

**Федеральное государственное автономное образовательное
учреждение высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»**

УТВЕРЖДЕНО

**Директор физтех-школы
прикладной математики и
информатики
А.М. Райгородский**

	Рабочая программа дисциплины (модуля)
по дисциплине:	Теория кодирования
по направлению:	Информатика и вычислительная техника
профиль подготовки:	Прикладная математика и информатика Физтех-школа Прикладной Математики и Информатики кафедра дискретной математики
курс:	1
квалификация:	магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

Аудиторных часов: 60 всего, в том числе:

лекции: 0 час.

семинары: 60 час.

лабораторные занятия: 0 час.

Самостоятельная работа: 30 час.

Всего часов: 90, всего зач. ед.: 2

Программу составил: А.Б. Дайняк, канд. физ.-мат. наук, доцент

Программа обсуждена на заседании кафедры дискретной математики 05.03.2020

Аннотация

В курсе излагаются в основном вопросы алгебраической теории кодирования: способам использования комбинаторных и алгебраических структур для построения схем кодирования информации, обеспечивающих устойчивость к ошибкам. Основным рассматриваемый тип ошибок — ошибки замещения с детерминированным ограничением, однако рассмотрены и некоторые другие модели и задачи. В частности, теоремы Шеннона об энтропийной оценке пропускной способности вероятностного канала связи, теорема Хафмана, коды Варшамова—Тененгольца для исправления ошибок вставки/выпадения, а также ошибки синхронизации.

1. Цели и задачи

Цель дисциплины

освоение основных современных методов теории кодирования.

Задачи дисциплины

- освоение студентами базовых знаний (понятий, концепций, методов и моделей) в области теории кодирования;
- приобретение теоретических знаний и практических умений и навыков в области теории кодирования;
- оказание консультаций и помощи студентам в проведении собственных теоретических ис-следований в области теории кодирования.

2. Перечень формируемых компетенций

Освоение дисциплины направлено на формирование следующих компетенций:

Код и наименование компетенции	Индикаторы достижения компетенции
УК-3 Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	УК-3.1 Организует и координирует работу участников проекта, способствует конструктивному преодолению возникающих разногласий и конфликтов
УК-6 Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1 Умеет решать задачи собственного личностного и профессионального развития, определять и реализовывать приоритеты совершенствования собственной деятельности
ОПК-3 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области информатики и вычислительной техники, учитывая особенности и ограничения различных методов решения	ОПК-3.1 Способен анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области математики, естественных наук и информационно-коммуникационных технологий	ОПК-4.1 Способен применять знания и навыки по использованию информационно-коммуникационных технологий для поиска и изучения научной литературы, применения прикладных программных продуктов

3. Перечень планируемых результатов обучения по дисциплине (модулю)

В результате освоения дисциплины обучающиеся должны знать:

фундаментальные понятия, законы, теории части дискретной математики и теории кодирования;
 современные проблемы соответствующих разделов дискретной математики и теории кодирования;
 понятия, аксиомы, методы доказательств и доказательства основных теорем в разделах, входящих в базовую часть цикла;
 основные свойства соответствующих математических объектов;
 аналитические и численные подходы и методы для решения типовых прикладных задач дискретной математики и теории кодирования.

уметь:

понять поставленную задачу;
 использовать свои знания для решения фундаментальных и прикладных задач;
 оценивать корректность постановок задач;
 строго доказывать или опровергать утверждение;
 самостоятельно находить алгоритмы решения задач, в том числе и нестандартных, и проводить их анализ;
 самостоятельно видеть следствия полученных результатов;
 точно представить математические знания в области в устной и письменной форме.

владеть:

навыками освоения большого объема информации и решения задач (в том числе, сложных);
 навыками самостоятельной работы и освоения новых дисциплин;
 культурой постановки, анализа и решения математических и прикладных задач, требующих для своего решения использования математических подходов и методов;
 предметным языком дискретной математики и навыками грамотного описания решения задач и представления полученных результатов.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1. Разделы дисциплины (модуля) и трудоемкости по видам учебных занятий

№	Тема (раздел) дисциплины	Трудоемкость по видам учебных занятий, включая самостоятельную работу, час.			
		Лекции	Семинары	Лаборат. работы	Самост. работа
1	Алфавитное кодирование		12		6
2	Коды БЧХ		12		6
3	Линейные коды		12		6
4	Свёрточные коды		12		6
5	Сложность задачи декодирования линейных кодов		12		6
Итого часов			60		30
Подготовка к экзамену		0 час.			
Общая трудоёмкость		90 час., 2 зач.ед.			

