# Information-Theoretic Characterization of Sparse Recovery

**Supplementary Notes**

We use lower-case $p(Y|X_S)$ notation for the conditional outcome distribution given the true subset of variables averaged over the latent variable $\beta_S$. In some cases when we would like to distinguish between the outcome distribution conditioned on different sets of variables we use $p_\omega(\,\cdot\,|\,\cdot\,)$ notation, to emphasize that the conditional distribution is conditioned on the given variables, assuming the true set $S$ is $S_\omega$. W.l.o.g. we assume the true set is $S_1$ for below proofs. Define $\mathcal{I} = \{1, \ldots, \binom{D}{K}\}$ as the collection of sets $\omega$ of size $K$.

## 1 Proof of Theorem 2.1

First, note that $P(E) \leq \sum_{i=1}^{K} P(E_i)$, for $E$ and $E_i$ as defined. If we show separately for each $i$ and any $0 \leq \delta \leq 1$ that the following bound holds, then the theorem follows:

$$P(E_i) \leq 2^{-N\left(E_o(\delta) - \delta \frac{\log \binom{D-K}{i}\binom{K}{i}}{N}\right)}. \tag{1}$$

Instead of the above bound, we prove a slightly weaker bound for expositional clarity, which is

$$P(E_i) \leq 2^{-N\left(E_o(\delta) - \frac{\log \binom{D-K}{i}\binom{K}{i}}{N}\right)}. \tag{2}$$

Note that the main difference between the above equation and the previous bound is the missing $\delta$ term multiplying the binomial expression. The main result follows along the same lines and we refer the reader to [1] for further details.

To prove this result we denote by $\mathcal{A}_i$ the set of indices corresponding to sets of $K$ variables that differ from the true set $S_1$ in exactly $i$ variables, i.e.,

$$\mathcal{A}_i = \{\omega \in \mathcal{I} : |S_{1^c,\omega}| = i, |S_\omega| = K\} \tag{3}$$

We can establish that,

$$\begin{aligned}
\Pr[E_i|\omega_0 = 1, X^N_{S_1}, Y^N, \theta] &\leq \sum_{\omega \in \mathcal{A}_i} \sum_{X^N_{S_{1^c,\omega}}} P(X^N_{S_{1^c,\omega}}|\theta) \frac{p_\omega(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1^c,\omega}})^s}{p_1(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1,\omega^c}})^s} \qquad (4) \\
&= \sum_{S_{1,\omega}} \sum_{S_{1^c,\omega}} \sum_{X^N_{S_{1^c,\omega}}} P(X^N_{S_{1^c,\omega}}|\theta) \frac{p_\omega(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1^c,\omega}})^s}{p_1(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1,\omega^c}})^s}.
\end{aligned}$$

Inequality (4) is established separately in the following section. It follows that,

$$\begin{aligned}
\Pr[E_i|\omega_0 = 1, X^N_{S_1}, Y^N, \theta] &\leq \left( \sum_{S_{1,\omega}} \sum_{S_{1^c,\omega}} \sum_{X^N_{S_{1^c,\omega}}} P(X^N_{S_{1^c,\omega}}|\theta) \frac{p_\omega(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1^c,\omega}})^s}{p_1(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1,\omega^c}})^s} \right)^\delta && (5)\\
&\leq \left( \sum_{S_{1,\omega}} \binom{D-K}{i} \sum_{X^N_{S_{1^c,\omega}}} P(X^N_{S_{1^c,\omega}}|\theta) \frac{p_\omega(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1^c,\omega}})^s}{p_1(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1,\omega^c}})^s} \right)^\delta && (6)\\
&\leq \binom{D-K}{i} \sum_{S_{1,\omega}} \left( \sum_{X^N_{S_{1^c,\omega}}} P(X^N_{S_{1^c,\omega}}|\theta) \frac{p_\omega(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1^c,\omega}})^s}{p_1(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1,\omega^c}})^s} \right)^\delta, \ \forall s > 0, \ 0 \leq \delta \leq 1. && (7)
\end{aligned}$$

Inequality (5) follows from the fact that $\Pr[E_i|\omega_0 = 1, X^N_{S_1}, Y^N, \theta] \leq 1$. Consequently, if $U$ is an upper bound of this probability then it follows that, $\Pr[E_i|\omega_0 = 1, X^N_{S_1}, Y^N, \theta] \leq U^\delta$ for $\delta \in [0, 1]$. Inequality (6) follows from symmetry, namely, the inner summation is only dependent on the values of $X^N_{S_{1^c,\omega}}$ and not on the items in the set $S_{1^c,\omega}$. There are exactly $\binom{D-K}{i}$ possible sets $S_{1^c,\omega}$ hence the binomial expression. Note that the sum over $S_{1,\omega}$ cannot be further simplified. This is due to the fact that $X^N_{S_{1,\omega}}$ is already specified since we have conditioned on $X^N_{S_1}$. Since $X^N_{S_1}$ is fixed, the inner sum need not be equal for all sets $S_{1,\omega}, \omega \in \mathcal{A}_i$. Finally, (7) follows from standard observation that sum of positive numbers raised to $\delta$-th power for $\delta < 1$ is smaller than the sum of the $\delta$-th power of each number.

We now substitute for the conditional error probability derived above and follow the steps below:

$$\begin{aligned}
P(E_i) &= \int \sum_{X^N_{S_1}} \sum_{Y^N} P(\theta) P(X^N_{S_1}|\theta) p_1(Y^N|X^N_{S_1}) \Pr[E_i|\omega_0 = 1, X^N_{S_1}, Y^N, \theta] \, \mathrm{d}\theta \\
&\leq \binom{D-K}{i} \int \sum_{S_{1,\omega}} \sum_{Y^N} \sum_{X^N_{S_1}} P(\theta) P(X^N_{S_1}|\theta) p_1(Y^N|X^N_{S_1}) \left( \sum_{X^N_{S_{1^c,\omega}}} P(X^N_{S_{1^c,\omega}}|\theta) \frac{p_\omega(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1^c,\omega}})^s}{p_1(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1,\omega^c}})^s} \right)^\delta \mathrm{d}\theta
\end{aligned}$$

Due to symmetry the summation over sets $S_{1,\omega}$ does not depend on $\omega$. Since there are $\binom{K}{K-i}$ sets $S_{1,\omega}$ we get,

$$\begin{aligned}
P(E_i) &\leq \binom{D-K}{i} \binom{K}{i} \int \sum_{Y^N} \sum_{X^N_{S_1}} P(\theta) P(X^N_{S_1}|\theta) p_1(Y^N|X^N_{S_1}) \left( \sum_{X^N_{S_{1^c,\omega}}} P(X^N_{S_{1^c,\omega}}|\theta) \frac{p_\omega(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1^c,\omega}})^s}{p_1(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1,\omega^c}})^s} \right)^\delta \mathrm{d}\theta \\
&\leq \binom{D-K}{i} \binom{K}{i} \int \sum_{Y^N} \sum_{X^N_{S_{1,\omega^c}}} \sum_{X^N_{S_{1,\omega}}} P(\theta) P(X^N_{S_1}|\theta) p_1^{1-s\delta}(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1,\omega^c}}) \\
&\qquad \left( \sum_{X^N_{S_{1^c,\omega}}} P(X^N_{S_{1^c,\omega}}|\theta) p_\omega(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1^c,\omega}})^s \right)^\delta \mathrm{d}\theta \\
&= \binom{D-K}{i} \binom{K}{i} \int \sum_{Y^N} \sum_{X^N_{S_{1,\omega}}} P(\theta) P(X^N_{S_{1,\omega}}|\theta) \left( \sum_{X^N_{S_{1,\omega^c}}} P(X^N_{S_{1,\omega^c}}|\theta) p_1^{1/(1+\delta)}(Y^N|X^N_{S_{1,\omega}}, X^N_{S_{1,\omega^c}}) \right)^{1+\delta} \mathrm{d}\theta
\end{aligned}$$

where the last step follows by letting $s = \frac{1}{1+\delta}$ and noting that from symmetry $X^N_{S_{1^c,\omega}}$ is just a dummy variable and can be replaced by $X^N_{S_{1,\omega^c}}$. This establishes the weaker bound in (2), by letting $\mathcal{S}^1 = S_{1,\omega^c}$ and $\mathcal{S}^2 = S_{1,\omega}$.

$\square$

**Proof of Equation 4**

Let $\zeta_\omega$, $\omega \in \mathcal{A}_i$ denote the event where $\omega$ is more likely than 1. Then, from the definition of $\mathcal{A}_i$, the 2 encoded messages differ in $i$ variables. Hence

$$\Pr[E_i|\omega_0 = 1, X^N_{S_1}, Y^N, \theta] \leq P(\bigcup_{\omega \in \mathcal{A}_i} \zeta_\omega) \leq \sum_{\omega \in \mathcal{A}_i} P(\zeta_\omega)$$

Now note that $X^N_{S_1}$ shares $(K-i)$ variables with $X^N_{S_\omega}$. Following the introduced notation, the common partition is denoted $X^N_{S_{1,\omega}}$, which is a $N \times (K-i)$ submatrix. The remaining $i$ rows which are in $X^N_{S_1}$ but not in $X^N_{S_\omega}$ are $X^N_{S_{1,\omega^c}}$. Similarly, $X^N_{S_{1^c,\omega}}$ corresponds to variables in $X^N_{S_\omega}$ but not in $X^N_{S_1}$. In other words $X^N_{S_1} = (X^N_{S_{1,\omega}}, X^N_{S_{1,\omega^c}})$ and $X^N_{S_\omega} = (X^N_{S_{1,\omega}}, X^N_{S_{1^c,\omega}})$, where the notation $(F^{N\times n_1}; G^{N \times n_2})$ denotes an $N \times (n_1 + n_2)$ matrix with a submatrix $F$ in the first $n_1$ columns and $G$ in the remaining $n_2$ columns. Thus,

$$\begin{aligned} P(\zeta_\omega) &= \sum_{X^N_{S_\omega}: p(Y^N|X^N_{S_\omega}) \geq p(Y^N|X^N_{S_1})} P(X^N_{S_\omega}|X^N_{S_1}, \theta) \\ &\leq \sum_{X^N_{S_{1^c,\omega}}} P(X^N_{S_{1^c,\omega}}|\theta) \frac{p(Y^N|X^N_{S_\omega})^s}{p(Y^N|X^N_{S_1})^s} \qquad \forall s > 0, \ \forall \omega \in \mathcal{A}_i \end{aligned} \tag{8}$$

where by exchangeability, we have $P(X^N_{S_\omega}|X^N_{\mathcal{S}^1}, \theta) = P(X^N_{S_{1^c,\omega}}|X^N_{\mathcal{S}^1}, \theta) = P(X^N_{S_{1^c,\omega}}|\theta)$ and $\frac{p(Y^N|X^N_{S_\omega})^s}{p(Y^N|X^N_{S_1})^s} \geq 1$ for all $s > 0$, since $\frac{p(Y^N|X^N_{S_\omega})}{p(Y^N|X^N_{S_1})} \geq 1$.

