Energy-limited Massive Random Access via Noisy Group Testing

Huseyin A. Inan, Peter Kairouz and Ayfer Ozgur
Stanford University
Email: {hinan1, kairouzp, aozgur}@stanford.edu

***Abstract*—We consider a random access scheme with a massive number of low-energy wireless devices, where a small but arbitrary subset of them can be active at a given time. We develop a solution to this problem via the noisy group testing framework, where the goal is to identify with high probability a small set of defectives in a large set of items with minimum number of noisy binary tests. We translate the power constraint for the wireless devices to a constraint on the number of tests each item can participate in the group testing problem. We present fundamental upper and lower bounds on the length of the codewords in terms of the total number of devices, the number of active devices, the error probability, the noise and the power constraint. We conclude with a scheme that enjoys low decoding complexity under this model.**

I. Introduction

The number of devices that are connected to the Internet has recently surpassed the global human population, and is projected to hit trillions within the next two decades [1]. This unprecedented growth in connectivity is fueled by the emergence of low-energy, low-cost wireless devices that combine sensing, computation, and communication. While these devices are expected to enable a plethora of smart technologies, they present a number of new challenges to communication engineers. Our work addresses one of them: *how do we allow a massive number of sporadically active low-energy wireless devices with small payloads to access the spectrum with minimal coordination and channel estimation overheads*?

In this context, energy efficiency, as opposed to spectral efficiency, becomes important because the devices are often expected to be powered by tiny batteries or energy harvesting technologies. Existing wireless access technologies are designed for a very different regime: high data rates, large payloads, and much less stringent constraints on-device processing power and energy. They thus seem fundamentally inadequate for delivering the needed energy efficiency and scalability in IoT-type settings.

In this paper, we consider a setting where there are n devices that are connected to a single central node but only d of them are active within a communication round, where d and n can be both large but typically $d \ll n$. An active device transmits a binary codeword of length t and the central node detects whether or not there is energy in the channel in each time-slot. Since multiple devices can be active at the same time, the aforementioned energy-detection model leads to a noisy Boolean OR-channel; i.e., the received signal is the Boolean OR of the transmitted binary codewords with occasional bit flips. The bit flips are due to channel and circuit noise. The goal of the central node is to recover the (up to d) simultaneously transmitted codewords from their noisy Boolean OR. In [2], we argue that this simple model can be used to study a number of seemingly different problems in a single framework: device discovery (communicating active device identities), joint discovery and data transmission, and data transmission with non-identifiable users. The simple one-off modulation scheme combined with energy detection converts the familiar linear Gaussian channel model to a noisy Boolean OR model. While this can be information-theoretically suboptimal, this simple scheme can be preferable in practice since it eliminates the need for channel state information at the transmitter and/or receiver. This in turn eliminates the need for channel training and estimation, which can be especially desirable when data payloads are small and devices are only sporadically active.

The problem statement above corresponds to the noisy *non-adaptive group testing* (NAGT) problem. Group testing consists of identifying a small set of d (or less) defective items from a large population of size n by performing tests on groups of items, rather than on individual items. For the noisy NAGT problem, it has been shown that $t = \theta(d \log n)$ is necessary and sufficient when the error and noise parameters (ϵ and q, respectively) are treated as fixed constants [3], [4]. These constructions perform $\theta(\log n)$ tests on average on each item. In the multiple access setting, the number of tests per item corresponds to the number of 1's in the corresponding codeword and is therefore proportional to the energy required to transmit the codeword. Therefore the average transmission power constraint for a wireless device can be translated to a constraint on the number of tests an item participates in. Motivated by this observation, we tackle the following question: how small can t be if we place a Hamming weight constraint on the codewords? More specifically, for fixed d, n, ϵ, q, and ρ, how small can t be if each codeword can have at most ρt 1's, where ρ corresponds to the power constraint at the transmitter?

