

**ИЗВЕСТИЯ ВЫСШИХ УЧЕБНЫХ ЗАВЕДЕНИЙ
ПОВОЛЖСКИЙ РЕГИОН
ТЕХНИЧЕСКИЕ НАУКИ**

№ 4 (20)

2011

СОДЕРЖАНИЕ

**ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ
ТЕХНИКА И УПРАВЛЕНИЕ**

<i>Волчихин В. И., Зинкин С. А.</i> Абстрактное и структурное моделирование сетей хранения и обработки данных	3
<i>Лебедев В. Б., Паришина Е. В., Песошин В. А.</i> Эффективная модель анализа данных на основе линейных векторных решеток	19
<i>Данилов А. М., Лапшин Э. В., Беликов Г. Г., Лебедев В. Б.</i> Методологические принципы организации многопоточковой обработки данных с распараллеливанием вычислительных процессов	26
<i>Архангельский А. А., Гурин Е. И., Савельев Б. А.</i> Подключение специализированных процессоров к персональному компьютеру с применением интерфейса USB.....	35
<i>Щербаков М. А.</i> Итерационный метод оптимальной нелинейной фильтрации изображений	43
<i>Горбаченко В. И., Матвеева Н. О., Гурин Е. И.</i> Распараллеливание нейросетевого алгоритма решения краевых задач математической физики на многоядерных процессорах NVIDIA.....	57
<i>Малынкин К. В., Мухин И. Б.</i> Ситуационная адаптация пространственных данных для решения бортовых геоинформационных задач	66
<i>Чумаченко Е. Н., Кулагин В. П., Хвостиков С. А.</i> Об улучшении качества прогнозирования распространения растительности на территории России по данным космического зондирования (на примере модели SEVER).....	76
<i>Гай В. Е., Утробин В. А.</i> Методика выделения в звуковом сигнале участков речевой активности	85
<i>Бистерфельд О. А.</i> Аналитико-имитационный метод оценки временных характеристик для различных вариантов представления отношений категоризации в реляционных базах данных.....	95
<i>Финогеев А. Г., Овечкин Р. М.</i> Человеко-машинное взаимодействие в системе мониторинга и поддержки принятия решений в области торговли кредитными деривативами	106
<i>Смогунов В. В., Митрохина Н. Ю.</i> Системный анализ методов проектирования автомобильных дорог	116

ЭЛЕКТРОНИКА, ИЗМЕРИТЕЛЬНАЯ ТЕХНИКА И РАДИОТЕХНИКА

<i>Байдаров С. Ю., Комаров В. В., Паршуков М. Ю., Светлов А. В.</i> Методика обработки результатов измерений параметров операционного усилителя.....	128
<i>Куликовский К. Л., Якунина В. В.</i> Ультразвуковая информационно-измерительная система объемного расхода чистой нефти	141
<i>Новрузов И. И., Комаров В. В.</i> Исследование электромагнитных и тепловых полей в ближней зоне контактного волноводного аппликатора	151
<i>Хлынов М. А., Баранов В. Н., Рашиев Р. Р.</i> Принципы построения и аппаратная реализация многофункциональных фототерапевтических устройств	159
<i>Ашанин В. Н., Чувькин Б. В., Тер-Аракелян Р. Г.</i> Оптимизация проектирования Σ Δ -модуляторов высоких порядков по критерию устойчивости.....	167

МАШИНОСТРОЕНИЕ И МАШИНОВЕДЕНИЕ

<i>Кревчик В. Д., Симонов Н. П., Соколов А. В.</i> Кавитационный механизм управления качеством поверхностного слоя деталей при использовании смазывающе-охлаждающих жидкостей с нанодобавками в условиях звукокапиллярного эффекта	180
<i>Артемов И. И., Кревчик В. Д., Симонов Н. П.</i> Квантовый механизм образования нанокластеров в поверхностном слое металла с цепочкой микротрещин в условиях кавитационного режима звукокапиллярного эффекта	192
<i>Добросельский К. Г.</i> Теплофизические и динамические характеристики начального участка вертикальной низкоскоростной струи в поперечном потоке	205
<i>Грабовский А. А.</i> Анализ уравниваемости двигателей внутреннего сгорания с кривошипно-шатунным механизмом со сдвоенными кинематическими связями.....	214
<i>Брякин А. Л., Кожевников В. В., Кочкин С. В.</i> Двухкоординатный датчик положения	225

ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

УДК 681.324

В. И. Волчихин, С. А. Зинкин

АБСТРАКТНОЕ И СТРУКТУРНОЕ МОДЕЛИРОВАНИЕ СЕТЕЙ ХРАНЕНИЯ И ОБРАБОТКИ ДАННЫХ

Аннотация. Рассматриваются проблемы использования формальных представлений распределенных процессов и объектов, взаимодействующих через общее структурированное пространство памяти. Использован алгебраический подход к определению операционной семантики распределенных систем хранения и обработки данных, основанный на определении данных систем сетями абстрактных машин. Решение поставленных задач основано на интеграции методов искусственного интеллекта с методами формально определенного распределенного поведенческого моделирования. На формальном уровне описан переход от централизованных систем внешнего хранения и обработки данных к функционально и топологически децентрализованным сетям.

Ключевые слова: хранение данных, обработка данных, формальное определение распределенных процессов, алгебраический подход, сети абстрактных машин, распределенные поведенческие модели, топологически децентрализованные сети.

Abstract. The paper considers the use of formal representations of distributed processes and objects that interact through a common structured memory space. The authors resort to an algebraic approach to the definition of an operating semantics of distributed systems, data storage and processing, based on the definition of data networks with systems of abstract machines. The solution of these problems is based on integration of artificial intelligence techniques with formally defined distributed behavioral model. At the formal level the researchers describe the transition from centralized systems of external data storage and processing to functionally and topologically decentralized networks.

Key words: data storage, data processing, formal representations of distributed processes, algebraic approach, networks of abstract machines, distributed behavioral models, topologically decentralized networks.

Введение

В сетях внешних запоминающих устройств (ВЗУ) совокупность сетевых дисковых модулей может быть представлена некоторым коллективно используемым распределенным ресурсом. Доступ к подобному ресурсу с несколькими единицами осуществляется с помощью запросов различных видов. Различие запросов характерно, например, для распределенных RAID-массивов и реплицированных распределенных баз данных, реализуемых на базе сетей ВЗУ.

Вопросы физической реализации систем и сетей хранения и обработки данных были рассмотрены, например, в [1, 2], однако в этих и других работах рассматривались в основном технические аспекты построения систем и сетей хранения данных и практически не затрагивались вопросы связи процессов абстрактного и структурного проектирования.

Логико-алгебраические модели в форме сетей абстрактных машин (CeAM) [3–5] являются удобным средством для построения имитационных поведенческих моделей систем и сетей внешнего хранения и обработки данных. Формальная определенность данных моделей позволяет получить исполняемые спецификации, необходимые для практической реализации систем.

1. Выбор необходимого уровня абстракции при моделировании распределенного ресурса – многофункциональной системы внешней памяти, подключаемой к вычислительной сети

В качестве примера ресурса R с несколькими единицами рассмотрим многофункциональную систему внешней памяти, подключаемую к вычислительной сети. Данная система может выполнять функции RAID-массива или хранить реплицированные или различные базы данных. Актуальной является также решаемая при этом задача функционально-топологической децентрализации систем и сетей хранения и обработки данных. Каждый центр обслуживания представляет собой локальную подсистему хранения и обработки данных. В качестве характерного примера подобного центра обслуживания выберем ВЗУ на магнитных дисках со встроенной распределенной кэш-памятью и модулями распределенного процессора базы данных [3].

Построение логико-алгебраических моделей многофункциональной системы внешней памяти основано на принципах, изложенных в работах [4, 5]. В отличие от перечисленных работ, в настоящей работе решается задача функционально-топологической децентрализации архитектуры системы хранения и обработки данных на основе преобразования формально определенных имитационных поведенческих моделей.

При проектировании систем и сетей хранения и обработки данных представляет интерес построение выражений для модулей сетей CeAM как с использованием квантифицированных операторов выбора $\exists!$, $\exists!!$, $\tilde{\forall}$ и $\tilde{\forall}!!$ [4, 5], так и без них. Применение квантифицированных операторов, а также дополнительных «связывающих» предикатов может привести к существенному уменьшению необходимого числа применяемых модулей CeAM и упрощению описывающих выражений. Учитывая, что формализмы сетей абстрактных машин мы предлагаем использовать в качестве непосредственно интерпретируемых спецификаций при создании нового аппаратного, микропрограммного и программного обеспечения систем хранения и обработки данных, следует рассмотреть различные формы записи выражений в алгебре модулей CeAM.

Дополнительно отметим, что все квантифицированные операторы выбора являются блокирующими, т.е. при несоблюдении условий их выполнения работа агента-сервера по реализации модуля блокируется. Так, работа модуля блокируется, если при попытке выполнения оператора $\exists!$ (оператора выборки произвольного кортежа из области истинности предиката) в области

истинности предиката, описываемого выражением справа, нет ни одного кортежа. Работа модуля, использующего оператор $\tilde{\exists}!!$, блокируется, если имеющийся в области истинности предиката кортеж не является единственным или эта область вообще не содержит ни одного кортежа. Оператор $\tilde{\forall}$ выборки всех кортежей из области истинности предиката блокирует работу модуля, если область истинности предиката пуста, т.е. не содержит ни одного кортежа, а оператор $\tilde{\forall}!!$ также блокирует работу модуля, если в момент проверки область определения предиката не совпадает с областью его истинности. Во всех случаях подразумевается, что предикат описывается выражением, стоящим справа от символа квантифицированного оператора.

2. Абстрактная и структурная модели топологически централизованной системы хранения и обработки данных

Рассмотрим абстрактный распределенный ресурс R со многими единицами. Определим данный ресурс сетью абстрактных машин, формализующей процессы, происходящие в многофункциональной системе ВЗУ, структура которой представлена на рис. 1. Здесь приведена система коммутаций, подключаемая через серверы А, В и С к глобальной сети системы ВЗУ, где СА – сетевые адаптеры; УУ – устройства управления сетевой дисковой памятью; УУОИ – устройства управления обменом информацией; УМ – управляющие модули; УУКП – устройства управления кэш-памятью; МКП – модуль кэш-памяти; МПБД – модули, выполняющие некоторые функции ячеек процессора базы данных, встроенных в систему ВЗУ; УУМПБД – устройства управления для МПБД. Система включает также четыре коммутатора, или коммуникационные сети, способные соединять порты по принципу «каждый – с каждым».

Многофункциональная система ВЗУ содержит пять коммуникационных сетей, или коммутаторов, посредством которых устанавливаются соединения между группами устройств. Для удобства свяжем наименования основных групп устройств с предметными переменными t, v, c и z , принимающими значения в множествах $T = \{t_1, t_2, \dots, t_n\}$, $V = \{v_1, v_2, \dots, v_q\}$, $Y = \{y_1, y_2, \dots, y_m\}$, $C = \{c_1, c_2, c_3\}$ и $Z = \{z_1, z_2, \dots, z_p\}$ соответственно.