$\square$

## 2 Proof of Theorem 2.2

We first derive the sufficiency bound, using the results of Theorem 2.1. To achieve that, we derive a sufficient condition for the error exponent of the error probability $P(E_i)$ in (1) to be positive and to drive the error probability to zero as $D \to \infty$. Specifically,

$$Nf(\delta) = NE_o(\delta) - \delta \log \binom{D-K}{i}\binom{K}{i} \to \infty \tag{9}$$

where

$$f(\delta) = E_o(\delta) - \delta \frac{\log \binom{D-K}{i}\binom{K}{i}}{N}.$$

To establish the sufficiency bound we follow the argument in [4]. Note that $f(0) = 0$. Since the function $f(\delta)$ is differentiable and has a power series expansion, for a sufficiently small $\delta$, we get by Taylor series expansion in the neighborhood $\delta = 0$ that,

$$f(\delta) = f(0) + \delta \frac{df}{d\delta}\Big|_{\delta=0} + O(\delta^2)$$

Note that

$$\frac{\partial E_o}{\partial \delta}\Big|_{\delta=0} = \frac{I(X^N_{\mathcal{S}^1}; Y^N | X^N_{\mathcal{S}^2}, \theta)}{N}, \tag{10}$$

which is shown in the next section.

We can further decompose $I(X^N_{\mathcal{S}^1};Y^N|X^N_{\mathcal{S}^2},\theta)$ using the following chain of equalities:

$$\begin{aligned}I(X^N_{\mathcal{S}^1};Y^N|X^N_{\mathcal{S}^2},\theta)+I(\beta_S;X^N_{\mathcal{S}^1}|X^N_{\mathcal{S}^2},Y^N,\theta) &= I(X^N_{\mathcal{S}^1};Y^N,\beta_S|X^N_{\mathcal{S}^2},\theta)=I(X^N_{\mathcal{S}^1};\beta_S|\theta)+I(X^N_{\mathcal{S}^1};Y^N|X^N_{\mathcal{S}^2},\beta_S,\theta)\\ &= NI(X_{\mathcal{S}^1};Y|X_{\mathcal{S}^2},\beta_S,\theta),\end{aligned}$$

where the last equality is due to $X$ and $\beta_S$ being independent and $(X^N,Y^N)$ pairs being independent over $n$ given $\beta_S$. Therefore we have

$$\frac{\partial E_o}{\partial \delta}\bigg|_{\delta=0}=\frac{I(X^N_{\mathcal{S}^1};Y^N|X^N_{\mathcal{S}^2},\theta)}{N}=I(X_{\mathcal{S}^1};Y|X_{\mathcal{S}^2},\beta_S,\theta)-\frac{I(\beta_S;X^N_{\mathcal{S}^1}|X^N_{\mathcal{S}^2},Y^N,\theta)}{N}. \tag{11}$$

Now assume that $N$ satisfies

$$N>(1+\epsilon)\frac{\log\binom{D-K}{i}\binom{K}{i}}{I(X_{\mathcal{S}^1};Y|X_{\mathcal{S}^2},\beta_S,\theta)}. \tag{12}$$

for any constant $\epsilon>0$. We note that from the Lagrange form of the Taylor Series expansion (an application of the mean value theorem) we can write $E_o(\delta)$ in terms of its first derivative evaluated at zero and a remainder term, i.e.,

$$E_o(\delta)=E_o(0)+\delta E_o'(0)+\frac{\delta^2}{2}E_o''(\psi)$$

for some $\psi\in[0,\delta]$. Hence, for the choice of $N$ in (12) and using (11) we have

$$Nf(\delta)\geq N\left(\delta\frac{\epsilon}{1+\epsilon}I(X_{\mathcal{S}^1};Y|X_{\mathcal{S}^2},\beta_S,\theta)-\delta^2CI(X_{\mathcal{S}^1};Y|X_{\mathcal{S}^2},\beta_S,\theta)-\delta\frac{I(\beta_S;X^N_{\mathcal{S}^1}|X^N_{\mathcal{S}^2},Y^N,\theta)}{N}\right) \tag{13}$$

where $C=\frac{|E_o''(\psi)|}{2I(X_{\mathcal{S}^1};Y|X_{\mathcal{S}^2},\beta_S,\theta)}$ which might depend on $K$.

A preliminary analysis of the necessary condition that we establish in the next section reveals that $N=\Omega(K\log D)$ is necessary, since $\log\binom{D-K+i}{i}=\Theta(i\log D)$ and $I(X_{\mathcal{S}^1};Y|X_{\mathcal{S}^2},\beta_S,\theta)\leq H(Y)=O(1)$. Also, $I(\beta_S;X^N_{\mathcal{S}^1}|X^N_{\mathcal{S}^2},Y^N,\theta)\leq H(\beta_S)$, which is constant with respect to $D$ since the observation model is only dependent on $K$ variables, due to the sparsity assumption of the observation model $P(Y|X)$. So we see that

$$\frac{I(\beta_S;X^N_{\mathcal{S}^1}|X^N_{\mathcal{S}^2},Y^N,\theta)}{N}=O\left(\frac{1}{\log D}\right)$$

which is always dominated by $I(X_{\mathcal{S}^1};Y|X_{\mathcal{S}^2},\beta_S,\theta)$, which we assumed to be $\omega(1/\log D)$. Therefore we can rewrite (13) as

$$Nf(\delta)\geq N\left(\delta\left(\frac{\epsilon}{1+\epsilon}-o(1)\right)I(X_{\mathcal{S}^1};Y|X_{\mathcal{S}^2},\beta_S,\theta)-\delta^2CI(X_{\mathcal{S}^1};Y|X_{\mathcal{S}^2},\beta_S,\theta)\right).$$

Finally, if we choose $\delta\leq\frac{\epsilon'}{C}$, where $\epsilon'=\frac{\epsilon}{1+\epsilon}$, then $f(\delta)=\eta$ for some $\eta>0$ which does not depend on $D$ or $N$. It follows that $Nf(\delta)\to\infty$ as $D\to\infty$.

We have just shown that for fixed $K$,

$$N>(1+\epsilon)\cdot\frac{\log\binom{D-K}{i}\binom{K}{i}}{I(X_{\mathcal{S}^1};Y|X_{\mathcal{S}^2},\beta_S,\theta)}$$

is sufficient to ensure an arbitrarily small $P(E_i)$. Now note that

$$(1+\epsilon)\binom{D-K+i}{i}\geq\binom{D-K}{i}\binom{K}{i}$$

asymptotically as $D \to \infty$ and $K$ is fixed, for any constant $\epsilon > 0$, which can be incorporated into the previous $\epsilon$ as both are arbitrary. Since the average error probability $P(E) \le \sum_{i=1}^{K} P(E_i)$, it follows that if

$$N > (1+\epsilon) \max_{i=1,\dots,K} \frac{\log \binom{D-K+i}{i}}{I(X_{\mathcal{S}^1}; Y | X_{\mathcal{S}^2}, \beta_S, \theta)}$$

then for any fixed $K$, $\lim_{D\to\infty} P(E) = 0$. Consequently, since this is true for any $K$, $\lim_{K\to\infty} \lim_{D\to\infty} P(E) = 0$.

□

It is important to highlight the main difference between the analysis of the error probability for the problem considered herein and the channel coding problem. In contrast to channel coding, the codewords of a candidate set and the true set are not independent since the two sets could be overlapping. To overcome this difficulty, we separate the error events $E_i$, $i = 1, \dots, K$, of misclassifying the true set in $i$ items. Then, for every $i$ we average over realizations of ensemble of codewords for every candidate set while holding fixed the partition common to these sets and the true set of variables.