We first note from the aforementioned results that when $\rho = \theta(1/d)$, $t = \theta(\log n/\rho)$ is necessary and sufficient so we are in general interested in the region where $\rho = O(1/d)$. It can be easily seen that when $\rho = o(1/d)$, $t = \theta(\log n/\rho)$

This work was supported in part by NSF Grant #1514538 and the Stanford SystemX Alliance.

can be trivially achieved by simply taking the construction designed for $\rho = \theta(1/d)$ and adding extra rows of zeros, until we get $t = \theta(\log n/\rho)$ in which case the weight constraint ρt per codeword is satisfied. Note that with this approach the weight of the codewords remains the same, $\theta(\log n)$, but the density ρ is decreased to the desired level by increasing t. However, intuitively one would hope to do better by utilizing the sparsity of the codewords and potentially reduce the weights of the codewords below $\theta(\log n)$ and their length below $t = \theta(\log n/\rho)$. Our main result is to show that we can do better than this trivial achievability by providing upper and lower bounds on the number of tests in terms of the system parameters $(d, n, \epsilon, q, \rho)$.

The remainder of our paper is organized as follows. We start by providing a precise statement of the problem in Section II. We present our main results and discuss their implications in Section III. We then investigate efficiently-decodable, low-energy group testing strategies in Section IV. We conclude with a few remarks and non-trivial extensions in Section V.

A. Related work

The literature on the NAGT framework includes both deterministic and random test designs [5]. Achievability and lower bound results on the number of tests are presented for exact identification of defective sets with both zero-error guarantees [6]–[9] and ϵ-error probability guarantees [4], [10], [11]. The main focus has been on minimizing the number of tests without any constraint on the number of tests each item can participate in. Very recently, [12] and [2] considered the model where the number of ones in columns and rows of the test design matrix are constrained by an independent parameter that does not scale with the codeword length t. They obtained corresponding upper and lower bounds on the number of tests in the noiseless setting. In this work, we focus on a more practical noisy model with power constraints, and aim to characterize the dependence of the number of tests on system parameters.

Our channel model is closely related to the one studied in [13]. However, the authors of [13] allow for a partial recovery of the set of active items and do not place constraints on the columns of the testing matrix. Our approach is also related to the recent work of [14]. Similar to [14], we investigate low-energy massive access schemes. However, [14] assumes that the signals are all received at the same power level and thus considers the linear additive Gaussian noise channel.

II. System Model

We consider a setting where n devices are connected to a central node but only $d = o(n)$ of them are active in a given communication round. Let the non-zero entries of $X \in \{0,1\}^n$ represent the set of active devices which are uniformly distributed among the n devices. Therefore, X is uniformly distributed over the set of $\binom{n}{d}$ possible d-sparse vectors in $\{0,1\}^n$. The problem formulation is as follows. Given n devices, design a length t binary signature with at most ρt number of ones for each device (i.e., $M_i \in \{0,1\}^t$ s.t. $|M_i| \leq \rho t$ for $i = 1, \ldots, n$) with a decoding procedure such that we can estimate the vector $\hat{X} \in \{0,1\}^n$ from the noisy measurements

$$\hat{Y} = Y \oplus v = \bigvee_{i:X_i=1} M_i \oplus v.$$

with ϵ average error probability, i.e., $\Pr[\hat{X} \neq X] \leq \epsilon$, where $\bigvee$ denotes the Boolean-OR of the columns, v is the noise vector with i.i.d. Bernoulli(q) entries, Y is the noiseless measurement vector and $\oplus$ represents modulo-2 addition that reflects the unknown bit flips.

III. Main Results

In this section, we formally present our results and discuss their implications. The detailed proofs are deferred to the Appendix. We begin with a lower bound on the number of tests required by our model. The lower bound is based on an information-theoretic approach (a similar approach is used in [4], [15] for the case without energy constraints).