Каждой единице распределенного ресурса R поставим в соответствие процесс обращения к модулю дисковой памяти или накопителю на магнитных дисках (НМД). Определим предметную переменную t , пробегающую по элементам множества «жетонов» $T = \{t_1, t_2, \dots, t_n\}$, обладание которыми позволяет запросам различных типов воспользоваться единицами ресурса R .

В сети абстрактных машин множество T будет соответствовать также и множеству НМД. Таким образом, при построении логико-алгебраической модели для функционально-централизованной системы хранения данных за основу построения сети абстрактных машин взяты процессы устройств хранения данных – модулей ВЗУ на магнитных дисках. В модели дисковые модули ВЗУ будут представлены значениями предметной переменной t , пробегающей по множеству $T = \{t_1, t_2, \dots, t_n\}$. Тогда верхний индекс i при имени каждого модуля будет свидетельствовать о причастности данного модуля к процессу i -го модуля ВЗУ, в выполнении которого участвуют все остальные устройства.

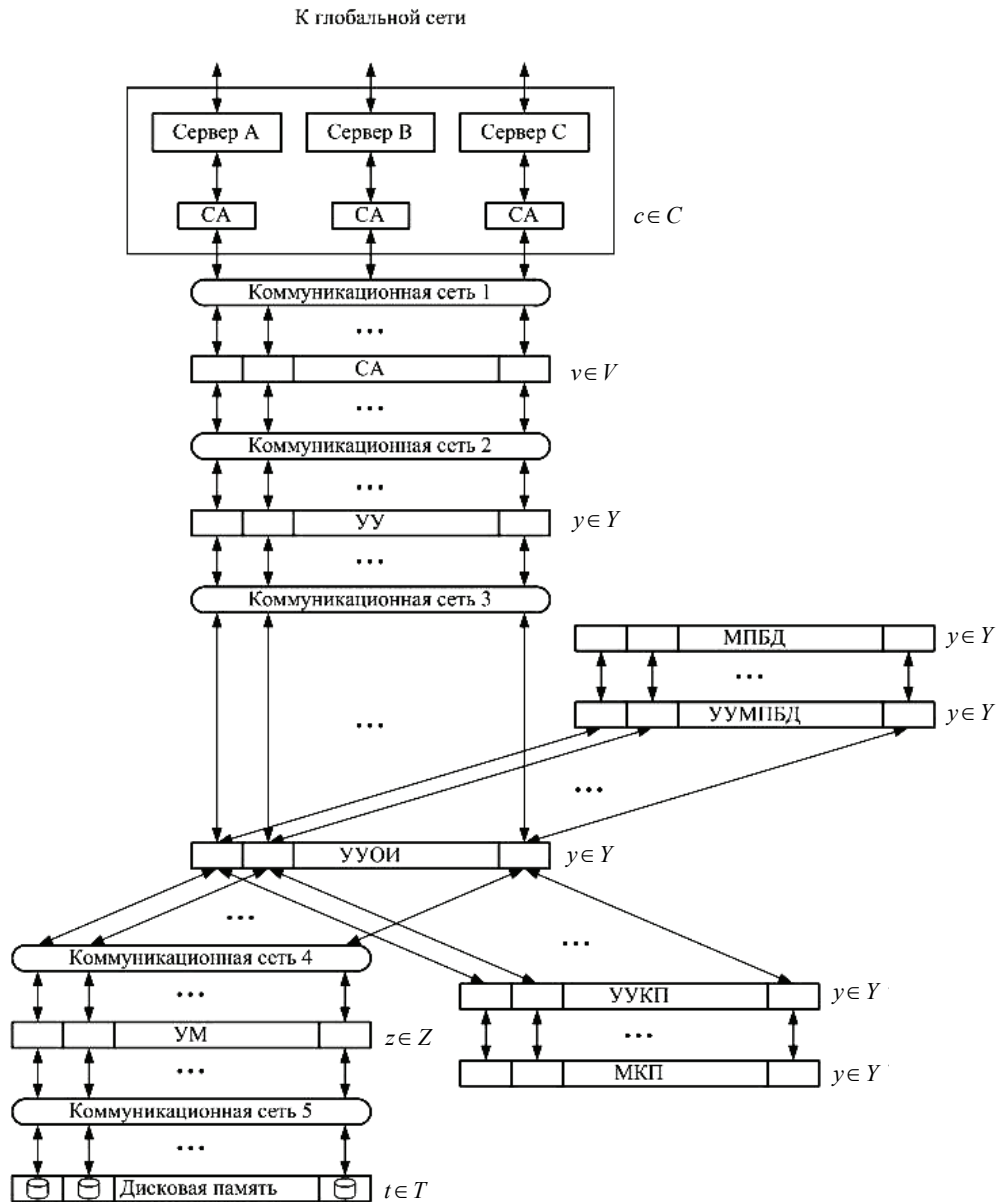


Рис. 1. Многофункциональная система ВЗУ, подключаемая к сети

Управляющие модули абстрактных машин, составляющие общую сеть СеАМ, связываются с модулями, которые будут приведены далее, посредством унарного предиката $f_{work}(t)$. При реализации сети СеАМ сетью виртуальных машин предикат $f_{work}(t)$ представляется информационным объектом в виртуальном FS-пространстве функций и предикатов.

Модули СеАМ $m_{zapkp_begin}^{(i)}$ и $m_{zapkp_end}^{(i)}$ реализуют начальную и конечную фазы операции записи данных в кэш-память МКП(y), $y \in Y$, через сети коммуникаций 1, 2, 3, 4, сетевые адаптеры СА(v), $v \in V$, устройства управления УУ(y), устройства управления обменом информацией УУОИ(y), устройства управления кэш-памятью УУКП(y):

$$\begin{aligned}
 m_{zapkp_begin}^{(i)} &= [f_k(c_j)]([f_{work}(t_i) \& (f_{oper}(t_i) = w) \& f_{st}(t_i)]([\exists! v \in V] f_{sa}(v)] \\
 &([\exists! y \in Y] f_{uu}(y)]([f_{uuoi}(y)]([f_{uukp}(y)]([f_{mkp}(y)] \\
 &(\{f_k(c_j) \leftarrow false, f_{st}(t_i) \leftarrow false, f_{sa}(v) \leftarrow false, f_{uu}(y) \leftarrow false, \\
 &f_{uuoi}(y) \leftarrow false, f_{uukp}(y) \leftarrow false, f_{mkp}(y) \leftarrow false, f_{zapkp}(t_i) \leftarrow true, \\
 &f_{uusa}(t_i, y, v) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \\
 m_{zapkp_end}^{(i)} &= [f_{zapkp}(t_i)]([\exists! (t_i \in T, y \in Y, v \in V) f_{uusa}(t, y, v)] \\
 &(\{f_k(c_j) \leftarrow true, f_{sa}(v) \leftarrow true, f_{uu}(y) \leftarrow true, f_{uuoi}(y) \leftarrow true, \\
 &f_{uukp}(y) \leftarrow true, f_{mkp}(y) \leftarrow true, f_{zapkp}(t_i) \leftarrow false, \\
 &f_{poisk}(t_i) \leftarrow true\} \vee R^E) \vee R^E).
 \end{aligned}$$

В выражениях для модулей $m_{zapkp_begin}^{(i)}$ и $m_{zapkp_end}^{(i)}$ используются следующие предикаты и функция f_{oper} :

$f_{work}(t)$ – предикат, определяющий связь сети, представляющей ресурс R , с управляющей сетью абстрактных машин;

$f_{oper}(t)$ – функция, область значений которой составляют две константы w и r , определяющие в модели операции записи и чтения данных;

$f_{st}(t)$ – предикат, определяющий готовность процессов обращения к дисковым модулям;

$f_{sa}(v)$ – предикат, определяющий готовность согласующих СА;

$f_{uu}(y)$ – предикат, определяющий готовность УУ;

$f_{uuoi}(y)$ – предикат, определяющий готовность УУОИ;

$f_{uukp}(y)$ – предикат, определяющий готовность УУКП;

$f_{mkp}(y)$ – предикат, определяющий готовность МКП;

$f_k(c)$ – предикат, определяющий готовность серверных СА;

$f_{zapkp}(t)$ – каузальный предикат модуля;

$f_{uusa}(t, y, v)$ – предикат, связывающий процесс запроса с УУ и СА;

$f_{poisk}(t)$ – каузальный предикат модуля.

Модули $m_{poisk_begin}^{(i)}$ и $m_{poisk_end}^{(i)}$ реализуют операцию установки-поиска нужного цилиндра, дорожки и сектора в дисковом модуле сети ВЗУ:

$$\begin{aligned}
 m_{poisk_begin}^{(i)} &= [f_{poisk}(t_i)]([\exists! (t_i \in T, y \in Y, v \in V) f_{uusa}(t, y, v)] \\
 &([f_{uu}(y)]([f_{uuoi}(y)]([\exists! z \in Z] f_{um}(z)]([f_d(t_i)] \\
 &(\{f_d(t_i) \leftarrow false, f_{poisk}(t_i) \leftarrow false, \\
 &f_p(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \\
 m_{poisk_end}^{(i)} &= [f_p(t_i)](\{f_p(t_i) \leftarrow false, f_{pered}(t_i) \leftarrow true\} \vee R^E).
 \end{aligned}$$

В выражении для данных модулей дополнительно используются следующие предикаты:

$f_{um}(z)$ – предикат, определяющий готовность УМ;
 $f_d(t)$ – предикат, определяющий готовность дискового модуля НМД;
 $f_p(t)$ – каузальный предикат модуля;
 $f_{pered}(t)$ – каузальный предикат модуля.

С помощью модулей $m_{pered_begin}^{(i)}$ и $m_{pered_end}^{(i)}$ в системе ВЗУ реализуется операция переписи данных из МКП в дисковый модуль сети ВЗУ, после чего сервер, выдавший запрос на запись, извещается об окончании операции. Выражения для модулей $m_{pered_begin}^{(i)}$ и $m_{pered_end}^{(i)}$ имеют следующий вид:

$$\begin{aligned} m_{pered_begin}^{(i)} = & [f_{pered}(t_i)]([\exists!(t_i \in T, y \in Y, v \in V) f_{uusa}(t, y, v)] \\ & ([f_{uu}(y)]([f_{uuoi}(y)]([f_{uukp}(y)]([f_{mkp}(y)]([\exists!z \in Z) f_{um}(z)] \\ & (\{f_{pered}(t_i) \leftarrow false, f_{uu}(y) \leftarrow false, f_{uuoi}(y) \leftarrow false, f_{uukp}(y) \leftarrow false, \\ & f_{mkp}(y) \leftarrow false, f_{um}(z) \leftarrow false, f_{uuum}(t_i, y, z) \leftarrow true, \\ & f_{per}(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \end{aligned}$$

$$\begin{aligned} m_{pered_end}^{(i)} = & [f_{per}(t_i)]([\exists!(t_i \in T, y \in Y, v \in V) f_{uusa}(t, y, v)] \\ & ([\exists!(t_i \in T, y \in Y, z \in Z) f_{uuum}(t, y, z)] \\ & (\{f_{per}(t_i) \leftarrow false, f_{uu}(y) \leftarrow true, f_{uuoi}(y) \leftarrow true, f_{uukp}(y) \leftarrow true, \\ & f_{mkp}(y) \leftarrow true, f_{um}(z) \leftarrow true, f_{uusa}(t_i, y, v) \leftarrow false, \\ & f_{uuum}(t_i, y, z) \leftarrow false, f_d(t_i) \leftarrow true, f_{st}(t_i) \leftarrow true, \\ & f_{work}(t_i) \leftarrow false\} \vee R^E) \vee R^E) \vee R^E). \end{aligned}$$

В данных выражениях дополнительно используются следующие предикаты:

$f_{uuum}(t, y, z)$ – предикат, связывающий процесс запроса с УУ и УМ;
 $f_{per}(t)$ – каузальный предикат.