4.2. Содержание дисциплины (модуля), структурированное по темам (разделам)

Семестр: 2 (Весенний)

1. Алфавитное кодирование

Достаточные условия однозначности декодирования: равномерность, префиксность, суффиксность. Распознавание однозначности: критерий Маркова. Оценка длины неоднозначно декодируемого слова.

2. Коды БЧХ

Задача восстановления синхронизации. Восстановление синхронизации для смежных классов циклических кодов.

3. Линейные коды

Определения. Порождающая и проверочная матрицы. Связь кодового расстояния с проверочной матрицей. Граница Варшавова—Гилберта. Систематическое кодирование. Декодирование по синдрому. Коды Хемминга.

4. Сверточные коды

Матрицы Адамара. Конструкции Сильвестра и Пэли. Коды на основе матриц Адамара.

5. Сложность задачи декодирования линейных кодов

Графы-расширители. Вероятностное доказательство существования расширителей. Коды на основе двудольных графов. Кодовое расстояние кодов на основе расширителей. Алгоритм декодирования Сипсера—Шпильмана.

5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Учебная аудитория, оснащенная компьютером и мультимедийным оборудованием (проектор, звуковая система).

6. Перечень рекомендуемой литературы

Основная литература

1. Заметки по теории кодирования [Текст] / А. Е. Ромащенко, А. Ю. Румянцев, А. Шен .— [Учебное изд.] .— М : МЦНМО, 2011 .— 80 с.
2. Введение в алгебраические коды [Текст] : учеб. пособие для вузов / Ю. Л. Сагалович ; М-во образования и науки Рос. Федерации, Моск. физ.-техн. ин-т (гос. ун-т), Ин-т проблем передачи информации им. А. А. Харкевича РАН .— М. : Изд-во МФТИ, 2007 .— 262 с.
3. Основы кодирования [Текст] / М. Вернер ; пер. с немец. Д. К. Зигангирова - М. Техносфера, 2004, 2006

Дополнительная литература

1. Дискретный анализ. Основы высшей алгебры [Текст] : учеб. пособие для вузов / Ю. И. Журавлев, Ю. А. Флеров, М. Н. Вялый ; М-во образования и науки Рос. Федерации, Моск. физ.-техн. ин-т (гос. ун-т .— 2-е изд., испр. и доп. — М. : МЗ Пресс, 2007 .— 224 с.

7. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

<http://dm.fizteh.ru/>

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

На занятиях используются мультимедийные технологии, включая демонстрацию презентаций.

В процессе самостоятельной работы обучающихся возможно использование таких программных средств, как Mathcad, MATLAB, Maple и др.

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Рекомендуется успешно сдавать контрольные работы, так как это упрощает итоговую аттестацию по предмету.

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ (МОДУЛЮ)

по направлению: Информатика и вычислительная техника
профиль подготовки: Прикладная математика и информатика
Физтех-школа Прикладной Математики и Информатики
кафедра дискретной математики
курс: 1
квалификация: магистр

Семестр, формы промежуточной аттестации: 2 (весенний) - Дифференцированный зачет

Разработчик: А.Б. Дайняк, канд. физ.-мат. наук, доцент

1. Компетенции, формируемые в процессе изучения дисциплины

Код и наименование компетенции	Индикаторы достижения компетенции
УК-3 Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	УК-3.1 Организует и координирует работу участников проекта, способствует конструктивному преодолению возникающих разногласий и конфликтов
УК-6 Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.1 Умеет решать задачи собственного личностного и профессионального развития, определять и реализовывать приоритеты совершенствования собственной деятельности
ОПК-3 Способен выбирать и (или) разрабатывать подходы к решению типовых и новых задач в области информатики и вычислительной техники, учитывая особенности и ограничения различных методов решения	ОПК-3.1 Способен анализировать задачу, планировать пути решения, предлагать и комбинировать способы решения
ОПК-4 Способен успешно реализовывать решение поставленной задачи, провести анализ результата и представить выводы, применяя знания и навыки в области математики, естественных наук и информационно-коммуникационных технологий	ОПК-4.1 Способен применять знания и навыки по использованию информационно-коммуникационных технологий для поиска и изучения научной литературы, применения прикладных программных продуктов

