**Теория кодирования**
**МФТИ, весна 2015**
**Лектор: Александр Дайняк**

# 1. Однозначно декодируемые коды. Неравенство Макмиллана.

В следующих заданиях под кодом понимается множество кодовых слов.

**Задача 1.** Установите, является ли следующие коды однозначно декодируемыми:
*(a)* $\{01,\ 201,\ 112,\ 122,\ 0112\}$,
*(b)* $\{aaa,\ abaa,\ ba,\ baab,\ aabaaba\}$,
*(c)* $\{aa,\ ab,\ aba,\ aca,\ bca,\ aaca,\ cabc\}$,
*(d)* $\{ab,\ aba,\ caa,\ baaa,\ aa\}$.
Для этого постройте специальный граф («граф Маркова», его структура описана в первой половине первой лекции) и посмотрите, есть ли в графе цикл, проходящий через вершину $\varepsilon$.

**Задача 2.** Подумайте, как с помощью графа Маркова научиться для неоднозначного кода отыскивать слово *минимальной длины*, которое можно декодировать двумя разными способами. Приведите пример, показывающий, что такое слово не всегда соответствует циклу в графе Маркова, имеющему минимальное число вершин.

**Задача 3.** Для следующих кодов в общем случае (для произвольного $k$) установите, являются ли они однозначно декодируемыми. Запись вида $w^k$ обозначает, что слово $w$ записывается подряд $k$ раз.
*(a)* $\{0,\ (10)^{k+1},\ (01)^k\}$,
*(b)* $\{001,\ 011,\ 100,\ 110,\ (1100)^k\}$.

**Задача 4.** Пусть $C$ — код, число слов в котором равно $20$, а максимальная длина слов равна $10$. Известно, что любое двоичное слово длины $\leqslant 2013$ можно не более чем одним способом разбить на слова из кода $C$. Можно ли из этого заключить, что код $C$ однозначный?

**Задача 5.** Пользуясь неравенством Крафта—Макмиллана, установите, при каких $n$ следующие наборы чисел могут являться наборами длин кодовых слов однозначно декодируемых двоичных кодов. Тот же вопрос для троичных кодов.
*(a)* $\{2,\ 3,\ n,\ n^2,\ \lceil \log_2 n \rceil\}$,
*(b)* $\{2,\ n+1,\ n+1,\ n+3,\ 2n\}$.

**Задача 6.** Пусть $T$ — набор натуральных чисел, в котором числа не повторяются, кроме, быть может, самого большого (да и то не более одного раза). Докажите, что $T$ может быть набором длин слов однозначного двоичного кода.

**Задача 7.** Постройте двоичный префиксный код из шести слов, длины слов которого равны
*(a)* $1, 2, 4, 4, 4, 4$,
*(b)* $2, 2, 3, 3, 3, 3$.

**Задача 8.** Верно ли, что для любого однозначного кода найдётся *суффиксный* код с тем же набором длин слов?

**Задача 9.** *Булевым $n$-мерным кубом* называется множество $\{0,\ 1\}^n$. *Гранью булева куба размерности $l$* называется множество наборов вида $\{(x_1,\ \ldots,\ x_n) \mid x_{i_1} = \alpha_1,\ \ldots,\ x_{i_{n-l}} = \alpha_{n-l}\}$. Иными словами, грань булева куба — это множество наборов, у которых все, кроме некоторых $l$ координат, фиксированы, а остальные координаты пробегают всевозможные значения. Докажите, что если набор чисел $n,\ l_1,\ l_2,\ \ldots,\ l_m$ таков, что $\sum_{i=1}^{m} 2^{l_i} \leqslant 2^n$, то в $n$-мерном булевом кубе можно выделить набор непересекающихся граней размерностей $l_1,\ \ldots,\ l_m$ соответственно.

## 2. Двоичные коды с минимальной избыточностью

В задачах ниже сокращение «к. м. и.» означает *код с минимальной избыточностью*. В этом разделе мы рассматриваем только двоичные коды.

**Задача 10.** Из доказанной на лекции теоремы Хаффмана о редукции вытекает следующий способ построения к. м. и. для произвольного набора вероятностей. Пусть дан набор $p_1 \geqslant \ldots \geqslant p_{n-1} \geqslant p_n$. «Склеим» в этом наборе две наименьших частоты $p_{n-1}$ и $p_n$, получим новый набор частот $p_1, \ldots, p_{n-2}, p_{n-1}+p_n$. В полученном наборе чисел опять сложим два наименьших, придя к набору из $(n-2)$ чисел. Так будем действовать, пока не придём к набору из двух чисел. Для такого набора к. м. и., очевидно, состоит из двух слов длины $1$ каждое: слова «0» и «1». Остаётся вернуться к исходному набору $p_1, \ldots, p_n$, «расклеивая» частоты обратно и пользуясь теоремой о редукции. Пользуясь описанными выше соображениями, постройте к. м. и. для набора частот

*(a)* $\{0.6;\ 0.1;\ 0.08;\ 0.08;\ 0.04;\ 0.04;\ 0.03;\ 0.03\}$,

*(b)* $\{0.3;\ 0.2;\ 0.2;\ 0.08;\ 0.08;\ 0.08;\ 0.06\}$,

*(c)* $\{0.4;\ 0.1;\ 0.1;\ 0.1;\ 0.1;\ 0.1;\ 0.1\}$.

