ОТЗЫВ

научного руководителя о диссертации

Щукина Владислава Юрьевича

«Дизъюнктивные коды со списочным декодированием» (Disjunctive List-Decoding Codes),
представленной на соискание ученой степени кандидата
физико - математических наук по специальности
01.01.05 - «теория вероятностей и математическая статистика».

Дизъюнктивным каналом множественного доступа (ДКМД) называется канал множественного доступа (КМД) с одним выходом и s двоичными входами, в котором двоичный сигнал на выходе задается как булева сумма (дизъюнкция) s двоичных входных сигналов. Кодом для ДКМД объема t, длины N со списочным декодированием с параметрами $s, L = 1, 2, \ldots$, где $2 \leq s + L < t$, (кратко СД s_L-кодом) называется двоичная $(N \times t)$-матрица инцидентности семейства t подмножеств N-множества, интерпретируемая как двоичный код, состоящий из t кодовых слов длины N, для которого объединение (дизъюнкция) произвольных s подмножеств не содержит объединение каких-либо других L подмножеств данного семейства. В наиболее важном для приложений частном случае $L = 1$ определение СД s_1-кода, которое было дано в 1964 году У. Каутсом и Р. Синглтоном как дизъюнктивного s-кода (ДК), означает, что объединение любых s членов семейства покрывает те и только те члены семейства, из которых составлено это объединение.

При $L \geq 2$ определение СД s_L-кода было введено в 1983 году в моей совместной статье с В.В. Рыковым для анализа верхних и нижних границ сложности двухступенчатых алгоритмов групповых проверок в описываемой ДКМД модели поиска неизвестного подмножества дефектных элементов (дефектов) $\mathcal{S}$, $\mathcal{S} \subset [t] = \{1, 2, \ldots, t\}$, состоящего из $|\mathcal{S}|$, $|\mathcal{S}| \leq s$, элементов множества $[t]$. Это определение означало, что применение на первой ступени двухступенчатой процедуры N групповых проверок, задаваемых двоичной $(N \times t)$-матрицей СД s_L-кода, позволяет по двоичным результатам проверок, получаемых на выходе ДКМД, т.е. по булевой сумме $\leq s$ столбцов СД s_L-кода с номерами $j \in \mathcal{S}$ однозначно выделить $\leq s + L - 1$ элементов $[t]$, среди которых находится все множество дефектов $\mathcal{S}$. Очевидно, что по результатам последующих на второй ступени $\leq s + L - 1$ статических проверок поодиночке элементов, выделенных на первой ступени, можно найти искомое множество $\mathcal{S}$ дефектов за общее число проверок на двух ступенях $\leq N + s + L - 1$. В частности, в этой статье 1983 года с помощью метода случайного кодирования для ансамбля с *независимыми компонентами* кодовых слов было впервые показано, что при соответствующем выборе параметра $L \geq 2$ применение такой двухступенчатой процедуры групповых проверок с СД s_L-кодом на первой ступени находит любое подмножество $|\mathcal{S}|$, $\mathcal{S} \subset [t]$, $|\mathcal{S}| \leq s$, дефектов за общее число проверок, не превосходящее $O(s \log_2 t)$. Далее в работах других авторов в 2000-х годах этот принципиальный результат лишь повторялся и незначительно уточнялся.

В моих совместных с В.В. Рыковым статьях 1980-х годов изучалась логарифмическая асимптотика $(N \to \infty)$ максимального объема $t(N, s, L)$ СД s_L-кодов т.е. строились верхние и нижние границы для функции целочисленных параметров $s, L = 1, 2, \ldots$:

$$R_L(s) \triangleq \overline{\lim_{N\to\infty}} \frac{\log_2 t(N, s, L)}{N}, \tag{1}$$

называемой, следуя традиции комбинаторной теории кодирования, *скоростью* СД s_L-кодов. В частности, из верхней границы для скорости $R_1(s)$, установленной в фундаментальной статье 1982 года, следует, что для одноступенчатой процедуры статических групповых проверок в рассматриваемой модели поиска $\le s$ дефектных элементов множества $[t]$ при $t \to \infty$ необходимо по крайней мере $s^2 \log_2 t/(2\log_2 s)\,(1 + o(1))$ проверок.