2. Показатели оценивания компетенций

В результате изучения дисциплины «Теория кодирования» обучающийся должен:

знать:

фундаментальные понятия, законы, теории части дискретной математики и теории ко-дирования;
современные проблемы соответствующих разделов дискретной математики и теории кодирования;
понятия, аксиомы, методы доказательств и доказательства основных теорем в разделах, входящих в базовую часть цикла;
основные свойства соответствующих математических объектов;
аналитические и численные подходы и методы для решения типовых прикладных задач дискретной математики и теории кодирования.

уметь:

понять поставленную задачу;
использовать свои знания для решения фундаментальных и прикладных задач;
оценивать корректность постановок задач;
строго доказывать или опровергать утверждение;
самостоятельно находить алгоритмы решения задач, в том числе и нестандартных, и проводить их анализ;
самостоятельно видеть следствия полученных результатов;
точно представить математические знания в области в устной и письменной форме.

владеть:

навыками освоения большого объема информации и решения задач (в том числе, сложных);
навыками самостоятельной работы и освоения новых дисциплин;
культурой постановки, анализа и решения математических и прикладных задач, требующих для своего решения использования математических подходов и методов;
предметным языком дискретной математики и навыками грамотного описания решения задач и представления полученных результатов.

3. Перечень типовых (примерных) вопросов, заданий, тем для подготовки к текущему контролю

Текущий контроль

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.

Булевым n -мерным кубом называется множество $\{0,1\}^n$. Гранью булева куба размерности l называется множество наборов вида $\{(x_1, \dots, x_n) \mid x_{i_1} = \alpha_1, \dots, x_{i_{n-l}} = \alpha_{n-l}\}$. Иными словами, грань булева куба — это множество наборов, у которых все, кроме некоторых l координат, фиксированы, а остальные координаты пробегают всевозможные значения. Используя префиксные коды, докажите, что если набор чисел n_1, n_2, \dots, n_m таков, что $\sum_{i=1}^m 2^{n_i} \leq 2^n$, то в n -мерном булевом кубе можно выделить набор непересекающихся граней размерностей n_1, \dots, n_m соответственно.

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

Докажите, что если код является к.м.и. для некоторого набора частот, то неравенство Макмиллана обращается для этого кода в равенство.

Пусть числа n_1, \dots, n_m удовлетворяют соотношению $\sum_{k=1}^m 2^{n_k} = 1$. Докажите, что существует код, являющийся к.м.и. для некоторого набора частот и имеющий длины слов n_1, \dots, n_m .

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 $\frac{(n-1)(n+2)}{2}$.

Докажите, что код способен исправлять одновременно t_1 ошибок замещения и t_2 ошибок стирания тогда и только тогда, когда $d(C) \geq 2t_1 + t_2$.

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.

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. Говорят, что слова расширенного кода получаются из слов исходного кода 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.

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

Код Варшавова—Тененгольца длины n определяется как множество слов $[a_1 \dots a_n \mid \sum_{i=1}^n a_i \equiv 0 \pmod{n+1}]$. 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 до корректного слова в коде Варшавова—Тененгольца.)

Пусть $l > n$, и пусть l простое. Рассмотрим множество двоичных слов $[a_1 \dots a_n \mid \sum_{i=1}^n a_i \equiv \sum_{i=1}^n i^2 a_i \equiv 0 \pmod{l}]$. Докажите, что это множество является кодом, способным исправлять две ошибки замещения вида $0 \rightarrow 1$.

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 $\forall \mathbf{b} \in \{0,1\}^n, \sum_{\mathbf{a} \in C} f_{\mathbf{a}}(\mathbf{b}) \equiv 1 \pmod{2}$ и $\forall \mathbf{a} \in C, \sum_{\mathbf{b} \in \{0,1\}^n} f_{\mathbf{a}}(\mathbf{b}) \geq s$. Покажите, что в этом случае $|C| \leq \frac{2^n}{s}$. Чему равны функции $f_{\mathbf{a}}$ в доказательстве границы Хемминга?