С помощью модулей $m_{analiz_begin}^{(i)}$ и $m_{analiz_end}^{(i)}$ производится анализ содержимого таблиц модулей кэш-памяти сети ВЗУ:

$$\begin{aligned} m_{analiz_begin}^{(i)} = & [f_k(c_j)]([f_{work}(t_i) \& (f_{oper}(t_i) = r) \& f_{st}(t_i)] \\ & ([\exists!v \in V) f_{sa}(v)]([\exists!y \in Y) f_{uu}(y)]([f_{uuoi}(y)]([f_{uukp}(y)]([f_{mkp}(y)] \\ & (\{f_{uu}(y) \leftarrow false, f_{uuoi}(y) \leftarrow false, f_{uukp}(y) \leftarrow false, f_{mkp}(y) \leftarrow false, \\ & f_{st}(t_i) \leftarrow false, f_{analiz}(t_i) \leftarrow true, f_k(c_j) \leftarrow false, \\ & f_{uutv}(t_i, y, v) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \end{aligned}$$

$$\begin{aligned} m_{analiz_end}^{(i)} = & [f_{analiz}(t_i)]([\exists!(t_i \in T, y \in Y, v \in V) f_{uutv}(t, y, v)] \\ & ([\exists!h \in \{true, false\}) f_{dkp}(h)] \end{aligned}$$

$$\begin{aligned} &(\{f_{uu}(y) \leftarrow true, f_{analiz}(t_i) \leftarrow false, f_{uuoi}(y) \leftarrow true, \\ &f_{uukp}(y) \leftarrow true, f_{mkp}(y) \leftarrow true, f_{alpha}(t_i) \leftarrow h, f_k(c_j) \leftarrow true, \\ &f_{an}(t_i) \leftarrow true\} \vee R^E) \vee R^E \vee R^E). \end{aligned}$$

Здесь дополнительно используются следующие предикаты:

$f_{analiz}(t)$ – каузальный предикат;

$f_{uutv}(t, y, v)$ – предикат, связывающий процесс запроса с МКП и СА;

$f_{dkp}(h)$ – предикат, используемый для выбора признака успешного поиска в кэш-памяти;

$f_{alpha}(t)$ – предикат, используемый для сохранения признака поиска в кэш-памяти;

$f_{an}(t)$ – каузальный предикат.

Модули $m_{read_kp_begin}^{(i)}$ и $m_{read_kp_end}^{(i)}$ управляют процессом чтения данных из кэш-памяти, если сведения о наличии необходимых данных получены при анализе справочных таблиц:

$$\begin{aligned} m_{read_kp_begin}^{(i)} &= [f_{an}(t_i)]([f_{alpha}(t_i)]([f_k(c_j)]([\exists!(t_i \in T, y \in Y, v \in V) f_{uutv}(t, y, v)] \\ &([f_{sa}(v)]([f_{uuoi}(y)]([f_{uukp}(y)]([f_{mkp}(y)] \\ &(\{f_k(c_j) \leftarrow false, f_{sa}(v) \leftarrow false, f_{uu}(y) \leftarrow false, f_{uuoi}(y) \leftarrow false, \\ &f_{uukp}(y) \leftarrow false, f_{mkp}(y) \leftarrow false, f_{an}(t_i) \leftarrow false, \\ &f_{read_kp}(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \end{aligned}$$

$$\begin{aligned} m_{read_kp_end}^{(i)} &= [f_{read_kp}(t_i)]([\exists!(t_i \in T, y \in Y, v \in V) f_{uutv}(t, y, v)] \\ &(\{f_{read_kp}(t_i) \leftarrow false, f_k(c_j) \leftarrow true, f_{sa}(v) \leftarrow true, f_{uu}(y) \leftarrow false, \\ &f_{uuoi}(y) \leftarrow false, f_{uukp}(y) \leftarrow false, f_{mkp}(y) \leftarrow false, \\ &f_{uutv}(t_i, y, v) \leftarrow false, f_{work}(t_i) \leftarrow false, f_{st}(t_i) \leftarrow true\} \vee R^E) \vee R^E), \end{aligned}$$

где $f_{read_kp}(t)$ – каузальный предикат.

Модули $m_{ust_begin}^{(i)}$ и $m_{ust_end}^{(i)}$ выполняют начальную и конечную фазы операции установки головок записи-считывания и выбора сектора в дисковом модуле сети ВЗУ:

$$\begin{aligned} m_{ust_begin}^{(i)} &= [f_{an}(t_i)]([\neg f_{alpha}(t_i)]([f_k(c_j)]([\exists! y \in Y) f_{uu}(y)] \\ &([f_{uuoi}(y)]([\exists! z \in Z) f_{um}(z)]([f_d(t_i)] \\ &(\{f_d(t_i) \leftarrow false, f_{ust}(t_i) \leftarrow true, f_{an}(t_i) \leftarrow false, f_k(c_j) \leftarrow false\} \vee \\ &\vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \end{aligned}$$

$$m_{ust_end}^{(i)} = [f_{ust}(t_i)](\{f_{ust_got}(t_i) \leftarrow true, f_{ust}(t_i) \leftarrow false, f_k(c_j) \leftarrow true\} \vee R^E),$$

где $f_{ust}(t), f_{ust_got}(t)$ – каузальные предикаты.

Модули $m_{dkp_begin}^{(i)}$ и $m_{dkp_end}^{(i)}$ выполняют начальную и конечную фазы операции копирования данных из дискового модуля в кэш-память:

$$\begin{aligned}
 m_{dkp_begin}^{(i)} &= [f_{ust_got}(t_i)]([\tilde{\exists}!y \in Y)f_{uu}(y)]([\tilde{\exists}!z \in Z)f_{um}(z)] \\
 &([f_{uui}(y)]([f_{uukp}(y)]([f_{mkp}(y)] \\
 &(\{f_{uu}(y) \leftarrow false, f_{uui}(y) \leftarrow false, f_{um}(z) \leftarrow false, f_{uukp}(y) \leftarrow false, \\
 &f_{mkp}(y) \leftarrow false, f_{ust_got}(t_i) \leftarrow false, f_{dkp_copy}(t_i) \leftarrow true, \\
 &f_{uuuu}(t_i, y, z) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \\
 m_{dkp_end}^{(i)} &= [f_{dkp_copy}(t_i)]([\tilde{\exists}!(t_i \in T, y \in Y, z \in Z)f_{uuuu}(t, y, z)] \\
 &(\{f_{uu}(y) \leftarrow true, f_{uui}(y) \leftarrow true, f_{um}(z) \leftarrow true, \\
 &f_{uukp}(y) \leftarrow true, f_{mkp}(y) \leftarrow true, f_{dkp_copy}(t_i) \leftarrow false, \\
 &f_{dkp_got}(t_i) \leftarrow true, f_d(t_i) \leftarrow true\} \vee R^E) \vee R^E).
 \end{aligned}$$

Здесь $f_{dkp_copy}(t)$, $f_{dkp_got}(t)$ – каузальные предикаты.

Модули $m_{kpc_begin}^{(i)}$ и $m_{kpc_end}^{(i)}$ управляют передачей данных из кэш-памяти в сервер c_j , сделавший запрос на чтение данных из сети ВЗУ:

$$\begin{aligned}
 m_{kpc_begin}^{(i)} &= [f_{dkp_got}(t_i)]([f_k(c_j)]([\tilde{\exists}!(t_i \in T, y \in Y, z \in Z)f_{uuuu}(t, y, z)] \\
 &([f_{uui}(y)]([f_{uukp}(y)]([f_{mkp}(y)]([\tilde{\exists}!v \in V)f_{sa}(v)](\{f_k(c_j) \leftarrow false, \\
 &f_{sa}(v) \leftarrow false, f_{uu}(y) \leftarrow false, f_{uui}(y) \leftarrow false, f_{uusa}(t_i, y, v) \leftarrow true, \\
 &f_{uukp}(y) \leftarrow false, f_{mkp}(y) \leftarrow false, f_{dkp_got}(t_i) \leftarrow false, \\
 &f_{kpc}(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \\
 m_{kpc_end}^{(i)} &= [f_{kpc}(t_i)]([\tilde{\exists}!(t_i \in T, y \in Y, z \in Z)f_{uuuu}(t, y, z)] \\
 &([\tilde{\exists}!(t_i \in T, y \in Y, v \in V)f_{uusa}(t_i, y, v)] \\
 &(\{f_k(c_j) \leftarrow true, f_{uu}(y) \leftarrow true, f_{uui}(y) \leftarrow true, f_{sa}(v) \leftarrow true, \\
 &f_{uukp}(y) \leftarrow true, f_{mkp}(y) \leftarrow true, f_{kpc}(t_i) \leftarrow false, \\
 &f_{work}(t_i) \leftarrow false, f_{st}(t_i) \leftarrow true, f_{uusa}(t_i, y, v) \leftarrow false, \\
 &f_{uuuu}(t_i, y, z) \leftarrow false\} \vee R^E) \vee R^E) \vee R^E),
 \end{aligned}$$

где $f_{kpc}(t)$ – каузальный предикат.

