

МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ, СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«СИБИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ТЕЛЕКОММУНИКАЦИЙ И ИНФОРМАТИКИ»

Вестник СибГУТИ Том 16, № 1, 2022

Выпускается ежеквартально, выходит с 2007 г.

Учредитель –

федеральное государственное бюджетное образовательное учреждение высшего образования
«Сибирский государственный университет телекоммуникаций и информатики»

Председатель редакционного совета А. Н. Фионов, д.т.н., проф.

Редакционный совет:

С. С. Абрамов, д.т.н., доц.	С. Н. Новиков, д.т.н., доц.
В. Б. Барахнин, д.т.н., доц.	В. И. Носов, д.т.н., проф.
В. М. Белов, д.т.н., проф.	С. В. Поршнев, д.т.н., проф.
В. Н. Васюков, д.т.н., проф.	А. С. Родионов, д.т.н., доц.
В. Ю. Васильев, д.х.н., доц.	А. И. Романенко, д.ф.-м.н., проф.
Н. И. Горлов, д.т.н., проф.	Б. Я. Рябко, д.т.н., проф.
Н. Л. Казначеева, д.э.н., доц.	И. И. Рябцев, д.ф.-м.н., чл.-корр. РАН
В. С. Канев, д.т.н., проф.	Э. Сименс, д.т.н.
А. И. Карпович, д.э.н., проф.	О. В. Стукач, д.т.н., доц.
М. Г. Курносов, д.т.н., проф.	В. К. Трофимов, д.т.н., проф.
В. В. Лебедянцева, д.т.н., проф.	А. И. Фалько, д.т.н., проф.
А. В. Лихачёв, д.т.н.	С. В. Федоренко, д.т.н., доц.
С. Н. Мамоиленко, д.т.н., доц.	Б. Г. Хаиров, д.э.н., доц.
В. Г. Мартынец, д.ф.-м.н.	А. Г. Черевко, к.ф.-м.н., доц.
О. Г. Мелентьев, д.т.н., проф.	С. В. Шидловский, д.т.н.
Р. В. Мещеряков, д.т.н., проф.	Ю. И. Шокин, д.ф.-м.н., акад. РАН
И. Г. Неизвестный, д.ф.-м.н., чл.-корр. РАН	В. П. Шувалов, д.т.н., проф.

Редакция:

А. Н. Фионов (главный редактор), М. Ю. Галкина (заведующая редакцией),
Н. А. Двуреченская (технический и литературный редактор, компьютерная вёрстка),
Т. А. Алфёрова (лингвист-корректор).

Адрес редакции и издателя

630102, г. Новосибирск, ул. Кирова, д. 86

e-mail: vestnik@sibsutis.ru

Информация о журнале доступна в сети Internet по адресу <http://vestnik.sibsutis.ru>.

Журнал включён в Перечень ВАК российских рецензируемых научных журналов, в которых должны быть опубликованы основные научные результаты диссертаций на соискание учёных степеней доктора и кандидата наук.

Журнал зарегистрирован Федеральной службой по надзору за соблюдением законодательства в сфере массовых коммуникаций и охране культурного наследия, свидетельство о регистрации ПИ № ФС77-25835 от 29.09.2006. ISSN 1998-6920. Подписной индекс в объединённом каталоге «Пресса России» – 82519.

Дата выхода в свет 20.05.2022. Отпускная (редакционная) цена 800 руб.

Отпечатано в издательском центре СибГУТИ по адресу: 630102, г. Новосибирск, ул. Кирова, д. 86.

Бумага офсетная, формат А4. Тираж 300 экз.

© СибГУТИ, 2022 г.

СОДЕРЖАНИЕ

В. А. Карболин, В. И. Носов

Рекомендации по выбору индекса модуляции в КСШП-системах радиосвязи с модуляцией BPPM при работе внутри помещений 3

Т. В. Авдеенко, Н. В. Пустовалова

Многофакторный анализ влияния особенностей студентов на академическую успеваемость для персонализации образовательной среды 18

В. Н. Лиманский, Б. З. Горбунов

Метод селекции ионов для их идентификации в газовых средах 35

Е. П. Мачикина, В. Г. Насенник

Конструкция равновесных корректирующих кодов с синхронизацией 46

В. Г. Арсентьев, Г. И. Криволапов

Подавление помех судоходства при гидроакустическом поиске подводного объекта 56

А. Н. Паршуков

Математическая модель технологического процесса абсорбции природного газа ... 68

А. Ю. Филатов, В. В. Михеев

Анализ эффективности потоково-локальной сборки мусора в распределённых системах хранения и обработки данных 77

М. П. Базилевский

Исследование поведения относительных вкладов переменных в общую детерминацию в оцененном на основе метода выпрямления искаженных коэффициентов регрессионном уравнении 89

Г. Ы. Токтошов

Методология выбора трасс для прокладки сетей и коммуникаций 97

Рекомендации по выбору индекса модуляции в КСШП-системах радиосвязи с модуляцией ВРРМ при работе внутри помещений

В. А. Карболин, В. И. Носов

Проведено исследование вероятности битовой ошибки для модуляции ВРРМ в зависимости от индекса модуляции для каналов внутри помещений КСШП-систем радиосвязи. Приводятся рекомендации по выбору индекса модуляции, оптимального с точки зрения минимальной вероятности битовой ошибки в многолучевом канале передачи данных. Получено аналитическое выражение для вероятности битовой ошибки для модуляции ВРРМ. Представлены графики зависимости вероятности битовой ошибки от отношения сигнал/шум для различных скоростей передачи данных, вычисленных с использованием аналитического выражения и полученных с помощью компьютерного моделирования в системе MATLAB.

Ключевые слова: короткоимпульсные сверхширокополосные системы радиосвязи, помехоустойчивость, индекс модуляции, двоичная позиционно-импульсная модуляция, ВРРМ, UWB, Indoor UWB, КСШП.

1. Введение

Известно, что для двоичной позиционно-импульсной модуляции (дПИМ; Bipolar Pulse Position Modulation, ВРРМ) в канале АБГШ короткоимпульсных сверхширокополосных систем радиосвязи (КСШП) существует оптимальный с точки зрения минимума вероятности битовой ошибки индекс модуляции δ , который определяется исходя из минимума нормированной функции АКФ [1]:

$$\delta = \arg \left\{ \max_{\delta} \{R(0) - R(\delta)\} \right\}, \quad (1)$$

где δ – индекс модуляции;

$$R(\tau) = \frac{1}{R(0)} \int_{-\infty}^{\infty} (p(t) \cdot p(t - \tau)) d\tau = \frac{1}{E_p} \int_{-\infty}^{\infty} (p(t) \cdot p(t - \tau)) d\tau \quad - \text{ функция нормированной}$$

АКФ;

$p(t)$ – форма модулированного КСШП-символа; $E_p = \int_{-\infty}^{\infty} p^2(t) dt$ – энергия импульса.

Из графика функции нормированной АКФ для канала без многолучёвости (рис. 1) следует, что оптимальный индекс модуляции $\delta = 0.9 \times 10^{-10}$ с.

Также известно, что для каналов с многолучевым распространением полученное вышеописанным способом значение индекса модуляции не будет являться оптимальным [1–3]. Более того, вероятность битовой ошибки зависит от импульсной характеристики (ИХ) канала передачи, и при модуляции дПИМ в каналах с многолучевым распространением эта вероят-

ность подвержена асимметрии при приёме нулевого и единичного символов. Также в [4] упоминается связь вероятности ошибки в КСШП-системах связи в многолучевой среде с интервалом $[A, B]$, который зависит от скорости передачи. Таким образом, существует зависимость вероятности ошибки в КСШП-системах связи в многолучевой среде от: ИХ канала радиосвязи; скорости передачи; формы используемых импульсов; выбранного вида модуляции.

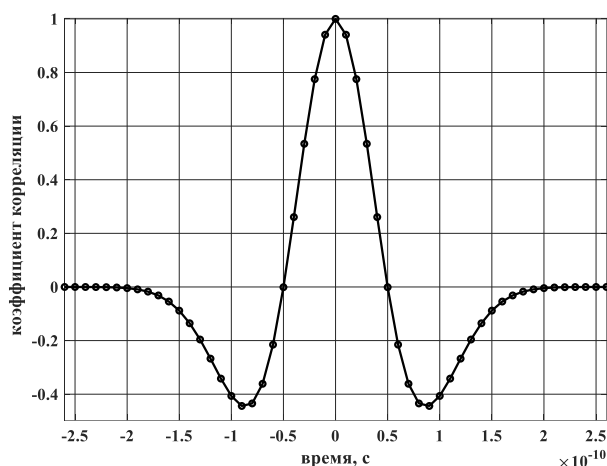


Рис. 1. Функция нормированной АКФ

При действии систем КСШП радиосвязи внутри помещений образуется множество путей распространения сигнала от источника к приёмнику, поэтому необходимо тщательно подходить к выбору индекса модуляции. В данной работе предлагаются рекомендации по выбору оптимального индекса модуляции в зависимости от скорости передачи на основе анализа вероятности ошибки в многолучевой среде распространения сигнала.

2. Вид модуляции, импульсная характеристика канала передачи, структурная схема приёмника

Канал передачи данных и его ИХ, применяемые в исследовании, построены по рекомендациям, указанным в [5] (модель канала СМ-1). Подробное описание применяемого вида модуляции, канала передачи и модели системы радиосвязи дано в [6], в данной работе приведём краткие сведения. На рис. 2а представлен вид символов двоичной позиционно-импульсной модуляции, соответствующих передаче «0» и «1» во входном цифровом потоке. На рис. 2б приведена импульсная характеристика многолучевого канала, соответствующая модели СМ-1. Радиоприёмное устройство, используемое в модели системы радиосвязи (рис. 3), построено на основе схемы корреляционного приёма. Данная схема характеризуется наличием локальной копии принимаемого сигнала (шаблона для корреляции).

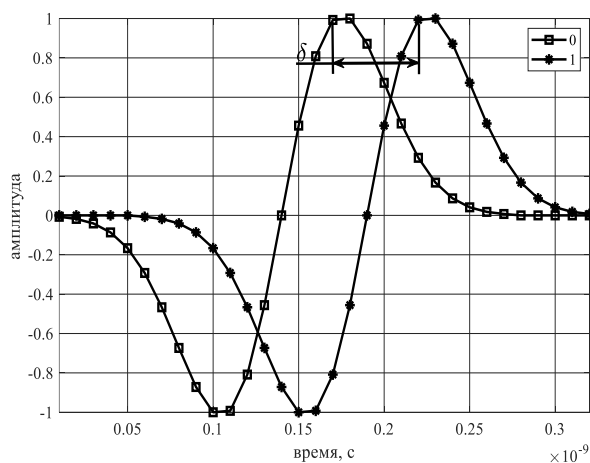


Рис. 2а. Канальные символы модуляции

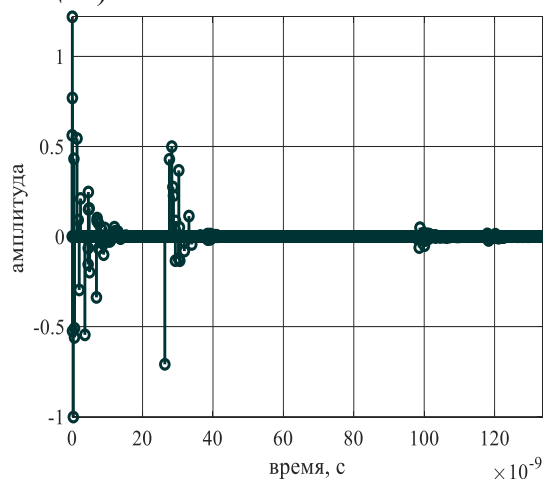


Рис. 2б. ИХ радиоканала (СМ-1)

3. Определение вероятности ошибки для каналов с многолучевым распространением при модуляции ВРРМ

Сигнал с модуляцией ВРРМ можно представить в виде:

$$s(t) = \sum_m p(t - mT - d_m \cdot \delta), \quad (2)$$

где d_m – принимает значения 0 и 1;

δ – индекс модуляции;

T – длительность символа;

$p(t)$ – форма передаваемого символа, в данном случае вторая производная импульса Гаусса

$$\left(g_2(t) = \frac{d(g_1(t))}{dt} = -K_2 \cdot (2t/\tau^2) \cdot e^{-(t/\tau)^2}, \quad K_2 = \sqrt{\frac{\tau E_1}{\sqrt{\pi/2}}} \right).$$

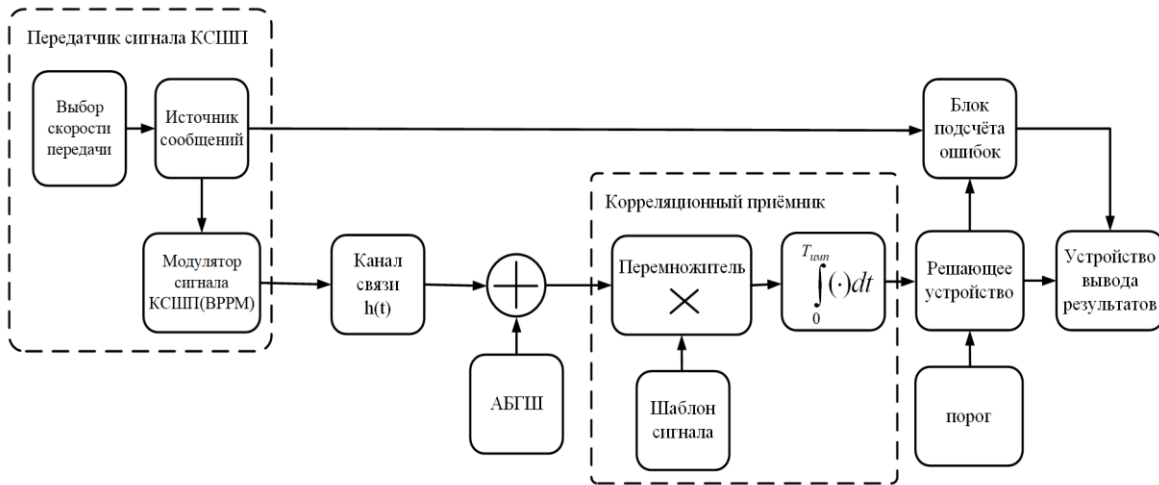


Рис. 3. Модель КСШП-системы радиосвязи

В корреляционном приёмнике определяется коэффициент корреляции для модуляции ВРРМ:

$$R(\tau) = \frac{1}{E_p} \int_{-\infty}^{\infty} p(t)p(t-\tau)dt. \quad (3)$$

График коэффициента корреляции (нормированной АКФ) представлен на рис. 1. Принятый сигнал, прошедший через канал связи, можно представить выражением:

$$r(t) = s(t) * h(t) + n(t) = \sum_{l=1}^L \sum_{m=1}^{\infty} \alpha_l \cdot p(t - mT - d_m \cdot \delta - \tau_l) + n(t), \quad (4)$$

где $h(t)$ – ИХ радиоканала;

α_l , τ_l – амплитудные коэффициенты и времена прихода лучей ИХ в точку приёма соответственно;

$n(t)$ – АБГШ; $p(t)$ – форма передаваемого символа;

$*$ – операция свёртки.

Обычно для двоичной ПИМ шаблон для корреляции в приёмнике задаётся выражением:

$$\omega(t) = p(t) - p(t - \delta), \quad (5)$$

где $p(t)$ – форма передаваемого символа;

δ – индекс модуляции.

Тогда напряжение на выходе коррелятора приёмника будет равно:

$$g(mT) = \int_0^{T_{symp}} r(t) \cdot \omega(t) \cdot dt = \int_0^{T_{symp}} r(t) \cdot [p(t - mT - \tau_0) - p(t - \delta - mT - \tau_0)] \cdot dt, \quad (6)$$

где τ_0 – время прихода первого луча в приёмник.

Решение принимается на основе:

$$\begin{aligned} g(mT) \geq 0 &\rightarrow \text{символ } 0 \\ g(mT) < 0 &\rightarrow \text{символ } 1 \end{aligned} \quad (7)$$

Необходимо заметить, что если бы длительность ИХ была меньше длительности символа T , то при передаче символа среднестатистическое значение напряжения на выходе коррелятора:

$$u_0 = M\{g(t)/0\} = E_p \sum_{l=0}^{L-1} \alpha_l \cdot \{R(\tau_l - \tau_0) - R(\tau_l - \tau_0 - \delta)\}, \quad (8a)$$

соответственно, при передаче единичного символа среднестатистическое значение напряжения на выходе коррелятора:

$$u_1 = M\{g(t)/1\} = E_p \sum_{l=0}^{L-1} \alpha_l \cdot \{R(\tau_l - \tau_0 + \delta) - R(\tau_l - \tau_0)\}, \quad (8б)$$

где $M\{\cdot\}$ – операция математического ожидания; $R(t)$ – функция коэффициента корреляции.

Но поскольку длительность ИХ намного больше длительности символа, то возникает межсимвольная интерференция (МСИ). В выражениях (8а), (8б), так же как в работах [2, 3], учитывается интерференция от отсчётов ИХ канала связи в пределах одного символа (длительность ИХ радиоканала меньше длительности символа). Чтобы учесть МСИ, необходимо изменить выражения (8а), (8б).

Рассмотрим процесс многолучевого распространения сигнала через анализ ИХ. Для простоты рассмотрим двоичную амплитудную импульсную модуляцию. На графике, изображённом на рис. 4, представлена тестовая ИХ канала (а) и тестовая импульсная последовательность (модуляция двоичная амплитудная импульсная (дАИМ, ВРАМ)) (б) для передачи через канал.

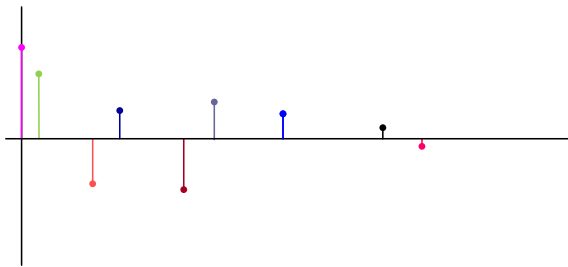


Рис. 4а. Тестовая ИХ канала

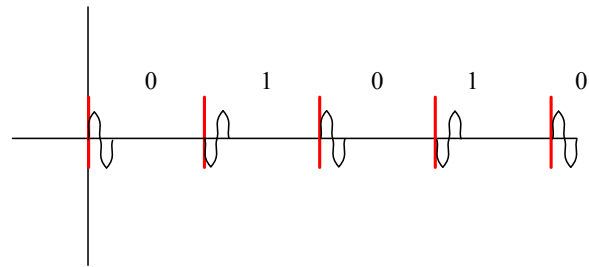


Рис. 4б. Тестовая последовательность

На рис. 5 представлены отклики канала на входную тестовую последовательность от 1-го (а), 2-го (б), 5-го (в) лучей и совмещённые отклики от всех лучей с окном интегрирования в приёмнике (г).

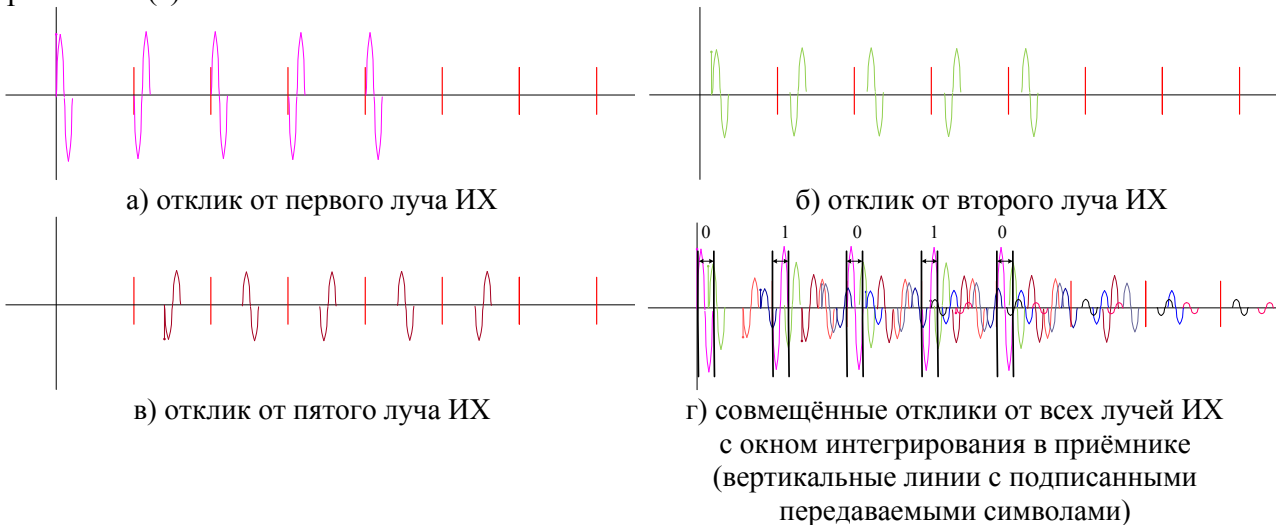
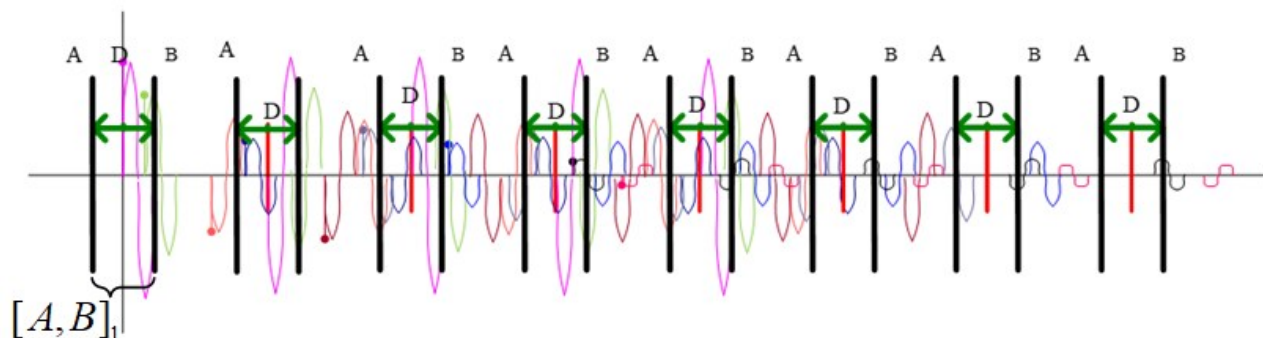


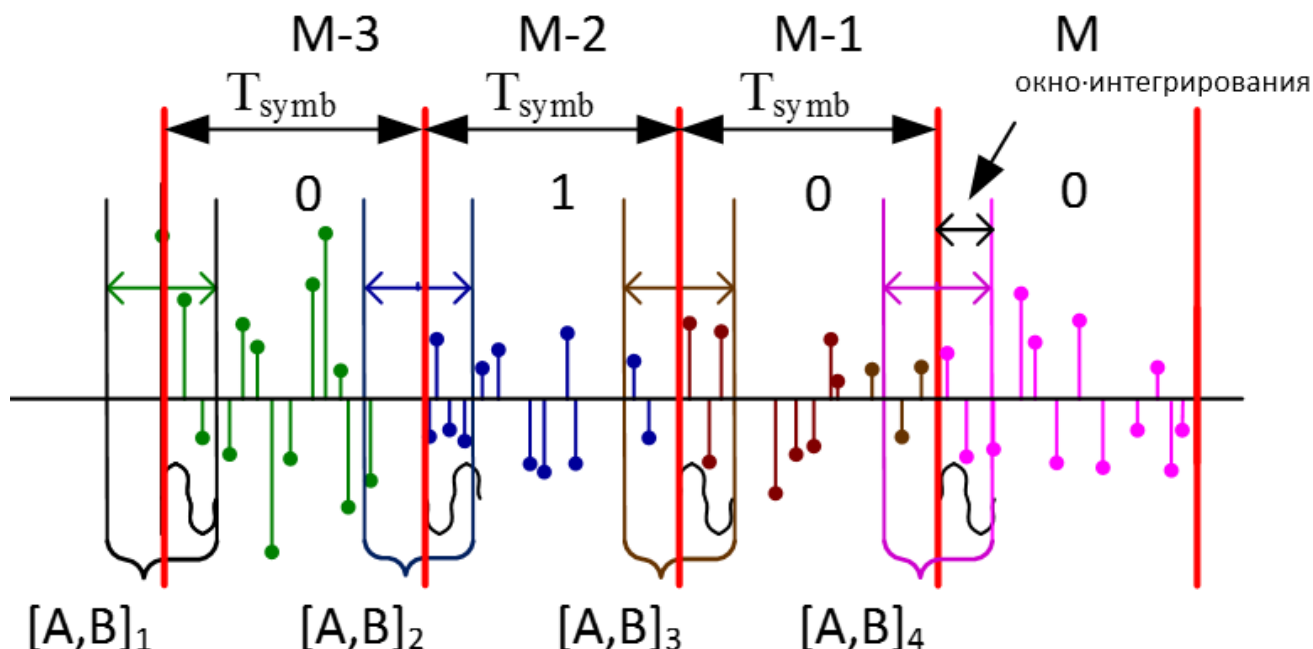
Рис. 5. Отклики канала на входную тестовую последовательность

В работе [4] введено понятие интервала $[A, B]$ – это интервал, при попадании отсчётов ИХ в который появляется МСИ. Этот интервал равен удвоенному значению длительности импульса $p(t)$. Это понятие можно использовать для оценки количества отсчётов ИХ, которые будут влиять на МСИ. Используя это понятие, можно увидеть, что при длительной передаче, намного большей по времени, чем длительность ИХ, все отсчёты ИХ с кратностью периода символа, находящиеся в пределах интервала $[A, B]$, будут вносить вклад в МСИ. На рис. 6 показаны интервалы $[A, B]_k$ для тестовой передаваемой последовательности и ИХ, представленных на рис. 4. Точка D является началом окна интегрирования в приёмнике.

Рис. 6. Интервалы $[A, B]$

Заметим, что отсчёты ИХ, входящие в первый интервал $[A, B]_1$ (во вторую его половину, интервал $[D, B]_1$ (рис. 6)), будут давать вклад, который усиливает интерференцию от текущего передаваемого символа, остальные же отсчёты ИХ, относящиеся к интервалам $[A, B]_{k \neq 1}$, следующим с кратностью периода символов, будут вносить искажения от предшествующих символов передаваемой последовательности. Данное обстоятельство продемонстрировано на рис. 7. Формула для нахождения отклика канала на интервале окна интегрирования символа M :

$$\text{значение}_{\text{ОИИМ}}(M) = s_M(t) * [A, B]_1 + s_{M-1}(t) * [A, B]_2 + s_{M-2}(t) * [A, B]_3 + s_{M-3}(t) * [A, B]_4 \quad (9)$$

Рис. 7. Интервалы $[A, B]_{i=1...4}$, создающие МСИ на интервале окна интегрирования символа M

Чтобы учесть МСИ от интервалов $[A, B]_k$ для дПИМ, необходимо увеличить интервал $[A, B]$ на удвоенную величину индекса модуляции: $2 \times \delta$. На рис. 8 показан интервал $[A, B]$ для двоичной ПИМ ($[A, B]_{\text{дПИМ}}$) и интервал $[A, B]$ для двоичной АИМ ($[A, B]_{\text{дАИМ}}$).

$$\left[E_p \sum_{m=1}^{K-1} \sum_{l=0}^{L-1} \alpha_l \cdot \{R(\tau_l - \tau_0 - mT + \delta) - R(\tau_l - \tau_0 - mT)\} \right]_{\text{для } [A, B]_{k, k \neq 1}}.$$

Соответственно, при передаче символа «1» среднестатистическое значение напряжения на выходе коррелятора:

$$\begin{aligned} u_1 &= M \{g(t)/1\}_{\text{для } [A, B]_1} + \\ &+ M \left\{ \begin{aligned} &\{g(t)/0\}_{\text{для } [A, B]_{k, k \neq 1}}, \text{ если предшествующий передаваемый символ – ноль } \{0\} \\ &\{g(t)/1\}_{\text{для } [A, B]_{k, k \neq 1}}, \text{ если предшествующий передаваемый символ – единица } \{1\} \end{aligned} \right\} = \\ &= M \left(\left[E_p \sum_{l=0}^{L-1} \alpha_l \cdot \{R(\tau_l - \tau_0 + \delta) - R(\tau_l - \tau_0)\} \right]_{\text{для } [A, B]_1} \right) + \\ &+ M \left(\left[\begin{aligned} &\left[E_p \sum_{m=1}^{K-1} \sum_{l=0}^{L-1} \alpha_l \cdot \{R(\tau_l - \tau_0 - mT) - R(\tau_l - \tau_0 - mT - \delta)\} \right]_{\text{для } [A, B]_{k, k \neq 1}} \{0\} \\ &\left[E_p \sum_{m=1}^{K-1} \sum_{l=0}^{L-1} \alpha_l \cdot \{R(\tau_l - \tau_0 - mT + \delta) - R(\tau_l - \tau_0 - mT)\} \right]_{\text{для } [A, B]_{k, k \neq 1}} \{1\} \end{aligned} \right] \right) \end{aligned} \quad (11)$$

Условная дисперсия шума при передаче «0» и «1»:

$$\sigma_0^2 = \text{var} \{g/0\} = \text{var} \{g/1\} = \sigma_1^2 = \sigma^2 = (N_0/2) E_p \cdot (1 - \rho), \quad (12)$$

где $\rho = R(\delta)$ – значение функции коэффициента корреляции при смещении, равном индексу модуляции δ .

Условная плотность вероятности при передаче «0»:

$$p(g/0) = \left(1/\sqrt{2\pi\sigma^2}\right) \exp\left\{-(g-u_0)^2/2\sigma^2\right\}. \quad (13)$$

Условная плотность вероятности при передаче «1»:

$$p(g/1) = \left(1/\sqrt{2\pi\sigma^2}\right) \exp\left\{-(g-u_1)^2/2\sigma^2\right\}. \quad (14)$$

Условная вероятность ошибки при передаче «0»:

$$P_E(g/0) = \int_{-\infty}^0 p(g/0) dg = \int_{u_0/\sigma}^{\infty} (1/\sqrt{2\pi}) \exp\left\{-x^2/2\right\} dx = Q(u_0/\sigma). \quad (15)$$

Условная вероятность ошибки при передаче «1»:

$$P_E(g/1) = \int_0^{\infty} p(g/1) dg = \int_{-u_1/\sigma}^{\infty} (1/\sqrt{2\pi}) \exp\left\{-x^2/2\right\} dx = Q(-u_1/\sigma). \quad (16)$$

Так как передача «0» и «1» равновероятна ($P_0 = P_1 = 0.5$), для двоичной системы передачи имеем вероятность ошибки:

$$P_E = 0.5 \cdot Q(u_0/\sigma) + 0.5 \cdot Q(-u_1/\sigma), \quad (17)$$

где $Q(z) = \int_z^{\infty} (1/\sqrt{2\pi}) \exp\left\{-z^2/2\right\} dz$.

Используя выражения (10)–(17), можно рассчитать вероятности ошибки для различных индексов модуляции, скоростей передачи.

4. Результаты проведения экспериментальной части

4.1. Эксперимент № 1

1. В качестве основы для проведения расчётов выбрали отношение сигнал/шум (SNR) равным вектору значений от -16 дБ до +16 дБ с шагом 0.5 дБ. Отношение сигнал/шум в логарифмическом масштабе ($SNR_{дБ}$) соответствует этому отношению в линейном масштабе $SNR_{lin} = 10^{SNR(дБ)/10}$.

2. Вычислили соотношение E_b/N_0 (E_b – энергия бита, N_0 – спектральная плотность мощности (СПМ) шума) для различных скоростей передачи данных, вектор скорости был выбран с дискретными значениями: $R_b = (100; 400; 1000; 1400)$ Мбит/с. Отношение SNR_{lin} связано с отношением в линейном масштабе $(E_b/N_0)_{lin}$:

$$SNR_{lin} = \frac{E_b \cdot R_b}{(N_0/2) \cdot B} = \frac{E_b \cdot (1/T_b)}{(N_0/2) \cdot B} = 2 \cdot \frac{E_b/N_0}{\tau_u \cdot q \cdot B}, \quad T_b = \frac{1}{R_b}, \quad (18)$$

где B – полоса, в которой сосредоточена СПМ шума, в данном случае её можно принять равной частоте дискретизации сигнала $F_s = 10^{11}$ Гц [7];

τ_u – длительность импульса;

$q = T_b/\tau_u$ – скважность.

Необходимо отметить: поскольку модуляция двоичная, то время передачи одного бита равно времени передачи одного символа, следовательно, имеем: $T_b = T_{symb}$, $R_b = R_{symb}$, $E_b = E_{symb}$, где T_{symb} , R_{symb} , E_{symb} – время передачи, скорость и энергия символа соответственно. Из (18) получаем:

$$(E_b/N_0)_{lin} = 0.5 \cdot SNR_{lin} \cdot \tau_u \cdot q \cdot F_s. \quad (19)$$

3. Создали массив индексов модуляции $\delta = (5; 6; 7; 8; 9; 10; 11; 25) \cdot 10^{-11}$ с.

4. Создали вектор из 1000 двоичных элементов, передали через радиоканал с заданной ИХ (рис. 2б) с разными скоростями передачи из вектора R и с разными отношениями сигнал/шум (ОСШ), а также изменяли индекс модуляции из вектора δ . В результате рассчитали экспериментальную и аналитическую вероятности ошибки, используя формулы (10)–(17). Чтобы результат, полученный аналитически, соответствовал результату, полученному экспериментально, отношения SNR вычислялись до прохождения канала, т.е. вычислялась энергия символа по правилам численного интегрирования: $E_{symb} = T_s \cdot \sum_{n=0}^{L-1} x_n^2$, где T_s – интервал дискретизации; x_n – дискретные отсчёты сигнала во времени; L – количество отсчётов в символе (активная часть сигнала и пауза). Рассчитывали мощность передаваемого символа: $P_{symb} = E_{symb}/T_{symb} = T_s \cdot \sum_{n=0}^{L-1} x_n^2 / L \cdot T_s = \sum_{n=0}^{L-1} x_n^2 / L$. Соответственно, мощность шума равна $N = P_{symb}/SNR$. Модель шума определяется по выражению: $n = \sqrt{N} \cdot randn(1, L)$, $randn(1, L)$ – выборка из стандартного нормального распределения размером $(1 \times L)$. В итоге получили одну реализацию вероятности ошибки для каждого значения векторов скорости и шума.

5. Значения для вероятности ошибки аналитическим способом получены в результате усреднения ансамбля реализаций в п. 4 по 1000 независимым испытаниям (различным реализациям вектора двоичных символов), аналогично получены значения в результате проведения эксперимента с использованием модели, усредненные по 30 независимым испытаниям. Данные представлены в виде графиков вероятности ошибки для различных скоростей передачи данных и индексов модуляции в зависимости от SNR (рис. 9–12).

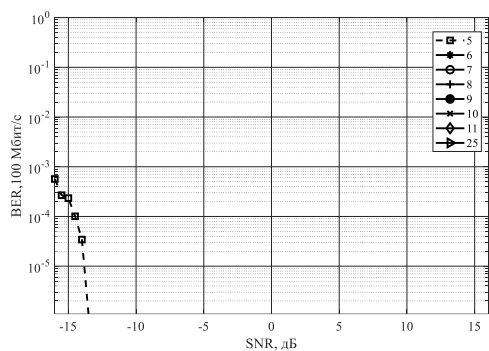


Рис. 9а. Скорость 100 Мбит/с (данные получены экспериментально с помощью модели на рис. 3, усреднение по 30 реализациям)

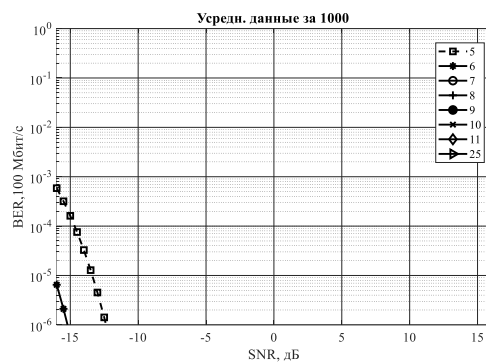


Рис. 9б. Скорость 100 Мбит/с (данные получены через аналитические зависимости (10)–(17), усреднение по 1000 реализациям)

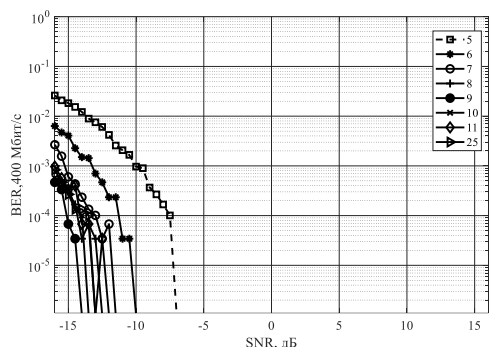


Рис. 10а. Скорость 400 Мбит/с (данные получены экспериментально с помощью модели на рис. 3, усреднение по 30 реализациям)

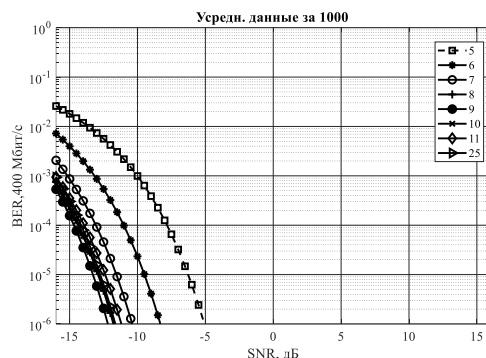


Рис. 10б. Скорость 400 Мбит/с (данные получены через аналитические зависимости (10)–(17), усреднение по 1000 реализациям)

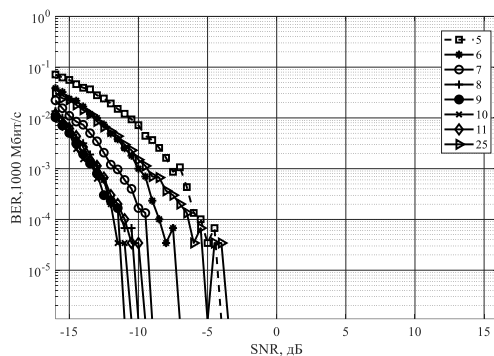


Рис. 11а. Скорость 1000 Мбит/с (данные получены экспериментально с помощью модели на рис. 3, усреднение по 30 реализациям)

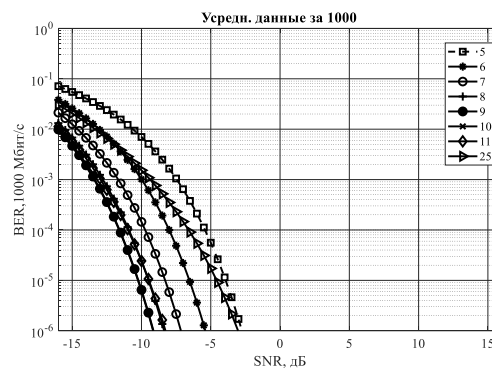


Рис. 11б. Скорость 1000 Мбит/с (данные получены через аналитические зависимости (10)–(17), усреднение по 1000 реализациям)

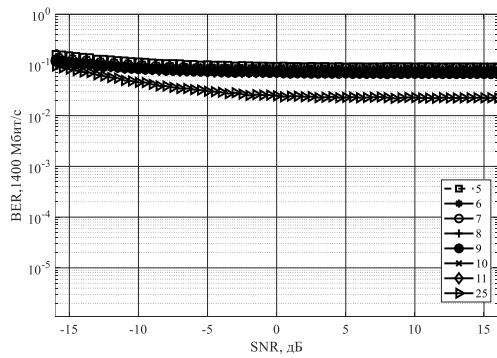


Рис. 12а. Скорость 1400 Мбит/с (данные получены экспериментально с помощью модели на рис.3, усреднение по 30 реализациям)

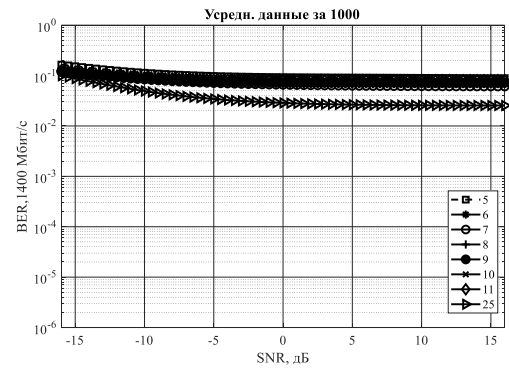


Рис. 12б. Скорость 1400 Мбит/с (данные получены через аналитические зависимости (10)–(17), усреднение по 1000 реализациям)

4.2. Эксперимент № 2

В этом эксперименте задались вектором значений E_b/N_0 от -16 дБ до +16 дБ с шагом 0.5 дБ. Этапы проведения эксперимента соответствуют этапам эксперимента № 1 из п. 4.1. Отметим факт расчёта мощности шума при использовании модели на рис. 3. Зная вектор значений E_b/N_0 , перешли к вектору значений в линейном масштабе $(E_b/N_0)_{lin}$. Зная значение энергии сигнала, вычислили СПМ шума N_0 . Для вещественных сигналов формула для расчёта требуемой мощности: $N = 0.5 \times N_0 \times F_s$ [7]. Далее, следуя п. 4 подраздела 4.1, получили вектор шума на интервале длительности символа: $n = \sqrt{N} \cdot randn(1, L) = \sqrt{0.5 \cdot N_0 \cdot F_s} \cdot randn(1, L)$.

В результате получили зависимости вероятности ошибки аналитическим способом путём усреднения ансамбля реализаций аналогично п. 4 эксперимента № 1 по 1000 независимым испытаниям (различным реализациям вектора двоичных символов). Также полученные в результате проведения эксперимента с использованием модели значения усреднены по 30 независимым испытаниям. Данные представлены в виде графиков вероятности ошибки для различных скоростей передачи и различных индексов модуляции в зависимости от E_b/N_0 (рис. 13 – 16).

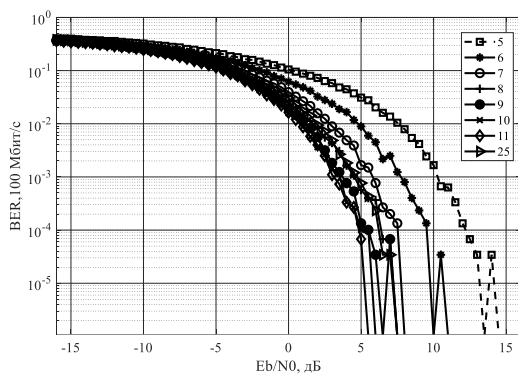


Рис. 13а. Скорость 100 Мбит/с (данные получены экспериментально с помощью модели на рис. 3, усреднение по 30 реализациям)

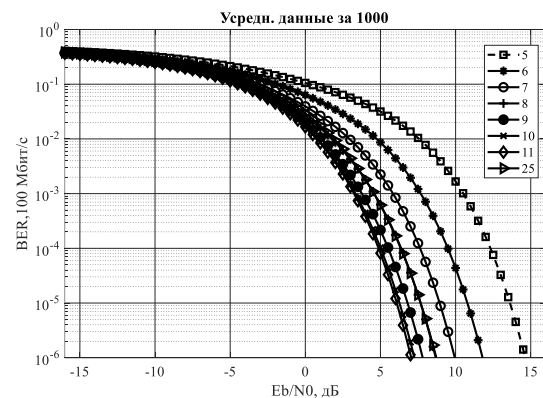


Рис. 13б. Скорость 100 Мбит/с (данные получены через аналитические зависимости (10)–(17), усреднение по 1000 реализациям)

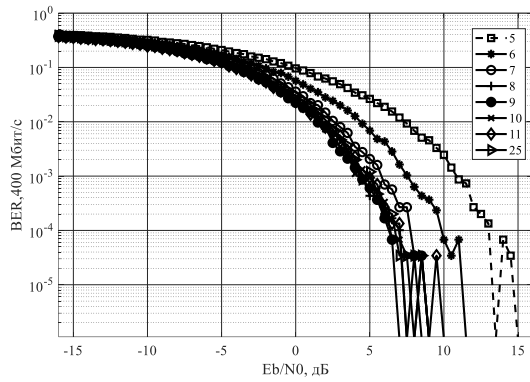


Рис. 14а. Скорость 400 Мбит/с (данные получены экспериментально с помощью модели на рис. 3, усреднение по 30 реализациям)

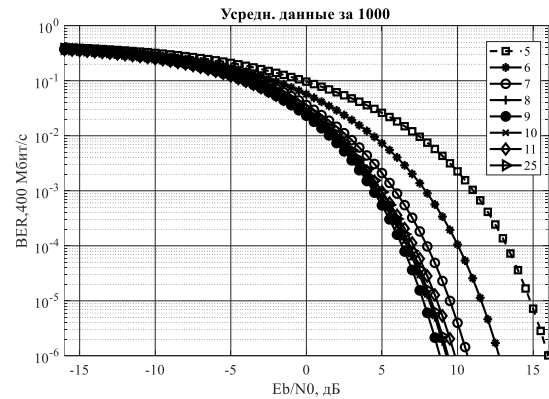


Рис. 14б. Скорость 400 Мбит/с (данные получены через аналитические зависимости (10)–(17), усреднение по 1000 реализациям)

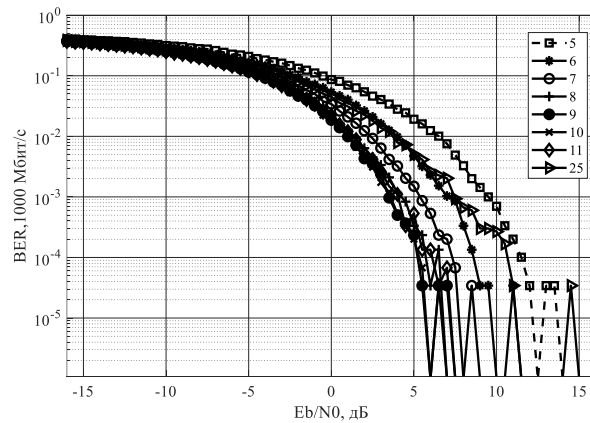


Рис. 15а. Скорость 1000 Мбит/с (данные получены экспериментально с помощью модели на рис. 3, усреднение по 30 реализациям)

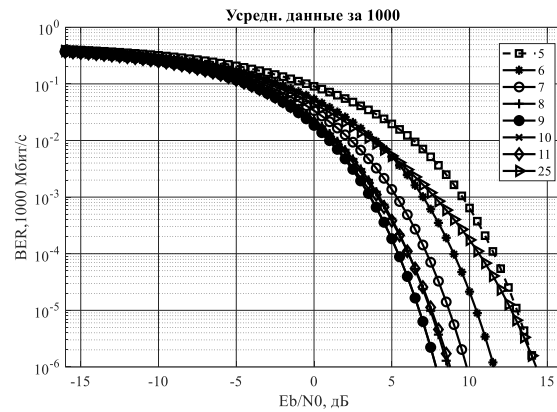


Рис. 15б. Скорость 1000 Мбит/с (данные получены через аналитические зависимости (10)–(17), усреднение по 1000 реализациям)

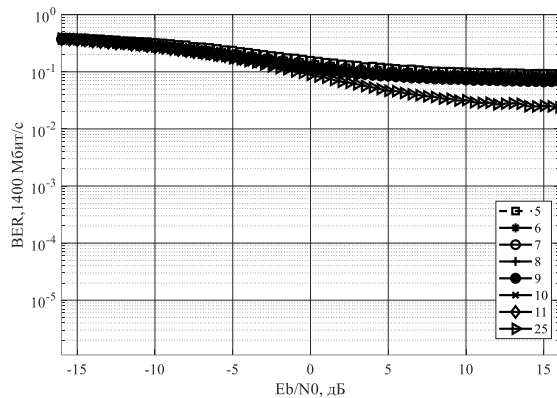


Рис. 16а. Скорость 1400 Мбит/с (данные получены экспериментально с помощью модели на рис. 3, усреднение по 30 реализациям)

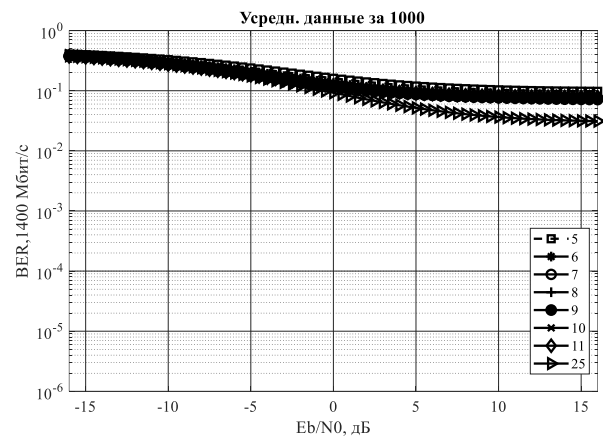


Рис. 16б. Скорость 1400 Мбит/с (данные получены через аналитические зависимости (10)–(17), усреднение по 1000 реализациям)

5. Анализ полученных результатов

Данные с усреднением по 30 реализациям при проведении эксперимента, представленные на графиках (рис. 9а – 16а), демонстрирует сходимость к кривым, полученным с помощью аналитического выражения (рис. 9б – 16б). Это свидетельствует о том, что расчётные значения

зависимости вероятности ошибки от ОСШ с использованием формул (10)–(17) верны. Следовательно, выражения можно использовать для оценки помехоустойчивости системы с дПИМ-модуляцией.

