## A. Proof of Theorem 2

Before we presenting the main proof idea, we first analyze the moment characteristic of our proposed Wave Soliton Distribution. For simplicity, we use $d$ to denote $mn$ in the sequel.

**Lemma 1.** *Let a random variable $X$ follows the following Wave Soliton distribution $P_w = [p_1, p_2, \dots, p_d]$.*

$$p_k = \begin{cases} \frac{\tau}{d}, k = 1 \\ \frac{\tau}{70}, k = 2 \\ \frac{\tau}{k(k-1)}, 3 \le k \le d \end{cases}. \tag{1}$$

*Then all orders moment of $X$ is given by*

$$\mathbb{E}[X^s] = \begin{cases} \Theta\left(\tau \ln(d)\right), s = 1 \\ \Theta\left(\frac{\tau}{s} d^{s-1}\right), s \ge 2. \end{cases} \tag{2}$$

*Proof.* Based on the definition of moment for discrete random variable, we have

$$\mathbb{E}[X] = \sum_{k=1}^{d} k p_k = \frac{\tau}{d} + \frac{\tau}{35} + \sum_{k=3}^{d} \frac{\tau}{k-1} = \Theta\left(\tau \ln(d)\right). \tag{3}$$

Note that in the last step, we use the fact that $1 + 1/2 + \cdots + 1/d = \Theta(\ln(d))$.

$$\begin{aligned} \mathbb{E}[X^s] &= \sum_{k=1}^{d} k^s p_k = \frac{\tau}{d} + \frac{\tau 2^s}{70} + \sum_{k=3}^{d} \frac{\tau k^{s-1}}{k-1} \\ &\overset{(a)}{=} \Theta\left(\frac{\tau}{s} d^{s-1}\right). \end{aligned} \tag{4}$$

The step (a) uses the Faulhaber's formula that $\sum_{k=1}^{d} k^s = \Theta(d^{s+1}/(s+1))$. □

The technical idea in the proof of this theorem is to use the Hall's theorem. Assume that the bipartite graph $G(V_1, V_2, P_w)$ does not have a perfect matching. Then by Hall's condition, there exists a violating set $S \subseteq V_1$ or $S \subseteq V_2$ such that $|N(S)| < |S|$, where the neighboring set $N(S)$ is defined as $N(S) = \{y | (x, y) \in E(G) \text{ for some } x \in S\}$. Formally, by choosing such $S$ of smallest cardinality, one immediate consequence is the following technical statement.

**Lemma 2.** *If the bipartite graph $G(V_1, V_2, P_w)$ does not contain a perfect matching and $|V_1| = |V_1| = d$, then there exists a set $S \subseteq V_1$ or $S \subseteq V_2$ with the following properties.*

1. $|S| = |N(S)| + 1$.
2. *For each vertex $t \in N(S)$, there exists at least two adjacent vertices in $S$.*
3. $|S| \le d/2$.

Figure 1 illustrates two simple examples of structure $S$ satisfying above three conditions.

**Case 1:** We consider that $S \subseteq V_1$. Define an event $E(V_1)$ is that there exists a set $S \subseteq V_1$ satisfying above three conditions.

**Case 1.1:** We consider $S \subseteq V_1$ and $|S| = 1$.

In this case, we have $|N(S)| = 0$ and need to estimate the probability that there exists one isolated vertex in partition $V_1$. Let random variable $X_i$ be the indicator function of the event that vertex $v_i$ is isolated. Then we have the probability that

$$\mathbb{P}(X_i = 1) = \left(1 - \frac{\alpha}{d}\right)^d,$$

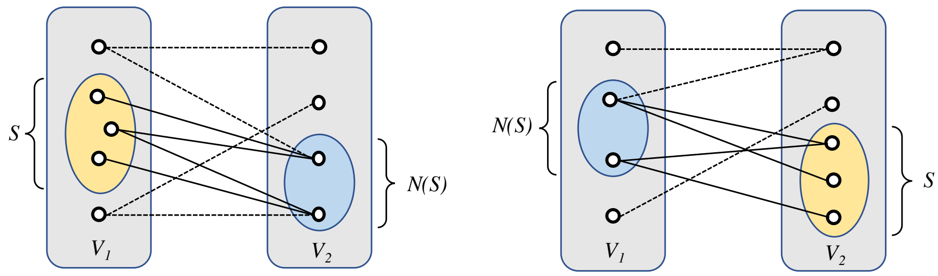


*Figure 1.* Example of structure $S \in V_1$, $N(S) \in V_2$ and $S \in V_2$, $N(S) \in V_1$ satisfying condition 1,2 and 3. One can easily check that there exists no perfect matching in these two examples.

where $\alpha$ is the average degree of a node in the partition $V_2$ and $\alpha = \Theta(\tau \ln(d))$ from Lemma 1. Let $X$ be the total number of isolated vertices in partition $V_1$. Then we have

$$\mathbb{E}[X] = E\left[\sum_{i=1}^{d} X_i\right] = d\left(1 - \frac{\alpha}{d}\right)^d = d\left(1 - \frac{\tau \ln(d)}{d}\right)^d = \Theta\left(\frac{1}{d^{\tau-1}}\right) \overset{(a)}{=} o(1). \tag{5}$$

The above, step (a) is based on the fact that $\tau > 1.94$.