3. Промежуточный этап построения сети абстрактных машин для децентрализованной системы хранения данных

При описании сети абстрактных машин в разд. 2 использовались предметные переменные t, v, y, z , пробегающие по множествам $T = \{t_1, t_2, \dots, t_n\}$,

$V = \{v_1, v_2, \dots, v_q\}$, $Y = \{y_1, y_2, \dots, y_m\}$ и $Z = \{z_1, z_2, \dots, z_p\}$ соответственно. Для перехода к сетевой архитектуре многофункциональной системы ВЗУ на промежуточном этапе заменим в указанной сети вхождения предметных переменных v , y и z на переменную t , а затем удалим все тернарные предикаты. В выражении для каждого i -го модуля будем использовать соответственно единственное значение t_i предметной переменной t , что будет соответствовать тому факту, что каждому модулю дисковой памяти $D(i)$ будут соответствовать устройства УМ(i), МКП(i), УУКП(i), УУОИ(i), УУМПБД(i), МПБД(i), УУ(i), СА(i), представленные на рис. 1. При этом коммутаторы 2, 3, 4 и 5 утрачивают свое значение и могут быть заменены непосредственными связями соответствующих устройств с одинаковыми значениями индекса i . Данное преобразование абстрактной сети СеАМ, построенной в разд. 2, в новую сеть не будет эквивалентным, так как здесь мы намеренно отказываемся от коммутации типа «каждый – с каждым» и используем ограниченные топологически децентрализованной сетевой архитектурой связи. Новая сеть абстрактных машин будет содержать одноименные предыдущей сети модули, описываемые следующими выражениями:

$$\begin{aligned} m_{zapkp_begin}^{(i)} &= [f_k(c_j)]([f_{work}(t_i) \& (f_{oper}(t_i) = w) \& f_{st}(t_i)] \\ &([f_{sa}(t_i)]([f_{uu}(t_i)]([f_{uuoi}(t_i)]([f_{uukp}(t_i)]([f_{mkp}(t_i)] \\ &(\{f_k(c_j) \leftarrow false, f_{st}(t_i) \leftarrow false, f_{sa}(t_i) \leftarrow false, f_{uu}(t_i) \leftarrow false, \\ &f_{uuoi}(t_i) \leftarrow false, f_{uukp}(t_i) \leftarrow false, f_{mkp}(t_i) \leftarrow false, f_{zapkp}(t_i) \leftarrow true\} \vee \\ &\vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \end{aligned}$$

$$\begin{aligned} m_{zapkp_end}^{(i)} &= [f_{zapkp}(t_i)](\{f_k(c_j) \leftarrow true, f_{sa}(t_i) \leftarrow true, \\ &f_{uu}(t_i) \leftarrow true, f_{uuoi}(t_i) \leftarrow true, f_{uukp}(t_i) \leftarrow true, \\ &f_{mkp}(t_i) \leftarrow true, f_{zapkp}(t_i) \leftarrow false, f_{poisk}(t_i) \leftarrow true\} \vee R^E); \end{aligned}$$

$$\begin{aligned} m_{poisk_begin}^{(i)} &= [f_{poisk}(t_i)]([f_{uu}(t_i)]([f_{uuoi}(t_i)]([f_{um}(t_i)]([f_d(t_i)] \\ &(\{f_d(t_i) \leftarrow false, f_{poisk}(t_i) \leftarrow false, \\ &f_p(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \end{aligned}$$

$$m_{poisk_end}^{(i)} = [f_p(t_i)](\{f_p(t_i) \leftarrow false, f_{pered}(t_i) \leftarrow true\} \vee R^E);$$

$$\begin{aligned} m_{pered_begin}^{(i)} &= [f_{pered}(t_i)]([f_{uu}(t_i)]([f_{uuoi}(t_i)]([f_{uukp}(t_i)]([f_{mkp}(t_i)]([f_{um}(t_i)] \\ &(\{f_{pered}(t_i) \leftarrow false, f_{uu}(t_i) \leftarrow false, f_{uuoi}(t_i) \leftarrow false, f_{uukp}(t_i) \leftarrow false, \\ &f_{mkp}(t_i) \leftarrow false, f_{um}(t_i) \leftarrow false, f_{per}(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee \\ &\vee R^E) \vee R^E) \vee R^E); \end{aligned}$$

$$\begin{aligned} m_{pered_end}^{(i)} &= [f_{per}(t_i)](\{f_{per}(t_i) \leftarrow false, f_{uu}(t_i) \leftarrow true, \\ &f_{uuoi}(t_i) \leftarrow true, f_{uukp}(t_i) \leftarrow true, f_{mkp}(t_i) \leftarrow true, \end{aligned}$$

$$f_{um}(t_i) \leftarrow true, f_d(t_i) \leftarrow true, f_{st}(t_i) \leftarrow true, \\ f_{work}(t_i) \leftarrow false \} \vee R^E);$$

$$m_{analiz_begin}^{(i)} = [f_k(c_j)]([f_{work}(t_i) \& (f_{oper}(t_i) = r) \& f_{st}(t_i)] \\ ([f_{sa}(t_i)]([f_{uu}(t_i)]([f_{uuoi}(t_i)]([f_{uukp}(t_i)]([f_{mkp}(t_i)] \\ (\{f_{uu}(t_i) \leftarrow false, f_{uuoi}(t_i) \leftarrow false, f_{uukp}(t_i) \leftarrow false, f_{mkp}(t_i) \leftarrow false, \\ f_{st}(t_i) \leftarrow false, f_{analiz}(t_i) \leftarrow true, f_k(c_j) \leftarrow false\} \vee \\ \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E);$$

$$m_{analiz_end}^{(i)} = [f_{analiz}(t_i)]([\exists! h \in \{true, false\}) f_{dkp}(h)] \\ (\{f_{uu}(t_i) \leftarrow true, f_{analiz}(t_i) \leftarrow false, f_{uuoi}(t_i) \leftarrow true, \\ f_{uukp}(t_i) \leftarrow true, f_{mkp}(t_i) \leftarrow true, f_{alpha}(t_i) \leftarrow h, f_k(c_j) \leftarrow true, \\ f_{an}(t_i) \leftarrow true\} \vee R^E) \vee R^E);$$

$$m_{read_kp_begin}^{(i)} = [f_{an}(t_i)]([f_{alpha}(t_i)]([f_k(c_j)] \\ ([f_{sa}(t_i)]([f_{uuoi}(t_i)]([f_{uukp}(t_i)]([f_{mkp}(t_i)] \\ (\{f_k(c_j) \leftarrow false, f_{sa}(t_i) \leftarrow false, f_{uu}(t_i) \leftarrow false, f_{uuoi}(t_i) \leftarrow false, \\ f_{uukp}(t_i) \leftarrow false, f_{mkp}(t_i) \leftarrow false, f_{an}(t_i) \leftarrow false, \\ f_{read_kp}(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E);$$

$$m_{read_kp_end}^{(i)} = [f_{read_kp}(t_i)] \\ (\{f_{read_kp}(t_i) \leftarrow false, f_k(c_j) \leftarrow true, f_{sa}(t_i) \leftarrow true, f_{uu}(t_i) \leftarrow false, \\ f_{uuoi}(t_i) \leftarrow false, f_{uukp}(t_i) \leftarrow false, f_{mkp}(t_i) \leftarrow false, \\ f_{work}(t_i) \leftarrow false, f_{st}(t_i) \leftarrow true\} \vee R^E);$$

$$m_{ust_begin}^{(i)} = [f_{an}(t_i)]([\neg f_{alpha}(t_i)]([f_k(c_j)]([f_{uu}(t_i)]([f_{uuoi}(t_i)] \\ ([f_{um}(t_i)]([f_d(t_i)](\{f_d(t_i) \leftarrow false, f_{ust}(t_i) \leftarrow true, \\ f_{an}(t_i) \leftarrow false, f_k(c_j) \leftarrow false\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee \\ \vee R^E) \vee R^E) \vee R^E);$$

$$m_{ust_end}^{(i)} = [f_{ust}(t_i)](\{f_{ust_got}(t_i) \leftarrow true, f_{ust}(t_i) \leftarrow false, f_k(c_j) \leftarrow true\} \vee R^E);$$

$$m_{dkp_begin}^{(i)} = [f_{ust_got}(t_i)]([f_{uu}(t_i)]([f_{um}(t_i)]([f_{uuoi}(t_i)] \\ ([f_{uukp}(t_i)]([f_{mkp}(t_i)](\{f_{uu}(t_i) \leftarrow false, f_{uuoi}(t_i) \leftarrow false, \\ f_{um}(t_i) \leftarrow false, f_{uukp}(t_i) \leftarrow false, f_{mkp}(t_i) \leftarrow false,$$

$$f_{ust_got}(t_i) \leftarrow false, f_{dkp_copy}(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E);$$

$$m_{dkp_end}^{(i)} = [f_{dkp_copy}(t_i)](\{f_{uu}(t_i) \leftarrow true, f_{uuoi}(t_i) \leftarrow true, f_{um}(t_i) \leftarrow true, f_{uukp}(t_i) \leftarrow true, f_{mkp}(t_i) \leftarrow true, f_{dkp_copy}(t_i) \leftarrow false, f_{dkp_got}(t_i) \leftarrow true, f_d(t_i) \leftarrow true\} \vee R^E);$$

$$m_{kpc_begin}^{(i)} = [f_{dkp_got}(t_i)]([f_k(c_j)]([f_{uuoi}(t_i)]([f_{uukp}(t_i)]([f_{mkp}(t_i)](\{f_k(c_j) \leftarrow false, f_{sa}(t_i) \leftarrow false, f_{uu}(t_i) \leftarrow false, f_{uuoi}(t_i) \leftarrow false, f_{uukp}(t_i) \leftarrow false, f_{mkp}(t_i) \leftarrow false, f_{dkp_got}(t_i) \leftarrow false, f_{kpc}(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E);$$

$$m_{kpc_end}^{(i)} = [f_{kpc}(t_i)](\{f_k(c_j) \leftarrow true, f_{uu}(t_i) \leftarrow true, f_{uuoi}(t_i) \leftarrow true, f_{sa}(t_i) \leftarrow true, f_{uukp}(t_i) \leftarrow true, f_{mkp}(t_i) \leftarrow true, f_{kpc}(t_i) \leftarrow false, f_{work}(t_i) \leftarrow false, f_{st}(t_i) \leftarrow true\} \vee R^E).$$

4. Сеть абстрактных машин для топологически-децентрализованной (сетевой) системы хранения данных

Проведем дальнейшее преобразование сети абстрактных машин, построенной в разд. 2 и подвергшейся промежуточному преобразованию в разд. 3. В выражениях для модулей абстрактной сети, полученной в разд. 3, использованы предикаты и функции, которым соответствуют информационные объекты в виртуальном FS-пространстве. Недостатком организации данной сети СеАМ является то, что при ее реализации сетью виртуальных машин многие модули обращаются к предикатам и функциям, конкурируя между собой. Для снижения данных издержек проведем декомпозицию основных предикатов и функций, введя для новых предикатов и функций символы, снабженные верхними индексами, соответствующими индексу i каждого модуля:

$$m_{zapkp_begin}^{(i)} = [f_k(c_j)]([f_{work}^{(i)}(t_i) \& (f_{oper}^{(i)}(t_i) = w) \& f_{st}^{(i)}(t_i)]([f_{sa}^{(i)}(t_i)]([f_{uu}^{(i)}(t_i)]([f_{uuoi}^{(i)}(t_i)]([f_{uukp}^{(i)}(t_i)]([f_{mkp}^{(i)}(t_i)](\{f_k(c_j) \leftarrow false, f_{st}^{(i)}(t_i) \leftarrow false, f_{sa}^{(i)}(t_i) \leftarrow false, f_{uu}^{(i)}(t_i) \leftarrow false, f_{uuoi}^{(i)}(t_i) \leftarrow false, f_{uukp}^{(i)}(t_i) \leftarrow false, f_{mkp}^{(i)}(t_i) \leftarrow false, f_{zapkp}^{(i)}(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E);$$

$$m_{zapkp_end}^{(i)} = [f_{zapkp}^{(i)}(t_i)](\{f_k(c_j) \leftarrow true, f_{sa}^{(i)}(t_i) \leftarrow true,$$