Также видно, что помехоустойчивость системы сильно зависит от индекса модуляции. Например, из графика на рис. 13б видно, что на уровне вероятности ошибки, равном 10^{-5} , изменение индекса модуляции может приводить к изменению отношения E_b/N_0 на 8 дБ. Это происходит потому, что в ИХ радиоканала имеются пути распространения, в которых разница между временем прихода луча в точку приёма (τ_l) и временем начала очередного символа ($m \cdot T_{\text{сymb}}$) по модулю совпадает с индексом модуляции (рис. 17), т.е.

$$|\tau_l - mT_{\text{сymb}}| = \delta. \quad (20)$$

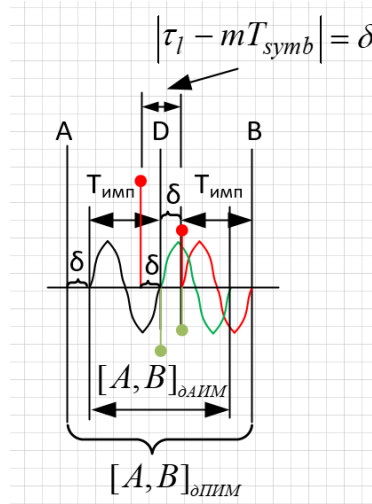


Рис. 17. К определению лучей модуляции дПИМ, дающих значительный вклад в МСИ

Ситуация осложняется, когда амплитуды интерферирующих с основным символом лучей сравнимы с амплитудой основного луча. Рассматривая интервалы $[A, B]$ ИХ радиоканала, можно определять помехоустойчивость системы. Как предлагается в работе [8], удобно использовать понятие совмещённого интервала $[A, B]_u$. Этот интервал получается путём объединения всех интервалов $[A, B]_k$ $k = 1..K$ импульсной характеристики (рис. 18).

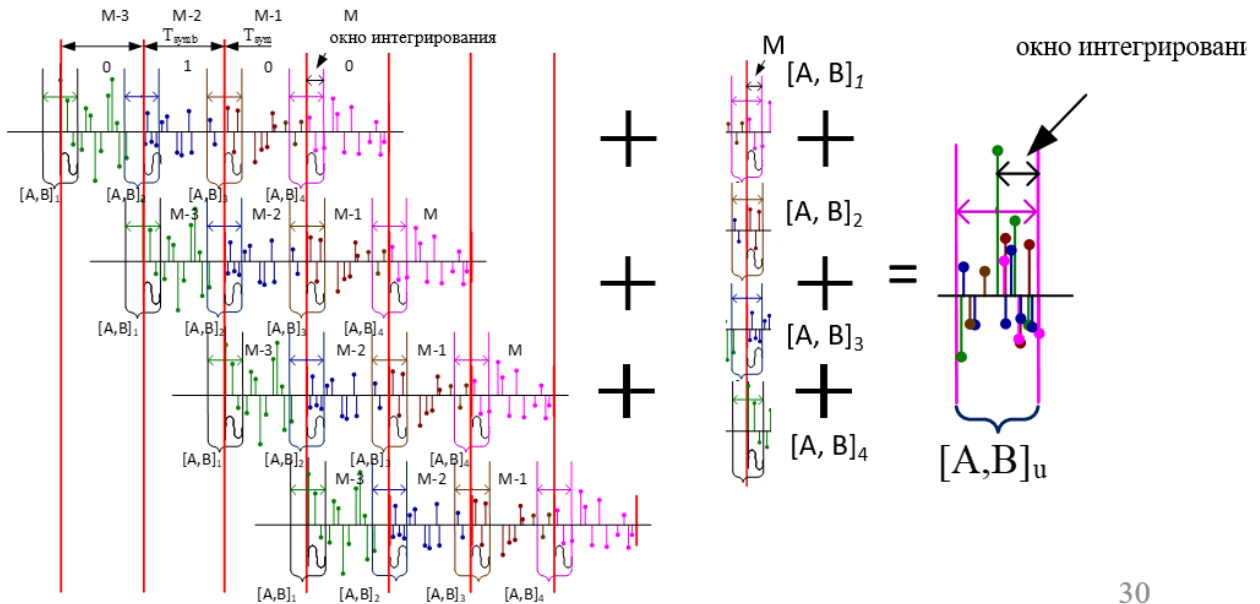


Рис. 18. К определению объединённого интервала $[A, B]_u$

Также следует обратить внимание, что совмещённый интервал $[A, B]_u$ и обычные интервалы $[A, B]$ являются зависимыми от скорости передачи данных. Это подтверждается тем фактом, что при одном и том же индексе модуляции для различных скоростей передачи данных может быть разная вероятность ошибки, как, например, зависимость вероятности ошибки на рис. 14б и 15б, соответствующая скоростям передачи 400 Мбит/с и 1000 Мбит/с для индексов модуляции 6×10^{-11} с и 25×10^{-11} с. Система с индексом модуляции 25×10^{-11} с обладает лучшей помехоустойчивостью на скорости передачи данных 400 Мбит/с по сравнению с системой с индексом модуляции 6×10^{-11} с, а на скорости 1000 Мбит/с ситуация меняется на противоположную. Интересно отметить факт, что индекс модуляции 25×10^{-11} с имеет полностью неперекрывающиеся во временной области модулирующие формы сигналов для единицы и нуля – для них $\rho = R(\delta) = 0$. Таким образом, оптимальные индексы модуляции для различных скоростей передачи данных представлены в табл. 1

Таблица 1. Значения оптимальных индексов модуляции в многолучевом радиоканале с модуляцией дПИМ

Скорость передачи, Мбит/с	Индекс модуляции δ , с
100	11×10^{-11}
400	9×10^{-11}
1000	9×10^{-11}
1400	25×10^{-11}

При выборе индекса модуляции для КСШП-системы радиосвязи с модуляцией дПИМ для обеспечения лучшей помехоустойчивости на заданной скорости необходимо стремиться к выполнению следующих условий:

– не допускать выполнения равенства (20);

– стремиться, чтобы выполнялось равенство условной вероятности ошибки при передаче единичного символа и условной вероятности ошибки при передаче нулевого символа [3]:

$$P_E(g/1) = P_E(g/0).$$

6. Заключение

Результатом данного исследования является установление аналитической зависимости вероятности битовой ошибки от ОСШ в многолучевом канале при работе внутри помещений для модуляции дПИМ при различных индексах модуляции, что позволяет осуществлять выбор наиболее подходящих параметров КСШП-системы радиосвязи (индекс модуляции, скорость передачи данных) для достижения минимума вероятности ошибки в условиях меняющейся ИХ канала радиосвязи или, для обратной задачи, вычислять вероятность ошибки при заданных параметрах КСШП-системы радиосвязи. Достоверность расчётов подтверждается совпадением графиков вероятности битовой ошибки от ОСШ для различных скоростей передачи данных, полученных с помощью компьютерного моделирования в системе MATLAB.

На основе полученных результатов можно заключить, что помехоустойчивость КСШП-систем радиосвязи с модуляцией дПИМ чувствительна к индексу модуляции, и оптимальный выбор его значения может улучшить производительность до 8 дБ по ОСШ.

Дано обоснование источникам ошибок, возникающих при передаче, с позиции рассмотрения интервалов $[A, B]$, характеризующих ИХ радиоканала.

Приведены рекомендации по выбору индекса модуляции в многолучевом канале.

Таким образом, в условиях меняющейся ИХ канала радиосвязи можно определять оптимальные параметры КСШП-системы радиосвязи через определённые интервалы времени, делая систему адаптивной к рабочим условиям.

Литература

1. Arslan H., Chen Z. N., Di Benedetto M.-G. Ultra wideband wireless communication. John Wiley and Sons, 2006. 500 p.
2. Lijia G., Guangrong Y., Soffiène A. On the BER performance of pulse-position-modulation UWB radio in multipath channels // Proc. IEEE Conference on Ultra Wideband Systems and Technologies. Baltimore, MD, USA, 21–23 May, 2002. P. 231–234.
3. Guangrong Y., Lijia G., Shaoqian L. Performance of UWB time-hopping spread-spectrum impulse radio in multipath environments // Proc. 57th IEEE Semiannual Vehicular Technology Conference, Jeju, South Korea, 22–25 April, 2003. V. 3. P. 1644–1648.
4. Карболин В. А., Носов В. И. Исследование влияния скорости передачи и частоты дискретизации импульсной характеристики на помехоустойчивость КСШП-системы радиосвязи // Вестник СибГУТИ. 2018. № 2. С. 71–83.
5. Foerster J. Channel Modeling Sub-Committee Report Final. IEEE P802.15-02/368r5-SG3a, 2002. 40 p.
6. Карболин В. А., Носов В. И. Исследование помехоустойчивости КСШП-системы радиосвязи с двоичной позиционной импульсной модуляцией при многолучевом распространении (INDOOR режим) // Материалы РНТК «Современные проблемы телекоммуникаций», СибГУТИ, 25–26 апреля 2019 г. С. 227–230.
7. Документация математического пакета MATLAB [электронный ресурс]: URL: <https://www.mathworks.com/help/comm/ref/awgnchannel.html> (дата обращения: 24.05.2021).
8. Karbolin V. A., Nosov V. I., and Kalinin V. O. Performance Analysis of UWB Communication Receiver in Multipath Environment Based on RAKE Receiver // Proc. 21st International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices (EDM), Novosibirsk, June 29–July 3, 2020. P. 98–103.

Статья поступила в редакцию 17.12.2021.

Карболин Виталий Анатольевич

аспирант СибГУТИ (630102, Новосибирск, ул. Кирова, 86), e-mail: rosinop@yandex.ru.

Носов Владимир Иванович

д.т.н., профессор кафедры цифрового телерадиовещания и систем радиосвязи СибГУТИ (630102, Новосибирск, ул. Кирова, 86), тел. (383) 269-82-54, e-mail: nvi@sibguti.ru.

References

1. Arslan H., Chen Z. N., Di Benedetto M.-G. Ultra wideband wireless communication. John Wiley and Sons, 2006. 500 p.
2. Lijia G., Guangrong Y., Soffiène A. On the BER performance of pulse-position-modulation UWB radio in multipath channels. *Proc. IEEE Conference on Ultra Wideband Systems and Technologies*, Baltimore, MD, USA, 21–23 May, 2002, pp. 231–234.
3. Guangrong Y., Lijia G., Shaoqian L. Performance of UWB time-hopping spread-spectrum impulse radio in multipath environments. *Proc. 57th IEEE Semiannual Vehicular Technology Conference*, Jeju, South Korea, 22–25 April, 2003, vol. 3, pp. 1644–1648.
4. Karbolin V. A., Nosov V. I. *Issledovanie vlijaniya skorosti peredachi i chastoty diskre-tizacii impul'snoj harakteristiki na pomehoustojchivost' KSSHП-sistemy radiosvyazi* [Investigation of the influence of transmission rate and sampling frequency of impulse response for noise immunity of a short-pulse ultra-wideband radio communication system]. *Vestnik SibGUTI*. 2018, no. 2, pp. 71–83.
5. Foerster J. Channel Modeling Sub-Committee Report Final. *IEEE P802.15-02/368r5-SG3a*, 2002. 40 p.

6. Karbolin V. A., Nosov V. I. *Issledovanie pomehoustojchivosti KSShP-sistemy radiosvjazi s dvoichnoj pozicionnoj impul'snoj moduljaciej pri mnogoluchevom rasprostranenii (INDOOR rezhim)* [Investigation of the noise immunity of a short-pulse ultra-wideband radio communication system with binary position pulse modulation in multipath propagation (INDOOR mode)]. *Materialy RNTK «Sovremennye problemy telekommunikacij»*, SibSUTIS, April 25-26, 2019, pp. 227-230.
7. *Dokumentacija matematicheskogo paketa MATLAB [jelektronnyj resurs]* [MATLAB software package documentation]. URL: <https://www.mathworks.com/help/comm/ref/awgnchannel.html> (access date: 24.05.2021).
8. Karbolin V. A., Nosov V. I., and Kalinin V. O. Performance Analysis of UWB Communication Receiver in Multipath Environment Based on RAKE Receiver. *Proc. 21st International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices (EDM)*, Novosibirsk, June 29-July 3, 2020, pp. 98-103.

Recommendations for the modulation index choice of UWB radio systems with BPPM in indoor channels

Vitaly A. Karbolin

Postgraduate student, Siberian State University of Telecommunications and Information Sciences (SibSUTIS, Novosibirsk, Russia), rosinop@yandex.ru.

Vladimir I. Nosov

Doctor of technical sciences, Professor, Siberian State University of Telecommunications and Information Sciences (SibSUTIS, Novosibirsk, Russia), nvi@sibguti.ru.

The impact of modulation index on the BER performance of BPPM in UWB radio is analyzed in multipath indoor channels. The recommendations for the optimal modulation index choice are drawn up with respect to minimum bit error rate in multipath indoor channel. The theoretical expression for BER performance of BPPM is derived. The BER performance figures for different data rates as a function of a signal to a noise ratio produced by using the theoretical expression and computer simulations are presented. MATLAB is used for computer simulations.

Keywords: short-pulse UWB radio system, BER (bit error rate) performance, modulation index, binary pulse position modulation, BPPM, UWB, Indoor UWB.

Многофакторный анализ влияния особенностей студентов на академическую успеваемость для персонализации образовательной среды

Т. В. Авдеенко, Н. В. Пустовалова¹

Современные университеты, разрабатывая цифровую среду, опираются на студенто-центрический подход, актуальные педагогические и дидактические практики, что позволяет учитывать индивидуальные особенности студентов и адаптировать к ним образовательный процесс. Персонализация образовательной среды, таким образом, призвана улучшить качество образовательного процесса, подняв его на новый уровень. В данной статье приведены результаты применения методов многофакторного моделирования для выявления общих связей и закономерностей между успеваемостью студентов и их психометрическими характеристиками. Исследуемая выборка включает результаты тестирования студентов Новосибирского государственного технического университета (НГТУ), определяющие когнитивные, мотивационные и личностные особенности обучающихся, а также информацию о всех прошедших аттестационных испытаниях, предусмотренных учебным планом. Отличительной чертой данной работы является разнообразие тестов, которым подвергались студенты, с целью найти сочетания составляющих их подтестов (факторов), наиболее сильно влияющих на успешность обучения, для последующего использования выявленных закономерностей при проектировании индивидуальных траекторий, например, выдачи образовательного контента с учетом особенности личности студента.

Ключевые слова: персональная образовательная среда, регрессионный анализ, факторный анализ, структурное моделирование, модель обучаемого, психометрические характеристики.

1. Введение

Сегодня в свете таких глобальных тенденций, как студенто-центрический подход, гуманизация, инклюзивное образование [1–3] перед вузами стоит задача не просто создать цифровую среду, а сделать ее комфортной и дружелюбной для пользователей. Современное видение процесса подготовки кадров предполагает, что университеты не просто дают определенный набор сведений, а позволяют каждому обучаемому сформировать систему знаний, умений и навыков (hard, soft и digital skills), обеспечивающих его дальнейшее личностное и профессиональное развитие [4]. Именно персонализация образовательной среды должна улучшить качество образовательного процесса, поднять его на новый уровень.

На сегодняшний день все университеты используют специализированные программы для автоматизации своей деятельности, в том числе для оценки учебных достижений студентов, контроля посещаемости, самодиагностики, сбора и анализа отзывов. На рынке существует много различных программных систем, которые можно приобрести «под ключ» (готовые решения от 1С), или же конструкторы (Битрикс), или пакеты прикладных программ с базовым

¹ Работа выполнена при финансовой поддержке Министерства науки и высшего образования в рамках госзадания (проект № FSUN-2020-0009).

функционалом и открытым кодом, которые можно доработать под свои нужды (MOODLE). Вместе с тем простой автоматизации уже недостаточно.

Современный образовательный процесс уделяет все больше внимания личности обучаемого, его индивидуальным особенностям и потребностям. Персонализированная образовательная среда (Personal learning environment, сокращенно PLE) – «совокупность объектов образовательного процесса, полученная из информационно-коммуникационной образовательной среды путем адаптации в соответствии с целями, содержанием и планируемыми результатами обучения, потребностями и способностями обучаемого и выступающая средством персонализации личности» [5]. Персонализация цифровой образовательной среды происходит параллельно с цифровизацией, внедрением дистанционного обучения и виртуальных технологий. PLE важна не только для обучаемых, но также для аналитиков и исследователей. Данные о персональных особенностях студентов прогрессивные вузы (например, Университет 2035, МИФИ и др.) используют в педагогическом дизайне образовательных программ и результатов. Так, при регистрации на платформе «Университет 2035» обучающимся предлагается пройти входную диагностику. Эта диагностика включает тесты на скорость реакции, умение адаптироваться, мотивацию, стиль взаимодействия с окружающими и т.д. Такие платформы, как «Coursera» используют данные о том, как слушатели проходят разные курсы, для того чтобы управлять контентом и определенным образом мотивировать обучающихся для завершения изучения курса.

В связи с этим перед разработчиками информационных систем для образовательных организаций встает вопрос о том, какие данные о студентах нужно собирать. Ведь неверно сформулированные требования к входной информации могут привести к тому, что система будет оперировать большими объемами данных, что ведет к увеличению затрат, и при этом не будет предоставлять адекватной информации для принятия управленческих решений. Поэтому очень важно тщательно спроектировать входные данные, т.е. отобрать те индивидуальные характеристики обучаемых, которые будут учитываться университетом для улучшения своей деятельности, а также методы их определения и оценки.

Статья организована следующим образом. Раздел 2 содержит постановку задачи, объясняя важность выбора показателей для формирования PLE. В разделе 3 описаны результаты регрессионного анализа данных, полученных в результате тестирования студентов. Раздел 4 включает результаты факторного анализа и структурного моделирования уравнений. Раздел 5 – заключение.

2. Проектирование модели обучаемого для PLE с использованием данных психометрических измерений студентов

Анализ источников [6–8] показал, что для оценки индивидуальных особенностей обучаемых широко используются психометрические характеристики студентов (скорость реакции, эмоциональный интеллект, IQ и т.д.). Этому способствует тот факт, что психология предоставляет целый набор стандартизированных тестов, проводимых по устоявшимся методикам, которые позволяют оценивать некоторые параметры для выборки обучаемых, анализировать и сравнивать их между собой (Амтхауэр, Равен, Кеттл, Айзенк и др). Разные авторы пытаются анализировать самые разные зависимости, для чего широко используется аппарат прикладной статистики.

Проанализировав работы современных исследователей, приходим к выводу, что для совершенствования учебного процесса необходимо правильно выявить и оценить множество психологических факторов, положительно или отрицательно влияющих на успешность образовательной деятельности студентов. Эти сведения необходимы для построения модели обучаемого, которая лежит в основе архитектуры интеллектуальных обучающих систем. Вместе с ней также используются модель тьютора, содержащая сведения о том, как обучать, и модель эксперта, включающая сведения о том, чему обучать [9]. Для создания модели обучаемого

необходимо определить те измеримые характеристики, которые можно собрать в рамках образовательного процесса и которые будут содержать значимую информацию для принятия управленческих решений в ходе его реализации.

При выполнении исследования мы сталкивались с большим количеством различных показателей, которые описывают некие ненаблюдаемые конструкции, включающие в себя наблюдаемые переменные. В качестве примера можно привести понятие «интеллект», его нельзя измерить напрямую, вместо этого исследователи разрабатывают гипотезу об интеллекте и создают инструменты измерения в виде вопросов или заданий, предназначенных для его оценки в соответствии с их гипотезой. В нашем случае имеется множество характеристик, измеренных тестовыми заданиями, объединение которых в более общие факторы может способствовать более качественной интерпретации их влияния на успеваемость студента. Для выявления скрытых факторов, отвечающих за объединение сильно коррелированных наблюдаемых переменных, было принято решение использовать методы факторного анализа.

В ходе исследования психологических особенностей личности определенную сложность представляет собой интерпретация результатов, зависящая от обоснованного структурирования признакового пространства. Последнее осложняется тем, что в области психологии до сих пор нет единого устоявшегося определения того, что такое интеллект и какова его структура. Однако на основании анализа источников и сравнения различных тестовых материалов было принято решение выделить в структуре интеллекта конструкты, отвечающие за когнитивные (память, скорость реакции и т.д.), личностные (когнитивный стиль, нейротизм и т.д.) и мотивационные (стремление к избеганию неудач и т.д.) характеристики. Представляется, что такое разделение позволит выбрать из показателей, полученных в результате выполнения различных тестов, некоторый набор, отражающий значимые составляющие интеллекта, отвечающие не только за «вычислительные» способности человека, но и за его умение работать в команде, ставить цели и добиваться их, что соответствует современной точке зрения на образовательный процесс и отражает необходимость развивать не только *hard*, но и *soft skills*. Эта парадигма заложена в основу модели обучаемого, разрабатываемой для проектирования и последующей реализации PLE (рис. 1).

В процессе проектирования целесообразно использовать возможности анализа и моделирование для повышения качества артефактов данного этапа. Поэтому было принято решение выполнить анализ исходных данных с применением статистических методов, в частности факторного анализа, а также построить структурную модель. Всё это должно позволить выявить те предикторы и факторы, которые оказывают заметное влияние на зависимую переменную — академическую успеваемость, интерпретировать их и на основе полученных результатов определить входные данные для PLE, в частности, реализовать модель обучаемого в виде онтологии, в которой результаты отдельных тестов будут экземплярами классов и будут связаны с экземплярами класса *Learner* (обучаемый) отношениями. В этом случае для проектирования отношений необходима информация о том, какие предикторы оказывают влияние на академическую успеваемость как по отдельности, так и вместе.

Модель образовательного контента, предложенная в [10], также реализована в виде онтологии *Protege* и используется для представления данных о видах, способах реализации, а также стратегиях выдачи образовательного контента. Она используется совместно с онтологией «Обучаемый» для установления соответствия между совокупностью персональных характеристик, присущих студенту, и наиболее подходящими ему формой и стратегией предоставления контента.

Для построения модели обучаемого (рис.1), интегрирующей набор характеристик субъектов обучения, влияющих на возможности восприятия ими (обучаемыми) информации и ее качественного усваивания, а также связей между этими характеристиками и целями обучения, мы предлагаем использовать методы многомерного анализа данных. Так, в качестве входных данных для оценки когнитивных характеристик предполагается использовать результаты тестов Амтхауэра, Равена. Для оценки личностных характеристик — результаты тестов «Большой

шая пятерка», Бачард. Мотивационные характеристики предполагается оценивать посредством краткого опросника Грэя. Данные об академической успеваемости уже собираются в LMS университета на протяжении нескольких лет, методика их сбора и расчета показателей успеваемости устоялась. В качестве зависимой переменной в исследовании выступает академическая успеваемость, которая рассчитывается как балл по шкале ECTS без учета пересдач, полученный за сдачу экзаменов, зачетов и курсовых.



Рис. 1. Использование модели обучаемого при проектировании PLE

Структура классов модели обучаемого «Learner model», в которой отражены составляющие личности студента, оказывающие влияние на академическую успеваемость, приведена на рис. 2. Однако в большинстве работ, посвященных исследованию влияния персональных свойств на успеваемость, используется или какой-то один тест, или несколько разных текстов, оценивающих один и тот же параметр. В настоящей работе мы исследуем, как влияют на успеваемость различные комбинации свойств личности студентов, на основе анализа данных множества тестов, оценивающих когнитивные, мотивационные и личностные свойства, с целью последующего выявления возможностей влияния на эти показатели посредством создания более персонифицированных форм представления образовательного контента.

Исследование организовано следующим образом. Сначала мы проводим анализ данных, в основе которого лежит построение многомерных линейных моделей. Проводится исследование влияния объясняющих переменных на общую успеваемость, а также успеваемость по группам предметов (гуманитарные, математические, информационные), в результате чего определяются значимые связи, которые будут включены в модель обучаемого. Далее, на основе предварительного проведения факторного анализа предикторов, мы предлагаем и исследуем приемлемые с точки зрения критериев качества структурные модели, интегрирующие исследуемые показатели в их сложной взаимосвязи, с целью последующего включения в модель обучаемого как самих показателей, так и их структурной организации.

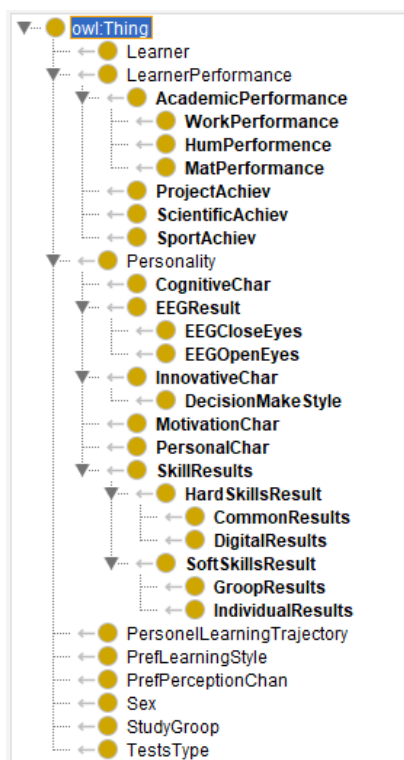


Рис. 2. Онтология «Learner model», реализующая модель обучаемого на основе результатов психометрического тестирования студентов

3. Выделение значимых характеристик обучаемых на основе регрессионного анализа

3.1. Описание и разведочный анализ входных данных

Для исследования на первом этапе были использованы результаты, полученные по следующим методикам оценки (тестовые материалы): Амтхауэра, Равена, Бачард, краткого опросника Грэя и «Большой пятерки». Результат каждого теста представляет собой один или несколько численных показателей (табл. 1). Также в исследуемом наборе данных присутствуют две дискретные переменные: «gender» – пол испытуемого и «group» – группа, в которой обучается студент. При обработке данных каждому студенту предварительно присвоен идентификационный номер (ID), а ФИО удалены.

Выборка включает результаты тестирования 72 человек – студентов третьего курса факультета прикладной математики и информатики НГТУ. Данные содержат 26 переменных, включая три группы характеристик, данные об академической успеваемости. Объекты с пропусками в данных были удалены из анализа.

Таблица 1. Переменные показателей, полученных в результате тестирования обучаемых

Наименование теста	Имя переменной (сокращение)	Описание переменной
Тест Амтхауэра	iq1	логическое мышление и общая осведомленность
	iq2	способность к абстрагированию
	iq3	комбинаторные способности (слова)
	iq4	лексический запас
	iq5	математическое мышление
	iq6	способность оперировать числами и индуктивное мышление
	iq7	комбинаторные способности (фигуры)

	iq8	пространственное воображение
	iq9	способность к кратковременному хранению информации
Тест Равена	RIQ	флюидный интеллект
Тест Бачард	eq1	выражение положительных эмоций
	eq2	выражение отрицательных эмоций
	eq3	внимание к эмоциям
	eq4	роль эмоций в принятии решений
	eq5	отзывчивость на радость
	eq6	отзывчивость к несчастьям
	eq7	эмоциональное участие
Краткий опросник Грэя	BAS	система активации поведения (стимулы награды)
	BIS	система торможения поведения (стимулы наказания)
«Большая пятерка»	Neuroticism (nrt)	нейротизм
	Extraversion (ext)	экстраверсия
	openess_experience (op_)	открытость опыту
	Agreeableness (agr)	доброжелательность
	Conscientiousness (cons)	добросовестность

Для того чтобы более детально оценить влияние нижеперечисленных объясняющих переменных на успеваемость (AVG – средний балл по всем пройденным предметам), учебные предметы дополнительно были разделены на три группы: средний балл по математическим предметам MAT (напр., «Математический анализ»), предметам, связанным с профессиональной деятельностью PRG (напр., «Основы программирования»), гуманитарным предметам GUM (напр., «Иностранный язык»).

На первом этапе разведочного анализа проведено определение и удаление выбросов, содержащихся в исходных данных, на основе графика boxplot (точки-отскоки на рис. 3).

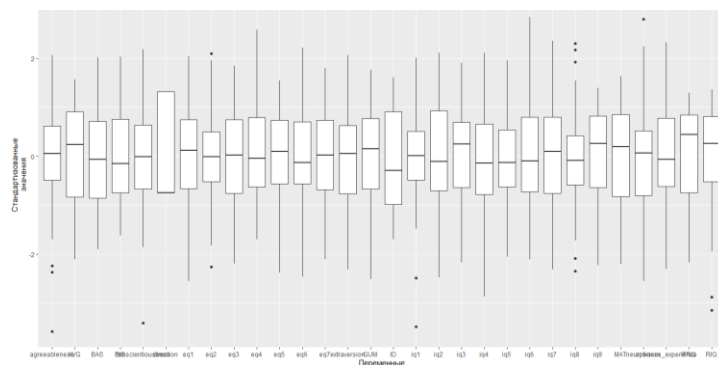


Рис. 3. График boxplot для стандартизованных значений объясняющих переменных

Анализ мультиколлинеарности проведен на основе фактора инфляции дисперсии VIF (табл. 2), в результате которого были исключены избыточные предикторы eq1 и eq6.

Таблица 2. Фактор инфляции дисперсии для исходных данных

Переменная	iq1	iq2	iq3	iq4	iq5	iq6	iq7	iq8	iq9	eq1	eq2	eq3	eq4
VIF	1.39	1.95	2.02	1.67	1.72	1.86	1.88	1.78	1.44	3.04	2.59	1.64	2.03

Переменная	eq5	eq6	eq7	RIQ	BAS	BIS	nrt	ext	op ₋	agr	cons
VIF	2.35	4.42	2.48	1.82	1.60	3.20	2.84	3.02	1.86	2.78	2.69

3.2. Многомерный регрессионный анализ

Регрессионный анализ позволяет построить линейную модель вида

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon, \quad (1)$$

где y – зависимая переменная модели, определяющая качество обучения, x_i – объясняющие переменные (предикторы), представляющие собой характеристики обучаемых, β_i – неизвестные параметры модели, подлежащие оцениванию на основе собранных данных, ε – случайная неизвестная ошибка, удовлетворяющая необходимым требованиям.

Построение регрессионной модели (1) осуществляется при помощи функции `lm` пакета для анализа данных R. Результаты применения функции `lm` ко множеству всех объясняющих переменных с показателем среднего балла *AVG* в качестве зависимой переменной представлено в табл. 3. Как видно, ни одна из переменных не выявляет значимости, что является показателем избыточности предикторов.

Таблица 3. Результат применения функции `lm` для зависимой переменной *AVG*

Предикторы	Коэффициенты	Стандартные ошибки	Значение t-критерия	p-значение
(Intercept)	-78.2106	57.0132	-1.37	0.176
iq1	0.3897	0.2580	1.51	0.137
iq2	-0.1035	0.2730	-0.38	0.706
iq3	0.2720	0.2360	1.15	0.255
iq4	0.2372	0.2720	0.87	0.387
iq5	-0.0592	0.1443	-0.41	0.684
iq6	0.1509	0.2017	0.75	0.458
iq7	0.0141	0.2535	0.06	0.956
iq8	0.1270	0.2183	0.58	0.563
iq9	0.0866	0.1727	0.50	0.618
eq2	0.1189	0.2752	0.43	0.668
eq3	0.1356	0.2167	0.63	0.534
eq4	0.0866	0.2524	0.34	0.733
eq5	-0.2319	0.2124	-1.09	0.280
eq7	-0.0216	0.3510	-0.06	0.951
BAS	0.0518	0.4350	0.12	0.906
BIS	0.3460	0.3718	0.93	0.357
RIQ	0.1074	0.1630	0.66	0.513
neuroticism	-0.0811	0.2421	-0.34	0.739
extraversion	0.0754	0.1897	0.40	0.693
openess experience	-0.1194	0.2358	-0.51	0.615
agreeableness	0.1254	0.2896	0.43	0.667
conscientiousness	0.5504	0.2429	2.27	0.028

Для получения модели незначимые предикторы были последовательно исключены из модели (табл. 4). В результате получаем статистически значимую модель ($F_{4,67} = 7.17$, $p = 0.0000719$) зависимости среднего балла *AVG* от iq1, iq3 теста Амтхауэра, системы торможения поведения и добросовестности. Флюидный и эмоциональный интеллекты не оказывают существенного влияния на средний балл. Функциональный вид модели представлен формулой (2), коэффициент детерминации $R^2 = 0.26$.

Таблица 4. Результат применения функции lm для значимых предикторов

Предикторы	Коэффициенты	Стандартные ошибки	Значение t-критерия	p-значение
(Intercept)	-31.494	26.113	-1.21	0.23204
iq1	0.380	0.208	1.83	0.07174
iq3	0.401	0.157	2.56	0.01260
BIS	0.542	0.222	2.44	0.01726
conscientiousness	0.562	0.146	3.86	0.00026

$$AVG = -31.494 + 0.38 * iq1 + 0.401 * iq3 + 0.542 * BIS + 0.562 * conscientiousness . \quad (2)$$

Аналогичным образом была построена регрессионная модель для среднего балла по математическим предметам. В результате получаем модель (3), отражающую статистически значимую зависимость МАТ от iq3 теста Амтхауэра и добросовестности ($F_{2,69} = 8.55$, $p = 0.000482$). Модель описывает 18 % общей изменчивости успеваемости по математическим предметам. В отличие от общего среднего балла AVG, свою значимость потерял показатель BIS (стимулы наказания) краткого опросника Грэя.

$$MAT = -4.67 + 0.604 * iq3 + 0.443 * conscientiousness . \quad (3)$$

Регрессионная модель для среднего балла PRG по предметам, связанным с профессиональной деятельностью, описана формулой (4). Значимость предикторов iq1 и iq9 невысока, но при их исключении процент общей изменчивости сильно падает, поэтому при условии, что по F-критерию модель является значимой ($F_{3,68} = 4.85$, $p = 0.0409$), она была принята в качестве приемлемой. Процент описываемой изменчивости – 14 %.

$$PRG = -34.864 + 0.536 * iq1 + 0.341 * iq9 + 0.585 * conscientiousness . \quad (4)$$

И, наконец, регрессионная модель (5) отражает зависимость среднего балла по гуманитарным предметам от исследуемых предикторов ($F_{3,68} = 9.33$, $p = 0.0000301$). В отличие от предыдущих результатов, помимо iq3 из теста Амтхауэра и добросовестности, заметно обозначено влияние флюидного интеллекта. Процент объясненной изменчивости – 26 %.

$$GUM = -23.14 + 0.437 * iq3 + 0.368 * RIQ + 0.395 * conscientiousness . \quad (5)$$

Помимо множества количественных объясняющих переменных, рассмотренных ранее, в исследуемой выборке изначально присутствовали две категориальные переменные: пол и направление подготовки, однако они были исключены в силу незначимого влияния на показатели академической успеваемости. При добавлении категориальных предикторов к модели значимого взаимодействия ни с одной из зависимых переменных, определяющих качество обучения, не выявлено.

При интерпретации полученных результатов в первую очередь обращает на себя внимание тот факт, что «добросовестность» (*conscientiousness*) и «комбинаторные способности (слова)» (iq3) оказались значимыми предикторами в трех случаях из четырех. На практике можно наблюдать, что выдающиеся когнитивные способности не приведут к значимым академическим результатам, если обучаемый несвоевременно или не в полном объеме выполняет предусмотренные задания, так как получает меньшие баллы или штрафы, не может вовремя сдать сессию и т.д. При этом не секрет, что обучаемые, не очень хорошо знающие предмет, но обладающие хорошими речевыми навыками, могут получить более высокий балл, создав иллюзию понимания, расположив к себе экзаменатора или имея более выгодный вид на фоне знающего, но «неговорящего» одноклассника. Так что полученный результат не противоречит наблюдениям за обучаемыми. Далее, предиктор «логическое мышление и общая осведомленность» (iq1) оказался значимым как для модели с общей успеваемостью, так и для модели с успеваемостью по профессионально-ориентированным дисциплинам. Стоит отметить, что в современной науке и образовании междисциплинарный подход имеет большое значение. Поэтому успешность (высокая успеваемость) может быть результатом того, что обучаемый обладает кругозором и умеет сопоставлять факты, относящиеся к разным предметным областям,

для решения задач. Примечательно, что для гуманитарных дисциплин существенным оказалось влияние флюидного интеллекта RIQ, который свидетельствует об умении решать новые задачи без опоры на старые знания [14] и считается очень важным для успешного обучения. Также для общей успеваемости значимым оказался предиктор «система торможения поведения (стимулы наказания)» (BIS). При отсутствии значимых предикторов положительной мотивации это может говорить о более существенном вкладе страха наказания или провала, а значит, о меньшей действенности стимулов, связанных с поощрением. Для многих обучаемых очень трудно подобрать действенные положительные мотивационные стимулы (стипендия, похвала, чувство удовлетворенности проделанной работой и т.д.), в то время как отрицательные стимулы действенны и очевидны (звонок родителям, отчисление и т.д.).

4. Многофакторное моделирование

4.1. Факторный анализ

В первую очередь необходимо оценить возможность использования факторного анализа, конкретно, метода главных компонент (principal component analysis, PCA). Для этого определим критерий сферичности Бартлетта и критерий Кайзера–Мейера–Олкина. В результате выполнения теста Бартлетта получено значение 905, $p = 5.64 \times 10^{-68}$, следовательно, PCA имеет перспективы хорошего сжатия. Тест Кайзера–Мейера–Олкина показал значение 0.63, что оценивает возможность выполнения факторного анализа как посредственную. Однако, учитывая результаты теста Бартлетта, проведем анализ.

На основе графика «Каменистая осыпь» (рис. 4) было принято решение о числе главных компонент в модели. Точка на графике каменистой осыпи (рис. 3), где убывание собственных значений максимально замедляется, соответствует числу факторов, равному 4, однако суммарная доля объясненной дисперсии в этом случае равна 47.9 %.

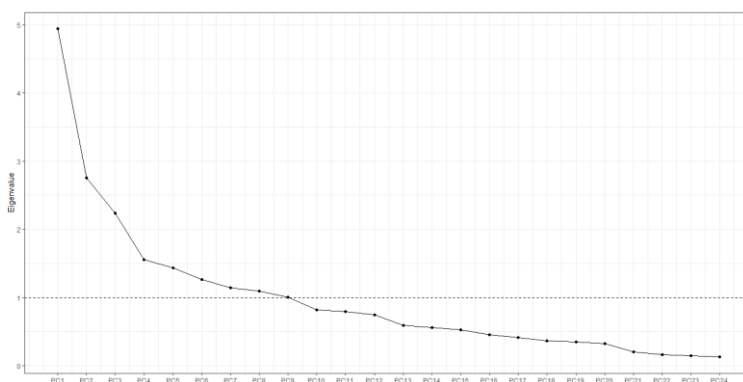


Рис. 4. График «Каменистая осыпь»

После вращения факторов методом ргмах получаем таблицу нагрузок (рис. 5a). Доля общей дисперсии, объясняемой каждым фактором, и накопление этих дисперсий представлены на рис. 6a. Для гипотезы о достаточности заданного числа факторов получаем $p\text{-value} = 0.0693$, что означает, что гипотеза не отвергается, но, возможно, число факторов стоит увеличить для достижения большего уровня значимости и приближения доли общей объясненной дисперсии к 50 %. Поэтому был рассмотрен альтернативный вариант увеличения числа факторов до 6. Результат после вращения методом ргмах отражен на рис. 5b.

Для проверки гипотезы о достаточности заданного числа факторов $p\text{-value} = 0.553$, что с большей уверенностью позволяет сказать, что шести факторов достаточно. Доля общей дисперсии, объясняемой каждым фактором, и накопление этих дисперсий представлены на рис. 6b.

Loadings:					Loadings:						
	Factor1	Factor2	Factor3	Factor4		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
eq6	0.905	0.658	-0.007	-0.082	eq6	0.799	-0.052	-0.213	-0.030	0.386	0.088
BIS	0.780	0.152	-0.003	0.015	eq7	0.820	0.171	-0.169	-0.021	-0.089	0.104
neuroticism	0.582	-0.040	0.079	-0.213	iq2	-0.237	0.564	-0.036	0.015	-0.137	0.373
eq7	0.421	0.668	0.175	-0.027	iq3	0.235	0.780	0.024	-0.045	0.050	0.093
BAS	0.079	-0.529	0.030	0.098	eq1	0.097	0.075	0.618	-0.131	0.000	0.272
openess_experience	0.016	0.544	0.217	0.106	extraversion	-0.210	-0.105	0.998	0.015	0.072	-0.137
agreeableness	0.233	0.526	0.048	0.313	conscientiousness	0.019	0.038	0.032	-1.024	0.097	0.006
iq2	-0.089	-0.433	0.621	0.308	BIS	0.090	0.007	0.116	-0.091	1.002	0.010
iq3	0.130	0.213	0.816	0.143	eq2	0.099	0.100	-0.018	0.000	0.023	0.998
iq7	-0.113	0.076	0.502	0.118	eq3	0.080	0.110	0.322	-0.093	0.142	0.128
eq1	-0.054	0.079	0.168	0.951	eq4	0.274	-0.332	0.149	0.299	0.021	0.068
eq2	0.413	-0.200	0.116	0.570	eq5	0.422	-0.269	0.250	0.008	-0.090	-0.045
extraversion	-0.228	0.103	-0.177	0.593	BAS	-0.355	0.020	0.120	0.417	0.032	0.121
eq3	0.062	0.125	0.081	0.350	iq1	0.083	0.260	-0.176	0.083	-0.033	-0.115
eq4	0.327	0.134	-0.309	0.151	iq4	-0.007	0.473	-0.035	-0.117	0.212	-0.082
eq5	0.123	0.422	-0.221	0.293	iq5	-0.236	0.466	0.051	0.105	-0.038	-0.096
iq1	0.009	0.123	0.239	-0.302	iq6	-0.255	0.400	0.100	-0.008	0.091	-0.226
iq4	0.069	0.048	0.478	-0.053	iq7	0.185	0.459	0.010	0.129	-0.265	-0.036
iq5	-0.188	-0.175	0.434	-0.077	iq8	0.159	0.315	-0.071	0.047	-0.052	-0.009
iq6	-0.180	-0.158	0.371	-0.063	iq9	-0.005	0.085	0.208	0.154	0.194	0.038
iq8	0.051	0.105	0.304	-0.059	RIQ	0.095	0.224	-0.073	-0.010	0.274	-0.071
iq9	0.270	0.024	0.033	0.040	neuroticism	0.199	0.055	-0.194	0.417	0.302	0.007
RIQ	0.234	0.141	0.257	-0.082	openess_experience	0.495	0.203	0.195	0.029	-0.137	-0.250
conscientiousness	-0.346	0.255	0.026	0.289	agreeableness	0.456	0.085	0.451	0.060	0.044	-0.008

а) Таблица нагрузок для 4 факторов

б) Таблица нагрузок для 6 факторов

Рис. 5. Таблица нагрузок для 4 и 6 факторов

Для визуального сравнения наполнение факторов, полученных в результате построения обеих моделей, приведено в табл. 5. Стоит отметить, что при отнесении переменных к факторам учитывались значения нагрузок более 0.4 (без учета знака).

