Forward-Backward Decoding Equations for A Class of Insertion Deletion and Substitution Channels

Javad Haghighat

In these notes, we present forward-backward (FB) decoding equations for a class of insertion, deletion, and substitution (IDS) channels characterized by the joint probability distribution of insertion events and deletion/substitution events. FB decoding has already been formulated in several works, including [1]. These notes simply present the equations in a slightly different manner. In the accompanying MATLAB simulations, we follow the notation introduced here.

We begin by presenting the system model in Section I and deriving some preliminary results in Section II. We present the main results in Section III for which the proofs are given in Section IV. The results are for q-array transmission when marker symbols are employed to maintain synchronization. In Section V we show how to obtain bit-level LLRs and also how to extend the derivations for the case when a watermark sequence is employed instead of marker symbols.

I. System Model

We consider transmission of T symbols, $x_1 ... x_T$ through an IDS channel, where $x_t \in \{0 : q-1\}$ is a realization of a random variable X_t and $\{0 : q-1\}$ denotes the set $\{0, \ldots, q-1\}$. For clarity of notation, we define $\mathcal{Q} = \{0 : q-1\}$. J out of the T transmitted symbols are marker symbols with known values $m_j \in \mathcal{Q}$ for $j \in \{1 : J\}$. Let $\tau_j \in \{1 : T\}$ denote the position of the jth marker symbol, and define $\mathcal{M} = \{\tau_1, \ldots, \tau_J\}$. Then, X_{τ_j} can be regarded as a degenerate random variable for which all of the probability is assigned to the single point, m_j, i.e., $P\left(X_{\tau_j} = m_j\right) = 1$. For non-marker positions, $t \in \{1 : T\} - \mathcal{M}$, X_t's are independent and identically distributed (i.i.d.) random variables with a probability distribution $P\left(X_t = a\right) = \rho_a$, $a \in \mathcal{Q}$. When we employ marker symbols, it is common practice to take a uniform distribution with $\rho_a = \frac{1}{q}$. However, we keep the general representation of ρ_a to facilitate extension of this model for systems that employ watermark sequences instead of marker symbols.

Before proceeding to describe the IDS channel model, let us introduce a few notation and definitions as follows:

For integers i, j, let X_i^j denote the vector $X_i, X_{i+1}, \ldots, X_j$. Also, let us generalize the definition and define X_i^j as an empty vector when $i > j$.

The binary random variable $\mathbf{1}_{\mathcal{E}}$ corresponding to an event $\mathcal{E}$ is defined such that its realization is 1 if $\mathcal{E}$ occurs and is 0 otherwise.

Let I_1^T be a set of i.i.d. random variables, where I_k takes non-negative integers as its realizations.

Let Z_1^T be a set of i.i.d. binary random variables, where Z_k takes 0 or 1 as its realization.

Define the joint probability distribution function of I_k and Z_k as:

$$\mu_{l,b} = P\left(I_k = l, Z_k = b\right) \tag{1}$$

for $l \geq 0$ and $b \in \{0, 1\}$.

Also define events $\Phi_{l,b}^k$ for $k \in \{1 : T\}$, such that:

$$\Phi_{l,b}^k : \text{ the event } I_k = l \,\text{and}\, Z_k = b \tag{2}$$

and define random variables N_k, $k \in \{1 : T\}$ where:

$$N_k = \sum_{t=1}^{k} \left(I_t + Z_t\right) \tag{3}$$

We extend the definition of N_k by defining N_0 as a degenerate random variable that takes value of 0 with probability one, i.e., $P\left(N_0 = 0\right) = 1$.

A. Channel Model

The IDS channel is modeled as follows. For an input symbol $x_k \in \mathcal{Q}$, the channel generates $l + b$ output symbols, where l is a realization of I_k and b is a realization of Z_k. The first l received symbols are referred to as inserted symbols where each inserted symbol is a realization of a random variable with a uniform distribution over $\mathcal{Q}$. When $b = 1$, the $(l + 1)$th received symbol is a noisy version of x_k after passing through a substitution channel that generates an output symbol a with a transition probability $f\left(x_k, a\right)$ defined as:

$$f(x_k, a) = \begin{cases} 1 - p_s; & a = x_k \\ \frac{p_s}{q-1}; & a \in \mathcal{Q} - \{x_k\} \end{cases} \tag{4}$$

On the other hand, when $b = 0$, x_k is deleted by the channel.

Assume that a sequence y_1^R is received after passing x_1^T through the IDS channel. According to the IDS channel model expressed above, y_n, $n \in \{1 : R\}$,is a realization of a random variable Y_n which can be expressed as:

$$Y_n = \left(1 - \sum_{t=1}^{T} \mathbf{1}_{Z_t N_t = n}\right) V_n \oplus \sum_{t=1}^{T} \left(\mathbf{1}_{Z_t N_t = n} (X_t \oplus S_t)\right) \tag{5}$$

where $\oplus$ denotes modulo-q addition, V_1^R are i.i.d. random variables with uniform distribution over $\mathcal{Q}$ and S_1^T are i.i.d. random variables such that $P(S_t = 0) = 1 - p_s$ and $P(S_t = a) = \frac{p_s}{q-1}$ for $a \in \mathcal{Q} - \{0\}$.