В 1989 году для случая $L = 1$ я построил наилучшую к настоящему времени *нижнюю* границу скорости $R_1(s)$ дизъюнктивных s-кодов, основанную на технике оценивания вероятностей больших уклонений в методе случайного кодирования для ансамбля кодов, в котором слова выбираются независимо с равномерным распределением из множества, состоящего из всех $\binom{N}{w}$ двоичных слов длины N и веса w, $1 \le w \le N$. В методе случайного кодирования для такого ансамбля, называемого *равновесным* ансамблем, логарифмическая асимптотика вероятностей больших уклонений вычисляется существенно сложнее, чем для используемого большинством авторов ансамбля с независимыми компонентами кодовых слов, где можно применять более простые методы оценивания логарифмической асимптотики, но получать при этом менее точные нижние оценки скорости $R_1(s)$.

Диссертация В.Ю. Щукина состоит из введения и 4 глав.

Целью первой главы является разработка вероятностных и комбинаторных методов для построения новых нижних границ для скорости СД s_L-кодов. Во второй главе вводятся теоретико-информационных аналоги СД s_L-кодов, называемые почти СД s_L-кодами, и изучается их применение для построения верхней границы вероятности ошибки двухступенчатых алгоритмов поиска дефектов. В третьей главе исследуется уровень значимости (вероятность ошибки) в задаче проверки гипотезы о фиксированной верхней границе мощности случайного множества дефектов в естественном предположении, что разные множества дефектов одинакового объема появляются с одной и той же вероятностью. В четвертой главе В.Ю. Щукин рассматривает границы для q-ичных, $q \ge 2$, обобщений СД s_L-кодов, называемых q-ичными СД s_L-гиперкодами, которые в предыдущих работах других авторов изучались лишь для частного случая $L = 1$, либо для частного случая $q = 2$. Эти коды возникают при разработке методов защиты авторских прав на цифровую продукцию, методов архивации данных и в задачах кодирования-декодирования при передаче сообщений для некоторых важных для приложений моделей симметричного КМД с q-ичными входными символами, передаваемыми по КМД методом частотной модуляции.

В **первой главе** диссертации получена давно ожидаемая мной от учеников нижняя граница $\underline{R}_L(s)$ скорости $R_L(s)$ СД s_L-кодов, представляющая собой границу случайного кодирования для ансамбля равновесных кодов. С этой задачей В.Ю. Щукин успешно справился, существенно обобщив аналитическую технику анализа вероятностей больших уклонений, которая у меня в 1989 году была лишь для частного случая $L = 1$. Для всех фиксированных значений параметров s, $s \ge 2$, и L, $L \ge 2$, построенная им нижняя граница $\underline{R}_L(s)$ существенно улучшает найденные ранее другими авторами границы случайного кодирования на ансамбле с независимыми двоичными компонентами кодовых слов. Применение этих результатов к двухступенчатым процедурам поиска, задаваемым ДКМД, означает, что при $s \ge 3$ наилучшая известная асимптотическая ($t \to \infty$) верхняя оценка минимально возможного числа $N_0(\le s,t)$ групповых проверок для двухступенчатых процедур безошибочного (приводящего к однозначному решению) поиска $\le s$ дефектных элементов из множества $[t]$ имеет вид

$$N_0(\le s,t) \le \frac{\log_2 t}{c_s}\,(1+o(1)), \quad c_s = \lim_{L\to\infty} \underline{R}_L(s) = \log_2\left[\frac{(s-1)^{s-1}}{s^s}+1\right], \quad s\ge 1, \qquad (2)$$

Коэффициент c_s в правой части неравенства (2) можно назвать нижней границей шенноновской пропускной способности двухступенчатых процедур поиска с нулевой ошибкой. Отметим, что при больших s предельное значение $c_s \sim \frac{\log_2 e}{es} = \frac{0.5307}{s}$. Для частного случая $s=2$ возможность улучшения коэффициента $c_2 = \log_2 5 - 2 = 0.3219$ в границе (2) была доказана в 2009 году П. Дамашке, который для построения соответствующей двухступенчатой процедуры поиска не использовал СД 2_L-коды.