**Proof of Equation 10**

We have

$$E_o(\delta) = -\frac{1}{N} \log \sum_{\theta} \sum_{Y^N} \sum_{X^N_{\mathcal{S}^2}} P(X^N_{\mathcal{S}^2}|\theta) P(\theta) \left[ \sum_{X^N_{\mathcal{S}^1}} P(X^N_{\mathcal{S}^1}|\theta) p(Y^N | X^N_{\mathcal{S}^1}, X^N_{\mathcal{S}^2})^{\frac{1}{1+\delta}} \right]^{1+\delta} \qquad 0 \le \delta \le 1$$

where its derivative at $\delta = 0$ can be written as

$$\frac{\partial E_o}{\partial \delta}\Big|_{\delta=0} = -\frac{1}{N} \frac{1}{\sum_{Y^N, X^N_{\mathcal{S}^1}, X^N_{\mathcal{S}^2}, \theta} P(X_{\mathcal{S}^1}|\theta) P(X_{\mathcal{S}^2}|\theta) P(\theta) p(Y^N | X^N_S)}$$
$$\sum_{Y^N, X^N_{\mathcal{S}^2}, \theta} P(X^N_{\mathcal{S}^2}|\theta) P(\theta) \left( \sum_{X^N_{\mathcal{S}^1}} P(X^N_{\mathcal{S}^1}|\theta) p(Y^N|X^N_S) \left[ \log \left( \sum_{X^N_{\mathcal{S}^1}} P(X^N_{\mathcal{S}^1}|\theta) p(Y^N|X^N_S) \right) - \log p(Y^N|X^N_S) \right] \right).$$

Noting that $P(X^N_{\mathcal{S}^1}|\theta) P(X^N_{\mathcal{S}^2}|\theta) = P(X^N_{\mathcal{S}^1}, X^N_{\mathcal{S}^2}|\theta)$ by the representation theorem and $p(Y^N|X^N_S) = p(Y^N|X^N_S, \theta)$ by the independence of $Y$ and $\theta$ given $X_S$, above equality simplifies to

$$\begin{aligned}
\frac{\partial E_o}{\partial \delta}\Big|_{\delta=0} &= -\frac{1}{N} \frac{1}{\sum_{Y^N, X^N_{\mathcal{S}^1}, X^N_{\mathcal{S}^2}, \theta} P(Y^N, X^N_S, \theta)} \sum_{Y^N, X^N_{\mathcal{S}^1}, X^N_{\mathcal{S}^2}, \theta} P(Y^N, X^N_S, \theta) \left( \log P(Y^N | X^N_{\mathcal{S}^2}, \theta) - \log P(Y^N | X^N_S, \theta) \right) \\
&= \frac{1}{N} \sum_{Y^N, X^N_{\mathcal{S}^1}, X^N_{\mathcal{S}^2}, \theta} P(Y^N, X^N_S, \theta) \log \frac{P(Y^N, X^N_{\mathcal{S}^2} | X^N_{\mathcal{S}^1}, \theta)}{P(Y^N, X^N_{\mathcal{S}^2} | \theta)} \\
&= \frac{I(X^N_{\mathcal{S}^1}; X^N_{\mathcal{S}^2}, Y^N | \theta)}{N} \\
&= \frac{I(X^N_{\mathcal{S}^1}; Y^N | X^N_{\mathcal{S}^2}, \theta)}{N},
\end{aligned}$$

where the second equality follows by noting the first denominator is equal to 1 and by adding and subtracting $\log P(X^N_{\mathcal{S}^2}|\theta)$ inside the parenthesis. The third equality follows from the definition of mutual information. The final equality follows from the independence of $X_{\mathcal{S}^1}$ and $X_{\mathcal{S}^2}$ given $\theta$.

## 2.1 Necessity bound

The vector of outcomes $Y^N$ is probabilistically related to the index $\omega \in \mathcal{I} = \{1, 2, \ldots, \binom{D}{K}\}$. Suppose $K - i$ elements of the salient set are revealed to us, denoted by $\mathcal{S}^2$. From $X^N$ and $Y^N$ we estimate the set index $\omega$. Let the estimate be $\hat{\omega} = g(X^N, Y^N)$. Define the probability of error

$$P_e = P(E) = \Pr[\hat{\omega} \neq \omega].$$

$E$ is a binary random variable that takes the value 1 in case of an error i.e., if $\hat{\omega} \neq \omega$, and 0 otherwise, then using the chain rule of entropies [2] we have

$$\begin{aligned} H(E, \omega | Y^N, X^N, \mathcal{S}^2) &= H(\omega | Y^N, X^N, \mathcal{S}^2) + H(E | \omega, Y^N, X^N, \mathcal{S}^2) \\ &= H(E | Y^N, X^N, \mathcal{S}^2) + H(\omega | E, Y^N, X^N, \mathcal{S}^2). \end{aligned} \tag{14}$$

The random variable $E$ is fully determined given $X^N$, $Y^N$, $\omega$ and $\mathcal{S}^2$. It follows that $H(E|\omega, Y^N, X^N, \mathcal{S}^2) = 0$. Since $E$ is a binary random variable $H(E|Y^N, X^N, \mathcal{S}^2) \leq 1$. Consequently, we can bound $H(\omega|E, Y^N, X^N, \mathcal{S}^2)$ as follows,

$$\begin{aligned} H(\omega | E, Y^N, X^N, \mathcal{S}^2) &= P(E = 0) H(\omega | E = 0, Y^N, X^N, \mathcal{S}^2) + P(E = 1) H(\omega | E = 1, Y^N, X^N, \mathcal{S}^2) \\ &\leq (1 - P_e)\, 0 + P_e \log \left( \binom{D - K + i}{i} - 1 \right) \\ &\leq P_e \log \binom{D - K + i}{i}. \end{aligned} \tag{15}$$

The first inequality follows from the fact that revealing $K - i$ entries, and given that $E = 1$, the conditional entropy can be upper bounded by the logarithm of the number of outcomes. From (14), we obtain the genie aided Fano's inequality

$$H(\omega | Y^N, X^N, \mathcal{S}^2) \leq 1 + P_e \log \binom{D - K + i}{i} \tag{16}$$

Note that for the left hand term, we have

$$\begin{aligned} H(\omega | Y^N, X^N, \mathcal{S}^2) &= H(\omega | \mathcal{S}^2) - I(\omega; Y^N, X^N | \mathcal{S}^2) \\ &= H(\omega | \mathcal{S}^2) - I(\omega; X^N | \mathcal{S}^2) - I(\omega; Y^N | X^N, \mathcal{S}^2) \\ &\overset{(a)}{=} H(\omega | \mathcal{S}^2) - I(\omega; Y^N | X^N, \mathcal{S}^2) \\ &\overset{(b)}{=} H(\omega | \mathcal{S}^2) - (H(Y^N | X^N, \mathcal{S}^2) - H(Y^N | X^N, \omega)) \\ &\overset{(c)}{=} H(\omega | \mathcal{S}^2) - (H(Y^N | X^N, \mathcal{S}^2, \theta) - H(Y^N | X^N, \omega, \theta)) \\ &\overset{(d)}{\geq} H(\omega | \mathcal{S}^2) - (H(Y^N | X^N_{\mathcal{S}^2}, \theta) - H(Y^N | X^N_{S_\omega}, \theta)) \\ &\overset{(e)}{=} H(\omega | \mathcal{S}^2) - I(X^N_{\mathcal{S}^1}; Y^N | X^N_{\mathcal{S}^2}, \theta) \end{aligned}$$

where (a) follows from the fact that $X^N$ is independent of $\mathcal{S}^2$ and $\omega$; (b) follows from the fact that conditioning with respect to $\omega$ includes conditioning with respect to $\mathcal{S}^2$; (c) follows from the independence of $Y$ and $\theta$ given $X$; (d) follows from the fact that $Y^N$ depends on $\mathcal{S}^2$ only through $X^N_{\mathcal{S}^2}$ and similarly for the second term $Y^N$ depends on $\omega$ only through $X^N_{S_\omega}$; the argument for (e) follows by definition.

From (16), it then follows that

$$H(\omega | \mathcal{S}^2) - I(X^N_{\mathcal{S}^1}; Y^N | X^N_{\mathcal{S}^2}, \theta) \leq 1 + P_e \log \binom{D - K + i}{i}$$

and since the set $\mathcal{S}^2$ of $K-i$ variables is revealed, $\omega$ is uniformly distributed over the set of indices that correspond to sets of size $K$ containing $\mathcal{S}^2$. It follows that

$$\log\binom{D-K+i}{i} - I(X^N_{\mathcal{S}^1}; Y^N | X^N_{\mathcal{S}^2}, \theta) \leq 1 + P_e \log\binom{D-K+i}{i}.$$

Rewriting the above inequality, we have

$$P_e \geq 1 - \frac{I(X^N_{\mathcal{S}^1}; Y^N | X^N_{\mathcal{S}^2}, \theta) + 1}{\log\binom{D-K+i}{i}}. \tag{17}$$

Thus, for the probability of error to be asymptotically bounded away from zero, it is necessary that

$$\log\binom{D-K+i}{i} \leq I(X^N_{\mathcal{S}^1}; Y^N | X^N_{\mathcal{S}^2}, \theta) = NI(X_{\mathcal{S}^1}; Y | X_{\mathcal{S}^2}, \beta_S, \theta) - I(\beta_S; X^N_{\mathcal{S}^1} | X^N_{\mathcal{S}^2}, Y^N, \theta). \tag{18}$$

Using (11), we can see that

$$N \geq \max_{i=1,\dots,K} \frac{\log\binom{D-K+i}{i}}{I(X_{\mathcal{S}^1}; Y | X_{\mathcal{S}^2}, \beta_S, \theta) - \frac{I(\beta_S; X^N_{\mathcal{S}^1} | X^N_{\mathcal{S}^2}, Y^N, \theta)}{N}}$$

is a necessary condition for the number of samples $N$. Finally, since $I(\beta_S; X^N_{\mathcal{S}^1} | X^N_{\mathcal{S}^2}, Y^N) \geq 0$, the following expression is a lower bound to the expression above, proving that it is a necessary condition for recovery,

$$N \geq \max_{i=1,\dots,K} \frac{\log\binom{D-K+i}{i}}{I(X_{\mathcal{S}^1}; Y | X_{\mathcal{S}^2}, \beta_S, \theta)}.$$

□

# 3 Continuous Variables

Even though the results and proof ideas that were used in the above sections are fairly general, the proofs provided for sufficiency bounds were stated for discrete variables and outcomes. In this section we make the necessary generalizations to extend these proofs to continuous variable and observation models. We follow the methodology in [4] and [3].

To simplify the exposition, we consider the extension to continuous variables in the special case of fixed and known $\beta_S$ and i.i.d. variables. Let $Q(X) = \prod_{i=1}^{D} Q(X_i)$ denote the joint distribution of variables $X$. The extensions to random $\beta_S$ and conditionally i.i.d. variables are straightforward. In this case, $I(X_{\mathcal{S}^1}; Y | X_{\mathcal{S}^2}, \beta_S)$ reduces to $I(X_{\mathcal{S}^1}; Y | X_{\mathcal{S}^2})$ and $E_o(\delta)$ reduces to

$$E_o(\delta) = -\log \sum_{Y} \sum_{X_{\mathcal{S}^2}} \left[ \sum_{X_{\mathcal{S}^1}} Q(X_{\mathcal{S}^1}) p(Y, X_{\mathcal{S}^2} | X_{\mathcal{S}^1})^{\frac{1}{1+\delta}} \right]^{1+\delta} \qquad 0 \leq \delta \leq 1 \tag{19}$$

with $\left.\frac{\partial E_o(\delta)}{\partial \delta}\right|_{\delta=0} = I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y) = I(X_{\mathcal{S}^1}; Y | X_{\mathcal{S}^2})$, since $(X^{(n)}, Y^{(n)})$ pairs are independent across $n$ for fixed $\beta_S$.