Before we presenting the results in the case $2 \le |S| \le d/2$, we first define the following three events.

**Definition 1.** *Given a set $S \subseteq V_1$ and $|S| = s$, for each vertex $v \in V_2$, define an event $S_0^s$ is that $v$ has zero adjacent vertex in $S$, an event $S_1^s$ is that $v$ has one adjacent vertex in $S$ and an event $S_{\ge 2}^s$ is that $v$ has at least two adjacent vertices in $S$.*

Then we can upper bound the probability of event $E$ by

$$\begin{aligned}
\mathbb{P}(E(V_1)) &= \mathbb{P}(\text{there exists } S \in V_1 \text{ and } N(S) \in V_2 \text{ such that conditions } 1, 2 \text{ and } 3 \text{ are satisfied}) \\
&\overset{(a)}{\le} \mathbb{P}(\text{there exists an isoloted vertex in } V_1) + \sum_{s=2}^{d/2} \binom{d}{s}\binom{d}{s-1} \cdot \mathbb{P}(S_{\ge 2}^s)^{s-1} \cdot \mathbb{P}(S_0^s)^{d-s+1} \\
&= o(1) + \sum_{s=2}^{d/2} \binom{d}{s}\binom{d}{s-1} \cdot \mathbb{P}(S_{\ge 2}^s)^{s-1} \cdot \mathbb{P}(S_0^s)^{d-s+1}.
\end{aligned} \tag{6}$$

The above, step (a) is based on the union bound. Formally, given $|S| = s$ and fixed vertex $v \in V_2$, we can calculate $\mathbb{P}(S_0^s)$ via the law of total probability.

$$\begin{aligned}
\mathbb{P}(S_0^s) &= \sum_{k=1}^{d} \mathbb{P}(S_0^s | \deg(v) = k) \cdot \mathbb{P}(\deg(v) = k) \\
&= \sum_{k=1}^{d-s} p_k \binom{d-s}{k} \cdot \binom{d}{k}^{-1} \\
&= \sum_{k=1}^{d-s} p_k \left(1 - \frac{k}{d}\right)\left(1 - \frac{k}{d-1}\right)\left(1 - \frac{k}{d-2}\right) \cdots \left(1 - \frac{k}{d-s+1}\right).
\end{aligned} \tag{7}$$

Similarly, the probability $\mathbb{P}(S_1^s)$ is given by the following formula.

$$\begin{aligned}
\mathbb{P}(S_1^s) &= \sum_{k=1}^{d} \mathbb{P}(S_1^s | \deg(v) = k) \cdot \mathbb{P}(\deg(v) = k) \\
&= \sum_{k=1}^{d-s+1} p_k \binom{d-s}{k-1} \cdot \binom{s}{1} \cdot \binom{d}{k}^{-1}
\end{aligned}$$

$$= \sum_{k=1}^{d-s+1} p_k \left(1 - \frac{k}{d}\right)\left(1 - \frac{k}{d-1}\right)\cdots\left(1 - \frac{k}{d-s+2}\right) \cdot \frac{sk}{d-s+1}. \tag{8}$$

Then, the probability $\mathbb{P}(S^s_{\geq 2})$ is given by the following formula.

$$\mathbb{P}(S^s_{\geq 2}) = 1 - \mathbb{P}(S^s_0) - \mathbb{P}(S^s_1). \tag{9}$$

The rest is to utilize the formula (7), (8) and (9) to estimate the order of (6) under several scenarios.

**Case 1.2:** We consider $S \subseteq V_1$ and $|S| = \Theta(1)$.

Based on the result of (7), we have

$$\begin{aligned}
\mathbb{P}(S^s_0) &\leq \sum_{k=1}^{d} p_k \left(1 - \frac{k}{d}\right)^s = \sum_{k=1}^{d} p_k \sum_{i=0}^{s} \binom{s}{i} (-1)^i \frac{k^i}{d^i} \\
&\overset{(a)}{=} \sum_{k=1}^{d} p_k \left(1 - \frac{sk}{d}\right) + \sum_{i=2}^{s} \binom{s}{i} \frac{(-1)^i}{d^i} \sum_{k=1}^{d} k^i p_k \\
&\overset{(b)}{=} 1 - \Theta\left(\frac{s\tau \ln(d)}{d}\right) + \Theta\left(\frac{1}{d}\right) = 1 - \Theta\left(\frac{s\tau \ln(d)}{d}\right).
\end{aligned} \tag{10}$$