**Задача 11.** Приведите пример набора частот, для которого существуют к. м. и. с различными наборами длин слов.

**Задача 12.** Существует способ построения кода для набора частот, называемый методом Шеннона—Фано. Он состоит в следующем. Пусть задан набор частот $p_1, \ldots, p_n$. Разобьём его на две части, так, чтобы сумма частот в одной из частей была максимально близка к сумме частот в другой части. Далее каждую из этих частей рекурсивно продолжим разбивать по тому же принципу... Получим дерево, вершинами которого являются поднаборы набора $p_1, \ldots, p_n$, и при этом рёбра идут из набора $X$ в наборы $Y$ и $Z$, если в ходе рекурсивных разбиений мы разбили $X$ на $Y$ и $Z$. Корнем этого бинарного дерева будет набор $\{p_1, \ldots, p_n\}$, а листьями — одноэлементные наборы $\{p_1\}$, ... $\{p_n\}$. Условно будем считать, что у каждой нелистовой вершины есть «левое» и «правое» поддеревья (какое из них двух выбирается в качестве «левого», не играет роли). Теперь для каждой частоты соответствующее кодовое слово можно построить так: спускаемся от корня дерева к листу этой частоты, и каждый раз, идя в левое поддерево, пишем «0», а идя в правое — «1». Выписанная последовательность нулей и единиц и будет кодовым словом. Убедитесь, что для набора частот $\{0.35;\ 0.17;\ 0.17;\ 0.16;\ 0.15\}$ метод Шеннона—Фано даёт код, *не являющийся к. м. и.*, и постройте к. м. и. методом Хаффмана.

**Задача 13.** Докажите, что если префиксный код является к. м. и. для некоторого набора частот, то в нём найдутся два слова, отличающиеся только в последней координате.

**Задача 14.** *(a)* Докажите, что если код является к. м. и. для некоторого набора частот, то неравенство Макмиллана обращается для этого кода в равенство.

*(b)* Пусть числа $l_1, \ldots, l_n$ удовлетворяют соотношению $\sum_{k=1}^{n} 2^{-l_k} = 1$. Докажите, что существует код, являющийся к. м. и. для некоторого набора частот и имеющий длины слов $l_1, \ldots, l_n$.

**Задача 15.** Докажите, что если код является к. м. и. для некоторого набора частот, то в нём чётное количество слов максимальной длины.

**Задача 16.** Индукцией по $n$ докажите, что если код является к. м. и. для некоторого набора из $n$ частот, то сумма длин всех кодовых слов не превосходит величины $\frac{(n-1)(n+2)}{2}$.

## 3. Типы ошибок. Кодовое расстояние.

По умолчанию везде далее под «расстоянием» подразумевается расстояние Хемминга.

**Задача 17.** Когда происходит ошибка замещения, в переданном в канал слове некоторый бит заменяется на противоположный. Рассмотрим ещё один тип ошибок: *ошибки стирания*. При ошибке стирания бит слова заменяется на специальный символ ?, отсутствующий в кодовом алфавите. В этом случае задача *обнаружения* ошибок стирания, конечно, не ставится. Зато, по-прежнему, имеет смысл задача *исправления* ошибок.

*(a)* Докажите, что код $C$ способен исправлять $t$ ошибок стирания т. и т. т., когда $d(C) > t$.

*(b)* Докажите, что код способен исправлять одновременно $t_1$ ошибок замещения и $t_2$ ошибок стирания т. и т. т., когда $d(C) > 2t_1 + t_2$.

**Задача 18.** Эта задача показывает, что ошибки выпадения и вставки «взаимозаменимы».

*(a)* Докажите, что если код способен исправлять одну ошибку вставки, то он способен исправлять одну ошибку выпадения.

*(b)* Пусть $n_1,\ n_2,\ n_3,\ n_4$ — произвольные фиксированные неотрицательные числа, и пусть $n_1 + n_2 = n_3 + n_4$. Докажите, что если некоторый код способен исправлять одновременно $n_1$ ошибок выпадения и $n_2$ ошибок вставки, то этот же код способен исправлять одновременно $n_3$ ошибок выпадения и $n_4$ ошибок вставки.

**Задача 19.** Рассмотрим *код проверки чётности*: $C := \{\mathbf{a} \in \{0,1\}^n \mid \sum_{i=1}^n a_i \equiv 0 \pmod 2\}$. Найдите кодовое расстояние этого кода и определите, сколько ошибок замещения может этот код обнаруживать и исправлять.

## 4. Коды Варшамова—Тененгольца

**Задача 20.** Код Варшамова—Тененгольца длины $n$ определяется как множество слов

$$\left\{a_1 \dots a_n \mid \sum_{i=1}^{n} i a_i \equiv 0 \pmod{(n+1)}\right\}.$$

Докажите, что если $n = 2^m - 1$ для некоторого $m$, то число кодовых слов равно $2^{n-m}$. Указание: рассмотрите позиции кодового слова с номерами $1,\ 2,\ 4,\ \dots,\ 2^{m-1}$ и докажите, что если все остальные позиции заданы произвольным образом, то указанные позиции однозначно доопределяются.

**Задача 21.** Пусть $l > n$, и пусть $l$ простое. Рассмотрим множество двоичных слов

$$\left\{a_1 \dots a_n \mid \sum_{i=1}^{n} i a_i \equiv \sum_{i=1}^{n} i^2 a_i \equiv 0 \pmod{l}\right\}.$$

Докажите, что это множество является кодом, способным исправлять две ошибки замещения вида $0 \to 1$.

## 5. Элементарные методы построения кодов

**Задача 22.** Укажите, как из кода с параметрами $(n, M, d)_q$, где $d \geqslant 2$, получить код с параметрами $(n-1, M, d')_q$, где $d' \geqslant d - 1$.