Theorem 1: Any group-testing algorithm achieving a probability of error of at most ϵ where each item is included in at most ρt tests for $\rho \leq 1/d$ requires at least

$$t \geq \frac{(1-\epsilon)d\log(n/d) - 1}{\rho d \log(e/\rho d)}$$

number of tests (regardless of the value of q and therefore including $q = 0$).

Recall that without any constraints on the number of tests each item can participate in, prior work [3], [4], [10] shows that $t = \theta(d \log n)$ is achievable with $w = \theta(\log n)$ weight for each codeword. This corresponds to $\rho = \theta(1/d)$, or in other words $\rho = \theta(1/d)$ is sufficient to achieve the minimal t. Note that a straightforward scheme for any $\rho = o(1/d)$ can be obtained by simply taking the matrix M designed for $\rho = \theta(1/d)$ and adding extra rows of zeros, until we get $t = \theta(\log n/\rho)$ in which case the weight constraint ρt per codeword is satisfied. Therefore, the scaling $t = \theta(\log n/\rho)$ can be trivially achieved. However, we observe from the lower bound in Theorem 1 that one can hope to do better. The lower bound in Theorem 1 states that there is a potential gain of $\log(1/\rho d)$ which can be significant in the regime where $\rho = o(1/d)$. For instance, when $\rho = \theta(1/d^2)$, we can potentially gain a factor of $\log(d)$ over the trivial approach. Therefore, the natural question is how tight this lower bound is. Regarding this question, we consider two extreme cases. We first show that the lower bound is tight in the noiseless case, i.e., when $q = 0$ and then we consider the noisy case when q is a fixed constant. Our next result shows that the lower bound in Theorem 1 is tight under the noiseless model.

Theorem 2: Let n be the size of the population of items for which there is a defective set of cardinality d. Constructing a $t \times n$ matrix M by choosing each column i.i.d. uniformly over the vectors in $\{0,1\}^t$ with weight ρt where $\rho < 1/d$ ensures the existence of a group testing strategy that achieves an average probability of error of at most ϵ under the noiseless model provided that

$$t \geq \frac{\log(n/\epsilon)}{\rho \log(1/e\rho d)}.$$

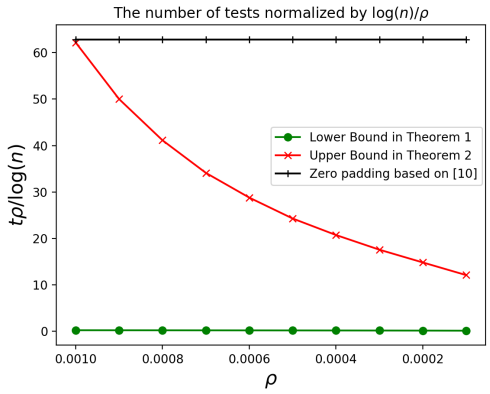

Fig. 1. The number of tests scaled with $\log(n)/\rho$ as a function of ρ. The model parameters are chosen as $n = 2^{30}$, $d = 100$, $\epsilon = 0.01$, $q = 0.1$.

We observe from Theorem 1 and Theorem 2 that $t = \theta\left(\frac{\log(n)}{\rho \log(1/\rho d)}\right)$ is optimal under the sparsity constraint ρ in the noiseless case. We next consider the noisy case where each test outcome flips with probability q where q is a fixed constant.

Theorem 3: Let n be the size of the population of items for which there is a defective set of cardinality d. Constructing a $t \times n$ matrix M by choosing each column i.i.d. uniformly over the vectors in $\{0,1\}^t$ with weight ρt where $\rho < 1/(2ed)$ ensures the existence of a group testing strategy that achieves an average probability of error of at most ϵ provided that

$$t > \max\left(\frac{\log 2/\epsilon + \log d}{2q^2\delta^2\rho}, \frac{\log 4/\epsilon + \log(n-d)}{2(1-q(1+\delta)-q-\gamma(1-2q))^2\rho}, \frac{\log 4/\epsilon + \log(n-d)}{\log(1/2e(d\rho)^\gamma)\rho}\right),$$

where q is the fixed bit flip probability constant and δ and γ are free parameters to be chosen such that $0 < \delta < \frac{(1-2q)(1-\gamma)}{q}$ and $2e(d\rho)^\gamma < 1$.