To verify (5), note that (i) $\mathbf{1}_{Z_k N_k = n} = 1$ if and only if y_n is received as a noisy version of x_k through a substitution channel. The reason is that $Z_k N_k = n$ occurs when $N_k = n$ and $Z_k = 1$, i.e., when the kth transmission interval ends at time step n and x_k is not deleted by the channel. (ii) There can be at most one k for which $\mathbf{1}_{Z_k N_k = n} = 1$[1].

Using (i), (ii), we may verify (5) as follows. (a) When $\sum_{k=1}^{T} \mathbf{1}_{Z_k N_k = n} = 0$, y_n is an inserted symbol and must be modeled as the realization of a uniformly distributed random variable. From (5) it is clear that when $\sum_{k=1}^{T} \mathbf{1}_{Z_k N_k = n} = 0$, $Y_n = V_n$ which does have a uniform distribution. Also, when there exists a k for which $\mathbf{1}_{Z_k N_k = n} = 1$, y_n is the output of the substitution channel with input x_k; hence, y_n can be modeled as a realization of $X_k \oplus S_k$. This completes verification of (5).

Before closing this section, let us note that the class of IDS channels modeled here by the joint probability distributions $\mu_{l,b} = P(I_k = l, Z_k = b)$, include existing channel models in the literature as special cases. For example, it is easy to verify that the Markov-like channel model with insertion, deletion and substitution probabilities p_i, p_d, p_s which is broadly used in the literature, is obtained by setting $\mu_{l,0} = p_i^l p_d$ and $\mu_{l,1} = p_i^l p_t$ for $l \geq 0$ (where it is easy to verify that by setting $p_t = 1 - p_i - p_d$, $\left(\sum_{l=0}^{\infty} \sum_{b=0}^{1} \mu_{l,b}\right) = 1$). Also, Gallager's model, which is another model employed in the literature, is found by setting $\mu_{0,0} = p_d$, $\mu_{2,0} = p_i$, $\mu_{0,1} = p_t$ and $\mu_{l,b} = 0$ otherwise. As another example, a deletion-

[1]this can be proved by contradiction. Assume that there are k, k' such that $\mathbf{1}_{Z_{k'} N_{k'} = n} = \mathbf{1}_{Z_k N_k = n} = 1$, then $N_{k'} = N_k = n$, and $Z_{k'} = Z_k = 1$. Without loss of generality, assume that $k' > k$. Then, $N_{k'} = N_k + \sum_{t=k+1}^{k'} (I_t + Z_t)$, and since $Z_{k'} = 1$, we will have $N_{k'} \geq N_k + 1$. This is a contradiction.

substitution channel is found by setting $\mu_{0,0} = p_d, \mu_{0,1} = 1 - p_d$, where p_d is the deletion probability. Furthermore, letting $p_s = 0$ gives a pure deletion channel for this example.

Also, it is reasonable to assume that there exists a value l_{max} such that $\mu_{l,b}$ becomes negligible for all $l \geq l_{max}$. For instance, for the Markov-like channel model where $\mu_{l,0} = p_i^l p_d$, and $\mu_{l,1} = p_i^l p_t$, assuming $p_i = 0.02$, we will have $\mu_{4,b} \leq 1.6 \times 10^{-7}$, hence, we may take $l_{max} = 4$. Therefore, in our derivations, we assume that $l \in \{0 : l_{max}\}$ for a reasonable value of l_{max}. Setting this bound is required to limit the number of summation terms and implement the FB algorithm in practice.

II. Preliminary Results

In this section we prove some properties that will help us derive our main results.

Let us define:

$$\mathcal{E}_i^j = \begin{cases} \text{the event that } Y_i^j = y_i^j & , i \leq j \\ \Omega & , i > j \end{cases} \tag{6}$$

where Ω denotes the universal event and the condition is justified by noting that when $i > j$, by definition Y_i^j, y_i^j are both empty vectors, therefore $P\left(Y_i^j = y_i^j\right) = 1$. This generalized definition of $\mathcal{E}_i^j$ will help us to express the forward-backward equations in a more concise manner.

Lemma 1. *For $l \geq 1$, given the events $\Phi_{l,b}^{k+1}$ and $N_k = n$:*

(i) In both cases of $b = 0$ and $b = 1$, for all $n' \in \{n+1 : n+l\}$ we have $\sum_{t=1}^{T} \mathbf{1}_{Z_t N_t = n'} = 0$, hence, $Y_{n+1}^{n+l} = V_{n+1}^{n+l}$.