**Задача 23.** Укажите, как из кода с параметрами $(n, M, d)_q$ получить код с параметрами $(n-1, M', d')_q$, где $M' \geqslant \lceil M/q \rceil$ и $d' \geqslant d$.

**Задача 24.** *Расширенным кодом* для двоичного кода $C$ называется код $C'$, слова которого получены дописыванием к словам кода $C$ одного бита, равного сумме (по модулю 2) остальных битов слова. То есть расширенный код получается из исходного добавлением бита контроля чётности. Докажите, что если $C$ является $(n, M, d)$-кодом и $d$ нечётно, то $C$ будет $(n+1, M, d+1)$-кодом.

**Задача 25.** Пусть $C$ — некоторый $(n, M, d)_q$-код, причём кодовое расстояние $d$ достигается не более чем на двух парах кодовых слов. Покажите, как по коду $C$ построить $(n+1, M, d+1)_q$-код.

**Задача 26.** Пусть $C_1$ и $C_2$ имеют параметры соответственно $(n, M_1, d_1)$ и $(n, M_2, d_2)$. Докажите, что множество слов $\{(\mathbf{a}, \mathbf{a}+\mathbf{b}) \mid \mathbf{a} \in C_1, \mathbf{b} \in C_2\}$ является $(2n,\ M_1 \cdot M_2,\ \min\{2d_1, d_2\})$-кодом (суммирование здесь побитовое). Это утверждение принадлежит Плоткину (M. Plotkin), а соответствующая конструкция кодов называется *конструкцией Плоткина*. Указание: рассмотрите пару слов вида $(\mathbf{a}', \mathbf{a}'+\mathbf{b}')$, $(\mathbf{a}'', \mathbf{a}''+\mathbf{b}'')$, и разберите случаи $\mathbf{b}' = \mathbf{b}''$ и $\mathbf{b}' \neq \mathbf{b}''$.

# 6. Границы мощностей кодов

Пусть $C$ — $q$-ичный код с длиной слов $n$. *Скоростью* кода $C$ называется величина $\mathrm{rate}(C) := \frac{\log_q |C|}{n}$. Эта величина показывает, насколько «разбухают» данные, если они кодируются с помощью кода $C$, во время *блочного кодирования*. Если кодируются $q$-ичные данные, то каждому $q$-ичному слову длины $\lfloor \log_q |C| \rfloor$ можно однозначно сопоставить кодовое слово. При *блочном кодировании* данных последовательность, которую нужно закодировать, разбивается на куски одинаковой длины, и затем каждый кусок кодируется своим кодовым словом. Если исходные данные имели размер $N$, то после кодирования они будут иметь размер $\frac{N}{\lfloor \log_q |C| \rfloor} \cdot n \approx \frac{N}{\mathrm{rate}(C)}$.

Пусть $C$ — код с длиной слов $n$ и кодовым расстоянием $d$. *Относительное кодовое расстояние* кода $C$ — это величина $\delta(C) := \frac{d}{n}$. Она характеризует, какую долю от длины кодовых слов могут составлять ошибки, чтобы их всё ещё можно было обнаружить.

Очевидно, $\mathrm{rate}(C) \in [0, 1]$ и $\delta(C) \in [0, 1]$ для любого кода $C$. При асимптотическом сравнении границ и оценке качества семейств кодов чаще используются указанные нормированные величины $\delta(C)$ и $\mathrm{rate}(C)$, нежели $d(C)$ и $|C|$.

Напомним основные границы, доказанные на лекциях:

- **Граница Хемминга.** Для любого $(n, M, d)_q$-кода выполнено неравенство $M \leqslant \frac{q^n}{|S_{\lfloor (d-1)/2 \rfloor}|}$.
- **«Анти-Хемминг».** Если $M \leqslant \frac{q^n}{|S_d|}$, то существует $(n, M, d)_q$-код.
- **Граница Синглтона.** Для любого $(n, M, d)_q$-кода выполнено неравенство $M \leqslant q^{n-d+1}$.
- **Граница Плоткина.** Для любого $(n, M, d)_q$-кода, где $\delta(C) = \delta > 1 - 1/q$, выполнено неравенство $M \leqslant \frac{\delta}{\delta - \frac{q-1}{q}}$.
- **Граница Элайеса—Бассалыго.** Для любого двоичного $(n, M, d)$-кода, где $\delta(C) = \delta \leqslant \frac{1}{2}$, выполнено неравенство $M \leqslant \frac{n \cdot 2^n}{|S_{\lfloor \tau n - 1 \rfloor}|}$, где $\tau = \frac{1}{2}(1 - \sqrt{1 - 2\delta})$.

**Задача 27.** При доказательстве границы Элайеса—Бассалыго мы переходили от двоичных слов к точкам евклидова пространства, так, чтобы расстояния между точками легко пересчитывались через расстояние между словами. Однако точного совпадения расстояний в этом случае мы не требовали. Докажите, что при $n > 1$ хеммингово пространство размерности $n$ (т.е. множество $\{0,1\}^n$ с метрикой Хемминга) нельзя вложить в $\mathbb{R}^m$ ни для какого $m$, так, чтобы евклидово расстояние между образами любой пары слов было равно хеммингову расстоянию между словами.