Assume the continuous joint variable probability density $Q(X)$ with joint cumulative density function $F$ and the conditional probability density $p(Y = y | X_S = x)$ for the observation model, which is assumed to be a continuous function of both $x$ and $y$.

Let $X' \in \mathcal{X}'^N$ be the random vector and $Y' \in \mathcal{Y}'$ be the random variable generated by the quantization of $X \in \mathcal{X}^N = \mathbb{R}^N$ and $Y \in \mathcal{Y} = \mathbb{R}$ respectively, where each variable in $X$ is quantized to $L$ values and $Y$ quantized to $J$ values. Let $F'$ be the joint cumulative density function of $X'$. As before, let $\hat{S}(X^N, Y^N)$

be the ML decoder with continuous inputs with probability of making $i$ errors in decoding denoted by $P(E_i)$. Let $\hat{S}(X'^N, Y'^N)$ be the ML decoder that quantizes inputs $X^N$ and $Y^N$ to $X'^N$ and $Y'^N$ and has a corresponding probability of error $P'(E_i)$. Define

$$E_o(\delta, X', Y') = -\log \sum_{y' \in \mathcal{Y}'} \sum_{x'_{\mathcal{S}^2} \in \mathcal{X}'^{K-i}} \left[ \sum_{x'_{\mathcal{S}^1} \in \mathcal{X}'^i} Q(x'_{\mathcal{S}^1}) p(y', x'_{\mathcal{S}^2} | x'_{\mathcal{S}^1})^{\frac{1}{1+\delta}} \right]^{1+\delta},$$

$$E_o(\delta, X, Y) = -\log \int_{\mathcal{Y}} \int_{\mathcal{X}^{K-i}} \left[ \int_{\mathcal{X}^i} Q(x_{\mathcal{S}^1}) p(y, x_{\mathcal{S}^2} | x_{\mathcal{S}^1})^{\frac{1}{1+\delta}} \, \mathrm{d}x_{\mathcal{S}^1} \right]^{1+\delta} \mathrm{d}x_{\mathcal{S}^2} \, \mathrm{d}y,$$

where the indexing denotes the random variates which the error exponents are computed with respect to.

Utilizing the results in the proof of Theorem 3.1 for the discrete models, we will show the following for the continuous model

$$P(E_i) \leq 2^{-N\left(E_o(\delta, X, Y) - \delta \frac{\log \binom{D-K}{i}\binom{K}{i}}{N}\right)}. \tag{20}$$

The rest of the proof will then follow as in the discrete case, by noting that $\frac{\partial E_o(\delta, X, Y)}{\partial \delta}\Big|_{\delta=0} = I(X_{\mathcal{S}^1}; X_{\mathcal{S}^2}, Y)$, with the mutual information definition for continuous variables [2].

Our strategy will be the following: we will increase the number of quantization levels for $Y'$ and $X'$ respectively and since discrete result (1) holds for any number of quantization levels, by taking limits we will be able to show that

$$P'(E_i) \leq 2^{-N\left(E_o(\delta, X, Y) - \delta \frac{\log \binom{D-K}{i}\binom{K}{i}}{N}\right)}. \tag{21}$$

Since $\hat{S}(X^N, Y^N)$ is the minimum probability of error decoder, any upper bound for $P'(E_i)$ will also be an upper bound for $P(E_i)$, proving (20).

Assume $Y$ is quantized with the quantization boundaries denoted by $a_1, \ldots, a_{J-1}$, with $Y' = a_j$ if $a_{j-1} < Y \leq a_j$. For convenience denote $a_0 = -\infty$ and $a_J = \infty$. Furthermore assume quantization boundaries are equally spaced, i.e. $a_j - a_{j-1} = \Delta_J$ for $2 \leq j \leq J-1$. Now we can write the following

$$\begin{aligned}
E_o(\delta, X', Y') &= -\log \sum_{j=1}^{J} \sum_{x'_{\mathcal{S}^2}} \left[ \sum_{x'_{\mathcal{S}^1}} Q(x'_{\mathcal{S}^1}) \left( \int_{a_{j-1}}^{a_j} p(y, x'_{\mathcal{S}^2} | x'_{\mathcal{S}^1}) \, \mathrm{d}y \right)^{\frac{1}{1+\delta}} \right]^{1+\delta} \\
&= -\log \Bigg\{ \sum_{j=2}^{J-1} \Delta_J \sum_{x'_{\mathcal{S}^2}} \left[ \sum_{x'_{\mathcal{S}^1}} Q(x'_{\mathcal{S}^1}) \left( \frac{\int_{a_{j-1}}^{a_j} p(y, x'_{\mathcal{S}^2} | x'_{\mathcal{S}^1}) \, \mathrm{d}y}{\Delta_J} \right)^{\frac{1}{1+\delta}} \right]^{1+\delta} \\
&\quad + \sum_{x'_{\mathcal{S}^2}} \left[ \sum_{x'_{\mathcal{S}^1}} Q(x'_{\mathcal{S}^1}) \left( \int_{-\infty}^{a_1} p(y, x'_{\mathcal{S}^2} | x'_{\mathcal{S}^1}) \, \mathrm{d}y \right)^{\frac{1}{1+\delta}} \right]^{1+\delta} \\
&\quad + \sum_{x'_{\mathcal{S}^2}} \left[ \sum_{x'_{\mathcal{S}^1}} Q(x'_{\mathcal{S}^1}) \left( \int_{a_{J-1}}^{\infty} p(y, x'_{\mathcal{S}^2} | x'_{\mathcal{S}^1}) \, \mathrm{d}y \right)^{\frac{1}{1+\delta}} \right]^{1+\delta} \Bigg\}.
\end{aligned}$$

Let $J \to \infty$ and for each $J$ choose the sequence of quantization boundaries such that $\lim \Delta_J = 0$, $\lim a_{J-1} = \infty$, $\lim a_1 = -\infty$. Then the last two terms disappear and using the fundamental theorem of calculus, we obtain

$$\lim_{J \to \infty} E_o(\delta, X', Y') = E_o(\delta, X', Y) = -\log \int_{\mathcal{Y}} \sum_{x'_{\mathcal{S}^2}} \left[ \sum_{x'_{\mathcal{S}^1}} Q(x'_{\mathcal{S}^1}) p(y, x'_{\mathcal{S}^2} | x'_{\mathcal{S}^1})^{\frac{1}{1+\delta}} \right]^{1+\delta} \mathrm{d}y. \tag{22}$$

Although it is not necessary for our proof, it can also be shown that $E_o(\delta, X', Y')$ increases for finer quantizations of $Y'$, therefore $E_o(\delta, X', Y)$ gives the smallest upper bound over $P'(E_i)$ over the quantizations of $Y$.

We repeat the same procedure for $X$. Assume each variable $X_n$ in $X$ is quantized with the quantization boundaries denoted by $b_1, \ldots, b_{L-1}$, with $X'_n = b_l$ if $b_{l-1} < X_n \leq b_l$. For convenience denote $b_0 = -\infty$ and $b_L = \infty$. Furthermore assume quantization boundaries are equally spaced, i.e. $b_l - b_{l-1} = \Delta_L$ for $2 \leq l \leq L-1$. Then we can write

$$\begin{aligned}
E_o(\delta, X', Y) &= -\log \int_Y \sum_{l=1}^{L} \left[ \sum_{x'_{\mathcal{S}^1}} Q(x'_{\mathcal{S}^1}) \left( \int_{b_{l-1}}^{b_l} p(y, x_{\mathcal{S}^2} | x'_{\mathcal{S}^1}) \, \mathrm{d}x_{\mathcal{S}^2} \right)^{\frac{1}{1+\delta}} \right]^{1+\delta} \mathrm{d}y \\
&= -\log \int_{\mathcal{Y}} \sum_{l=1}^{L} \left[ \int_{\mathcal{X}^i} \left( \int_{b_{l-1}}^{b_l} p(y, x_{\mathcal{S}^2} | x_{\mathcal{S}^1}) \, \mathrm{d}x_{\mathcal{S}^2} \right)^{\frac{1}{1+\delta}} \mathrm{d}F'(x_{\mathcal{S}^1}) \right]^{1+\delta} \mathrm{d}y \qquad (23) \\
&= -\log \int_{\mathcal{Y}} \Bigg\{ \sum_{l=2}^{L-1} \Delta_L \left[ \int_{\mathcal{X}^i} \left( \frac{\int_{b_{l-1}}^{b_l} p(y, x_{\mathcal{S}^2} | x_{\mathcal{S}^1}) \, \mathrm{d}x_{\mathcal{S}^2}}{\Delta_L} \right)^{\frac{1}{1+\delta}} \mathrm{d}F'(x_{\mathcal{S}^1}) \right]^{1+\delta} \\
&\qquad + \int_{\mathcal{X}^i} \left( \int_{-\infty}^{b_1} p(y, x_{\mathcal{S}^2} | x_{\mathcal{S}^1}) \, \mathrm{d}x_{\mathcal{S}^2} \right)^{\frac{1}{1+\delta}} \mathrm{d}F'(x_{\mathcal{S}^1}) \\
&\qquad + \int_{\mathcal{X}^i} \left( \int_{b_{L-1}}^{\infty} p(y, x_{\mathcal{S}^2} | x_{\mathcal{S}^1}) \, \mathrm{d}x_{\mathcal{S}^2} \right)^{\frac{1}{1+\delta}} \mathrm{d}F'(x_{\mathcal{S}^1}) \Bigg\} \mathrm{d}y.
\end{aligned}$$

where (23) follows with $F'(x_{\mathcal{S}^1})$ being the step function which represents the cumulative density function of the quantized variables $X'_{\mathcal{S}^1}$.