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

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 $(2n, M_1 \cdot M_2, \min\{d_1, d_2\})$ -кодом (суммирование здесь побитовое). Это утверждение принадлежит Плоткину (M. Plotkin), а соответствующая конструкция кодов называется конструкцией Плоткина. Указание: рассмотрите пару слов вида $(\mathbf{a}; \mathbf{a} + \mathbf{b})$ и $(\mathbf{a}; \mathbf{a} + \mathbf{b}')$, и разберите два случая, — когда $\mathbf{b} = \mathbf{b}'$ и когда $\mathbf{b} \neq \mathbf{b}'$.

Исследуйте связь между конструкцией Плоткина для кодов и конструкцией Сильвестра для матриц 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:

$$\begin{bmatrix} A \cdot B & A \cdot B & \dots & A \cdot B \\ A \cdot B & A \cdot B & \dots & A \cdot B \\ \vdots & \vdots & \ddots & \vdots \\ A \cdot B & A \cdot B & \dots & A \cdot B \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & \dots & a_{11}b_{1n} \\ a_{12}b_{11} & a_{12}b_{12} & \dots & a_{12}b_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m}b_{11} & a_{1m}b_{12} & \dots & a_{1m}b_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{11} & a_{m1}b_{12} & \dots & a_{m1}b_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{mn}b_{11} & a_{mn}b_{12} & \dots & a_{mn}b_{1n} \end{bmatrix}$$

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

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 Якобштала порядка q (при указанном q эта матрица будет симметрична). Далее рассмотрим матрицу $\begin{bmatrix} 0 & 1 & \dots & 1 \\ 1 & & & \\ \vdots & & & \\ 1 & & & \end{bmatrix}$ (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 $\begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}$, а элементы, равные ± 1 , заменим блоками $\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ соответственно.

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.

Проверьте, что если двоичный код линейен, то и построенный по нему расширенный код [то есть код, полученный добавлением к словам бита проверки чётности; см. Задачи 2-й недели] линейен.

Проверьте, что если исходные коды в конструкции Плоткина [см. Задачи 4-й недели] линейны, то и результирующий код тоже линейный.

Пусть CC — $[n, k, d]_q$ -код. Пусть слово $\mathbf{b} \in \mathbb{F}_q^n$ таково, что $d(\mathbf{a}, \mathbf{b}) \geq d$ 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 CC к коду C' называется пополнением кода CC .

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

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

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 $\begin{bmatrix} G_1 \\ G_2 \end{bmatrix}$ является $[n_1 + n_2, k, d']_q$ -кодом, где $d' \geq d_1 + d_2$. Приведите пример, когда $d' > d_1 + d_2$.

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 $\begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix}$.

Пусть \mathcal{C} и \mathcal{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 \mathcal{C} и из $\mathbb{F}_q^{k'}$ в \mathcal{C}' . Произведением кодов \mathcal{C} и \mathcal{C}' называется код, строящийся следующим образом (опишем это как отображение, заданное на $\mathbb{F}_q^{k'}$). Произвольное слово из $\mathbb{F}_q^{k'}$ разобьём на части $a_1, \dots, a_{k''}$ из \mathbb{F}_q^k . Закодируем каждую из них в коде \mathcal{C} . Получим k'' слов длины n каждое. Запишем их построчно в $(k'' \times n)$ -матрицу. Теперь каждый столбец этой матрицы закодируем в коде \mathcal{C}' . Получим n столбцов длины n' каждый. Их конкатенация и будет словом в коде-произведении. Докажите, что получаемый код линеен и имеет параметры $[n'n, k'k', d'd']_q$.

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

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 Нордстрёма—Робинсона имеет параметры $(16, 256, 6)$. 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 Нордстрёма—Робинсона). (Указание: примените последовательно теорему Соломона—Штиффлера об остаточном коде и границу Хемминга.) Проверьте, что одна лишь граница Грайсмера—Соломона—Штиффлера не позволяет доказать несуществование $[16, 8, 6]$ -кода.