**Задача 28.** Докажите следующее обобщение границы Хемминга для двоичных кодов. Пусть существует такой набор функций $\{f_{\mathbf{a}}\}_{\mathbf{a} \in C}$, определённых на $\{0,1\}^n$, что

$$\forall \mathbf{b} \in \{0,1\}^n \sum_{\mathbf{a} \in C} f_{\mathbf{a}}(\mathbf{b}) \leqslant 1 \quad \text{и} \quad \forall \mathbf{a} \in C \sum_{\mathbf{b} \in \{0,1\}^n} f_{\mathbf{a}}(\mathbf{b}) \geqslant s\,.$$

Покажите, что в этом случае $|C| \leqslant \frac{2^n}{s}$. Чему равны функции $f_{\mathbf{a}}$ в доказательстве границы Хемминга?

**Задача 29.** *(a)* Покажите, что объём шара радиуса $r$ в $\mathbb{A}_q^n$ (с метрикой Хемминга) равен $\sum_{k=0}^{r} \binom{n}{k}(q-1)^k$.

*(b)* Найдите асимптотику этой величины при $n \to \infty$ и фиксированных $q, r$.

*(c)* Найдите асимптотику логарифма этой величины при $q = 2$, $n \to \infty$ и фиксированном $\rho = \frac{r}{n}$.

**Задача 30.** *(a)* Какая из границ Хемминга и Синглтона лучше при $n \to \infty$ и фиксированных $q, d$?
*(b)* Какая из границ Хемминга и Синглтона лучше при $n \leqslant q$?

**Задача 31.** Пусть величины $\delta(C)$ и $\mathrm{rate}(C)$ неизменны, а $n$ стремится к бесконечности.
*(a)* Переформулируйте границу Синглтона в терминах $\mathrm{rate}(C)$ и $\delta(C)$.
*(b)* Докажите, что для двоичных кодов граница Хемминга даёт неравенство $\mathrm{rate}(C) \leqslant 1 - H\left(\frac{\delta(C)}{2}\right)$, где $H$ — *функция двоичной энтропии*: $H(p) = -p\log_2 p - (1-p)\log_2(1-p)$.
*(c)* Убедитесь, что граница Элайеса—Бассалыго лучше границы Хемминга при $n \to \infty$ и фиксированном $\delta < \frac{1}{2}$.

**Задача 32.** Граница Плоткина, доказанная на лекции, подходит только для оценки мощности кодов, имеющих очень большое кодовое расстояние: $\delta > 1 - \frac{1}{q}$. Тем не менее, из неё можно получить оценку, применимую при меньших $\delta$.
*(a)* По аналогии с задачей 23 докажите, что если существует $(n,\ M,\ d)_q$-код, то существует и $(n - t,\ \frac{M}{q^t},\ d)_q$-код.
*(b)* Пусть величины $q$, $\delta(C)$ и $\mathrm{rate}(C)$ зафиксированы, а $n$ стремится к бесконечности. Воспользуйтесь результатом предыдущего пункта и границей Плоткина для доказательства утверждения: если $\delta(C) \leqslant 1 - \frac{1}{q}$, то $\mathrm{rate}(C) \leqslant 1 - \frac{q}{q-1} \cdot \delta(C) + o(1)$.

**Задача 33.** *(a)* Пусть $m$ — произвольное натуральное число. Опираясь на конструкцию Плоткина из задачи 26, постройте код с параметрами $(2^m,\ 2^{m+1},\ 2^{m-1})$. Указание: примените конструкцию Плоткина рекурсивно, взяв в качестве вспомогательного кода тривиальный $(2^{m-1},\ 2,\ 2^{m-1})$-код.
*(b)* Пользуясь результатом задачи 23 и границей Плоткина, докажите, что если у $(n, M, d)$-кода $d = \frac{n}{2}$, то $M \leqslant 2n$. Заметьте, что число слов построенного в п. a кода достигает этой границы.

**Задача 34.** Насколько сильно (линейно/полиномиально/экспоненциально… по $n$) расходятся между собой граница Хемминга и обратное утверждение («анти—Хемминг»), при фиксированном $d$? Тот же вопрос при $\frac{d}{n} = \mathrm{const}$.

## 7. Линейные коды: общие свойства

**Задача 35.** Пусть двоичный линейный код содержит хотя бы одно слово нечётного веса. Докажите, что количество кодовых слов нечётного веса равно половине от числа всех кодовых слов.

**Задача 36.** *(a)* Проверьте, что если двоичный код линеен, то и построенный по нему расширенный код линеен. (см. задачу 24)
*(b)* Проверьте, что если исходные коды в конструкции Плоткина (см. задачу 26) линейны, то и результирующий код тоже линейный.

**Задача 37.** Пусть $C$ — $[n, k, d]_q$-код. Пусть слово $\mathbf{b} \in \mathbb{F}_q^n$ таково, что $d(\mathbf{a}, \mathbf{b}) \geqslant d$ для любого $\mathbf{a} \in C$. Покажите, что множество $C' := \{(\mathbf{a} + \beta\mathbf{b}) \mid \mathbf{a} \in C,\ \beta \in \mathbb{F}_q\}$ является $[n, k+1, d]_q$-кодом. Переход от кода $C$ к коду $C'$ называется *пополнением* кода $C$.

**Задача 38.** Докажите, что если $C$ — линейный код, то для любых $a, b \in C$ и любого $t \in \mathbb{N}$ число кодовых слов на расстоянии $t$ от $a$ равно числу слов на расстоянии $t$ от $b$.

**Задача 39.** Докажите или опровергните: если $[n, k, d]$-коды $C_1$ и $C_2$ таковы, что $|C_1 \cap C_2| \geqslant k$, то $C_1 = C_2$.