Let $L \to \infty$, for each $L$ choose a set of quantization point such that $\lim \Delta_L = 0$, $\lim b_{L-1} = \infty$, $\lim b_1 = -\infty$. Again, the second and third terms disappear and the first sum converges to the integral over $X_{\mathcal{S}^2}$. Note that $p(y, x_{\mathcal{S}^2} | x_{\mathcal{S}^1})$ is a continuous function of all its variables since it was assumed that $Q(x)$ and $p(y|x)$ were continuous. Also note that $\lim_{L\to\infty} F' = F$, which implies the weak convergence of the probability measure of $X'$ to the probability measure of $X$. Given these facts, using the portmanteau theorem we obtain that $E_{F'}\left[p(Y, X_{\mathcal{S}^2} | X_{\mathcal{S}^1})\right] \to E_F\left[p(Y, X_{\mathcal{S}^2} | X_{\mathcal{S}^1})\right]$, which leads to

$$\lim_{L\to\infty} E_o(\delta, X', Y) = -\log \int_{\mathcal{Y}} \int_{\mathcal{X}^{K-i}} \left[ \int_{\mathcal{X}^i} p(y, x_{\mathcal{S}^2} | x_{\mathcal{S}^1})^{\frac{1}{1+\delta}} \, \mathrm{d}F(x_{\mathcal{S}^1}) \right]^{1+\delta} \mathrm{d}x_{\mathcal{S}^2} \, \mathrm{d}y = E_o(\delta, X, Y). \qquad (24)$$

This leads to the following result, completing the proof.

$$P(E_i) \leq P'(E_i) \leq \lim_{J,L\to\infty} 2^{-N\left( E_o(\delta, X', Y') - \delta \frac{\log \binom{D-K}{i}\binom{K}{i}}{N} \right)} = 2^{-N\left( E_o(\delta, X, Y) - \delta \frac{\log \binom{D-K}{i}\binom{K}{i}}{N} \right)}. \qquad (25)$$

# 4 Proof of Theorem 3.1

To derive the upper bound on error probability, we compute $E_o(\delta)$ explicitly and replace it in Theorem 2.1. First we compute for the easier case, with fixed $\beta_S = \sigma$. In this case, note that $(X, Y)$ pairs are independent across samples and

$$E_o(\delta) = -\log \int_{\theta} P(\theta) \int_Y \int_{X_{\mathcal{S}^2}} P(X_{\mathcal{S}^2} | \theta) \left[ \int_{X_{\mathcal{S}^1}} P(X_{\mathcal{S}^1} | \theta) p(Y | X_{\mathcal{S}^1}, X_{\mathcal{S}^2})^{\frac{1}{1+\delta}} \, \mathrm{d}X_{\mathcal{S}^1} \right]^{1+\delta} \mathrm{d}X_{\mathcal{S}^2} \, \mathrm{d}Y \, \mathrm{d}\theta, \quad 0 \leq \delta \leq 1.$$

For the correlated Gaussian variables, this reduces to

$$E_o(\delta) = -\log \int_{\mu} \mathcal{N}(\mu; 0, \rho/N) \int_{Y} \int_{X_{\mathcal{S}^2}} \mathcal{N}(x_2; (K-i)\mu, (K-i)(1-\rho)/N)$$
$$\left[ \int_{X_{\mathcal{S}^1}} \mathcal{N}(x_1; i\mu, i(1-\rho)/N) \mathcal{N}(y - \sigma(x_1 + x_2); 0, 1/\mathrm{SNR})^{\frac{1}{1+\delta}} \, \mathrm{d}x_1 \right]^{1+\delta} \mathrm{d}x_2 \, \mathrm{d}y \, \mathrm{d}\mu.$$

As the first step, we input the Gaussian distributions and take the integral inside the brackets over $x_1$, which gives us

$$\Big[ \int_{X_{\mathcal{S}^1}} \mathcal{N}(x_1; i\mu, i(1-\rho)/N) \mathcal{N}(y - \sigma(x_1 + x_2); 0, 1/\mathrm{SNR})^{\frac{1}{1+\delta}} \, \mathrm{d}x_1 \Big]^{1+\delta}$$
$$= \frac{\left(\sqrt{\frac{1}{\mathrm{SNR}\sigma^2}}\right)^{\delta}}{\sigma\sqrt{2\pi}\left(\sqrt{\frac{i(1-\rho)}{N(1+\delta)} + \frac{1}{\mathrm{SNR}\sigma^2}}\right)^{1+\delta}} \exp\left( -\frac{(\frac{y}{\sigma} - x_2 - i\mu)^2}{2\left(\frac{i(1-\rho)}{N(1+\delta)} + \frac{1}{\mathrm{SNR}\sigma^2}\right)} \right).$$

By plugging in this expression and integrating over $x_2$, we then have

$$\int_{X_{\mathcal{S}^2}} \mathcal{N}(x_2; (K-i)\mu, (K-i)(1-\rho)/N) \Big[ \int_{X_{\mathcal{S}^1}} \mathcal{N}(x_1; i\mu, i(1-\rho)/N) \mathcal{N}(y - \sigma(x_1 + x_2); 0, 1/\mathrm{SNR})^{\frac{1}{1+\delta}} \, \mathrm{d}x_1 \Big]^{1+\delta} \mathrm{d}x_2$$
$$= \frac{1}{\sigma\sqrt{2\pi}} \left( \frac{1}{\sqrt{1 + \frac{i(1-\rho)\mathrm{SNR}\sigma^2}{N(1+\delta)}}} \right)^{\delta} \frac{1}{\sqrt{\frac{(K-i)(1-\rho)}{N} + \frac{i(1-\rho)}{N(1+\delta)} + \frac{1}{\mathrm{SNR}\sigma^2}}} \exp\left( -\frac{(y - \sigma K\mu)^2}{2\sigma^2\left(\frac{(K-i)(1-\rho)}{N} + \frac{i(1-\rho)}{N(1+\delta)} + \frac{1}{\mathrm{SNR}\sigma^2}\right)} \right)$$

Integrating the above expression over $y$, we are left with

$$\left( \frac{1}{\sqrt{1 + \frac{i(1-\rho)\mathrm{SNR}\sigma^2}{N(1+\delta)}}} \right)^{\delta},$$

which no longer depends on $\mu$, therefore the expectation over $\mu$ is equal to the above expression and finally we have

$$E_o(\delta) = \frac{\delta}{2} \log\left( 1 + (1-\rho)\frac{i\sigma^2\mathrm{SNR}}{N(1+\delta)} \right),$$

for any $0 \leq \delta \leq 1$.

Now we will show a lower bound on the error exponent $E_o(\delta)$ for the case where $\beta_S$ is random and IID $\mathcal{N}(0, \sigma^2)$. In this case, $Y^{(n)}$ are not independent across $n$. In order to lower bound $E_o$, we first upper bound the observation probability such that,

$$p(Y^N | X^N_{\mathcal{S}^1}, X^N_{\mathcal{S}^2})^{\frac{1}{1+\delta}} = \left( \int_{\beta_S} P(\beta_S) P(Y^N | X^N_{\mathcal{S}^1}, X^N_{\mathcal{S}^2}, \beta_S) \, \mathrm{d}\beta_S \right)^{\frac{1}{1+\delta}} \leq \int_{\beta_S} P(\beta_S)^{\frac{1}{1+\delta}} P(Y^N | X^N_{\mathcal{S}^1}, X^N_{\mathcal{S}^2}, \beta_S)^{\frac{1}{1+\delta}} \, \mathrm{d}\beta_S$$

which follows from the subadditivity of exponent $\frac{1}{1+\delta}$. A lower bound on $E_o$ is then given by

$$E_o(\delta) \geq -\frac{1}{N} \log M^{1+\delta} \int_{\theta^N} P(\theta^N) \int_{Y^N} \int_{X^N_{\mathcal{S}^2}} P(X^N_{\mathcal{S}^2} | \theta^N)$$
$$\left[ \int_{\beta_S} \frac{P(\beta_S)^{\frac{1}{1+\delta}}}{M} \int_{X^N_{\mathcal{S}^1}} P(X^N_{\mathcal{S}^1} | \theta^N) P(Y^N | X^N_{\mathcal{S}^1}, X^N_{\mathcal{S}^2}, \beta_S)^{\frac{1}{1+\delta}} \, \mathrm{d}X^N_{\mathcal{S}^1} \, \mathrm{d}\beta_S \right]^{1+\delta} \mathrm{d}X^N_{\mathcal{S}^2} \, \mathrm{d}Y^N \, \mathrm{d}\theta$$

where $M = \int P(\beta_S)^{\frac{1}{1+\delta}} \, \mathrm{d}\beta_S$ and then by Jensen's inequality, it follows that

$$E_o(\delta) \geq -\frac{1}{N} \log M^{\delta} \int_{\beta_S} P(\beta_S)^{\frac{1}{1+\delta}} \Bigg( \int_{\theta} P(\theta) \int_Y \int_{X_{\mathcal{S}^2}} P(X_{\mathcal{S}^2}|\theta) \\ \left[ \int_{X_{\mathcal{S}^1}} P(X_{\mathcal{S}^1}|\theta) P(Y|X_{\mathcal{S}^1}, X_{\mathcal{S}^2}, \beta_S)^{\frac{1}{1+\delta}} \, \mathrm{d}X_{\mathcal{S}^1} \right]^{1+\delta} \mathrm{d}X_{\mathcal{S}^2} \, \mathrm{d}Y \, \mathrm{d}\theta \Bigg)^N \mathrm{d}\beta_S \tag{26}$$

where we also used the independence of $(X^{(n)}, Y^{(n)})$ across $n$ given $\beta_S$.