Во **второй главе** диссертации автор разрабатывает принципиально важный теоретико-информационный подход для комбинаторных моделей групповых проверок в задачах поиска, когда неизвестное искомое множество дефектов $\mathcal{S}$, $\mathcal{S}\subset[t]$, имеющее заданный объем $|\mathcal{S}|=s$, интерпретируется как случайная величина принимающая равновероятные значения на множестве всех $\binom{t}{s}$ s-подмножеств множества $[t]$. При этом в качестве естественной характеристики двухступенчатой процедуры поиска s дефектов во множестве $[t]$ рассматривается *вероятность ошибки* $\mathcal{P}_N(s,t)$, которая определяется как *доля* числа s-подмножеств среди всех $\binom{t}{s}$ s-подмножеств множества $[t]$, появление которых в качестве дефектных s-наборов приводит к неоднозначному решению (отказу от решения) после проведения всех N групповых проверок этой двухступенчатой процедуры.

Цель данной главы - построить верхнюю границу для вероятности ошибки $\mathcal{P}_N(s,t)$, которая при $t,N\to\infty$ и некотором фиксированном параметре $R\sim\log_2 t/N$, $0<R<1$, называемом скоростью, экспоненциально убывает с ростом N и имеет вид:

$$\mathcal{P}_N(s,t) \le \exp\{-N[E(s,R)+o(1)]\}, \quad c_s < R < C_s, \qquad (3)$$

где c_s определена в правой части неравенства (2), а монотонно убывающую функцию $E(s,R)>0$ параметра R, $c_s<R<C_s$, можно рассматривать как критерий двухступенчатых процедур поиска s дефектов в множестве $[t]$ со скоростью R.

Для вывода границы (3) вводится понятие почти СД s_L-кода, как двоичной $(N\times t)$-матрицы X, для которой *доля* $\mathcal{P}_N(s,L,t)$ числа s-наборов ее кодовых слов, таких что их дизъюнкция покрывает хотя бы одну дизъюнкцию других L, $L\ge 1$, кодовых слов, удовлетворяет условию: $\mathcal{P}_N(s,L,t)\to 0$ при $t,N\to\infty$ и $R\sim\log_2 t/N$, где $R>0$ - фиксированная скорость кода X. Величина $\mathcal{P}_N(s,L,t)$ называется вероятностью ошибки почти СД s_L-кодов. Применяя разработанную им нетривиальную технику оценивания для метода случайного кодирования в ансамбле равновесных кодов, В.Ю. Щукин доказывает существование почти СД s_L-кодов X со скоростью $R>0$ и вероятностью ошибки

$$\mathcal{P}_N(s,L,t) \le \exp\{-N[E_L(s,R)+o(1)]\}, \quad 0<R<C_s, \qquad C_s \sim \frac{\ln 2}{s}, \quad s\to\infty, \qquad (4)$$

где для любых $s\ge 1$ и $L\ge 1$ функция $E_L(s,R)$ параметра R, $0<R<C_s$, монотонно убывает, $E_L(s,R)>0$ и при $R=C_s$ значение $E_L(s,C_s)=0$. Максимальное значение скорости C_s можно назвать границей случайного кодирования для пропускной способности почти СД s_L-кодов. Пусть $L=L(s,R)$ - минимальное среди чисел $\ell\ge 1$, при которых достигается

$\max_{\ell \geq 1} E_\ell(s, R) = E_L(s, R)$. Поэтому граница (3) при $E(s, R) = E_L(s, R)$ вытекает из существования двухступенчатой процедуры поиска, где N групповых проверок на первой ступени задаются почти СД s_L-кодом, $L = L(s, R)$, с вероятностью ошибки, удовлетворяющей (4).

В **третьей главе** диссертации для фиксированных значений t и s, $1 \leq s < t$, неизвестное множество дефектных элементов $\mathcal{S}$, $\mathcal{S} \subset [t]$, $[t] = \{1, 2, \dots t\}$ интерпретируется как случайное множество, о распределении вероятностей которого делается естественное предположение, что разные множества дефектов одинакового объема появляются с одной и той же вероятностью. Исследуется задача проверки гипотезы о справедливости фиксированной верхней границы s, $s \ll t$, для мощности $|\mathcal{S}|$. Для этого проводится N неадаптивных групповых тестов, задаваемых двоичной $(N \times t)$-матрицей X. Автор вводит два алгоритма принятия решение о гипотезе $|\mathcal{S}| \leq s$ против альтернативы $|\mathcal{S}| \geq s + 1$, основанные на N результатах этих тестов, которые представляют собой двоичный столбец Σ длины N, равный булевой сумме столбцов $(N \times t)$-матрицы X с номерами, образующими случайное (неизвестное) множество $\mathcal{S}$.