**Задача 40.** Столбцы проверочной матрицы $[7, 4, 3]$-кода Хемминга — всевозможные ненулевые векторы из $\mathbb{F}_2^3$. Предъявите какую-нибудь порождающую матрицу этого кода и приведите её к каноническому виду.

**Задача 41.** Обобщите конструкцию кодов Хемминга на $q$-ичный случай. Достигается ли граница Хемминга на недвоичных кодах Хемминга?

**Задача 42.** *(a)* Пусть $G_1$ и $G_2$ — порождающие матрицы кодов с параметрами $[n_1, k, d_1]_q$ и $[n_2, k, d_2]_q$ соответственно. Покажите, что код с порождающей матрицей $(G_1|G_2)$ является $[n_1{+}n_2,\ k,\ d']$-кодом, где $d' \geqslant d_1 + d_2$. Приведите пример, когда $d' > d_1 + d_2$.

*(b)* Пусть $G_1$ и $G_2$ — порождающие матрицы кодов с параметрами $[n_1, k_1, d_1]_q$ и $[n_2, k_2, d_2]_q$ соответственно. Найдите параметры кода с порождающей матрицей $\left(\begin{smallmatrix} G_1 & 0 \\ 0 & G_2 \end{smallmatrix}\right)$.

**Задача 43.** Пусть $C'$ и $C''$ — линейные коды с параметрами $[n', k', d']_q$ и $[n'', k'', d'']_q$ соответственно. Будем считать, что у этих кодов зафиксированы некоторые порождающие матрицы, а значит и способы отображения слов из $\mathbb{F}_q^{k'}$ в $C'$ и из $\mathbb{F}_q^{k''}$ в $C''$. *Произведением* кодов $C'$ и $C''$ называется код, строящийся следующим образом (опишем это как отображение, заданное на $\mathbb{F}_q^{k'k''}$). Произвольное слово из $\mathbb{F}_q^{k'k''}$ разобьём на части $\mathbf{a}_1, \ldots, \mathbf{a}_{k''} \in \mathbb{F}_q^{k'}$. Закодируем каждую из них в коде $C'$. Получим $k''$ слов длины $n'$ каждое. Запишем их построчно в $(k'' \times n')$-матрицу. Теперь каждый столбец этой матрицы закодируем в коде $C''$. Получим $n'$ столбцов длины $n''$ каждый. Их конкатенация и будет словом в коде-произведении. Докажите, что получаемый код линеен и имеет параметры $[n'n'',\ k'k'',\ d'd'']_q$.

**Задача 44.** Пусть порождающая матрица систематического кода имеет вид: $G = \left(I^k \middle| \widetilde{G}\right)$, где $I^k$ — единичная матрица. Докажите, что матрица $\left(-\widetilde{G}^T \middle| I^{n-k}\right)$ является проверочной матрицей этого кода.

**Задача 45.** Естественный вопрос: верно ли, что если существует какой-то код с заданными параметрами $(n, M, d)$, то найдётся и линейный код с такими же или лучшими параметрами? Оказывается, не всегда; в некоторых случаях линейные коды проигрывают нелинейным по числу слов. Известный *нелинейный* двоичный код Нордстрома—Робинсона имеет параметры $(16,\ 256,\ 6)$.

*(a)* Докажите, что не существует *линейного* двоичного кода с параметрами $[16,\ 8,\ 6]$ (если бы он существовал, то число кодовых слов в нём было то же, что у кода Нордстрома—Робинсона). Указание: примените последовательно теорему Соломона—Штиффлера об остаточном коде и границу Хемминга.

*(b)* Убедитесь, что одна лишь граница Грайсмера—Соломона—Штиффлера не позволяет доказать несуществование $[16,\ 8,\ 6]$-кода.

**Задача 46.** С помощью теоремы Варшамова—Гилберта докажите, что существуют коды с параметрами $[16,\ 9,\ 4]$, $[16,\ 6,\ 5]$, $[16,\ 4,\ 6]$ и $[21,\ 8,\ 6]$.

## 8. Коды на основе двудольных графов

**Задача 47.** $G$ — двудольный граф без изолированных вершин, с долями $L$ и $R$, и пусть каждой вершине $v \in R$ поставлен в соответствие некоторый линейный код $C_v$, длина кодовых слов которого равна $\deg v$. *Кодом Таннера (R. M. Tanner), построенным по графу $G$ и семейству кодов* $(C_v \mid v \in R)$, называется код, определяемый следующим образом. Каждой вершине $u \in L$ ставим в соответствие переменную $x_u$. Считаем, что слово $(a_u \mid u \in L)$ принадлежит коду Таннера, если для каждой $v \in R$ слово, образованное соседями $v$, принадлежит коду $C_v$.

*(a)* Докажите, что код Таннера является линейным кодом.

*(b)* Докажите, что если каждый из кодов $C_v$ является $[n_v, k_v, d_v]$-кодом, то размерность кода Таннера не меньше $|L| - \sum_{v \in R}(n_v - k_v)$.

*(c)* Докажите, что кодовое расстояние кода Таннера не меньше $\min_{v \in R} d(C_v)$.

**Задача 48.** Пусть $C$ — код Таннера, построенный на $(n, m, \Delta, \alpha, c)$-расширителе (код «Таннера—Сипсера—Шпильмана»), в котором степени вершин в доле $R$ равны $l$. Пусть для каждого $v \in R$ код $C_v$ является $[l, k, d]$-кодом. Докажите, что при $c > \Delta/d$ кодовое расстояние кода $C$ не меньше $\alpha n$. Указание: доказательство проводится аналогично тому, что было на лекции; находим вершину в $R$, в которую входит меньше, чем $d$ рёбер из вершин-единиц предполагаемого ненулевого «кодового» слова веса $< \alpha n$ — противоречие.