The above, step (a) is based on exchanging the order of summation; step (b) utilizes the result of Lemma 1 and the fact that $s$ is the constant. Similarly, we have

$$\mathbb{P}(S^s_{\geq 2}) \leq 1 - \mathbb{P}(S^s_0) \leq 1 - \sum_{k=1}^{d-s} p_k \left(1 - \frac{k}{d-s}\right)^s = \Theta\left(\frac{s\tau \ln(d)}{d}\right). \tag{11}$$

Combining the upper bound (6) and estimation of $\mathbb{P}(S^s_0)$ and $\mathbb{P}(S^s_{\geq 2})$, we have

$$\begin{aligned}
&\sum_{s=\Theta(1)} \binom{d}{s}\binom{d}{s-1} \cdot \mathbb{P}(S^s_{\geq 2})^{s-1} \cdot \mathbb{P}(S^s_0)^{n-s+1} \\
\overset{(a)}{\leq} &\sum_{s=\Theta(1)} \frac{e^{2s-1} d^{2s-1}}{s^{2s-1}} \cdot \Theta\left(\frac{s\tau \ln(d)}{d}\right)^{s-1} \cdot \left[1 - \Theta\left(\frac{s\tau \ln(d)}{d}\right)\right]^{d-s+1} \\
= &\sum_{s=\Theta(1)} \Theta\left(\frac{\ln^{s-1}(d)}{d^{(\tau-1)s}}\right) = o(1).
\end{aligned} \tag{12}$$

The above, step (a) utilizes the inequality $\binom{d}{s} \leq (ed/s)^s$.

**Case 1.3:** We consider $S \subseteq V_1$, $|S| = \Omega(1)$ and $|S| = o(d)$ .

Based on the result of (7), we have

$$\begin{aligned}
\mathbb{P}(S^s_0) &\leq \sum_{k=1}^{d} p_k \left(1 - \frac{k}{d}\right)^s \\
&\overset{(a)}{=} \sum_{ks=o(d)} p_k \left(1 - \frac{ks}{d}\right) + \sum_{ks=\Theta(d) \text{ or } \Omega(d)} p_k \left(1 - \frac{k}{d}\right)^s \\
&\leq \sum_{k=1}^{d/s} p_k \left(1 - \frac{ks}{d}\right) + \sum_{ks=\Theta(d) \text{ or } \Omega(d)} p_k \\
&\overset{(b)}{\leq} 1 - \frac{\tau(s-1)}{d} - \Theta\left(\frac{\tau s \ln(d/s)}{d}\right) + \Theta\left(\frac{\tau s}{d}\right)
\end{aligned}$$

$$\overset{(c)}{=} 1 - \Theta\left(\frac{\tau s \ln(d/cs)}{d}\right). \tag{13}$$

The above, step (a) is based on summation over different orders of $k$. In particular, when $ks = o(d)$ and $s = \Omega(1)$, we have $(1-k/d)^s = \Theta(e^{-ks/d}) = 1 - \Theta(ks/d)$. Step (b) utilizes the partial sum formula $1 + 1/2 + \cdots + s/d = \Theta(\ln(d/s))$. The parameter $c$ of step (c) is a constant. Similarly, we have

$$\begin{aligned}
\mathbb{P}(S^s_{\geq 2}) &\leq 1 - \mathbb{P}(S^s_0) \\
&\overset{(a)}{\leq} 1 - \left[\sum_{ks=o(d)} p_k \left(1 - \frac{ks}{d-s}\right) + \sum_{ks=\Theta(d), k \leq d/s-1} p_k e^{-\frac{ks}{d-s}}\right] \\
&\overset{(b)}{\leq} 1 - \sum_{k=1}^{d/s-1} p_k \left(1 - \frac{ks}{d-s}\right) \\
&\overset{(c)}{=} \Theta\left(\frac{\tau s \ln(d/c's)}{d}\right).
\end{aligned} \tag{14}$$

The above, step (a) is based on summation over different orders of $k$, and abandon the terms when $k \geq d/s$. In particular, when $ks = \Theta(d)$ and $s = \Omega(1)$, we have $(1 - k/(d-s))^s = \Theta(e^{-ks/(d-s)})$. The step (b) utilizes the inequality $e^{-x} \geq 1 - x, \forall x \geq 0$. The parameter $c'$ of step (c) is a constant. Combining the upper bound (6) and estimation of $\mathbb{P}(S^s_0)$ and $\mathbb{P}(S^s_{\geq 2})$, we have