По первому алгоритму, который называется s-дизъюнктивным, решение в пользу гипотезы $|\mathcal{S}| \leq s$ $(|\mathcal{S}| \geq s + 1)$ принимается, когда Σ покрывает $\leq s$ $(\geq s + 1)$ столбцов матрицы X. Автор показывает, что уровень значимости (вероятность ошибки) такого решения совпадает с вероятностью ошибки $\mathcal{P}_N(s, 1, t)$ почти СД s_1-кодов, исследуемых в главе 2. При $t, N \to \infty$ и фиксированной скорости $R \sim \log_2 t/N$, $0 < R < 1$, экспоненциально убывающая с ростом N верхняя граница $\alpha_N^1(s, R)$ данной вероятности, построенная в главе 2 методом случайного кодирования на ансамбле равновесных кодов, имеет вид:

$$\mathcal{P}_N(s, 1, t) \leq \alpha_N^1(s, R) = \exp\{-N[E_1(s, R) + o(1)]\}, \quad 0 < R < C_s \sim \frac{\ln 2}{s}, \quad s \to \infty,$$

где монотонно убывающую функцию $E_1(s, R) > 0$ параметра R, $0 < R < C_s$, можно назвать критерием s-дизъюнктивного алгоритма для кодов X со скоростью R.

По второму алгоритму, который называется T-пороговым, где T, $0 < T < N$, – фиксированное целое число, решение в пользу гипотезы $|\mathcal{S}| \leq s$ $(|\mathcal{S}| \geq s + 1)$ принимается, когда число единиц (вес) двоичного столбца Σ не превышает порог T (превышает порог $T + 1$). Ранее в теории групповых проверок T-пороговый алгоритм не изучался. Основным аналитическим результатом главы 3 является вычисление для ансамбля равновесных кодов логарифмической асимптотики при $N \to \infty$ и фиксированном s границы случайного кодирования $\alpha_N^2(s, T)$ для уровня значимости (вероятности ошибки) T-порогового алгоритма при оптимальном выборе порога $T \sim \tau_s N$, $0 < \tau_s < 1$. Автор устанавливает, что

$$\alpha_N^2(s, \tau_s N) = \exp\{-N[E_{Thr}(s) + o(1)]\}, \qquad E_{Thr}(s) \sim \frac{\log_2 e}{4s^2}, \quad s \to \infty.$$

Показатель экспоненты $E_{Thr}(s)$ можно рассматривать как критерий T-порогового алгоритма. Его сравнение с критерием $E_1(s, R)$, $0 < R < C_s$, позволяет найти пороговое значение скорости $R_{Thr}(s) = \sup\{R : E_1(s, R) \geq E_{Thr}(s)\}$, $0 < R_{Thr}(s) < C_s$, такое, что при любом значении R, $R_{Thr}(s) < R < C_s$, s-дизъюнктивный алгоритм для кодов со скоростью R будет менее эффективным, чем $\tau_s N$-пороговый алгоритм.

В **четвертой главе** диссертации В.Ю. Щукин переносит комбинаторные постановки задач по исследованию границ скорости двоичных кодов со списочным декодированием, которые в первой главе изучались для ДКМД, на детерминированный симметричный канала множественного доступа, называемый q-ичным, $q \geq 2$, гиперканалом множественного доступа, который имеет один выход и s входов. Входными сигналами являются элементы некоторого q-ичного алфавита, а выходной сигнал для любого s-набора входных q-ичных символов определяется как подмножество данного q-ичного алфавита, представляющее собой объединение элементов этого s-набора. Код объема t и длины N из элементов q-ичного алфавита, являющийся аналогом двоичного СД s_L-кода, называется q-ичным СД s_L-гиперкодом и определяется условием: для любого s-подмножества кодовых слов и любого L-подмножества из каких-либо других кодовых слов среди N строк кода найдется хотя бы одна строка, в которой L-подмножество содержит хотя бы один элемент, не совпадающий ни с одним из элементов этой строки, входящих в s-подмножество.