(ii) In case of $b = 1$, we have $\mathbf{1}_{(Z_{k+1} N_{k+1} = n+l+1)} = 1$, hence, $Y_{n+l+1} = X_{k+1} \oplus S_{k+1}$

Proof: For $n' \in \{n+1 : n+l\}$ and $t \in \{1 : T\}$, we have:

$$\begin{aligned} & P\left(Z_t N_t = n' | \Phi_{l,b}^{k+1}, N_k = n\right) \\ &= P\left(Z_t = 1, N_t = n' | \Phi_{l,b}^{k+1}, N_k = n\right) \\ &= P\left(N_t = n' | N_{k+1} = n + l + b, N_k = n, Z_{k+1} = b\right) \times P\left(Z_t = 1 | \Phi_{l,b}^{k+1}, N_k = n, N_t = n'\right) \\ &= 0 \end{aligned}$$

The reason is that for an $n' \in \{n+1 : n+l\}$, given $N_{k+1} = n + l + b, N_k = n$, the event $N_t = n'$ occurs only when $t = k+1, b = 0$. However, given $\Phi_{l,0}^{k+1}$, we have $Z_{k+1} = 0$, hence:

$$P\left(Z_{k+1}=1|\Phi_{l,0}^{k+1},N_k=n,N_t=n'\right)=0$$

Therefore, $Z_tN_t=n'$ is the impossible event and $\mathbf{1}_{Z_tN_t=n'}=0$ for all t.

When $b=1$, given $\Phi_{l,1}^{k+1}$ and $N_k=n$ we have $N_{k+1}=n+l+1$ and $Z_{k+1}=1$, hence $\mathbf{1}_{(Z_{k+1}N_{k+1}=n+l+1)}=1$. ■

Corollary 2. *For $a\in\mathcal{Q}$ define:*

$$\zeta\left(k,a\right)=\sum_{a'=0}^{q-1}P\left(X_k=a'\right)\times f\left(a',a\right) \tag{7}$$

Then for $l\geq 0$ and $b\in\{0,1\}$:

$$P\left(\mathcal{E}_{n+1}^{n+l+b}|\Phi_{l,b}^{k+1},N_k=n,X_{k+1}=a\right)=\left(\frac{1}{q}\right)^l\left(f\left(a,y_{n+l+1}\right)\right)^b \tag{8}$$

$$P\left(\mathcal{E}_{n+1}^{n+l+b}|\Phi_{l,b}^{k+1},N_k=n\right)=\left(\frac{1}{q}\right)^l\left(\zeta\left(k+1,y_{n+l+1}\right)\right)^b \tag{9}$$

Proof: When $l=b=0$, by definition, $\mathcal{E}_{n+1}^{n+l+b}$ is the universal event, hence the probabilities are 1 and the equalities hold. For $l\geq 1$ and $b=0$, from Lemma 1 it is clear that $Y_{n+1}^{n+l}=V_{n+1}^{n+l}$, hence $\mathcal{E}_{n+1}^{n+l}$ is the event that $V_{n+1}^{n+l}=y_{n+1}^{n+l}$ which occurs with probability $\left(\frac{1}{q}\right)^l$. Also, when $b=0$, $\left(f\left(a,y_{n+l+1}\right)\right)^b=1$, hence, (8) holds.

For $b=1$, from Lemma 1 we have $Y_{n+1}^{n+l}=V_{n+1}^{n+l}$ and $Y_{n+l+1}=X_{k+1}\oplus S_{k+1}$. Therefore:

$$P\left(\mathcal{E}_{n+1}^{n+l+b}|\Phi_{l,b}^{k+1},N_k=n,X_k=a\right)=P\left(Y_{n+1}^{n+l}=V_{n+1}^{n+l}\right)\times P\left(X_{k+1}\oplus S_{k+1}=y_{n+l+1}|X_k=a\right)$$

where $P\left(Y_{n+1}^{n+l}=V_{n+1}^{n+l}\right)=\left(\frac{1}{q}\right)^l$ and it is straightforward to show that:

$$P\left(X_{k+1}\oplus S_{k+1}=y_{n+l+1}|X_{k+1}=a\right)=P\left(S_{k+1}=a\oplus y_{n+l+1}\right)=f\left(a,y_{n+l+1}\right)$$

Hence, (8) holds.

Equation (9) is derived by summing over all possible realizations of X_{k+1}, and using the definition of function $\zeta\left(.,.\right)$. ■

Lemma 3. *Given $N_k = n$, for all $l \geq 0$,* and $b \in \{0,1\}$, and $a \in \mathcal{Q}$ *we have:*

(i) $\mathcal{E}_1^n$ and $\mathcal{E}_{n+1}^R$ are independent,

(ii) $\mathcal{E}_1^n$ is independent of $\Phi_{l,b}^{k+1}$ and $X_{k+1} = a$,

(iii) $\mathcal{E}_{n+1}^R$ is independent of $\Phi_{l,b}^k$ and $X_k = a$.

Proof: For all $t > k$, and all $n' \in \{1 : n\}$:

$$P\left(Z_t N_t = n' | N_k = n\right) = P\left(N_t = n' | N_k = n\right) \times P\left(Z_t = 1 | N_t = n', N_k = n\right) = 0$$

where the equality is proved by noting that $P\left(N_t = n' | N_k = n\right) \neq 0$ only when $n' = n$; however, given $N_t = n$ and $N_k = n$ we have $Z_t = 0$.

Therefore, for any $n' \in \{1 : n\}$ we have:

$$Y_{n'} = \left(1 - \sum_{t=1}^{k} \mathbf{1}_{Z_t N_t = n'}\right) V_{n'} \oplus \sum_{t=1}^{k} \left(\mathbf{1}_{Z_t N_t = n'} \times (X_t \oplus S_t)\right) \tag{10}$$

where $N_t = \sum_{t'=1}^{t} (I_{t'} + Z_{t'})$. Thus, realizations of Y_1^n are determined by realizations of $V_1^n, Z_1^k, I_1^k, X_1^k, S_1^k$.