Note that the dependence of t on the constraint ρ in Theorem 3 is complicated as ρ also implicitly impacts the optimal choice of the free parameters δ and γ. In order to illustrate the dependence of t on ρ, and in particular how our lower bound improves over the trivially achievable scheme with zero padding, in Figure 1 we plot the lower and upper bounds on t in Theorem 1 and 3 by normalizing these bounds by $\log(n)/\rho$. We see that as ρ decreases the normalized t in our upper bound also decreases, meaning that t in our achievable scheme does not increase as fast as $\log(n)/\rho$ with decreasing ρ. As a reference, we also provide a trivial scheme obtained by zero padding, i.e. adding additional zero rows to the design matrix M of the scheme developed in [10] so that it satisfies a chosen ρ constraint. Note that the scheme in [10] was developed without constraints on the number of tests per item.

Next we will discuss about decoding. We note that the decoding procedure we use in the proof of Theorem 3 has complexity of $O(tn)$ which is the complexity of "cover decoder" type implementations in the literature. However, for the envisioned IoT applications, due to the massive number of nodes, this is not quite practical. In the next section, we will address this issue by a simple modification of a recent work [16] where the authors introduce a construction based on left regular bipartite graphs and utilizing constant-rate expander code [17] with a low complexity decoding procedure.

IV. Decoding

There is a recent research effort towards practical decoding schemes while preserving the order of t as much as possible. For the statistical group testing problem, in [18] the requirement of recovering all the defective items with high probability achieves $O(d \log d \log n)$ number of tests and time complexity which has an additional $\log d$ factor. Another recent result is [16] where the model is the same as we consider in this paper without the energy constraints and the introduced algorithm requires $O(d \log d \log n)$ number of tests with $O(d(\log^2 d + \log n))$ decoding complexity. Note that there is an additional $\log d$ factor in the required number of tests, however, the decoding complexity reduces to $O(d \log n)$ when $d = O(poly(\log n))$ which is order optimal. There is no result yet that achieves both $O(d \log n)$ decoding complexity and number of tests at the same time to the best of our knowledge.

Our next result is a simple modification of [16] that gets the hit of additional $\log d$ factor in the number of tests, however, enjoys a much better decoding complexity hence the trade-off would be worthy in practical applications.

Theorem 4: Let n be the size of the population of items for which there is a defective set of cardinality d. There exists a group testing strategy with the number of tests $t = O\left(\frac{1}{\rho} \log d \log n\right)$ and decoding complexity $O\left(\frac{1}{\rho} \log d \log \frac{1}{\rho} + d \log n\right)$ that achieves a probability of error of at most ϵ and each column weight is bounded with ρt.

We skip the proof here due to space constraints. We note that there is an additional factor of $\log d$ in the number of tests, however, it comes with a great reduction in the decoding complexity of the model. We can observe in general that in the region where $\rho = \theta(1/d^\alpha)$ for some constant $\alpha \geq 1$ and $d = O(poly(\log n))$, the decoding complexity is $O(poly(\log n))$ which is substantially more efficient compared to the "cover decoder" type algorithm we utilized in the previous section.

V. Conclusion

We studied the noisy group testing setting with a weight constraint on the columns of the group testing design matrix. We proved lower and upper bounds on the minimum number of tests. We also presented a low decoding complexity strategy with nearly optimal codeword length. Many questions remain to be addressed including tightening the gap between the upper and lower bounds, and providing efficiently decodable strategies with optimal codeword length.

References

[1] P. Sparksr, "White paper: The economics of a trillion connected devices," 2017.