## 9. Многочлены и конечные поля

**Задача 49.** Для следующих многочленов докажите, что они являются неприводимыми и выполните вычисления в соответствующем конечном поле.

*(a)* $x^3 + x + 1 \in \mathbb{Z}_2[x]$, вычислите $(x+1) \cdot (x^2+1)^{-1} + x^2$ в поле $\mathbb{Z}_2[x]/(x^3 + x + 1)$;

*(b)* $x^4 + 2 \in \mathbb{Z}_5[x]$, вычислите $(x^2+1) \cdot (x^2 + 4x + 2) + 3x^3 + 1$ в поле $\mathbb{Z}_5[x]/(x^4 + 2)$.

*Сопровождающей матрицей* нормированного многочлена $Q(x) = c_0 + c_1 x + \ldots + c_{m-1}x^{m-1} + + x^m \in \mathbb{Z}_p[x]$ называется матрица из $\mathbb{Z}_p^{m \times m}$ вида

$$A_Q := \begin{pmatrix} 0 & 0 & \ldots & 0 & -c_0 \\ 1 & 0 & \ldots & 0 & -c_1 \\ 0 & 1 & \ldots & 0 & -c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -c_{m-1} \end{pmatrix}.$$

Через $I$ обозначим единичную матрицу из $\mathbb{Z}_p^{m \times m}$. Пусть $\mathcal{A}$ — линейная оболочка (где линейные комбинации берутся с коэффициентами из $\mathbb{Z}_p$) множества матриц $\{I,\ A_Q,\ A_Q^2,\ A_Q^3,\ \ldots\}$. Известно, что $\mathcal{A}$ является полем относительно обычных сложения и умножения матриц. Это поле оказывается изоморфным $\mathbb{Z}_p/Q$. Таким образом, при необходимости вычислять в конечном поле, можно пользоваться знакомой арифметикой матриц, достаточно быстро выполняемой алгоритмически. При этом, однако, элементы поля $\mathbb{F}_{p^m}$ приходится хранить матрицами размера $m \times m$. Оптимальный по памяти способ хранения — хранить коэффициенты многочленов из $\mathbb{Z}_p/Q$, то есть векторы из $\mathbb{Z}_p^m$.

**Задача 50.** *(a)* Убедитесь, что многочлен $x^2 + x + 1 \in \mathbb{Z}_2[x]$ неприводим над полем $\mathbb{Z}_2$.

*(b)* Постройте таблицы сложения и умножения поля $\mathbb{Z}_2[x]/(x^2 + x + 1)$.

*(c)* Сопоставьте элементам поля набор матриц $\mathcal{A} \subset \mathbb{Z}_2^{2\times 2}$, такой, что сумме и произведению элементов поля соответствует сумма и произведение матриц из $\mathcal{A}$.

*(d)* Укажите в полученном поле *примитивный элемент*, т. е. элемент $\lambda$, такой, что любой элемент поля можно представить как $\lambda^t$ для подходящего $t$.

**Задача 51.** Постройте конечное поле из $9$ элементов.

**Задача 52.** Покажите, что $x$ является примитивным элементом поля $\mathbb{Z}_2[x]/(x^4 + x + 1)$.

**Задача 53.** Докажите, что в поле $\mathbb{Z}_p[x]/Q$ (где $Q$ — неприводимый над $\mathbb{Z}_p$ многочлен) для любых $a, b \in \mathbb{Z}_p[x]/Q$ справедливо равенство $(a+b)^p = a^p + b^p$.

# 10. Коды Рида—Маллера и Рида—Соломона. Мажоритарное декодирование.

**Задача 54.** Цель этой задачи — показать, что $[7, 4, 3]$-код Хемминга допускает мажоритарное декодирование. Рассмотрим порождающую матрицу этого кода:

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Любой кодовый вектор $\mathbf{a} = (a_1, \ldots, a_7)$ может быть представлен в виде $\mathbf{a} = c_1\mathbf{e}_1 + \ldots + c_4\mathbf{e}_4$, где $\mathbf{e}_i$ — строки матрицы. Допустим, что мы передали вектор $\mathbf{a}$ по каналу, и на выходе получили вектор $\widehat{\mathbf{a}}$, отличный от исходного не более чем в одной координате.

*(a)* Предъявите три равенства вида $c_2 = \sum_s a_{j_s}$, где множества слагаемых сумм в правых частях равенств не пересекаются. Предъявите аналогичные системы из трёх равенств для $c_3$ и $c_4$.

*(b)* Пользуясь равенствами из предыдущего пункта, декодирование можно осуществлять так: вычисляем для каждого $i \in \{2, 3, 4\}$ три суммы для $c_i$ (вместо $a_{j_s}$ беря $\widehat{a}_{j_s}$ — координаты вектора $\widehat{\mathbf{a}}$) и берём в качестве $c_i$ то значение, которое получилось по крайней мере в двух суммах из трёх. Осталось найти $c_1$. Для этого строим вектор $\widehat{\mathbf{a}} - c_2\mathbf{e}_2 - c_3\mathbf{e}_3 - c_4\mathbf{e}_4$. Этот вектор должен отличаться от вектора $c_1\mathbf{e}_1$ не более чем в одной координате, — отсюда тривиально определяется $c_1$. Декодируйте описанным способом слово $(1001010)$.

**Задача 55.** Обнаружьте и, при возможности, исправьте ошибки в следующих словах кода Рида—Маллера с параметрами $r = 1$, $m = 3$:

*(a)* $11010110$,

*(b)* $11100001$.