On the other hand, for all $n'' \in \{n+1 : R\}$ and all $t \leq k$, we have $P\left(N_t = n'' | N_k = n\right) = 0$, therefore,$\mathbf{1}_{Z_t N_t = n''} = 0$ and we have:

$$Y_{n''} = \left(1 - \sum_{t=k+1}^{T} \mathbf{1}_{Z_t N_t = n''}\right) V_{n''} \oplus \sum_{t=k+1}^{T} \left(\mathbf{1}_{Z_t N_t = n''} \times (X_t \oplus S_t)\right) \tag{11}$$

for all $n'' \in \{n+1 : R\}$.

Furthermore, given $N_k = n$, for all $t \in \{k+1 : T\}$, N_t is expressed as:

$$N_t = n + \sum_{t'=k+1}^{T} (I_{t'} + Z_{t'})$$

Therefore, given $N_k = n$, realizations of Y_{n+1}^R are determined by realizations of $V_{n+1}^R, Z_{k+1}^T, I_{k+1}^T, X_{k+1}^T, S_{k+1}^T$. Hence, the events $Y_1^n = y_1^n$ and $Y_{n+1}^R = y_{n+1}^R$ which are denoted as $\mathcal{E}_1^n$ and $\mathcal{E}_{n+1}^R$, are independent, and also $\mathcal{E}_1^n$ is independent of $X_{k+1} = a$, and $\mathcal{E}_{n+1}^R$ is independent of $X_k = a$.

To show the independence of $\Phi_{l,b}^{k+1}$ and $\mathcal{E}_1^n$, we note that $\Phi_{l,b}^{k+1}$occurs when $I_{k+1} = l, Z_{k+1} = b$; i.e., its occurrence does not depend on realizations of $V_1^n, Z_1^k, I_1^k, X_1^k, S_1^k$. Similarly, to show the independence of $\Phi_{l,b}^k$ and $\mathcal{E}_{n+1}^R$, we note that $\Phi_{l,b}^k$occurs when $I_k = l, Z_k = b$; i.e., its occurrence does not depend on

realizations of $V_{n+1}^R, Z_{k+1}^T, I_{k+1}^T, X_{k+1}^T, S_{k+1}^T$. This completed the proof. ■

Corollary 4. *Let $i, j \in \{1:n\}$, $i', j' \in \{n+1:R\}$, $i \leq j$, and $i' \leq j'$. Then, given $N_k = n$, $\mathcal{E}_i^j$ and $\mathcal{E}_{i'}^{j'}$ are independent events.*

Proof: In Lemma 3 we showed that given $N_k = n$, realizations of Y_1^n (which includes Y_i^j) are determined by realizations of $V_1^n, Z_1^k, I_1^k, X_1^k, S_1^k$; whereas realizations of Y_{n+1}^R (which includes $Y_{i'}^{j'}$) are determined by realizations of $V_{n+1}^R, Z_{k+1}^T, I_{k+1}^T, X_{k+1}^T, S_{k+1}^T$. This completes the proof. ■

III. Main Results

Let us define:

$$\alpha\left(k,n\right) = P\left(\mathcal{E}_1^n, N_k = n\right) \tag{12}$$

$$\beta\left(k,n,R\right) = P\left(\mathcal{E}_{n+1}^R | N_k = n\right) \tag{13}$$

Then, for $a \in \mathcal{Q}$:

$$\begin{aligned} &P\left(Y_1^R = y_1^R | X_k = a\right) = \\ &\sum_{n=0}^{U_{k-1}} \sum_{l=0}^{l_{max}} \sum_{b=0}^{1} \frac{\mu_{l,b}}{q^l} \left(f\left(a, y_{n+l+1}\right)\right)^b \alpha\left(k-1,n\right) \beta\left(k, n+l+b, R\right) \end{aligned} \tag{14}$$

where $U_{k-1} = \min(R, (1+l_{max})(k-1))$.

Furthermore, $\alpha\left(k,n\right)$ and $\beta\left(k,n,R\right)$ satisfy the recursive equations:

$$\alpha\left(k,n\right) \;\; = \sum_{l=0}^{l_{max}} \sum_{b=0}^{1} \frac{\mu_{l,b}}{q^l} \left(\zeta\left(k, y_n\right)\right)^b \alpha\left(k-1, n-l-b\right) \tag{15}$$

$$\beta\left(k,n,R\right) = \sum_{l=0}^{l_{max}} \sum_{b=0}^{1} \frac{\mu_{l,b}}{q^l} \left(\zeta\left(k+1, y_{n+l+1}\right)\right)^b \beta\left(k+1, n+l+b, R\right) \tag{16}$$

and the initial conditions:

$$\begin{cases} \alpha(0,0) = 1 \\ \alpha(0,n) = 0 & ,n \neq 0 \\ \alpha(k,0) = \mu_{0,0}^{k} & ,k \geq 1 \\ \alpha(k,n) = 0 & ,k \geq 0, n < 0 \end{cases} \tag{17}$$

$$\begin{cases} \beta(T,R,R) = 1 \\ \beta(T,n,R) = 0 & ,n \neq R \\ \beta(k,R,R) = \mathbf{1}_{\mu_{0,0}>0} & ,k < T \\ \beta(k,n,R) = 0 & ,k \leq T, n > R \end{cases} \tag{18}$$

We derive these results in the next section. Also, a MATLAB code simulating this decoding algorithm is provided in the appendix.