	Factor1	Factor2	Factor3	Factor4		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
Proportion Var	0.113	0.101	0.099	0.096	Proportion Var	0.107	0.095	0.085	0.067	0.066	0.06
Cumulative Var	0.113	0.214	0.314	0.410	Cumulative Var	0.107	0.202	0.287	0.354	0.419	0.48

а) для 4 факторов

б) для 6 факторов

Рис. 6. Доля общей дисперсии и ее накопление для 4 и 6 факторов

Таблица 5. Сравнение наполнения факторов двух моделей

Фактор	Переменная	Расшифровка	Фактор	Переменная	Расшифровка
F1	eq6	отзывчивость к несчастьям	F1	eq6	отзывчивость к несчастьям
	BIS	стимулы наказания		eq7	эмоциональное участие
	nrt	нейротизм		eq5	отзывчивость на радость
	op_	открытость опыту		agr	доброжелательность
	eq7	эмоциональное участие		iq3	комбинаторные способности (слова)
F2	eq7	эмоциональное участие	F2	iq4	лексический запас
	BAS	стимулы награды		iq5	математическое мышление
	op_	открытость опыту		iq6	способность оперировать числами и индуктивное мышление
	agr	доброжелательность		iq7	комбинаторные способности (фигуры)
	eq5	отзывчивость на радость		iq2	способность к абстрагированию
F3	iq2	способность к абстрагированию	F3	eq1	логическое мышление и общая осведомленность
	iq3	комбинаторные способности (слова)		ext	экстраверсия
	iq7	комбинаторные способности (фигуры)		agr	доброжелательность

	iq4	лексический запас			
	iq5	математическое мышление			
F4	eq1	выражение положительных эмоций	F4	BAS	стимулы награды
	ext	экстраверсия		nrt	нейротизм
	eq2	выражение отрицательных эмоций			
-			F5	BIS	стимулы наказания
-			F6	eq2	выражение отрицательных эмоций

Полученное распределение исходных переменных по группам, определяемым факторами, достаточно разнообразно. Факторы объединяют переменные, полученные для разных групп тестов. Однако попытаемся дать интерпретацию полученных разбиений.

В четырехфакторной модели возможна следующая интерпретация факторов.

Фактор 1. Включает переменные из разных тестов: eq6, BIS, eq2, eq7 и nrt. Все вместе можно интерпретировать как проявления тревожной натуры, чутко реагирующей на неудачи, возможность провала, наказания и т.д.

Фактор 2. Включает переменные, связанные с выражением положительных эмоций от взаимодействия с окружающим миром (eq5, eq6, eq7, op_, agr) и от получаемых стимулов награды (BAS).

Фактор 3. Объединяет наиболее важные переменные из группы когнитивных способностей по Амтхауэру, ответственные за аналитические способности.

Фактор 4. Объединяет переменные, направленные на выражение эмоций окружающим – как положительных, так и отрицательных.

Для шестифакторной модели можно интерпретировать факторы следующим образом.

Фактор 1. Переменные, объединенные этим фактором, связаны с открытостью, доброжелательностью и отзывчивостью. Общим мотивом для этих переменных может быть сотрудничество и эмоциональный отклик от такого рода взаимодействия.

Фактор 2. Объединяет переменные, связанные с различными компонентами теста Амтхауэра. В модели с четырьмя факторами очень схожим по составу получился фактор 3.

Фактор 3. Самая большая нагрузка в составе этого фактора приходится на такие переменные, как «экстраверсия», «логическое мышление и общая осведомленность», а также «доброжелательность». В отличие от фактора 1, такой состав фактора выглядит более ориентированным на взаимодействие с профессиональной точки зрения, так как эмоциональная составляющая в нем сочетается с логической.

Фактор 4. Включает реакцию на стимулы награды – BAS и нейротизм – nrt. Нейротизм как параметр указывает на высокий уровень восприимчивости к стимулам вообще, что говорит об усиленном восприятии стимулов награды.

Фактор 5. В данном факторе самая существенная нагрузка приходится на «реакцию на стимулы наказания» – BIS. Нагрузка на остальные переменные по сравнению с этой очень мала. Таким образом, очевидно сильное влияние стимулов наказания. Выше уже были упомянуты результаты, указывающие на корреляцию между BIS и успеваемостью.

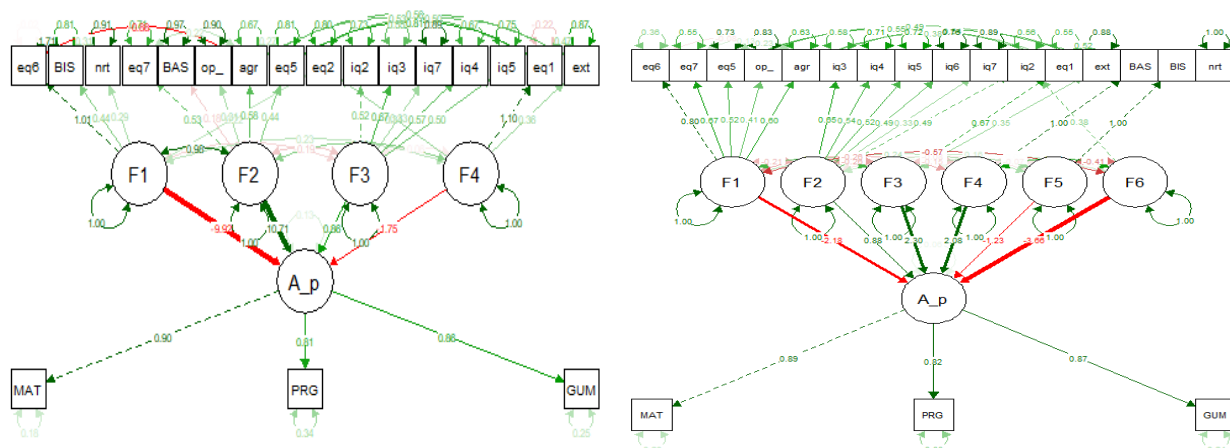
Фактор 6. Основная нагрузка падает на одну переменную – eq2 (выражение отрицательных эмоций), хотя в этом факторе нагрузку, близкую к 0.4, имеют переменные iq2 (способность к абстрагированию) и iq6 (способность оперировать числами и индуктивное мышление).

Говоря о сравнении четырехфакторной и шестифакторной моделей, стоит отметить, что BIS и BAS в обеих моделях относятся к разным факторам. Это говорит о том, что стимулы награды и наказания представляют собой отдельные переменные, не зависящие друг от друга, а также о важности мотивационной составляющей в принципе. Кроме того, в обеих моделях был выделен фактор, в составе которого объединены переменные, отвечающие за аналитические способности в структуре интеллекта по Амтхауэру. Это говорит о важности и независимости группы показателей, отвечающих за когнитивные способности.

4.2. Построение структурной модели

При построении структурной модели зависимости переменных, отвечающих за академическую успеваемость, от когнитивных, личностных и мотивационных характеристик обучаемых воспользуемся результатами факторного анализа. Факторный анализ позволяет построить модель с учетом знания о структуре разбиения предикторов по группам. В структурном моделировании используется несколько статистических тестов, позволяющих определить, соответствует ли модель данным и насколько хорошо. Как правило, мы не можем сделать вывод о том, «правильная» модель или нет, оценив качество соответствия между моделью и данными. Также мы не можем оценить, какую часть изменчивости она объясняет. Правильно подобранная модель означает только то, что модель правдоподобна.

Полученные модели для 4- и 6-факторного представления структуры объясняющих переменных представлены на рис. 7.



а) 4-факторная структурная модель

б) 6-факторная структурная модель

Рис. 7. Структурные модели, отражающие зависимость успеваемости от 4 (а) и 6 (б) факторов

Для оценки состоятельности структурных моделей обычно используются следующие показатели. CFI (сравнительный индекс соответствия) – при изучении исходных сравнений CFI в значительной степени зависит от среднего размера корреляций в данных. Если средняя корреляция между переменными невысока, CFI не будет очень высоким. Желательно, чтобы значение CFI было 0.95 и выше. RMSEA (среднеквадратичная ошибка аппроксимации) – индекс соответствия, где нулевое значение указывает на лучшее соответствие. Принято считать, что RMSEA 0.1 и выше указывает на плохое соответствие модели. SRMR (стандартизованный среднеквадратичный остаток) – популярный индикатор абсолютного соответствия. Значение менее 0.1 говорит о хорошем соответствии модели.

Значения указанных показателей для построенных моделей представлены в табл. 6.

Таблица 6. Сравнение характеристик двух моделей

Индекс	Значение для четырехфакторной модели	Значение для шестифакторной модели
CFI	0.858	0.910
RMSEA	0.082	0.066
SRMR	0.130	0.107

По данным табл. 6 видно, что шестифакторная модель обладает несколько лучшими характеристиками, нежели четырехфакторная. Тем не менее, обе модели можно считать приемлемыми с учетом размера выборки.

В обеих моделях с общей успеваемостью (рис. 9 и 10) присутствует фактор, где наибольшая нагрузка приходится на группу переменных, характеризующих интеллект по методологии Амтхауэра (логическое и математическое мышление, способность оперировать в рамках категорий и т.д.). Отметим тот факт, что и для четырехфакторной, и для шестифакторной модели

эти факторы (F3 и F2 соответственно) напрямую влияют на успеваемость, но при этом есть факторы, вклад которых больше.

Если обратить внимание на F2 и F1 в четырехфакторной и шестифакторной моделях соответственно, можно заметить, что в обоих случаях факторы, включающие переменные стимулов наказания (BIS), имеют знак минус. То есть чем меньше влияние этого фактора, тем лучше показатель успеваемости. Важным кажется тот факт, что в обеих моделях факторы, связанные с эмоциональной составляющей и мотивацией (F1 и F6 соответственно), оказывают существенное влияние на общую успеваемость и при этом входят в модель со знаком минус. Это можно интерпретировать с двух точек зрения: чем меньше факторов, тревожащих обучаемых, тем лучше они учатся; успешные обучаемые умеют интегрироваться в образовательный процесс, понимают, как он устроен, поэтому испытывают меньше тревог и учатся лучше.

Для уточненного анализа построим структурные модели на основе подтверждающего 6-факторного анализа, в которых зависимыми переменными являются средние баллы по различным группам дисциплин (математические, профессионально-ориентированные, гуманитарные), на которые влияют, предположительно, несовпадающие группы объясняющих переменных.

Структурные модели влияния на различные виды успеваемости представлены на рис. 8. Параметры для оценки моделей, приведенных на рис. 8, представлены в табл. 7. Качество моделей приемлемое.

Таблица 7. Значения критериев качества для моделей влияния на успеваемость по группам дисциплин

	Значение, математические предметы	Значение, профессиональные предметы	Значение, гуманитарные предметы
Кол-во параметров	62	62	62
Степени свободы	91	91	91
X2	126.180	129.476	126.878
CFI	0.880	0.872	0.881
RMSEA	0.073	0.077	0.074
SRMR	0.114	0.111	0.113

При интерпретации шестифакторных моделей в разрезе математических, профессионально-ориентированных и гуманитарных предметов (рис. 8) можно отметить следующее.

На успеваемость по математическим предметам (рис. 8а) наибольшее влияние оказывает фактор, включающий компоненты теста Амтхауэра, соответственно, большее влияние имеют когнитивные способности обучаемого. Аналогичная ситуация наблюдается для модели с успеваемостью по профессионально-ориентированным дисциплинам (рис. 8б), но в этом случае можно видеть, что F1 и F6 не имеют влияния на успеваемость. В этом случае можно предположить, что успешность в дисциплинах, связанных с математикой и профессиональной деятельностью, зависит исключительно от когнитивных способностей обучаемого и требует особого склада ума.

В разрезе гуманитарных предметов (рис. 8с) успеваемость больше всего определяется фактором F6, который также связан с переменной интеллекта. При этом достаточно большое влияние оказывают факторы F1 и F5, где присутствуют предикторы, определяющие различные показатели эмоционального интеллекта (то есть мотивационные и личностные характеристики). F4 присутствует в модели с обратным знаком, что говорит об обратном влиянии этого фактора, где основной вклад дает предиктор BAS – стимулы награды. Можно предположить, что обучение по данным дисциплинам (куда в учебном плане входят родной и иностранный языки, история и т.д.), не вызывает у студентов затруднений. Она воспринимается как рутинная, каждодневная действительность, и награда, похвала за ее успешное выполнение не выглядит особенно ценной в глазах обучаемых.

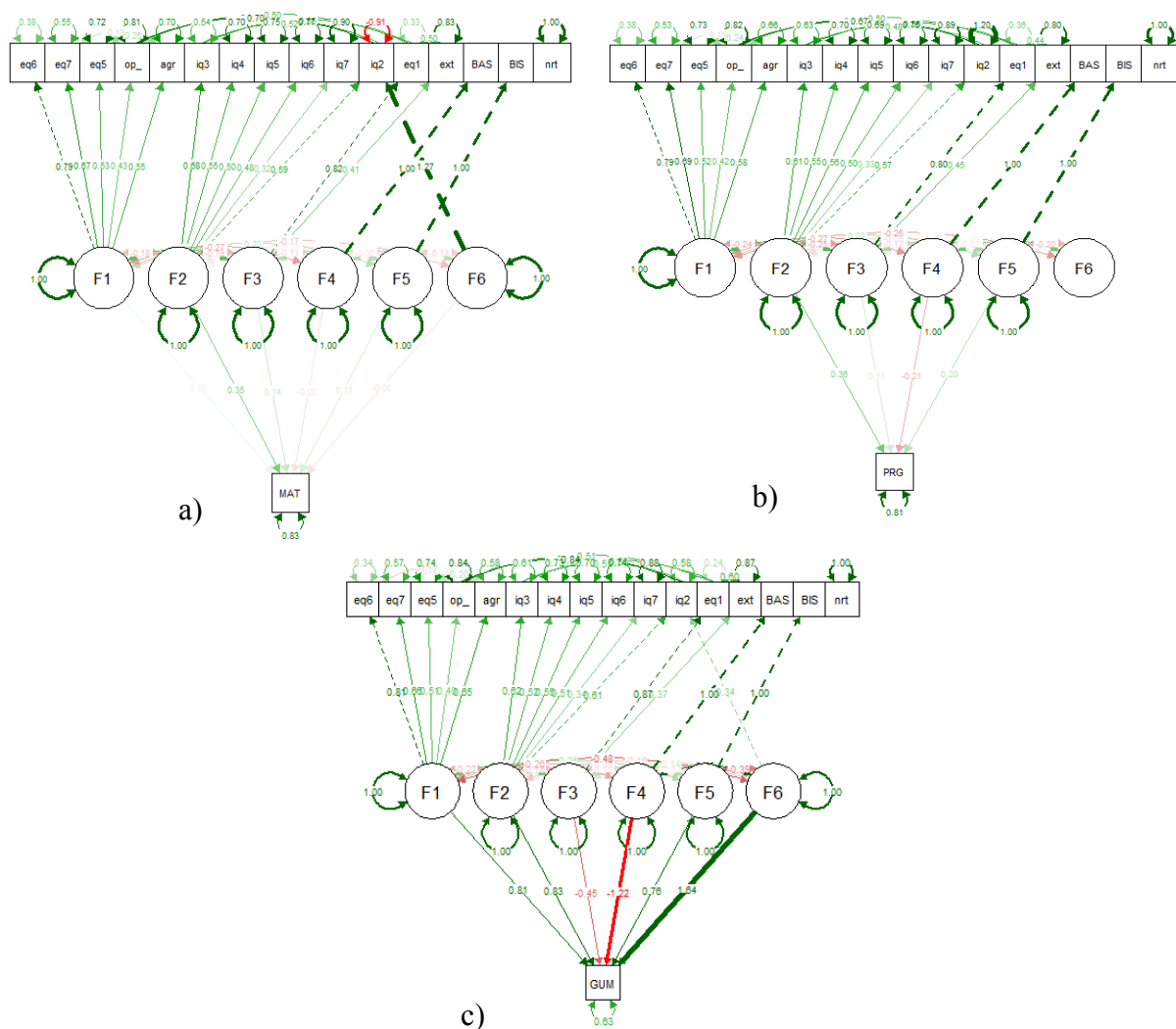


Рис. 8. Структурные модели влияния на успеваемость математических (а), профессиональных (б), и гуманитарных (с) дисциплин

Во всех трех моделях в разрезе групп предметов факторы стимулов награды и наказания представлены в двух разных факторах. Последнее может говорить о том, что использование мотивационных стимулов, как отрицательных, так и положительных, требует определенной стратегии с учётом особенностей обучаемых.

5. Заключение

Выполненное исследование позволило определить степень влияния основных когнитивных, личностных и мотивационных характеристик студентов на различные виды успеваемости. Выявленный набор показателей, оказывающих значимое влияние на успеваемость и при этом не дублирующих друг друга, предполагается использовать для входного диагностического тестирования студентов при адаптации модели обучаемого в реализуемой PLE.

Кроме того, при реализации PLE полученная модель будет включена в архитектуру информационных технологий университета в виде онтологии «Модель обучаемого» (организация интерфейса посредством SWI-Prolog, JSON, API). Совместно с онтологией «Модель образовательного контента» её предполагается использовать для рекомендации студентам образовательного контента, наилучшим образом учитывающего их индивидуальные особенности. Мы предполагаем персонализировать выдачу контента, используя методы интеллектуальной поддержки на основе онтологических моделей. «Модель обучаемого» объединяет характери-

стики, отвечающие за представление различных аспектов личности, что позволяет использовать ее для описания особенностей поведения студентов и последующей их классификации, например, в зависимости от стратегии использования контента. Онтология «Образовательный контент» содержит описание различных признаков современного контента, используемого в образовательном процессе. На основании правил и аксиом, а также знаний, представленных двумя онтологиями в среде Protégé, разрабатываемая система будет рекомендовать форму и стратегию выдачи образовательного контента для конкретного студента или группы студентов, обладающих схожими характеристиками.

Онтология, представляющая модель обучаемого (Learner model), учитывает не только экспертные знания, изначально заложенные в систему, но и результаты анализа статистических данных, которые можно оперативно изменять с учетом изменения контингента студентов. Таким образом, используя стратегию самообучения интеллектуальной системы, можно обновлять знания системы путем добавления экземпляров, а также корректировки связей между классами и экземплярами, что позволит оперативно отслеживать изменения и своевременно отражать их в PLE.

Литература

1. Олейникова О. Н., Редина Ю. Н. Макротренды развития системы высшего образования в мире. [Электронный ресурс]. URL: <http://www.vestnik.vsu.ru/pdf/educ/2018/03/2018-03-02.pdf> (дата обращения: 11.12.2021).
2. Россия 2025: от кадров к талантам. [Электронный ресурс]. URL: http://d-russia.ru/wp-content/uploads/2017/11/Skills_Outline_web_tcm26-175469.pdf (дата обращения: 11.12.2021).
3. Глобальная конкурентоспособность российского образования. [Электронный ресурс]. URL: <https://publications.hse.ru/books/223161350> (дата обращения: 11.12.2021).
4. Драгунова Е. В., Пустовалова Н. В., Вальдман И. А. Инновационные технологии в дизайне образовательных экосистем // Материалы МНТК «Актуальные проблемы электронного приборостроения (АПЭП-2018)». Т. 8. С. 62–69.
5. Васильченко С. Х. Формирование персональной образовательной среды на основе информационных технологий для реализации индивидуальных траекторий обучения (на примере корпоративного обучения): автореф. дисс. ... канд. пед. наук. М., 2012.
6. Кочергина Е. В., Най Дж. В. К., Орёл Е. А. Факторы «Большой пятерки» как психологические предикторы академической успеваемости студентов вузов // Психологические исследования. 2013. Т. 6. № 27. [Электронный ресурс]. URL: <http://psystudy.ru/index.php/num/2013v6n27/776-kochergina27.html> (дата обращения: 11.12.2021).
7. Збанацкая А. Б. Личностные характеристики студентов вуза как основа индивидуального стиля учебной деятельности: дисс. ...канд. псих. наук: М., 2007.
8. Малахова С. И. Связь психометрического интеллекта с личностной саморегуляцией студентов: автореф. дисс. ... канд. пед. наук. М., 2013.
9. Gilmore P., Self J. The application of machine learning to intelligent tutoring systems. Artificial Intelligence and Human Learning. Intelligent Computer-aided Instruction. Chapman and Hall, London, 1988. pp. 179–196.
10. Пустовалова Н. В. Онтологическая модель процесса взаимодействия образовательного контента и обучаемого в архитектуре интеллектуальных обучающих систем // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2021. № 4. С. 113–125.
11. Ржанова И. Е., Бритова В. С., Алексеева О. С., Бурдукова Ю. А. Флюидный интеллект: обзор зарубежных исследований // Клиническая и специальная психология. 2018. Т. 7,

- № 4. С. 19–43. [Электронный ресурс]. URL: doi:10.17759/cpse.2018070402 (дата обращения: 11.12.2021).
12. Математические методы в психологии. Основы применения (Mathematical Methods in Psychology: Basics of Applying) / Санкт-Петербургский государственный университет. [Электронный ресурс]. URL: <https://www.coursera.org/learn/matematicheskiye-metody-v-psikhologii> (дата обращения: 11.12.2021).
13. RDocumentation [Электронный ресурс]. URL: <https://www.rdocumentation.org/> (дата обращения: 11.12.2021).

Статья поступила в редакцию 19.12.2021;
переработанный вариант – 18.02.2022.

Авдеенко Татьяна Владимировна

д.т.н., профессор, профессор кафедры теоретической и прикладной информатики НГТУ (630073, Новосибирск, пр. К. Маркса, 20), тел. (383) 3-460-600, e-mail: tavdeenko@mail.ru.

Пустовалова Наталья Викторовна

аспирант, старший преподаватель кафедры автоматизированных систем управления НГТУ, тел. (383) 3-461-100, e-mail: NVPustovalova@gmail.com.

References

1. Olejnikova O. N., Redina Ju. N. *Makrotrendy razvitija sistemy vysshego obrazovanija v mire. [Jelektronnyj resurs]*. [Macrotrends in the development of the higher education system in the world]. URL: <http://www.vestnik.vsu.ru/pdf/educ/2018/03/2018-03-02.pdf> (access date: 11.12.2021).
2. *Rossija 2025: ot kadrov k talantom. [Jelektronnyj resurs]* [Russia 2025: from personnel to talents]. URL: http://d-russia.ru/wp-content/uploads/2017/11/Skills_Out-line_web_tcm26-175469.pdf (access date: 11.12.2021).
3. *Global'naja konkurentosposobnost' rossijskogo obrazovanija. [Jelektronnyj resurs]* [Global competitiveness of Russian education]. URL: <https://publications.hse.ru/books/223161350> (access date: 11.12.2021).
4. Dragunova E. V., Pustovalova N. V., Val'dman I. A. *Innovacionnye tehnologii v dizajne obrazovatel'nyh jekosistem* [Innovative technologies in the design of educational ecosystems]. *Materialy MNTK «Aktual'nye problemy jelektronного priborostroenija (APJeP-2018)»* [Materials of the MNTK "Actual problems of electronic instrumentation (APEP-2018)"]. vol. 8, pp. 62-69.
5. Vasil'chenko S. H. *Formirovanie personal'noj obrazovatel'noj sredy na osnove informacionnyh tehnologij dlja realizacii individual'nyh traektorij obuchenija (na primere korporativnogo obuchenija)* [Formation of a personal educational environment based on information technologies for the implementation of individual learning paths (on the example of corporate training)]: *avtoref. diss. ... kand. ped. nauk*. Moscow, 2012.
6. Kochergina E. V., Naj Dzh. V. K., Orjol E. A. *Faktory «Bol'shoj pjaterki» kak psihologicheskie prediktory akademicheskoy uspevaemosti studentov vuzov* [Factors of the "Big Five" as psychological predictors of academic performance of university students]. *Psihologicheskie issledovanija*. 2013, vol.6, no. 27. [Electronic resource]. URL: <http://psystudy.ru/index.php/num/2013v6n27/776-kochergina27.html> (access date: 11.12.2021).
7. Zbanackaja A. B. *Lichnostnye harakteristiki studentov vuza kak osnova individual'nogo stilja uchebnoj dejatel'nosti* [Personal characteristics of university students as the basis of individual style of learning activity]: *diss. ...kand. psih. nauk*. Moscow, 2007.

8. Malahova S. I. *Svjaz' psihometricheskogo intellekta s lichnostnoj samoreguljaciej studentov* [Relationship of psychometric intelligence with personal self-regulation of students]: *avtoref. diss. ... kand. ped. nauk.* Moscow, 2013.
9. Gilmore P., Self J. The application of machine learning to intelligent tutoring systems. *Artificial Intelligence and Human Learning. Intelligent Computer-aided Instruction*. Chapman and Hall, London, 1988, pp. 179-196.
10. Pustovalova N. V. *Ontologicheskaja model' processa vzaimodejstvija obrazovatel'nogo kontenta i obuchaemogo v arhitekture intellektual'nyh obuchajushhih sistem* [Ontological model of the process of interaction between educational content and the student in the architecture of intelligent learning systems]. *Vestnik Astrahanskogo gosudarstvennogo tehničeskogo universiteta. Serija: Upravlenie, vychislitel'naja tehnika i informatika*. 2021, no. 4, pp. 113-125.
11. Rzhanova I. E., Britova V. S., Alekseeva O. S., Burdukova Ju. A. *Fljuidnyj intellekt: obzor zarubezhnyh issledovanij* [Fluid intelligence: a review of foreign studies] *Klinicheskaja i special'naja psihologija*. 2018, vol. 7, no. 4, pp. 19-43. [Electronic resource]. URL: doi:10.17759/cpse.2018070402 (access date: 11.12.2021).
12. *Matematicheskie metody v psihologii. Osnovy primenenija* [Mathematical Methods in Psychology: Basics of Applying]. *Sankt-Peterburgskij gosudarstvennyj universitet*. [Electronic resource]. URL: <https://www.coursera.org/learn/matematicheskiye-metody-v-psikhologii> (access date: 11.12.2021).

Multivariate analysis of the students' characteristics influences on academic progress for the personalization of educational environment

Tatiana N. Avdeenko

Doctor of technical sciences, Professor, Novosibirsk State Technical University (Novosibirsk, Russia), tavdeenko@mail.ru.

Natalia V. Pustovalova

Postgraduate student, Novosibirsk State Technical University (Novosibirsk, Russia), NVPustovalova@gmail.com.

Modern universities developing a digital environment rely on a student-centered approach, relevant pedagogical and didactic practices taking into account learners' characteristics and adapting it to the educational process. Personalization of the educational environment should improve the quality of the educational process raising it to a new level. This article presents the results of the multivariate modeling methods to identify standard relations and patterns between academic performance and students' psychometric characteristics. The results of students testing at the Novosibirsk State Technical University (NSTU) determining their cognitive, motivational, and personal characteristics are considered in this paper and information about all past certification tests is provided for the curriculum.

Keywords: regression analysis, principal component analysis, structural equation modeling, learner's model, psychometric characteristics.

Метод селекции ионов для их идентификации в газовых средах

В. Н. Лиманский, Б. З. Горбунов

Данный метод относится к задаче выделения ионов с заданной подвижностью из их смеси с электрически нейтральными молекулами. В частности, он позволяет извлекать интересные ионы из потока анализируемого газа, чтобы их можно было направить в поток чистого газа, где их можно было бы определить количественно. Метод обеспечивает более эффективное разделение, увеличивает разрешение устройства селекции ионов и, следовательно, позволяет количественно оценивать сложные смеси при низких концентрациях молекул в газах, представляющих интерес.

Ключевые слова: идентификация ионов, подвижность ионов, обнаружение следовых количеств в воздухе.

1. Предпосылки для создания новой методики

Количественное определение следовых количеств соединений в воздухе часто требует разделения сложной смеси и выбора интересующих молекул для анализа. Одним из методов, используемых при этом для разделения ионов при атмосферном давлении, является спектрометрия подвижности ионов [1], которая широко используется для многих приложений безопасности, например, для обнаружения взрывчатых веществ в аэропортах.

Для разделения электрически заряженных аэрозольных частиц в воздухе использовалась технология дифференциального анализатора подвижности [2]. Недавно этот метод был расширен до разделения ионов при атмосферном давлении [2, 3].

В [4] предложен тип анализатора мобильности ионов в потоке, содержащего по меньшей мере два электрода, расположенных так, чтобы создавать электрическое поле между ними, и поток газа, который направлен вдоль вектора напряжённости электрического поля. Ионы переносятся через канал потоком газа, причём ионы определённой подвижности захватываются электрическим полем и попадают на электрод, а поток газа отводится в стороны. Когда ионы достигают конца канала, детектор на конце канала видит непрерывный поток ионов с выбранной подвижностью. Ионы с той или иной подвижностью отбираются путём изменения электрического поля и/или скорости потока.

Другой метод дифференциального анализатора подвижности с поперечным потоком описан в [5]. Этот метод обеспечивает выполнение спектрометрии подвижности ионов или частиц, детектируя их с помощью многоканального селектора. Система сначала получает образец для анализа, затем ионизирует образец и вводит ионизированный образец в ламинарный газовый поток. Электрическое поле пересекает ламинарный газовый поток, так что ламинарный газовый поток и электрическое поле объединяются, чтобы пространственно разделить ионы анализируемого газа на основе ионной подвижности и чтобы затем пространственно разделённые ионы контактировали с различными элементами многоканального электрода. Далее система анализирует ток многоканального электрода, чтобы определить подвижность ионов анализируемого газа.

В дифференциальных анализаторах подвижности ионов разрешающая способность (РС) определяется отношением зазора между электродами к толщине пучков траекторий ионов в потоке несущего газа. Траектории ионов и нейтральных молекул регулируются законами движения сплошных сред, которые приводят к выражению: $PC = Q_{sh}/Q_i$, где Q_{sh} – величина объёмного расхода воздуха в зазоре между электродами, Q_i – объёмный расход потока анализируемого образца. Таким образом, для уменьшения толщины пучков траекторий ионов необходимо увеличивать скорость потока несущего газа через зазор между электродами или/и уменьшать скорость потока ионов образца.

На практике для увеличения разрешения РС поток в зазоре увеличивается. Чтобы достичь разрешающей способности, достаточной для практических приложений, поток в зазоре между электродами обычно должен быть намного больше, чем поток анализируемого образца ионов, в результате чего скорость потока газа в зазоре между электродами часто может быть близка к скорости звука. Это создает две существенные проблемы. Во-первых, для создания таких высоких скоростей потока газа требуются мощные и, следовательно, большие и дорогие насосы. Во-вторых, это приводит к высоким числам Рейнольдса, – таким образом, поток становится турбулентным. Турбулентность оказывает сильное влияние на разрешающую способность, уменьшая её из-за образования вихрей и увеличивая уширение траекторий ионов. Увеличивать РС за счет уменьшения потока пробы также нежелательно, потому что это уменьшает количество ионов, выходящих из анализатора, и, следовательно, снижает чувствительность.

2. Новый метод

Настоящий метод обеспечивает концептуально другой способ улучшения селекции ионов и увеличения разрешающей способности за счёт использования входа образца для фокусирования ионов перед попаданием в зазор между электродами. Это сужает пучки траекторий ионов и позволяет избежать или уменьшить необходимость увеличения скорости потока через зазор между электродами.

Соответственно, в одном аспекте метод содержит установку для отделения ионов с заданной подвижностью от газовой смеси ионов с различной подвижностью с использованием устройства анализатора дифференциальной подвижности, при этом устройство анализатора содержит полость (камеру) разделения ионов, имеющую:

- (а) вход для потока измеряемого газа, содержащего ионы;
- (б) фокусирующую камеру, отверстие на одном конце которой служит впускным отверстием для потока измеряемого газа, через которое измеряемый газ с ионами может поступать в камеру разделения ионов;
- (с) впускное отверстие для газа в камеру разделения ионов, соединенное или подсоединенное к источнику газа для поддержания требуемой скорости и объёмного расхода Q_{sh} в камере;
- (d) выпускное отверстие для газа из камеры разделения ионов;
- (е) выпускное отверстие для газа, содержащего ионы с заданной подвижностью;
- (f) два электрода, предназначенных для создания электрического поля разделения ионов внутри камеры разделения ионов.

При этом зона фокусировки предусмотрена в камере фокусировки, причём зона фокусировки содержит по меньшей мере две поверхности, на которых может создаваться неоднородная плотность электрического заряда.

Эта установка основана на новом способе выделения ионов с заданной подвижностью ионов который включает:

- (i) введение потока чистого (без ионов) газа в камеру разделения ионов через впускное отверстие с заданной скоростью потока (или объёмного расхода), Q_{sh} ;

(ii) введение потока газа-пробы, содержащего ионное облако, в камеру фокусировки и через неё, с заданной скоростью потока (или объёмного расхода) (Qi) в зону фокусировки;

(iii) воздействие на облако ионов в пробе газа в зоне фокусировки неоднородным электрическим полем, создаваемым неоднородной плотностью электрического заряда, так что электрическое поле изменяет траектории ионов в пробе газа таким образом, что они сходятся, образуя поток ионов уменьшенной ширины;

(iv) направление потока исследуемого газа, содержащего ионный поток уменьшенной ширины, из фокусирующей камеры через впускное отверстие для потока измеряемого газа в камеру разделения ионов;

(v) выбор напряжённости разделяющего поля таким образом, чтобы привлекать ионы с заданной подвижностью ионов к выходному отверстию для ионов.

В большинстве известных типов дифференциальных анализаторов подвижности (как показано на рис. 1) поток газа пробы входит в камеру разделения ионов, так что направление его потока совпадает с направлением потока газа в зазоре между электродами. Это достигается с помощью перегородки, которая направляет поток измеряемого газа, так что он течёт в направлении, по существу, параллельном потоку газа в зазоре, прежде чем слиться с потоком газа в зазоре на конце перегородки. Одно существенное различие между известными устройствами дифференциальных анализаторов подвижности и настоящим способом разделения ионов состоит в том, что в настоящем способе нет перегородки, направляющей поток газа пробы в направлении, параллельном потоку газа в оболочке. Вместо этого исследуемый газ проходит через камеру фокусировки, которая ориентирована под углом девяносто градусов относительно направления потока газа в зазоре вдоль камеры разделения ионов и непосредственно через входное отверстие для газа-пробы в камеру разделения ионов. Таким образом, поток анализируемого газа входит в камеру разделения ионов под углом девяносто градусов по отношению к направлению потока газа в зазоре.

В новом методе используется существенно неоднородное фокусирующее электрическое поле, создаваемое электрическим зарядом неоднородной плотности. Электрическое поле имеет компоненту, которая ориентирована преимущественно поперёк потока измеряемого газа и имеет свойство изменять траектории ионов в потоке измеряемого газа таким образом, чтобы они сходились, то есть перемещались ближе к центральной линии, проходящей через зону фокусировки.

Электрический заряд с неоднородной плотностью может быть областью с высоким локализованным зарядом. Области локализованного заряда могут быть обеспечены электродами, расположенными внутри фокусирующей камеры. Неоднородное фокусирующее электрическое поле может быть создано двумя или более разнесёнными заряженными областями одинаковой полярности – например, пара разнесённых положительно заряженных областей. Две или более разнесённые области с одинаковой полярностью заряда создают противоположные электрические поля. Ионы, проходящие через противоположно направленные электрические поля, – одинаковой полярности, поэтому подвергаются силам отталкивания, которые заставляют их двигаться к центральной точке между разнесёнными друг от друга заряженными областями или к точке, где отталкивающие эффекты разнесённых друг от друга заряженных областей уравновешиваются.

Степень изменения траекторий ионов при их прохождении через фокусирующее электрическое поле будет зависеть как от силы электрического поля, так и от времени, проведённого ионами в электрическом поле, которое, в свою очередь, будет зависеть, по крайней мере, в некоторой степени от скорости прохождения газа через электрическое поле. Если скорость потока газа пробы слишком высока, ионы не будут подвергаться воздействию электрического поля достаточно долго, чтобы существенно повлиять на их траектории. Поэтому скорость потока измеряемого газа выбирается таким образом, чтобы ионы подвергались воздействию электрического поля в течение достаточно длительного времени, чтобы могла произойти фокусировка.

Явление модификации траекторий ионов таким образом, что они сходятся, давая ионный поток уменьшенной ширины, может быть обозначено здесь как «свёртка», а действие изменения траекторий ионов таким образом, что они сходятся, может быть названо как «сближение».

Первый критерий сближения определяет взаимосвязь между скоростью потока пробы, скоростью потока газа в зазоре, длиной разделительной камеры анализатора, которая косвенно связана с напряженностью электрического поля, а геометрия зоны фокусировки может быть выражена следующим образом:

$$(Hin/L) \cdot (Qsh/Qi) > Pf,$$

где Hin – толщина потока ионов образца, L – длина камеры, Qsh – объёмный расход потока чистого воздуха в зазоре между электродами, Qi – объёмный расход потока ионов, а Pf – фактор фокусировки ($1 < Pf < 30$). Данное выражение основано на экспериментальных наблюдениях за работой дифференциальных анализаторов подвижности различной геометрии, работающих в различных режимах. Это наиболее важная особенность метода, показывающая, что геометрия входа влияет на фокусировку и, следовательно, на разрешающую способность.

Как указано выше, расход пробы газа (который равен Qi в приведённом выше выражении) влияет на продолжительность воздействия на ионы фокусирующего электрического поля и, следовательно, на степень изменения траекторий ионов. Кроме того, ширина потока измеряемого газа (которая равна Hin в приведённом выше выражении) будет определять начальную ширину или толщину потока ионов, движущихся к камере разделения ионов. Чем больше ширина потока измеряемого газа, тем большая степень фокусировки требуется для получения хорошего разрешения.

Так, например, эксперименты, проведённые с различной шириной потока пробы (Hin), показали, как на разрешающую способность влияет геометрия входа (табл. 1).

Таблица 1. Разрешающая способность, измеренная для различных Hin в экспериментах с ионами ацетонитрила в азоте ($L = 10$ мм, $Qsh = 4$ л/мин и $Qi = 0.25$ л/мин были постоянными)

Hin (мм)	PC	
	Измерения с фокусировкой	Расчёт без фокусировки
0.1	14	16
0.2	16	16
0.4	18	16
0.8	21	16
1.5	22	16
3.0	21	16

Увеличение Hin с 0.1 мм до 3 мм показывает увеличение разрешающей способности с четырнадцати до двадцати одного. Этот эффект нельзя объяснить в терминах обычного анализатора, где $Rp = Qsh/Qi = 16$. Точность измерения PC составила ± 1.2 . Таким образом, небольшие отклонения в измеренных значениях Rp на 0.1 и 3.0 мм можно объяснить точностью измерений.

В этом методе используется электрическое поле, составляющая которого ориентирована поперёк потока измеряемого газа. При условии, что вышеупомянутый критерий фокусировки (свёртывания) выполняется, ионы в потоке пробы будут подталкиваться к центру потока, образуя узкие пучки ионов, без необходимости увеличения скорости потока через камеру разделения. Способ нового метода настолько эффективен, что позволяет формировать очень узкий поток ионов даже при отсутствии потока газа в камере, хотя, конечно, на практике поток газа в камере используется при разделении и сборе ионов различной подвижности.

Следует отметить, что критерий свёртки содержит, в дополнение к двум параметрам, которые контролируют толщину пучка траекторий ионов в обычном анализаторе (то есть Qsh и Qi), два дополнительных параметра (Hin и L), которые определяют фокусировку ионов в соответствии с настоящим методом. Соответствие критерию свёртки позволяет улучшить разрешение дифференциальных анализаторов подвижности за счёт увеличения значения другого

отношения – H_{in}/L . В обычном устройстве на разрешение влияет только отношение Q_{sh}/Q_i . Это еще одно отношение, позволяющее улучшить разрешение, не увеличивая Q_{sh} настолько, чтобы поток воздуха в камере разделения становился турбулентным. Это позволяет устройству быть меньше, дешевле и обеспечивать более высокое разрешение.

В данном методе обычно используется зона свёртки траекторий ионов, включающая камеру фокусировки (которая может также упоминаться здесь как впускной накопитель ионов), примыкающей к камере разделения ионов, по крайней мере, с двумя поверхностями, где может быть создана неоднородная плотность электрического заряда, через которую может быть направлен поток образца. Неоднородная плотность электрического заряда приводит к возникновению, по существу, неоднородного фокусирующего электрического поля, имеющего компоненту, ориентированную преимущественно поперёк потока образца, и имеет свойство быть способным сворачивать траектории ионов в потоке газа пробы через камеру фокусировки, так что ионы перемещаются ближе к центральной линии зоны свёртки.

Второй критерий фокусирования/свёртывания, который определяет взаимосвязь между расходом измеряемого газа, напряженностью электрического поля или на практике разностью потенциалов и геометрией зоны фокусировки, может быть выражен следующим образом:

$$\left(1 + \frac{\Delta V_{in}}{\Delta Z_{in}} \cdot \frac{H}{\Delta V}\right) \cdot \frac{H_{in}}{L} \cdot \frac{Q_{sh}}{Q_i} > P_f$$

где H_{in} – толщина потока ионов пробы, L – длина камеры разделения, H – зазор между электродами камеры разделения ионов, ΔV – разность потенциалов между электродами камеры разделения ионов, Q_{sh} – поток чистого газа в зазоре, Q_i – поток газа пробы, ΔV_{in} – разность потенциалов на поверхности входа для образца, ΔZ_{in} – глубина скважины зоны свертки, где применяется разность потенциалов ΔV_{in} , а P_f – фактор фокусировки. Обычно для коэффициента фокусировки P_f попадает в диапазон ($1 < P_f < 40$).

Следует отметить, что во втором критерии свертки $\Delta V_{in} = 0$, это соответствует первому критерию свертки. Зона фокусировки (свёртки) может иметь различные формы.

Она может быть создана пучками электрических зарядов, расположенными на конце камеры фокусировки, то есть там, где отверстие в конце камеры фокусировки служит входом пробы газа. Например, отверстие в конце камеры фокусировки может совпадать с прорезью, зазором или выемкой в одном из электродов камеры разделения ионов, так что края электрода образуют ободок отверстия и обеспечивают пучки электрических зарядов, необходимых для создания фокусирующего электрического поля. Следует понимать, что в этом варианте напряженность электрического поля, создаваемого электрическими зарядами, будет направлена как вверх по потоку назад вдоль камеры фокусировки, так и внутрь камеры разделения ионов.

В одном варианте зона фокусировки обеспечивается электропроводящей полосой, расположенной в стенке фокусирующей камеры, при этом к этой проводящей полосе приложена разность потенциалов ΔV_{in} . Полоска изолирована окружающим электрическим изолятором от любых других внутренних проводящих поверхностей камеры фокусировки и расположена в месте Z_{in} от отверстия в конце фокусировки. Вертикальный размер, или ширина полосы, равен ΔZ_{in} . Изолятор также изолирует полосу от электродов в камере разделения ионов. Подходящие значения для ΔZ_{in} , координаты Z_{in} и напряжения ΔV_{in} можно найти обычным методом проб и ошибок. Этот вариант осуществления выгоден тем, что имеет большую эффективность фокусировки, поскольку дополнительное напряжение ΔV_{in} обеспечивает дополнительное усиление фокусирующего электрического поля в фокусирующей камере.

В другом варианте осуществления способа фокусирующая камера снабжена множеством электрически проводящих полос, расположенных на её стенке, при этом к каждой электрически проводящей полосе приложена разность потенциалов ΔV_{in} , тем самым создавая фокусирующее электрическое поле, причём каждая полоса изолирована окружающим электрическим изолятором от любых других проводящих полос и внутренних проводящих поверхностей фокусирующей камеры.

Таким образом, в этом варианте осуществления метода имеется несколько дополнительных электропроводящих полос (подобных тем, которые описаны в предыдущем варианте),

расположенных в определенных местах (координатах) Z_{in} , $Z_{in\alpha}$ и $Z_{in\beta}$. Каждая полоска имеет приложенную к ней разность потенциалов: ΔV_{in} , $\Delta V_{in\alpha}$ и $\Delta V_{in\beta}$ соответственно. Полоски изолированы друг от друга и от других внутренних поверхностей фокусирующей камеры электрическим изолятором. Расположение верха полосковых электродов относительно отверстия в конце фокусирующей камеры Z_{in} , $Z_{in\alpha}$ и $Z_{in\beta}$, их вертикальные размеры (ширина) ΔZ_{in} , $\Delta Z_{in\alpha}$ и $\Delta Z_{in\beta}$ и напряжения ΔV_{in} , $\Delta V_{in\alpha}$ и $\Delta V_{in\beta}$ могут быть определены эмпирически. Этот вариант осуществления выгоден тем, что каждая из дополнительных полосок, дополнительных напряжений ΔV_{in} , $\Delta V_{in\alpha}$ и $\Delta V_{in\beta}$ обеспечивает дополнительное усиление фокусирующего электрического поля и таким образом улучшает разрешающую способность, обеспечиваемую устройством.

В другом варианте осуществления метода камера фокусировки на своей внутренней поверхности снабжена слоем проводящего материала, который изолирован от электродов камеры разделения ионов с помощью электроизоляционного материала, и при этом пара проводящих электродов соединена с проводящим электродом. Материал и указанные электроды, в свою очередь, подключены к источнику разности потенциалов, чтобы создать желаемый профиль потенциала вдоль внутренней поверхности фокусирующей камеры. Таким образом, для вертикальной скважины желаемый профиль разности потенциалов $V(Z)$ будет создан внутри скважины как функция Z . Это создаёт неоднородное электрическое поле с фокусирующими свойствами для повышения разрешения.

На рис. 1 показан известный тип устройства дифференциального анализатора подвижности, имеющего вход для образца ионов (1), ведущий к камере разделения ионов (2), и перегородку (3), которая разделяет поток газа образца, проходящий через вход (1), от потока чистого газа (4) оболочки, вводимого через вход потока газа оболочки (5). В камере разделения ионов (2) два электрода расположены сверху (6) и снизу (7) камеры разделения ионов, соответственно, и они создают электрическое поле (8), которое при использовании притягивает ионы выбранной подвижности ионов, так что их траектории отклоняются выходной перегородкой для ионов (10) к выходу для ионов (9). Камера разделения ионов также имеет выход (11) для потока газа, содержащего невыбранные ионы.