$$\begin{aligned}
&\sum_{s=\Omega(1), s=o(d)} \binom{d}{s}\binom{d}{s-1} \cdot \mathbb{P}(S^s_{\geq 2})^{s-1} \cdot \mathbb{P}(S^s_0)^{d-s+1} \\
\leq &\sum_{s=\Omega(1), s=o(d)} \frac{e^{2s-1} d^{2s-1}}{s^{2s-1}} \cdot \Theta\left(\frac{s\tau \ln(d/c's)}{d}\right)^{s-1} \cdot \left[1 - \Theta\left(\frac{s\tau \ln(d/cs)}{d}\right)\right]^{d-s+1} \\
= &\sum_{s=\Omega(1), s=o(d)} \Theta\left(\left(\frac{s}{d}\right)^{(\tau-1)s} c^{\tau s} \tau^{s-1} e^{2s-1} \ln^{s-1}(d/c's)\right) \\
= &\sum_{s=\Omega(1), s=o(d)} \Theta\left(\frac{s \ln^{1.06}(d/s)}{d}\right)^{(\tau-1)s} = o(1).
\end{aligned} \tag{15}$$

**Case 1.4:** We consider $S \subseteq V_1$ and $|S| = \Theta(d) = cd$ .

Based on the result of (7) and Stirling's approximation, we have

$$\begin{aligned}
\mathbb{P}(S^s_0) &= \sum_{k=1}^{d-s} p_k \left(1 - \frac{k}{d}\right)^{d-k+\frac{1}{2}} \left(1 + \frac{k}{d-s-k}\right)^{d-s-k+\frac{1}{2}} (1-c)^k \\
&\overset{(a)}{=} \sum_{k=o(d)} p_k (1-c)^k + \sum_{k=\Theta(d), k \leq d-s} o\left((1-c)^k\right) \\
&= p_1(1-c) + p_2(1-c)^2 + \tau \sum_{k \geq 3, k=o(d)} \frac{(1-c)^k}{k(k-1)} + o(1) \\
&\overset{(b)}{=} p_1(1-c) + p_2(1-c)^2 + \tau\left[\frac{1}{2}(1-c^2) + c\ln(c)\right] \triangleq f_0(c).
\end{aligned} \tag{16}$$

The above, step (a) is based on summation over different orders of $k$. The step (b) is based on the following partial sum formula.

$$\sum_{k=3}^{q} \frac{(1-c)^k}{k(k-1)} = \frac{1}{2q}\left[2c(1-c)q(1-c)^q \Phi(1-c, 1, q+1) - 2(1-c)^{q+1} + q(1-c^2) + 2cq\ln(c)\right]. \tag{17}$$

where the function $\Phi(1-c,1,q+1)$ is the Lerch Transcendent, defined as

$$\Phi(1-c,1,q+1)=\sum_{k=0}^{\infty}\frac{c^k}{k+q+1}. \tag{18}$$

Let $q=\Omega(1)$, we arrive at the step (b). Similarly, utilizing the result of (8), we have

$$\begin{aligned}\mathbb{P}(S_1^s)&=\frac{c}{1-c}\sum_{k=o(d)}p_k k(1-c)^k\\&=p_1c+2p_2c(1-c)+\tau c(c-1-\ln(c))\triangleq f_1(c).\end{aligned} \tag{19}$$

Therefore, utilizing the upper bound (6), we arrive at

$$\begin{aligned}&\sum_{s=\Theta(d),s\le d/2}\binom{d}{s}\binom{d}{s-1}\cdot\mathbb{P}(S_{\ge 2}^s)^{s-1}\cdot\mathbb{P}(S_0^s)^{n-s+1}\\\le&\sum_{s=cd,s\le d/2}\left[\left(\frac{1}{c}\right)^{2c}\left(\frac{1}{1-c}\right)^{2(1-c)}[1-f_0(c)-f_1(c)]^c\,[f_0(c)]^{1-c}\right]^d\\=&\sum_{s=cd,s\le d/2}(1-\Theta(1))^d=o(1).\end{aligned} \tag{20}$$

Therefore, combining the results in the above four cases, we conclude that $\mathbb{P}(E(V_1))=o(1)$.

**Case 2:** We consider that $S\subseteq V_2$. We relax the condition 2 in Lemma 2 to the following condition.

$2'$. *For each vertex $t\in S$, there exists at least one adjacent vertex in $N(S)$.*

Define an event $E(V_2)$ is that there exists a set $S\subseteq V_2$ satisfying condition 1, 2, 3, and an event $E'$ is that there exists a set $S$ satisfying above condition $1,2'$ and 3. One can easily show that the event $E(V_2)$ implies the event $E'$ and $\mathbb{P}(E(V_2))\le\mathbb{P}(E')$. Then we aim to show that the probability of event $E'$ is $o(1)$.