IV. Derivation of main results

In this section, we provide proofs for equations (14) through (18).

A. Derivation of (14)

Let n_k be a realization of N_k. Since $l \leq l_{max}$ and $b \leq 1$, n_k cannot exceed $(l_{max}+1)\,k$. Also, since n_k is the total number of symbols received by transmission of x_1^k, and it is assumed that a total of R symbols are received, then $n_k \leq R$. Therefore, if we define $U_k = \min(R, (l_{max}+1)\,k)$, we have $n_k \in \{0 : U_k\}$.

Now, since the event $Y_1^R = y_1^R$ is denoted by $\mathcal{E}_1^R$, we write:

$$P\left(Y_1^R = y_1^R | X_k = a\right) = \sum_{n=0}^{U_{k-1}} \sum_{l=0}^{l_{max}} \sum_{b=0}^{1} P\left(\mathcal{E}_1^R, N_{k-1} = n, \Phi_{l,b}^k | X_k = a\right) \tag{19}$$

where:

$$
\begin{aligned}
&P\left(\mathcal{E}_1^R, N_{k-1}=n, \Phi_{l,b}^k | X_k=a\right) \\
&= P\left(\mathcal{E}_1^n, \mathcal{E}_{n+1}^{n+l+b}, \mathcal{E}_{n+l+b+1}^R, N_{k-1}=n, \Phi_{l,b}^k | X_k=a\right) \\
&= P\left(\mathcal{E}_1^n, N_{k-1}=n | X_k=a\right) \\
&\times P\left(\mathcal{E}_{n+1}^{n+l+b}, \Phi_{l,b}^k | X_k=a, \mathcal{E}_1^n, N_{k-1}=n\right) \\
&\times P\left(\mathcal{E}_{n+l+b+1}^R | \mathcal{E}_1^{n+l+b}, N_{k-1}=n, \Phi_{l,b}^k, X_k=a\right)
\end{aligned} \tag{20}
$$

Now, we note that (i) $N_{k-1}=n$ and $X_k=a$ are independent, and (ii) given $N_{k-1}=n$, $\mathcal{E}_1^n$ and $X_k=a$ are independent (Lemma 3). Therefore, by applying the chain rule, we obtain:

$$
\begin{aligned}
&P\left(\mathcal{E}_1^n, N_{k-1}=n | X_k=a\right) = P\left(N_{k-1}=n | X_k=a\right) \times P\left(\mathcal{E}_1^n | N_{k-1}=n, X=a\right) \\
&= P\left(N_{k-1}=n\right) P\left(\mathcal{E}_1^n | N_{k-1}=n\right) = P\left(\mathcal{E}_1^n, N_{k-1}=n\right) \\
&= \alpha\left(k-1, n\right)
\end{aligned} \tag{21}
$$

Also, we note that (i) given $N_{k-1}=n$, the events $\Phi_{l,b}^k$ and $\mathcal{E}_1^n$ are independent (Lemma 3); and by definition, $\Phi_{l,b}^k$ is independent of $N_{k-1}=n$ and $X_k=a$ (ii) given $N_{k-1}=n$, $\mathcal{E}_{n+1}^{n+l+b}$ and $\mathcal{E}_1^n$ are independent (Corollary 4). Therefore, we may apply the chain rule to obtain:

$$
\begin{aligned}
&P\left(\mathcal{E}_{n+1}^{n+l+b}, \Phi_{l,b}^k | X_k=a, \mathcal{E}_1^n, N_{k-1}=n\right) \\
&= P\left(\Phi_{l,b}^k | X_k=a, \mathcal{E}_1^n, N_{k-1}=n\right) \\
&\times P\left(\mathcal{E}_{n+1}^{n+l+b} | \Phi_{l,b}^k, X_k=a, \mathcal{E}_1^n, N_{k-1}=n\right) \\
&= P\left(\Phi_{l,b}^k\right) P\left(\mathcal{E}_{n+1}^{n+l+b} | \Phi_{l,b}^k, X_k=a, N_{k-1}=n\right) \\
&= \mu_{l,b} \times \left(\tfrac{1}{q}\right)^l \left(f\left(a, y_{n+l+1}\right)\right)^b
\end{aligned} \tag{22}
$$

where the fourth line is obtained using (i), (ii), and the fifth line is obtained using Corollary 2.

Also, the condition $N_{k-1}=n, \Phi_{l,b}^k$, can be written as $N_k=n+l+b$, $\Phi_{l,b}^k$. Furthermore, given $N_k=n+l+b$, the event $\mathcal{E}_{n+l+b+1}^R$ is independent of the events $\mathcal{E}_1^{n+l+b}$, $\Phi_{l,b}^k$, $X_k=a$ (Lemma 3). Therefore:

$$
\begin{aligned}
&P\left(\mathcal{E}_{n+l+b+1}^R | \mathcal{E}_1^{n+l+b}, N_{k-1}=n, \Phi_{l,b}^k, X_k=a\right) \\
&= P\left(\mathcal{E}_{n+l+b+1}^R | N_k=n+l+b\right) = \beta\left(k, n+l+b, R\right)
\end{aligned} \tag{23}
$$

By plugging (21), (22), (23) into (20) and then plugging the result into (19), we obtain (14). This completes the proof.