We start by taking the integral inside the square brackets. For the linear model set-up we have,

$$\int_{X_{\mathcal{S}^1}} P(X_{\mathcal{S}^1}|\theta) P(Y|X_{\mathcal{S}^1}, X_{\mathcal{S}^2}, \beta_S)^{\frac{1}{1+\delta}} \, \mathrm{d}X_{\mathcal{S}^1} = \int_{\mathbb{R}^i} \mathcal{N}\left(x; \mu 1_i, \frac{1-\rho}{N} I_i\right) \mathcal{N}\left(y - x^\top \beta_1 - x_2^\top \beta_2; 0, 1/\mathrm{SNR}\right)^{\frac{1}{1+\delta}} \mathrm{d}x$$
$$= \left(\frac{1}{\sqrt{2\pi A}}\right)^i \left(\frac{\sqrt{\mathrm{SNR}}}{\sqrt{2\pi}}\right)^{\frac{1}{1+\delta}} \int_{\mathbb{R}^i} \exp\left(-\frac{(x-\mu 1_i)^\top (x - \mu 1_i)}{2A}\right) \exp\left(-\frac{(y - x^\top \beta_1 - x_2^\top \beta_2)^2}{2B}\right) \mathrm{d}x$$
$$= \left(\frac{1}{\sqrt{2\pi A}}\right)^i \left(\frac{\sqrt{\mathrm{SNR}}}{\sqrt{2\pi}}\right)^{\frac{1}{1+\delta}} \int_{\mathbb{R}^i} \exp\left(-\frac{x^\top x}{2A} - \frac{(x^\top \beta_1 + C)^2}{2B}\right) \mathrm{d}x$$
$$= \left(\frac{1}{\sqrt{2\pi A}}\right)^i \left(\frac{\sqrt{\mathrm{SNR}}}{\sqrt{2\pi}}\right)^{\frac{1}{1+\delta}} \int_{\mathbb{R}^i} \exp\left(-\frac{1}{2}(x + (BD)^{-1} AC\beta_1)^\top \frac{D}{A} (x + (BD)^{-1} AC\beta_1)\right) \exp\left(-\frac{C^2}{2E}\right) \mathrm{d}x$$

where $A = \frac{1-\rho}{N}$, $B = \frac{1+\delta}{\mathrm{SNR}}$, $C = x_2^\top \beta_2 + \mu 1_i^\top \beta_1 - y$, $D = I_i + \frac{A}{B}\beta_1 {\beta_1}^\top$ and $E = \frac{B}{1 - \frac{A}{B}{\beta_1}^\top D^{-1} \beta_1}$. Then taking the integral, some terms on the left cancel and we have

$$\int_{X_{\mathcal{S}^1}} P(X_{\mathcal{S}^1}|\theta) P(Y|X_{\mathcal{S}^1}, X_{\mathcal{S}^2}, \beta_S)^{\frac{1}{1+\delta}} \, \mathrm{d}X_{\mathcal{S}^1} = \left(\frac{\sqrt{\mathrm{SNR}}}{\sqrt{2\pi}}\right)^{\frac{1}{1+\delta}} \frac{1}{\sqrt{|D|}} \exp\left(-\frac{C^2}{2E}\right). \tag{27}$$

Writing the second integral that is over $X_{\mathcal{S}^2} = x_2$, we then have

$$\int_{X_{\mathcal{S}^2}} P(X_{\mathcal{S}^2}|\theta) \left[ \int_{X_{\mathcal{S}^1}} P(X_{\mathcal{S}^1}|\theta) P(Y|X_{\mathcal{S}^1}, X_{\mathcal{S}^2}, \beta_S)^{\frac{1}{1+\delta}} \, \mathrm{d}X_{\mathcal{S}^1} \right]^{1+\delta} \mathrm{d}X_{\mathcal{S}^2}$$
$$= \sqrt{\frac{\mathrm{SNR}}{2\pi}} \frac{1}{\sqrt{|D|}^{(1+\delta)}} \int_{\mathbb{R}^{K-i}} \mathcal{N}(x; \mu 1_{K-i}, A I_{K-i}) \exp\left(-\frac{(x^\top \beta_2 + \mu 1_i^\top \beta_1 - y)^2}{2E'}\right) \mathrm{d}x$$
$$= \sqrt{\frac{\mathrm{SNR}}{2\pi}} \frac{1}{\sqrt{|D|}^{(1+\delta)}} \left(\frac{1}{\sqrt{2\pi A}}\right)^{K-i} \int_{\mathbb{R}^{K-i}} \exp\left(-\frac{x^\top x}{2A} - \frac{(x^\top \beta_2 + F)^2}{2E'}\right) \mathrm{d}x$$
$$= \sqrt{\frac{\mathrm{SNR}}{2\pi}} \frac{1}{\sqrt{|D|}^{(1+\delta)}} \left(\frac{1}{\sqrt{2\pi A}}\right)^{K-i} \int_{\mathbb{R}^{K-i}} \exp\left(-\frac{1}{2}(x + (E'G)^{-1} AF\beta_2)^\top \frac{G}{A} (x + (E'G)^{-1} AF\beta_2)\right) \exp\left(-\frac{F^2}{2H}\right) \mathrm{d}x$$

where $E' = \frac{E}{1+\delta}$, $F = \mu 1_K^\top \beta_S - y$, $G = 1 + \frac{A}{E'}\beta_2 {\beta_2}^\top$ and $H = \frac{E'}{1 - \frac{A}{E'}{\beta_2}^\top G^{-1} \beta_2}$. Again, evaluating the integral, we obtain

$$\int_{X_{\mathcal{S}^2}} P(X_{\mathcal{S}^2}|\theta) \left[ \int_{X_{\mathcal{S}^1}} P(X_{\mathcal{S}^1}|\theta) P(Y|X_{\mathcal{S}^1}, X_{\mathcal{S}^2}, \beta_S)^{\frac{1}{1+\delta}} \, \mathrm{d}X_{\mathcal{S}^1} \right]^{1+\delta} \mathrm{d}X_{\mathcal{S}^2} = \sqrt{\frac{\mathrm{SNR}}{2\pi}} \frac{1}{\sqrt{|D|}^{(1+\delta)}} \frac{1}{\sqrt{|G|}} \exp\left(-\frac{F^2}{2H}\right).$$

Integrating the above expression w.r.t. $Y = y$, we see that the result is independent of $\theta = \mu$, and therefore

$$\int_\theta P(\theta)\int_Y\int_{X_{\mathcal{S}^2}} P(X_{\mathcal{S}^2}|\theta)\left[\int_{X_{\mathcal{S}^1}} P(X_{\mathcal{S}^1}|\theta)P(Y|X_{\mathcal{S}^1},X_{\mathcal{S}^2},\beta_S)^{\frac{1}{1+\delta}}\,\mathrm{d}X_{\mathcal{S}^1}\right]^{1+\delta}\mathrm{d}X_{\mathcal{S}^2}\,\mathrm{d}Y\,\mathrm{d}\theta = \frac{\sqrt{\mathrm{SNR}}}{\sqrt{|D|}^{(1+\delta)}}\sqrt{\frac{H}{|G|}}.$$

By the matrix determinant lemma, we have $|D| = 1 + \frac{A}{B}{\beta_1}^\top\beta_1$ and by the Sherman-Morrison formula, $D^{-1} = I_i - \frac{\beta_1{\beta_1}^\top}{\frac{B}{A}+{\beta_1}^\top\beta_1}$. Similarly, $|G| = 1 + \frac{A}{E'}{\beta_2}^\top\beta_2$ and $G^{-1} = I_i - \frac{\beta_2{\beta_2}^\top}{\frac{E'}{A}+{\beta_2}^\top\beta_2}$. By plugging in these expressions, we can then see that $E' = \frac{B|D|}{1+\delta}$ and $H = E'|G|$. We simplify the above expression to obtain

$$\frac{\sqrt{\mathrm{SNR}}}{\sqrt{|D|}^{(1+\delta)}}\sqrt{\frac{H}{|G|}} = \frac{\sqrt{\mathrm{SNR}}}{\sqrt{|D|}^{(1+\delta)}}\sqrt{\frac{B|D|}{1+\delta}} = \left(\frac{1}{\sqrt{|D|}}\right)^{\delta} = \left(1+(1-\rho)\frac{\mathrm{SNR}{\beta_1}^\top\beta_1}{N(1+\delta)}\right)^{-\frac{\delta}{2}}. \tag{28}$$

Note that this expression is analogous to the bound we obtained for the fixed case, since $E[{\beta_1}^\top\beta_1] = i\sigma^2$. With the above bound, we will now show a lower bound on $E_o(\delta)$ for $\delta = 1$ and $\sigma^2 = \frac{1}{8\pi}$. We note that we choose this $\sigma^2$ without loss of generality, since for any value or scaling of $\sigma$ can be incorporated into the SNR of the problem to obtain an equivalent model, such that $\mathrm{SNR}\sigma^2$ is fixed. This result can also be shown without the assumption on $\sigma^2$, but the specific bounding methods we use utilize this assumption. To this effect, we analyze the equivalent problem with parameters $\mathrm{SNR}' = \mathrm{SNR}\sigma^2 8\pi$ and $\sigma'^2 = \frac{1}{8\pi}$. Note that with this choice of $\sigma'^2$ and $\delta$, $M = \int_{\mathbb{R}^K} P(\beta_S)^{\frac{1}{2}}\,\mathrm{d}\beta_S = 1^K = 1$. Using (26), we now write,

