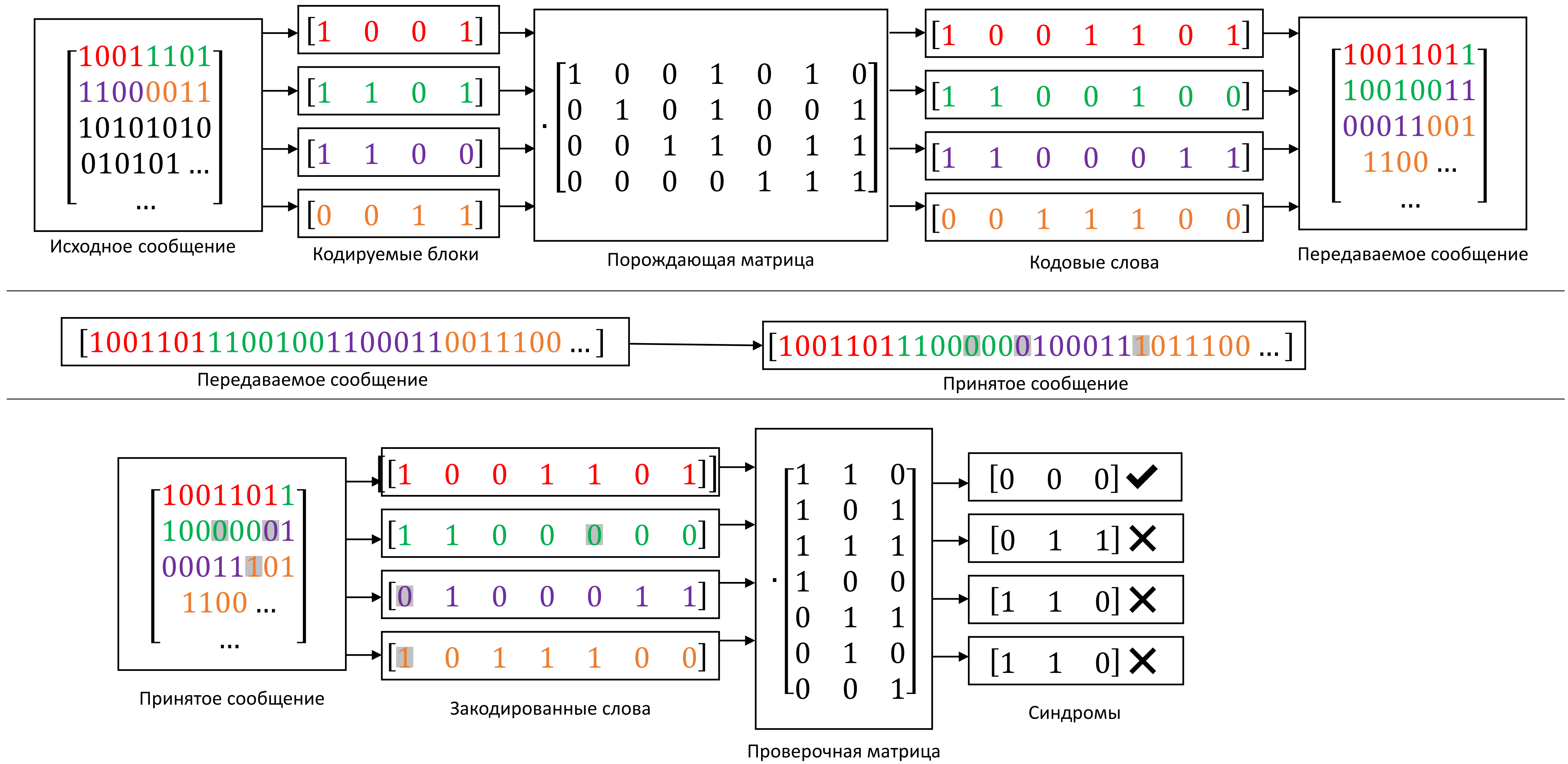
Теория кодирования

Гошин Егор Вячеславович, к.т.н., доцент
кафедры суперкомпьютеров и общей информатики

Общая схема кодирования, передачи и обнаружения ошибки



Систематические коды

Рассмотрим матрицу $k \times n$, где $k < n$, первые k столбцов которой представляют собой единичную матрицу I_k, т.е. матрицу вида:

$$G = [I_k|X]$$

Строки этой матрицы линейно независимы, а сама матрица представлена в приведённом ступенчатом виде. Следовательно, любая матрица G такого вида (он называется стандартным) является порождающей для некоторого линейного кода (n, k, d).