$$\begin{aligned}
 & f_{uu}^{(i)}(t_i) \leftarrow true, f_{uuoi}^{(i)}(t_i) \leftarrow true, f_{uukp}^{(i)}(t_i) \leftarrow true, \\
 & f_{mkp}^{(i)}(t_i) \leftarrow true, f_{zapkp}^{(i)}(t_i) \leftarrow false, f_{poisk}^{(i)}(t_i) \leftarrow true \} \vee R^E); \\
 & m_{poisk_begin}^{(i)} = [f_{poisk}^{(i)}(t_i)]([f_{uu}^{(i)}(t_i)]([f_{uuoi}^{(i)}(t_i)]([f_{um}^{(i)}(t_i)]([f_d^{(i)}(t_i)] \\
 & (\{f_d^{(i)}(t_i) \leftarrow false, f_{poisk}^{(i)}(t_i) \leftarrow false, \\
 & f_p^{(i)}(t_i) \leftarrow true \} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \\
 & m_{poisk_end}^{(i)} = [f_p^{(i)}(t_i)](\{f_p^{(i)}(t_i) \leftarrow false, f_{pered}^{(i)}(t_i) \leftarrow true \} \vee R^E); \\
 & m_{pered_begin}^{(i)} = [f_{pered}^{(i)}(t_i)]([f_{uu}^{(i)}(t_i)]([f_{uuoi}^{(i)}(t_i)]([f_{uukp}^{(i)}(t_i)]([f_{mkp}^{(i)}(t_i)]([f_{um}^{(i)}(t_i)] \\
 & (\{f_{pered}^{(i)}(t_i) \leftarrow false, f_{uu}^{(i)}(t_i) \leftarrow false, f_{uuoi}^{(i)}(t_i) \leftarrow false, f_{uukp}^{(i)}(t_i) \leftarrow false, \\
 & f_{mkp}^{(i)}(t_i) \leftarrow false, f_{um}^{(i)}(t_i) \leftarrow false, f_{per}^{(i)}(t_i) \leftarrow true \} \vee R^E) \vee R^E) \vee R^E) \vee \\
 & \vee R^E) \vee R^E) \vee R^E); \\
 & m_{pered_end}^{(i)} = [f_{per}^{(i)}(t_i)](\{f_{per}^{(i)}(t_i) \leftarrow false, f_{uu}^{(i)}(t_i) \leftarrow true, \\
 & f_{uuoi}^{(i)}(t_i) \leftarrow true, f_{uukp}^{(i)}(t_i) \leftarrow true, f_{mkp}^{(i)}(t_i) \leftarrow true, \\
 & f_{um}^{(i)}(t_i) \leftarrow true, f_d^{(i)}(t_i) \leftarrow true, f_{st}^{(i)}(t_i) \leftarrow true, \\
 & f_{work}^{(i)}(t_i) \leftarrow false \} \vee R^E); \\
 & m_{analiz_begin}^{(i)} = [f_k(c_j)]([f_{work}^{(i)}(t_i) \& (f_{oper}^{(i)}(t_i) = r) \& f_{st}^{(i)}(t_i)] \\
 & ([f_{sa}^{(i)}(t_i)]([f_{uu}^{(i)}(t_i)]([f_{uuoi}^{(i)}(t_i)]([f_{uukp}^{(i)}(t_i)]([f_{mkp}^{(i)}(t_i)] \\
 & (\{f_{uu}^{(i)}(t_i) \leftarrow false, f_{uuoi}^{(i)}(t_i) \leftarrow false, f_{uukp}^{(i)}(t_i) \leftarrow false, f_{mkp}^{(i)}(t_i) \leftarrow false, \\
 & f_{st}^{(i)}(t_i) \leftarrow false, f_{analiz}^{(i)}(t_i) \leftarrow true, f_k(c_j) \leftarrow false \} \vee R^E) \vee R^E) \vee \\
 & \vee R^E) \vee R^E) \vee R^E) \vee R^E); \\
 & m_{analiz_end}^{(i)} = [f_{analiz}^{(i)}(t_i)]([\exists! h \in \{true, false\}] f_{dkp}^{(i)}(h)] \\
 & (\{f_{uu}^{(i)}(t_i) \leftarrow true, f_{analiz}^{(i)}(t_i) \leftarrow false, f_{uuoi}^{(i)}(t_i) \leftarrow true, \\
 & f_{uukp}^{(i)}(t_i) \leftarrow true, f_{mkp}^{(i)}(t_i) \leftarrow true, f_{alpha}^{(i)}(t_i) \leftarrow h, \\
 & f_k(c_j) \leftarrow true, f_{an}^{(i)}(t_i) \leftarrow true \} \vee R^E) \vee R^E); \\
 & m_{read_kp_begin}^{(i)} = [f_{an}^{(i)}(t_i)]([f_{alpha}^{(i)}(t_i)]([f_k(c_j)] \\
 & ([f_{sa}^{(i)}(t_i)]([f_{uuoi}^{(i)}(t_i)]([f_{uukp}^{(i)}(t_i)]([f_{mkp}^{(i)}(t_i)]
 \end{aligned}$$

$$\begin{aligned}
 &(\{f_k(c_j) \leftarrow false, f_{sa}^{(i)}(t_i) \leftarrow false, f_{uu}^{(i)}(t_i) \leftarrow false, f_{uuoi}^{(i)}(t_i) \leftarrow false, \\
 &f_{uukp}^{(i)}(t_i) \leftarrow false, f_{mkp}^{(i)}(t_i) \leftarrow false, f_{an}^{(i)}(t_i) \leftarrow false, \\
 &f_{read_kp}^{(i)}(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \\
 &m_{read_kp_end}^{(i)} = [f_{read_kp}^{(i)}(t_i)](\{f_{read_kp}^{(i)}(t_i) \leftarrow false, \\
 &f_k(c_j) \leftarrow true, f_{sa}^{(i)}(t_i) \leftarrow true, f_{uu}^{(i)}(t_i) \leftarrow false, \\
 &f_{uuoi}^{(i)}(t_i) \leftarrow false, f_{uukp}^{(i)}(t_i) \leftarrow false, f_{mkp}^{(i)}(t_i) \leftarrow false, \\
 &f_{work}^{(i)}(t_i) \leftarrow false, f_{st}^{(i)}(t_i) \leftarrow true\} \vee R^E); \\
 &m_{ust_begin}^{(i)} = [f_{an}^{(i)}(t_i)]([\neg f_{alpha}^{(i)}(t_i)]([f_k(c_j)]([f_{uu}^{(i)}(t_i)]([f_{uuoi}^{(i)}(t_i)] \\
 &([f_{um}^{(i)}(t_i)]([f_d^{(i)}(t_i)](\{f_d^{(i)}(t_i) \leftarrow false, f_{ust}^{(i)}(t_i) \leftarrow true, \\
 &f_{an}^{(i)}(t_i) \leftarrow false, f_k(c_j) \leftarrow false\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee \\
 &\vee R^E) \vee R^E) \vee R^E); \\
 &m_{ust_end}^{(i)} = [f_{ust}^{(i)}(t_i)](\{f_{ust_got}^{(i)}(t_i) \leftarrow true, f_{ust}^{(i)}(t_i) \leftarrow false, f_k(c_j) \leftarrow true\} \vee R^E); \\
 &m_{dkp_begin}^{(i)} = [f_{ust_got}^{(i)}(t_i)]([f_{uu}^{(i)}(t_i)]([f_{um}^{(i)}(t_i)]([f_{uuoi}^{(i)}(t_i)] \\
 &([f_{uukp}^{(i)}(t_i)]([f_{mkp}^{(i)}(t_i)](\{f_{uu}^{(i)}(t_i) \leftarrow false, f_{uuoi}^{(i)}(t_i) \leftarrow false, \\
 &f_{um}^{(i)}(t_i) \leftarrow false, f_{uukp}^{(i)}(t_i) \leftarrow false, f_{mkp}^{(i)}(t_i) \leftarrow false, \\
 &f_{ust_got}^{(i)}(t_i) \leftarrow false, f_{dkp_copy}^{(i)}(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee \\
 &\vee R^E) \vee R^E) \vee R^E) \vee R^E); \\
 &m_{dkp_end}^{(i)} = [f_{dkp_copy}^{(i)}(t_i)](\{f_{uu}^{(i)}(t_i) \leftarrow true, f_{uuoi}^{(i)}(t_i) \leftarrow true, \\
 &f_{um}^{(i)}(t_i) \leftarrow true, f_{uukp}^{(i)}(t_i) \leftarrow true, f_{mkp}^{(i)}(t_i) \leftarrow true, f_{dkp_copy}^{(i)}(t_i) \leftarrow false, \\
 &f_{dkp_got}^{(i)}(t_i) \leftarrow true, f_d^{(i)}(t_i) \leftarrow true\} \vee R^E); \\
 &m_{kpc_begin}^{(i)} = [f_{dkp_got}^{(i)}(t_i)]([f_k(c_j)]([f_{uuoi}^{(i)}(t_i)]([f_{uukp}^{(i)}(t_i)]([f_{mkp}^{(i)}(t_i)] \\
 &(\{f_k(c_j) \leftarrow false, f_{sa}^{(i)}(t_i) \leftarrow false, f_{uu}^{(i)}(t_i) \leftarrow false, f_{uuoi}^{(i)}(t_i) \leftarrow false, \\
 &f_{uukp}^{(i)}(t_i) \leftarrow false, f_{mkp}^{(i)}(t_i) \leftarrow false, f_{dkp_got}^{(i)}(t_i) \leftarrow false, \\
 &f_{kpc}^{(i)}(t_i) \leftarrow true\} \vee R^E) \vee R^E) \vee R^E) \vee R^E) \vee R^E); \\
 &m_{kpc_end}^{(i)} = [f_{kpc}^{(i)}(t_i)](\{f_k(c_j) \leftarrow true, f_{uu}^{(i)}(t_i) \leftarrow true,
 \end{aligned}$$

$$\begin{aligned} f_{uoi}^{(i)}(t_i) &\leftarrow true, f_{sa}^{(i)}(t_i) \leftarrow true, f_{uukp}^{(i)}(t_i) \leftarrow true, \\ f_{mkp}^{(i)}(t_i) &\leftarrow true, f_{kpc}^{(i)}(t_i) \leftarrow false, f_{work}^{(i)}(t_i) \leftarrow false, \\ f_{st}^{(i)}(t_i) &\leftarrow true \} \vee R^E). \end{aligned}$$

Преобразования сетей абстрактных машин, проведенные в разд. 3 и 4, не являются эквивалентными, так как при переходе от централизованной к сетевой децентрализованной архитектуре многофункциональной системы ВЗУ целенаправленно упрощены связи между устройствами.

Полученная система выражений для модулей сети СеАМ соответствует сетевой архитектуре, представленной на рис. 2. Сеть хранения и обработки данных построена на основе коммутатора сети МФВЗУ (многофункциональных внешних запоминающих устройств). Выражения для модулей СеАМ позволяют не только построить имитационную поведенческую модель данной сети, но и могут быть положены в основу построения программного и микропрограммного обеспечения сети ВЗУ.