$$\begin{aligned}
E_o(1) &\geq -\frac{1}{N}\log M\int_{\mathbb{R}^K} P(\beta_S)^{\frac{1}{2}}\left(1+(1-\rho)\frac{\mathrm{SNR}{\beta_1}^\top\beta_1}{2N}\right)^{-\frac{N}{2}}\mathrm{d}\beta_S\\
&= -\frac{1}{N}\log\int_{\mathbb{R}^i} P(\beta_1)^{\frac{1}{2}}\left(1+(1-\rho)\frac{\mathrm{SNR}'{\beta_1}^\top\beta_1}{2N}\right)^{-\frac{N}{2}}\mathrm{d}\beta_1\\
&= -\frac{1}{N}\log(\sqrt{4})^{\frac{i}{2}}\int_{\mathbb{R}^i}\exp\left[-\frac{{\beta_1}^\top\beta_1}{4\sigma'^2}\right]\left(1+(1-\rho)\frac{\mathrm{SNR}'{\beta_1}^\top\beta_1}{2N}\right)^{-\frac{N}{2}}\mathrm{d}\beta_1\\
&\geq -\frac{1}{N}\log(\sqrt{4})^{\frac{i}{2}}\left(\sqrt{8\pi\sigma'^2}\right)^{\frac{iN}{2}}\int_{\mathbb{R}^i}\left[\left(\frac{1}{\sqrt{8\pi\sigma'^2}}\right)^{i}\exp\left[-\frac{{\beta_1}^\top\beta_1}{8\sigma'^2}\right]\left(1+(1-\rho)\frac{\mathrm{SNR}'{\beta_1}^\top\beta_1}{2N}\right)\right]^{-\frac{N}{2}}\mathrm{d}\beta_1\\
&\geq -\frac{1}{N}\log 4^{\frac{i}{4}}\left[\int_{\mathbb{R}^i}\left(\frac{1}{\sqrt{8\pi\sigma'^2}}\right)^{i}\exp\left[-\frac{{\beta_1}^\top\beta_1}{8\sigma'^2}\right]\left(1+(1-\rho)\frac{\mathrm{SNR}'{\beta_1}^\top\beta_1}{2N}\right)\mathrm{d}\beta_1\right]^{-\frac{N}{2}}\\
&= \frac{1}{2}\log\left(1+(1-\rho)\frac{2i\mathrm{SNR}\sigma^2}{N}\right) - \frac{i}{4N}\log 4.
\end{aligned}$$

The first equality follows by taking $\beta_2$ out of the integral and noting that $\int_{\mathbb{R}^{K-i}} P(\beta_2)^{\frac{1}{2}}\,\mathrm{d}\beta_2 = 1^{K-i} = 1$. We obtain the second equality by expanding $P(\beta_1)^{\frac{1}{2}}$. We upper bound $\exp\left[-\frac{{\beta_1}^\top\beta_1}{8\sigma'^2}\right]$ by $\exp\left[-\frac{{\beta_1}^\top\beta_1}{8\sigma'^2}\right]^{-\frac{N}{2}}$ where $0 \leq \exp\left[-\frac{{\beta_1}^\top\beta_1}{8\sigma'^2}\right] \leq 1$ to obtain the first inequality and the second one follows by the superadditivity of exponentiating with $-\frac{N}{2}$. Finally, we note that the integral is an expectation w.r.t. $\beta_1 \sim \mathcal{N}(0, 4\sigma'^2 I_i)$ and obtain the last equality, where we also replace $\mathrm{SNR}'$ and $\sigma'^2$.

□

# 5 Proof of Lemma 3.1

Note the following equalities,

$$\begin{aligned}
I(X_{\mathcal{S}^1}; Y | X_{\mathcal{S}^2}, \beta_S, \mu) &= h(Y|X_{\mathcal{S}^2}, \beta_S, \mu) - h(Y|X_S, \beta_S, \mu) \\
&= h\left(X_{\mathcal{S}^1}^\top \beta_{\mathcal{S}^1} + W | \beta_{\mathcal{S}^1}, \mu\right) - h(W) \\
&= E_{\beta_{\mathcal{S}^1}, \mu}\left[\frac{1}{2}\ln\left(2\pi e\left(\operatorname{var}\left(X_{\mathcal{S}^1}^\top \beta_{\mathcal{S}^1} | \beta_{\mathcal{S}^1}, \mu\right)\ \beta_{\mathcal{S}^1}^\top \beta_{\mathcal{S}^1} + \frac{1}{\mathrm{SNR}}\right)\right)\right] - \frac{1}{2}\ln\left(2\pi e\,\frac{1}{\mathrm{SNR}}\right) \\
&= E_{\beta_{\mathcal{S}^1}}\left[\frac{1}{2}\ln\left(1 + (1-\rho)\frac{\beta_{\mathcal{S}^1}^\top \beta_{\mathcal{S}^1}\mathrm{SNR}}{N}\right)\right],
\end{aligned}$$

where the second equality follows from the independence of $X_{\mathcal{S}^1}$ and $X_{\mathcal{S}^2}$ given $\mu$ and the last equality follows from the fact that $\operatorname{var}(X_{\mathcal{S}^1}^\top \beta_{\mathcal{S}^1} | \beta_{\mathcal{S}^1}, \mu) = \beta_{\mathcal{S}^1}^\top E[U_{\mathcal{S}^1} U_{\mathcal{S}^1}^\top]\beta_{\mathcal{S}^1} = \beta_{\mathcal{S}^1}^\top \beta_{\mathcal{S}^1}\frac{1-\rho}{N}$.

# 6 Proof of Theorem 3.2

We first show that $\mathrm{SNR} = \log D$ is a necessary condition. For any $D$, $K$ or SNR assume $N$ is much larger such that

$$E\left[\ln\left(1 + (1-\rho)\frac{\beta_{\mathcal{S}^1}^\top \beta_{\mathcal{S}^1}\mathrm{SNR}}{N}\right)\right] \asymp E\left[(1-\rho)\frac{\beta_{\mathcal{S}^1}^\top \beta_{\mathcal{S}^1}\mathrm{SNR}}{N}\right] = (1-\rho)\frac{i\sigma^2\mathrm{SNR}}{N}.$$

Then the necessary condition given by Theorem 3.1 is

$$N > C \max_i \frac{\log\binom{D-K}{i}\binom{D}{i}}{(1-\rho)\frac{i\sigma^2\mathrm{SNR}}{N}}$$

which readily leads to the condition that

$$\mathrm{SNR} > C \max_i \frac{\log\binom{D-K}{i}\binom{D}{i}}{(1-\rho)i\sigma^2} \asymp \log D \tag{29}$$

for $\sigma$ constant.

From the upper bound given by Theorem 3.1, the sufficiency bound in Theorem 3.2 is obtained in a straightforward manner, by looking at conditions where $Nf(\rho)$ goes to infinity. So for each $i$, we have

$$P(E_i) \leq 2^{-\left(N\frac{1}{2}\log\left(1+(1-\rho)\frac{2i\sigma^2\mathrm{SNR}}{N}\right) - \frac{i}{4}\log 4 - \log\binom{D-K}{i}\binom{K}{i}\right)},$$

then, as $\log\binom{D-K}{i}\binom{K}{i} = \Theta(i\log(D/i))$ dominates $\frac{i}{4}\log 4$ we can see that the following is a sufficient condition on $N$ for exact support recovery:

$$N > (1+\epsilon)\max_{i=1,\ldots,K} \frac{2\log\binom{D-K}{i}\binom{K}{i}}{\log\left(1 + (1-\rho)\frac{2i\mathrm{SNR}\sigma^2}{N}\right)}. \tag{30}$$

Assume $\mathrm{SNR} > C\,\frac{\log D}{(1-\rho)\sigma^2}$. Also assume $N = \Omega\left(\frac{K\log(D/K)}{\log(1+(1-\rho)\sigma^2)}\right)$, as in the theorem statement. Then, the bound in (30) becomes

$$\max_{i=1,\ldots,K} \frac{2\log\binom{D-K}{i}\binom{K}{i}}{\log\left(1 + (1-\rho)\frac{2i\mathrm{SNR}\sigma^2}{N}\right)} \asymp \max_{i=1,\ldots,K} \frac{i\log(D/i)}{\log\left(1 + 2C\frac{i}{K}\log(1+(1-\rho)\sigma^2)\frac{\log D}{\log(D/K)}\right)},$$

where we assume $\sigma^2$ constant, w.l.o.g., since the scaling of elements of $\beta_S$ can instead be incorporated into SNR to obtain an equivalent model as we did in the proof of Theorem 3.1.

First, consider the case $K = o(D)$. Then the sufficient condition reduces to

$$N > \max_{i=1,\ldots,K} \frac{i \log D}{\log\left(1 + 2C\frac{i}{K}\log(1+(1-\rho)\sigma^2)\right)},$$

which, for the case $i = o(K)$ is

$$N > \frac{i \log D}{C\frac{i}{K}\left(\log(1+(1-\rho)\sigma^2)\right)} \asymp \frac{K\log(D/K)}{\log(1+(1-\rho)\sigma^2)},$$

which is satisfied for chosen $N$. For $i = \Theta(K)$, asymptotically, we have

$$N > \frac{K \log D}{\log\left(1+2C\log(1+(1-\rho)\sigma^2)\right)} \asymp \frac{K\log(D/K)}{\log\left(1+\log(1+(1-\rho)\sigma^2)\right)},$$

which is also satisfied by $N$.

Second, consider the case $K = \Theta(D)$. We then have the condition

$$N > \max_{i=1,\ldots,K} \frac{i \log(D/i)}{\log\left(1 + 2C\frac{i}{K}\log(1+(1-\rho)\sigma^2)\log D\right)},$$

which for $i = o(K)$, is asymptotically equivalent to

$$N > \frac{i \log D}{2C\frac{i}{K}\log(1+(1-\rho)\sigma^2)\log D} = \frac{K}{2C\log(1+(1-\rho)\sigma^2)} \asymp \frac{K\log(D/K)}{\log(1+(1-\rho)\sigma^2)}$$

which is satisfied for chosen $N$. For $i = \Theta(K)$, asymptotically we have the condition

$$N > \frac{K\log(D/K)}{\log\left(1+\log(1+(1-\rho)\sigma^2)\log D\right)},$$

which is also satisfied for chosen $N$.