[2] H. A. Inan, P. Kairouz, and A. Ozgur, "Sparse Combinatorial Group Testing for Low-Energy Massive Random Access," *ArXiv e-prints*, 2017.

[3] D. Sejdinovic and O. Johnson, "Note on noisy group testing: Asymptotic bounds and belief propagation reconstruction," *CoRR*, vol. abs/1010.2441, 2010.
[4] G. K. Atia and V. Saligrama, "Boolean compressed sensing and noisy group testing," *Information Theory, IEEE Transactions on*, vol. 58, no. 3, pp. 1880–1901, 2012.
[5] D.-Z. Du and F. K. Hwang, *Combinatorial group testing and its applications*, vol. 12, World Scientific, 2000.
[6] A. G. D'yachkov and V. V. Rykov, "Bounds on the length of disjunctive codes," *Problemy Peredachi Informatsii*, vol. 18, no. 3, pp. 7–13, 1982.
[7] W. Kautz and R. Singleton, "Nonrandom binary superimposed codes," *IEEE Transactions on Information Theory*, vol. 10, no. 4, pp. 363–377, October 1964.
[8] E. Porat and A. Rothschild, "Explicit non-adaptive combinatorial group testing schemes," *Automata, Languages and Programming*, pp. 748–759, 2008.
[9] P. Indyk, H. Q. Ngo, and A. Rudra, "Efficiently decodable non-adaptive group testing," in *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2010.
[10] C. L. Chan, S. Jaggi, V. Saligrama, and S. Agnihotri, "Non-adaptive group testing: Explicit bounds and novel algorithms," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 3019–3035, May 2014.
[11] Jonathan Scarlett and Volkan Cevher, "Phase Transitions in Group Testing," in *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2016.
[12] V. Gandikota, E. Grigorescu, S. Jaggi, and S. Zhou, "Nearly optimal sparse group testing," in *Allerton*, 2016, pp. 401–408.
[13] J. Luo and D. Guo, "Neighbor discovery in wireless ad hoc networks based on group testing," in *Allerton*, 2008, pp. 791–797.
[14] Y. Polyanskiy, "A perspective on massive random-access," in *ISIT*, June 2017, pp. 2523–2527.
[15] Arkadii G. D'yachkov, "Lectures on designing screening experiments," *CoRR*, vol. abs/1401.7505, 2014.
[16] S. Cai, M. Jahangoshahi, M. Bakshi, and S. Jaggi, "Efficient algorithms for noisy group testing," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2113–2136, April 2017.
[17] D. A. Spielman, "Linear-time encodable and decodable error-correcting codes," *IEEE Transactions on Information Theory*, vol. 42, no. 6, pp. 1723–1731, Nov 1996.
[18] K. Lee, R. Pedarsani, and K. Ramchandran, "Saffron: A fast, efficient, and robust framework for group testing based on sparse-graph codes," in *ISIT*, 2016, pp. 2873–2877.

APPENDIX A
THE PROOF OF THEOREM 1

We have the power constraint over all codewords, i.e., $w_i \leq \rho t$ for all $i \in [n]$. We consider the case $\rho \leq 1/d$.

We note that $X \to Y \to \hat{Y} \to \hat{X}$ forms a Markov chain. By standard information-theoretic definitions we have

$$H(X) = H(X|\hat{X}) + I(X;\hat{X}).$$

Since X is uniformly distributed over $\binom{n}{d}$ defective sets, $H(X) = \log\binom{n}{d}$. By Fano's inequality, $H(X|\hat{X}) \leq 1 + \epsilon \log\binom{n}{d}$. Also, we have $I(X;\hat{X}) \leq I(Y;\hat{Y})$ by the data processing inequality. We further have $I(Y;\hat{Y}) = H(Y) - H(Y|\hat{Y}) \leq H(Y)$. Note that Y can have at most $\rho t d \leq t$ ones since there are d defective items and each one has at most ρt number of ones, therefore, we have

$$H(Y) \leq \log\left(\binom{t}{0} + \ldots + \binom{t}{\rho t d}\right) \leq \rho t d \log\left(\frac{e}{\rho d}\right).$$

where the last inequality is due to $\sum_{i=0}^{k}\binom{n}{i} \leq (en/k)^k$. Combining the above inequalities, we obtain

$$(1-\epsilon)\log\binom{n}{d} \leq 1 + \rho t d \log\left(\frac{e}{\rho d}\right).$$