Как только поток анализируемого газа покидает перегородку (3), на движение ионов в анализируемом газе влияет электрическое поле (8), в результате чего траектории ионов (12) и (13) отклоняются от преимущественно горизонтальных линий тока (4) в сторону электрода (6). Ионы с заданной подвижностью ионов будут притягиваться к выходу выбранных ионов (9) через секцию камеры (2), примыкающую ко второй перегородке (10).

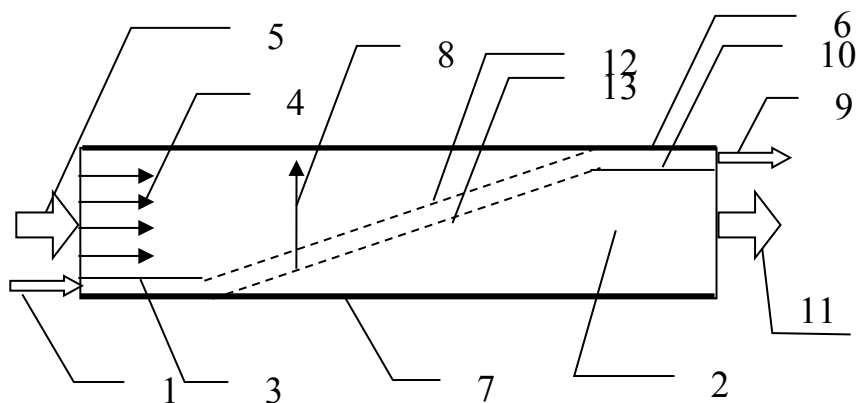


Рис.1. Известный тип устройства дифференциального анализатора подвижности

Ионы с более высокой подвижностью пересекут линию траектории (12) в направлении электрода (6) и будут разряжены. Ионы с меньшей подвижностью будут пересекать линию траектории (13) и будут уноситься потоком газа оболочки к выпускному отверстию для потока оболочки (11). Сканируя разность напряжений между электродами (6) и (7), можно получить спектр подвижности ионов в образце.

Следует принять во внимание, что рис. 1 представляет собой упрощённое двумерное изображение случая однородной скорости потока в камере разделения ионов (2). На самом деле профиль скорости $v(X, Y, Z)$ более сложен из-за граничных условий на внутренних поверхностях камеры: $v(X, Y, Z) = 0$. Здесь X , Y и Z – декартовы координаты внутри камеры (2), где X – горизонтальная координата по длине камеры разделения ионов, Z – вертикальная координата по высоте камеры разделения ионов, а Y – координата, ортогональная к X и Z .

Разрешающая способность устройства, показанного на рис. 1, определяется отношением скорости потока через зазор между электродами к скорости потока ионной пробы Q_{sh}/Q_{in} . Часто отношение Q_{in}/Q_{sh} используется для характеристики разрешения анализаторов, потому что оно представляет ширину пика иона в спектре подвижности. Расстояние между перегородкой 3 и электродом 7 обычно выбирается в соответствии с принципом непрерывности потока, чтобы исключить разницу скоростей между потоком измеряемого газа, содержащего ионы, и потоком в зазоре. Разрешающая способность может быть выражена как отношение Q_{in}/Q_{sh} , или чаще используют отношение: $PS = Q_{sh}/Q_{in}$. В обычном дифференциальном анализаторе подвижности на разрешение не влияет геометрия входов и выходов.

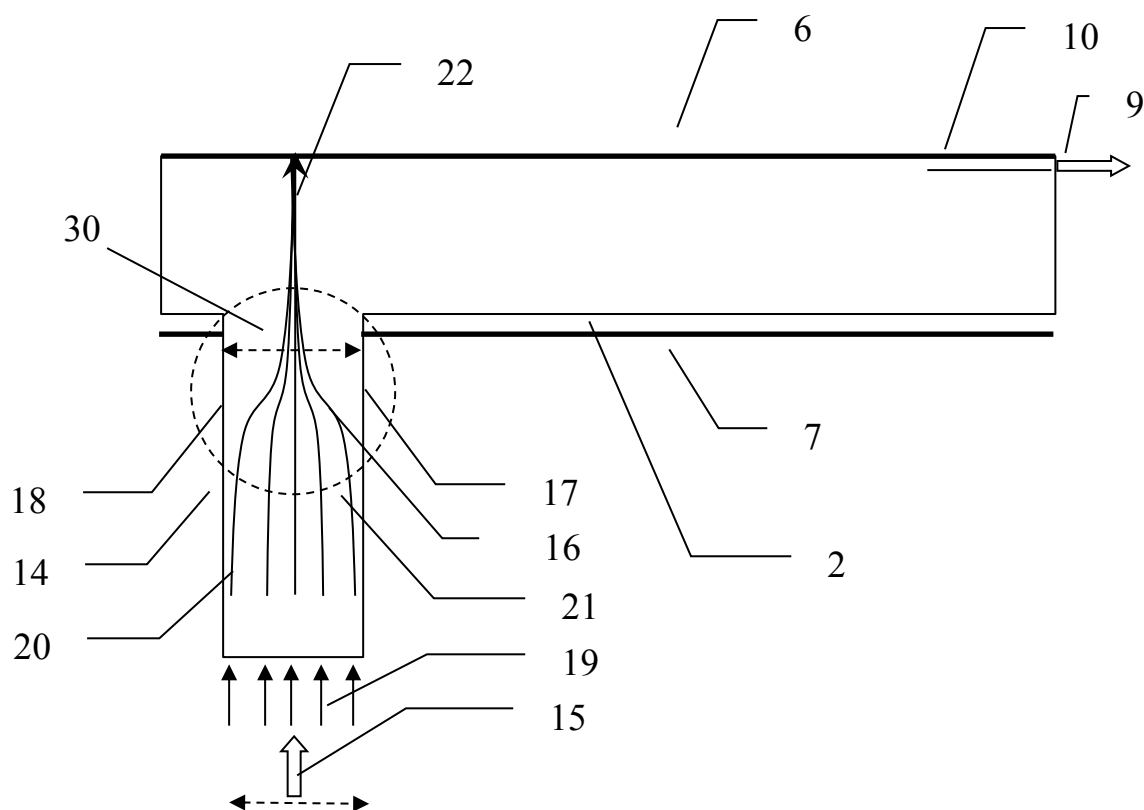


Рис. 2. Схема анализатора согласно первому варианту реализации

На рис. 2 показано устройство анализатора согласно первому варианту реализации. На рисунке показан аппарат без газового потока в зазоре, чтобы более наглядно продемонстрировать свёртку (схождение) траекторий ионов в зоне фокусировки траекторий ионов (14). Таким образом, устройство содержит камеру (2) разделения ионов, подключённую к источнику потока газа пробы (15), содержащего представляющие интерес ионы. Сопутствующие средства (например, насос) для поддержания скорости потока газа пробы в устройство не показаны. Устройство дополнительно содержит камеру для фокусировки ионов (16) с двумя поверхностями (17) и (18), каждая из которых имеет неоднородное распределение плотности электрического заряда, при этом стенки разнесены на расстояние H_{in} (показано двойной пунктирной линией). Камера разделения ионов (2) анализатора с электродами (6) и (7) также включает средства для создания разности напряжений между ними и тем самым создания электрического поля для разделения ионов (не показаны).

Hin, причём каждая из этих двух поверхностей имеет неоднородное распределение плотности электрического заряда одинаковой полярности (например, положительный заряд) на них. Электрические заряды на поверхностях (17) и (18) вызывают неоднородное электрическое поле, которое действует как зона фокусировки (14) для ионов в пробе газа.

Вход (15) камеры фокусировки (16) может быть соединён со средством (например, насосом, не показан) для поддержания потока газа пробы в камеру фокусировки (16).

На рис. 3 показана работа устройства, аналогичного изображённому на рис. 2, с включённым потоком газа в зазоре. Как и в варианте на рис. 2, проба газа, содержащая ионы, входит в отверстие (15) в фокусирующую камеру (16) и проходит вдоль фокусирующей камеры к выходу (30) и в камеру разделения ионов (2). Когда газ пробы проходит по фокусирующей камере (2), начальные траектории ионов (19) фокусируются в узкий пучок траекторий (22), как описано выше в отношении рис. 2. Газовый поток в зазоре (4) входит через входное отверстие для газа (5) и создаёт поток чистого газа через камеру (2). Поток измеряемого газа, содержащий сфокусированный поток ионов, движется в камеру (2) и уносится вниз по потоку газом потока в зазоре. По мере того как они проходят по камере (2), траектории ионов изменяются за счёт взаимодействия с электрическим полем и притягиваются к электроду (6). Степень, с которой ионы перемещаются к электроду (6), зависит от напряжённости поля и подвижности ионов в соответствии с их классом, а также от скорости объединённого потока газа в оболочке и потока газа пробы при заданной напряжённости электрического поля, к выходу попадают только ионы с определённой подвижностью (9). Ионы с большей подвижностью движутся быстрее и нейтрализуются на поверхности электрода (6). Ионы с меньшей подвижностью движутся медленнее и выносятся потоком оболочки в выходное отверстие камеры разделения (11). Напряжённость электрического поля можно регулировать, изменяя разность напряжений, приложенных к электродам (6) и (7), так, чтобы ионы с определённой подвижностью ионов собирались на выходе (9). Изменяя напряжённость электрического поля, можно получить спектр ионов с различной подвижностью. Описанный выше режим работы аналогичен режиму обычного анализатора, но отличается тем, что фокусировка ионов в фокусирующей камере (16) уменьшает толщину пучка ионных траекторий (22), что позволяет уменьшить поток в оболочке без снижения разрешения или улучшения разрешения без увеличения скорости потока оболочки до соотношения скорости потока пробы.

На практике было замечено, что первый критерий свёртки позволяет достичь $PC = 30$ при очень небольшом расходе через зазор: $Q_{sh} = 4$ л/мин. Это значительно меньше, чем поток в зазоре, необходимый для получения той же разрешающей способности с обычным инструментом, где требуется $Q_{sh} > 100$ л/мин. Следовательно, с дифференциальным анализатором подвижности, разработанным согласно настоящему предложению на основе критерия свёртки, можно построить небольшое и дешёвое портативное устройство.

Второй критерий свёртки позволяет ещё больше увеличить разрешающую способность. Таким образом, устройство, разработанное в соответствии со вторым критерием свёртки, продемонстрирует ещё больший потенциал миниатюризации и увеличения разрешения.

Следует уточнить, что критерии свёртки и значение Pf не предсказывают количественно разрешающую способность. Они должны указать направление увеличения РС. Разрешающая способность – это монотонная функция фокусирующего фактора $PC(Pf)$, которую можно представить в дифференциальной форме $dPC(Pf)/dPf > 0$. В общем случае зависимость РС от Pf является сложной функцией.

3. Эксперименты

Устройство селекции ионов согласно варианту осуществления, показанному на рис. 2 и 3, было построено и испытано. Исследованы различные размеры входного отверстия для образца. Все металлические части, включая электрод (6) и (7), были изготовлены из нержавеющей стали. Прокладка (перегородка) 10 мм между электродами (6) и (7) была изготовлена из

ПТФЭ. В одном варианте ширина фокусирующей камеры (16) составляла $H_{in} = 1$ мм. Зазор между электродами в камере анализатора (2) составлял $H = 8$ мм. Длина зоны селекции, или расстояние между входом для ионов и выходом для ионов, составляла 10 мм. Зазор между верхним электродом (6) и выходной перегородкой (10) на входе (9) составлял 0.5 мм. В камере анализатора были две пористые мембраны, пропускающие воздух через поток оболочки, чтобы уменьшить турбулентность и создать режим ламинарного потока, не показанный на чертежах. Мембраны располагались слева от входа для ионов и справа от выхода для ионов и были изготовлены из полиамида (нейлона) толщиной 0.5 мм. Выходная перегородка изготовлена из нержавеющей стали толщиной 0.2 мм. Остальные размеры устройства были такими, как описано в [7].

Напряжение между электродами подавалось пилообразным генератором, специально разработанным для данного анализатора и позволяющим достичь линейного изменения напряжения от 1 В до 1000 В за время сканирования от 10 до 1000 с. Это позволило записать количество отсчётов в зависимости от напряжения сканирования с использованием ионного счётчика типа, описанного в US 7,372,020 [8]. Полученные данные можно использовать для расчёта количества отсчётов в зависимости от электрической подвижности. Процедура этого расчёта широко описана в технических учебниках, а также см. [7].

Типичный спектр образца, содержащего ацетонитрил, записанный с помощью описанного выше устройства, показан на рис. 4.

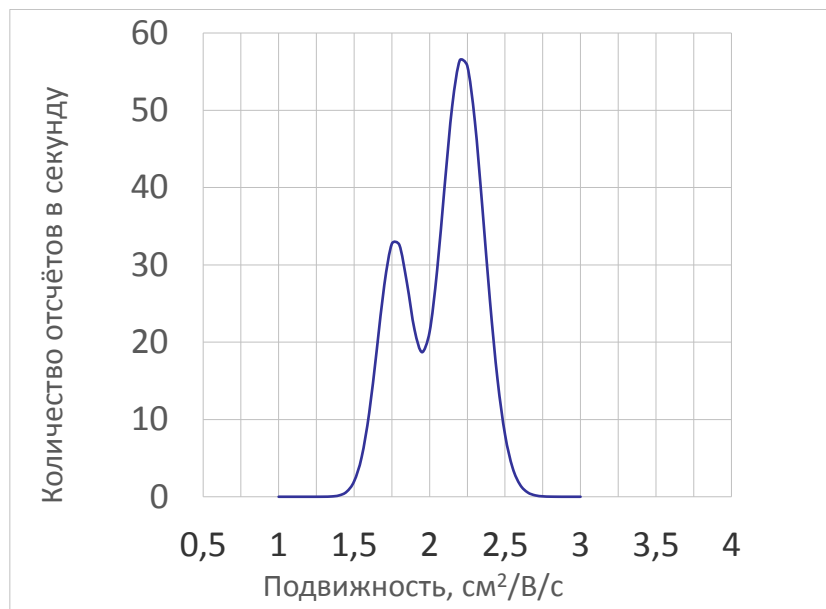


Рис. 4. Спектр подвижности ионов, характерный для изложенного метода

Тот факт, что в спектре присутствуют два пика, демонстрирует высокое разрешение, обеспечиваемое предложенным устройством. Без фокусировки ионов, обеспечиваемой устройством по настоящему методу, два таких пика не могут быть разрешены.

Литература

1. *Eiceman G. A.* Ion-mobility spectrometry as a fast monitor of chemical composition // *Trends in Analytical Chemistry*. 2002. V. 21, № 4. P. 259–275.
2. *Martínez-Lozano P., Fernández de la Mora J.* Resolution improvements of a nano-DMA operating transonically // *J. Aerosol Sci.* 2006. V. 37. P. 500–512.
3. *Santos J. P., E. Hontanón, E. Ramiro and M. Alonso* Performance evaluation of a high-resolution parallel-plate differential mobility analyser // *Atmos. Chem. Phys.* 2009. V. 9. P. 2419–2429.

4. Патент США 2005/0006578; A1, Alan L. Rockwood, Edgar D. Lee, Nosa Agbonkonkon, Milton L. Lee.
5. Патент США 2006/0054804; A1, Anthony S. Wexler.
6. Патент США 7,928,374; B2, Juan Rus-Perez and Juan Fernandez de la Mora.
7. Steer B., Gorbunov B., Muir R., Ghimire A. and Rowles J. Portable Planar DMA: Development and Tests // *Aerosol Science and Technology*. 2014. V. 48. P. 250–259.
8. Патент США 7,372,020 B2. Счетчик ионов; Б. Горбунов.

Статья поступила в редакцию 19.11.2021.

Лиманский Владимир Николаевич

к.т.н., доцент кафедры ЦТРВиСПС СибГУТИ (630102, Новосибирск, Кирова, 86),
e-mail: v_lim@mail.ru.

Горбунов Борис Захарович

д.ф.-м.н., Ancon Technologies Limited, CIC, Canterbury, Kent CT2 7FG, UK,
e-mail: gorbunov1@aol.com.

References

1. Eiceman G. A. Ion-mobility spectrometry as a fast monitor of chemical composition. *Trends in Analytical Chemistry*. 2002, vol. 21, no 4, pp. 259-275.
2. Martínez-Lozano P., Fernández de la Mora J. Resolution improvements of a nano-DMA operating transonically. *J. Aerosol Sci.* 2006, vol. 37, pp. 500-512.
3. Santos J. P., E. Hontanón, E. Ramiro and M. Alonso Performance evaluation of a high-resolution parallel-plate differential mobility analyser. *Atmos. Chem. Phys.* 2009, vol. 9, pp. 2419-2429.
4. Patent SShA [US Patent] 2005/0006578; A1, Alan L. Rockwood, Edgar D. Lee, Nosa Agbonkonkon, Milton L. Lee.
5. Patent SShA [US Patent] 2006/0054804; A1, Anthony S. Wexler.
6. Patent SShA [US Patent] 7,928,374; B2, Juan Rus-Perez and Juan Fernandez de la Mora.
7. Steer B., Gorbunov B., Muir R., Ghimire A. and Rowles J. Portable Planar DMA: Development and Tests. *Aerosol Science and Technology*. 2014, vol. 48, pp. 250-259.
8. Patent SShA [US Patent] 7,372,020 B2. *Schetchik ionov* [Ion counter]; B. Gorbunov.

An ion selecting method for identification of ions in gaseous media

Vladimir N. Limanskii

Candidate of technical sciences, Docent, Siberian State University of Telecommunications and Information Sciences (SibSUTIS, Novosibirsk, Russia), v_lim@mail.ru.

Boris Z. Gorbunov

Doctor of physical and mathematical sciences, Ancon Technologies Limited, CIC (Canterbury, Kent CT2 7FG, UK), gorbunov1@aol.com.

This method relates to selecting ions with a predetermined ion mobility from its mixture containing ions and electrically neutral molecules. More particularly, this method enables to extract ions of interest from a flow of analyzed gas so that they can be directed into a clean gas flow where they can be quantified. The method provides more effective separation, increases the resolution of ion selecting device and therefore enables to quantify complex mixtures at low concentrations molecules of interest in gases.

Keywords: ion identification, ion mobility, detection of trace amounts in air.

Конструкция равновесных корректирующих кодов с синхронизацией

Е. П. Мачикина, В. Г. Насенник

В статье предлагается конструкция корректирующего блочного кода с возможностью синхронизации при пакетной последовательной передаче данных в зашумленном канале, который имеет невысокую сложность реализации кодера и декодера. Проведено исследование корректирующей способности построенных кодов в сравнении с некоторыми каскадными кодами.

Ключевые слова: равновесные коды, коды с синхронизацией, исправление ошибок, DC-free, DC-balanced, RLL.

1. Введение

При разработке электронной аппаратуры для промышленного оборудования часто возникает задача передачи по последовательному каналу связи высокоскоростных команд и данных между блоками аппаратуры в условиях сильных электромагнитных помех от расположенных рядом узлов и кабелей промышленного оборудования – сервоприводов, коллекторных двигателей, искрящих контактов реле. Для защиты от импульсных помех применяется помехоустойчивое кодирование. Для защиты от кондуктивных помех по общему проводу небольшой амплитуды применяются дифференциальные линии связи, но при кондуктивных помехах большой амплитуды необходимо применение трансформаторной или конденсаторной гальваноразвязки или использование оптических линий связи. По сравнению с медными линиями оптоволоконные линии связи требуют более дорогих трансиверов, коннекторов и оборудования для монтажа коннекторов, а также имеют ограниченное применение в буксируемых кабельных цепях, где кабели подвергаются регулярным деформациям. Для совместимости с гальваноразвязкой применяемый код не должен иметь постоянной составляющей, быть равновесным – при биполярном кодировании количество нулей и единиц в потоке должно быть равным. Кроме того, предъявляется ограничение на величину динамического разбаланса (running disparity), чтобы не происходило насыщения трансформаторов или конденсаторов, обеспечивающих гальваническую развязку – эта величина должна быть как можно меньше и как можно быстрее возвращаться к нулю. Для высокоскоростных оптических трансиверов также часто требуется, чтобы применяемый код не имел постоянной составляющей.

На выбор способа кодирования информации для таких каналов обмена данными влияет необходимость решения следующих задач:

- обеспечение синхронизации данных (битовой, байтовой, блочной и т.д.);
- исправление и обнаружение ошибок;
- обеспечение совместимости с каким-либо видом гальванической развязки.

Кроме того, для методов кодирования могут выдвигаться и дополнительные требования, связанные со сложностью реализации кодера и декодера данных при передаче информации. Известны блочные и свёрточные коды, которые одновременно обеспечивают исправление ошибок и равновесность [1, 2].

Известные в литературе методы блочного кодирования данных не обеспечивают одновременной реализации всех упомянутых требований, поэтому часто на практике применяются комбинированные (каскадные) коды. Внутренний код обычно обеспечивает синхронизацию и совместимость с гальванической развязкой, а внешний код – обнаружение и исправление ошибок, возникающих в канале.

В качестве внутреннего кода обычно используют равновесные коды, которые обеспечивают совместимость с гальванической развязкой за счёт равенства частоты появления 0 и 1 в потоке и ограничения на динамический разбаланс.

Битовая синхронизация передачи данных обычно достигается за счёт частого переключения между передаваемыми символами, т.е. ограничением максимальной длины последовательности одинаковых символов в потоке (run length limit, RLL) [3, 4]. Это свойство похоже на ограничение динамического разбаланса, но не тождественно ему. Примерами кодов, которые обеспечивают оба этих свойства, могут служить манчестерское кодирование, 4B6B, 8B10B, 64B/66B и т.д. [3, 5] Похожие требования на ограничение длины последовательности одинаковых бит возникают в магнитной или оптической записи информации.

В качестве внешнего кода можно использовать свёрточные или блочные коды с исправлением ошибок – код Хэмминга, код Голея, коды Боуза–Чоудхури–Хоквингема (БЧХ), коды Рида–Маллера, коды Рида–Соломона и др. Однако решение каждой из упомянутых проблем по отдельности увеличивает избыточность кода и заметно усложняет реализацию кодера/декодера данных.

Получающиеся кодовые конструкции можно сравнивать между собою по трём основным характеристикам:

1. относительная скорость кода (code rate), т.е. отношение количества «полезных» бит к количеству передаваемых бит. Обратная к ней величина называется избыточностью кода.
2. корректирующая способность, т.е. зависимость вероятности ошибочного приёма блока от вероятности ошибки бита в канале.
3. сложность реализации или объем аппаратных ресурсов, требуемых для реализации кода. Например, сложность можно оценивать как количество регистров (FF) и комбинаторных логических элементов (LUT) в программируемых логических матрицах (FPGA).

В этой статье предлагается конструкция корректирующего блочного кода с возможностью синхронизации, т.е. решающего одновременно все упомянутые задачи и при этом имеющего невысокую сложность реализации кодера и декодера, а также вполне конкурентоспособные характеристики относительной скорости и корректирующей способности в сравнении с каскадными кодами.

2. Конструкция кода и его свойства

Обозначим триплет (тройку битов) с одной единицей на i -том месте как $\alpha^i, i = 0, 1, 2$, т.е. $\alpha^0 = (001), \alpha^1 = (010), \alpha^2 = (001)$. Триплеты с двумя единицами, которые являются инверсиями триплетов $\alpha^i, i = 0, 1, 2$, будут обозначаться как $\bar{\alpha}^i, i = 0, 1, 2$. Например, $\bar{\alpha}^2 = (011)$.

В качестве кодовых слов используются комбинации триплетов в определенном порядке, т.е. кодовое слово строится как

$$\begin{aligned} c(i, j) &= \alpha^i \bar{\alpha}^j \alpha^{f(i, j)} \bar{\alpha}^{g(i, j)}, \bar{c}(i, j) = \bar{\alpha}^i \alpha^j \bar{\alpha}^{f(i, j)} \alpha^{g(i, j)} \\ f(i, j) &= (a_0 + a_1 \cdot i + a_2 \cdot j) \bmod 3, \\ g(i, j) &= (b_0 + b_1 \cdot i + b_2 \cdot j) \bmod 3, \end{aligned} \quad (1)$$

где $a_0, b_0 \in \{0, 1, 2\}$, $a_1, a_2, b_1, b_2 \in \{1, 2\}$, $i, j = 0, 1, 2$. Параметры $a = (a_0, a_1, a_2)$ и $b = (b_0, b_1, b_2)$ являются одинаковыми для всех кодовых слов одного кода. Код, построенный по

правилу (1), обозначим как $\text{Code}(a, b)$. Нетрудно видеть, что $|\text{Code}(a, b)| = 18$, поскольку $i, j = 0, 1, 2$ и код $\text{Code}(a, b)$ содержит инверсию каждого кодового слова. Такая мощность кода дает возможность однозначно кодировать четырёхбитовые последовательности.

Покажем, что различным параметрам $a = (a_0, a_1, a_2)$ и $b = (b_0, b_1, b_2)$ соответствуют различные коды, построенные согласно (1). Сравним кодовые слова кодов $\text{Code}(a, b)$ и $\text{Code}(a', b')$ для $a \neq a'$ и $b \neq b'$. Предположим, что для всех значений i, j кодовые слова из $\text{Code}(a, b)$ и $\text{Code}(a', b')$ совпадают. Поскольку все кодовые слова имеют одинаковое строение как комбинации триплетов, сравнивать будем «показатели» для третьего и четвертого триплетов для всех $i, j = 0, 1, 2$

$$\begin{aligned}(a_0 + a_1 \cdot i + a_2 \cdot j) \bmod 3 &= (a'_0 + a'_1 \cdot i + a'_2 \cdot j) \bmod 3, \\ (b_0 + b_1 \cdot i + b_2 \cdot j) \bmod 3 &= (b'_0 + b'_1 \cdot i + b'_2 \cdot j) \bmod 3.\end{aligned}\tag{2}$$

Очевидно, что (2) возможно, только если $a_k - a'_k \bmod 3 = 0$ и $b_k - b'_k \bmod 3 = 0$ для всех $k = 0, 1, 2$. Поскольку все параметры могут принимать только значения от 0 до 2, то $a_k = a'_k$ и $b_k = b'_k$, $k = 0, 1, 2$. Такое однозначное соответствие между набором параметров (a, b) и множеством кодов $\text{Code}(a, b)$ доказывает следующее утверждение

Утверждение 1. *Количество различных кодов $\text{Code}(a, b)$, построенных согласно (1), равно 144.*

Различные свойства построенных кодов сформулированы в утверждениях 2–5.

Утверждение 2. *Код $\text{Code}(a, b)$ является равновесным.*

Доказательство. Действительно, поскольку кодовые слова, построенные согласно (1), состоят из двух триплетов с одной единицей и из двух триплетов с двумя единицами, то каждое кодовое слово содержит ровно 6 нулей и 6 единиц. $\square$

Утверждение 3. *В последовательности кодовых слов кода $\text{Code}(a, b)$ количество последовательных 1 или 0 не больше 4.*

Доказательство. Поскольку кодовые слова, построенные согласно (1), состоят из чередующихся триплетов с одной единицей и с двумя единицами, то внутри каждого кодового слова содержится не более трех последовательных одинаковых символов. Тройки нулей или единиц получаются на стыках триплетов. Последовательности из четырех одинаковых символов могут возникать только на стыках кодовых слов. Например, когда кодовое слово с четвертым триплетом (011) стыкуется с кодовым словом, которое начинается на (110). Последовательности из пяти и более одинаковых символов в кодовой последовательности невозможны. $\square$

Утверждение 4. *Динамический разбаланс кода $\text{Code}(a, b)$ не больше 2.*

Доказательство. Рассмотрим динамический разбаланс в потоке битов. Предполагаем, что используется биполярное представление – бит 1 передается импульсом напряжения положительной полярности единичной амплитуды, а бит 0 передается импульсом напряжения отрицательной полярности такой же амплитуды. Перед началом передачи динамический разбаланс полагается равным нулю.

После передачи первого бита динамический разбаланс по своей абсолютной величине становится равным 1.

Если второй бит отличается от первого бита, то в этом случае динамический разбаланс обнуляется. Если же второй бит такой же, как и первый, то в этот момент динамический разбаланс достигает абсолютной величины 2.

Поскольку в триплете всегда два нуля и одна единица или две единицы и один нуль, то после передачи третьего бита триплета динамический разбаланс становится равен по абсолютной величине 1.

В зависимости от четвертого бита динамический разбаланс либо обнуляется, либо достигает абсолютной величины 2. После пятого бита динамический разбаланс становится равным по абсолютной величине 1.

Поскольку кодовые слова, построенные согласно (1), состоят из пар триплетов с разными инверсиями, то каждый второй триплет пары восстанавливает баланс, нарушаемый первым триплетом, т.е. после шестого бита динамический разбаланс обнуляется.

Таким образом, наибольший динамический разбаланс достигается только после второго или после четвертого бита и достигает абсолютной величины 2. Во всех остальных случаях он по абсолютной величине меньше 2. $\square$

Утверждение 5. Для кодового расстояния кода $\text{Code}(a, b)$ верно, что

$$d(\text{Code}(a, b)) = \begin{cases} 6, & (a_1 + a_2 + b_1 + b_2) \bmod 2 = 1, \\ 4, & (a_1 + a_2 + b_1 + b_2) \bmod 2 = 0. \end{cases} \quad (3)$$

Доказательство. Поскольку кодовые слова состоят из триплетов, то расстояние Хэмминга [6] между двумя кодовыми словами кода будет складываться из расстояний между соответствующими триплетами. Рассмотрим несколько возможных случаев.

Случай 1. Для двух кодовых слов $c(i, j) = \alpha^i \bar{\alpha}^j \alpha^{f(i,j)} \bar{\alpha}^{g(i,j)}$ и $c(k, l) = \alpha^k \bar{\alpha}^l \alpha^{f(k,l)} \bar{\alpha}^{g(k,l)}$ определим минимально возможное расстояние Хэмминга между ними. Обозначим $\Delta i = (i - k) \bmod 3$, $\Delta j = (j - l) \bmod 3$. Если $\Delta i \neq 0$ и $\Delta j \neq 0$, то кодовые слова $c(i, j)$ и $c(k, l)$ точно отличаются в первом и втором триплетях, а в третьем и четвертом триплетях могут совпадать, т.е. $d(c(i, j), c(k, l)) \geq 4$. Равенство будет достигаться, только если

$$\begin{aligned} f(i, j) - f(k, l) &= (a_1 \cdot \Delta i + a_2 \cdot \Delta j) \bmod 3 = 0, \\ g(i, j) - g(k, l) &= (b_1 \cdot \Delta i + b_2 \cdot \Delta j) \bmod 3 = 0. \end{aligned} \quad (4)$$

Тогда если $\Delta i = \Delta j \neq 0$, то соотношения (4) выполняются при $(a_1 + a_2) \bmod 3 = 0$, $(b_1 + b_2) \bmod 3 = 0$, т.е. a_1 и a_2 имеют разную четность, b_1 и b_2 имеют разную четность. Значит, если $(a_1 + a_2 + b_1 + b_2) - \text{четное}$, то $d(c(i, j), c(k, l)) = 4$. Аналогичное рассуждение, если $\Delta i \neq \Delta j$. Если $\Delta i = 0$ или $\Delta j = 0$, то $f(i, j) \neq f(k, l)$, $g(i, j) \neq g(k, l)$ и $d(c(i, j), c(k, l)) \geq 6$.

Случай 2. Для двух кодовых слов $\bar{c}(i, j) = \bar{\alpha}^i \alpha^j \bar{\alpha}^{f(i,j)} \alpha^{g(i,j)}$ и $\bar{c}(k, l) = \bar{\alpha}^k \alpha^l \bar{\alpha}^{f(k,l)} \alpha^{g(k,l)}$ рассуждения такие же, как и в случае 1.

Случай 3. Найдем расстояние между кодовыми словами $c(i, j) = \alpha^i \bar{\alpha}^j \alpha^{f(i,j)} \bar{\alpha}^{g(i,j)}$ и $\bar{c}(k, l) = \bar{\alpha}^k \alpha^l \bar{\alpha}^{f(k,l)} \alpha^{g(k,l)}$. Обозначим $\Delta i = (i - k) \bmod 3$, $\Delta j = (j - l) \bmod 3$. Если $\Delta i = 0$, то $f(i, j) \neq f(k, l)$, $g(i, j) \neq g(k, l)$ и кодовые слова отличаются во всех триплетях, причем $d(c(i, j), \bar{c}(k, l)) = d(\alpha^i, \bar{\alpha}^i) + d(\alpha^j, \bar{\alpha}^j) + d(\alpha^{f(i,j)}, \bar{\alpha}^{f(i,l)}) + d(\alpha^{g(i,j)}, \bar{\alpha}^{g(i,l)}) = 3 + 1 + 1 + 1 = 6$. Если $\Delta j = 0$, то так же $d(c(i, j), \bar{c}(k, j)) = 6$. Если $\Delta i \neq 0$ и $\Delta j \neq 0$, то кодовые слова будут отличаться во всех четырех триплетях как минимум в одной позиции, т.е. $d(c(i, j), \bar{c}(k, j)) \geq 4$. Для того, чтобы $d(c(i, j), \bar{c}(k, j)) \geq 6$, нужно, чтобы в кодовых словах третьи или четвертые триплеты были различны, т.е. $(a_1 \cdot \Delta i + a_2 \cdot \Delta j) \bmod 3 = 0$ или $(b_1 \cdot \Delta i + b_2 \cdot \Delta j) \bmod 3 = 0$. Рассуждения, как в случае 1, приводят к выводу, что при нечетном $(a_1 + a_2 + b_1 + b_2)$ расстояние $d(c(i, j), \bar{c}(k, l)) \geq 6$. $\square$

Таким образом, среди построенных кодов имеются 72 кода с кодовым расстоянием 6 и 72 кода с кодовым расстоянием 4. Как известно из литературы [3, 6, 7, 8], для того, чтобы исправить t ошибок в кодовом слове, код должен иметь кодовое расстояние $d \geq 2t + 1$. Кодовое расстояние 6 позволяет гарантированно исправить не более 2 ошибок замещения в кодовом слове, а кодовое расстояние 4 – не более одной ошибки. Для кода с кодовым расстоянием 6 кодовые слова, имеющие три ошибки, часто имеют расстояние Хэмминга 3 до нескольких кодовых слов, так что они не могут быть однозначно декодированы, поэтому такие ошибки могут быть обнаружены, но не исправлены. Кодовое расстояние 4 позволяет обнаруживать 2 ошибки в слове.

Полностью аналогичные рассуждения можно провести и для кодов с симметричной инверсией триплетов – $c(i, j) = \alpha^i \bar{\alpha}^j \bar{\alpha}^{f(i, j)} \alpha^{g(i, j)}$ и $\bar{c}(i, j) = \bar{\alpha}^i \alpha^j \alpha^{f(i, j)} \bar{\alpha}^{g(i, j)}$. Отличие заключается только в том, что для этих кодов иначе выглядит утверждение 3, поскольку последовательность из 4 одинаковых бит может встречаться не только на стыках кодовых слов, но и в середине кодового слова. Такая схема позволяет построить ещё 72 кода с кодовым расстоянием 6.

3. Блочная синхронизация

Для достижения блочной синхронизации в литературе предлагаются различные методы [3, 8]. Например, в коде 8B10B используются комма-коды, которыми являются 10 кодовых слов, содержащих 5 одинаковых бит подряд, при этом 5 одинаковых бит подряд не встречаются ни в каких других кодовых словах или стыках кодовых слов [5].

Математический анализ стыков кодовых слов представляется достаточно сложной задачей [2], поэтому для исследования кодов $\text{Code}(a, b)$ была написана программа на языке C, с помощью которой были проанализированы все возможные последовательности кодовых слов для каждого из построенных кодов с расстоянием 6. Анализ получившихся кодов $\text{Code}(a, b)$ с кодовым расстоянием 6 показал, что ни в одном из кодов уникальных комма-кодов нет.

Далее был проведен программный анализ комбинаций из двух кодовых слов. После исследования всех 144 кодов $\text{Code}(a, b)$ в некоторых кодах были найдены пары кодовых слов, которые никогда не могут встретиться в потоке данных при неправильной блочной синхронизации со смещением от 1 до 11 бит. Всего было выявлено 10 таких кодов – 4 кода с симметричной инверсией триплетов и 6 кодов с антисимметричной инверсией. В этих кодах найдено по 2 пары кодовых слов, которые не только не встречаются в потоке произвольных данных, но и, более того, находятся на расстоянии Хэмминга не менее 4 до ближайшей комбинации. Наихудшая ситуация достигается только при смещении на 6 бит. Таким образом, найденные пары кодовых слов пригодны как комма-коды для установления блочной синхронизации, в том числе и при наличии помех — до 1 ошибки замещения в двойном кодовом слове.

Дополнительное сравнение этих найденных 10 кодов между собой позволило заметить следующее.

Во-первых, в кодах с симметричной инверсией триплетов 4 одинаковых бита могут встретиться как на стыках кодовых слов, так и в середине, а в кодах с асимметричной инверсией триплетов 4 одинаковых бита могут встретиться только на стыках кодовых слов. Таким образом, анализ статистики встречаемости последовательностей из 4 одинаковых бит подряд может быть использован для целей синхронизации или обнаружения потери синхронизации, причём коды с асимметричной инверсией имеют некоторое преимущество перед кодами с симметричной инверсией.

Во-вторых, в 4 кодах с антисимметричной инверсией может встречаться регулярная последовательность из чередующихся троек битов 000 и 111, затрудняющая однозначную интерпретацию, что может рассматриваться как слабость.

После исключения таких кодов из рассмотрения остаются два эквивалентных кода без упомянутых слабостей с параметрами $a = (2, 2, 2)$, $b = (2, 1, 2)$ и $a = (2, 1, 1)$, $b = (2, 2, 1)$.

Обозначим найденные коды как 4B12B_1 и 4B12B_2, соотношения (1) для этих кодов будут иметь вид: $f(i, j) = (2 + 2 \cdot i + 2 \cdot j) \bmod 3$, $g(i, j) = (2 + i + 2 \cdot j) \bmod 3$ для кода 4B12B_1 и $f(i, j) = (2 + i + j) \bmod 3$, $g(i, j) = (2 + 2 \cdot i + j) \bmod 3$ для кода 4B12B_2.

Коды 4B12B_1 и 4B12B_2 приведены в табл. 1 и 2. Кодовые слова записаны в предположении, что передача битов в канал связи происходит слева направо, начиная с крайнего левого бита.

Анализ расстояния Хэмминга от всех возможных 12-битных слов до кодовых слов 4B12B

Т а б л и ц а 1. Код 4B12B_1

i	j	$f(i, j)$	$g(i, j)$	$c(i, j)$	$\bar{c}(i, j)$
0	0	2	2	001 110 100 011	110 001 011 100
0	1	1	1	001 101 010 101	110 010 101 010
0	2	0	0	001 011 001 110	110 100 110 001
1	0	1	0	010 100 010 110	101 011 101 001
1	1	0	2	010 101 001 011	101 010 110 100
1	2	2	1	010 011 100 101	101 100 011 010
2	0	0	1	100 110 001 101	011 001 110 010
2	1	2	0	100 101 100 110	011 010 011 001
2	2	1	2	100 011 010 011	011 100 101 100

Т а б л и ц а 2. Код 4B12B_2

i	j	$f(i, j)$	$g(i, j)$	$c(i, j)$	$\bar{c}(i, j)$
0	0	2	2	001 110 100 011	110 001 011 100
0	1	0	0	001 101 001 110	110 010 110 001
0	2	1	1	001 011 010 101	110 100 101 010
1	0	0	1	010 110 001 101	101 001 110 010
1	1	1	2	010 101 010 011	101 010 101 100
1	2	2	0	010 011 100 110	101 100 011 001
2	0	1	0	100 110 010 110	011 001 101 001
2	1	2	1	100 101 100 101	011 010 011 010
2	2	0	2	100 011 001 011	011 100 110 100

показывает, что некоторые слова оказываются на расстоянии Хэмминга 3 только до 1 кодового слова из 4B12B и поэтому могут быть однозначно декодированы. Таким образом, для каждого кодового слова из 4B12B можно однозначно декодировать 1 слово без ошибок, $C_{12}^1 = 12$ – слов с одной ошибкой, $C_{12}^2 = 66$ – слов с двумя ошибками и 12 из $C_{12}^3 = 220$ – слов с тремя ошибками.

Одну из найденных пар кодовых слов можно выделить в качестве комма-кодов и обозначить как K_0 и K_1 , остальные кодовые слова – $D_0, \dots, D_{15}$. Для кода 4B12B_1 в качестве K_0 и K_1 подходят кодовые слова $c(0, 0)$ и $\bar{c}(0, 2)$, а для кода 4B12B_2 – кодовые слова $c(2, 2)$ и $\bar{c}(0, 0)$.

Полученные коды 4B12B_1 и 4B12B_2 пригодны для использования при пакетной передаче сообщений. В канале связи постоянно передаются кодовые слова, при этом пакеты данных кодируются с использованием кодовых слов $D_0, \dots, D_{15}$, а межпакетные интервалы заполняются чередующимися кодовыми словами K_0 и K_1 . Кодовое слово D , встречающееся после слова K , интерпретируется как начало пакета данных.

Коды 4B12B_1 и 4B12B_2 не исправляют ошибок синхронизации (удаление и вставка битов), но для этих кодов возможно восстановление синхронизации по паре комма-кодов и обнаружение потери блочной синхронизации. Приёмник устанавливает или восстанавливает блочную синхронизацию при обнаружении в потоке пары $K_0 K_1$.

Потеря синхронизации может быть обнаружена при помощи подсчёта статистики встречаемости троек битов 000 и 111 в последовательности передаваемых бит в канале. При правильной синхронизации такие триплеты не встречаются в потоке, и приёмник может их встретить только в результате ошибок замещения, а при смещении на 1 или 2 бита такие тройки встречаются часто. Соответственно, для детектирования потери блочной синхронизации достаточно иметь всего три счётчика, в которых подсчитывать каждый случай возникновения таких триплетов – со сдвигом на 0, 1 и 2 бита. Если синхронизация правильная, то счётчик со сдвигом 0 регистрирует триплеты 000 и 111 наиболее редко, только когда их возникновение обусловлено ошибками. Если же счётчик со сдвигом 0 опережает два других счётчика на некую достаточно большую величину, это можно интерпретировать как признак потери блочной синхронизации. Авторы экспериментально подобрали величину 30. Уменьшение этой величины приводит к увеличению риска ложного обнаружения потери блочной синхронизации, увеличение же приводит к неоправданно затянутому принятию решения о потере блочной синхронизации.

4. Сравнение с каскадными кодами

Как было указано во введении, кодовые конструкции можно сравнивать между собой по относительной скорости, по корректирующей способности и по сложности реализации. Для сравнительного анализа построенного кода 4B12B были выбраны каскадные коды БЧХ(15, 7) + манчестер и RS(15, 11) + манчестер. В этих кодах в качестве внутреннего кода для обеспечения равновесности и RLL используется манчестерское кодирование, а в качестве внешнего кода для коррекции ошибок — широко известные и подробно исследованные коды БЧХ(15, 7) и RS(15, 11). Коды БЧХ(15, 7) + манчестер и RS(15, 11) + манчестер имеют близкие значения относительной скорости – 0.367 и 0.233, в то время как код 4B12B – 0.333. Однако хотя коды БЧХ(15, 7) и RS(15, 11) и обладают близкими сопоставимыми характеристиками по корректирующей способности и сложности реализации, они вообще не имеют свойства блочной синхронизации, так что сравнение, очевидно, будет не вполне корректное.

Оценим объем аппаратных ресурсов, необходимых для реализации кодирования и декодирования кодом 4B12B. Благодаря малой мощности кода 4B12B реализация процедур кодирования и декодирования сообщений не представляет затруднений и требует пренебрежимо малых ресурсов FPGA. Для реализации табличного декодирования хорошо подходят блоки BRAM, входящие в состав FPGA. Декодер, реализованный с использованием BRAM, требует всего 2 блока BRAM16, входящих в состав FPGA Spartan 6, или 1 блок BRAM36, входящий в состав более поздних семейств FPGA.

Полностью параллельный декодер кода 4B12B требует 94 LUT FPGA Spartan 6 при декодировании только до 2 ошибок на слово или 183 LUT, если включить декодирование ещё и слов с 3 ошибками, которые можно однозначно декодировать к ближайшему кодовому слову (здесь и далее объем фактически затраченных аппаратных ресурсов приведен по результатам компиляции). При реализации существенным образом использовалось то обстоятельство, что полноценные сумматоры здесь не обязательны, а вычисление расстояния Хэмминга можно осуществлять только в рамках множества {0, 1, 2, много}, при этом значительно сокращается количество требующихся ресурсов FPGA.

Последовательно-параллельный декодер кода 4B12B требует 73 FF и 108 LUT и строится на использовании того факта, что инверсия кодового слова тоже является кодовым словом, поэтому можно хранить только одно слово из пары. Аналогичное свойство проявляется в отношении 12 слов с тремя ошибками, которые допускают однозначное декодирование. Это позволило вдвое сократить количество хранимых данных. Таким образом, если данные поступают последовательно по 1 биту за такт, то за 9 тактов можно успеть сопоставить принятое слово со всеми

кодowymi словами по 2 кодовых слова за такт, и декодирование осуществляется со скоростью поступления данных.

