1+c)(\delta+1)}, \tag{53}$$

where $y = t/(\delta+1)$. Letting $N \triangleq \frac{n-k}{4(1+c)(\delta+1)}$, the solution in y to the last equation is $y = \frac{N}{(3W(\frac{N^{\frac{2}{3}}}{3}))}$, which leads to

$$t \leq \frac{n-k}{(12(1+c)W\left(\frac{1}{3}\left(\frac{n-k}{4(1+c)(\delta+1)}\right)^{\frac{2}{3}}\right)}$$

Funding B. Colombier, V. Grosso and PL. Cayrel were funded by a French national grant managed by the Agence Nationale de la Recherche (ANR): project PQ-TLS reference ANR-22-PETQ-0008 through France 2023 program

Declarations

Conflicts of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

1. Adam, N.R., Worthmann, J.C.: Security-control methods for statistical databases: A comparative study. ACM Comput. Surv. **21**(4), 515–556 (1989)
2. Albrecht, M.R., Bernstein, D.J., Chou, T., Cid, C., Gilcher, J., Lange, T., Maram, V., von Maurich, I., Misoczki, R., Niederhagen, R., Paterson, K.G., Persichetti, E., Peters, C., Schwabe, P., Sendrier, N.,

Szefer, J., Tjhai, C.J., Tomlinson, M., Wang, W.: Classic McEliece. Technical report, National Institute of Standards and Technology, available at https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions. (2020)

3. Aragon, N., Barreto, P., Bettaieb, S., Bidoux, L., Blazy, O., Deneuville, J.-C., Gaborit, P., Gueron, S., Guneysu, T., Melchor, C.A., Misoczki, R., Persichetti, E., Sendrier, N., Tillich, J.-P., Zémor, G., Vasseur, V., Ghosh, S.: BIKE. Technical report, National Institute of Standards and Technology, available at https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions. (2020)
4. Augot, D., Finiasz, M., Sendrier, N.: A fast provably secure cryptographic hash function. IACR Cryptol. ePrint Arch. **2003**, 230 (2003)
5. Becker, A., Joux, A., May, A., Meurer, A.: Decoding random binary linear codes in $2^{n/20}$: How $1+1=0$ improves information set decoding. In: Advances in Cryptology - EUROCRYPT 2012, Lecture Notes in Comput. Sci. Springer (2012)
6. Berlekamp, E., McEliece, R., van Tilborg, H.: On the inherent intractability of certain coding problems. IEEE Trans. Inform. Theory **24**(3), 384–386 (1978)
7. Bernstein, D.J., Jeffery, S., Lange, T., Meurer, A.: Quantum algorithms for the subset-sum problem. In: Post-quantum cryptography 2011, vol. 7932 of Lecture Notes in Comput. Sci. Limoges, France, pp. 16–33, (2013)
8. Bernstein, D.J., Lange, T., Peters, C.: Attacking and defending the McEliece cryptosystem. In: Post-Quantum Cryptography 2008, vol. 5299 of Lecture Notes in Comput. Sci. pp. 31–46 (2008)
9. Bernstein, D.J., Lange, T., Peters C.: Smaller decoding exponents: ball-collision decoding. In: Advances in cryptology - CRYPTO 2011, volume 6841 of Lecture Notes in Comput. Sci. pp. 743–760 (2011)
10. Bootle, J., Delaplace, C., Espitau, T., Fouque, P-A., Tibouchi, M.: LWE without modular reduction and improved side-channel attacks against BLISS. In: Peyrin, T., Galbraith, S. (eds.) Advances in cryptology – ASIACRYPT 2018, Cham, pp. 494–524 (2018). Springer International Publishing
11. Both, L., May, A.: Decoding linear codes with high error rate and its impact for LPN security. In: Lange, T., Steinwandt, R. (eds.) Post-Quantum Cryptography - 9th International Conference, PQCrypto 2018, Fort Lauderdale, FL, USA, Proceedings, vol. 10786 of Lecture Notes in Computer Science, Springer, pp. 25–46 (2018). Accessed 9–11 Apr 2018
12. Canteaut, A., Chabaud, F.: A new algorithm for finding minimum-weight words in a linear code: Application to McEliece's cryptosystem and to narrow-sense BCH codes of length 511. IEEE Trans. Inform. Theory **44**(1), 367–378 (1998)
13. Cao, C.-C., Li, C., Sun, X.: Quantitative group testing-based overlapping pool sequencing to identify rare variant carriers. BMC Bioinformatics **15**(195), 1–14 (2014)
14. Carrier, K., Debris-Alazard, T., Meyer-Hilfiger, C., Tillich, J-P.: Statistical decoding 2.0: Reducing decoding to LPN (2022)
15. Cayrel, P.-L., Colombier, B., Drăgoi, V.-F., Menu, A., Bossuet, L.: Message-recovery laser fault injection attack on the classic McEliece cryptosystem. In: Canteaut, A., Standaert, F.-X. (eds.) Advances in cryptology - EUROCRYPT 2021 - 40th annual international conference on the theory and applications of cryptographic techniques, Zagreb, Croatia, Proceedings, Part II, vol. 12697 of Lecture Notes in Computer Science, Springer, pp. 438–467 (2021). Accessed 17–21 Oct 2021
16. Colombier, B., Drăgoi, V.-F., Cayrel, P.-L., Grosso, V.: Message-recovery profiled side-channel attack on the classic McEliece cryptosystem. Cryptology ePrint Archive, Report 2022/125 (2022)
17. Corless, R.M., Gonnet, G.H., Hare, D.E.G., Jeffrey, D.J., Knuth, D.E.: On the LambertW function. Adv. Comput. Math. **5**, 329–359 (1996)
18. Debris-Alazard, T., Tillich, J.-P.: Statistical decoding. In: 2017 IEEE International symposium on information theory (ISIT), pp 1798–1802, (2017)
19. Dinur, I., Nissim, K.: Revealing information while preserving privacy. In: Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS '03, New York, pp. 202–210 (2003). Association for Computing Machinery
20. Drăgoi, V-F., Cayrel, P.-L., Colombier, B., Bucerzan, D., Hoara, S.: Solving a modified syndrome decoding problem using integer programming. Int. J. Comput. Commun. Control **15**(5), (2020)
21. Drăgoi, V.,-F., Colombier, B., Cayrel, P.-L., Grosso, V.: Integer syndrome decoding in the presence of noise. In: 2022 IEEE Information theory workshop (ITW), pp. 482–487 (2022)
22. Dumer, I.: Two decoding algorithms for linear codes. Probl. Inf. Transm. **25**(1), 17–23 (1989)
23. Dumer, I.: On minimum distance decoding of linear codes. In: Proc. 5th Joint Soviet-Swedish Int. Workshop Inform. Theory, Moscow, pp. 50–52, (1991)
24. Esser, A., Bellini, E.: Syndrome decoding estimator. In: Hanaoka, G., Shikata, J., Watanabe, Y. (eds.) IACR International conference on practice and theory of public-key cryptography, vol 13177 of Lecture Notes in Computer Science, pp 112–141, virtual event, March 2022. Springer