We also note that $\log\binom{n}{d} \geq d\log(n/d)$. Therefore

$$t \geq \frac{(1-\epsilon)d\log(n/d) - 1}{\rho d \log(e/\rho d)}.$$

APPENDIX B
THE PROOF OF THEOREM 2

We consider the region where $\rho < 1/d$. We choose the columns of M i.i.d. uniformly over the vectors in $\{0,1\}^t$ with weight ρt. Note that this set has cardinality of $\binom{t}{\rho t}$ and all the columns will have the same weight $w = \rho t$. Let P_e denote the average probability of error, averaged over all codebooks M and over all defective sets of size d. By the symmetry of the codebook construction, the average error probability does not depend on the defective set, hence we can assume without loss of generality that the defective set is the first d items.

We apply the cover decoder which identifies an item as defective if it is covered by the noiseless test outcome vector Y. Note that an error happens when a non-defective item is covered by Y, i.e., the union of the defective items. Since the union of the defective items can at most include $d\rho t \leq t$ number of ones, this probability is bounded by $\binom{d\rho t}{\rho t}/\binom{t}{\rho t}$. We apply the union bound over the non-defective items and obtain

$$P_e \leq n\frac{\binom{d\rho t}{\rho t}}{\binom{t}{\rho t}} \leq n(ed\rho)^{\rho t},$$

where the last inequality is due to $(n/k)^k \leq \binom{n}{k} \leq (en/k)^k$. We conclude the proof by bounding the last term by ϵ.

APPENDIX C
THE PROOF OF THEOREM 3

We focus on the region where $\rho < 1/(2ed)$. We can similarly assume without loss of generality that the defective set is the first d items, i.e., $S = \{1, 2, \ldots, d\}$.

We next describe the decoding procedure. For a defective item with weight w, we observe that the corresponding w entries in the noiseless outcome will be all ones. In expectation, wq of these entries will be flipped to zeros due to the noise. Therefore, it is natural to apply the following decoding rule. For an item $j \in [n]$, let w_j be the number of non-zero entries of an item (weight) and $\tilde{w}_j$ be the number of non-zero entries of an item for which the corresponding entry of $\hat{Y}$ is also non-zero. The decoder declares the item j as defective if $\tilde{w}_j \geq w_j(1 - q(1+\delta))$ for some δ that we will specify later, else the decoder declares the item j as non-defective. Note that for $q = 0$, the decoding rule reduces to the "cover decoder". We further note that this decoding rule has been considered in the work of [10] (the **No-CoMa** algorithm).

We note that decoding error happens either by declaring an item in $\{1, 2, \ldots, d\}$ as non-defective or declaring an item in $\{d+1, \ldots, n\}$ as defective. Using union bound, we have the following

$$P_e \leq d\sum_{M} \Pr(M)\Pr(\tilde{w}_1 < w(1-q(1+\delta))|M) + (n-d)\sum_{M}\Pr(M)\Pr(\tilde{w}_{d+1} \geq w(1-q(1+\delta))|M).$$

We denote the first term as P_e^- and second term as P_e^+. Note that the term $\Pr(\tilde{w}_1 < w(1-q(1+\delta))|M)$ does not depend on the randomness of construction, it only depends on the randomness of noise because a defective item has w ones and in the noiseless case this will give w ones in the corresponding positions of the outcomes so the number of matching non-zero entries when the noise is added only depends on the number of bit flips due to error. Therefore, we bound the first term as

$$\begin{aligned} P_e^- &\leq d \ \Pr(\text{Number of bit flips} > wq(1+\delta)) \\ &\overset{(i)}{\leq} d\exp(-2wq^2\delta^2) \end{aligned}$$

where (i) is due to Hoeffding's Inequality. Therefore, $P_e^- \leq \epsilon/2$ when $w \geq \frac{\log 2/\epsilon + \log d}{2q^2\delta^2}$.