Заключение

1. Обоснована необходимость разработки новых методов и средств поддержки функционально-структурной и функционально-топологической декомпозиции архитектурных моделей систем и сетей хранения и обработки данных. Данные модели и методы должны соответствовать современной технологической инфраструктуре параллельных и распределенных баз данных, кластерных и мультипроцессорных систем, сетевых соединений, интеллектуальных контроллеров и накопителей информации. Показано, что сети абстрактных машин и сценарные представления, формализуемые логико-алгебраическими моделями, удобны для описания и реализации декомпозиционно-композиционных методов проектирования систем и сетей хранения и обработки данных.

2. Разработаны новые формальные модели организации функционирования распределенного ресурса – многофункциональной системы внешней памяти, подключаемой к вычислительной сети. В отличие от известных, в данных моделях эффективно реализуются взаимодействия компонент, что позволяет разработчику модифицировать сложные модели, получая новые качественные результаты. Рассмотрены подобные трансформации формальных моделей, позволившие перейти от централизованной архитектуры систем хранения и обработки данных к децентрализованной (сетевой).

3. Предложены принципы построения систем управления применительно к объектам, функционирование которых представимо в классе сетей абстрактных машин. На основе логико-алгебраического подхода, формальных и неформальных моделей и методов искусственного интеллекта, теории иерархически структурированных абстрактных машин и сценарных представлений предложена методология построения виртуализированных систем и сетей внешнего хранения и обработки данных, узлы которых сочетают функции виртуализации с функциями обработки структурированных данных.

4. Реализация предлагаемых принципов позволяет создавать новые архитектуры систем и сетей хранения и обработки данных и соответствующее программное обеспечение на основе подходов, существенно отличных от фон-неймановских [6] и определенных парадигмой согласования процессов и

объектов, которая в большой степени соответствует принципам решения задач в основном невычислительного характера: информационного поиска, обработки структурированных данных, управления данными и знаниями в экспертных системах.

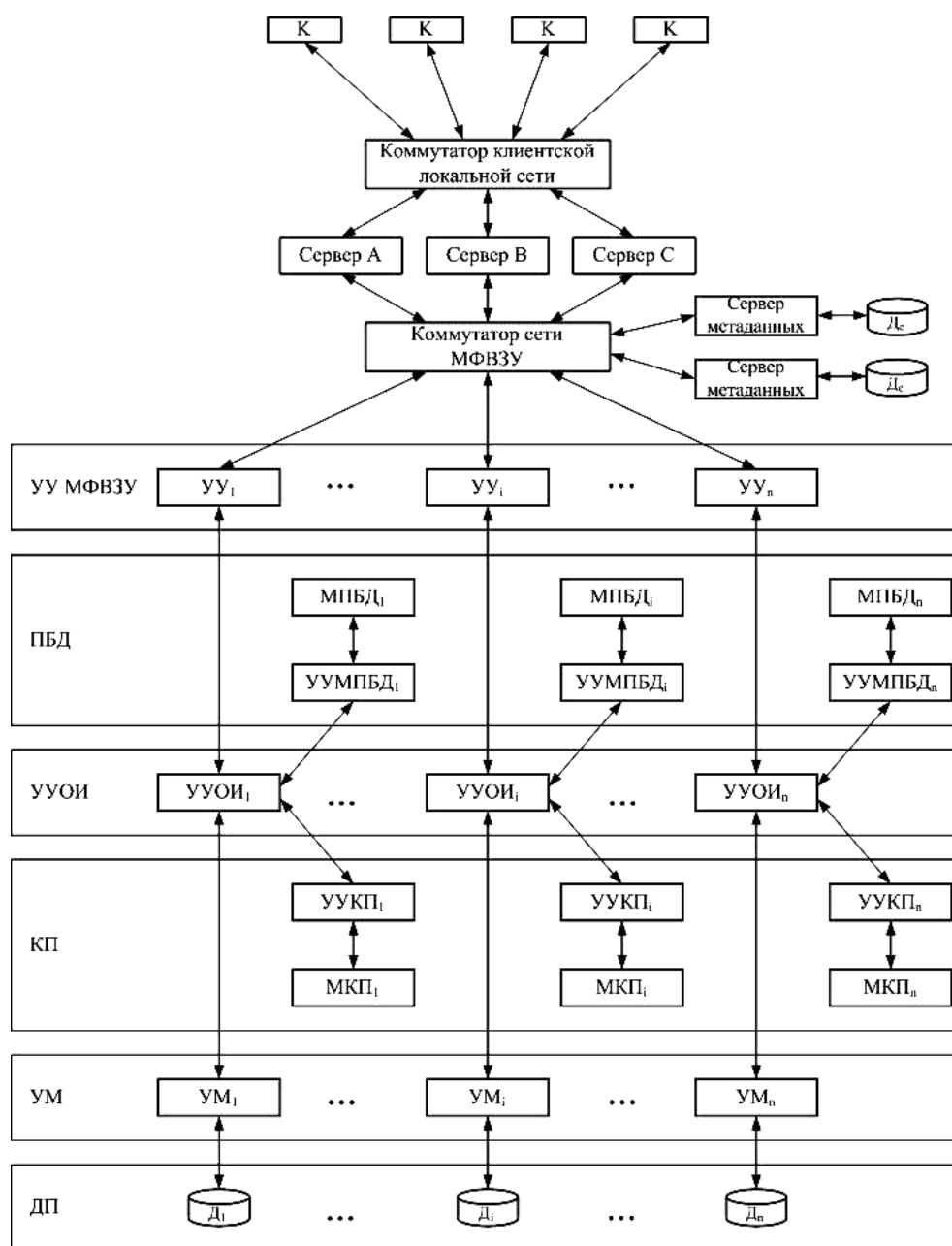


Рис. 2. Результирующая сеть многофункциональных ВЗУ

Список литературы

1. **Фарли, М.** Сети хранения данных / М. Фарли. – М. : Лори, 2003. – 550 с.
2. **Наик, Д.** Системы хранения данных в Windows / Д. Наик. – М. : Вильямс, 2005. – 432 с.

3. **Зинкин, С. А.** Сетевая FS-архитектура машин баз данных и исполняемые формальные спецификации / С. А. Зинкин // Вестник Пермского государственного университета. Серия «Математика. Механика. Информатика». – 2008. – № 4 (20). – С. 169–181.
4. **Зинкин, С. А.** Сети абстрактных машин высших порядков в проектировании систем и сетей хранения и обработки данных (базовый формализм и его расширения) / С. А. Зинкин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2007. – № 3. – С. 13–22.
5. **Зинкин, С. А.** Сети абстрактных машин высших порядков в проектировании систем и сетей хранения и обработки данных (механизмы интерпретации и варианты использования) / С. А. Зинкин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2007. – № 4. – С. 37–51.
6. **Волчихин, В. И.** Развитие нефоннеймановских концепций в компьютерной и сетевой архитектуре / В. И. Волчихин, С. А. Зинкин // Новые информационные технологии и системы : труды IX Международной научно-технической конференции. – Пенза, 2010. – Ч. 2. – С. 48–63.

Волчихин Владимир Иванович

доктор технических наук, профессор,
ректор Пензенского государственного
университета

E-mail: rectorat@pnzgu.ru

Volchikhin Vladimir Ivanovich

Doctor of engineering sciences, professor,
rector of Penza State University

Зинкин Сергей Александрович

доктор технических наук, профессор,
кафедра вычислительной техники,
Пензенский государственный
университет

E-mail: zsa49@yandex.ru

Zinkin Sergey Alexandrovich

Doctor of engineering sciences, professor,
sub-department of computer engineering,
Penza State University

УДК 681.324

Волчихин, В. И.

Абстрактное и структурное моделирование сетей хранения и обработки данных / В. И. Волчихин, С. А. Зинкин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 3–18.

ЭФФЕКТИВНАЯ МОДЕЛЬ АНАЛИЗА ДАННЫХ НА ОСНОВЕ ЛИНЕЙНЫХ ВЕКТОРНЫХ РЕШЕТОК

Аннотация. Рассмотрены вопросы представления структур данных в виде решеток четырех типов. Показано, что эффективность моделирования повышается при использовании модели данных в виде линейной векторной решетки. Рассмотрены свойства таких решеток и особенности алгоритма построения модели данных. Приводится пример анализа данных.

Ключевые слова: теоретико-решетчатая модель анализа данных, модель данных в виде линейной векторной решетки, особенности алгоритма построения модели данных, пример анализа данных.

Abstract. The article considers the problems of representation of data structures in the form of lattices of four types. It is shown that the efficiency of simulation increases with the use of the data model as a linear vector lattice. The researchers describe the properties of these lattices, and, especially, the algorithm for constructing a data model. An example of data analysis is also represented.

Key words: theoretical lattice-form model of data analysis, data model as a linear vector lattice, features of model data construction algorithm, data analysis example.

Введение

Методы комбинаторно-упорядоченного моделирования (КУМ), основанные на теории решеток, построенных с помощью оператора замыкания специального типа, как показано в работах [1–5], можно использовать для эффективного решения задач анализа данных в различных областях, в частности: при структурном анализе систем управления; в технологии добычи данных (Data Mining); для классификации объектов при распознавании образов; при анализе модели знаний в автоматизированных обучающих системах; при автоматизированном биллинге услуг связи; при анализе и синтезе электронных схем и др. Особенностью метода КУМ является высокая адекватность представления структуры данных в виде решетки, образованной оператором замыкания. В частности, структура исходных данных в виде семейства порождающих множеств может быть адекватно представлена (с точностью до изоморфизма) полной решеткой, упорядоченной включением [1]. В большинстве случаев метод КУМ позволяет повысить эффективность анализа данных за счет адекватного представления их структуры в виде решетки.

Вместе с тем возникает ряд проблем, связанных с реализацией алгоритма построения решеток, который в некоторых случаях обладает высокой трудоемкостью. В работе [2] рассмотрены пути повышения эффективности метода КУМ, основанные на преобразовании представления данных в виде полной решетки в характеристическую линейную векторную решетку. В настоящей работе рассмотрены некоторые особенности представления моделей данных в виде линейных векторных решеток.

1. Типология теоретико-решетчатых моделей структур данных

Для анализа данных применяется решетка основного типа L_{Ψ} , который образуется с помощью оператора замыкания $\hat{A} = \bigcap R_A$, заданного на исход-

ном порождающем семействе множеств $\{R\} \subseteq 2^P$ (например, семействе атрибутов исходных данных), где $A \subseteq \bigcup R$, $A \subseteq R_A$, $R_A \in \{R\}$, $P = \bigcup_{R \in \{R\}} R$ – носитель модели L_Ψ [1]. Элементами решетки являются все замкнутые множества, удовлетворяющие условию $A = \hat{A}$. Семейство замкнутых подмножеств $A \subseteq P$, удовлетворяющих условию $A = \hat{A}$, образует относительно включения полную решетку L_Ψ с теоретико-решетчатыми операциями $A \vee B = A \cup B$ и $A \wedge B = A \cap B$, где $A, B \in L_\Psi$ и $\hat{}$ обозначает операцию замыкания объединения множеств. Структурные нуль $\mathbf{0}$ и единица $\mathbf{1}$ решетки L_Ψ задаются выражениями $\mathbf{0} = \hat{\emptyset} = \bigcap R_{\emptyset} = \bigcap_{R \in \{R\}} R$, $\mathbf{1} = P = \bigcup R = \bigcup_{R \in \{R\}} R$. Решетка L_Ψ является частным булевой решетки $\mathbf{B}(P)$ относительно оператора замыкания $\hat{A} = \bigcap R_A$, причем нижние грани в L_Ψ совпадают с нижними гранями в $\mathbf{B}(P)$ [6].