B. Derivation of (15)

We write:

$$\alpha\left(k,n\right)=\sum_{l=0}^{l_{max}}\sum_{b=0}^{1}P\left(\mathcal{E}_{1}^{n},N_{k}=n,\Phi_{l,b}^{k}\right) \tag{24}$$

Where:

$$P\left(\mathcal{E}_{1}^{n},N_{k}=n,\Phi_{l,b}^{k}\right)=P\left(\mathcal{E}_{1}^{n-l-b},\mathcal{E}_{n-l-b+1}^{n},N_{k-1}=n-l-b,\Phi_{l,b}^{k}\right)$$

Now, we note that (i) $N_{k-1}=n-l-b$ is independent of $\Phi_{l,b}^{k}$ and given $N_{k-1}=n-l-b$, $\mathcal{E}_{1}^{n-l-b}$ is independent of $\Phi_{l,b}^{k}$ (Lemma 3). Therefore, the joint event $\mathcal{E}_{1}^{n-l-b},N_{k-1}=n-l-b$ is independent of $\Phi_{l,b}^{k}$. (ii) Given $N_{k-1}=n-l-b$, the event $\mathcal{E}_{n-l-b+1}^{n}$ is independent of $\mathcal{E}_{1}^{n-l-b}$ (Lemma 3). Using (i), (ii), we may employ the chain rule to write:

$$\begin{aligned}
&P\left(\mathcal{E}_{1}^{n-l-b},\mathcal{E}_{n-l-b+1}^{n},N_{k-1}=n-l-b,\Phi_{l,b}^{k}\right)\\
&=P\left(\Phi_{l,b}^{k}\right)\times P\left(\mathcal{E}_{1}^{n-l-b},N_{k-1}=n-l-b|\Phi_{l,b}^{k}\right)\\
&\times P\left(\mathcal{E}_{n-l-b+1}^{n}|\mathcal{E}_{1}^{n-l-b},N_{k-1}=n-l-b,\Phi_{l,b}^{k}\right)\\
&=\mu_{l,b}\times P\left(\mathcal{E}_{1}^{n-l-b},N_{k-1}=n-l-b\right)\times P\left(\mathcal{E}_{n-l-b+1}^{n}|N_{k-1}=n-l-b,\Phi_{l,b}^{k}\right)\\
&=\mu_{l,b}\times\alpha\left(k-1,n-l-b\right)\times\left(\tfrac{1}{q}\right)^{l}\left(\zeta\left(k,y_{n}\right)\right)^{b}
\end{aligned} \tag{25}$$

where the fourth line is found using (i), (ii), and the fifth line is found by employing Corollary 2. By plugging (25) into (24) we obtain (15).

C. Derivation of (16)

We write:

$$\beta\left(k,n,R\right)=\sum_{l=0}^{l_{max}}\sum_{b=0}^{1}P\left(\mathcal{E}_{n+1}^{R},\Phi_{l,b}^{k+1}|N_{k}=n\right) \tag{26}$$

where:

$$P\left(\mathcal{E}_{n+1}^{R},\Phi_{l,b}^{k+1}|N_{k}=n\right)=P\left(\mathcal{E}_{n+1}^{n+l+b},\mathcal{E}_{n+l+b+1}^{R},\Phi_{l,b}^{k+1}|N_{k}=n\right)$$

Now, we note that (i) $\Phi_{l,b}^{k+1}$ is independent of $N_k = n$, (ii) the joint event $\Phi_{l,b}^{k+1}, N_k = n$ can be rewritten as the joint event $\Phi_{l,b}^{k+1}, N_{k+1} = n + l + b$. Furthermore, given $N_{k+1} = n + l + b$, the event $\mathcal{E}_{n+l+b+1}^{R}$ is independent of $\mathcal{E}_{n+1}^{n+l+b}$ and $\Phi_{l,b}^{k+1}$ (Corollary 4).

We may apply the chain rule along with (i), (ii) and the result of Corollary 2 to obtain:

$$\begin{aligned}
&P\left(\mathcal{E}_{n+1}^{n+l+b}, \mathcal{E}_{n+l+b+1}^{R}, \Phi_{l,b}^{k+1} | N_k = n\right) \\
&= P\left(\Phi_{l,b}^{k+1} | N_k = n\right) \times P\left(\mathcal{E}_{n+1}^{n+l+b} | \Phi_{l,b}^{k+1}, N_k = n\right) \\
&\times P\left(\mathcal{E}_{n+l+b+1}^{R} | N_{k+1} = n + l + b, \Phi_{l,b}^{k+1}, \mathcal{E}_{n+1}^{n+l+b}\right) \\
&= P\left(\Phi_{l,b}^{k+1}\right) \times \left(\tfrac{1}{q}\right)^{l} \left(\zeta\left(k+1, y_{n+l+1}\right)\right)^{b} \times P\left(\mathcal{E}_{n+l+b+1}^{R} | N_{k+1} = n + l + b\right) \\
&= \tfrac{\mu_{l,b}}{q^{l}} \left(\zeta\left(k+1, y_{n+l+1}\right)\right)^{b} \times \beta\left(k+1, n+l+b\right)
\end{aligned} \tag{27}$$

By plugging (27) into (26) we obtain (16).