В 4 главе диссертант устанавливает новые и существенно улучшающие предыдущие результаты других авторов верхние $\overline{R}_L^q(s)$ и нижние $\underline{R}_L^q(s)$ границы для скорости q-ичных СД s_L-гиперкодов $R_L^q(s)$, определяемой равенством (1), где логарифм максимального объема q-ичных СД s_L-гиперкодов берется по основанию q. Для больших значений s при фиксированных $L \geq 1$ и $q \geq 2$ полученная автором верхняя граница $\overline{R}_L^q(s)$ имеет вид

$$\overline{R}_L^q(s) = \frac{2L(q-1)\log_q s}{s^2}(1+o(1)), \quad s \to \infty,$$

и основана на уже отмеченной выше верхней границе $\overline{R}_1(s) = 2\log_2 s/s^2$, построенной мной и В.В. Рыковым в статье 1982 года для скорости $R_1(s)$ классических дизъюнктивных s-кодов.

Две нижние границы $\underline{R}_L^q(s)$, представленные в этом разделе, получены методом случайного кодирования. Для первой границы $\underline{R}_L^2(s)$, доказанной для частного случая $q = 2$ надо подчеркнуть технически весьма непростой вывод с использованием равновесного ансамбля двоичных кодов. Но на мой взгляд наибольший интерес в этой главе представляет вторая нижняя граница $\underline{R}_L^q(s)$, для доказательства которой автор существенно развивает и обобщает метод случайного кодирования, предложенный мной в 2003 году для дизъюнктивных двоичных кодов со списочным декодированием. Важно отметить предельную форму этой нижней границы

$$\underline{R}_\infty^q(s) = \lim_{L\to\infty} \underline{R}_L^q(s) = \frac{(q-1)\log_q e}{es}(1+o(1)), \quad s \to \infty,$$

где главный член правой части можно интерпретировать как нижнюю границу пропускной способности с нулевой ошибкой для разработанной мной и В.В. Рыковым в статье 1981 года сети связи с большим числом t пользователей, центральной станцией и КМД, в котором при передаче q-ичных символов в каждый момент времени применяется частотная модуляция, а для необходимой за время передачи N идентификации s, $s \ll t$, активных пользователей сети связи применяется q-ичный СД s_L-гиперкод объема t и длины N.

Из приведенного перечня результатов и моего комментария видно, что стержнем, соединяющим все 4 главы диссертации является систематическое применение техники оценивания вероятностей больших уклонений в методе случайного кодирования для двоичного ансамбля равновесных кодов. Существенные аналитические трудности здесь связаны с нетривиальным применением стандартного метода множителей Лагранжа при решении необходимо возникающих задач оптимизации выпуклых функций. Для двоичного дизъюнктивного канала множественного доступа, а также для q-ичного гиперканала множественного доступа, это позволило построить новые существенно более точные нижние границы скорости оптимальных кодов, чем нижние границы для аналогичных оптимальных кодов, ранее полученные другими авторами, использующими ансамбль с независимыми компонентами кодовых слов. При этом автором диссертации проявлена большая изобретательность при решении конкретных задач, а также свободное владение различными вероятностными, комбинаторными, аналитическими и числовыми методами. Отмечу его самостоятельность в главе 3 как при постановке, так и при решении новой для теории дизъюнктивных кодов и теории групповых проверок теоретико-вероятностной задачи проверки гипотезы о справедливости фиксированной верхней границы для объема случайного множества дефектов.

Диссертация В.Ю. Щукина несомненно удовлетворяет всем требованиям «Положения о порядке присуждения ученых степеней» Высшей аттестационной комиссии Министерства образования и науки Российской Федерации, а ее автор, Щукин Владислав Юрьевич, заслуживает присуждения ему ученой степени кандидата физико - математических наук по специальности 01.01.05 - «теория вероятностей и математическая статистика».

Научный руководитель:
доктор физико - математических наук
по специальности 01.01.05,
профессор кафедры теории вероятностей
механико-математического факультета
ФГБОУ ВО «Московский государственный
университет им. М.В. Ломоносова» — Дьячков Аркадий Георгиевич
тел. +7(495)939-1403,
электронная почта: agd-msu@yandex.ru — 21.09.2016

Подпись профессора А.Г. Дьячкова заверяю
и.о. декана механико - математического факультета МГУ,
доктор физико - математических наук,
профессор — Чубариков Владимир Николаевич