Код с такой порождающей матрицей называется систематическим.

Систематические коды

Одна из причин, по которой систематические коды имеют преимущество по сравнению с несистематическими – это процесс формирования проверочной матрицы.

Можно показать, что по алгоритму 3 (с предыдущей лекции) если матрица G имеет стандартную форму

$$G = [I_k|X]$$

то проверочная матрица будет иметь вид

$$H = \begin{bmatrix} X \\ I_{n-k} \end{bmatrix}$$

Кодовые слова систематического кода

Все кодовые слова систематического кода будут иметь вид

$$v = uG = u[I_k\ X] = [uI_k\ \ uX] = [u\ \ uX]$$

Теорема. Если C – линейный код длины n и размерности k с порождающей матрицей G в стандартном виде, тогда первые k разрядов кодового слова $v = uG$ формируют слово u.

Для систематического кода первые k разрядов называют информационными, а последние $n - k$ проверочными.

Приведение к стандартному виду

Что делать, если порождающая матрица не может быть приведена к стандартному виду?

Рассмотрим два кода:

$$C_1 = \{000, 100, 001, 101\}$$

и

$$C_2 = \{000, 100, 010, 110\}$$

Тогда

$$G_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad G_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Эквивалентные коды

Если C блочный код длины n, можно всегда получить новый блочный код C' длины n одинаковой перестановкой разрядов во всех кодовых словах из C.

Такой код называют эквивалентным.

Теорема. Для любого кода C существует эквивалентный ему линейный код C' с порождающей матрицей в стандартном виде.

Пример

$$G = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix} \quad H = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Пример

$$G' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix} \quad H' = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Классы смежности

Если C – линейный код длины n, и если u – произвольное слово длины n, определим класс смежности C по u как множество всех слов вида $v + u$, где v – кодовые слова из С.

Будем обозначать класс смежности как $C + u$

$$C + u = \{v + u \mid v \in C\}$$

Пусть $C = \{000, 111\}, u = 101$

тогда

$$C + 101 = \{101, 010\}$$

Пример

$$C = \{0000, 1011, 0101, 1110\}$$

$$C + 1000 = \{1000, 0011, 1101, 0110\}$$

$$C + 0100 = \{0100, 1111, 0001, 1010\}$$

$$C + 0010 = \{0010, 1001, 0111, 1100\}$$

Свойства классов смежности

Пусть C – линейный код длины n.
Пусть u и v – слова длины n.

Тогда:

1. Если u входит в класс смежности $C + v$, тогда $C + u = C + v$.
2. Слово u входит в класс смежности $C + u$.
3. Если $u + v$ входит в C, то u и v входят в один класс смежности.
4. Если $u + v$ не входят в C, то u и v входят в различные классы смежности.

5. Каждое слово из K^n входит в один и ровно один класс смежности по C. Таким образом, либо $C + u = C + v$, либо $C + u$ и $C + v$ не имеют общих слов.

6. $|C + u| = |C|$. Число слов в каждом классе смежности одинаково и равно числу слов в C.

7. Если C имеет размерность k, тогда существует ровно 2^{n-k} различных классов смежности по C и каждый из них содержит 2^k слов.

8. Код C является одним из собственных классов смежности.

Пример

$$C = \begin{Bmatrix} 000000, 100110, 010011, 001111, \\ 110101, 101001, 011100, 111010 \end{Bmatrix}$$

$$C + 000100 = \begin{Bmatrix} 000100, 100010, 010111, 001011, \\ 110001, 101101, 011000, 111110 \end{Bmatrix}$$

$$C + 100000 = \begin{Bmatrix} 100000, 000110, 110011, 101111, \\ 010101, 001001, 111100, 011010 \end{Bmatrix}$$

$$C + 000010 = \begin{Bmatrix} 000010, 100100, 010001, 001101, \\ 110111, 101011, 011110, 111000 \end{Bmatrix}$$

$$C + 010000 = \begin{Bmatrix} 010000, 110110, 000011, 011111, \\ 100101, 111001, 001100, 101010 \end{Bmatrix}$$

$$C + 000001 = \begin{Bmatrix} 000001, 100111, 010010, 001110, \\ 110100, 101000, 011101, 111011 \end{Bmatrix}$$

$$C + 001000 = \begin{Bmatrix} 001000, 101110, 011011, 000111, \\ 111101, 100001, 010100, 110010 \end{Bmatrix}$$

$$C + 000101 = \begin{Bmatrix} 000101, 100011, 010110, 001010, \\ 110000, 101100, 011001, 111111 \end{Bmatrix}$$

ММП для линейных кодов

Пусть C – линейный код. Предположим, что передано слово $v \in C$ и получено слово w, возникшее в результате ошибки $u = v + w$. Тогда $w + u = v$, и это означает, что w и u находятся в одном классе смежности по C.