D. Derivation of (17)

Since N_0 is a degenerate random variable where $P\left(N_0 = 0\right) = 1$, then $N_0 = 0$ is the universal event. Also, by definition, $\mathcal{E}_1^0$ is the universal event. Therefore, $\alpha\left(0,0\right) = P\left(\mathcal{E}_1^0, N_0 = 0\right) = 1$.

Also, since $N_0 = n$ is the impossible event for $n \neq 0$, then $\alpha\left(0,n\right) = P\left(\mathcal{E}_1^0, N_0 = n\right) = 0$ for $n \neq 0$.

Since for all $k \geq 0$ we have $P\left(N_k < 0\right) = 0$, then $N_k = n$ is the impossible event for $n < 0$. Therefore, $\alpha\left(k,n\right) = 0$ for $k \geq 0, n < 0$.

To show that for $k \geq 1$, $\alpha\left(k,0\right) = \mu_{0,0}^{k}$, we note that $\alpha\left(k,0\right) = P\left(\mathcal{E}_1^0, N_k = 0\right) = P\left(N_k = 0\right)$ (since $\mathcal{E}_1^0$ is the universal event). Also, $P\left(N_k = 0\right) = P\left(\bigcap_{t=1}^{k} \Phi_{0,0}^{t}\right) = \prod_{t=1}^{k} P\left(\Phi_{0,0}^{t}\right) = {\mu_{0,0}^{k}}^{2}$.

E. Derivation of (18)

Since it is assumed that R symbols are received as a result of transmission of T symbols over the channel, we may take N_T as a degenerate random variable with $P\left(N_T = R\right) = 1$. Also, by definition, $\mathcal{E}_{R+1}^{R} = \Omega$, the universal event. Therefore, $\beta\left(T,R,R\right) = P\left(\mathcal{E}_{R+1}^{R} | N_T = R\right) = 1$.

Also, $N_T = n$ is the impossible event for $n \neq R$. Therefore, $\beta\left(T,n,R\right) = P\left(\mathcal{E}_{R+1}^{R} | N_T = n\right) = 0$ for $n \neq R$.

Also, for all $k \leq T$ and all $n > R$, we know that the joint event $N_k = n, N_T = R$ is impossible. Therefore:

[2]Note that we may not combine the cases of $k = 0$ and $k > 0$ and write $\alpha\left(k,0\right) = \mu_{0,0}^{k}$ for $k \geq 0$. The reason is that for some specific models, we may have $\mu_{0,0} = 0$; i.e., a pure deletion event may not occur. In such cases, still $\alpha\left(0,0\right) = 1$; hence, we need to treat the case of $k = 0$ separately.

$$P\left(N_k = n\right) = P\left(N_k = n, N_T = R\right) + P\left(N_k = n, N_T \neq R\right) = 0$$

Therefore, $N_k = n$ is the impossible event and $\beta\left(k, n, R\right) = 0$.

Finally, for $k < T$, by noting that $N_T = R$ is the universal event, we have:

$$\begin{aligned} &P\left(N_k = R\right) = P\left(N_k = R, N_T = R\right) \\ &= P\left(\cap_{t=k+1}^{T} \Phi_{0,0}^{t}\right) = \prod_{t=k+1}^{T} P\left(\Phi_{0,0}^{t}\right) \\ &= \mu_{0,0}^{T-k} \end{aligned}$$

Therefore, since:

$$\beta\left(k, R, R\right) = P\left(\mathcal{E}_{R+1}^{R} | N_k = R\right) = P\left(\Omega | N_k = R\right)$$

we have $\beta\left(k, R, R\right) = 1$ if $\mu_{0,0} \neq 0$, and $\beta\left(k, R, R\right) = 0$ otherwise, which can be expressed as $\beta\left(k, R, R\right) = \mathbf{1}_{\mu_{0,0}>0}$.

This completes the derivation of initial conditions for $\beta\left(.,.,.\right)$.

V. Bit-Level Likelihoods and Extension to Watermarks

In this section, we show how to find bit-level likelihoods to be passed to outer binary decoders. We also show how to extend the scheme for the case where watermarks are employed instead of markers. We conclude the section by providing complementary discussions.