**Задача 56.** Цель задачи — показать связь кодов Рида—Маллера с кодами Хемминга.

*(a)* Докажите, что коды Рида—Маллера с параметрами $(r, m)$ и с параметрами $(m - r - 1, m)$ ортогональны (скалярное произведение двух слов из различных кодов равно нулю). Указание: вычислить скалярное произведение кодовых слов — это то же самое, что взять произведение многочленов из различных кодов и вычислить сумму значений полученного многочлена по всем наборам значений переменных.

*(b)* Докажите, что *расширенный код Хемминга*, полученный добавлением бита проверки чётности к словам $[2^m - 1, 2^m - m - 1, 3]$-кода Хемминга, эквивалентен коду Рида—Маллера с параметрами $(m - 2, m)$. Указание: посмотрите, как выглядит проверочная матрица расширенного кода Хемминга и используйте утверждение предыдущего пункта.

**Задача 57.** Обнаружьте и исправьте ошибки в следующих словах $[5, 3, 3]_5$-кода Рида—Соломона. Предполагается, что поле, в котором работаем — это поле вычетов по модулю $5$, а кодовое слово построено как $(P(0), P(1), P(2), P(3), P(4))$.

*(a)* $14134$,

*(b)* $42244$,

*(c)* $20324$.

**Задача 58.** Задачи по лемме Липтона—ДеМилло—Шварца—Зиппеля.

*(a)* Докажите, что заключение леммы остаётся верным, если $s_1, \ldots, s_m$ выбираются не из одного и того же множества, а каждое из своего множества мощности $N$.

*(b)* Как можно обобщить лемму, если каждое $s_i$ выбирается из некоторого множества мощности $N_i$?

*(c)* Как можно уточнить лемму, если добавить знание о величинах $\deg_{x_i} P$?

**Задача 59.** Эта задача касается параметров $\delta$ и $\mathrm{rate}$, определяемых в начале разд. 6. Докажите, что при стремящейся к бесконечности длине слов и при фиксированном $\delta > 0$ у кодов Рида—Маллера оказывается $\mathrm{rate} \to 0$.

**Задача 60.** Код Рида—Соломона над алфавитом $\mathbb{F}_{2^t}$ можно рассматривать как двоичный код, заменив каждую координату кодового слова на двоичный вектор длины $t$.

*(a)* Какие параметры будет иметь полученный код?

*(b)* Покажите, что при стремящейся к бесконечности длине слов и при фиксированном $\delta > 0$ у полученного кода оказывается $\mathrm{rate} \to 0$.

## 11. Циклические коды. Коды БЧХ.

**Задача 61.** На лекции было доказано, что если циклический код $C$ порождается многочленом $g$, и если $f \in C$, то корни многочлена $g$ являются и корнями многочлена $f$. Докажите обратное утверждение: если у порождающего многочлена $g$ ровно $\deg g$ различных корней, и если многочлен $f$ таков, что среди его корней есть все корни $g$, то $f \in C$.

**Задача 62.** Проверочная матрица циклического кода, эквивалентного $[7, 4, 3]$-коду Хемминга, имеет вид:

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}$$

Найдите проверочный и порождающий многочлены этого кода и выпишите его порождающую матрицу.

**Задача 63.** Определите, какие из следующих многочленов могут быть порождающими многочленами циклического двоичного кода длины $12$:

*(a)* $x^3 + 1$,

*(b)* $x^2 + x + 1$,

*(c)* $x^3 + x + 1$,

*(d)* $x^4 + 1$.

**Задача 64.** В поле $\mathbb{F}_{2^m}$ справедливо равенство $(a + b)^2 = a^2 + b^2$ (см. задачу 53). Пользуясь этим, докажите следующее утверждение. Пусть $\lambda$ — примитивный элемент поля $\mathbb{F}_{2^m}$, пусть $t > 0$, и пусть $g$ — порождающий многочлен циклического кода $C$, причём каждый из коэффициентов $g$ равен нулю либо единице. Тогда, если $g$ имеет корни $\lambda, \lambda^2, \ldots, \lambda^{2t+1}$, то справедлива оценка $d(C) \geqslant 2t + 3$. (Указание: эта оценка немного улучшает границу БЧХ.)

**Задача 65.** В $\mathbb{F}_2[x]$ справедливо равенство $x^{15} - 1 = (x+1)(x^2+x+1)(x^4+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1)$. Известно, что для примитивного элемента $\alpha \in \mathbb{F}_{2^4}$ следующие многочлены являются минимальными для соответствующих элементов:

- $x^4 + x + 1$ — минимальный многочлен для $\alpha$, $\alpha^2$, $\alpha^4$, $\alpha^8$,
- $x^4 + x^3 + x^2 + x + 1$ — минимальный многочлен для $\alpha^3$, $\alpha^6$, $\alpha^9$, $\alpha^{12}$,
- $x^4 + x^3 + 1$ — минимальный многочлен для $\alpha^7$, $\alpha^{11}$, $\alpha^{13}$, $\alpha^{14}$,
- $x^2 + x + 1$ — минимальный многочлен для $\alpha^5$, $\alpha^{10}$.

Для каждого из следующих двоичных циклических кодов длины $15$ с порождающим многочленом $g(x)$ найдите размерность кода, максимально точно оцените кодовое расстояние, выпишите проверочный многочлен кода.

*(a)* $g(x) = x^4 + x + 1$,

*(b)* $g(x) = (x+1)(x^2+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1)$,
*(c)* $g(x) = (x+1)(x^2+x+1)(x^4+x^3+1)$.