**Definition 2.** *Given a set $S\subseteq V_2$ and $|S|=s$, for each vertex $v\in V_2$, define an event $N_{\ge 1}^s$ is that $v$ has at least one adjacent vertex in $N(S)$ and $v$ does not connect to any vertices in $V_1/N(S)$.*

Then we can upper bound the probability of event $E'$ by

$$\begin{aligned}\mathbb{P}(E')&=\mathbb{P}(\text{there exists } S\in V_2 \text{ and } N(S)\in V_1 \text{ such that condition } 1,2' \text{ and 3 are satisfied})\\&\overset{(a)}{\le}\sum_{s=2}^{d/2}\binom{d}{s}\binom{d}{s-1}\cdot\mathbb{P}(N_{\ge 1}^s)^s\\&\le\frac{e^{2s-1}d^{2s-1}}{s^{2s-1}}\cdot\mathbb{P}(N_{\ge 1}^s)^s.\end{aligned} \tag{21}$$

The above, step (a) is based on the fact that any vertices in set $V_2$ has degree at least one according to the definition of the Wave Soliton distribution. Given $|S|=s$ and fixed vertex $v\in S$, we can calculate $\mathbb{P}(N_{\ge 1}^s)$ via the law of total probability.

$$\begin{aligned}\mathbb{P}(N_{\ge 1}^s)&=\sum_{k=1}^{d}\mathbb{P}(N_{\ge 1}^s|\deg(v)=k)\cdot\mathbb{P}(\deg(v)=k)\\&=\sum_{k=1}^{s-1}p_k\binom{s-1}{k}\cdot\binom{d}{k}^{-1}\\&=\sum_{k=1}^{s-1}p_k\frac{s-1}{d}\cdot\frac{s-2}{d-1}\cdot\frac{s-3}{d-2}\cdot\dots\cdot\frac{s-k}{d-k+1}\\&\le\sum_{k=1}^{s-1}p_k\frac{s^k}{d^k}.\end{aligned} \tag{22}$$

**Case 2.1:** We consider $S \subseteq V_2$ and $|S| = \Theta(1)$.

Based on the result of (22), we have

$$\mathbb{P}(N_{\geq 1}^s) \leq \frac{\tau s}{d^2} + p_2 \frac{s^2}{d^2} + \sum_{k=3}^{s-1} p_k \frac{s^k}{d^k} \leq \frac{\tau s}{d^2} + p_2 \frac{s^2}{d^2} + \frac{\tau s^3}{d^3}\left(\frac{1}{2} - \frac{1}{s-1}\right) < \frac{s^2}{d^2}\left[\frac{1}{36} + \frac{\tau}{s} + \frac{\tau s}{d}\left(\frac{1}{2} - \frac{1}{s-1}\right)\right] \quad (23)$$

Then we have

$$\sum_{s=\Theta(1)} \frac{e^{2s-1} d^{2s-1}}{s^{2s-1}} \cdot \mathbb{P}(N_{\geq 1}^s)^s = \Theta\left(\frac{1}{d}\right). \quad (24)$$

**Case 2.2:** We consider $S \subseteq V_2$, $|S| = \Omega(1)$ and $|S| = o(d)$.

Similarly, using the result in Case 2.1 and upper bound (21), we arrive

$$\sum_{s=o(d)} \binom{d}{s}\binom{d}{s-1} \cdot \mathbb{P}(N_{\geq 1}^s)^s \leq \sum_{s=o(d)} \frac{se^{2s-1}}{d}\left[\frac{1}{36} + \frac{\tau}{s} + \frac{\tau s}{d}\left(\frac{1}{2} - \frac{1}{s-1}\right)\right]^s \overset{(a)}{=} o(1). \quad (25)$$

The above, step (a) is based on the fact that $1/36 + o(1) < e^{-2}$.

**Case 2.3:** We consider $S \subseteq V_2$ and $|S| = \Theta(d) = cd$. Based on the result of (22), we have

$$\mathbb{P}(N_{\geq 1}^s) \leq \sum_{k=1}^{s-1} p_k c^k \overset{(a)}{\leq} p_1 c + p_2 c^2 + \tau\left(c - \frac{c^2}{2} + (1-c)\ln(1-c)\right) \triangleq f_2(c). \quad (26)$$

The above, step (a) utilizes the partial sum formula (17). Using the upper bound (21), we arrive

$$\sum_{s=\Theta(d)} \binom{d}{s}\binom{d}{s-1} \cdot \mathbb{P}(N_{\geq 2}^s)^s = \left[\left(\frac{1}{c}\right)^{2c}\left(\frac{1}{1-c}\right)^{2(1-c)} [f_2(c)]^c\right]^d = \sum_{s=\Theta(d)} (1-\Theta(1))^d = o(1). \quad (27)$$

Combining the results in the above three cases, we have $\mathbb{P}(E') = o(1)$. Therefore, the theorem follows.