A. Deriving Bit-level Likelihoods

After calculating $P\left(\mathcal{E}_1^R | X_k = a\right)$ according to (14), we may calculate:

$$P\left(\mathcal{E}_1^R, X_k = a\right) = P\left(X_k = a\right) \times P\left(\mathcal{E}_1^R | X_k = a\right)$$

and

$$P\left(\mathcal{E}_1^R\right) = \sum_{a=0}^{q-1} P\left(\mathcal{E}_1^R, X_k = a\right)$$

Now, let $b_{k,j}$ denote the jth bit of x_k. We assume that $b_{k,j}$ is a realization of a random variable $B_{k,j}$. Define:

$$\mathcal{A}_{j,0} = \{a \in \mathcal{Q} \text{ s.t. the } j\text{-th bit of } a \text{ is zero}\}$$
$$\mathcal{A}_{j,1} = \{a \in \mathcal{Q} \text{ s.t. the } j\text{-th bit of } a \text{ is one}\}$$

Then:

$$P\left(\mathcal{E}_1^R, B_{k,j} = 0\right) = \sum_{a \in \mathcal{A}_{j,0}} P\left(\mathcal{E}_1^R, X_k = a\right)$$
$$P\left(\mathcal{E}_1^R, B_{k,j} = 1\right) = \sum_{a \in \mathcal{A}_{j,1}} P\left(\mathcal{E}_1^R, X_k = a\right)$$

and the likelihood ratio for this bit can be found as $\frac{P\left(\mathcal{E}_1^R, B_{k,j}=1\right)}{P\left(\mathcal{E}_1^R, B_{k,j}=0\right)}$.

B. Extension to watermarks

Assume that $x_t = \tilde{x}_t \oplus w_t$, where $w_1^T \in \mathcal{Q}^T$ is a watermark sequence that is known at the decoder, and $\tilde{x}_t$ is a realization of a random variable $\tilde{X}_t$. Let $\tilde{X}_1 \ldots \tilde{X}_T$ be i.i.d. with a (non-uniform) distribution $P\left(\tilde{X}_t = a\right) = \tilde{\rho}_a$ for $a \in \mathcal{Q}$. Then $P\left(X_t = a\right) = \rho_a$ where $\rho_a = \tilde{\rho}_{a \oplus w_t}$ for $a \in \mathcal{Q}$. The same forward-backward equations may be employed to find the likelihoods in this case, by noting that:

$$P\left(Y_1^R = y_1^R | \tilde{X}_k = a\right) = P\left(Y_1^R = y_1^R | X_k = a \oplus w_k\right)$$

for all $a \in \mathcal{Q}$.

When watermarks are employed, we usually take $\mathcal{M} = \{\}$, i.e., no marker symbols are inserted. However, if marker symbols are inserted within the sequence $\tilde{x}_1^T$ at positions $\mathcal{M} = \{\tau_1, \ldots, \tau_J\}$, then $P\left(\tilde{X}_{\tau_j} = \tilde{m}_j\right) = 1$, i.e., $\tilde{X}_{\tau_j}$ is a degenerate random variable that certainly takes the marker value $\tilde{m}_j \in \mathcal{Q}$. In such a case, X_{τ_j} is also a degenerate random variable with $P\left(X_{\tau_j} = m_j\right) = 1$ for $m_j = \tilde{m}_j \oplus w_{\tau_j}$.

In some papers, even when markers are employed and X_t's are uniformly distributed, still a sequence w_1^T is added which is called the bias sequence. Such a case corresponds to the scenario where $\tilde{\rho}_a = \rho_a = \frac{1}{q}$, $\mathcal{M} = \{\tau_1, \ldots, \tau_J\}$, and $m_j = \tilde{m}_j \oplus w_{\tau_j}$.

C. Role of function $\zeta(.,.)$ in synchronization

As a final discussion, we note that the function $\zeta(k, y_n) = \sum_{a=0}^{q-1} P\left(X_k = a\right) \times f\left(a, y_n\right)$ helps align the received sequence with the transmitted sequence according to the marker pattern, as follows. If the kth symbol is a marker symbol with a value of a', then X_k is a degenerate random variable that takes

value of a' with probability 1. Therefore, $\zeta(k, y_n)$ is maximized when $y_n = a'$, i.e., in positions n for which the received symbol matches the marker pattern (by assuming $p_s < \frac{1}{q}$).

Alternatively, when a watermark is employed, we can show that $\zeta(k, y_n)$ is maximized when $y_n = w_k$. For simplicity, assume that $q = 2$, and let $P\left(\tilde{X}_k = 0\right) = p_0 > 0.5$. Then, by noting that $P(X_k = a) = P\left(\tilde{X}_k = w_k \oplus a\right)$ we have:

$$\zeta(k, 0) = P\left(\tilde{X}_k = w_k\right)(1 - p_s) + P\left(\tilde{X}_k = w_k \oplus 1\right) p_s$$
$$\zeta(k, 1) = P\left(\tilde{X}_k = w_k\right) p_s + P\left(\tilde{X}_k = w_k \oplus 1\right)(1 - p_s)$$

Since $p_s < 0.5$, it is clear that (i) when $w_k = 0$, $\zeta(k, 0) > \zeta(k, 1)$, (ii) when $w_k = 1$, $\zeta(k, 1) > \zeta(k, 0)$. From (i), (ii), we observe that $\zeta(k, y_n)$ is maximized when $y_n = w_k$, i.e., for positions n in which the received symbol matches the watermark symbol[3]

REFERENCES

[1] M. C. Davey and D. J. Mackay, "Reliable communication over channels with insertions, deletions and substitutions," IEEE Trans. Inf. Theory, vol. 47, no. 2, pp. 687–698, Feb. 2001.

[3]In this example, each symbol contains one bit, but the discussion can be readily extended for a general value of $q > 2$ with $p_s < \frac{1}{q}$.