Сравним теперь коды по корректирующей способности. Для двоичного симметричного канала $[3, 6, 8]$ с вероятностью ошибки бита p , $0 \leq p \leq 1$ вероятность успешного декодирования слова для кода 4B12B составляет:

$$s(p) = C_{12}^0 \cdot (1-p)^{12} + C_{12}^1 \cdot p \cdot (1-p)^{11} + C_{12}^2 \cdot p^2 \cdot (1-p)^{10} + 12 \cdot p^3 \cdot (1-p)^9, \quad (5)$$

а вероятность ошибочного декодирования слова (word error rate):

$$\text{WER}(p) = 1 - s(p). \quad (6)$$

Для проведения сравнения эффективности кода 4B12B с каскадными кодами, построенными из кода БЧХ(15, 7) поверх манчестерского и кода Рида–Соломона (15, 11) поверх манчестерского, была написана программа на языке С, результаты которой представлены на графике (рис. 1). Для каждого кода программно эмулировалась передача 10^8 кодовых слов по каналу связи с заданной вероятностью ошибки бита $0 \leq p \leq 1$, а значение $\text{WER}(p)$ оценивалось как отношение количества неправильно декодированных слов к общему количеству переданных. Результат вычислительного эксперимента для кода 4B12B показал прекрасное совпадение с теоретически полученным выражением для $\text{WER}(p)$.

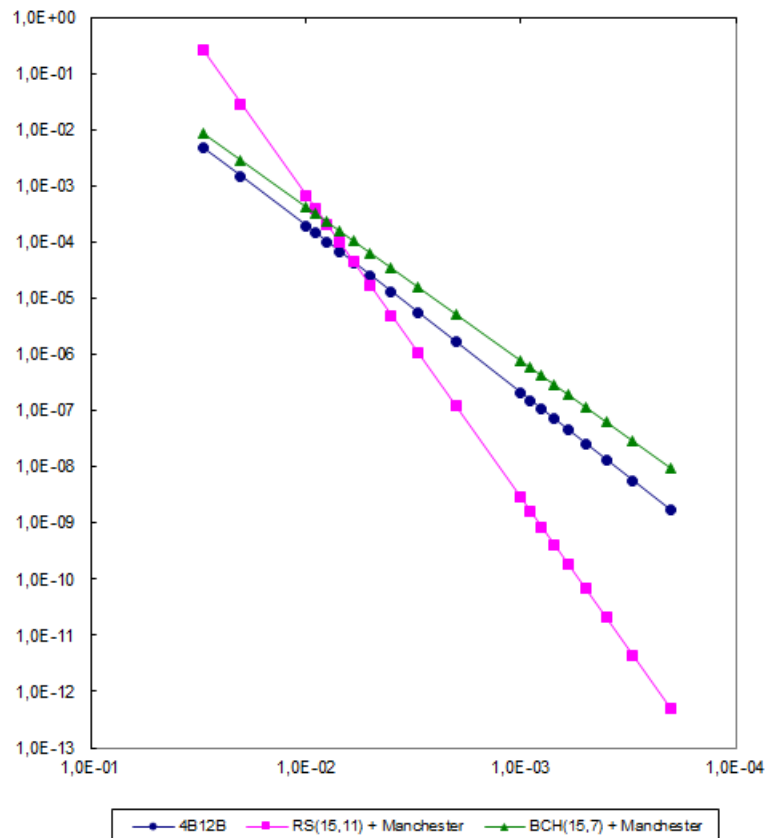


Рис. 1. График вероятности ошибочного приёма слова в зависимости от вероятности ошибки бита в канале.

Видно, что код RS(15,11)+манчестер за счёт значительно большей длины блока (120 бит) имеет преимущество перед кодом 4B12B при вероятностях ошибки менее 0.005 и даже имеет чуть более высокую относительную скорость: 0.367 против 0.333 у 4B12B, но при этом декодер

кода RS(15, 11) требует гораздо больших вычислительных ресурсов, чем декодер 4B12B. Декодер RS(15, 11) со стираниями версии 7.1 из Xilinx ISE 14.7 требует 310 FF и 269 LUT, при этом имеет задержку обработки, равную 45 тактов, так что для обработки непрерывного потока потребуется использовать три таких декодера или втрое ограничивать скорость передачи данных по сравнению с максимально возможной для приёма и обработки в FPGA.

По сравнению с кодом БЧХ(15, 7) + манчестер код 4B12B имеет некоторое преимущество (примерно в 2 раза) в вероятности ошибочного приёма слова и даже чуть более высокую относительную скорость: 0.333 против 0.233 у БЧХ(15, 7) + манчестер.

Литература

1. Deng R.H., Herro M.A. DC-free coset codes // *IEEE Transactions on Information theory*. 1988. V.34. P. 786–792.
2. Wadayama T., Han Vinck A. J. DC-Free Binary Convolutional Coding // *IEEE Transactions on Information theory*. 2002. V.48, №1, P. 162–172.
3. Скляр Б. Цифровая связь. Теоретические основы и практическое применение. Второе издание, исправленное. М.: Вильямс, 2003.
4. Mercier H., Bhargava V. K., Tarokh V. A survey of error-correcting codes for channels with symbol synchronization errors // *IEEE Communications Surveys & Tutorials*. 2010. V.12, №1, P. 87–96.
5. Widmer A. X., Franaszek P. A. A DC-Balanced, Partitioned-Block, 8B/10B Transmission Code // *IBM Journal of Research and Development*. 1983. V.27, №5, P. 440–451.
6. Хэмминг Р. В. Теория кодирования и теория информации. Перевод с английского С. И. Гельфанда под редакцией Б. С. Цыбакова. М.: Радио и связь, 1983.
7. Левенштейн В. И. Оценки для кодов, обеспечивающих исправление ошибок и синхронизацию // *Проблемы передачи информации*. 1969. Т. 5, В. 2, с. 3–13.
8. Питерсон У., Уэлдон Э. Коды, исправляющие ошибки. Перевод с английского под редакцией Р. Л. Добрушина и С. И. Самойленко. М.: МИР, 1976.

Статья поступила в редакцию 20.12.2021;
переработанный вариант – 28.02.2022.

Мачикина Елена Павловна

к.ф.-м.н., доцент; доцент кафедры прикладной математики и кибернетики СибГУТИ, (630102, Новосибирск, ул. Кирова, 86), тел. (383) 269-82-72, e-mail: machikina@sibguti.ru.

Насенник Виталий Геннадьевич

e-mail: vitaly.nasennik@gmail.com, ORCID: 0000-0002-7654-6953.

References

1. Deng R. H., Herro M. A. DC-free coset codes. *IEEE Transactions on Information Theory*. 1988, vol. 34, pp. 786-792.
2. Wadayama T., Han Vinck A. J. DC-Free Binary Convolutional Coding. *IEEE Transactions on Information Theory*. 2002, vol. 48, No. 1, P. 162-172.
3. Skljar B. *Cifrovaja svjaz'. Teoreticheskie osnovy i prakticheskoe primeneniye*. [Digital communication. Theoretical foundations and practical application]. Second edition, revised. Moscow: Williams, 2003.
4. Mercier H., Bhargava V. K., Tarokh V. A survey of error-correcting codes for channels with symbol synchronization errors. *IEEE Communications Surveys & Tutorials*. 2010, vol. 12, no. 1. pp. 87-96.

5. Widmer A. X., Franaszek P. A. A DC-Balanced, Partitioned-Block, 8B/10B Transmission Code. *IBM Journal of Research and Development*. 1983, vol. 27, no. 5, pp. 440–451.
6. Hjemming R. V. *Teorija kodirovanija i teorija informacii. Perevod s anglijskogo S. I. Gel'fanda pod redakciej B. S. Cybakova* [Coding theory and information theory. Translation from English]. Moscow: Radio and communication, 1983.
7. Levenshtejn V. I. *Ocenki dlja kodov, obespechivajushhih ispravlenie oshibok i sinhronizaciju* [Estimates for codes providing error correction and synchronization]. *Problemy peredachi informacii* [Problems of Information Transmission]. 1969, vol. 5, no. 2. pp. 3-13.
8. Piterson U., Ujeldon Je. *Kody, ispravljajushhie oshibki. Perevod s anglijskogo pod redakciej R. L. Dobrushina i S. I. Samojlenko* [Error-correcting codes. Translation from English]. Moscow: Mir, 1976.

A Construction of Synchronizable DC-balanced Error-Correcting Codes

Elena P. Machikina

Candidate of physical and mathematical sciences, Siberian State University of Telecommunications and Information Sciences (SibSUTIS, Novosibirsk, Russia), machikina@sibgti.ru.

Vitaly G. Nasennik

ORCID: 0000-0002-7654-6953, vitaly.nasennik@gmail.com.

The paper proposes a construction of error-correcting block codes with synchronization in packet serial data transmission over a noisy channel, which has a low complexity of encoder and decoder implementations. A study of corrective abilities of constructed codes in comparison with some concatenated codes is presented.

Keywords: DC-free codes, DC-balanced codes, synchronizable codes, forward error correction, RLL codes.

Подавление помех судоходства при гидроакустическом поиске подводного объекта

В. Г. Арсентьев, Г. И. Криволапов

Рассматривается способ подавления помех мешающего судоходства при гидроакустическом поиске с надводного судна автономного необитаемого подводного объекта, находящегося на дне известной морской акватории.

Ключевые слова: поисковое судно, навигационный сигнал, когерентная компенсация сосредоточенных помех.

1. Введение

В публикациях [1], [2] рассмотрен способ скрытного гидроакустического поиска автономного необитаемого подводного объекта, в основе которого лежат следующие основополагающие принципы:

- для передачи сигналов управления в гидроакустическом канале между поисковым судном и подводным объектом используются импульсные фрагменты сонограмм записей звуков морских животных, рыб и ракообразных с заданными амплитудными, временными и частотными параметрами, схожими по акустическому восприятию с возможными естественными шумами морской среды известной акватории;

- гармоническая навигационная посылка сигнала маяка-пингера подводного объекта в процессе поиска маскируется квазибелым шумом, близким по акустическому восприятию к возможному фоновому шуму мелкого или глубокого моря, а её обработка на поисковом судне осуществляется с использованием процедуры синфазирования, позволяющей позиционировать подводный объект относительно поискового судна с необходимой точностью в условиях малых отношений сигнал/маскирующий шум;

- информационный обмен между объектами взаимодействия «поисковое судно – подводный объект» организуется на основе асинхронного режима приёма-передачи сеансовых сигналов поискового судна и подводного объекта с необходимым различением в условиях преимущественно флуктуационных шумов морской среды акватории поиска.

В процессе гидроакустического поиска подводного объекта в пределах акватории поиска возможно появление различных надводных судов, проходящих мимо поискового судна на различных расстояниях и создающих помехи приёму навигационного сигнала маяка-пингера подводного объекта на поисковом судне, возникающие за счёт отражения шумовых гидроакустических сигналов надводных судов от морского дна.

На рис. 1 представлен пример возникновения двух помех от надводных судов при гидроакустическом поиске автономного подводного аппарата, где обозначены: 1 – надводное поисковое судно; 2 – автономный подводный аппарат; 3 – пеленгационная антенна поискового судна; 4 – илистое морское дно; 5, 6 – проходящие надводные суда; 7 – скальные фрагменты морского дна.

Военные корабли и гражданские суда, курсирующие в различных акваториях мирового океана, являются весьма сильными источниками помех в процессе рассматриваемого гидроакустического поиска подводных объектов. Создаваемые этими объектами шумы можно разделить на три группы: шумы машин и вспомогательных механизмов, шумы, связанные с работой гребных винтов, и гидродинамические шумы. На рис. 2 показаны наиболее типичные спектры шумов различных надводных судов, взятые из источников [3].

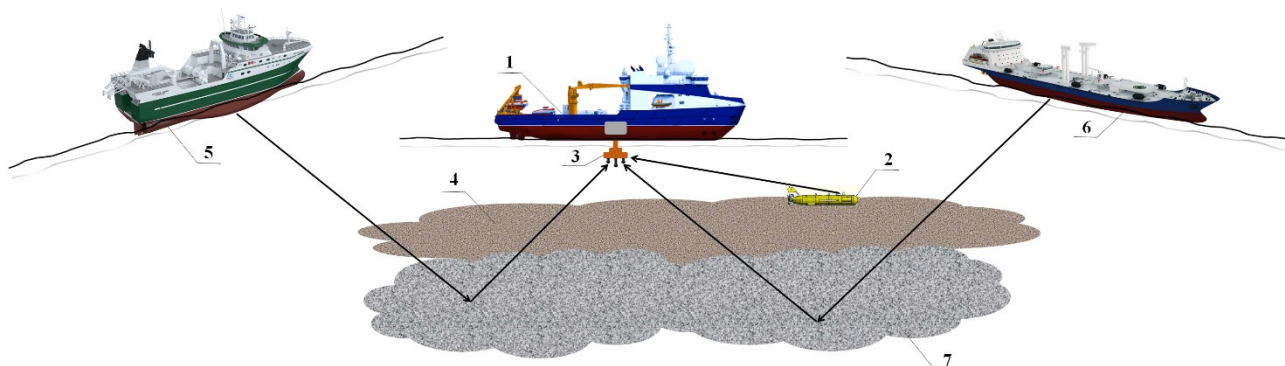


Рис. 1. Возникновение помех от проходящих судов при гидроакустическом поиске автономного подводного аппарата

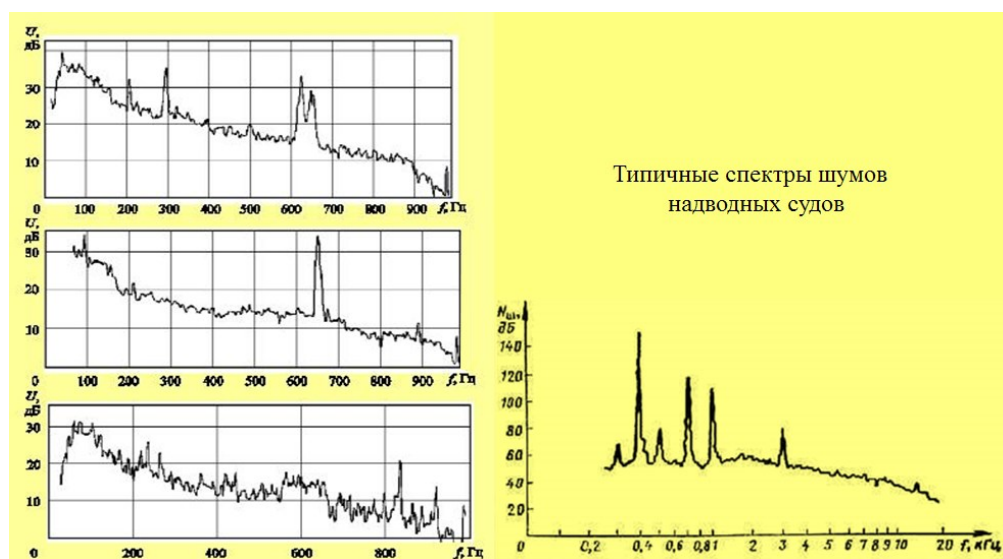


Рис. 2. Спектры шумов надводных судов

Шумовые помехи надводных судов согласно имеющейся статистике можно отнести к категории сосредоточенных (узкополосных) помех в области частот от десятков герц до одного килогерца с затуханием акустических колебаний при распространении в морской среде, позволяющим иметь уровни таких помех в местах возможного приёма навигационных сигналов при поиске подводных объектов, превышающие уровень флуктуационных шумов морской среды на $80 \div 100$ дБ. При математической оценке эффективности подавления сосредоточенных помех в различных радиотехнических системах часто используется модель помехи в виде узкополосного случайного процесса с изменяющимися амплитудой и фазой колебаний заданной средней частоты её энергетического спектра [4].

Рекомендуемые в способе [1] скрытного гидроакустического поиска подводного объекта частотно-временные параметры сеансового сигнала, излучаемого маяком-пингером, представлены на рис. 3. Указанный сигнал общей длительностью не более 2.22 с излучается в процессе гидроакустического поиска с периодичностью не менее 8 с, то есть обладает достаточно большой скважностью – временным интервалом, на котором могут присутствовать только помехи: флуктуационные шумы морской среды и помехи судоходства, но отсутствует сигнал маяка-пингера подводного объекта.

Подавление помех судоходства базируется на принципе пространственно-временной обработки [5] сигнала маяка-пингера с когерентной компенсацией помех в парциальных трактах приёма пеленгационной антенны поискового судна, использующем пространственные, временные и частотные различия навигационного сигнала и помех судоходства.

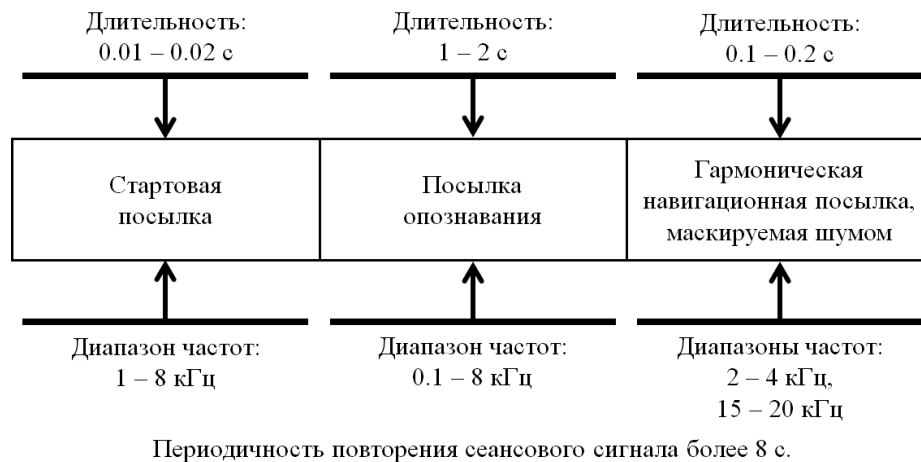


Рис. 3. Частотно-временная структура сигнала маяка-пингера подводного объекта

2. Обработка навигационного сигнала на поисковом судне

Обработка навигационного сигнала маяка-пингера подводного объекта на поисковом судне в процессе скрытного гидроакустического поиска предполагает выполнение определённого порядка действий, процедур и операций, которые для наглядности восприятия функциональных связей представлены на рис. 4.

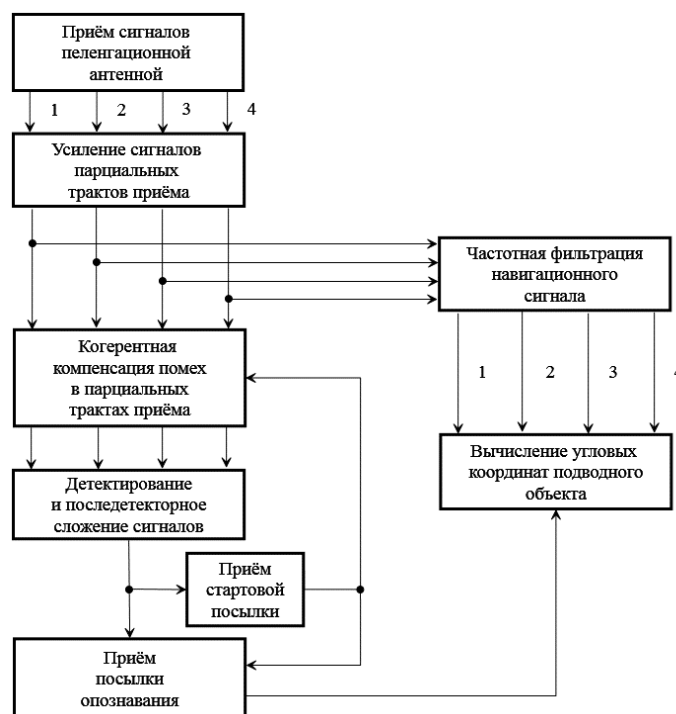


Рис. 4. Обработка сигнала маяка-пингера на поисковом судне

1. Гидроакустические сигналы водной среды акватории поиска принимаются пеленгационной антенной диаметрально-ортогональной геометрии расположения четырёх приёмных гидрофонов. Такими сигналами являются продолжительные по времени флуктуационные шумы морской среды, помехи судоходства и кратковременный сигнал маяка-пингера.

2. Принятые гидрофонами пеленгационной антенны сигналы парциальных трактов приёма усиливаются (с необходимой частотной фильтрацией) до уровня, требующегося для их дальнейшей обработки.

3. Осуществляется адаптивная когерентная компенсация [6] помех судоходства в парциальных трактах приёма пеленгационной антенны с выполнением следующих действий.

3.1. Возможный частотный диапазон помех судоходства разделяется на три поддиапазона (по максимальному количеству помех, которые можно подавить с использованием четырёхэлементной пеленгационной антенны) для частотной фильтрации помех разных надводных источников, рассредоточенных по акватории поиска, спектры частот которых попадают в соответствующие поддиапазоны.

На рис. 5 схематично показано соотношение диапазонов частот широкополосной посылки опознавания сигнала маяка-пингера и узкополосных помех судоходства, где обозначены: ΔF_s – диапазон частот посылки опознавания сигнала маяка-пингера; ΔF_p – возможный диапазон частот помех судоходства; $\Delta F_1, \Delta F_2, \Delta F_3$ – выбранные для частотной фильтрации поддиапазоны частот помех.

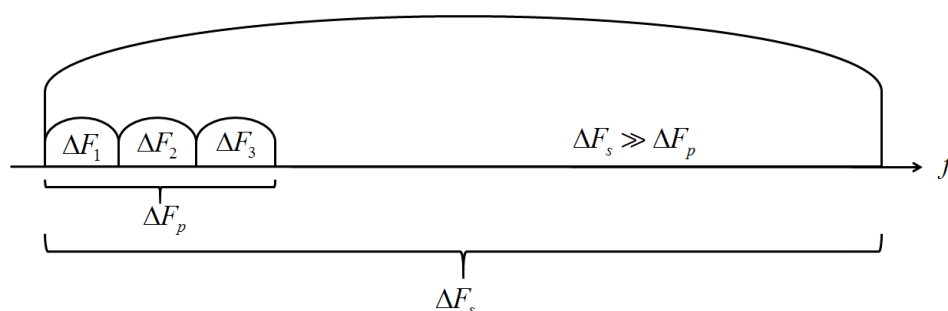


Рис. 5. Диапазоны частот посылки опознавания сигнала маяка-пингера и помех судоходства

В соответствии с частотными параметрами сигнала, излучаемого маяком-пингером подводного объекта, и типичными спектрами шумов надводных судов, представленными на рис. 3 и 2 соответственно, а также предпочтительной схмотехнической реализацией октавных полосовых фильтров низкочастотного диапазона возможен выбор следующих величин диапазона частот посылки опознавания сигнала маяка-пингера и поддиапазонов частот помех судоходства: $\Delta F_s = 0.1 \dots 8$ кГц, $\Delta F_1 = 0.1 \dots 0.2$ кГц, $\Delta F_2 = 0.2 \dots 0.4$ кГц, $\Delta F_3 = 0.4 \dots 0.8$ кГц.

3.2. Из четырёх парциальных трактов приёма пеленгационной антенны организуются четыре группы каналов обработки сигналов из четырёх каналов приёма в каждой группе, по принципу «три для одного» – для подавления помех в одном, основном, канале приёма группы используют три соответствующих вспомогательных канала приёма.

3.3. В каждой группе каналов обработки сигналов во время отсутствия излучения коротковременно повторяющегося в процессе гидроакустического поиска сигнала маяка-пингера осуществляется когерентная компенсация колебаний помех основного канала приёма взвешенными колебаниями помех вспомогательных каналов приёма с использованием операции адаптивного комплексного взвешивания, при этом сигналы каждого вспомогательного канала приёма подвергаются частотной фильтрации в соответствующем поддиапазоне возможного частотного диапазона помех судоходства.

4. После когерентной компенсации помех судоходства выполняется линейное амплитудное детектирование и последетекторное сложение сигналов парциальных трактов приёма пеленгационной антенны с целью обеспечения достаточного для достоверного приёма посылки опознавания сигнала маяка-пингера отношения сигнал/шум.

5. С появлением излучения маяка-пингера осуществляется последетекторный приём и обработка стартовой посылки навигационного сигнала.

6. После приёма и обработки стартовой посылки навигационного сигнала приостанавливается на известное время, соответствующее суммарной длительности посылки опознавания

и навигационной посылки сигнала маяка-пингера, выполнение операции адаптивного комплексного взвешивания с сохранением ранее полученных коэффициентов комплексного взвешивания, что позволяет избежать подавления в результате когерентной компенсации принимаемой посылки опознавания сигнала маяка-пингера.

7. Выполняется приём и обработка посылки опознавания сигнала маяка-пингера.

8. После вынесения решения о достоверном приёме посылки опознавания сигнала маяка-пингера реализуется процедура определения угловых координат подводного объекта относительно поискового судна с предварительной полосовой фильтрацией сигнала маяка-пингера в полосе частот его замаскированной шумом гармонической навигационной посылки и подавлением вне полосы фильтрации сосредоточенных помех судоходства.

3. Когерентная компенсация сосредоточенных помех

Когерентная компенсация помех в парциальных трактах приёма пеленгационной антенны с группировкой по принципу «три для одного» может быть представлена укрупнёнными функциональными операциями, показанными на рис. 6, где обозначены: 1 – усиление (с частотной фильтрацией в диапазоне частот ΔF_s) сигналов парциальных трактов приёма пеленгационной антенны; 2, 3, 4 – частотная фильтрация колебаний помех в поддиапазонах частот $\Delta F_1, \Delta F_2, \Delta F_3$; 5 – формирование коэффициентов комплексного взвешивания для когерентной компенсации колебаний помех в соответствующем парциальном тракте приёма, 6 – комплексное взвешивание колебаний помех соответствующих парциальных трактов приёма трёх частотных поддиапазонов, 7 – вычитание из колебаний помех соответствующего парциального тракта приёма взвешенных колебаний помех трёх частотных поддиапазонов возможного частотного диапазона помех судоходства.

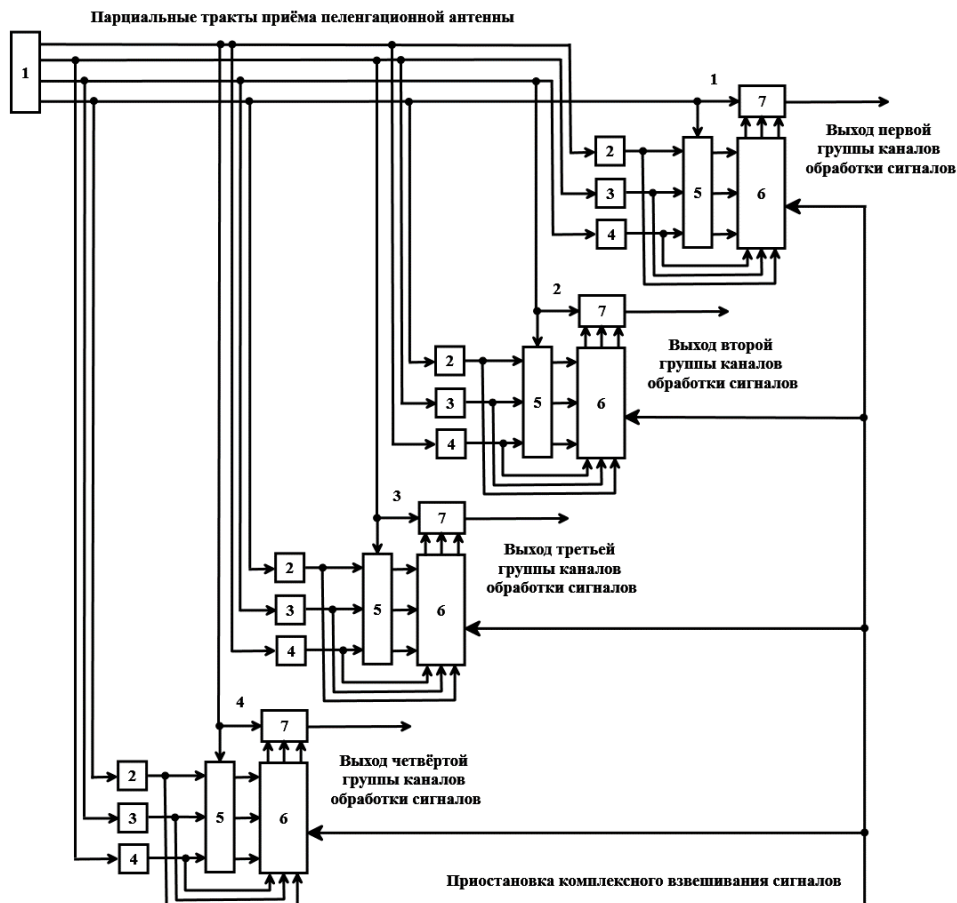


Рис. 6. Функциональные операции когерентной компенсации трёх узкополосных помех

Пример технической реализации группы каналов обработки сигналов для осуществления адаптивной когерентной компенсации помех мешающего судоходства в каждом парциальном тракте приёма пеленгационной антенны поискового судна показан на рис. 7, где обозначены: $\Phi 1, \Phi 2, \Phi 3$ – фильтры с полосами частот $\Delta F_1, \Delta F_2, \Delta F_3$ соответственно; КП – квадратурные преобразователи; П1 – умножители; П2 – умножители с предельным ограничением умножаемых сигналов; И – управляемые интеграторы; С – сумматор; Инв – инвертор.

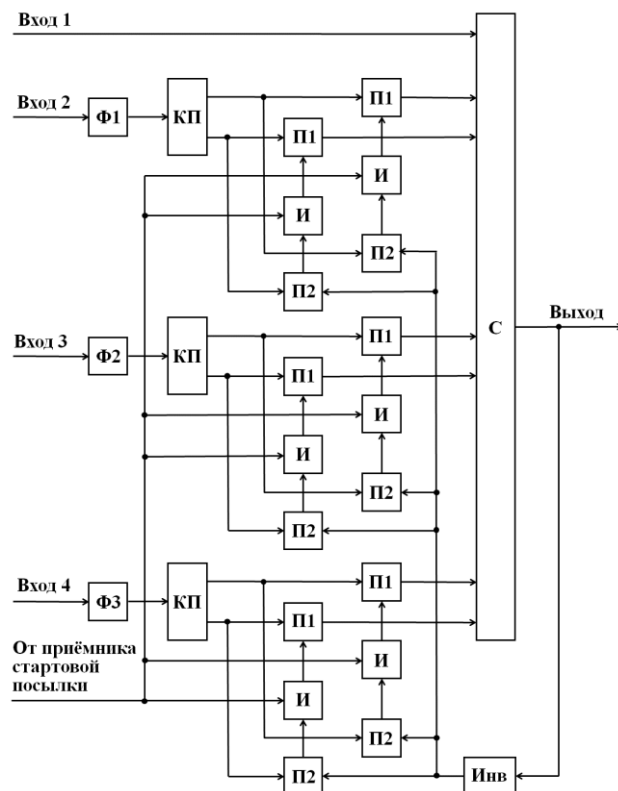


Рис. 7. Техническая реализация группы каналов обработки сигналов

Представленная на рис. 7 структура функциональных связей элементов группы каналов обработки сигналов реализует адаптивную когерентную компенсацию сигналов, поступающих на Вход 1, взвешенными сигналами, поступающими на Входы 2, 3, 4, на основе метода наискорейшего градиентного спуска [6]. Для осуществления когерентной компенсации помех в трактах приёма пеленгационной антенны поискового судна организуются подключения Входов 1÷4 каждой из четырёх групп каналов обработки сигналов к соответствующим трактам приёма пеленгационной антенны. Порядок таких подключений приведён в табл. 1.

Таблица 1. Подключения каналов обработки сигналов

Группа каналов обработки сигналов	Вход группы каналов обработки сигналов	Парциальный тракт приёма пеленгационной антенны
Первая	Вход 1	1
	Вход 2	2
	Вход 3	3
	Вход 4	4
Вторая	Вход 1	2
	Вход 2	1
	Вход 3	3
	Вход 4	4
Третья	Вход 1	3
	Вход 2	1
	Вход 3	2
	Вход 4	4

Четвёртая	Вход 1	4
	Вход 2	1
	Вход 3	2
	Вход 4	3

4. Энергетическая эффективность когерентной компенсации помех

При математическом описании операций когерентной компенсации узкополосных помех возникает закономерный вопрос выбора формы представления помех судосудства (далее – помех) и сигнала посылки опознавания маяка-пингера подводного объекта (далее – сигнала), который необходимо достоверно принимать на поисковом судне в условиях воздействия помех судосудства и флуктуационного шума морской среды.

Форма представления сигнала и помех должна способствовать уменьшению громоздкости используемых математических преобразований с целью исключения возможных ошибок при вычислениях, в необходимой мере отражать параметры реальных сигналов и помех, имеющих место в процессе гидроакустического поиска, и не приводить к недопустимому искажению конечных результатов из-за введённых математических допущений.

С учётом необходимой корректности математических преобразований и физических особенностей гидроакустического канала связи введём следующие допущения:

- сигнал и помехи в четырёх парциальных трактах приёма пеленгационной антенны представляются комплексными числами в показательной форме, при этом модуль комплексного числа интерпретируется как амплитуда действительного сигнала, а мнимый аргумент – как его фаза;

- уровни принимаемых сигнала, помех и флуктуационного шума морской среды в каждом парциальном тракте приёма пеленгационной антенны принимаются соответственно одинаковыми, это обусловлено малым пространственным разнесением приёмных гидрофонов пеленгационной антенны;

- флуктуационные шумы морской среды в парциальных трактах приёма пеленгационной антенны считаются некоррелированными, этот физический аспект образования шумов морской среды для используемого диапазона частот подтверждён экспериментальными исследованиями;

- корреляционные связи между сигналом и помехами в парциальных трактах приёма пеленгационной антенны отсутствуют вследствие независимости порождающих их источников.

Запишем сигнально-помеховые процессы в первом (основном) парциальном тракте приёма пеленгационной антенны поискового судна в следующем, пригодном для дальнейших преобразований, виде:

$$z_1 = X e^{i\psi_1} + Y_1 e^{i\phi_1} + Y_2 e^{i\xi_1} + Y_3 e^{i\phi} + \sqrt{\sigma_\xi^2}, \quad (1)$$

где X, Y_1, Y_2, Y_3 – амплитуды сигнала и трёх помех соответственно; $\psi_1, \phi_1, \xi_1, \phi$ – фазы сигнала и трёх помех, обусловленные пространственным разнесением источников сигнала и помех, а также гидрофонов пеленгационной антенны поискового судна; σ_ξ^2 – дисперсия флуктуационного шума морской среды.

Сигнально-помеховые процессы во вспомогательных парциальных трактах приёма (втором, третьем и четвёртом) представим соответственно:

$$\begin{aligned} z_2 &= X_1 e^{i\psi_{21}} + Y_1 e^{i\phi_2} + \sqrt{\gamma_1 \sigma_\xi^2}; \quad z_3 = X_2 e^{i\psi_{32}} + Y_2 e^{i\xi_3} + \sqrt{\gamma_2 \sigma_\xi^2}; \\ z_4 &= X_3 e^{i\psi_{43}} + Y_3 e^{i\phi_4} + \sqrt{\gamma_3 \sigma_\xi^2}; \quad \gamma_1 = \frac{\Delta F_1}{\Delta F_s}; \quad \gamma_2 = \frac{\Delta F_2}{\Delta F_s}; \quad \gamma_3 = \frac{\Delta F_3}{\Delta F_s}, \end{aligned} \quad (2)$$

где ΔF_s – полоса частот сигнала; X_1, X_2, X_3 – амплитуды сигнала после частотной фильтрации в поддиапазонах частот $\Delta F_1, \Delta F_2, \Delta F_3$ соответственно; ψ_{21} – фаза сигнала второго парциального тракта приёма после частотной фильтрации в поддиапазоне частот ΔF_1 ; ψ_{32} – фаза

сигнала третьего парциального тракта приёма после частотной фильтрации в поддиапазоне частот ΔF_2 ; ψ_{43} – фаза сигнала четвёртого парциального тракта приёма после частотной фильтрации в поддиапазоне частот ΔF_3 ; φ_2 – фаза первой помехи второго парциального тракта приёма после частотной фильтрации в поддиапазоне частот ΔF_1 ; ξ_3 – фаза второй помехи третьего парциального тракта приёма после частотной фильтрации в поддиапазоне частот ΔF_2 ; ϕ_4 – фаза третьей помехи четвёртого парциального тракта приёма после частотной фильтрации в поддиапазоне частот ΔF_3 .

Диагональная квадратная матрица $\mathbf{A}$ корреляционных связей колебаний помех (трёх некоррелированных помех судоходства и флуктуационного шума морской среды) второго, третьего и четвёртого парциальных трактов приёма пеленгационной антенны имеет вид:

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} = \begin{pmatrix} Y_1^2 + \gamma_1 \sigma_\xi^2 & 0 & 0 \\ 0 & Y_2^2 + \gamma_2 \sigma_\xi^2 & 0 \\ 0 & 0 & Y_3^2 + \gamma_3 \sigma_\xi^2 \end{pmatrix}. \quad (3)$$

Вектор $\mathbf{S}$ корреляционных связей колебаний помех первого и второго, первого и третьего, первого и четвёртого парциальных трактов приёма пеленгационной антенны запишется соответственно:

$$\mathbf{S} = \begin{pmatrix} S_1 \\ S_2 \\ S_3 \end{pmatrix} = \begin{pmatrix} Y_1^2 e^{i(\varphi_1 - \varphi_2)} \\ Y_2^2 e^{i(\xi_1 - \xi_3)} \\ Y_3^2 e^{i(\phi_1 - \phi_4)} \end{pmatrix}. \quad (4)$$

Определяем элементы обратной матрицы $\mathbf{A}^{-1}$ корреляционных связей колебаний помех второго, третьего и четвёртого парциальных трактов приёма пеленгационной антенны:

$$\mathbf{A}^{-1} = \begin{pmatrix} \frac{1}{Y_1^2 + \gamma_1 \sigma_\xi^2} & 0 & 0 \\ 0 & \frac{1}{Y_2^2 + \gamma_2 \sigma_\xi^2} & 0 \\ 0 & 0 & \frac{1}{Y_3^2 + \gamma_3 \sigma_\xi^2} \end{pmatrix}. \quad (5)$$

Обозначим отношения помеха/флуктуационный шум в парциальных трактах приёма пеленгационной антенны:

$$\beta_1 = \frac{Y_1^2}{\sigma_\xi^2}; \quad \beta_2 = \frac{Y_2^2}{\sigma_\xi^2}; \quad \beta_3 = \frac{Y_3^2}{\sigma_\xi^2},$$

тогда коэффициенты комплексного взвешивания для когерентной компенсации колебаний помех в основном парциальном тракте приёма взвешенными колебаниями помех вспомогательных (второго, третьего и четвёртого) парциальных трактов приёма, которые формируются в группе каналов обработки сигналов (см. рис. 7) после завершения процесса адаптации по методу наискорейшего градиентного спуска, находим на основе известного в теории адаптивной обработки сигналов матричного уравнения Винера–Хопфа [5] из соотношений:

$$\begin{aligned}
W_{12}^{\bullet} &= A_{11}^{-1}S_1 + A_{12}^{-1}S_2 + A_{13}^{-1}S_3; \\
W_{13}^{\bullet} &= A_{21}^{-1}S_1 + A_{22}^{-1}S_2 + A_{23}^{-1}S_3; \\
W_{14}^{\bullet} &= A_{31}^{-1}S_1 + A_{32}^{-1}S_2 + A_{33}^{-1}S_3,
\end{aligned} \tag{6}$$

которые с учётом зависимостей (4) и (5) преобразуются к виду

$$\mathbf{W} = \begin{pmatrix} W_{12}^{\bullet} \\ W_{13}^{\bullet} \\ W_{14}^{\bullet} \end{pmatrix} = \begin{pmatrix} \frac{\beta_1}{\gamma_1 + \beta_1} e^{i(\varphi_1 - \varphi_2)} \\ \frac{\beta_2}{\gamma_2 + \beta_2} e^{i(\xi_1 - \xi_3)} \\ \frac{\beta_3}{\gamma_3 + \beta_3} e^{i(\phi_1 - \phi_4)} \end{pmatrix}. \tag{7}$$

Уровень колебаний помех для их когерентной компенсации в основном тракте приёма определяется суммой взвешенных колебаний трёх вспомогательных трактов приёма:

$$\begin{aligned}
y_{\Sigma} &= Y_1 e^{i\varphi_2} \frac{\beta_1}{\gamma_1 + \beta_1} e^{i(\varphi_1 - \varphi_2)} + Y_2 e^{i\xi_3} \frac{\beta_2}{\gamma_2 + \beta_2} e^{i(\xi_1 - \xi_3)} + Y_3 e^{i\phi_4} \frac{\beta_3}{\gamma_3 + \beta_3} e^{i(\phi_1 - \phi_4)} = \\
&= \frac{Y_1 \beta_1}{\gamma_1 + \beta_1} e^{i\varphi_1} + \frac{Y_2 \beta_2}{\gamma_2 + \beta_2} e^{i\xi_1} + \frac{Y_3 \beta_3}{\gamma_3 + \beta_3} e^{i\phi_1}.
\end{aligned} \tag{8}$$

Мощность некомпенсированного остатка помех основного тракта приёма пеленгационной антенны в результате выполнения когерентной компенсации составляет:

$$\Delta P_p = y_1 y_1^* - 2 y_1 y_{\Sigma}^* + y_{\Sigma} y_{\Sigma}^* = \sum_{j=1}^3 Y_j^2 \left(1 - \frac{\beta_j}{\gamma_j + \beta_j} \right)^2, \tag{9}$$

где $y_1 = Y_1 e^{i\varphi_1} + Y_2 e^{i\xi_1} + Y_3 e^{i\phi_1}$ – помехи основного парциального тракта приёма; * – знак комплексного сопряжения.

Возросшая в результате выполнения когерентной компенсации дисперсия флуктуационного шума морской среды основного парциального тракта приёма пеленгационной антенны определяется соотношением:

$$\Delta \sigma_{\xi}^2 = \sigma_{\xi}^2 (1 + \gamma_1 W_{12}^{\bullet} W_{12}^{\bullet*} + \gamma_2 W_{13}^{\bullet} W_{13}^{\bullet*} + \gamma_3 W_{14}^{\bullet} W_{14}^{\bullet*}) = \sigma_{\xi}^2 \left(1 + \sum_{j=1}^3 \frac{\gamma_j \beta_j^2}{(\gamma_j + \beta_j)^2} \right). \tag{10}$$

Введём дополнительные обозначения:

$$\alpha = \frac{X^2}{\sigma_{\xi}^2}; \quad m_1 = \frac{\beta_1}{\gamma_1 + \beta_1}; \quad m_2 = \frac{\beta_2}{\gamma_2 + \beta_2}; \quad m_3 = \frac{\beta_3}{\gamma_3 + \beta_3},$$

где α – отношения сигнал/флуктуационный шум в парциальных трактах приёма пеленгационной антенны.

При этом отношения сигнал/помеха (ОСП) в каждом из четырёх парциальных трактов приёма пеленгационной антенны после когерентной компенсации будут описываться следующей зависимостью:

$$\text{ОСП}_i = 10 \lg \left(\frac{\alpha}{1 + \sum_{j=1}^3 \beta_j (1 - m_j)^2 + \sum_{j=1}^3 \gamma_j m_j^2} \right) - H_i \text{ [дБ]}, \tag{11}$$

где $i = 1, 2, 3, 4$ – номер парциального тракта приёма; H_i [дБ] – коэффициент снижения мощности сигнала в парциальном тракте приёма за счёт уменьшения амплитуд спектральных составляющих в поддиапазонах частот $\Delta F_1, \Delta F_2, \Delta F_3$ фильтрации помех после операции вычитания сигналов.

На рис. 8 в качестве примера представлены зависимости отношения сигнал/помеха от отношений помеха/флуктуационный шум для разных отношений сигнал/флуктуационный шум при когерентной компенсации трёх узкополосных помех, рассчитанные с использованием выражения (11), при следующих параметрах сигнально-помеховых процессов в парциальных трактах приёма пеленгационной антенны: $H_1=H_2=H_3=H_4=3$ дБ, $\alpha=10, 20, 30$ дБ, $\Delta F_1=100$ Гц, $\Delta F_2=200$ Гц, $\Delta F_3=400$ Гц, $\Delta F_s=7.9$ кГц, $\beta_1=\beta_2=\beta_3=1 \dots 100$ дБ. В этом случае имеем $ОСП_1=ОСП_2=ОСП_3=ОСП_4$.

Как видно из рис. 8, отношения сигнал/помеха в четырёх парциальных трактах приёма пеленгационной антенны после когерентной компенсации не зависят от отношений помеха/флуктуационный шум и определяются уменьшенными на величины коэффициентов $H_1 \div H_4$ снижения мощностей сигнала отношениями сигнал/флуктуационный шум.

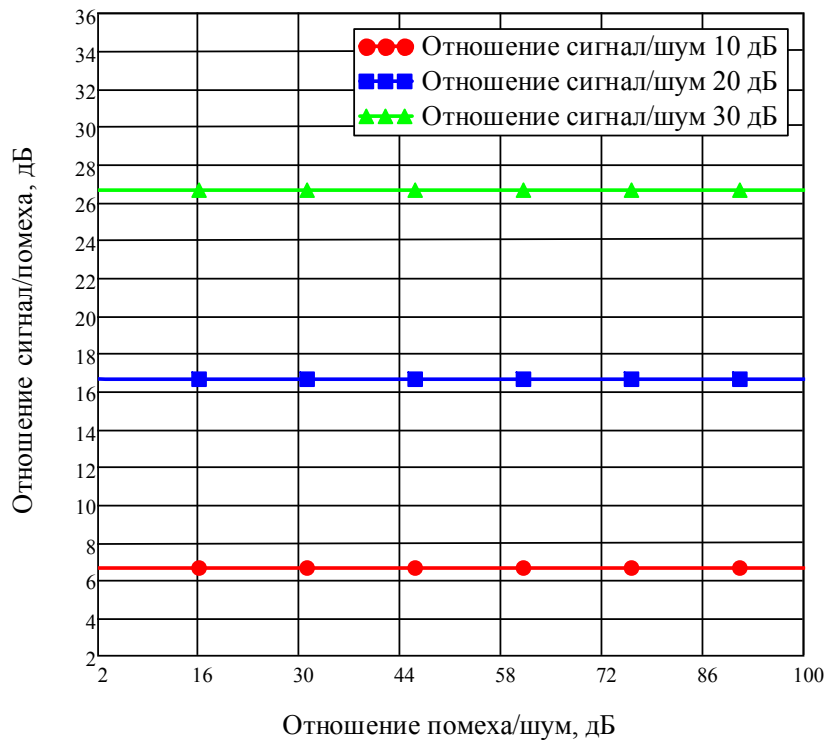


Рис. 8. Эффективность когерентной компенсации помех

Такой результат объясняется используемой при анализе математической моделью помех, предполагающей 100-процентную корреляцию колебаний компенсируемых помех, а значит, и их полное подавление в парциальных трактах приёма после операции вычитания (рис. 6), что применительно к реальным узкополосным помеховым процессам является вполне адекватным приближением.