## B. Proof of Lemma 3

Consider a random bipartite graph generated by degree distribution $P_w$ of the nodes in left partition $V_2$, define a left edge degree distribution $\lambda(x) = \sum_k \lambda_k x^{k-1}$ and a right edge degree distribution $\rho(x) = \sum_k \rho_k x^{k-1}$, where $\lambda_k$ ($\rho_k$) is the fraction of edges adjacent to a node of degree $k$ in the left partition $V_1$ (right partition $V_2$). The existing analysis in (Luby et al., 2001) provides a quantitative condition regarding the recovery threshold in terms of $\lambda(x)$ and $\rho(x)$.

**Lemma 3.** *Let a random bipartite graph be chosen at random with left edge degree distribution $\lambda(x)$ and right edge degree distribution $\rho(x)$, if*

$$\lambda(1 - \rho(1-x)) < x, x \in [\delta, 1], \quad (28)$$

*then the probability that peeling decoding process cannot recover $\delta d$ or more of original blocks is upper bounded by $e^{-cd}$ for some constant $c$.*

We first derive the edge degree distributions $\lambda(x) = \sum_k \lambda_k x^{k-1}$ and $\rho(x) = \sum_k \rho_k x^{k-1}$ via the degree distribution $\Omega_w(x)$. Suppose that the recovery threshold is $K$. The total number of edges is $K\Omega_w'(1)$ and the total number of edges that is adjacent to a right node of degree $k$ is $Kkp_k$. Then, we have $\rho_k = kp_k/\Omega_w'(1)$ and

$$\rho(x) = \Omega_w'(x)/\Omega_w'(1). \quad (29)$$

Fix a node $v_i \in V_1$, the probability that node $v_i$ is a neighbor of node $v_j \in V_2$ is given by

$$\sum_{k=1}^{d} p_k \binom{d-1}{k-1}\binom{d}{k}^{-1} = \frac{1}{d}\sum_{k=1}^{d} kp_k = \frac{\Omega_w'(1)}{d}.$$

Since $|V_2| = K$, the probability that node $v_i$ is the neighbor of exactly $l$ nodes in $V_2$ is $\binom{K}{l}(\Omega_w'(1)/d)^l(1-\Omega_w'(1)/d)^{K-l}$ and corresponding probability generating function is

$$\sum_{l=1}^{K}\binom{K}{l}\left(\frac{\Omega_w'(1)}{d}\right)^l\left(1-\frac{\Omega_w'(1)}{d}\right)^{K-l}x^l = \left[1-\frac{\Omega_w'(1)(1-x)}{d}\right]^K.$$

Then we can obtain $\lambda(x)$ as

$$\lambda(x) = \left[1-\frac{\Omega_w'(1)(1-x)}{d}\right]^{K-1}. \tag{30}$$

Further, we have

$$\lambda(1-\rho(1-x)) = \left[1-\frac{\Omega_w'(1-x)}{d}\right]^{K-1}. \tag{31}$$

Combining these results with Lemma 3, let $\delta = b/mn$, the lemma follows.

## C. Proof of Theorem 3

Suppose that $K = cd+1$, one basic fact is that

$$\lambda(1-\rho(1-x)) = \left[1-\frac{\Omega_w'(1-x)}{d}\right]^{K-1} \leq e^{-c\Omega_w'(1-x)} \tag{32}$$

Based on the results of Lemma 3, the rest is to show that $e^{-c\Omega_w'(x)} \leq 1-x$ for $x \in [0, 1-b/d]$. Based on the definition of our Wave Soliton distribution, we have

$$\begin{aligned}\Omega_w'(x) &= \frac{\tau}{d} + \frac{\tau x}{35} + \tau\sum_{k=2}^{d-1}\frac{x^k}{k} \\ &= \frac{\tau}{d} - \frac{34\tau x}{35} - \tau\ln(1-x) - \tau\sum_{k=d}^{\infty}\frac{x^k}{k} \overset{(a)}{\geq} \frac{\tau}{d} - \frac{34\tau x}{35} - \tau\ln(1-x) - \tau x^{10}.\end{aligned} \tag{33}$$

The above step (a) is utilizing the fact that $x^{10} \geq \sum_{k=d}^{\infty}\frac{x^k}{k}$ for $x \in [0, 1-b/d]$. It remains to show that there exists a constant $c$ such that

$$-c\left[\frac{\tau}{d} - \frac{34\tau x}{35} - \tau\ln(1-x) - \tau x^{10}\right] \leq \ln(1-x),\ \text{for } x \in [0, 1-b/d]. \tag{34}$$

which is verified easily.