25. Feige, U., Lellouche, A.: Quantitative group testing and the rank of random matrices. CoRR, abs/2006.09074, (2020)
26. Finiasz, M., Sendrier, N.: Security bounds for the design of code-based cryptosystems. In: Matsui, M. (ed.) Advances in Cryptology - ASIACRYPT 2009, vol. 5912 of Lecture Notes in Comput. Sci., Springer, pp. 88–105 (2009)
27. Fossorier, M., Kobara, K., Imai, H.: Modeling bit flipping decoding based on nonorthogonal check sums with application to iterative decoding attack of McEliece cryptosystem. IEEE Trans. Inf. Theor. **53**(1), 402–411 (2007)
28. Gaborit, P., Lauradoux, C., Sendrier, N.: SYND: a fast code-based stream cipher with a security reduction. In: 2007 IEEE International symposium on information theory, pp. 186–190 (2007)
29. Grosso, V., Cayrel, P.-L., Colombier, B., Drăgoi, V.-F.: Punctured syndrome decoding problem. In: Kavun, E.B., Pehl, M. (eds.) Constructive side-channel analysis and secure design, Cham, pp. 170–192 (2023). Springer Nature Switzerland
30. Guo, Q., Johansson, T., Mårtensson, E., Wagner, P.S.: Some cryptanalytic and coding-theoretic applications of a soft stern algorithm. Adv. Math. Commun. **13**(4), 559–578 (2019)
31. Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al.: Array programming with numpy. Nature **585**(7825), 357–362 (2020)
32. Johann Heinrich.: Observationes variae in mathesin puram. Acta Helvetica Physico-Mathematico-Anatomico-Bota-nico-Medica 3, 128–168 (1758)
33. Horlemann, A.-L., Puchinger, S., Renner, J., Schamberger, T., Wachter-Zeh, A.: Information-set decoding with hints. In: Wachter-Zeh, A., Bartz, H., Liva, G. (eds.) International workshop on code-based cryptography, vol. 13150 of Lecture Notes in Computer Science, Munich, Germany, Springer, pp. 60–83 (2021)
34. Jabri, A.A.: A statistical decoding algorithm for general linear block codes. In: Honary, B. (ed) Cryptography and Coding, Berlin, Heidelberg, pp. 1–8 (2001). Springer Berlin Heidelberg
35. Klar, B.: Bounds on tail probabilities of discrete distributions. Probab. Eng. Inform. Sci. **14**, 161–171 (2000)
36. Lee, P.J., Brickell, E.F.: An observation on the security of McEliece's public-key cryptosystem. In: Advances in cryptology - EUROCRYPT'88 vol. 330 of Lecture Notes in Comput. Sci. Springer, pp. 275–280 (1988)
37. Leon, J.: A probabilistic algorithm for computing minimum weights of large error-correcting codes. IEEE Trans. Inform. Theory **34**(5), 1354–1359 (1988)
38. Martins, J.P., Santos, R., Sousa, R.: Testing the Maximum by the Mean in Quantitative Group Tests, Publishing, Cham, pp. 55–63 (2014). Springer International
39. May, A., Meurer, A., Thomae, E.: Decoding random linear codes in $O(2^{0.054n})$. In: Lee, D.H., Wang, X. (eds.) Advances in Cryptology - ASIACRYPT 2011, vol. 7073 of Lecture Notes in Comput. Sci., Springer, pp. 107–124 (2011)
40. May, A., Ozerov, I.: On computing nearest neighbors with applications to decoding of binary linear codes. In: Oswald, E., Fischlin, M., (eds.) Advances in Cryptology - EUROCRYPT 2015, vol. 9056 of Lecture Notes in Comput. Sci., Springer, pp. 203–228 (2015)
41. Niebuhr, R.: Statistical decoding of codes over $\mathbf{F}_q$. In: Post-Quantum Cryptography 2011, vol. 7071 of Lecture Notes in Comput. Sci. Springer, pp. 217–227 (2011)
42. Niederreiter, H.: Knapsack-type cryptosystems and algebraic coding theory. Problems of Control and Information Theory **15**(2), 159–166 (1986)
43. Overbeck, R.: A new structural attack for GPT and variants. In: Mycrypt, vol. 3715 of Lecture Notes in Comput. Sci. pp. 50–63 (2005)
44. Overbeck, R.: Statistical decoding revisited. In: Batten, L., Safavi-Naini, R. (eds.) Information security and privacy?: 11^{th} Australasian conference, ACISP 2006. Lecture Notes in Comput, vol. 4058, pp. 283–294. Springer, Sci. (2006)
45. Prange, E.: Cyclic error-correcting codes in two symbols. Electronics Research Directorate, Air Force Cambridge Research Center, (1957). No. AFCRC-TN-57-103. ASTIA Document No. AD133749
46. Prange, E.: The use of information sets in decoding cyclic codes. IRE Transactions on Information Theory **8**(5), 5–9 (1962)
47. Shor, P.W.: Algorithms for quantum computation: Discrete logarithms and factoring. In: Goldwasser, S. (ed.) FOCS, pp. 124–134 (1994)
48. Shrestha, S.R., Kim, Y.S.: New McEliece cryptosystem based on polar codes as a candidate for post-quantum cryptography. In: 2014 14th International symposium on communications and information technologies (ISCIT), IEEE, pp. 368–372 (2014)