We continue with the second term. Consider the column $d+1$ which is non-defective. We introduce the event $\mathcal{A}_\alpha$ as the event of column $d+1$ being covered at exactly α positions by the defective items for $0 \leq \alpha \leq w$. We then have

$$\begin{aligned} P_e^+ = (n-d)\sum_M \sum_{\alpha=0}^{w} &\Pr(M) \\ &\cdot \Pr(\tilde{w}_{d+1} \geq w(1-q(1+\delta))|M, \mathcal{A}_\alpha)\Pr(\mathcal{A}_\alpha|M). \end{aligned}$$

We note that $\Pr(\tilde{w}_{d+1} \geq w(1-q(1+\delta))|M, \mathcal{A}_\alpha)$ only depends on noise since a non-defective item has w ones and given the event $\mathcal{A}_\alpha$ exactly α positions will be one and $w-\alpha$ positions will be zero correspondingly in the noiseless test outcomes. Hence, the number of matching non-zero entries when the noise is added only depends on the number of bit flips due to error. Therefore, $\tilde{w}_{d+1}$ is the number of no bit-flips among the covered α positions in addition to the number of bit-flips among the uncovered $w-\alpha$ positions. We define the event $\mathcal{B}_\alpha$ as the event of $\tilde{w}_{d+1}$ exceeding the threshold $w(1-q(1+\delta))$, i.e., the event of having $\tilde{w}_{d+1} \geq w(1-q(1+\delta))$. We have

$$P_e^+ = (n-d)\sum_{\alpha=0}^{w} \Pr(\mathcal{B}_\alpha) \sum_M \Pr(M)\Pr(\mathcal{A}_\alpha|M).$$

For any $\gamma \in [0,1]$, we can write this as

$$\begin{aligned} P_e^+ = (n-d)\sum_{\alpha=0}^{\gamma w} &\Pr(\mathcal{B}_\alpha) \sum_M \Pr(M)\Pr(\mathcal{A}_\alpha|M) \\ &+ (n-d)\sum_{\alpha=\gamma w+1}^{w} \Pr(\mathcal{B}_\alpha) \sum_M \Pr(M)\Pr(\mathcal{A}_\alpha|M). \end{aligned}$$

We denote the first term as P_1 and the second terms as P_2 and we begin with the analysis of P_1. Given that there are exactly α non-zero positions and $w-\alpha$ zero positions in the noiseless outcomes, the expected number of non-zero positions with the noise will be $\alpha(1-q)+(w-\alpha)q$. Note that since $q < 0.5$, we have $\alpha(1-q)+(w-\alpha)q = wq+\alpha(1-2q) \leq wq+w\gamma(1-2q)$ for $\alpha \in [0, \gamma w]$. We then have

$$\begin{aligned} &\Pr(\mathcal{B}_\alpha) = \Pr(\tilde{w}_{d+1} \geq w(1-q(1+\delta))) \\ &= \Pr(\tilde{w}_{d+1} - \alpha(1-q) - (w-\alpha)q \geq \\ &\qquad w(1-q(1+\delta)) - \alpha(1-q) - (w-\alpha)q) \\ &\leq \Pr(\tilde{w}_{d+1} - \alpha(1-q) - (w-\alpha)q \geq \\ &\qquad w(1-q(1+\delta)) - wq - w\gamma(1-2q)) \\ &\leq \exp(-2w(1-q(1+\delta)-q-\gamma(1-2q))^2) \end{aligned}$$