## D. Optimal Design of Sparse Code

We focus on determining the optimal degree distribution based on our previous analysis. Formally, we can formulate the following optimization problem.

$$\begin{aligned}\min\quad & \sum_{k=1}^{mn} kp_k \\ s.t.\quad & \mathbb{P}(M \text{ is full rank}) > p_c, \\ & \left[1-\tfrac{\Omega_w'(x)}{mn}\right]^{mn+c} \leq 1-x-c_0\sqrt{\tfrac{1-x}{mn}}, x \in [0, 1-b/mn], \\ & [p_k] \in \Delta_{mn},\end{aligned} \tag{35}$$

Here the coefficient matrix $M$ has row dimension $mn$ and column dimension $mn+c$. Due to the hardness of estimating the probability that $M$ is full rank, we relax this condition to the following condition

$$\mathbb{P}(G(V_1, V_2, P) \text{ contains a perfect matching}) > p_m, \tag{36}$$

*Table 1.* Optimizaed Degree Distribution for Various $mn$ (Numbers in brackets are Results from Robust Soliton Distribution.)

| $mn$ | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ | recovery threshold | average degree | rooting step |
|---|---|---|---|---|---|---|---|---|---|
| 6 | 0.0217 | 0.9390 | 0.0393 | 0.0 | 0.0 | 0.0 | 7.54 (7.61) | 2.01 (2.04) | 0.84 (0.47) |
| 9 | 0.0291 | 0.7243 | 0.2466 | 0.0 | 0.0 | 0.0 | 11.81 (12.15) | 2.21 (2.20) | 0.90 (0.56) |
| 12 | 0.0598 | 0.1639 | 0.7056 | 0.0707 | 0.0 | 0.0 | 14.19 (14.47) | 2.78 (2.78) | 1.47 (1.03) |
| 16 | 0.0264 | 0.3724 | 0.1960 | 0.4052 | 0.0 | 0.0 | 19.11 (19.83) | 2.98 (2.91) | 1.68 (1.08) |
| 25 | 0.0221 | 0.4725 | 0.1501 | 0.0 | 0.0 | 0.3553 | 28.71 (29.12) | 3.54 (3.55) | 2.35 (2.12) |

where $|V_1| = |V_2| = mn$ and $p_m$ is a given threshold. The basic intuition behind this relaxation comes from the analysis of Section 4.1: the probability that there exists a perfect matching in the balanced random bipartite graph $G(V_1, V_2, P)$ provides a lower bound of the probability that the coefficient matrix $M$ is of full rank. Based on this relaxation, the rest is to estimate the probability that $G(V_1, V_2, P)$ contains a perfect matching. Instead of estimating the lower bound of such a probability as in the proof of Theorem 2, here we provide an exact formula that is a function of the degree distribution $P$.

In the sequel, we denote $mn$ by $d$. Suppose the vertex in partition $V_2$ is denoted by $\{v_1, v_2, \ldots, v_d\}$. Define a degree distribution $P^{(s)} = [p_0^{(s)}, p_1^{(s)}, p_2^{(s)}, \ldots, p_d^{(s)}]$, where $p_k^{(s)}$ is the probability that any vertex $v \in V_2$ has exact $k$ adjacent vertices in a given set $S \subseteq V_1$ with $|S| = s$. For example, we have $P^{(d)} = P$, where $P$ is the original degree distribution. Let $E(d, V_1, P)$ be the event that $G(V_1, V_2, P)$ contains a perfect matching and $|V_1| = d$. In order to calculate $\mathbb{P}(E(d, V_1, P))$, we condition this probability on that there exists $v_1^n \in V_1$ matching with $v_1 \in V_2$. Then, we have

$$\begin{aligned}
\mathbb{P}(E(d, V_1, P)) =& \mathbb{P}(E(d, V_1, P) | \exists v_1^n \in V_1 \text{ matching with } v_1) \cdot \mathbb{P}(\exists v_1^n \in V_1 \text{ matching with } v_1) \\
=& \mathbb{P}(E(d-1, V_1 \backslash \{v_1^n\}, P^{(d-1)})) \cdot \left(1 - p_0^{(d)}\right) \\
=& \left(1 - p_0^{(d)}\right) \cdot \mathbb{P}(E(d-1, V_1 \backslash \{v_1^n\}, P^{(d-1)}) | \exists v_2^n \in V_1 \backslash \{v_1^n\} \text{ matching with } v_2) \\
& \qquad \mathbb{P}(\exists v_2^n \in V_1 \backslash \{v_1^n\} \text{ matching with } v_2) \\
=& \mathbb{P}(E(d-2, V_1 \backslash \{v_1^n, v_2^n\}, P^{(d-2)})) \left(1 - p_0^{(d-1)}\right) \cdot \left(1 - p_0^{(d)}\right) \\
=& \cdots \\
=& \left(1 - p_0^{(1)}\right) \left(1 - p_0^{(2)}\right) \cdots \left(1 - p_0^{(d-1)}\right) \left(1 - p_0^{(d)}\right).
\end{aligned} \tag{37}$$

The rest is to estimate the degree evolution $P^{(s)}$. Similarly, we have the following recursive formula to calculate the degree evolution: for all $0 \leq k \leq s, 1 \leq s \leq d-1$.

$$\begin{aligned}
p_k^{(s)} &= p_k^{(s+1)} \binom{s}{k} \binom{s+1}{k}^{-1} + p_{k+1}^{(s+1)} \binom{s}{k} \binom{s+1}{k+1}^{-1} \\
&= p_k^{(s+1)} \left(1 - \frac{k}{s+1}\right) + p_{k+1}^{(s+1)} \frac{k+1}{s+1}.
\end{aligned} \tag{38}$$

Utilizing the fact that $P^{(d)} = P$, we can get the exact formula of $P^{(s)}$, $1 \leq s \leq d-1$, and the formula of the probability that $G(V_1, V_2, P)$ contains a perfect matching.