Это значит, что если в каждом классе смежности выбрать слово наименьшей длины, это и будет наиболее вероятной ошибкой для любого принятого слова из этого класса смежности.

Пример

$$C = \begin{Bmatrix} 000000, 100110, 010011, 001111, \\ 110101, 101001, 011100, 111010 \end{Bmatrix}$$

$$C + 100000 = \begin{Bmatrix} 100000, 000110, 110011, 101111, \\ 010101, 001001, 111100, 011010 \end{Bmatrix}$$

$$C + 010000 = \begin{Bmatrix} 010000, 110110, 000011, 011111, \\ 100101, 111001, 001100, 101010 \end{Bmatrix}$$

$$C + 001000 = \begin{Bmatrix} 001000, 101110, 011011, 000111, \\ 111101, 100001, 010100, 110010 \end{Bmatrix}$$

$$C + 000100 = \begin{Bmatrix} 000100, 100010, 010111, 001011, \\ 110001, 101101, 011000, 111110 \end{Bmatrix}$$

$$C + 000010 = \begin{Bmatrix} 000010, 100100, 010001, 001101, \\ 110111, 101011, 011110, 111000 \end{Bmatrix}$$

$$C + 000001 = \begin{Bmatrix} 000001, 100111, 010010, 001110, \\ 110100, 101000, 011101, 111011 \end{Bmatrix}$$

$$C + 000101 = \begin{Bmatrix} 000101, 100011, 010110, 001010, \\ 110000, 101100, 011001, 111111 \end{Bmatrix}$$

Ведущий элемент класса смежности

Любое слово наименьшей длины в классе смежности называется ведущим элементом класса смежности.

Если используется неполный метод максимального правдоподобия и в классе смежности не более одного ведущего элемента, то этот ведущий элемент и будет ошибкой. Иначе – ошибка не может быть исправлена.

Процедура поиска класса смежности

Назовём wH синдромом w.

Для кода C матрица H – проверочная матрица. Если $w = 1101$, тогда синдром равен

$$wH = 1101H = 11$$

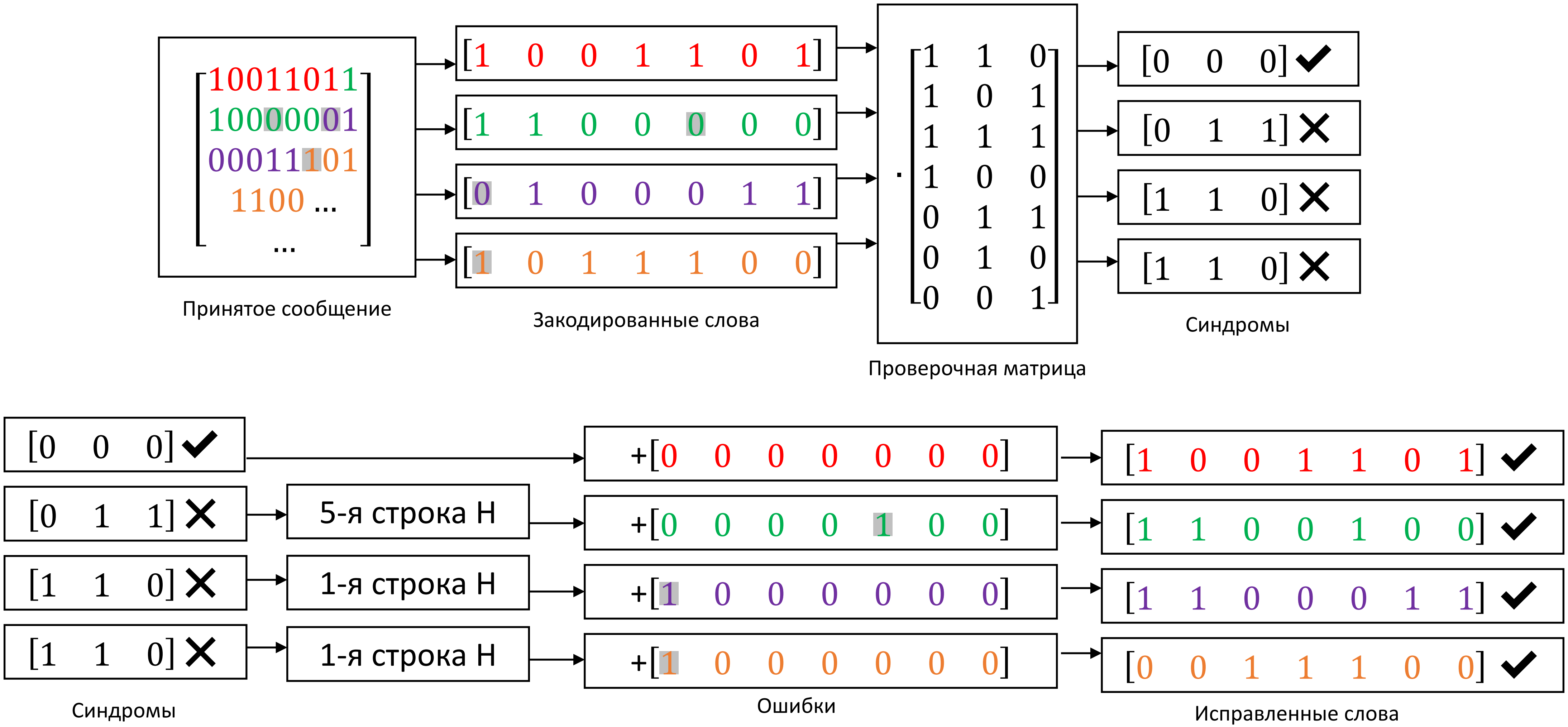
Слово наименьшей длины в классе смежности 1000 и его синдром также равен 11.