Второй (изоморфный) тип решетки L_Φ определяется кортежем семейства множеств допустимых значений признаков данных $\vec{K} = \langle K \rangle$ и порождающим семейством кортежей одноэлементных множеств значений этих признаков $\mathbf{K}^{(1)} = \left\{ \left\langle K^{(1)} \right\rangle \right\}$, задающим исходные данные, где $K^{(1)} \subseteq K$ и $|K^{(1)}| \leq 1$ (если какой-либо признак отсутствует, то полагают, что $K^{(1)} = \emptyset$) [3]. Изоморфизм решеток L_Ψ и L_Φ определяет отображение $\phi\left(\left\langle K^{(1)} \right\rangle\right) = A$, где $A \subseteq \bigcup R$.

Представление данных решетками типа L_Ψ или L_Φ зависит от решаемой задачи анализа данных. Например, в задачах анализа ассоциаций данных удобно использовать решетку типа L_Ψ [1], в задачах классификации данных и распознавания образов – решетку типа L_Φ [3].

Третий тип решеток L_ε задает отображение $\varepsilon(A) = (\dots, \varepsilon_b, \dots)$, где

$$\varepsilon_b = \begin{cases} 1, & \text{если } b \in A, \\ 0, & \text{если } b \notin A \end{cases} \quad (A \subseteq \bigcup R),$$

и определяет изоморфизм $L_\varepsilon \cong L_\Psi$ [2].

Вектор $(\dots, \varepsilon_b, \dots)$ называется характеристическим вектором подмножества $A \subseteq \bigcup R$, а множество характеристических векторов элементов решетки

L_Ψ , сохраняющее отношение следования, образует решетку L_ϵ , изоморфную решетке L_Ψ . Характеристические векторы удобны для организации вычислений при построении решеток. В решетке L_ϵ векторы упорядочены относительно операции включения, которая определяется естественным образом.

Вместе с тем множество характеристических векторов может быть упорядочено относительно операции доминирования, тогда образуется решетка четвертого типа: линейная векторная решетка $\tilde{L}_\epsilon$, состоящая из тех же элементов, что и L_ϵ , но имеющая лексикографическое упорядочение. В общем случае решетки $\tilde{L}_\epsilon$ и L_ϵ не являются изоморфными. Отношение лексикографического порядка $\leq$ на линейно упорядоченном множестве $\tilde{L}_\epsilon$ определяется следующим образом. Пусть $\epsilon, \tau \in \tilde{L}_\epsilon$ и $\epsilon = (\epsilon_1, \dots, \epsilon_n)$, $\tau = (\tau_1, \dots, \tau_n)$ ($n = |\bigcup R|$), тогда $\epsilon \leq \tau$, если выполняется условие: $\exists j$ такое, что $\epsilon_j < \tau_j$ и $\forall i (i < j)$ справедливо $\epsilon_i = \tau_i$. Согласно теореме Э. Шпильрайна любое отношение частичного порядка, заданное на некотором множестве, может быть продолжено до отношения линейного порядка [6]. В этом смысле отношение линейного порядка на векторной решетке $\tilde{L}_\epsilon$ является продолжением отношения включения на решетке L_ϵ .

2. Свойства теоретико-решетчатой модели данных типа линейной векторной решетки

Справедливы следующие свойства модели данных, основанные на решетках типа $\tilde{L}_\epsilon$.

Свойство 1. Пусть $\leq$ – отношение следования включением элементов в решетке L_ϵ , а $\leq$ – отношение лексикографического порядка в векторной решетке $\tilde{L}_\epsilon$, и решетки L_ϵ и $\tilde{L}_\epsilon$ состоят из одинаковых элементов, тогда справедливо $\alpha \leq \beta \Rightarrow \alpha \leq \beta$ (при $\alpha, \beta \in L_\epsilon$ и $\alpha, \beta \in \tilde{L}_\epsilon$). Обратная импликация в общем случае неверна.

Свойство 2. $(\alpha \wedge \beta = \alpha) \Rightarrow \alpha \leq \beta \Rightarrow \alpha \leq \beta$, где $\wedge$ – операция поразрядного пересечения векторов.

Свойство 3. Если $\alpha \wedge \beta = \gamma$, то $\gamma \leq \alpha$ и $\gamma \leq \beta$ ($\alpha, \beta \in L_\epsilon$).

Свойство 4. $\alpha \wedge \beta = \alpha \wedge \beta$, где $\wedge$ – операция взятия точной нижней грани в решетке L_ϵ [1], $\wedge$ – операция поразрядного пересечения векторов.

3. Особенности алгоритма построения модели данных в виде линейной векторной решетки

Алгоритмические методы построения КУМ в виде решетки L_Ψ являются весьма важными для практики. Однако исследования показали, что алгоритмы данного типа имеют комбинаторный характер и обычно обладают экспоненциальной асимптотической трудоемкостью, улучшить которую очень сложно. Вместе с тем решение ряда реальных примеров обнаружило,

что фактическая трудоемкость часто является невысокой и, более того, она значительно меньше трудоемкости многих других, известных алгоритмов.

Основной алгоритм построения решетки L_Ψ с помощью оператора за-

мыкания $\hat{A} = \bigcap R_A$ в качестве исходных данных использует семейство порождающих множеств $\{R\}$ решетки L_Ψ и еще, может быть, множество $P = \bigcup_{R \in \{R\}} R$, расширяющее (доопределяющее) семейство $\{R\}$. Исходные и

результатирующие данные алгоритма удовлетворяют условиям $\{R\} \subseteq 2^P$,

$|\{R\}| = m \leq 2^{|P|}$, $(A \in L_\Psi) \Leftrightarrow (A = \bigcap R_A)$. Обоснование алгоритма и оценка его

трудоемкости приводится в работе [1]. Трудоемкость алгоритма определяется в основном числом элементов решетки L_Ψ . Алгоритм является сходящимся и

строит решетку, общее число элементов в которой не превышает величины

$\min \left\{ 2^m, 1 + \sum_{i=0}^n \binom{|P|}{i} \right\}$ при $n < |P|$ или величины $\min \{ 2^m, 2^n \}$ при $n = |P|$, где

$m = |\{R\}|$, $n = \max_{R \in \{R\}} |R|$, $\{R\}$ – семейство порождающих множеств решетки

L_Ψ и $P = \bigcup R$ [1]. В практических задачах анализа данных обычно выполня-

ются условия $m \gg n$ и $n < |P|$, тогда трудоемкость алгоритма построения

решетки L_Ψ определяется величиной $O \left(\sum_{i=0}^n \binom{|P|}{i} \right)$, которая может быть су-

щественно уменьшена в случае задания ограничений на мощность получае-
мых элементов решетки.

Таким образом, предельная алгоритмическая сложность для любых способов теоретико-решетчатого представления структуры данных имеет асимптотический экспоненциальный характер, так как количество элементов решеток L_ϵ , L_ϕ , L_Ψ и $\bar{L}_\epsilon$ одинаково. Однако реализация структуры данных

линейной векторной решетки $\bar{L}_\epsilon$ достаточно просто может быть выполнена в

виде списка со связями [7]. В этом случае для обработки данных может быть

применена цифровая сортировка с трудоемкостью $O(n)$. Представление эле-

ментов решетки в виде двоичных векторов позволяет эффективно использо-

вать в алгоритмах параллельные поразрядные операции над векторами как

операции над множествами. В работе [8] рассмотрены вопросы применения

параллельных методов вычислений для уменьшения трудоемкости алгорит-
мов моделирования данных с помощью решеток.

4. Пример анализа данных

Рассмотрим пример базы данных, который приводится в качестве иллюстрации в различных работах по анализу ассоциаций данных [1]. База дан-
ных представлена как порождающее семейство множеств (транзакций)

$\{R\} = \{R_1, R_2, R_3, R_4\}$. Каждая транзакция состоит из множества позиций: $R_1 = \{1, 3, 4\}$, $R_2 = \{2, 3, 5\}$, $R_3 = \{1, 2, 3, 5\}$, $R_4 = \{2, 5\}$. Используя оператор замыкания $\hat{A} = \bigcap R_A$, путем построения таблиц, последовательного определения попарных пересечений множеств семейства $\{R\}$ и выделения существенных элементов с помощью алгоритма [1] построим решетку L_Ψ , которая используется для анализа ассоциаций данных. Структурный нуль и структурная единица решетки определяются следующими выражениями:

$$\begin{aligned} \mathbf{0} = \hat{\emptyset} &= \bigcap_{R \in \{R\}} R = R_1 \cap R_2 \cap R_3 \cap R_4 = \\ &= \{1, 3, 4\} \cap \{2, 3, 5\} \cap \{1, 2, 3, 5\} \cap \{2, 5\} = \emptyset, \\ \mathbf{1} = P &= \bigcup_{R \in \{R\}} R = R_1 \cup R_2 \cup R_3 \cup R_4 = \\ &= \{1, 3, 4\} \cup \{2, 3, 5\} \cup \{1, 2, 3, 5\} \cup \{2, 5\} = \{1, 2, 3, 4, 5\}. \end{aligned}$$

Диаграмма Хассе полученной решетки L_Ψ базы данных транзакций показана на рис. 1,а. Изоморфная решетка L_ϵ показана на той же диаграмме путем обозначения элементов решетки двоичными кодами, соответствующими характеристическим векторам решетки. На рис. 1,б приводится диаграмма Хассе линейной векторной решетки $\vec{L}_\epsilon$, в которую преобразована решетка L_ϵ . Штрихом связаны элементы, несравнимые в L_ϵ . Выделены транзитивно замыкающие дуги диаграммы линейной решетки $\vec{L}_\epsilon$, которые входят в диаграмму решетки L_ϵ . Векторы линейной решетки $\vec{L}_\epsilon$ (рис. 1,б) показаны с обратным порядком следования разрядов, что иногда удобно для представления вектора как двоичного числа (старшие разряды слева).