Разрез в графе — это разбиение его вершин $V = S \sqcup (V \setminus S)$. Размером разреза называется число рёбер между множеством S и его дополнением. Задача MAX-CUT состоит в том, чтобы по заданному графу определить разрез максимального размера. Известно, что задача MAX-CUT не только NP-трудна сама по себе, но также NP-трудной является задача нахождения разреза, имеющего размер 0.95 от максимального. Докажите, что задача MAX-CUT полиномиально сводится к задаче NCP и сделайте отсюда вывод о трудности приближённого решения задачи NCP.

Код Рида—Соломона над алфавитом \mathbb{F}_{2^t} можно рассматривать как двоичный код, заменив каждую координату кодового слова на двоичный вектор длины t . Какие параметры будет иметь полученный код? Покажите, что при стремящейся к бесконечности длине слов и при фиксированном $\delta > 0$ у полученного кода оказывается $\text{rate} \rightarrow 0$.

Докажите, что код, ортогональный коду Рида—Маллера с параметрами (r, m) , сам является кодом Рида—Маллера (какие у него параметры?).

Найдите асимптотику параметра $\text{rate}(\mathcal{C})$ для кода Рида—Маллера \mathcal{C} при фиксированном $\delta(\mathcal{C})$ и растущем m .

Цель этой задачи — показать, что $[7, 4, 3]$ -код Хемминга допускает мажоритарное декодирование. Рассмотрим порождающую матрицу этого кода: $\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$. Любой кодовый вектор $\mathbf{a} = (a_1, \dots, a_7)$ может быть представлен в виде $\mathbf{a} = c_1 \mathbf{e}_1 + \dots + c_4 \mathbf{e}_4$, где \mathbf{e}_i — строки матрицы. Допустим, что мы передали вектор \mathbf{a} по каналу, и на выходе получили вектор $\hat{\mathbf{a}}$, отличный от исходного не более чем в одной координате. 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 . Пользуясь этими равенствами, декодирование можно осуществлять так: вычисляем для каждого $i \in \{2, 3, 4\}$ три суммы для c_i (вместо a_{j_s} беря \hat{a}_{j_s} — координаты вектора $\hat{\mathbf{a}}$) 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 $\hat{\mathbf{a}} - c_2 \mathbf{e}_2 - c_3 \mathbf{e}_3 - c_4 \mathbf{e}_4$. 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) .

Пусть $n = 2^m$ и пусть \mathcal{C} — $\left(n, \frac{n}{2}, \frac{n}{2}\right)$ -код Боуза—Шрихане на основе $(n \times n)$ -матрицы Адамара, построенной по конструкции Сильвестра. Покажите, что код \mathcal{C} эквивалентен коду Рида—Маллера.

Докажите, что заключение леммы Липтона—ДеМилло—Шварца—Зиппеля остаётся верным, если s_1, \dots, s_m выбираются не из одного и того же множества, а каждое из своего множества мощности N . Как можно обобщить лемму, если каждое s_i выбирается из некоторого множества мощности N_i ? Как можно уточнить лемму, если добавить знание о величинах $\deg_{x_i} P$?

Выведите из доказанной нами «вероятностной» формы теоремы Шеннона (с оценкой вероятности успешного декодирования при случайном выборе кодируемого слова) теорему Шеннона в следующей формулировке: «Пусть ϵ — сколь угодно малое положительное число. Пусть вероятность ошибки на символ равна $p \in (0, 1/2)$. Тогда существуют такие n, k и такие функции кодирования E и декодирования D , что $\text{rate}(E, D) \geq 1 - H(p) - \epsilon$ и для любого слова $\mathbf{a} \in \{0, 1\}^k$ при случайном векторе ошибок \mathbf{e} имеет место оценка вероятности $\Pr[D(E(\mathbf{a}) + \mathbf{e}) \neq \mathbf{a}] \leq \epsilon$ ».

Для доказательства этого факта можно воспользоваться теоремой в исходной формулировке, а затем взять k на единицу меньше и на основе исходной функции E построить новую функцию кодирования $E': \{0, 1\}^{k-1} \rightarrow \{0, 1\}^n$, и соответствующую ей функцию декодирования.

Проследите связь между теоремами Шеннона и «теоремой, обратной к границе Хемминга», доказанной нами в третьем уроке второй недели.

Постройте с помощью каскадной конструкции и некоторых из рассматриваемых нами ранее в курсе кодов для построения двоичного кода с параметрами $(240, 256, d)$, где $d \geq 72$.