Для определения коэффициентов $H_1 \div H_4$ необходимо знать спектральную плотность помехи опознавания маяка-пингера автономного подводного объекта в диапазоне частот ΔF_s , которая будет характеризоваться спектральным составом выбранных для маскирования фрагментов сонограмм записей звуков морской среды.

Поскольку $\Delta F_s \gg \Delta F_p$ (см. рис. 5), где $\Delta F_p = \Delta F_1 + \Delta F_2 + \Delta F_3$, а спектральная плотность известных фрагментов сонограмм записей звуков морской среды далеко не равномерна, то

целесообразно при формировании посылки опознавания маяка-пингера выбирать фрагменты сонограмм с частотными спектрами, имеющими спад амплитуд спектральных составляющих в области возможных частот сосредоточенных помех мешающего судоходства.

5. Заключение

Рассмотренный способ позволяет осуществлять скрытный гидроакустический поиск автономного подводного объекта с надводного поискового судна в условиях воздействия помех судоходства за счёт использования в процессе поисковых действий когерентной компенсации сосредоточенных помех мешающего судоходства на поисковом судне, выполняемой в паузах между интервалами излучения навигационного сигнала с подводного объекта, которая включает в себя частотное разделение помех от разных надводных источников в трёх выбранных поддиапазонах частот, комплексное взвешивание в выбранных поддиапазонах частот колебаний сосредоточенных помех соответствующих парциальных трактов приёма пеленгационной антенны, вычитание из колебаний сосредоточенных помех каждого парциального тракта приёма трёх взвешенных в выбранных поддиапазонах частот колебаний сосредоточенных помех соответствующих вспомогательных парциальных трактов приёма, приостановку когерентной компенсации на время приёма навигационного сигнала маяка-пингера, осуществление после когерентной компенсации линейного амплитудного детектирования и последетекторного сложения сигналов парциальных трактов приёма пеленгационной антенны, что в совокупности позволяет реализовать в процессе гидроакустического поиска достоверный приём навигационного сигнала автономного подводного объекта на поисковом судне. В основу способа подавления помех положена пространственно-временная обработка сигналов маяка-пингера автономного подводного объекта, использующая пространственные, временные и частотные различия навигационного сигнала и сосредоточенных помех мешающего судоходства.

Литература

1. Патент РФ 2727331, МПК G01S 15/00. Способ гидроакустического поиска автономного донного подводного объекта / В. Г. Арсентьев, Г. И. Криволапов. Заявка 2019145191, заявлено 25.12.2019, опубликовано 21.07.2020.
2. *Арсентьев В. Г., Криволапов Г. И.* Гидроакустический поиск автономного необитаемого подводного объекта // Вестник СибГУТИ. 2020. № 3. С. 64–78.
3. Спектры шумов кораблей [Электронные ресурсы] // URL: https://myslide.ru/documents_4/7f018b920e76ca67a70f5e146c83f809/img13.jpg; https://present5.com/presentation/183941860_279148258/image-10.jpg (дата обращения: 10.01.2022).
4. Модели сигналов и помех [Электронный ресурс] // URL: https://scask.ru/i_book_nsp.php?id=37 (дата обращения: 10.01.2022).
5. *Уидроу Б., Стирнз С.* Адаптивная обработка сигналов. Пер. с англ. под ред. В. В. Шахгильдяна. М.: Радио и связь, 1989. 440 с.
6. Адаптивная компенсация помех в каналах связи / Ю. И. Лосев, А. Г. Бердников, Э. Ш. Гойхман, Б. Д. Сизов. М.: Радио и связь, 1988. 208 с.

*Статья поступила в редакцию 20.01.2022;
переработанный вариант – 10.03.2022.*

Арсентьев Виктор Георгиевич

к.т.н., ведущий научный сотрудник научно-технического центра специализированных информационных систем СибГУТИ (630008, Новосибирск, ул. Бориса Богаткова, 51), тел. (383) 2-693-938, e-mail: viktor.arsentev.51@mail.ru.

Криволапов Геннадий Илларионович

к.т.н., доцент, заведующий лабораторией, руководитель научно-технического центра специализированных информационных систем СибГУТИ (630008, Новосибирск, ул. Бориса Богаткова, 51), тел. (383) 2-693-942, e-mail: krivolapov@sibsutis.ru.

References

1. *Patent RF 2727331, MPK G01S 15/00. Sposob gidroakusticheskogo poiska avtonomnogo donnogo podvodnogo ob#ekta* [RF Patent. Method for hydroacoustic search of an autonomous bottom underwater object]. V. G. Arsent'ev, G. I. Krivolapov. *Zajavka 2019145191, zjavleno 25.12.2019, opublikovano 21.07.2020.*
2. Arsent'ev V. G., Krivolapov G. I. *Gidroakusticheskij poisk avtonomnogo neobitaemogo podvodnogo ob#ekta* [Hydroacoustic search for an autonomous uninhabited underwater object]. *Vestnik SibGUTI.* 2020, no. 3, pp. 64-78.
3. *Spektry шумов кораблей* [Ship noise spectra] [Electronic resource]. URL: https://myslide.ru/documents_4/7f018b920e76ca67a70f5e146c83f809/img13.jpg; https://present5.com/presentation/183941860_279148258/image-10.jpg (access date: 10.01.2022).
4. *Modeli signalov i pomeh* [Signal and interference models]. [Electronic resource]. URL: https://scask.ru/i_book_nsp.php?id=37 (access date: 10.01.2022).
5. Uidrou B., Stirnz S. *Adaptivnaja obrabotka signalov. Per. s angl. pod red. V. V. Shahgil'djana.* [Adaptive signal processing. Translation from English]. Moscow: Radio and communication, 1989, 440 p.
6. *Adaptivnaja kompensacija pomeh v kanalah svyazi* [Adaptive interference compensation in communication channels]. Ju. I. Losev, A. G. Berdnikov, Je. Sh. Goghman, B. D. Sizov. Moscow: Radio and communication, 1988, 208 p.

The suppression of navigation interference in the underwater sonar search of an object

Viktor G. Arsent'ev

Candidate of technical sciences, Leading researcher, Siberian State University of Telecommunications and Information Sciences (SibSUTIS, Novosibirsk, Russia), viktor.arsentev.51@mail.ru.

Gennagy I. Krivolapov

Candidate of technical sciences, Head of laboratory, Siberian State University of Telecommunications and Information Sciences (SibSUTIS, Novosibirsk, Russia), krivolapov@sibsutis.ru.

A method for suppressing hindering navigation interference while sonar searching of an autonomous uninhabited underwater object located at the bottom of a known sea area from a surface vessel is considered.

Keywords: search vessel, navigation signal, coherent compensation of concentrated interference.

Математическая модель технологического процесса абсорбции природного газа

А. Н. Паршуков

В работе получена математическая модель технологического процесса абсорбции как объекта управления. Данная модель описывает процесс абсорбции природного газа и предназначена для автоматического управления этим процессом. Основу модели составляют уравнения материального и теплового балансов. Результаты численного моделирования подтверждают чувствительность поведения процесса к автоматическому регулированию.

Ключевые слова: абсорбция природного газа, математическая модель абсорбера, автоматическое регулирование.

1. Введение

Одним из самых эффективных методов осушки природного газа является абсорбция. Абсорбцией называют процесс поглощения целевого компонента (в данном случае – воды) из газовой или парогазовой смеси специально подобранным жидким поглотителем – *абсорбентом*. Осушка газа обеспечивает непрерывную эксплуатацию газового оборудования и магистральных газопроводов и предотвращает гидратообразование, а также возникновение ледяных заторов в системах. Данный процесс является основным в технологии подготовки природного газа к транспорту на газоконденсатных месторождениях Крайнего Севера.

Анализ статей [1–12], посвящённых разработке автоматизированных систем управления процессами добычи, подготовки и транспортировки природного газа, показал, что в газовой отрасли реконструкции подвергается как технологическое оборудование, так и связанные с ним автоматизированные системы управления. Проведение данных работ вызвано высокими требованиями к качеству подготовки природного газа для его транспортировки, повышением надёжности технологического оборудования в связи с непрерывным характером производства и стремлением к оптимизации режимов работы установок комплексной подготовки газа.

Важным этапом создания современных автоматических систем управления для газовой отрасли является составление математических моделей всех технологических процессов и постановка вычислительных экспериментов.

Моделированию процесса абсорбции посвящено большое количество работ (см., напр., [3–12]). Данные модели, как правило, представляют собой систему дифференциальных уравнений в частных производных, описывающих тепло- и массообменные процессы внутри абсорбера и дополненных граничными условиями. Так, например, в работе [10] математическая модель абсорбера имеет вид следующей системы уравнений:

$$\begin{cases} \frac{\partial C_G}{\partial t} = -v_G \cdot \frac{\partial C_G}{\partial z} - R_G (C_G - C_G^0), & z \in (0; l_A) \\ \frac{\partial C_A}{\partial t} = f(u) \cdot \frac{\partial C_A}{\partial z} + R_A (C_G - C_G^0), \end{cases} \quad (1)$$

здесь $C_G = C_G(z, t)$ – концентрация газа на высоте z абсорбера в момент времени t ; $C_A = C_A(z, t)$ – концентрация абсорбента на высоте z в момент t ; C_G^0 – концентрация газа на высоте z в установившемся режиме (заранее неизвестная функция от переменной z , подлежит расчёту в процессе решения системы (1)); R_G и R_A – физико-технологические коэффициенты по газу и абсорбенту соответственно; v_G – скорость газа, она зависит от расхода газа Q_G , температуры газа T_G , давления газа P_G по закону:

$$v_G = \frac{V_m \cdot Q_G \cdot P_0 (T_0 + T_G)}{0.785 \cdot T_0 \cdot P_G \cdot D^2},$$

здесь $V_m = 22.4$ л – объем моля идеального газа при нормальных условиях; $T_0 = 273$ К⁰ – нормальная температура; $P_0 = 0.101$ МПа – нормальное давление; D – внутренний диаметр горизонтального сечения абсорбера (вертикальной ректификационной колонны); $f(u)$ – функция управления процессом; l_A – высота абсорбера.

Система уравнений (1) дополняется граничными условиями:

$$C_G(z, t)|_{z=0} = \bar{C}_G(t), \quad C_A(z, t)|_{z=l_A} = \bar{C}_A(t), \quad (2)$$

и начальными условиями:

$$C_G(z, t)|_{t=0} = C_G^0(z), \quad C_A(z, t)|_{t=0} = C_A^0(z). \quad (3)$$

Из системы уравнений (1)–(3) следует, что абсорбер представляет собой распределённый объект управления. Для управления таким объектом необходимо каждый раз решать систему уравнений (1) в частных производных с граничными и начальными условиями (2)–(3). Таким образом, математическую модель вида (1)–(3) невозможно использовать для оперативного управления технологическим процессом абсорбции, поскольку оперативное управление требует немедленной реакции (т.е. управления) на постоянно меняющуюся обстановку (в данном случае – на изменение граничных условий (3)).

Из проведённого анализа следует, что задача разработки новой математической модели абсорбционного процесса, которая позволяла бы управлять процессом в режиме реального времени, остаётся актуальной.

2. Математическая модель процесса абсорбции

Процесс абсорбции газа происходит в абсорберах, установленных на дожимных компрессорных станциях. Схема работы абсорбера показана на рис. 1. В качестве абсорбента применяют раствор диэтиленгликоля (ДЭГ). Газ и абсорбент контактируют внутри аппарата, перемещаясь противотоком. Газ поднимается снизу вверх, а абсорбент по тарелкам стекает сверху вниз. Последний насыщается водой из газа и отправляется вниз на регенерацию. Вверх колонны уходит уже осушенный газ.

При составлении математической модели технологического процесса абсорбции приняты следующие гипотезы:

- 1) извлечению подлежит один компонент газовой фазы (*целевой компонент*) – вода;
- 2) поток газа, направленный вверх, и поток жидкости, направленный вниз, перемещаются без какого-либо перемешивания между элементарными цилиндрическими слоями по оси потока (модель аппарата идеального вытеснения¹, по классификации, изложенной [13, с. 56–57]);
- 3) входные и выходные потоки газа равны, входные и выходные потоки жидкости также равны.

Другие, менее существенные, допущения будут изложены в процессе составления математической модели.

Введём следующие обозначения (см. рис. 1):

M – количество воды, поступающей в процессе абсорбции из газа в жидкость (абсорбент), моль/с;

F^G – поток газа, м³/с;

F^L – поток жидкости, м³/с;

Y_{IN} – начальная концентрация воды в газе, моль/м³;

Y_{OUT} – конечная концентрация воды в газе, моль/м³;

X_{IN} – начальная концентрация воды в жидкости, моль/м³;

X_{OUT} – конечная концентрация воды в жидкости, моль/м³;

T_{IN} – начальная температура жидкости, К⁰;

T_{OUT} – конечная температура жидкости, К⁰;

C – удельная теплоёмкость жидкости, Дж/(кг·К⁰);

ρ – плотность жидкости, кг/м³;

t – время, с.

В стационарном режиме уравнение материального баланса в абсорбере принимает вид:

$$M = F^G (Y_{IN} - Y_{OUT}) = F^L (X_{OUT} - X_{IN}), \quad \text{моль/с.} \quad (4)$$

Абсорбция является неизотермическим химическим процессом: при абсорбции происходит выделение теплоты. Количество выделившейся теплоты Q в стационарном режиме описывается уравнением теплового баланса

$$Q = q \cdot M = q F^L (X_{OUT} - X_{IN}), \quad \text{Дж/с,} \quad (5)$$

здесь q – теплота растворения одного моля воды в одном кубическом метре жидкости, Дж/моль.

¹ Модель аппарата идеального вытеснения означает, что линейная скорость потока целевого компонента в любой точке горизонтального сечения будет одинакова, а от сечения к сечению скорость потока может меняться (например, за счёт изменения объема при осуществлении физико-химического процесса абсорбции), и что перенос тепла путём теплопроводности незначителен.

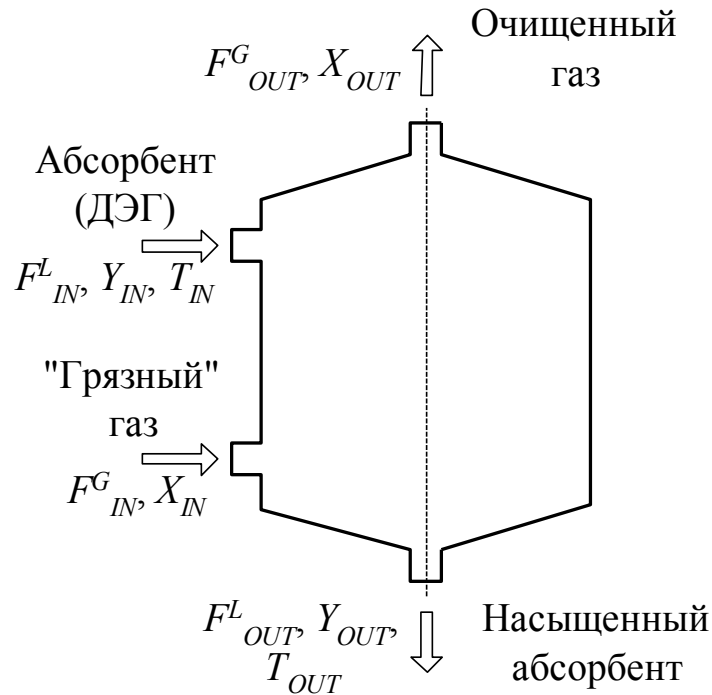


Рис. 1. Схема работы абсорбера

Примем, что вся выделившаяся теплота Q затрачивается на нагрев жидкости (то есть нагревом газа пренебрегаем), тогда

$$Q = c\rho F^L (T_{OUT} - T_{IN}). \quad (6)$$

Известно, что с ростом температуры растворимость газа уменьшается [13, с. 65], следовательно, выделение теплоты является негативным фактором. Негативное влияние повышения температуры стремятся компенсировать, подавая на вход абсорбера охлажденный абсорбент, то есть уменьшая T_{IN} в (6).

Таким образом, для расчёта стационарного режима работы абсорбера необходимо знать следующие параметры: расход газа F^G , начальную концентрацию воды в газе Y_{IN} , конечную концентрацию воды в газе Y_{OUT} , начальную концентрацию воды в абсорбенте X_{IN} (возможно, что $X_{IN}=0$), начальную температуру абсорбента T_{IN} . А вот расход абсорбента F^L , конечную концентрацию воды в абсорбенте X_{OUT} и конечную температуру абсорбента T_{OUT} заранее знать не можем.

Неизвестные X_{OUT} и F^L связаны уравнением материального баланса (4). Задавшись концентрацией X_{OUT}^* , можем вычислить минимально необходимый расход абсорбента F_{min}^L из уравнения

$$\frac{F_{min}^L}{F^G} = \frac{Y_{IN} - Y_{OUT}}{X_{OUT}^* - X_{IN}}.$$

Следовательно, для нормального протекания процесса абсорбции должно выполняться

$$F^L \geq F_{min}^L,$$

а параметр

$$l = \frac{F^L}{F_{min}^L}$$

назовём коэффициентом избытка абсорбента.

Рассматривая технологический процесс абсорбции как объект управления, можем выделить в качестве *контролируемых* (управляемых) *переменных* два наиболее важных физических параметра процесса – Y_{OUT} и T_{OUT} , в качестве *возмущения* – нестационарность Y_{IN} , в качестве управляющих воздействий (*управления*) – F^L и T_{IN} (см. рис. 2). На практике возможности управления по переменным F^L и T_{IN} довольно ограничены.

Цель управления состоит в стабилизации контролируемых переменных Y_{OUT} и T_{OUT} при изменениях Y_{IN} (колебания Y_{IN} относительно некоторого значения Y_{IN}^0 носят случайный характер, поэтому не могут быть учтены заранее).

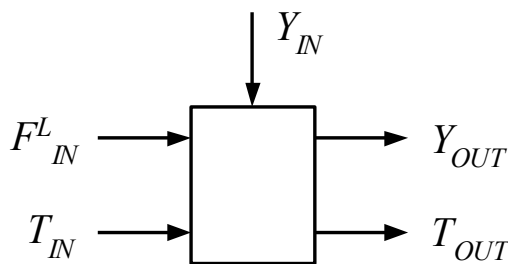


Рис. 2. Процесс абсорбции как объект управления

Математическую модель процесса абсорбции получим из уравнений (4), (5) и (6), выразив в них контролируемые переменные Y_{OUT} и T_{OUT} :

$$\begin{cases} Y_{OUT} = Y_{IN} - \frac{F^L}{F^G} (X_{OUT} - X_{IN}), \\ T_{OUT} = T_{IN} + \frac{q}{c\rho} (X_{OUT} - X_{IN}). \end{cases} \quad (7)$$

Переходные процессы в аппарате определяются скоростью газовой фазы (т.е. возникают и затухают практически мгновенно), поэтому следует ожидать, что данная математическая модель будет описывать управляемый процесс с удовлетворительной точностью. Полученная модель (7) позволяет рассчитывать изменения контролируемых переменных и может быть использована для синтеза законов автоматического регулирования процессом.

3. Пример моделирования переходных процессов в абсорбере

Из справочника [14, с. 247] для диэтиленгликоля ($C_4H_{10}O_3$) находим плотность (она задаётся ГОСТом [15]):

$$\rho = 1.118 \cdot 10^3, \quad \text{кг/м}^3,$$

удельную теплоёмкость

$$c = 2.093 \cdot 10^3, \quad \frac{\text{Дж}}{\text{кг} \cdot \text{K}^0},$$

и теплоту растворения одного моля воды в одном куб. м. ДЭГ:

$$q = 1.274 \cdot 10^6, \quad \text{Дж/моль}.$$

Для расчёта коэффициентов модели (4) примем следующие значения рабочего режима:

$$Y_{IN} = Y_{IN}^0 \equiv 0.556, \text{ моль/м}^3, \quad Y_{OUT} = Y_{OUT}^0 \equiv 0, \text{ моль/м}^3, \quad F^G = 6.67 \cdot 10^3, \text{ м}^3/\text{с}, \\ X_{IN} = 0, \text{ моль/м}^3, \quad X_{OUT} = 0.5, \text{ моль/м}^3.$$

Из 1-го уравнения системы (7) вычисляем значение F^L :

$$F^L = F^G \left(\frac{Y_{IN}^0 - Y_{OUT}^0}{X_{OUT} - X_{IN}} \right) = 8900.4, \text{ м}^3/\text{с},$$

а из 2-го уравнения – температуру абсорбента на входе в аппарат:

$$T_{IN} = T_{OUT} - \frac{q}{c\rho} (X_{OUT} - X_{IN}) = 287.0, \text{ К}^0.$$

Рассмотрим поведение процесса абсорбции при случайных колебаниях Y_{IN} в интервале $\pm 5\%$ от значения Y_{IN}^0 , то есть

$$Y_{IN}(t) = Y_{IN}^0 (1 + \zeta(t)),$$

график случайной величины ζ приведён на рис. 3.

Непосредственно из математической модели (7) находим значения управляющих воздействий:

$$F^L(t) = F^G \left(\frac{Y_{IN} - Y_{OUT}^0}{X_{OUT} - X_{IN}} \right), \text{ м}^3/\text{с}, \quad T_{IN}(t) = 287.0, \text{ К}^0.$$

Поскольку модель (7) статическая, вполне ожидаемо, что график управляющего воздействия F^L (см. рис. 4) меняется почти пропорционально текущему значению возмущения Y_{IN} .

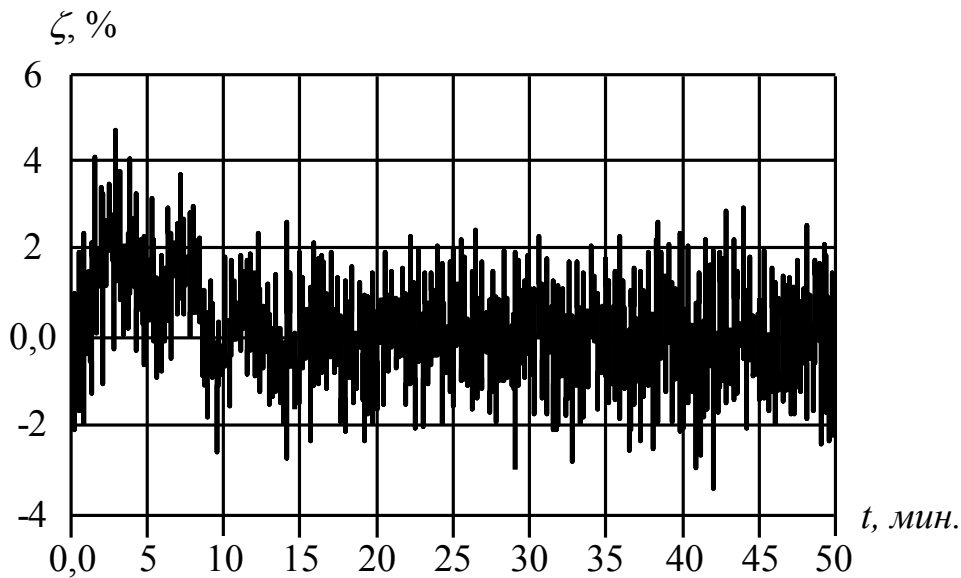
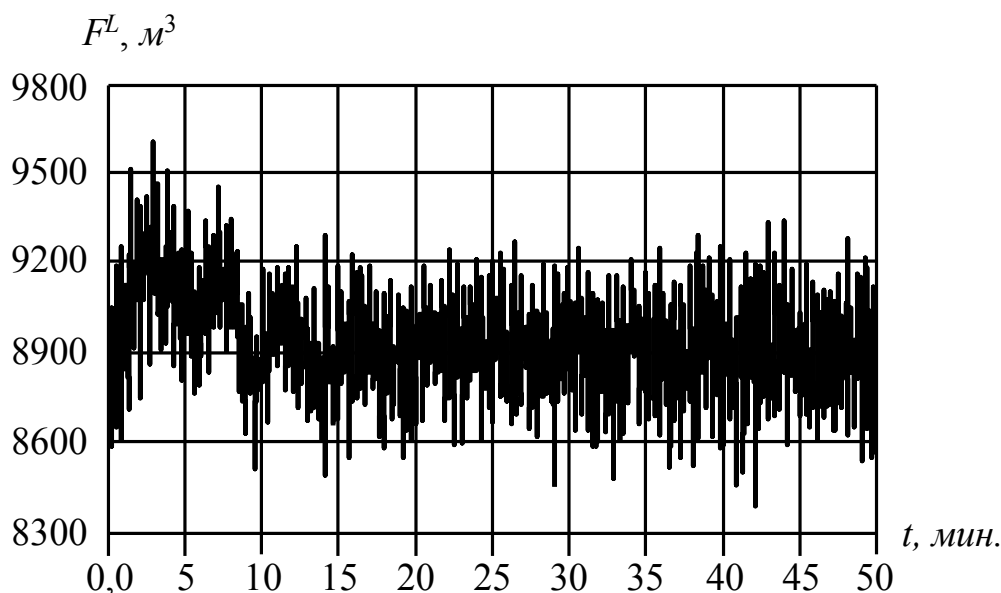


Рис. 3. График изменения возмущения ζ

Рис. 4. График управления F^L

4. Заключение

В настоящей работе предложена новая математическая модель процесса абсорбции. Данная модель предназначена для автоматического управления процессом абсорбции. В статье приведён пример расчета закона регулирования, гарантирующего стабилизацию контролируемых переменных – Y_{OUT} и T_{OUT} .

Литература

1. Мусакаев Н. Г., Хасанов М. К. Математическое моделирование процесса добычи газа из газогидратной залежи с учётом образования льда // Вестник Тюм. гос. ун-та: Физико-математические науки. Информатика. 2014. № 7. С. 43–50.
2. Шильев М. И., Толстых А. В. Моделирование процессов абсорбции газов в барботажных аппаратах // Теплофизика и аэромеханика. 2013. № 5. С. 575–586.
3. Абрамкин С. Е., Душин С. Е. Управление технологическими процессами газодобывающих комплексов // Материалы IX ВНК «Системный синтез и прикладная синергетика», п. Нижний Архыз, 24–27 сентября 2019. С. 430–439.
4. Абрамкин С. Е., Душин С. Е., Первухин Д. А. Проблемы разработки систем управления газодобывающими комплексами // Известия вузов. Приборостроение. 2019. Т. 62, № 8. С. 685–692.
5. Абрамкин С. Е., Душин С. Е., Сердитов Ю. Н. Исследование физических процессов в ректификационной колонне при регенерации диэтиленгликоля // Известия ЮФУ. Технические науки. 2018. № 5. С. 69–78.
6. Исмаилов Р. Н., Абрамкин С. Е., Душин С. Е. Состояние и перспективы развития автоматизации установок комплексной подготовки газа на УНГКМ // Перспективные направления развития Уренгойского комплекса: сб. науч. тр. ООО «Газпром добыча Уренгой». Москва: Недра, 2018. С. 271–281.
7. Баширцева Н. Ю. Переработка природных газов: настоящее и будущее // Газовая промышленность. 2013. № 6. С. 86–87.

8. Бодрый А. Б. Инновации в области адсорбционной осушки углеводородных газов // Материалы XXVI Всероссийского межотраслевого совещания. ОАО «НИПИгазпереработка», Краснодар, 2013. С. 54–56.
9. Бородин А. В. Добыча углеводородного сырья и перспективы интенсификации переработки попутного нефтяного газа // Нефтепереработка и нефтехимия. Москва: ОАО «ЦНИИТЭНЕФТЕХИМ». 2011. № 4. С. 56–59.
10. Абрамкин С. Е., Душин С. Е. Математическая модель массотеплообменных процессов технологического комплекса абсорбционной осушки газа // Материалы VI научной конференции «Управление и информационные технологии». Санкт-Петербург: СПбГЭТУ «ЛЭТИ». 2010. С. 220–224.
11. Бетлинский В. Ю., Жидков М. А., Овчинников В. П. Двухпоточные регулируемые вихревые трубы в промышленных установках очистки и осушки газов // Газовая промышленность. 2008. № 1. С. 72–75.
12. Балавин М. А. Математическое моделирование и оптимизация технологических процессов подготовки газа к транспорту: автореферат диссертации канд. техн. наук. Самара, 2000. 24 с.
13. Жоров Ю. М. Моделирование физико-химических процессов нефтепереработки и нефтехимии. Москва: Химия. 1978. 376 с.
14. Справочник химика / П. Б. Никольский [и др.]. 2-е изд., перераб. и доп. Ленинград: Химия, 1968. Т. 5. 976 с.
15. ГОСТ 10136-2019. Международный стандарт. Диэтиленгликоль. Технические условия. Введён 2020-05-01. Москва, 2019. 19 с.

*Статья поступила в редакцию 24.01.2022;
переработанный вариант – 23.02.2022.*

Паршуков Андрей Николаевич

к.т.н., доцент; доцент кафедры электроэнергетики Тюменского индустриального университета (625000, Тюмень, ул. Володарского, 38), e-mail: anparshukov@mail.ru.

References

1. Musakaev N. G., Hasanov M. K. *Matematicheskoe modelirovanie processa dobychi gaza iz gazogidratnoj zalezhi s uchjotom obrazovaniya l'da* [Mathematical modeling of the process of gas production from a gas hydrate deposit under the formation of ice]. *Vestnik Tjum. gos. un-ta: Fiziko-matematicheskie nauki. Informatika*. 2014, no. 7, pp. 43-50.
2. Shiljaev M. I., Tolstyh A. V. *Modelirovanie processov absorbcii gazov v barbotazhnyh apparatah* [Simulation of gas absorption processes in bubbling apparatuses]. *Teplofizika i aeromehanika*. 2013, no. 5, pp. 575-586.
3. Abramkin S. E., Dushin S. E. *Upravlenie tehnologicheskimi processami gazodobyvaju-shhih kompleksov* [Control of technological processes of gas production complexes]. *Materialy IX VNK «Sistemnyj sintez i prikladnaja sinergetika»* [Materials of the IX VNK "System synthesis and applied synergetics"], Nizhny Arkhyz village, September 24-27, 2019, pp. 430-439.
4. Abramkin S. E., Dushin S. E., Pervuhin D. A. *Problemy razrabotki sistem upravleniya gazodobyvajushimi kompleksami* [Problems of development of control systems for gas production complexes]. *Izvestija vuzov. Priborostroenie*. 2019, vol. 62, no. 8, pp. 685-692.
5. Abramkin S. E., Dushin S. E., Serditov Ju. N. *Issledovanie fizicheskikh processov v rektifikacionnoj kolonne pri regeneracii dijetilenglikolja 5* [Study of physical processes in a distillation column during the regeneration of diethylene glycol]. *Izvestija JuFU. Tehnicheskie nauki*. 2018, no. 5, pp. 69-78.

6. Ismagilov R. N., Abramkin S. E., Dushin S. E. *Sostojanie i perspektivy razvitija avtomatizacii ustanovok kompleksnoj podgotovki gaza na UNGKM* [State and prospects for the development of automation of complex gas treatment units at the UNGCF]. *Perspektivnye napravlenija razvitija Urengojnskogo kompleksa: sb. nauch. tr. OOO «Gazprom dobycha Urengoj»*. Moscow: Nedra, 2018, pp. 271-281.
7. Bashkirceva N. Ju. *Pererabotka prirodnyh gazov: nastojashhee i budushhee* [Processing of natural gases: present and future]. *Gazovaja promyshlennost'*. 2013, no. 6, pp. 86-87.
8. Bodryj A. B. *Innovacii v oblasti adsorbcionnoj osushki uglevodorodnyh gazov* [Innovations in the field of adsorption drying of hydrocarbon gases]. *Materialy XXVI Vserossijskogo mezhotraslevogo soveshhanija. OAO «NIPigazpererabotka»* [Materials of the XXVI All-Russian Intersectoral Conference. OJSC NIPigazpererabotka], Krasnodar, 2013, pp. 54-56.
9. Borodin A. V. *Dобыча углеводородного сырья i perspektivy intensifikacii pererabotki poputnogo nefljanogo gaza* [Production of hydrocarbon raw materials and prospects for intensification of associated petroleum gas processing]. *Neftepererabotka i neftehimija*. Moscow: OJSC «CNIITENEFTEKHIM». 2011, no. 4, pp. 56-59.
10. Abramkin S. E., Dushin S. E. *Matematicheskaja model' massoteploobmennyh processov tehnologicheskogo kompleksa absorbcionnoj osushki gaza* [Mathematical model of mass and heat transfer processes of the technological complex for absorption gas dehydration]. *Materialy VI nauchnoj konferencii «Upravlenie i informacionnye tehnologii»* [Materials of the VI Scientific Conference "Management and Information Technologies"]. St.Peterburg: St.Petersburg Electrotechnical University "LETI", 2010, pp. 220-224.
11. Betlinskij V. Ju., Zhidkov M. A., Ovchinnikov V. P. *Dvuhpotochnye reguliruemye vihrevye trubny v promyshlennyh ustanovkah ochistki i osushki gazov* [Double-flow adjustable vortex tubes in industrial gas cleaning and drying plants]. *Gazovaja promyshlennost'*. 2008, no. 1, pp. 72-75.
12. Balavin M. A. *Matematicheskoe modelirovanie i optimizacija tehnologicheskikh processov podgotovki gaza k transportu* [Mathematical modeling and optimization of technological processes of gas preparation for transport]: *avtoreferat dissertacii kand. tehn. nauk*. Samara, 2000, 24 p.
13. Zhorov Ju. M. *Modelirovanie fiziko-himicheskikh processov neftepererabotki i neftehimii* [Modeling of physical and chemical processes of oil refining and petrochemistry]. Moscow: Chemistry, 1978, 376 p.
14. *Spravochnik himika* [Chemist's Handbook]. P. B. Nikol'skij [i dr.]. 2-e izd., pererab. i dop. Leningrad: Chemistry, 1968, vol. 5, 976 p.
15. GOST 10136-2019. *Mezhdunarodnyj standart. Dijetilenglikol'. Tehnicheskie uslovija* [International standard. Diethylene glycol. Specifications.]. Introduced 2020-05-01. Moscow, 2019, 19 p.

Mathematical model of the natural gas absorption technological process

Andrey N. Parshukov

Candidate of technical sciences, Docent, Industrial University of Tyumen (Tyumen, Russia),
anparshukov@mail.ru.

In the paper, a mathematical model of the absorption technological process as an object of management is obtained. This model describes the process of natural gas absorption and is designed to control this process automatically. The model is based on the equations of material and thermal balances. The results of numerical simulation confirm the sensitivity of the process behavior to automatic regulation.

Keywords: natural gas absorption, mathematical model of the absorber, automatic regulation.

Анализ эффективности потоково-локальной сборки мусора в распределённых системах хранения и обработки данных

А. Ю. Филатов, В. В. Михеев

В данной работе рассматривается потоково-локальная сборка мусора – техника автоматического управления памятью, которая использует свойство объекта быть доступным только одному потоку исполнения. Возможность выделять и переиспользовать динамическую память независимо в каждом потоке, т.е. без дорогостоящих операций синхронизации, улучшает производительность и масштабируемость прикладных программ. В работе анализируется применимость данной техники к распределённым системам хранения и обработки данных, используемым в промышленности. Результаты свидетельствуют о том, что потоково-локальное управление памятью позволяет в полной мере задействовать ресурсы вычислительной системы и является многообещающим подходом к сборке мусора в высоконагруженных средах исполнения.

Ключевые слова: сборка мусора, виртуальная машина, потоково-локальные кучи, NUMA, JVM.

1. Введение

Распределённые системы хранения и обработки больших массивов данных получили широкое распространение в науке и промышленности. Программные комплексы Apache MapReduce и Apache Spark, реализованные на языках Java и Scala, благодаря своей настраиваемости, отказоустойчивости и высокой производительности стали основой для построения распределённых кластеров, состоящих из десятков, сотен и даже тысяч вычислительных узлов. Большое внимание уделяется увеличению производительности таких систем на низком уровне (выбор оборудования, конфигурация операционной системы, настройки сети) и на более высоком, связанном с тем, насколько эффективно управляемая среда (Java Virtual Machine, JVM) исполняет Java-байткод.

Пропускная способность упомянутых систем зависит от того, насколько эффективно программная среда задействует доступные ресурсы компьютера, особенно в контексте массово-параллельного исполнения в рамках одного процесса – все основные подсистемы JVM (управление памятью, генерация кода, динамическая загрузка классов) должны обеспечивать корректную работу сотен активных потоков исполнения. Важнейшей частью JVM, существенно влияющей на производительность программ, традиционно считается система автоматического управления памятью, которая обеспечивает эффективное выделение памяти для объектов и их своевременное удаление (т.н. «сборка мусора»).

Существуют алгоритмы сборки мусора, разделяющие объекты в динамической памяти программы на независимые группы – *локальные* объекты, доступные только одному потоку исполнения, и *глобальные* (или *разделяемые*), которые могут быть изменены несколькими потоками одновременно [1, 2]. Работа с потоково-локальными объектами не требует синхронизации потоков между собой, что позволяет выполнять операции по выделению памяти и сборке мусора

независимо, в каждом потоке отдельно, не внося излишних пауз в работу приложения. Также данный подход позволяет группировать потоково-локальные данные в соседних сегментах памяти, что улучшает производительность программ на оборудовании с неравномерным доступом к памяти (cache coherent non-uniform memory access, ccNUMA).

Заметим, что современные отказоустойчивые распределённые системы строятся по схожим принципам – они минимизируют разделяемое общее состояние, изолируя каждый из этапов вычислений от влияния других подсистем. Потоки программы внутри одного процесса выступают в роли независимых агентов, которые взаимодействуют друг с другом только с помощью редких сообщений, а значит, можно предположить, что в описанном сценарии потоково-локальная сборка мусора будет наиболее эффективной стратегией управления памятью.

Данная работа анализирует эффективность потоково-локального сборщика мусора (thread-local garbage collector, TLGC) в применении к распределённому кластеру, построенному на основе технологий Apache Hadoop [3] и Apache Spark [4]. Главным усовершенствованием системы управления памятью по сравнению с предшествующей работой авторов [5] стала поддержка в сборщике мусора т.н. «средних» объектов – объектов, чей размер варьируется в пределах от 104 байт до 8 килобайт. Использование продвинутых структур данных позволило минимизировать влияние внутренней фрагментации на производительность и в полной мере раскрыло потенциал TLGC в применении к многоуровневой распределённой системе, чувствительной ко времени работы сборщика мусора. Рассматривается изолированный (т.н. standalone) режим работы приложения – на одном выделенном вычислительном узле – и распределённый, предполагающий одновременное исполнение Spark-задачи на четырёх выделенных серверах, каждый из которых использует многопроцессорное ccNUMA-оборудование.

2. Архитектура системы управления памятью

Все объекты, хранимые в динамической памяти программы, разбиваются на две группы – локальные объекты, достижимые только из потока, который изначально выделил для них память, и глобальные (разделяемые) объекты, как проиллюстрировано на рис. 1. С каждым объектом ассоциирован один бит информации, позволяющий отличать локальные объекты от глобальных. Глобализация объекта может произойти только в определённых точках программы – при присваивании локального объекта в поле другого разделяемого объекта или в статическое поле класса, поэтому с каждым изменением объектного графа связана специализированная процедура – барьер на запись, который по необходимости изменяет «цвет» объекта с локального на разделяемый.

Стоит заметить, что алгоритм разделения объектов на потоково-локальные и разделяемые допускает несколько степеней свободы и для своей корректности требует отслеживания ряда инвариантов объектного графа. В исследовательской виртуальной машине использовался один из самых простых алгоритмов разграничения [2], который поддерживает множество разделяемых объектов транзитивно замкнутым – любой разделяемый объект может указывать только на разделяемые объекты, т.е. каждый сработавший барьер на запись помечает весь подграф объектов, достижимых от «утекающего», как глобальный.

2.1. Организация кучи

Для обеспечения эффективной работы приложения система автоматического управления памятью оперирует блоками непрерывного адресного пространства следующих видов:

- «Малые» блоки (16 килобайт) используются для размещения объектов фиксированного размера. Внутри такого блока располагаются только объекты одинакового размера, а сами блоки специализируются для размеров от 16 до 96 байт. Таким образом достигается

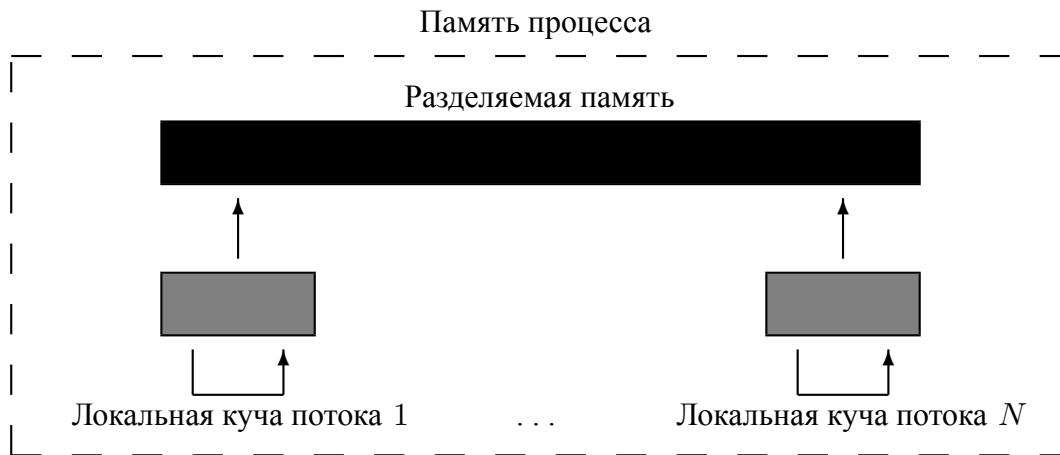


Рис. 1. Организация памяти в системах с потоково-локальными кучами

ся максимальная пропускная способность для случаев, когда приложение создаёт много сравнительно небольших объектов.

- «Средние» блоки (16 килобайт) используются для выделения объектов, чей размер варьируется от 104 байт до 8 килобайт. Из-за переменной величины запросов на выделение памяти такие блоки могут столкнуться с проблемой фрагментации, поэтому система управления памятью использует стратегию «лучший подходящий» (best-fit), которая позволяет уменьшить количество «вынужденно неиспользуемой» памяти.
- «Большие» блоки имеют произвольный размер и предназначены для хранения объектов, чей размер превышает 8 килобайт.

Стоит заметить, что исследуемая система получила возможность потоково-локально управлять памятью в средних блоках, что является её существенным улучшением по сравнению с работой [5]. Объекты среднего размера разбиты на независимые «кластеры» (используется подход, именуемый в литературе *segregated fit* [6]), а каждый запрос на выделение памяти приводит к поиску минимального по размеру свободного участка в соответствующем множестве с помощью сбалансированного дерева поиска специального вида [7]. Так как размеры запросов ограничены сверху, то даже в худшем случае поиск «лучшего подходящего» участка выполняется за константное время. Заметим, что введение потоково-локальных куч само по себе является источником фрагментации, а потому использование алгоритмов, минимизирующих внутреннюю фрагментацию, представляется оправданным выбором.

Потоково-локальный сборщик мусора разбивает доступные блоки на независимые группы:

- Общий пул больших блоков. Каждый поток может выделить память для большого объекта только с использованием специализированной критической секции. Так как выделение больших объектов происходит достаточно редко, то это не становится причиной уменьшения производительности системы.
- Потоково-локальные блоки (т.н. потоково-локальная куча). Каждому потоку принадлежит некоторый набор малых и средних блоков, которые могут использоваться для размещения новых объектов без какой-либо синхронизации с другими потоками.
- Общий пул пустых блоков. Если поток не может разместить объект в принадлежащем ему потоково-локальной блоке, то он запускает синхронизированную операцию по извлечению пустого блока из общего пула и добавлению его в потоково-локальную группу.

2.2. Локальная сборка мусора

Когда поток не может удовлетворить запрос на выделение памяти, он запускает потоково-локальную сборку мусора. Используемый подход является модификацией алгоритма маркировки с выметанием (*mark-sweep*) и состоит из следующих фаз:

- **Разметка.** Поток строит множество корневых объектов, хранящихся в регистрах и в памяти стековых кадров, и вычисляет множество объектов, достижимых от корневого множества. Разметка останавливается на разделяемых объектах, т.к. они по построению могут указывать только на другие разделяемые объекты.
- **Выметание.** Каждый потоково-локальный объект, который не был помечен на предыдущей фазе, считается мертвым, и занимаемая им память освобождается.

Если локальная сборка мусора не смогла освободить достаточный объём памяти, то запускается глобальная сборка мусора. Для этого все потоки приложения останавливаются в безопасных точках (*safe points*) [8], для них строится расширенное корневое множество (включающее в себя, в частности, объекты в статических полях классов) и запускается алгоритм сборки мусора, который способен переработать недостижимые разделяемые объекты и выполнить компактизацию кучи.

Глобальная сборка мусора обладает свойством *полноты* – каждый недостижимый объект будет удален, но требует остановки всех потоков приложения, что приводит к снижению пропускной способности системы и снижает её отзывчивость. Таким образом, основная цель локального сборщика мусора – уменьшить частоту глобальных сборок и снизить издержки по синхронизованному выделению памяти.

3. Экспериментальные результаты

Технологии распределённого хранения больших массивов данных и массово-параллельной их обработки Apache Hadoop и Apache Spark используются во многих современных промышленных системах. Исследователи используют эти системы для обработки так называемых «больших данных» с помощью различных алгоритмов. В основе данных инструментов лежат фреймворк MapReduce [9] и концепция Resilient Distributed Datasets [10], позволяющие компактно описать процедуру преобразования данных, которая может быть эффективно выполнена с участием распределённых агентов. Отказоустойчивость программных платформ достигается с помощью изоляции вычислений как на уровне процессов операционной системы, так и на уровне отдельных потоков виртуальной машины. Таким образом, каждый процесс-исполнитель использует несколько потоков для загрузки данных с жесткого диска и по сетевому протоколу, а также для выполнения пользовательских вычислений [11]. Можно сказать, что потоки в такой системе выступают как независимые агенты, которые разделяют между собой совсем небольшой объём данных.