Связь синдрома и позиции ошибки

Пусть C – линейный код длины n. Пусть H проверочная матрица C. Пусть w и u слова из K^n.

1. Синдром $wH = 0$ тогда и только тогда, когда w кодовое слово из C.
2. Синдромы $wH = uH$ тогда и только тогда, когда они в одном классе смежности.
3. Если u ошибка в принятом слове w, тогда uH – сумма строк H, соответствующих позициям, в которым возникла ошибка.

Исправление ошибки



Пример для ЛР 1-2

$$S = \{1001011, 1100001, 0011001, 1010101, 0011110\}$$

$$S_{matrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \rightarrow G = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

$$n = 7, k = 4$$

Пример для ЛР 1-2

$$G' = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$$H' = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Пример для ЛР 1-2

$$u = \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix}$$

$$v = uG' = \begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

Исходное сообщение: $\begin{bmatrix} 1 & 0 & 0 & 1 \end{bmatrix}$

Отправлено: $\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$

В ходе передачи возникла ошибка

Принято: $\begin{bmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}$

Декодирование без исправления: $\begin{bmatrix} 1 & 0 & 1 & 1 \end{bmatrix}$

Пример для ЛР 1-2

Исходное сообщение: $\begin{bmatrix}1 & 0 & 0 & 1\end{bmatrix}$

Отправлено: $\begin{bmatrix}1 & 0 & 0 & 1 & 1 & 0 & 0\end{bmatrix}$

В ходе передачи возникла ошибка

Принято: $\begin{bmatrix}1 & 0 & 1 & 1 & 1 & 0 & 0\end{bmatrix}$

$$vH' = \begin{bmatrix}1 & 0 & 1 & 1 & 1 & 0 & 0\end{bmatrix}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 0\\ 1 & 0 & 1\\ 0 & 1 & 1\\ 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix} = \begin{bmatrix}1 & 0 & 1\end{bmatrix}$$

Синдром: $\begin{bmatrix}1 & 0 & 1\end{bmatrix}$

В принятом сообщении есть ошибка

Пример для ЛР 1-2

Исходное сообщение: $[1 \quad 0 \quad 0 \quad 1]$

Отправлено: $[1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0]$

В ходе передачи возникла ошибка

Принято: $[1 \quad 0 \quad 1 \quad 1 \quad 1 \quad 0 \quad 0]$

Синдром: $[1 \quad 0 \quad 1]$

В принятом сообщении есть ошибка

$[1 \quad 0 \quad 1]$ – 3-я строка матрицы H'

Ошибка: $[0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0]$

Исправленное сообщение: $[1 \quad 0 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0]$

Декодированное сообщение: $[1 \quad 0 \quad 0 \quad 1]$