TABLE 1 shows several optimized degree distributions we have found using model (35) with specific choice of parameters $c, c_0, b$ and $p_c$. We also include the several performance results under above distribution and the Robust Soliton distribution (RSD). In traditional RSD, the degree 2 always have the highest probability mass, i.e., $p_2 \approx 0.5$. It is interesting to note that our optimized distribution has a different shape, which depends on $mn$ and choices of parameters. We can observe that, under the same average degree, the optimized distribution has a lower recovery threshold and larger number of rooting steps compared to the RSD. Another observation in solving the optimization problem is that, when the parameter $p_m$ is increased, the recovery threshold of proposed sparse code will be decreased, and the average degree and the number of rooting steps will be increased.

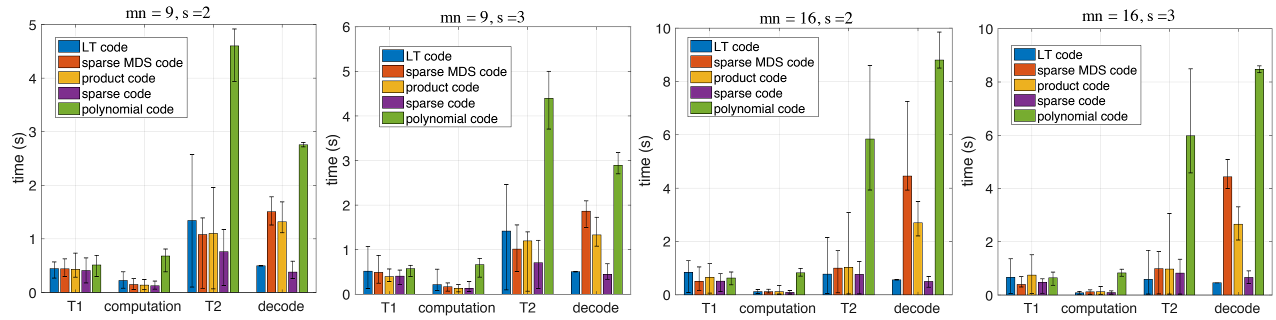


*Figure 2.* Simulation results for two $1.5E5 \times 1.5E5$ matrices with $6E5$ nonzero elements. T1 and T2 are the transmission times from master to worker, and the worker to master, respectively.

*Table 2.* Data Statistics for Different Input Matrix

| Data | $r$ | $s$ | $t$ | $nnz(A)$ | $nnz(B)$ | $nnz(C)$ |
|---|---|---|---|---|---|---|
| square | 1.5E5 | 1.5E5 | 1.5E5 | 6E5 | 6E5 | 2.4E6 |
| tall | 3E5 | 1.5E5 | 3E6 | 6E5 | 6E5 | 2.4E6 |
| fat | 1.5E5 | 3E5 | 1.5E5 | 6E5 | 6E5 | 1.2E6 |
| amazon-08 / web-google | 735320 | 735323 | 916428 | 5158379 | 4101329 | 28679400 |
| cont1 / cont11 | 1918396 | 1468599 | 1961392 | 2592597 | 5382995 | 10254724 |
| cit-patents / patents | 3774768 | 3774768 | 3774768 | 16518948 | 14970767 | 64796579 |
| hugetrace-00 / -01 | 4588484 | 4588484 | 12057440 | 13758266 | 13763443 | 38255405 |

## E. Extensive Simulation Results

We further compare our proposed sparse code with the existing schemes from the point of view of the time required to communicate inputs to each worker, compute the matrix multiplication in parallel, fetch the required outputs, and decode. As shown in Figure. 2, in all of these component times, the sparse code outperforms the product code and polynomial code, with the effects being more pronounced for the transmission time and the decoding time. Moreover, due to the efficiency of proposed hybrid decoding algorithm, our scheme achieves much less decoding time compared to the sparse MDS code and product code. Compared to the LT code, our scheme has lower transmission time because of the much lower recovery threshold. For example, when $m = n = 4$ and $s = 2$, the proposed sparse code requires $18$ workers, however, the LT code requires $24$ workers in average.

## References

Luby, Michael G, Mitzenmacher, Michael, Shokrollahi, Mohammad Amin, and Spielman, Daniel A. Efficient erasure correcting codes. *IEEE Transactions on Information Theory*, 47(2):569–584, 2001.