where the last step is by using Hoeffding's inequality assuming that $1-q(1+\delta)-q-\gamma(1-2q) > 0$ which is satisfied when $\delta < (1-2q)(1-\gamma)/q$. We can then bound the first term as

$$\begin{aligned} P_1 &\leq (n-d)\sum_{\alpha=0}^{\gamma w} \exp(-2w(1-q(1+\delta)-q-\gamma(1-2q))^2) \\ &\qquad \cdot \sum_M \Pr(M)\Pr(\mathcal{A}_\alpha|M) \\ &= (n-d)\exp(-2w(1-q(1+\delta)-q-\gamma(1-2q))^2) \\ &\qquad \cdot \sum_{\alpha=0}^{\gamma w}\sum_M \Pr(M)\Pr(\mathcal{A}_\alpha|M) \\ &\leq (n-d)\exp(-2w(1-q(1+\delta)-q-\gamma(1-2q))^2). \end{aligned}$$

Hence we have $P_1 \leq \epsilon/4$ when $w \geq \frac{\log 4/\epsilon + \log(n-d)}{2(1-q(1+\delta)-q-\gamma(1-2q))^2}$.

We continue with the second term P_2. We have

$$\begin{aligned} P_2 &= (n-d)\sum_{\alpha=\gamma w+1}^{w} \Pr(\mathcal{B}_\alpha)\sum_M \Pr(M)\Pr(\mathcal{A}_\alpha|M) \\ &\leq (n-d)\sum_{\alpha=\gamma w}^{w}\sum_M \Pr(M)\Pr(\mathcal{A}_\alpha|M). \end{aligned}$$

Note that the last term is the probability of having column $d+1$ being covered by at least γw positions by the defective set. We can upper bound this term as follows. Fix any γw positions among the position of ones in the defective set. The rest of the ones can be chosen $\binom{t-\gamma w}{w-\gamma w}$ different ways. Doing this for all such γw positions, we note that we would be counting some codewords multiple times. Since the number of ones is at most dw in the union of defective items, we have

$$\begin{aligned} P_2 &\leq (n-d)\frac{\binom{dw}{\gamma w}\binom{t-\gamma w}{w-\gamma w}}{\binom{t}{w}} \\ &\leq (n-d)\frac{\binom{dw}{\gamma w}\binom{t}{w-\gamma w}}{\binom{t}{w}} \\ &\overset{(i)}{\leq} (n-d)\frac{(ed/\gamma)^{\gamma w}\left(e/\rho(1-\gamma)\right)^{w-\gamma w}}{(1/\rho)^w} \\ &\overset{(ii)}{\leq} (n-d)[2e(d\rho)^\gamma]^w \\ &= \exp(\log(n-d) - w\log(1/2e(d\rho)^\gamma)), \end{aligned}$$

where (i) is due to $(\frac{n}{k})^k \leq \binom{n}{k} \leq (\frac{en}{k})^k$, and (ii) is from $(1/\gamma)^\gamma (1/(1-\gamma))^{1-\gamma} \leq 2$. We can choose γ such that $2e(d\rho)^\gamma < 1$ which is possible for $\rho < 1/2ed$. Hence we have $P_2 \leq \epsilon/4$ when $w \geq \frac{\log 4/\epsilon + \log(n-d)}{\log(1/2e(d\rho)^\gamma)}$. Therefore, the desired error probability ϵ is achieved when

$$\begin{aligned} w \geq \max\Bigg(&\frac{\log 2/\epsilon + \log d}{2q^2\delta^2}, \\ &\frac{\log 4/\epsilon + \log(n-d)}{2(1-q(1+\delta)-q-\gamma(1-2q))^2}, \frac{\log 4/\epsilon + \log(n-d)}{\log(1/2e(d\rho)^\gamma)}\Bigg) \end{aligned}$$

where δ and γ are free parameters to be chosen such that $0 < \delta < (1-2q)(1-\gamma)/q$ and $2e(d\rho)^\gamma < 1$.