49. Sidelnikov, V.M.: A public-key cryptosytem based on Reed-Muller codes. Discrete Math. Appl. **4**(3), 191–207 (1994)
50. Stern, J.: A method for finding codewords of small weight. In: Cohen, G.D., Wolfmann, J. (eds) Coding Theory and Applications, vol. 388 of Lecture Notes in Comput. Sci. Springer, pp. 106–113 (1988)
51. Stern, J.: A new identification scheme based on syndrome decoding. In: Stinson, D.R. (ed) Advances in cryptology - CRYPTO '93, 13th annual international cryptology conference, Santa Barbara, California, USA, August 22-26, 1993, Proceedings, vol. 773 of Lecture Notes in Comput. Sci. Springer, pp. 13–21 (1993)
52. Tiplea, F.L., Dragoi, V.F.: Generalized inverse based decoding. In: IEEE International Symposium on Information Theory, ISIT 2022, Espoo, Finland, June 26 - July 1, 2022, IEEE, pp. 2791–2796 (2022)
53. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J. et al.: Scipy 1.0: fundamental algorithms for scientific computing in python. Nature methods, **17**(3), 261–272 (2020)
54. Wang, C., Zhao, Q., Chuah, C.N.: Group testing under sum observations for heavy hitter detection. In: 2015 Information theory and applications workshop (ITA), pp. 149–153 (2015)
55. Wang, I.J., Huang, S.L., Lee, K.Y., Chen, K.C.: Data extraction via histogram and arithmetic mean queries: Fundamental limits and algorithms. In: 2016 IEEE International symposium on information theory (ISIT), pp. 1386–1390 (2016)
56. Winters, R.: Practical Predictive Analytics. Packt Publishing, Birmingham, England (2017)