**Задача 66.** Построим $[7, 4, 3]$-код Хемминга как циклический код. Пусть поле $\mathbb{F}_8$ построено как $\mathbb{F}_2[z]/(z^3+z+1)$. Пусть $\lambda := z^2+z+1 \in \mathbb{F}_8$.

*(a)* Убедитесь, что $\lambda$ — примитивный элемент в $\mathbb{F}_8$.

*(b)* Каждому элементу из $\mathbb{F}_8$ сопоставьте двоичный вектор длины $3$, так, чтобы это сопоставление «сохраняло сумму». Достаточно взять векторы коэффициентов многочленов из $\mathbb{F}_2[z]/(z^3+z+1)$, которыми представляются элементы поля.

*(c)* Предъявите минимальный многочлен $g \in \mathbb{F}_2[x]$ для элемента $\lambda$, найдя линейно зависимые векторы среди векторов, сопоставленных элементам $\lambda^0$, $\lambda^1$, $\lambda^2$, $\lambda^3$. Убедитесь, что этот многочлен является делителем многочлена $x^7 - 1 \in \mathbb{F}_2[x]$.

*(d)* Выпишите порождающую и проверочную матрицы циклического двоичного кода, построенного по порождающему многочлену $g$.

## 12. Восстановление синхронизации

**Задача 67.** Предложите двоичный код с длиной слов не более $23$ и количеством слов не менее $4096$, который способен исправлять $3$ ошибки замещения *либо* синхронизационный сдвиг вплоть до $10$ тактов.

**Задача 68.** *(a)* Покажите, что *почти все* двоичные слова длины $n$ обладают следующим свойством: любой циклический сдвиг слова (на $i$ позиций, где $0 < i < n$) не совпадает с самим словом.

*(b)* Докажите, что число слов в двоичном коде длины $n$, свободном от запятой, асимптотически не превосходит $\frac{2^n}{n}$.

## 13. Матрицы Адамара и коды Адамара

Пусть $A = (a_{ij})$, $B = (b_{ij})$ — квадратные матрицы порядков $m$ и $n$ соответственно. *Кронекеровым произведением* матриц $A$ и $B$ называется квадратная матрица порядка $mn$ вида:

$$A \times B = \begin{pmatrix} a_{11}\cdot B, & a_{12}\cdot B, & \dots & a_{1m}\cdot B \\ a_{21}\cdot B, & a_{22}\cdot B, & \dots & a_{2m}\cdot B \\ \vdots & \vdots & & \vdots \\ a_{m1}\cdot B, & a_{m2}\cdot B, & \dots & a_{mm}\cdot B \end{pmatrix} =$$

$$= \begin{pmatrix} a_{11}b_{11}, \dots a_{11}b_{1n}, & a_{12}b_{11}, \dots a_{12}b_{1n} & \dots & a_{1m}b_{11}, \dots a_{1m}b_{1n} \\ \vdots & \vdots & & \vdots \\ a_{11}b_{n1}, \dots a_{11}b_{nn}, & a_{12}b_{n1}, \dots a_{12}b_{nn} & \dots & a_{1m}b_{n1}, \dots a_{1m}b_{nn} \\ \vdots & \vdots & & \vdots \\ a_{m1}b_{n1}, \dots a_{m1}b_{nn}, & a_{m2}b_{n1}, \dots a_{m2}b_{nn} & \dots & a_{m2}b_{n1}, \dots a_{mm}b_{nn} \end{pmatrix}.$$

**Задача 69.** Пусть $H_m$ и $H_n$ — матрицы Адамара порядков $m$ и $n$ соответственно. Докажите, что их кронекерово произведение является матрицей Адамара порядка $mn$.

**Задача 70.** Пусть $n = 2^m$ и пусть $C$ — $\left(n, 2n, \frac{n}{2}\right)$-код на основе $(n \times n)$-матрицы Адамара, построенной по конструкции Сильвестра.

*(a)* Покажите, что код $C$ эквивалентен коду Рида—Маллера.

*(b)* Покажите, что код $C$ можно построить с помощью конструкции Плоткина (см. задачу 33).

**Задача 71.** Пусть $p > 2$ — простое число, и пусть $q = p^m$ таково, что $4 \mid (q-1)$. Известно, что при таких $q$ элемент $-1$ является квадратичным вычетом в $\mathbb{F}_q$. Покажите, что матрица Адамара порядка $2(q+1)$ может быть построена с помощью следующей конструкции. Пусть $J$ — матрица Якобшталя порядка $q$ (при указанном $q$ эта матрица будет симметрична). Далее рассмотрим матрицу

$$\begin{pmatrix} 0 & 1 & \dots & 1 \\ 1 & & & \\ \vdots & & J & \\ 1 & & & \end{pmatrix}$$

(то есть к $J$ приписываем строку из единиц сверху, столбец из единиц слева, и ноль слева-сверху). Затем в полученной матрице элементы, равные нулю (т.е. диагональные элементы) заменим на блок $\left(\begin{smallmatrix} 1 & -1 \\ -1 & -1 \end{smallmatrix}\right)$, а элементы, равные $\pm 1$, заменим блоками $\pm \left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$ соответственно.

## 14. Каскадные коды

**Задача 72.** С помощью каскадной конструкции постройте двоичные коды со следующими параметрами:

*(a)* $(2^{12},\, 2^7,\, 1500)$

*(b)* $[23 \cdot 2^{12},\, 3 \cdot (2^{14} - 8),\, 21]$