Производительность описанных систем можно оценить с помощью набора общедоступных тестов производительности *HiBench* [12], которые исследуют скорость работы различных алгоритмов обработки данных в распределённой среде. Стоит заметить, что программный комплекс состоит из множества взаимосвязанных компонент, поэтому выбранные тесты измеряют производительность вычислительного кластера в целом – эффективность работы сетевого интерфейса, операционной системы, подсистемы памяти, процессора, а также JVM, которая исполняет код самого фреймворка, написанного преимущественно на языках Java и Scala. Для исследования использовались две конфигурации: режим исполнения в рамках одного процесса на одном физическом сервере `hibench.standalone` и режим распределённого кластера `hibench.cluster`. В качестве модельных примеров были выбраны тесты **WordCount**, **KMeans** и **TeraSort**.

Т а б л и ц а 1. Характеристики тестового набора `hibench.standalone`

Тест	Выделенная память, 1 поток	Выделенная память, 64 потока	Размер локальной кучи
KMeans	23 377 Mb	23 675 Mb	70 Mb
WordCount	74 635 Mb	74 794 Mb	5 Mb
TeraSort	29 413 Mb	30 540 Mb	5 Mb

Исследование выполнялось на базе исследовательской виртуальной Java-машины, разрабатываемой в исследовательском центре компании Huawei, которая использует статический оптимизирующий компилятор для трансляции Java-байт-кода в машинный код. Предшествующая система управления памятью будет обозначаться далее как *базовая* система управления памятью. Основные характеристики использованного оборудования:

- CentOS Linux 7, 4.14.0 aarch64;
- HiSilicon Kunpeng-920 2600 MHz x 64, 2 NUMA nodes, 2 sockets, 32 cores per socket;
- L1d cache: 64K, L1i cache: 64K, L2 cache: 512K, L3 cache: 32768K;
- 500 Gb RAM.

3.1. Hibench.standalone

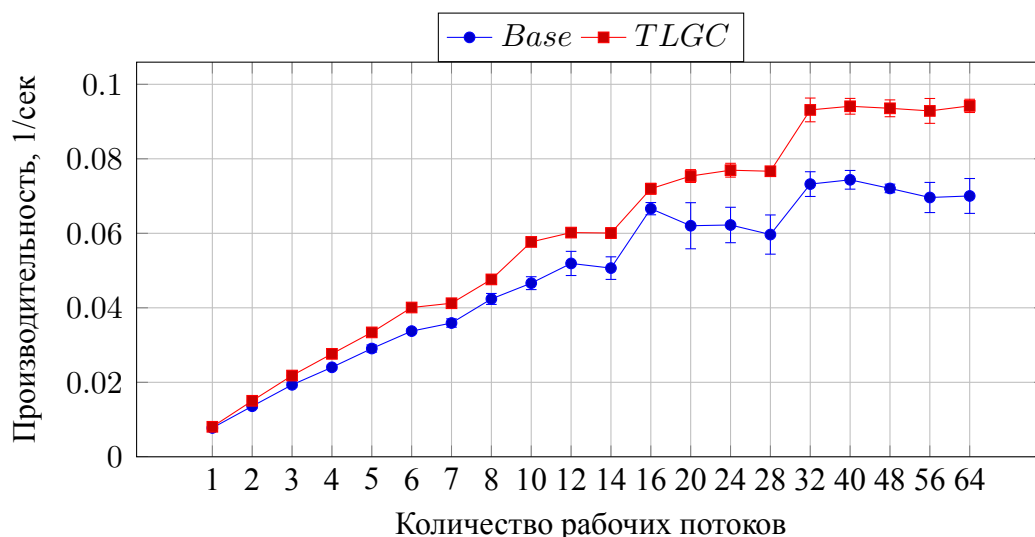
Режим запуска `standalone` предназначен для исполнения вычислительных задач в рамках одного процесса операционной системы на конкретном вычислительном узле. При этом подсистема балансировки нагрузки и обмена данными по сети не задействована. Работа с распределённым хранилищем данных HDFS заменяется работой непосредственно с файловой системой конкретного компьютера. Таким образом, данный режим позволяет уменьшить влияние настроек ОС, которые влияют на пропускную способность вычислительной системы, но при этом никаким образом не зависят от виртуальной машины. Мы использовали данный режим для того, чтобы оценить эффективность потоково-локального сборщика мусора в рамках одного процесса, запускающего большое число одновременно работающих потоков.

Тест выполняет фиксированную работу, используя указанное число рабочих потоков. В представленных графиках по оси ординат изображены не абсолютные значения производительности, а относительные – какая доля работы выполняется системой за одну секунду реального времени, больше – лучше.

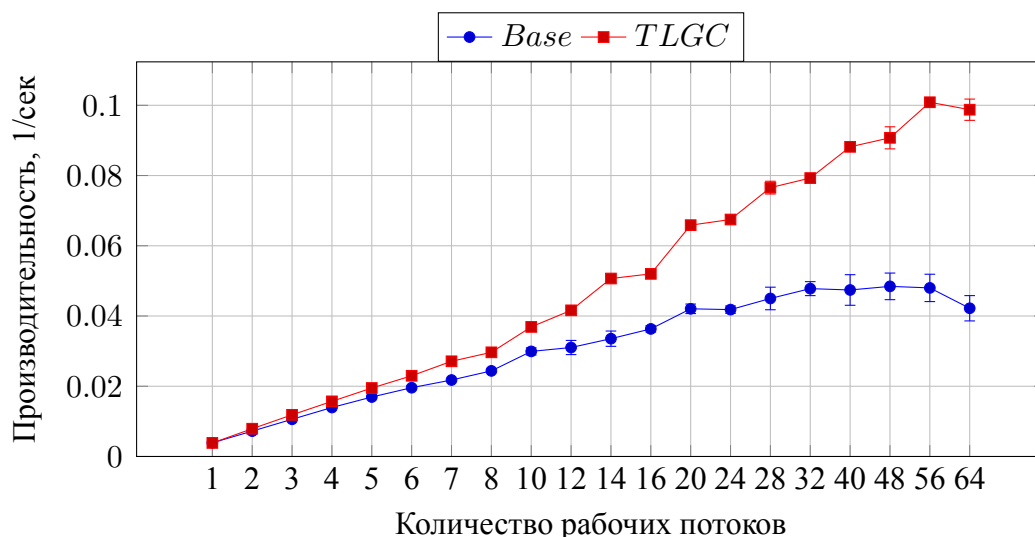
В табл. 1 собрана информация о количестве выделяемой динамической памяти и использованных при замерах настройках локального сборщика. Стоит заметить, что каждое из исследуемых приложений в многопоточном режиме работы создаёт большое количество объектов, и поэтому эффективность сборщика мусора становится определяющим фактором для итоговой производительности.

Тест **KMeans** измеряет производительность алгоритма кластеризации точек в многомерном пространстве методом k-средних [13]. Алгоритм реализован на языке Scala, который помогает программисту выражать преобразования данных в компактном декларативном стиле, используя парадигмы функционального программирования. Заметим, что при этом создается большое количество временных «служебных» объектов (`Tuple`, `Option` и т.п.). Также используются стандартные параметризованные коллекции языка, которые «заворачивают» (`box`) [14] примитивные типы (`int`, `float`, `double`) в экземпляры-обертки, создаваемые в динамической памяти.

Таким образом, производительность системы существенно зависит от используемого алгоритма сборки мусора, величины вносимой им задержки и поддерживаемой пространственной локальности данных. Рис. 2 показывает, что, независимо от используемого алгоритма сборки мусора, приложение имеет предел масштабируемости при 32–40 потоках. При этом TLGC поз-

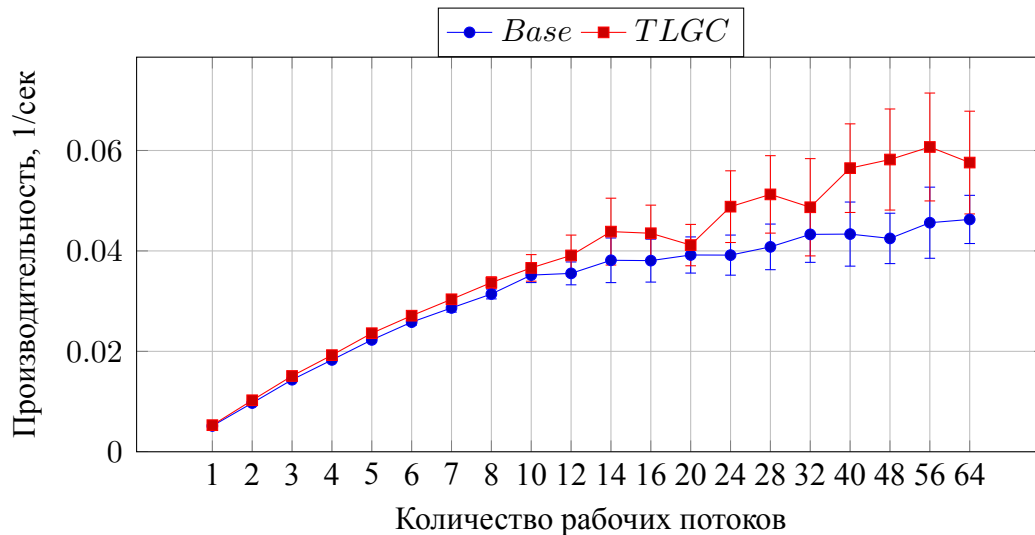
Рис. 2. Производительность теста `hibench.standalone:KMeans`

воляет увеличить итоговую пропускную способность более чем на 25 %.

Рис. 3. Производительность теста `hibench.standalone:WordCount`

Тест **WordCount** подсчитывает количество слов в корпусе англоязычных текстов. На первый взгляд кажется, что определяющим фактором производительности такого приложения должна быть эффективность работы с жестким диском, с которого считываются входные данные. Система выполняет чтение с диска в многопоточном режиме (файлы равномерно распределяются по множеству рабочих потоков), и практика показывает, что существенную долю времени работы занимает разбиение строк входного файла на отдельные слова с помощью библиотечного метода `java.lang.String.split`. Заметим, что создаваемые строки различны по длине, и поэтому важную роль играет способность системы управления памятью быстро выделять объекты варьирующихся размеров в многопоточном режиме.

Рис. 3 иллюстрирует выгоду от использования потоково-локального управления памятью для данного приложения. Благодаря лучшей масштабируемости TLGC увеличивает пропускную способность практически в два раза, позволяя эффективно использовать все доступные вычислительные ядра.

Рис. 4. Производительность теста `hibench.standalone:TeraSort`Т а б л и ц а 2. Производительность тестового набора `hibench.cluster`

Тест	<i>Base</i>	<i>TLGC</i>	Размер локальной кучи
KMeans	7.41	20.12 (271 %)	75 Mb
TeraSort	9.53	13.04 (136 %)	80 Mb
WordCount	31.17	42.16 (135 %)	80 Mb

Тест **TeraSort** сортирует входные данные, используя алгоритм внешней сортировки, что позволяет выполнять его в распределённом режиме [15]. Как показывает рис. 4, результаты измерений весьма нестабильны, а доверительные интервалы для среднего времени работы пересекаются. В таком случае мы не можем с уверенностью говорить о превосходстве одного из алгоритмов над другим. Это может объясняться как неэффективностью алгоритма потоково-локального управления памятью, так и внутренними свойствами приложения – его производительность может слабо зависеть от локальности данных.

3.2. Hibench.cluster

Режим запуска `cluster` предполагает, что вычисления запущены с помощью платформы Apache Spark, которая использует распределённое хранилище данных под управлением Apache Hadoop. При этом на каждом из вычислительных узлов запускается несколько рабочих процессов, в каждом из которых используется несколько потоков для выполнения вычислений. Промежуточные результаты пересылаются между узлами по сетевому протоколу. Заметим, что вспомогательные подсистемы – балансировщик нагрузки, система обмена сообщениями с внешними узлами и соседствующими процессами, интерфейс к распределённому хранилищу данных – существенно влияют на профиль исполнения приложения и создают дополнительную нагрузку на сборщик мусора. Таким образом, оптимальные настройки для `standalone` режима не обязательно будут оптимальны для режима работы в кластере.

Экспериментальные результаты получены на распределённом кластере, который состоял из 4 выделенных серверов, конфигурация каждого из которых совпадает с характеристиками, указанными в начале раздела 3. В табл. 2 представлены результаты измерений для базовой (*Base*) и потоково-локальной (*TLGC*) систем управления памятью. Пропускная способность измерялась в $\frac{\text{Гб}}{\text{мин} \cdot \text{число узлов}}$, больше – лучше.

Заметим, что оптимальный размер локальных куч в режиме `cluster` значительно больше,

чем в режиме standalone. Это объясняется тем, что рабочие процессы оперируют большими объёмами данных, а также «засоряют» динамическую память служебными объектами, отвечающими за работу с диском и сетью. Такие объекты принадлежат локальной компоненте, но их время жизни отличается от объектов, возникающих в процессе «полезных» вычислений тестируемого алгоритма, что затрудняет работу сборщика мусора.

Потоково-локальная система управления памятью демонстрирует пропускную способность, значительно превосходящую базовый режим на каждом из рассмотренных тестов. Таким образом, благодаря лучшей масштабируемости и локальности размещения данных TLGC показал себя как конкурентоспособный алгоритм, пригодный к применению в промышленных системах, основанных на многопроцессорном ccNUMA-оборудовании.

4. Обзор предшествующих работ

Современные объектно-ориентированные языки программирования допускают выполнение вычислений в нескольких потоках исполнения (threads), позволяя тем самым более эффективно задействовать все ресурсы вычислительной системы. Перед системой автоматического управления памятью ставится непростая задача по эффективному выделению памяти в многопроцессорной среде, которая может решаться различными путями. С одной стороны, алгоритм может минимизировать вероятность того, что одна и та же структура данных потребуется нескольким потокам одновременно, для чего потоки разбиваются на независимые друг от друга группы [16], каждая из которых получает некоторый участок памяти в эксклюзивное владение. Если каждая группа состоит только из одного потока, то синхронизация потребуется только при исчерпании памяти внутри этого участка, что позволяет найти компромисс между издержками на межпотокое взаимодействие и ростом внутренней фрагментации [17].

Некоторые системы [18] привязывают определённые участки памяти к вычислительному ядру (core-local allocation buffer, CLAB), процессору (processor-local allocation buffer, PLAB) и потоку программы (thread-local allocation buffer, TLAB). TLAB используется для выделения памяти без использования синхронизации, а в случае его исчерпания запускается алгоритм перераспределения памяти между разными уровнями: TLAB забирает память из CLAB, который, в свою очередь, использует PLAB. Важным преимуществом описанной многоуровневой системы (hierarchical allocation buffers, HABs) является то, что большая часть операций выполняется без блокировок, с помощью атомарных процессорных инструкций, что обеспечивает масштабируемость системы и низкие издержки при исполнении на многопроцессорном оборудовании.

Упомянутые подходы допускают только эффективное выделение памяти, предполагая, что сборка мусора происходит во всей куче на общих основаниях. Если же алгоритм предполагает потоково-локальное обнаружение недостижимых объектов, то речь идет о потоково-локальных кучах (thread local heaps), использующих инкрементальные алгоритмы сборки мусора. При этом динамическая память разделяется на эксклюзивно владеемую потоком область и разделяемую (глобальную кучу). За счет того, что объекты из локальной кучи недоступны другим потокам напрямую, локальная сборка не требует сложных алгоритмов синхронизации или атомарных операций, а при локальной сборке мусора останавливается только один поток, позволяя остальным частям системы продолжить свою работу.

Пионерской в данном направлении стала работа [1], предложившая использовать потоково-локальные кучи в языке Concurrent Caml Light. Важным ограничением стало то, что в локальную кучу помещались только неизменяемые с точки зрения языка объекты. Локальные кучи представляли собой выделенные области в адресном пространстве, и «глобализация» объекта сопровождалась его переносом в другую область памяти, т.н. «эвакуацией» объекта.

При анализе производительности функциональных программ, исполняющихся на много-

процессорном оборудовании, было замечено, что большое влияние на производительность оказывает локальность данных, т.е. в ряде случаев может быть выгодно минимизировать размер локальной компоненты. Особенно ярко такая оптимизация может проявить себя в системах автоматической параллелизации Haskell-подобных языков [19]. Сохраняя размер «рабочего пространства» небольшим, сборщик мусора выполняет все операции – выделение памяти, разметку достижимых объектов, зануление создаваемых объектов – с минимальной задержкой.

В исследовании [2] применяется динамический анализ графа объектов, реализуемый с помощью барьеров на запись. Классификация объектов выполняется с помощью тэгирования – с каждым объектом связан один бит, равный единице, только если объект стал разделяемым. Срабатывающий барьер выполняет однократный обход подграфа объектов, не требующий перемещения объектов и обновления указателей. К сожалению, при таком подходе глобальные объекты становятся источником фрагментации, т.к. локальный сборщик мусора не может переместить их, а значит, в ряде случаев память в локальной куче не может использоваться с максимальной эффективностью.

Динамический анализ [2] обладает определённой консервативностью – объект считается глобальным на основе достижимости через потенциально длинную цепочку указателей, при этом не принимается во внимание, был ли он действительно использован различными потоками исполнения. Работа [20] предлагает стратегию разделения объектов, которая не обладает эффектом «транзитивного заражения» и гарантирует, что любой барьер на запись будет выполнен за ограниченное время, не зависящее от числа живых объектов в программе. В более подробном анализе данного подхода [21] использовался протокол взаимодействия потоков с помощью арбитра (т.н. эвакуирующий поток) и структуры данных, запоминающей ссылки из одной локальной кучи в другую (remembered set). Практические измерения показали, что поддержание инвариантов с помощью барьеров на чтение и запись может приводить к двукратному замедлению приложения, но, по сравнению со стратегией [2], позволяет эффективнее обнаруживать локальные компоненты.

Таким образом, на сегодняшний день существует весьма мало исследований, посвящённых использованию потоково-локальных куч для управления памятью в объектно-ориентированных языках программирования. При этом современные распределённые системы, реализованные с помощью промышленных императивных языков программирования, ставят перед разработчиками алгоритмов сборки мусора всё более сложные задачи, связанные с высокими требованиями к производительности и масштабируемости системы управления памятью. Насколько известно авторам, ранее исследователи не анализировали применимость потоково-локального сборщика мусора к распределённым системам, построенным на основе технологий Apache Hadoop и Apache Spark. Данная работа в какой-то мере восполняет этот пробел.

5. Заключение

В рамках данного исследования был проведен анализ применимости системы потоково-локального управления памятью к распределённым отказоустойчивым комплексам обработки и хранения данных. Система обеспечила устойчивую работу многопоточных промышленных приложений на оборудовании со слабой моделью памяти, что говорит о зрелости и надежности представленной схемы сборки мусора. Предложенный подход позволил значительно сократить количество глобальных сборок мусора, уменьшая тем самым «время простоя» приложений.

Анализ производительности представительного набора программ выявил, что основные достоинства схемы заключаются в её масштабируемости, локальности размещения данных в динамической памяти и значительном уменьшении издержек на синхронизацию потоков между собой. Возможность с максимальной эффективностью использовать все вычислительные ядра

системы приводит к тому, что некоторые промышленные приложения получают прирост производительности более чем в два раза, оправдывая тем самым использование затратной процедуры разграничивания объектного графа.

Представленная реализация использует фиксированные параметры для конфигурирования размера локальных куч. Мы верим, что значения параметров можно вычислить с помощью адаптивных алгоритмов прямо во время исполнения программы, но представленная система перекладывает данную работу на пользователя системы, что является её недостатком.

Экспериментально подтверждено, что предложенный подход к управлению памятью позволяет улучшить производительность программ, исполняющихся на ссNUMA-оборудовании. Уменьшение паузы на глобальную сборку мусора и локальность данных приводят к повышению пропускной способности вычислительных систем, использующихся в промышленности. Таким образом, потоково-локальная организация системы управления памятью является многообещающим подходом к сборке мусора в распределённых системах обработки больших объёмов данных.

Литература

1. *Doligez D., Leroy X.* A concurrent, generational garbage collector for a multithreaded implementation of ml // Proc. 20th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, 1993. P. 113–123.
2. *Domani T., Goldshtein G., Kolodner E., Lewis E., Petrank E., Sheinwald D.* Thread-local heaps for java. // In SIGPLAN Not, 2002. P. 76–87.
3. *Apache Hadoop® official website* [Электронный ресурс] // Apache Hadoop: [сайт]. URL: <https://hadoop.apache.org/> (дата обращения: 01.01.2022).
4. *Apache Spark official website* [Электронный ресурс] // Apache Spark: [сайт]. URL: <https://spark.apache.org/> (дата обращения: 01.01.2022).
5. *Filatov A., Mikheev V.* Evaluation of thread-local garbage collection // Proc. 2020 Ivannikov Memorial Workshop (IVMEM), 2020. P. 15–21.
6. *Purdom P., Stigler S., Cheam T.* Statistical investigation of three storage allocation algorithms // BIT Numerical Mathematics. 1971. P. 187–195.
7. *Sleator D., Tarjan R.* Self-adjusting binary search trees // J. ACM. 1985. P. 682–686.
8. *Agesen O.* Gc points in a threaded environment // Technical report, SMLI TR-98-70. 1998.
9. *Dean J., Ghemawat S.* Mapreduce: Simplified data processing on large clusters // Commun. ACM, 2008. P. 107–113.
10. *Zaharia M., Chowdhury M., Franklin M., Shenker S., Stoica I.* Spark: Cluster computing with working sets // Proc. 2nd USENIX Conference on Hot Topics in Cloud Computing, 2010. P. 10.
11. *White T.* Hadoop: The Definitive Guide // O'Reilly Media, Inc., 4th ed., 2015.
12. *HiBench source repository* [Электронный ресурс] // GitHub: [сайт]. URL: <https://github.com/Intel-bigdata/HiBench> (дата обращения: 01.01.2022).
13. *MacQueen J.* Some methods for classification and analysis of multivariate observations // Berkeley Symposium on Mathematical Statistics and Probability, 1967. P. 281–297.
14. *Jones S., Launchbury J., Peyton Jones S.* Unboxed values as first class citizens // Proc. ACM Conference on Functional Programming and Computer Architecture, 1991. P. 636–666.
15. *O'Malley O.* TeraByte Sort on Apache Hadoop [Электронный ресурс] // TeraByte Sort: [сайт]. URL: <http://sortbenchmark.org/YahooHadoop.pdf> (дата обращения: 01.01.2022).
16. *Berger E., McKinley K., Blumofe R., Wilson P.* Hoard: A scalable memory allocator for multithreaded applications // SIGARCH Comput. Archit. News. 2000. P. 117–128.

17. Vee V.-Y., Hsu W.-J. A scalable and efficient storage allocator on shared memory multiprocessors // Proc. 1999 International Symposium on Parallel Architectures, Algorithms and Networks, 1999. P. 230.
18. Kirsch C., Payer H., Röck H. Hierarchical plabs, clabs, tlabs in hotspot // Proc. International Conference on Systems (ICONS), 2012.
19. Anderson A. Optimizations in a private nursery-based garbage collector // Proc. 2010 International Symposium on Memory Management, 2010. P. 21–30.
20. Mole M., Jones R., Kalibera T. A study of sharing definitions in thread-local heaps // ICIOOLPS, 2012.
21. Mole M. A study of thread-local garbage collection for multi-core systems // PhD thesis, University of Kent, 2015.

*Статья поступила в редакцию 28.02.2022;
переработанный вариант – 13.03.2022.*

Филатов Александр Юрьевич

ассистент кафедры программирования ММФ НГУ, ведущий инженер Новосибирского исследовательского центра Huawei, e-mail: a.filatov@g.nsu.ru.

Михеев Виталий Витальевич

старший преподаватель кафедры программирования ММФ НГУ, старший эксперт Новосибирского исследовательского центра Huawei, e-mail: mikheev.vitaly@huawei.com.

References

1. Doligez D., Leroy X. A concurrent, generational garbage collector for a multithreaded implementation of ml. Proc. 20th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, 1993, pp. 113–123.
2. Domani T., Goldshtein G., Kolodner E., Lewis E., Petrank E., Sheinwald D. Thread-local heaps for java. In SIGPLAN Not, 2002, pp. 76–87.
3. Apache Hadoop® official website [Electronic resource]. Apache Hadoop: [website]. URL: <https://hadoop.apache.org/> (access date: 01.01.2022).
4. Apache Spark official website [Electronic resource]. Apache Spark: [website]. URL: <https://spark.apache.org/> (access date: 01.01.2022).
5. Filatov A., Mikheev V. Evaluation of thread-local garbage collection. Proc. 2020 Ivannikov Memorial Workshop (IVMEM), 2020, pp. 15–21.
6. Purdom P., Stigler S., Cheam T. Statistical investigation of three storage allocation algorithms. *BIT Numerical Mathematics*. 1971, pp. 187–195.
7. Sleator D., Tarjan R. Self-adjusting binary search trees. *J. ACM*. 1985, pp. 682–686.
8. Agesen O. Gc points in a threaded environment. Technical report, SMLI TR-98-70. 1998.
9. Dean J., Ghemawat S. Mapreduce: Simplified data processing on large clusters.
10. Commun. ACM. 2008, pp. 107–113.
11. Zaharia M., Chowdhury M., Franklin M., Shenker S., Stoica I. Spark: Cluster computing with working sets. Proc. 2nd USENIX Conference on Hot Topics in Cloud Computing, 2010, p. 10.
12. White T. Hadoop: The Definitive Guide. O'Reilly Media, Inc., 4th ed., 2015.
13. HiBench source repository [Electronic resource]. GitHub: [website]. URL: <https://github.com/Intel-bigdata/HiBench> (access date: 01.01.2022).
14. MacQueen J. Some methods for classification and analysis of multivariate observations. Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297.

15. Jones S., Launchbury J., Peyton Jones S. Unboxed values as first class citizens. Proc. ACM Conference on Functional Programming and Computer Architecture, 1991, pp. 636–666.
16. O'Malley O. TeraByte Sort on Apache Hadoop [Electronic resource]. TeraByte Sort: [website]. URL: <http://sortbenchmark.org/YahooHadoop.pdf> (access date: 01.01.2022).
17. Berger E., McKinley K., Blumofe R., Wilson P. Hoard: A scalable memory allocator for multithreaded applications. SIGARCH Comput. Archit. News. 2000, pp. 117–128.
18. Vee V.-Y., Hsu W.-J. A scalable and efficient storage allocator on shared memory multiprocessors. Proc. 1999 International Symposium on Parallel Architectures, Algorithms and Networks, 1999, p. 230.
19. Kirsch C., Payer H., Röck H. Hierarchical plabs, clabs, tlabs in hotspot. Proc. International Conference on Systems (ICONS), 2012.
20. Anderson A. Optimizations in a private nursery-based garbage collector. Proc. 2010 International Symposium on Memory Management, 2010, pp. 21–30.
21. Mole M., Jones R., Kalibera T. A study of sharing definitions in thread-local heaps. ICPOOLPS, 2012, 21.
22. Mole M. A study of thread-local garbage collection for multi-core systems. PhD thesis, University of Kent, 2015.

Application of Thread-Local Garbage Collection to Distributed Systems for Large-Scale Data Processing

Alexander Yu. Filatov

Assistant, Novosibirsk State University, Leading engineer, Huawei Research Center (Novosibirsk, Russia), a.filatov@ng.su.ru.

Vitaly V. Mikheev

Lecturer, Novosibirsk State University, Senior expert, Huawei Research Center (Novosibirsk, Russia), mikheev.vitaly@huawei.com.

Thread-local garbage collection (TLGC) is a technique of automatic memory management that associates memory locations with a specific application thread. These memory areas (thread-local heaps) support independent processing allowing other mutator threads to be concurrently executed. Improved scalability and throughput make thread-local memory manager an attractive alternative to conventional GC algorithms. This paper discusses effectiveness of thread-local GC applied to distributed systems for large-scale data processing. Experimental results show that the proposed approach increases overall system throughput and proves that TLGC is a suitable choice for memory-intensive fault-tolerant distributed systems.

Keywords: garbage collection, JVM, thread local heaps, NUMA.

Исследование поведения относительных вкладов переменных в общую детерминацию в оцененном на основе метода выпрямления искаженных коэффициентов регрессионном уравнении

М. П. Базилевский

Для решения проблемы мультиколлинеарности в регрессионном анализе может применяться ранее разработанный автором метод выпрямления искаженных коэффициентов, основанный на построении модели полносвязной линейной регрессии. В статье для оценки степени влияния независимых переменных на зависимую переменную в полученном с помощью этого метода регрессионном уравнении предлагается использовать относительные вклады переменных в общую детерминацию. Доказано, что в таком уравнении в случае линейной функциональной зависимости входных переменных их относительные вклады в общую детерминацию равны. Тогда при сильной корреляции входных переменных их вклады распределяются примерно одинаково. Доказано, что задача оценивания полносвязной регрессии не зависит от выбора связующей переменной. Полученные результаты успешно продемонстрированы на примере моделирования внутреннего валового продукта (ВВП) России.

Ключевые слова: регрессионная модель, мультиколлинеарность, метод выпрямления искаженных коэффициентов, модель полносвязной линейной регрессии, относительные вклады переменных в общую детерминацию, ВВП России.

1. Введение

В регрессионном анализе [1, 2] одной из главных проблем является мультиколлинеарность [3–5]. Она искажает коэффициенты регрессии, что приводит к ошибочной интерпретации влияния тех или иных факторов на зависимую переменную. К настоящему времени проблема мультиколлинеарности не является до конца решенной, поэтому продолжается процесс поиск новых методов [6–8] борьбы с этим негативным явлением. В работе [9] автором была разработана модель полносвязной линейной регрессии (МПЛР), на основе которой в [10] был предложен метод выпрямления искаженных из-за мультиколлинеарности коэффициентов (МВИК). При сильной корреляции независимых переменных знаки коэффициентов в полученном с помощью этого метода регрессионном уравнении согласуются со знаками соответствующих коэффициентов корреляции с зависимой переменной. Поэтому в таком уравнении справедливо оценивать степень влияния независимых переменных на зависимую переменную с помощью известных характеристик – относительных вкладов переменных в коэффициент детерминации. Целью данной работы является исследование поведения этих характеристик и демонстрация полученных результатов на примере моделирования внутреннего валового продукта (ВВП) России.

2. Алгоритм МВИК

МПЛР является обобщением регрессии Деминга и имеет вид:

$$x_{ij} = x_{ij}^* + \varepsilon_i^{(x_j)}, \quad i = \overline{1, n}, \quad j = \overline{1, m}, \quad (1)$$

$$x_{ij}^* = a_j + b_j x_{im}^*, \quad i = \overline{1, n}, \quad j = \overline{1, m-1}, \quad (2)$$

где x_{ij} , $i = \overline{1, n}$, $j = \overline{1, m}$ – наблюдаемые значения m объясняющих (входных, независимых) переменных $x_1, x_2, \dots, x_m$; x_{ij}^* , $i = \overline{1, n}$, $j = \overline{1, m}$ – их истинные значения; n – объем выборки; a_j, b_j , $j = \overline{1, m-1}$ – неизвестные параметры; $\varepsilon_i^{(x_j)}$, $i = \overline{1, n}$, $j = \overline{1, m}$ – ошибки аппроксимации, которые вызваны неточностями при измерении значений переменных (никаких априорных сведений о вероятностной природе этих ошибок нет).

МПЛР (1), (2) оценивается с помощью взвешенного метода наименьших полных квадратов:

$$\sum_{j=1}^{m-1} \lambda_j \sum_{i=1}^n (x_{ij} - a_j - b_j x_{im}^*)^2 + \sum_{i=1}^n (x_{im} - x_{im}^*)^2 \rightarrow \min, \quad (3)$$

где λ_j , $j = \overline{1, m-1}$ – положительные весовые коэффициенты (лямбда-параметры), представляющие собой с вероятностно-статистической точки зрения отношения дисперсий ошибок переменных, т.е. $\lambda_j = \frac{\sigma_{\varepsilon^{(x_m)}}^2}{\sigma_{\varepsilon^{(x_j)}}^2}$, $j = \overline{1, m-1}$.

Переменную x_m^* в правой части равенств (2) будем называть связующей. Возникает вопрос: влияет ли выбор связующей переменной в равенствах (2) на решение задачи (3)? Ответ на него дает теорема 1.

Теорема 1. *Выбор связующей переменной в МПЛР (1), (2) не влияет на решение оптимизационной задачи (3).*

Доказательство. Пусть переменная x_1^* является связующей. Тогда в МПЛР равенства (2) будут иметь вид:

$$x_{ij}^* = a_j^0 + b_j^0 x_{i1}^*, \quad i = \overline{1, n}, \quad j = \overline{2, m}, \quad (4)$$

где a_j^0, b_j^0 , $j = \overline{2, m}$ – неизвестные параметры.

В таком случае требуется решить оптимизационную задачу

$$\sum_{i=1}^n (x_{i1} - x_{i1}^*)^2 + \sum_{j=2}^m \lambda_j^0 \sum_{i=1}^n (x_{ij} - a_j^0 - b_j^0 x_{i1}^*)^2 \rightarrow \min, \quad (5)$$

где $\lambda_j^0 = \frac{\sigma_{\varepsilon^{(x_1)}}^2}{\sigma_{\varepsilon^{(x_j)}}^2}$, $j = \overline{2, m}$.

Выразим из (4) при $j = m$ значения связующей переменной $x_{i1}^* = -\frac{a_m^0}{b_m^0} + \frac{1}{b_m^0} x_{im}^*$, $i = \overline{1, n}$, и

подставим их вместе с соотношениями дисперсий ошибок в (5):

$$\sum_{i=1}^n \left(x_{i1} + \frac{a_m^0}{b_m^0} - \frac{1}{b_m^0} x_{im}^* \right)^2 + \sum_{j=2}^{m-1} \frac{\sigma_{\varepsilon^{(x_1)}}^2}{\sigma_{\varepsilon^{(x_j)}}^2} \sum_{i=1}^n \left(x_{ij} - a_j^0 + b_j^0 \frac{a_m^0}{b_m^0} - \frac{b_j^0}{b_m^0} x_{im}^* \right)^2 +$$

$$+ \frac{\sigma_{\varepsilon^{(x_1)}}^2}{\sigma_{\varepsilon^{(x_m)}}^2} \sum_{i=1}^n \left(x_{im} - x_{im}^* \right)^2 \rightarrow \min. \quad (6)$$

Умножив выражение (6) на $\frac{\sigma_{\varepsilon^{(x_m)}}^2}{\sigma_{\varepsilon^{(x_1)}}^2}$ и обозначив $-\frac{a_m^0}{b_m^0} = a_1$, $\frac{1}{b_m^0} = b_1$, $a_j^0 - b_j^0 \frac{a_m^0}{b_m^0} = a_j$,

$\frac{b_j^0}{b_m^0} = b_j$, $j = \overline{2, m-1}$, получим выражение (3). Отсюда следует, что выбор связующей переменной в МПЛР (1), (2) не влияет на решение задачи (3). $\square$

Таким образом, в зависимости от выбора связующей переменной существует m равносильных форм записи МПЛР (1), (2). Дальнейшее изложение ведется исходя из того, что роль связующей переменной играет переменная x_m^* .

Задача (3) при известных лямбда-параметрах решается по следующему алгоритму.

1. Из нелинейной системы

$$b_p \left(D_{x_m} + \sum_{j=1}^{m-1} \lambda_j^2 b_j^2 D_{x_j} + 2 \sum_{j_1=1}^{m-2} \sum_{j_2=j_1+1}^{m-1} \lambda_{j_1} \lambda_{j_2} b_{j_1} b_{j_2} K_{x_{j_1} x_{j_2}} + 2 \sum_{j=1}^{m-1} \lambda_j b_j K_{x_j x_m} \right) =$$

$$= \left(1 + \sum_{j=1}^{m-1} \lambda_j b_j^2 \right) \left(\sum_{j=1}^{m-1} \lambda_j b_j K_{x_j x_p} + K_{x_m x_p} \right), \quad p = \overline{1, m-1},$$

с помощью предложенного в [10] численного метода находятся оценки $\tilde{l}_j$, $j = \overline{1, m-1}$.

2. Определяются оценки $\tilde{l}_j$, $j = \overline{1, m-1}$.

3. Вычисляются оценки истинных значений переменной x_m^* :

$$\tilde{x}_m = \sum_{j=1}^m A_j x_{ij}, \quad i = \overline{1, n}, \quad (7)$$

где $A_0 = - \left(1 + \sum_{j=1}^{m-1} \lambda_j \tilde{l}_j \right)^{-1}$; $A_j = \lambda_j \tilde{l}_j \left(\sum_{j=1}^{m-1} \lambda_j \tilde{l}_j \right)^{-1}$, $j = \overline{1, m-1}$;

$$A_m = \left(1 + \sum_{j=1}^{m-1} \lambda_j \tilde{l}_j \right)^{-1}.$$

В [10] установлено, что при сильной корреляции переменных $x_1, x_2, \dots, x_m$ знаки коэффициентов уравнения (7) A_j , $j = \overline{1, m-1}$ совпадают со знаками коэффициентов корреляции $r_{x_j x_m}$, $j = \overline{1, m-1}$, а $A_m > 0$. На основе этого в [10] предложен метод выпрямления искаженных коэффициентов (МВИК). Его алгоритм «Straight B» заключается в следующем.

1. Оценивается МПЛР (1), (2) при $\lambda_j = D_{x_m} / D_{x_j}$, $j = \overline{1, m-1}$ (в [10] показано, что выбор таких лямбда-параметров гарантирует максимум аддитивного коэффициента детерминации МПЛР).

2. С помощью метода наименьших квадратов (МНК) оценивается модель $y_i = c_0 + c_1 \tilde{x}_m$, $i = \overline{1, n}$, где y_i , $i = \overline{1, n}$ – значения объясняемой (выходной, зависимой) переменной; c_0, c_1 – неизвестные параметры; ε_i , $i = \overline{1, n}$ – ошибки аппроксимации.

3. В оцененную на предыдущем шаге регрессию подставляется выражение (7) и получается регрессионное уравнение:

$$\tilde{y} = \sum_{j=1}^m \theta_j x_j, \quad (8)$$

где $\theta_0 = \tilde{\gamma}$; $\theta_j = \tilde{\gamma}_j$, $j = \overline{1, m}$.

3. Относительные вклады переменных уравнения (8) в детерминацию

Коэффициент детерминации R^2 регрессии (8) находится по формуле:

$$R^2 = \sum_{j=1}^m \theta_j \frac{\sigma_{x_j}}{\sigma_y} r_{yx_j}. \quad (9)$$

При сильной корреляции переменных $x_1, x_2, \dots, x_m$ знаки коэффициентов уравнения (8) θ_j , $j = \overline{1, m}$, совпадают со знаками коэффициентов корреляции r_{yx_j} , $j = \overline{1, m}$, т.е. $\theta_j r_{yx_j} > 0$, $j = \overline{1, m}$. Тогда справедливы формулы для относительных вкладов переменных в детерминацию:

$$C_j^{\text{rel}} = \frac{100 \theta_j \frac{\sigma_{x_j}}{\sigma_y} r_{yx_j}}{R^2}, \quad j = \overline{1, m}. \quad (10)$$

Теорема 2. Если все пары объясняющих переменных $x_1, x_2, \dots, x_m$ связаны между собой линейными функциональными зависимостями и коэффициент корреляции переменных x_m и y отличен от 0, то относительные вклады C_j^{rel} , $j = \overline{1, m}$, переменных $x_1, x_2, \dots, x_m$ в общую детерминацию R^2 для уравнения (8), полученного на основе МВИК при $\lambda_j = D_{x_m} / D_{x_j}$, $j = \overline{1, m-1}$, равны.

Доказательство. Поделим первые $(m-1)$ равенств (10) на последнее:

$$\frac{C_j^{\text{rel}}}{C_m^{\text{rel}}} = \frac{\theta_j \sigma_{x_j} r_{yx_j}}{\theta_m \sigma_{x_m} r_{yx_m}}, \quad j = \overline{1, m-1}. \quad (11)$$

Поскольку $\theta_j = \tilde{\gamma}_j$, $j = \overline{1, m}$, то $\frac{\theta_j}{\theta_m} = \frac{A_j}{A_m}$, $j = \overline{1, m-1}$. Учитывая,

что $A_j = \lambda_j \tilde{l}_j$, $j = \overline{1, m-1}$, а $A_m = \left(1 + \sum_{j=1}^{m-1} \lambda_j \tilde{l}_j \right)^{-1}$, получим $\frac{\theta_j}{\theta_m} = \lambda_j \tilde{l}_j$, $j = \overline{1, m-1}$. Тогда (11) примет вид

$$\frac{C_j^{\text{rel}}}{C_m^{\text{rel}}} = \lambda_j \tilde{l}_j \frac{\sigma_{x_j} r_{yx_j}}{\sigma_{x_m} r_{yx_m}}, \quad j = \overline{1, m-1}. \quad (12)$$

С учетом того, что $\lambda_j = D_{x_m} / D_{x_j}$, $j = \overline{1, m-1}$, а при линейной функциональной зависимости переменных $x_1, x_2, \dots, x_m$ оценка $\tilde{l}_j$ равна МНК-оценке углового коэффициента модели парной линейной регрессии x_j от x_m , т.е. $\tilde{l}_j = \frac{\sigma_{x_j}}{\sigma_{x_m}}$, $j = \overline{1, m-1}$, формулу (12)

можно записать в виде:

$$\frac{C_j^{rel}}{C_m^{rel}} = \frac{D_{x_m}}{D_{x_j}} r_{x_m x_j} \cdot \frac{\sigma_{x_j}}{\sigma_{x_m}} \cdot \frac{\sigma_{x_j} r_{yx_j}}{\sigma_{x_m} r_{yx_m}} = r_{x_m x_j} \cdot \frac{r_{yx_j}}{r_{yx_m}}, \quad j = \overline{1, m-1}. \quad (13)$$

При линейной функциональной зависимости переменных $x_1, x_2, \dots, x_m$ коэффициенты корреляции $r_{x_m x_j}$, $j = \overline{1, m-1}$ могут принимать значения или $+1$, или -1 . Если $r_{x_m x_j} = 1$, то в любом случае $r_{yx_j} = r_{yx_m}$, поэтому отношение (13) принимает значение 1. Если же $r_{x_m x_j} = -1$, то $r_{yx_j} = -r_{yx_m}$, поэтому отношение (13) вновь принимает значение 1. Отсюда следует, что все относительные вклады переменных в детерминацию равны. $\square$

Следствием теоремы 2 является то, что при сильной корреляции переменных $x_1, x_2, \dots, x_m$ относительные вклады переменных в детерминацию примерно равны между собой. Таким образом, МВИК приводит к построению регрессионного уравнения (8), в котором общая детерминация примерно равномерно распределяется между всеми входящими в модель переменными.

4. Моделирование ВВП России

Проблема моделирования ключевых показателей экономики России является чрезвычайно актуальной. Одним из таких показателей является валовой внутренний продукт. Для построения регрессионной модели ВВП России были собраны статистические данные (<https://rosstat.gov.ru/>) за период 2000–2020 гг. по следующим переменным:

y – ВВП (в текущих ценах, млрд руб.);

x_1 – среднемесячная номинальная начисленная заработная плата работников по полному кругу организаций в целом по экономике РФ (руб.);

x_2 – численность занятых в возрасте 15–72 лет по РФ (тыс. чел.);

x_3 – численность безработных в возрасте 15–72 лет по РФ (тыс. чел.);

x_4 – наличие основных фондов в РФ на конец отчетного года по полной учетной стоимости (млн руб.);

x_5 – потребление электроэнергии по РФ (млн кВт·ч);

x_6 – продукция сельского хозяйства всех категорий по РФ (в фактических действовавших ценах; млрд руб.);

x_7 – количество введенных зданий жилого и нежилого назначения в РФ (тыс.);

x_8 – грузооборот железнодорожного транспорта (млрд т·км);

x_9 – оборот розничной торговли по РФ (млн руб.);

x_{10} – оборот оптовой торговли по РФ (млрд руб.);

x_{11} – инвестиции в основной капитал в РФ (млн руб.);

x_{12} – средние цены на первичном рынке жилья по РФ (на конец периода, руб. за 1 квадратный метр общей площади);

x_{13} – динамика денежной массы (M2) (млрд руб.);

x_{14} – внешняя торговля (экспорт, млн долл. США);

x_{15} – внешняя торговля (импорт, млн долл. США);

x_{16} – цена на нефть (долл. США);

x_{17} – добыча нефти в России (млн тонн).

Для однородности переменные x_{14} , x_{15} , x_{16} , измеряемые в долларах США, были переведены в рублевый эквивалент.

Найденная для переменных y , x_1 , x_2 , ..., x_{17} корреляционная матрица показала, что все они очень тесно коррелируют друг с другом. Так, например, самое малое по абсолютной величине значение коэффициента корреляции наблюдается между переменными x_3 и x_7 , равное -0,794.

Оцененная по исходной выборке с помощью МНК традиционная модель множественной линейной регрессии имеет вид:

$$\begin{aligned} \tilde{y} = & 10^6 + 67.734x_1 + 14.73x_2 + 15.508x_3 - 0.00841x_4 + 1.54x_5 + \\ & + 6.209x_6 + 1205.95x_7 + 115.685x_8 - 0.0319x_9 + 10.77x_{10} + 0.0085x_{11} - \\ & - 12.128x_{12} - 10.81x_{13} - 0.0197x_{14} - 0.0108x_{15} - 18.08x_{16} - 415.949x_{17}, \end{aligned} \quad (14)$$

(2.457) (2.676) (2.423) (-2.459) (2.609)
(1.323) (2.493) (2.09) (-2.432) (2.519) (2.774)
(-2.545) (-2.161) (-2.347) (-1.588) (-1.948) (-2.097)

для которой $R^2 = 0.999976$. В скобках под коэффициентами указаны соответствующие значения t-критерия Стьюдента.