4. Перечень типовых (примерных) вопросов и тем для проведения промежуточной аттестации обучающихся

1. Алфавитное кодирование. Свойства однозначно декодируемых кодов.
2. Распознавание однозначности. Оценка длины неоднозначно декодируемого слова. Префиксные коды. Неравенство Крафта—Макмиллана.
3. Существование префиксного кода с заданным набором длин слов. Коды с минимальной избыточностью. Теорема Хаффмана о редукции.
4. Задача исправления и обнаружения ошибок. Типы ошибок.
5. Коды Варшавова—Тененгольца, алгоритмы исправления одиночных ошибок выпадения и вставки символов. Графовая модель канала.
6. Шенноновское произведение графов, шенноновская ёмкость. Теорема о верхней оценке шенноновской ёмкости.
7. Симметричный канал с ошибками замещения.
8. Расстояние Хемминга. Кодовое расстояние.
9. Основные задачи теории кодов, исправляющих ошибки. Границы для параметров кодов, исправляющих ошибки: границы сферической упаковки, Синглтона, Плоткина. Вложение метрических пространств.
10. Лемма о числе векторов в евклидовом пространстве. Граница Элайеса—Бассальяго.
11. Линейные коды. Определения. Порождающая и проверочная матрицы. Связь кодового расстояния с проверочной матрицей. Систематическое кодирование.
12. Декодирование по синдрому. Коды Хемминга. Граница Варшавова—Гилберта. Остаточный код, граница Грайсмера—Соломона—Штиффлера.
13. Теорема Шеннона о пропускной способности симметричного двоичного канала (вторая часть — без доказательства).
14. Асимптотические параметры семейств кодов.
15. Сложность задачи декодирования линейных кодов: задача NCP. Графы-расширители. Вероятностное доказательство существования расширителей.
16. Коды на основе двудольных графов. Кодовое расстояние кодов на основе расширителей. Алгоритм декодирования Сипсера—Шпильмана.
17. Коды Рида—Соломона. Алгоритм декодирования Берлекэмпа—Велча. Коды Рида—Маллера: кодовое расстояние, алгоритм мажоритарного декодирования.
18. Лемма Липтона—ДеМилло—Шварца—Зиппеля. Понятие об алгеброгеометрических кодах.
19. Циклические коды. Проверочный и порождающий многочлены, критерий существования кода с заданным порождающим многочленом.

20. Вид порождающей и проверочной матриц. Систематическое кодирование. Граница Боуза—Чоудхури—Хоквингема.
21. Коды БЧХ. Задача восстановления синхронизации. Восстановление синхронизации для смежных классов циклических кодов.
22. Совершенные коды. Теорема Васильева.
23. Каскадные коды. Коды Форни.
24. Свёрточные коды.
25. Коды Адамара. Понятие о локальном декодировании.
26. Понятие о списочном декодировании. Граница Джонсона.
27. Некоторые приложения теоретико-кодовых результатов и подходов в теоретической информатике и инженерии.
28. Хранение данных на CD, криптографическая система МакЭлиса, схема разделения секрета Шамира, коммуникационная сложность, дерандомизация, пороговые структуры данных.

Критерии оценивания

- оценка «отлично (10)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений;
- оценка «отлично (9)» выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений;
- оценка «отлично (8)» выставляется студенту, показавшему всесторонние систематизированные, глубокие знания учебной программы дисциплины и умение применять их на практике при решении конкретных задач, и правильное обоснование принятых решений;
- оценка «хорошо (7)» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «хорошо (6)» выставляется студенту, если он знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «хорошо (5)» выставляется студенту, если он знает материал, и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задач некоторые неточности;
- оценка «удовлетворительно (4)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «удовлетворительно (3)» выставляется студенту, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет фрагментарно основными разделами учебной программы, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации;
- оценка «неудовлетворительно (2)» выставляется студенту, который не знает большей части основного содержания учебной программы дисциплины, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач;
- оценка «неудовлетворительно (1)» выставляется студенту, который не знает формулировок основных понятий дисциплины.

5. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Во время проведения дифференцированного зачета, обучающиеся могут пользоваться программой дисциплины.