The necessity bound is obtained by using the derived mutual information expression and looking at the case $i = K$. From Lemma 3.1, we have

$$I(X_{\mathcal{S}^1}; Y | X_{\mathcal{S}^2}, \beta_{\mathcal{S}}, \mu) \asymp E_{\beta_{\mathcal{S}^1}}\left[\log\left(1 + (1-\rho)\frac{\beta_{\mathcal{S}^1}^\top \beta_{\mathcal{S}^1}\mathrm{SNR}}{N}\right)\right],$$

which leads to the following necessary condition, as given by Theorem 2.2:

$$N \geq \max_{i=1,\ldots,K} \frac{\log\binom{D-K+i}{i}}{E_{\beta_{\mathcal{S}^1}}\left[\log\left(1 + (1-\rho)\frac{\beta_{\mathcal{S}^1}^\top \beta_{\mathcal{S}^1}\mathrm{SNR}}{N}\right)\right]}.$$

Note that $E_{\beta_{\mathcal{S}^1}}\left[\log\left(1 + (1-\rho)\frac{\beta_{\mathcal{S}^1}^\top \beta_{\mathcal{S}^1}\mathrm{SNR}}{N}\right)\right] \leq \log\left(E_{\beta_{\mathcal{S}^1}}\left[1 + (1-\rho)\frac{\beta_{\mathcal{S}^1}^\top \beta_{\mathcal{S}^1}\mathrm{SNR}}{N}\right]\right) = \log\left(1+(1-\rho)\frac{i\sigma^2\mathrm{SNR}}{N}\right)$ due to Jensen's inequality, therefore the following is also a necessary condition, where we consider only $i = K$:

$$N \geq \frac{\log\binom{D}{K}}{\log\left(1+(1-\rho)\frac{K\sigma^2\mathrm{SNR}}{N}\right)} \asymp \frac{K\log(D/K)}{\log\left(1+(1-\rho)\frac{K\sigma^2\mathrm{SNR}}{N}\right)}. \tag{31}$$

Assume $\mathrm{SNR} = \Theta(\log(D/K))$, which is given by (29) for $\sigma^2 = O(1)$ and $i = K$. It is then clear that (31) does not hold for $N = o(K\log(D/K))$, since $\frac{K\log(D/K)}{N} \geq \log\left(1+(1-\rho)\sigma^2\frac{K\log(D/K)}{N}\right)$ for $\sigma^2 = O(1)$. However for $N = \Omega(K\log(D/K))$, the condition (31) is

$$N = \Omega\left(\frac{K\log(D/K)}{\log\left(1+(1-\rho)\sigma^2\right)}\right),$$

which proves the lower bound in Theorem 3.2.

# 7 Proof of Theorem 3.3

To show the upper bound on error probability given in Theorem 3.3, we will write $P(Y|Z_{\mathcal{S}^1}, Z_{\mathcal{S}^2})$ and compute $E_o(\delta)$. For clarity, we consider fixed $\beta_S = \sigma$ as we did initially did in the proof of Theorem 3.1. Note that we can write

$$P(Y|Z_{\mathcal{S}^1}, Z_{\mathcal{S}^2}) = P(Y|Z_{\mathcal{S}^1}, Z_{\mathcal{S}^2}, \mu) = \int_{\mathbb{R}^K} P(Y|X_{\mathcal{S}^1}, X_{\mathcal{S}^2}) P(X_S|Z_S, \mu)\, \mathrm{d}X_S.$$

The first term is given by $\mathcal{N}(y - x^\top \beta_S; 0, 1/\mathrm{SNR})$ as before, for $Y = y$ and $X_S = x$. Let $\alpha = \frac{1}{1+\nu}$, then using the conditional probability of jointly Gaussian random vectors, we have

$$P(X_S = x|Z_S = z, \mu) = \mathcal{N}\left(x; (1-\alpha)\mu 1_K + \alpha z, \frac{1-\rho}{N}(1-\alpha) I_K\right),$$

then, considering only sums of $X_S$ and $Z_S$ as $x$ and $z$ since $\beta_S = \sigma$, as we did in the proof of Theorem 3.1, the integral is

$$\begin{aligned} P\left(Y = y \Big| \sum_{k\in S} Z_k = z, \mu\right) &= \int_{\mathbb{R}} \mathcal{N}(y - \sigma x; 0, 1/\mathrm{SNR})\, \mathcal{N}\left(x; (1-\alpha)K\mu + \alpha z, \frac{1-\rho}{N}(1-\alpha)K\right) \mathrm{d}x \\ &= \frac{1}{2\pi}\sqrt{\frac{1}{AB}} \int_{\mathbb{R}} \exp\left(-\frac{(x-C)^2}{2A}\right) \exp\left(-\frac{(y-\sigma x)^2}{2B}\right) \mathrm{d}x \\ &= \frac{1}{2\pi}\sqrt{\frac{1}{AB}} \int_{\mathbb{R}} \exp\left(-\frac{(x-G)^2}{2\frac{AB}{A\sigma^2+B}}\right) \exp\left(-\frac{(y-\sigma C)^2}{2(A\sigma^2+B)}\right) \mathrm{d}x \\ &= \frac{1}{2\pi}\sqrt{\frac{1}{AB}} \exp\left(-\frac{(y-\sigma C)^2}{2(A\sigma^2+B)}\right) \sqrt{\frac{AB}{2\pi(A\sigma^2+B)}} \\ &= \sqrt{\frac{1}{2\pi(A\sigma^2+B)}} \exp\left(-\frac{(y-\sigma C)^2}{2(A\sigma^2+B)}\right) \\ &= \mathcal{N}\left(y - \alpha\sigma z - \alpha(1-\alpha)\sigma K\mu; 0, \frac{1}{\mathrm{SNR}} + \frac{(1-\rho)(1-\alpha)K\sigma^2}{N}\right), \end{aligned}$$

where $A = \frac{1-\rho}{N}(1-\alpha)K$, $B = \frac{1}{\mathrm{SNR}}$, $C = (1-\alpha)\mu K + \alpha z$ and $G = \frac{A\sigma y + BC}{\sqrt{A\sigma^2+B}}$. The last equation follows through the steps used to show (27). For the first equality, we compute and replace the probability distributions w.r.t. sums $x$ and $z$. We obtain the third equality by rewriting the terms inside the exponentials to obtain a square term with $x$. Then, we take the second exponential outside the integral and compute the integral, which gives us the fourth equality. Note that $G$ does not affect the integration result. Finally in the last step we note that the resulting expression is a Gaussian distribution with the given form.

Note that the resulting probability distribution is the same as $P(Y|X_S)$ we used in the proof of Theorem 3.1 except a few differences: $\sigma$ is replaced with $\alpha\sigma$, $\frac{1}{\mathrm{SNR}}$ is replaced with $\frac{1}{\mathrm{SNR}} + \frac{(1-\rho)(1-\alpha)K\sigma^2}{N}$ and lastly there is an extra $(1-\alpha)\mu K$ term. This last term does not affect the resulting lower bound on the error exponent $E_o$, since it disappears in the integration over $Y$ like the other $\mu$ terms. We also note that $P(Z_{\mathcal{S}^1}|\mu)$ and $P(Z_{\mathcal{S}^2}|\mu)$ terms in the integral (26) are different than $P(X_{\mathcal{S}^1}|\mu)$ and $P(X_{\mathcal{S}^2}|\mu)$. To account for this difference, we need to replace the variance $\frac{1-\rho}{N}$ with $\frac{(1-\rho)(1+\nu)}{N}$ in the integrations w.r.t. $z_1$ and $z_2$ that follow.

Finally, by doing the necessary replacements outlined above and following the proof of Theorem 3.1, we obtain the following error exponent for fixed $\beta_S = \sigma$:

$$E_o(\delta) = \frac{\delta}{2}\log\left(1 + \frac{1-\rho}{1+\nu}\frac{i\sigma^2\mathrm{SNR}}{N(1+\delta)\xi}\right),$$

where $\xi = 1 + \frac{(1-\rho)\nu}{1+\nu}\frac{K\mathrm{SNR}\sigma^2}{N}$. Then the same analysis in the proof of Theorem 3.1 can be employed for random $\beta_S$, to obtain the lower bound,

$$E_o(1) \geq \frac{1}{2}\log\left(1 + \frac{1-\rho}{1+\nu}\frac{2i\sigma^2\mathrm{SNR}}{N\xi}\right) - \frac{i}{4N}\log 4,$$

which proves the upper bound on error probability given in Theorem 3.3.

# 8 Proof of Theorem 3.4

We analyze the upper bound given in Theorem 3.3 to obtain the sufficient condition on $N$. First, note that $\frac{(1-\rho)\nu}{1+\nu} \leq 1$, therefore $\xi \leq 1 + \frac{K\mathrm{SNR}\sigma^2}{N}$.

Let $\mathrm{SNR} = c\log(D/K)$ for now, which is more relaxed than the SNR condition we assume in the theorem and assume $N = \Omega\left(\frac{K\log(D/K)}{\log\left(1+\frac{1-\rho}{1+\nu}\sigma^2\right)}\right)$. Then it is easy to see that $\xi = O(1)$ for $\sigma^2 = O(1)$. As before, we can assume $\sigma^2 = O(1)$ w.l.o.g. since otherwise we can incorporate its scaling into SNR. Then for some constant $C > 0$, we have the lower bound

$$E_o(1) \geq \frac{1}{2}\log\left(1 + \frac{1-\rho}{1+\nu}\frac{2ci\sigma^2\mathrm{SNR}}{CN}\right) - \frac{i}{4N}\log 4,$$

and therefore we have

$$P(E_i) \leq 2^{-\left(N\frac{1}{2}\log\left(1+c'\frac{1-\rho}{1+\nu}\frac{i\sigma^2\mathrm{SNR}}{N}\right)-\frac{i}{4}\log 4-\log\binom{D-K}{i}\binom{K}{i}\right)},$$

for a constant $c' > 0$.

Following the arguments in the proof of Theorem 3.2, we can see that for $N$ chosen as above, $P(E)$ goes to zero, proving the theorem.

# References

[1] G. Atia and V. Saligrama. Boolean compressed sensing and noisy group testing. *IEEE Trans. Inf. Theory*, 58(3), March 2012.

[2] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* New York: John Wiley and Sons, Inc., 1991.

[3] R. G. Gallager. Information theory. In *Mathematics of Physics and Chemistry, Vol. 2.* Van Nostrand, Princeton, NJ, USA, 1964.

[4] R. G. Gallager. *Information Theory and Reliable Communication.* John Wiley & Sons, Inc., New York, NY, USA, 1968.