Как и ожидалось, из-за мультиколлинеарности знаки коэффициентов при переменных x_3 , x_4 , x_9 , x_{12} , x_{13} , x_{14} , x_{15} , x_{16} , x_{17} в уравнении (11) не соответствуют экономическому смыслу задачи.

Как следует из теоремы 2, в полученном на основе наших исходных данных с помощью МВИК регрессионном уравнении все относительные вклады переменных должны быть примерно равны $\frac{100}{17}\%$ (5.88%).

Оцененная по исходной выборке на основе МВИК регрессионная модель имеет вид:

$$\tilde{y} = 51 + 0.000296x_1 + 0.00154x_2 - 0.0033x_3 + 6.236 \cdot 10^{-8}x_4 + 4.99 \cdot 10^{-5}x_5 + \dots, \quad (15)$$

где

$$\begin{aligned} \tilde{y} = & 51 + 0.000296x_1 + 0.00154x_2 - 0.0033x_3 + 6.236 \cdot 10^{-8}x_4 + 4.99 \cdot 10^{-5}x_5 + \\ & + 0.0025x_6 + 0.0587x_7 + 0.0112x_8 + 3.925 \cdot 10^{-7}x_9 + 0.00017x_{10} + 7.208 \cdot 10^{-7}x_{11} + \\ & + 0.000214x_{12} + 0.000277x_{13} + 5.361 \cdot 10^{-7}x_{14} + 8.6 \cdot 10^{-7}x_{15} + 0.00329x_{16} + 0.0537x_{17}. \end{aligned} \quad (16)$$

Поставляя (16) в (15), получим регрессионное уравнение:

$$\begin{aligned} \tilde{y} = & 7.795 + 0.146x_1 + 0.761x_2 - 1.649x_3 + 3.088 \cdot 10^{-5}x_4 + 0.0247x_5 + \\ & + 1.237x_6 + 29.044x_7 + 5.557x_8 + 0.000194x_9 + 0.0846x_{10} + 0.000357x_{11} + \\ & + 0.106x_{12} + 0.137x_{13} + 0.000265x_{14} + 0.000426x_{15} + 1.631x_{16} + 26.592x_{17}, \end{aligned} \quad (17)$$

(6.27%) (5.58%) (4.75%) (6.17%) (5.98%)
(6.1%) (5.61%) (5.86%) (6.28%) (6.2%) (6.3%)
(5.23%) (6.19%) (6.19%) (6.32%) (5.86%) (5.11%)

для которого $R^2 = 0.97764$. В уравнении (17) под коэффициентами в скобках указаны относительные вклады переменных в общую детерминацию. Как видно, все объясняющие переменные в (17) поделили общую детерминацию примерно в равных долях, что подтверждает справедливость следствия из теоремы 2.

В модели (17) знаки абсолютно всех коэффициентов при объясняющих переменных соответствуют содержательному смыслу задачи. При этом качество модели (17) лишь незначительно ниже, чем у регрессии (14), поэтому её можно использовать не только для интерпретации, но и для прогнозирования.

Доказанные в работе теоремы могут быть использованы для обобщения МВИК для регрессионных моделей с различной степенью корреляции объясняющих переменных.

Литература

1. *Arkes J.* Regression analysis: a practical introduction. Routledge, 2019. 362 p.
2. *Westfall P. H., Arias A. L.* Understanding regression analysis: a conditional distribution approach. Chapman and Hall/CRC, 2020. 514 p.
3. *Thompson C.G., Kim R.S., Aloe A.M., Becker B.J.* Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results // *Basic and Applied Social Psychology*. 2017. V. 39, № 2. P. 81–90.
4. *Yoo C., Cho E.* Effect of multicollinearity on the bivariate frequency analysis of annual maximum rainfall events // *Water*. 2019. V. 11, № 5. P. 905.
5. *Lindner T., Puck J., Verbeke A.* Misconceptions about multicollinearity in international business research: Identification, consequences, and remedies // *Journal of International Business Studies*. 2020. V. 51. P. 283–298.
6. *Giacalone M., Panarello D., Mattera R.* Multicollinearity in regression: an efficiency comparison between Lp-norm and least squares estimators // *Quality & Quantity: International Journal of Methodology*. 2018. V. 52, № 4. P. 1831–1859.
7. *Babaie-Kafaki S., Roozbeh M.* A revised Cholesky decomposition to combat multicollinearity in multiple regression models // *Journal of Stat. Comp. and Simul.* 2017. V. 87. P. 2298–2308.
8. *Dawoud I.* A new improved estimator for reducing the multicollinearity effects // *Communication in Statistics – Simulation and Computation*. 2021. P. 1–12.
9. *Bazilevskiy M. P.* Multifactor fully connected linear regression models without constraints to the ratios of variables errors variances // *Inform. and its Applic.* 2020. V. 14, № 2. P. 92–97.
10. *Bazilevskiy M. P.* Method of straightening distorted due to multicollinearity coefficients in regression models // *Informatics and its Applications*. 2021. V. 15, № 2. P. 60–65.

*Статья поступила в редакцию 05.02.2022;
переработанный вариант –13.03.2022.*

Базилевский Михаил Павлович

к.т.н., доцент кафедры математики ИрГУПС (664074, Иркутск, ул. Чернышевского, 15),
e-mail: mik2178@yandex.ru.

References

1. *Arkes J.* Regression analysis: a practical introduction. Routledge, 2019, 362 p.
2. *Westfall P. H., Arias A. L.* Understanding regression analysis: a conditional distribution approach. Chapman and Hall/CRC, 2020, 514 p.
3. *Thompson C.G., Kim R.S., Aloe A.M., Becker B.J.* Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. *Basic and Applied Social Psychology*. 2017, vol. 39, no. 2, pp. 81-90.
4. *Yoo C., Cho E.* Effect of multicollinearity on the bivariate frequency analysis of annual maximum rainfall events. *Water*. 2019, vol. 11, no. 5, pp. 905.
5. *Lindner T., Puck J., Verbeke A.* Misconceptions about multicollinearity in international business research: Identification, consequences, and remedies. *Journal of International Business Studies*. 2020, vol. 51, pp. 283-298.
6. *Giacalone M., Panarello D., Mattera R.* Multicollinearity in regression: an efficiency comparison between Lp-norm and least squares estimators. *Quality & Quantity: International Journal of Methodology*. 2018, vol. 52, no. 4, pp. 1831-1859.

7. Babaie-Kafaki S., Roozbeh M. A revised Cholesky decomposition to combat multicollinearity in multiple regression models. *Journal of Stat. Comp. and Simul.* 2017, vol. 87, pp. 2298-2308.
8. Dawoud I. A new improved estimator for reducing the multicollinearity effects. *Communication in Statistics - Simulation and Computation.* 2021, pp. 1-12.
9. Bazilevskiy M. P. Multifactor fully connected linear regression models without constraints to the ratios of variables errors variances. *Inform. and its Applic.* 2020, vol. 14, no. 2, pp. 92-97.
10. Bazilevskiy M. P. Method of straightening distorted due to multicollinearity coefficients in regression models. *Informatics and its Applications.* 2021, vol. 15, no. 2, pp. 60-65.

Researching the behavior of variables relative contributions to the total determination in regression equation estimated using the method of distorted coefficients straightening

Mikhail P. Bazilevskiy

Candidate of technical sciences, Docent, Irkutsk State Transport University (Irkutsk, Russia), mik2178@yandex.ru.

To solve the problem of multicollinearity in regression analysis a distorted coefficients straightening method developed by the author and based on the construction of fully connected linear regression model can be used. In the article, to assess the degree of independent variables influence on the dependent variable in the regression equation obtained by using this method, it is proposed to use the variables relative contributions to the total determination. It is proved that in such an equation in the case of linear functional dependence of the input variables their relative contributions to the total determination are equal. Then, with a strong correlation of the input variables, their contributions are distributed approximately in the same way. It is proved that the problem of estimating a fully connected regression does not depend on the choice of connecting variable. The obtained results have been successfully demonstrated using the example of the Russia's GDP modeling.

Keywords: regression model, multicollinearity, method for straightening distorted coefficients, fully connected linear regression model, relative contributions of variables to the total determination, GDP of Russia.

Методология выбора трасс для прокладки сетей и коммуникаций

Г. Ы. Токтошов¹

Предлагается новая методика выбора трасс для прокладки инженерных коммуникаций, учитывающая насыщенность городской застройки различными инфраструктурными объектами инженерного обеспечения, вид и назначение проектируемой коммуникации, углы поворота коммуникации вокруг заданной оси при обходе препятствий и другие ограничения. В отличие от традиционного представления, трасса для прокладки коммуникаций рассматривается как объект в трехмерном пространстве, имеющий ограничение как по плану, так и по профилю – по высоте расположения сетевых объектов. Кроме того, область, на которой предполагается разместить сеть, рассматривается как трехмерное пространство и моделируется в виде трехмерной расчетной сети. Задача оптимизации коммуникаций поставлена в виде построения минимальной гиперсети при ограничениях и требованиях, накладываемых на проектируемую коммуникацию. Предложен модифицированный метод трассировки лучей, позволяющий выбрать трассу в зависимости от угловых и линейных координат, предопределяющих положение проектируемой коммуникации в пространстве.

Ключевые слова: трасса, инженерная коммуникация, инфраструктура, граф, гиперсеть, трехмерное преобразование.

1. Введение

Сети инженерных коммуникаций современного города обслуживают населенные пункты, промышленные зоны, торговые центры и культурно-бытовые предприятия. Всё это представляет собой комплексную систему, состоящую из инженерных коммуникаций, сооружений и специальных устройств. Современная инженерная коммуникация – это сложный территориально-распределенный комплекс, который выполняет такие жизненно важные функции, как обеспечение потребителей энергетическими и водными ресурсами, средствами связи, информацией, маршрутной сетью, транспортом и другими сетевыми услугами. Следует отметить, что развитие современного города характеризуется динамичностью и противоречивостью. Чрезвычайно быстрый рост городов создает немало социальных и экономических проблем, возникает необходимость построения новых населенных пунктов и соответствующих сетей инженерных коммуникаций для их обслуживания и т.п. Другими словами, для удовлетворения потребностей населенных пунктов и промышленных зон требуется создание новых или расширение существующих сетей инженерных коммуникаций, обеспечивающих передачу целевого продукта от источников к потребителям.

Отметим, что в настоящее время территория, на которой предполагается размещение инженерной коммуникации, чрезвычайно насыщена различными сетями инженерного назначения, такими как газо-, нефте-, и водопроводы, сети электросвязи и электропередач, сети автомобильных и железнодорожных линий и т.п. Все эти сети должны быть учтены на этапе проектирования сетей, так как на данном этапе определяется взаимное расположение и взаи-

¹ Исследования выполнены в рамках государственного задания ИВМиМГ СО РАН (0251-2021-0005).

модействие элементов различных сетей. В связи с этим возникает задача оптимального пространственного размещения сетей инженерных коммуникаций в условиях насыщенности городской территории.

Одной из задач, связанной с размещением сетей инженерных коммуникаций в пространстве, является выбор трассы их прокладки, учитывающей насыщенность городской застройки, вид и назначение проектируемой коммуникации, ограничения планового и профильного характера, углы поворота коммуникации вокруг заданной оси при обходе препятствий и другие ограничения. В настоящей работе предложена новая модель пространственного размещения коммуникаций, основанная на трехмерной сетке и методах трёхмерного преобразования.

2. Задачи о выборе трассы для прокладки коммуникаций

Отметим, что существуют различные постановки задачи о выборе трасс (как дискретные, так и непрерывные) для прокладки коммуникаций различного назначения между заданными парами точек области размещения D . Например, одна из задач, связанных с прокладкой автомобильной дороги из пункта A в пункт B , имеющая минимальную стоимость, формулируется следующим образом [1]: на двумерной области $D \in R^2$ задана непрерывная или кусочно-непрерывная функция $f(x, y) \geq 0$, соответствующая стоимости строительства дороги в точке $(x, y) \in D$. Тогда стоимость строительства автомобильной дороги из точки A в точку B по выбранному маршруту μ вычисляется следующим образом:

$$C(\mu) = \int_{\mu} f(x, y) d\mu.$$

Пусть $\mu(A, B)$ – множество всех возможных маршрутов прокладки автомобильной дороги из точки A в точку B , представляющие из себя непрерывные кривые в области $D \in R^2$. Тогда задача заключается в поиске такой кривой μ_0 среди всех $\mu(A, B)$, вдоль которой функция стоимости $C(\mu)$ принимает наименьшее значение:

$$\min_{\mu_0 \in \mu(A, B)} C(\mu) = \int_{\mu_0} f(x, y) d\mu.$$

Аналогичным образом в непрерывной постановке можно сформулировать задачи поиска трассы для других типов коммуникаций.

В качестве альтернативы для выбора трассы прокладки коммуникаций используют дискретные математические модели, суть которых заключается в формировании множества всех возможных маршрутов для прокладки коммуникаций и выборе среди них наиболее оптимального по заданному критерию. В дискретной постановке задачи выбор трассы для прокладки коммуникаций из пункта A в пункт B заключается в выборе маршрута, имеющего минимальную стоимость μ_{AB} среди всех маршрутов $\mu(A, B)$ между заданными парами вершин A и B на графе $G = (V, E)$, т.е.:

$$C(\mu_{AB}) = \min_{\mu(A, B)} \left(\sum_{(i, j) \in \mu} c_{ij} \right),$$

где $\mu(A, B)$ – множество всех возможных маршрутов из A в B ;

μ_{AB} – маршрут, обеспечивающий минимальность стоимости перехода из A в B ;

c_{ij} – стоимость перехода из i в j , являющихся звеном маршрута μ_{AB} , т.е. $(i, j) \in \mu_{AB}$.

Отметим, что использование дискретной математической модели для выбора маршрутов прокладки коммуникаций более целесообразно с точки зрения сложности и реализации вычислительного алгоритма, поскольку на практике число возможных вариантов трасс для прокладки сетей является конечным. В связи с этим в настоящей работе исследование задач о выборе трасс для прокладки сетей и коммуникаций было проведено в рамках моделей теории графов и гиперсетей.

3. Анализ моделей и методов оптимизации трассы

Разработано множество моделей и методов по выбору трасс для прокладки сетей и коммуникаций различного назначения. Так, например, в [2, 3] предложены модели по снижению стоимости строительных и эксплуатационных затрат магистральных и распределительных сетей, в [4, 5, 6] – выбор структуры газотранспортной сети по критерию минимума строительных затрат на основе методов линейного и нелинейного программирования, а также алгоритма имитации отжига, в [7, 8, 9] – выбор оптимальной структуры водопроводной сети на основе моделей имитационного и вероятностного анализа, а также модели гидравлической сети, и другие. С точки зрения дискретной оптимизации все задачи, связанные с выбором трассы для прокладки сетей и коммуникаций, обладают общими свойствами, описываемыми структурными моделями, такими как тензоры и морфологические таблицы [10, 11], избыточные схемы [12, 13], альтернативные деревья [14, 15], сэндвич-графы [16], вложенные графы [17], многослойные сети [18, 19], многоуровневые комплексные сети [20] и другие. Отметим, что все существующие модели и методы оптимизации трасс предполагают размещение проектируемой сети и коммуникации на заданной поверхности без учета их иерархичности. С другой стороны, в них не учитывается пространственное положение будущей коммуникации, зависящей от существующих сетей и коммуникаций, наличия естественных и искусственных преград, ограничения планового и профильного характера и т.п. В связи с этим в качестве математической модели для описания иерархичности сетей и коммуникаций в настоящей работе предлагается использовать гиперсетевую математическую модель [21], которая относительно просто и компактно описывает вложенность проектируемой коммуникации по выбранной трассе. При этом стоит отметить, что дискретный аналог городской территории моделируется графом первичной сети (primary network, PN), структура проектируемой сети – графом вторичной сети (secondary network, SN), а вложенность вторичной сети SN в первичную сеть PN – гиперсетью (hypernet, HN). Далее приведем достаточно простое определение двухуровневой гиперсети HN, которая определяется следующим образом.

Определение. Гиперсеть $HN = (X, V, R; P, W, F)$ – это иерархический математический объект, содержащий:

на уровне первичной сети PN:

$X = (x_1, x_2, \dots, x_n)$ – множество вершин;

$V = (v_1, v_2, \dots, v_g)$ – множество возможных трасс;

$P(v): V \rightarrow 2^X$ – отображение, сопоставляющее каждому элементу $v \in V$ множество $P(v) \subseteq X$ его вершин, определяющее граф первичной сети $PN = (X, V; P)$;

на уровне вторичной сети SN:

$Y = (y_1, y_2, \dots, y_p)$ – множество узловых элементов сети;

$R = (r_1, r_2, \dots, r_m)$ – множество ребер (линейные сооружения);

$W: r \rightarrow 2^{P(F(r))}$ – отображение, сопоставляющее каждому элементу $r \in R$ подмножество $W(r) \subseteq P(F(r))$ его вершин, где $P(F(r))$ – множество вершин PN, инцидентных ветвям $F(r) \subseteq V$, определяющее граф вторичной сети $SN = (Y \subseteq X, R; W)$. На практике SN соответствует структуре проектируемой сети.

Тогда отображение $F(r): R \rightarrow 2^V$, сопоставляющее каждому ребру $r \in R$ графа вторичной сети SN определенный маршрут из ветвей $v \in V$ в графе первичной сети PN , определяет гиперсеть HN . Таким образом, вложенность инженерной коммуникации по выбранным трассам определяется гиперграфом $FS = (V, R; F)$, т.е. ветвь $v \in V$ графа PN инцидентна ребру $r \in R$ тогда и только тогда, когда ребро r проходит (реализовано) по соответствующей ветви v , а отображения графа SN в PN называются гиперсетью $HN = (PN, SN; F)$.

На рис. 1 граф первичной сети PN представлен в виде решетки (рис. 1а), граф вторичной сети SN – в виде мультиграфа (рис. 1б), а одно из возможных отображений ребер $r \in R$ графа вторичной сети SN по определенным маршрутам $F(r) \in V$ в графе первичной сети PN – в виде гиперсети $HN = (PN, SN; F)$ (рис. 1с).

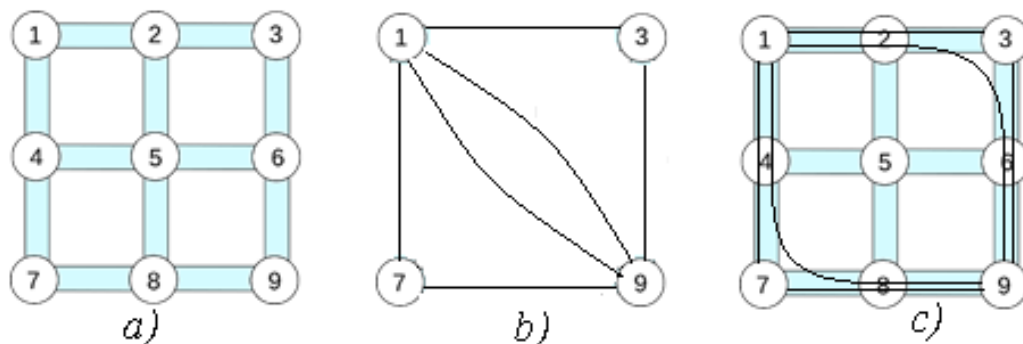


Рис. 1. Пример гиперсети

Отметим, что в нашем случае граф первичной сети PN представляет собой трехмерную расчетную сеть с шагами l_x , l_y и h , где белые кружочки (узлы) – это препятствия или запрещенные зоны, которые необходимо обходить при прокладке коммуникаций, а черные узлы (отмеченные буквами) – это свободные узлы, которые разрешено соединить трассой, находясь в одном из них. Например, на рис. 2 одна из возможных трасс между свободными узлами A и C выделена синим цветом.

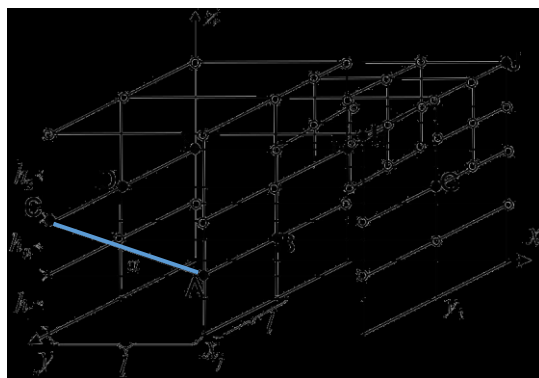


Рис. 2. Граф первичной сети PN

Таким образом, граф первичной сети PN в общем случае представляет собой дискретный аналог трехмерного пространства, в котором предполагается строить коммуникации. В зависимости от вида отображения $F: R \rightarrow 2^V$, т.е. от вида вложения вторичной сети в первичную сеть $SN \rightarrow PN$, возникают различные прикладные задачи в области проектирования и строительства сетей различного назначения.

4. Стратегия выбора маршрутов прокладки коммуникаций

Всякий маршрут μ между заданными парами точек в трехмерной сетке состоит из последовательности ветвей $v_i \in V, i=1, \dots, n$ графа первичной сети PN . Чтобы выбрать очередную i -ю ветвь $v_i \in V$, входящую в маршрут, сделаем следующее допущение: целевая (конечная) точка находится справа от начальной точки относительно оси Ox в трехмерном пространстве; линейные координаты $x_{нач}, y_{нач}, z_{нач}$ начала i -й ветви $v_k \in V$ совпадают с началом координат в местной системе $Ox_i y_i z_i$, а угловые координаты начала $\alpha_{нач} = \beta_{нач} = \gamma_{нач} = 0$.

Для координат начала $(x_{нач}, y_{нач}, z_{нач})$ и конца $(x_{кон}, y_{кон}, z_{кон})$ i -й ветви $v_k \in V$ введем следующие обозначения:

$$\Delta x = x_{нач} - x_{кон}; \quad \Delta y = y_{нач} - y_{кон}; \quad \Delta z = z_{нач} - z_{кон}$$

и предположим, что $\Delta x \geq 0$, $\Delta y \neq 0$ и $\Delta z \neq 0$, т.е. целевая точка находится справа от начальной точки относительно оси Ox в трехмерном пространстве. На рис. 3 показаны возможные варианты перехода из точки с координатами $(x_{нач}, y_{нач}, z_{нач})$ к соседней точке с координатами $(x_{кон}, y_{кон}, z_{кон})$ по принципу восьмисвязности в трехмерном пространстве.

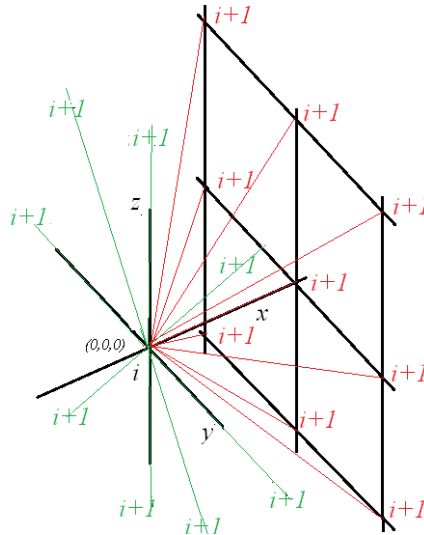


Рис. 3. Возможные направления выбора ветви

Красными линиями показаны возможные переходы из i -й местной системы координат $Ox_i y_i z_i$ в $(i+1)$ -й $Ox_{i+1} y_{i+1} z_{i+1}$ на один шаг по диагонали с одной ступенью и вдоль оси Ox . Зелеными линиями показаны возможные переходы в пределах i -й местной системы координат $Ox_i y_i z_i$, т.е. на плоскости yOz . Для решения некоторых практических задач допускается переход из данной точки к соседним, находящимся на расстоянии двух, трех или более ступеней.

Отметим, что при переходе из точки в точку необходимо соблюдать следующие ограничения и требования:

- $v_i \subset PN$ – выбранная ветвь графа первичной сети PN для прокладки ребер $r \in R$ графа вторичной сети SN должна находиться внутри области размещения;
- $v_i \not\subset PN_s$ – выбранная ветвь графа первичной сети PN для прокладки ребер $r \in R$ графа вторичной сети SN не должна проходить через запретные зоны в виде оврагов, водоемов, социальные объекты и т.п.;

- с) если $\alpha \leq \alpha_{up}$, то выбор i -й ветви $v_i \in V$ на графе первичной сети PN с допустимым углом наклона для прокладки ребер $r \in R$ графа вторичной сети SN из точки с меньшей отметкой в точку с большей отметкой разрешен, в противном случае – нет;
- д) если $\alpha \geq \alpha_{down}$, то выбор i -й ветви $v_i \in V$ на графе первичной сети PN с допустимым углом наклона для прокладки ребер $r \in R$ графа вторичной сети SN из точки с большей отметкой в точку с меньшей отметкой разрешен, в противном случае – нет;
- е) если $\alpha \leq |\alpha_{side}|$, то выбор i -й ветви $v_i \in V$ на графе первичной сети PN для прокладки ребер $r \in R$ графа вторичной сети SN горизонтально по поверхности склона разрешен, в противном случае – нет. Понятно, что если заданная область, на которой намечено строительство сетей, имеет склон, то $\{\alpha_{up}, \alpha_{down}, \alpha_{side}\} < 90^\circ$.

5. Гиперсетевая постановка задачи о выборе трассы

Пусть известны предполагаемые структуры первичной $PN = (X, V)$ и вторичной сети $SN = (Y, R)$ некоторой гиперсети HN . Тогда задача поиска трассы на графе первичной сети PN и отображения ребер графа вторичной сети SN по найденным трассам формулируется следующим образом.

Задача. Найти отображение

$$F^*(r): SN \rightarrow PN,$$

для которого

$$Q(HN) = \left(\sum_{v \in F^*(r), \forall r \in R} a(v) * T + \sum_{r \in R} b(r) \right) \rightarrow \min,$$

где $a(v)$ – стоимость ветвей первичной сети PN ;

$b(r)$ – стоимость ребер вторичной сети SN ;

при ограничениях и требованиях (а)–(е);

где $F^*(r) = \mu^*$ – оптимальный маршрут на графе первичной сети PN для прокладки ребра $r \in R$ графа вторичной сети SN ; T – композиция преобразований $T_{план}$ и $T_{профил}$ пространственной прямой (ветви) $v_i \in V$; α – угол между пространственной трассой, соединяющей некоторые пары узлов, и горизонтальной поверхностью.

Отметим, что задача построения гиперсети минимальной стоимости $Q(HN) \rightarrow \min$ даже в самой простой постановке без учёта матрицы пространственного преобразования и ограничений (а)–(е) является NP-трудной [22]. В связи с этим она может быть решена эвристическими методами, такими как генетический алгоритм, алгоритм муравьиной колонии, метод имитации отжига и т.п. В настоящей работе для решения поставленной задачи был предложен модифицированный алгоритм трассировки лучей, основанный на модели гиперсетей и методах трехмерного преобразования ветвей $v_i \in V$ графа PN , входящих в маршрут.

6. Модифицированный метод трассировки лучей

Для начала фиксируем план и профиль текущей ветви $v_i \in V$ (пространственной прямой) на соответствующие плоскости xOy и xOz . Согласно [23, 24] план пространственной прямой на плоскости xOy может быть получен путем поворота вокруг оси Y на угол β , затем поворота вокруг оси X на угол α и параллельного проецирования на плоскость $Z = 0$, а профиль на

плоскости xOz – путем поворота вокруг оси Z на угол γ , затем поворота вокруг оси X на угол α и параллельного проецирования на плоскость $Y=0$. Для этого применяются следующие виды преобразований:

$$T_{\text{план}} = T_{\text{пов. на } \beta \text{ вокр } y} \times T_{\text{пов. на } \alpha \text{ вокр } x} \times T_{\text{проец. на плоск. } z=0}$$

и

$$T_{\text{профил}} = T_{\text{пов. на } \gamma \text{ вокр } z} \times T_{\text{пов. на } \alpha \text{ вокр } x} \times T_{\text{проец. на плоск. } y=0},$$

где $T_{\text{пов. на } \alpha \text{ вокр } x}$, $T_{\text{пов. на } \beta \text{ вокр } y}$ и $T_{\text{пов. на } \gamma \text{ вокр } z}$ – матрицы преобразования вращения вокруг соответствующих осей симметрии.

Алгоритм состоит из двух ходов: прямого и обратного. В прямом ходе строятся всевозможные маршруты из начальной точки в целевую, в обратном ходе восстанавливается маршрут, соответствующий заданному критерию.

Прямой ход: пусть на $(i-1)$ -м шаге начальной точкой ветви $v_i \in V$ является $P_{i-1} = (x, y)$.

Предположим, что луч, исходящий из точки P_{i-1} , столкнется с точкой A и обратно вернется наблюдателю, т.е. точка A является препятствием для данного вида коммуникаций (не выполняется одной из условий (b)–(e)). Тогда обход препятствия через точки S_i или T_i производится по «жадной» стратегии, т.е.:

- если $(s - t) < 0$ то $P_i(x', y') = T_i(x + 1, y - 1)$, и $x' = x_i + 1, y' = y_i - 1$;
- если $(s - t) \geq 0$ то $P_i(x', y') = S_i(x + 1, y + 1)$ и $x' = x_i + 1, y' = y_i + 1$.

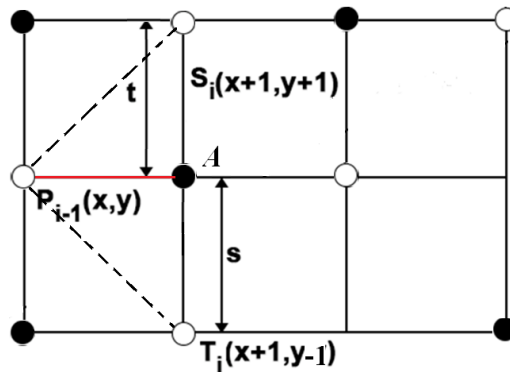


Рис. 4. Обход препятствий

В стратегии выбор точек S_i или T_i метрики s и t могут интерпретироваться как стоимость строительства коммуникаций из вершины P_{i-1} в S_i и T_i соответственно. Однако при решении практических задач вершины s и t могут быть выбраны неоднозначно, т.к. возможны случаи, когда $s - t = 0$. Это говорит о том, что иногда существует несколько различных минимальных путей между начальной и целевой точками.

Обратный ход: двигаясь в обратном направлении по заданному критерию оптимальности, например, по времени или стоимости строительства и эксплуатации коммуникаций, можно восстановить траекторию движения луча в неоднородной среде. Это производится путем восстановления прообразов вершин, входящих в маршрут, начиная от целевой точки и заканчивая начальной. Нетрудно заметить, что обход препятствия на плоскости xOz по профилю трассы осуществляется аналогичным образом, лишь с той разницей, что вместо координаты y фигурирует координата z .

В общем случае вычислительную сложность предлагаемого алгоритма можно оценить следующим образом:

прямой ход:

- для ветви $v_i \in V$, исходящей из некоторой начальной точки P_{i-1} , по N раз фиксируется план и профиль на плоскости проекций xOy и xOz , где N – число ветвей, входящих в трассу $F^*(r) \in V$. В итоге фиксация ветви $v_i \in V$ по плану осуществляется $j = 1, \dots, N$ раз, и при каждом изменении j фиксация по профилю также осуществляется $k = 1, \dots, N$ раз. Другими словами, при выполнении каждой из N итераций внешнего цикла по j внутренний цикл по k тоже выполняется N раз. Таким образом, общее количество итераций вложенного цикла по j и k равно $N \times N$, т.е. сложность преобразований пространственной прямой $v_i \in V$ на горизонтальную xOy и профильную xOz плоскости составляет $O(N^2)$;
- шаг за шагом выбираются целевые точки T_i или S_i в обход препятствия A (если они есть) согласно алгоритму трассировки лучей с вычислительной сложностью $O(N \times P)$ [25], в котором $P = |A|$, $P \leq N$ – число препятствий при выборе маршрута $F^*(r) \in V$.

обратный ход:

- список вершин графа PN , входящих в маршрут $F^*(r) \in V$, можно восстановить классическим алгоритмом Дейкстры за время $O(N^2)$.

Тогда сложность модифицированного алгоритма трассировки лучей приближенно составляет: $O(N^2) + O(N \times P) + O(N^2) \approx 3 \times O(N^2) \approx O(N^2)$.

7. Заключение

В настоящей работе была предложена новая методика выбора трасс для прокладки коммуникаций в трехмерном пространстве. В предлагаемой методике городская среда, в которой предполагается строительство коммуникаций, рассматривается как трехмерное пространство, содержащее существующие или вновь проектируемые градостроительные объекты. Как показано в работе, эти объекты, в свою очередь, либо способствуют, либо препятствуют строительству предполагаемой коммуникации. В связи с этим граф первичной сети PN целесообразно представить в виде трехмерной расчетной сетки, включающей в себя все запретные зоны и препятствия для прокладки проектируемой коммуникации. Для учета обхода препятствий в пространстве было предложено применить модели трёхмерного преобразования и теорию гиперсетей, а также методы трассировки лучей. Предложенная методика позволяет выбрать последовательность пространственных прямых, образующих маршрут из начальной точки в целевую в зависимости от их угловых и линейных координат, а также препятствия ситуационного и планового характера, которые предопределяют положение трассы в трехмерном пространстве. Так как задача является NP-трудной в сильном смысле этого слова, то она может быть решена эвристическими методами, являющимися комбинацией методов трехмерного преобразования, модифицированного алгоритма трассировки лучей и эволюционного синтеза.

Литература

1. Казаков А. Л., Лемперт А. А. Об одном подходе к решению задач оптимизации, возникающих в транспортной логистике // Автоматика и телемеханика. 2011. № 7. С. 50–57.

2. *Djebedjian B., El-Naggar M., and Shahin I.* Optimal design of gas distribution network: a case study // *Mansoura Engineering Journal (MEJ)*. 2011. V. 36, № 3. P. 35–51.
3. *Heidarifar M., and Ghasemi H.* A network topology optimization model based on substation and node-breaker modeling // *IEEE Trans. on Power Systems*. 2016. V. 31, № 1. P. 247–255.
4. *Li F., Liu Q., Guo X., and Xiao J.* A survey of optimization method for oil-gas pipeline network layout // *Proc. International Conference on Mechatronics, Electronic, Industrial and Control Engineering (MEIC)*, April 1–3, 2015, Shenyang, China. P. 257–260.
5. *Edgar T. F., Himmelblau D. M., and Bickel T. C.* Optimal design of gas transmission networks // *SPE J.* 1978. № 18. P. 96–104.
6. *Dong G., Qing T., Du R., Wang C., Li R. et al.* Complex network approach for the structural optimization of global crude oil trade system // *Journal of Cleaner Production*. 2020. V. 251. URL: <https://doi.org/10.1016/j.jclepro.2019.119366> (дата обращения: 19.01.2022).
7. *Liong S.-Y., and Atiquzzaman M.* Optimal design of water distribution network using shuffled complex evolution // *Journal of The Institution of Engineers (Singapore)*. 2004. V. 44, Is. 1. P. 93–107.
8. *Tricarico C., Gargano R., De Marinis G., Morley M. S., Kapelan Z., Savic D. A.* The influence of network topology on water distribution system performance // *Proc. 10th International Conference on Hydroinformatics (HIC)*, 2012, Hamburg, Germany, URL: https://www.researchgate.net/publication/230931471_The_Influence_of_Network_Topology_on_Water_Distribution_System_Performance (дата обращения: 28.01.2022).
9. *Lee H. M., Yoo D. G., Sadollah A., and Kim J. H.* Optimal cost design of water distribution networks using a decomposition approach // *Engineering Optimization*. 2016. V. 48, № 12. P. 2141–2156. URL: <https://doi.org/10.1080/0305215X.2016.1157689> (дата обращения: 03.02.2022).
10. *Лебедев В. В., Деревяшкин В. М.* Тензорная модель сети связи // *Вестник СибГУТИ*. 2014. № 2. С. 50–56.
11. *Одрин В. М.* Метод морфологического анализа технических систем. М.: ВНИИПИ, 1989. 312 с.
12. *Стенников В. А., Чемезов А. А.* Применение алгоритма перебора деревьев и метода имитации отжига для схемно-структурной оптимизации тепловых сетей // *Программные продукты и системы*. 2018. № 2 (31). С. 387–395.
13. *Наумов И. В., Ямщикова И. В.* Математическое обоснование выбора оптимизационной модели трассировки электрической сети // *Евразийский Союз Ученых (ЕСУ)*. 2015. № 7 (16). С. 123–127.
14. *Степанов В. П.* Оптимизация маршрутов на дорожной сети // *Наука и образование*. 2012. № 5. С. 1–12.
15. *Акимов С. В.* Модель морфологического множества уровня идентификации // *Труды учебных заведений связи*. 2005. № 172. С. 120–135.
16. *Golumbic M. C., Kaplan H., Shamir R.* Graph sandwich problems // *Journal of Algorithms*. 1995. № 19 (3). P. 449–473. DOI:10.1006/jagm.1995.1047.
17. *Poulovassilis A., Levene M.* A nested-graph model for the representation and manipulation of complex objects // *J. ACM Trans. Inf. Syst.* 1994. V. 12. P. 35–68.
18. *Orlowski S., Koster A. M. C. A., Raack C., Wessäly R.* Two-layer network design by branch-and-cut featuring MIP-based heuristics // *Proc. 3rd International Network Optimization Conference (INOC)*, Spa, Belgium, 2007. P. 114–119.
19. *Capone A., Carello G., Matera R.* Multi-layer network design with multicast traffic and statistical multiplexing // *Proc. IEEE Global Telecommunications Conference (IEEE GLOBECOM)*, Washington, USA, 2007. P. 2565–2570.
20. *Kurant M., Thiran P.* Layered complex networks // *J. Phys. Rev. Lett.* 2006. V. 96. P. 1–4.
21. *Попков В. К.* О моделировании городских транспортных систем гиперсетями // *Автоматика и телемеханика*. 2011. В. 72, № 6. С. 179–189.

22. Токтошов Г. Ы., Юргенсон А. Н., Мигов Д. А. О сложности задач оптимизации сетей инженерных коммуникаций // Т-Comm: Телекоммуникации и транспорт. 2020. Т. 14, № 9. С. 17–23.
23. Chen B., and Kaufman A. 3D volume rotation using shear transformations // Graphical Models. 2000. V. 62, Is. 4. P. 308–322.
24. Schmitt A., Muller H., and Leister W. Ray Tracing Algorithms – Theory and Practice // Theoretical Foundations of Computer Graphics and CAD. Jan 1988. P. 997–1030.
25. Трассировка лучей [Электронный ресурс]. URL: https://studme.org/156199/informatika/trassirovka_luchey (дата обращения: 20.03.2022).

Статья поступила в редакцию 15.02.2022;
переработанный вариант – 21.03.2022.

Токтошов Гулжигит Ысакович

к.т.н., доцент кафедры ММиЦРБС СибГУТИ (630102, Новосибирск, ул. Кирова, 86);
научный сотрудник лаборатории СМиО ИВМиМГ СО РАН (630090, Новосибирск, просп. Академика Лаврентьева, 6), e-mail: tgi_tok@rambler.ru.

References

1. Kazakov A. L., Lempert A. A. *Ob odnom podhode k resheniju zadach optimizacii, vozni-kajushhih v transportnoj logistike* [On one approach to solving optimization problems arising in transport logistics]. *Avtomatika i telemekhanika*. 2011, no. 7, pp. 50-57.
2. Djebedjian B., El-Naggar M., and Shahin I. Optimal design of gas distribution network: a case study. *Mansoura Engineering Journal (MEJ)*. 2011, vol. 36, no. 3, pp. 35-51.
3. Heidarifar M., and Ghasemi H. A network topology optimization model based on substation and node-breaker modeling. *IEEE Trans. on Power Systems*. 2016, vol. 31, no. 1, pp. 247-255.
4. Li F., Liu Q., Guo X., and Xiao J. A survey of optimization method for oil-gas pipeline network layout. *Proc. International Conference on Mechatronics, Electronic, Industrial and Control Engineering (MEIC)*, April 1-3, 2015, Shenyang, China, pp. 257-260.
5. Edgar T. F., Himmelblau D. M., and Bickel T. C. Optimal design of gas transmission networks. *SPE J.* 1978, no. 18, pp. 96-104.
6. Dong G., Qing T., Du R., Wang C., Li R. et al. Complex network approach for the structural optimization of global crude oil trade system. *Journal of Cleaner Production*. 2020, vol. 251. URL: <https://doi.org/10.1016/j.jclepro.2019.119366> (access date: 19.01.2022).
7. Liong S.-Y., and Atiquzzaman M. Optimal design of water distribution network using shuffled complex evolution. *Journal of The Institution of Engineers (Singapore)*. 2004, vol. 44, is. 1, pp. 93-107.
8. Tricarico S., Gargano R., De Marinis G., Morley M. S., Kapelan Z., Savic D. A. The influence of network topology on water distribution system performance. *Proc. 10th International Conference on Hydroinformatics (HIC)*, 2012, Hamburg, Germany, URL: https://www.researchgate.net/publication/230931471_The_Influence_of_Network_Topology_on_Water_Distribution_System_Performance (access date: 28.01.2022).
9. Lee H. M., Yoo D. G., Sadollah A., and Kim J. H. Optimal cost design of water distribution networks using a decomposition approach. *Engineering Optimization*. 2016, vol. 48, no. 12, pp. 2141-2156. URL: <https://doi.org/10.1080/0305215X.2016.1157689> (access date: 03.02.2022).
10. Lebedjancev V. V., Derevjashkin V. M. *Tenzornaja model' seti svyazi* [Tensor model of a communication network]. *Vestnik SibGUTI*. 2014, no. 2, pp. 50-56.
11. Odrin V. M. *Metod morfologicheskogo analiza tehnikeskikh system* [Method of morphological analysis of technical systems]. Moscow: VNIPI, 1989, 312 p.

12. Stennikov V. A., Chemezov A. A. *Primenenie algoritma perebora derev'ev i metoda imi-tacii otzhiga dlja shemno-strukturnoj optimizacii teplovyh setej* [Application of tree enumeration algorithm and annealing simulation method for scheme-structural optimization of heat networks]. *Programmnye produkty i sistemy*. 2018, no. 2 (31), pp. 387-395.
13. Naumov I. V., Jamshhikova I. V. *Matematicheskoe obosnovanie vybora optimizacionnoj modeli trassirovki jelektricheskoy seti* [Mathematical substantiation of the choice of an optimization model for tracing an electrical network]. *EvrAzijskij Sojuz Uchenyh (ESU)*. 2015, no. 7 (16), pp. 123-127.
14. Stepanov V. P. *Optimizacija marshrutov na dorozhnoj seti* [Route optimization on the road network]. *Nauka i obrazovanie*. 2012, no. 5, pp. 1-12.
15. Akimov S. V. *Model' morfologicheskogo mnozhestva urovnja identifikacii* [Model of the morphological set of the level of identification]. *Trudy uchebnyh zavedenij svyazi*. 2005, no. 172, pp. 120-135.
16. Golumbic M. C., Kaplan H., Shamir R. Graph sandwich problems. *Journal of Algorithms*. 1995, no. 19 (3), pp. 449-473. DOI:10.1006/jagm.1995.1047.
17. Poulouvassilis A., Levene M. A nested-graph model for the representation and manipulation of complex objects. *J. ACM Trans. Inf. Syst.* 1994, vol. 12, pp. 35-68.
18. Orlowski S., Koster A. M. C. A., Raack C., Wessäly R. Two-layer network design by branch-and-cut featuring MIP-based heuristics. Proc. 3rd International Network Optimization Conference (INOC), Spa, Belgium, 2007, pp. 114-119.
19. Capone A., Carello G., Matera R. Multi-layer network design with multicast traffic and statistical multiplexing. Proc. IEEE Global Telecommunications Conference (IEEE GLOBECOM), Washington, USA, 2007, pp. 2565-2570.
20. Kurant M., Thiran P. Layered complex networks. *J. Phys. Rev. Lett.* 2006, vol. 96, pp. 1-4.
21. Popkov V. K. *O modelirovanii gorodskih transportnyh sistem gipersetjami* [On the modeling of urban transport systems by hypernets]. *Avtomatika i telemekhanika*. 2011, vol. 72, no. 6, pp. 179-189.
22. Toktoshov G. Y., Jurgenson A. N., Migov D. A. *O slozhnosti zadach optimizacii setej inzhenernyh kommunikacij* [On the complexity of engineering network optimization problems] *T-Comm: Telekomunikacii i transport*. 2020, vol. 14, no. 9, pp. 17-23.
23. Chen B., and Kaufman A. 3D volume rotation using shear transformations. *Graphical Models*. 2000, vol. 62, is. 4, pp. 308-322.
24. Schmitt A., Muller H., and Leister W. Ray Tracing Algorithms - Theory and Practice. *Theoretical Foundations of Computer Graphics and CAD*. Jan 1988, pp. 997-1030.
25. Trassirovka lucej [Ray tracing] [Electronic resource]. URL: https://studme.org/156199/informatika/trassirovka_luchey (access date: 20.03.2022).

The route choosing methodology for networks and communications laying

Gulhzigit Y. Toktoshov

Candidate of technical sciences, Docent, Siberian State University of Telecommunications and Information Sciences (SibSUTIS), Researcher, The Institute of Computational Mathematics and Mathematical Geophysics SB RAS (Novosibirsk, Russia), tgi_tok@rambler.ru.

In this paper a new route selecting method for laying engineering communications taking into account the urban development saturation with various infrastructural objects, designed communications type and purpose, communications rotation angles around a given axis when avoiding obstacles, and other restrictions is proposed. The communications laying routes unlike the traditional representation are considered as three-dimensional object in space which has a restriction according both plan and profile. In addition, the communications placement region is considered as a three-dimensional space and is modeled as a three-dimensional computational network. The communications optimization task in the form of minimal hypernet structure with restrictions and requirements imposed on the designed communications is proposed. A modified ray tracing method enabling to choose a path depending on the angular and linear coordinates predetermining the communications position in space is proposed.

Keywords: route, communications, infrastructure, graph, hypernet, 3D transformation.