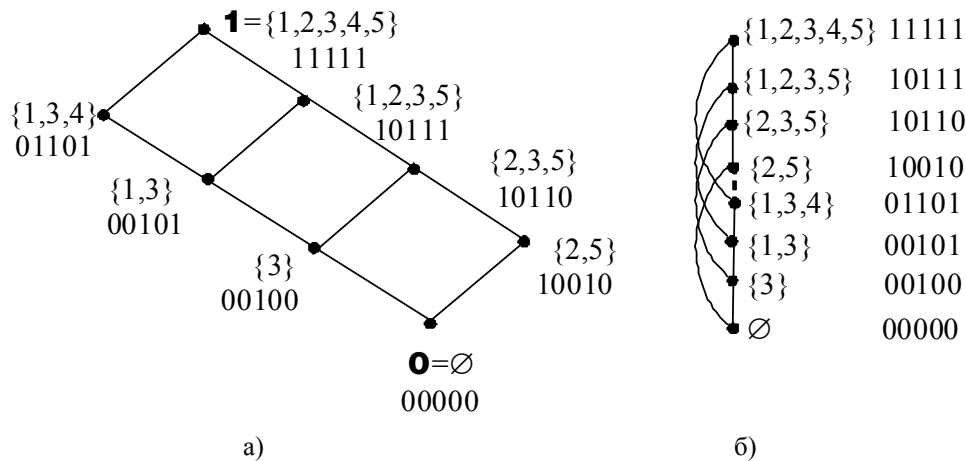


Рис. 1. Диаграммы решеток: а – L_Ψ и L_ϵ ; б – $\vec{L}_\epsilon$

Рассмотрим некоторые свойства приведенной модели данных. Например, в решетке $\vec{L}_\epsilon$ выполняется отношение $R_1 \leq R_4$, т.е. $(01101) \leq (10010)$, но в решетке L_ψ элементы R_1 и R_4 несравнимы. Если $R_4 \leq R_2$ в решетке L_ϵ или $(10010) \subseteq (10110)$, то согласно свойству 1 $R_4 \leq R_2$ в решетке $\vec{L}_\epsilon$ или $(10010) \leq (10110)$. Справедливо выражение $R_3 \wedge R_4 = R_4$, следовательно, согласно свойству 2 $R_4 \leq R_3$, или $(10010) \leq (10111)$ (рис. 1,б). В решетке $\vec{L}_\epsilon$ выполняется отношение $R_1 \wedge R_3 = \{1,3\} = (00101)$, тогда согласно свойствам 1 и 3 справедливо $\{1,3\} \leq R_1$ и $\{1,3\} \leq R_3$, или $(00101) \leq (01101)$, $(00101) \leq (10111)$, а согласно свойству 4, справедливо $R_1 \wedge R_3 = \{1,3\}$.

Заключение

Рассмотренный метод представления моделью в виде линейной векторной решетки $\vec{L}_\epsilon$ позволяет увеличить эффективность моделирования за счет увеличения быстродействия алгоритма построения решеток, что приводит к увеличению допустимой размерности решаемых задач анализа данных в среднем приблизительно на 30 %.

Список литературы

1. Лебедев, В. Б. Анализ ассоциаций данных методом комбинаторно-упорядоченного моделирования / В. Б. Лебедев // Известия высших учебных заведений. Поволжский регион. Сер. «Технические науки». – 2005. – № 5 (20). – С. 99–106.
2. Лебедев, В. Б. Моделирование структуры данных методами теории решеток / В. Б. Лебедев // Проблемы информатики в образовании, управлении, экономике и технике : сборник статей X Международной научно-технической конференции. – Пенза : Приволжский дом знаний, 2010. – С. 41–45.
3. Лебедев, В. Б. Построение изоморфных решеток в задаче дискретной классификации / В. Б. Лебедев, В. Е. Минаев // Университетское образование : сборник статей XII Международной научно-методической конференции. – Пенза : Изд-во ПГУ, 2008. – С. 247–249.
4. Лебедев, В. Б. Онтологическая модель знаний в АОС на основе решетки данных / В. Б. Лебедев, Е. В. Паршина // Университетское образование : сборник статей XIV Международной научно-методической конференции. – Пенза : Поволжский дом знаний, 2010. – С. 610–612.
5. Лебедев, В. Б. Анализ телекоммуникационных услуг методами теории решеток / В. Б. Лебедев, Е. В. Паршина // Проблемы информатики в образовании, управлении, экономике и технике : сборник статей XI Международной научно-технической конференции – Пенза : Приволжский дом знаний, 2011. – С. 212–214.
6. Айгнер, М. Комбинаторная теория / М. Айгнер. – М. : Мир, 1982. – 558 с.
7. Ахо, А. Построение и анализ вычислительных алгоритмов : пер. с англ. / А. Ахо, Дж. Хопкрофт, Дж. Ульман. – М. : Мир, 1979. – 536 с.
8. Лебедев, В. Б. О методах распараллеливания алгоритма дискретной классификации на основе комбинаторно-упорядоченных моделей / В. Б. Лебедев, В. Е. Минаев // Перспективы развития систем среднего и высшего профессионального образования в современном обществе : сборник статей Международной научно-практической конференции. – Пенза : Поволжский дом знаний, 2008. – Ч. 1. – С. 125–127.

Лебедев Виктор Борисович

доктор технических наук, профессор,
кафедра информационного обеспечения
управления и производства, Пензенский
государственный университет

E-mail: inoup@pnzgu.ru

Lebedev Viktor Borisovich

Doctor of engineering sciences, professor,
sub-department of dataware,
information management and production,
Penza State University

Паршина Елена Валерьевна

аспирант, Пензенский
государственный университет

E-mail: inoup@pnzgu.ru

Parshina Elena Valeryevna

Postgraduate student,
Penza State University

Песошин Валерий Андреевич

доктор технических наук, профессор,
заведующий кафедрой компьютерных
систем, Казанский национальный
исследовательский технический
университет им. А. Н. Туполева

E-mail: pesoshin@evm.kstu-kai.ru

Pesoshin Valeriy Andreevich

Doctor of engineering sciences, professor,
head of sub-department of computer system,
Kazan National Research Technical
University named after A. N. Tupolev

УДК 004.62

Лебедев, В. Б.

Эффективная модель анализа данных на основе линейных векторных решеток / В. Б. Лебедев, Е. В. Паршина, В. А. Песошин // Известия высших учебных заведений. Поволжский регион. Технические науки. — 2011. — № 4 (20). — С. 19–25.

МЕТОДОЛОГИЧЕСКИЕ ПРИНЦИПЫ ОРГАНИЗАЦИИ МНОГОПОТОКОВОЙ ОБРАБОТКИ ДАННЫХ С РАСПАРАЛЛЕЛИВАНИЕМ ВЫЧИСЛИТЕЛЬНЫХ ПРОЦЕССОВ

Аннотация. Предлагаются методологические принципы создания авиационных тренажеров с использованием многопроцессорных информационно-вычислительных комплексов с распараллеливанием вычислительных процессов в реальном масштабе времени. Приводится их реализация при разработке тренажера транспортного самолета.

Ключевые слова: многопоточковая обработка данных, модульная архитектура, многопроцессорные вычислительные комплексы.

Abstract. The article describes methodological principles of construction of flight simulators using multi-information-processing systems with parallel computing processes in real time. The authors present their implementation in the course of development of a transport aircraft simulator.

Key words: multi-threaded processing, modular architecture, multiprocessor computer systems.

Введение

Концепция модульности предполагает возможность объединения, разделения и модификации отдельных элементов без их влияния на систему в целом [1–5]. При модульной архитектуре систем модули могут создаваться независимо друг от друга и объединяться в блоки для получения необходимых результатов. Многие элементы модульности (модули пилотажных приборов, силовой установки, подвижности, кабина, вычислитель, пульт инструктора и др.) уже используются в современных тренажерах. Однако до последнего времени объединение модулей традиционно требовало больших временных и финансовых затрат (иногда ожидаемая выгода не достигалась или требовались чрезмерные усилия). Так, во многих пилотажных приборах используются аналоговые данные, а в ряде других приборов и вычислителях используются цифровые данные. Налицо обмен в авиационных тренажерах (АТ) большими потоками различной информации. Наибольшие трудности связаны с необходимостью выполнения всех операций в реальном масштабе времени.

1. Описание и постановка задачи

Модульный подход, облегчая некоторые трудности, налагает дополнительные ограничения, связанные с приведением данных в совместимую форму, на систему в целом. Однако если указанные проблемы будут решены, то присущая модульному подходу гибкость будет значительно перекрывать указанные ограничения по обработке данных. Вычисления могут распределяться между различными процессорами.

Существует потенциальная опасность выбора узкоспециализированного подхода с определением некоторой архитектуры ЭВМ и установкой жесткой структуры интерфейса с применением специального языка программиро-

вания. Должна существовать возможность создания новых необходимых модулей на основе единого подхода для обеспечения совместимости модулей друг с другом.

При модульном подходе модули и интерфейс могут рассматриваться с функциональной точки зрения (логический уровень) или система рассматривается как набор аппаратных и программных модулей (физический уровень). Указанное разделение позволяет достичь основной цели – разделить указанные уровни так, чтобы изменения на одном уровне не вызывали изменений на другом. Так, можно модернизировать модули акселерационных эффектов, визуализации, установить другие ЭВМ и т.д. без изменения логической структуры АТ или наоборот, использовать различные элементы физического уровня для создания конкретных АТ.

Решение вопроса о взаимодействии модулей на самом деле является решением задачи передачи данных от модуля к модулю. Для их связи необходима схема сопряжения (в идеале – универсальная).

Специального рассмотрения при модульном подходе требует вопрос программного обеспечения. Наибольшие усилия затрачены на обеспечение мобильности программного обеспечения. Доля стоимости математического обеспечения в общей стоимости АТ постоянно возрастает. Поэтому повышение эффективности труда программиста становится все более важным.

Во избежание сложности внедрения технических новшеств требования к модулям, интерфейсу и т.д. не должны быть чрезмерно жесткими и замыкаться на существующих технологиях. Залог успеха – в создании достаточно гибких модулей и системы в целом. Модуль должен определяться функционально. Типы используемых ЭВМ, длина слов, язык программирования и т.п. не должны входить в определение модуля. При определении модулей, интерфейса, утверждении стандартов (для исключения разногласий) головная организация, координируя деятельность подрядчиков, должна исполнять и роль эксперта.

Рассмотренный подход использовался при разработке ряда комплексных тренажеров самолетов Ил-76МД, Ил-96-300, Ту-204, Ан-72, Ан-74, широко используемых для подготовки летного состава в учебно-тренировочных центрах как гражданской, так и военной авиации.

2. Методы решения задач

Ограничимся рассмотрением модульного построения архитектуры комплексного тренажера на примере тренажера самолета Ту-204.

В состав тренажера входят:

- кабина с рабочими местами двух пилотов, бортинженера и инструктора, устанавливаемая на шестистепенном стенде имитатора акселерационных эффектов;
- базовый вычислитель, включающий ЭВМ и устройство комплексирования;
- модуль информационного обмена, состоящий из стоек абонентских терминалов, расположенных вне кабины и в кабине;
- модуль имитатора визуальной обстановки, включающий генератор изображения и модуль отображения визуальной обстановки;
- рабочее место инструктора вне кабины тренажера;
- две стойки с бортовыми ЦВМ;

- модуль имитатора метеонавигационной радиолокационной станции МНРЛС-85;
- модуль общего электропитания с автономным комплектом первичного электропитания для базового вычислителя тренажера;
- модуль кондиционирования воздуха в кабине тренажера.

На рис. 1 приводится модульная структура комплексного тренажера самолета Ту-204.

Возможны два варианта модульного построения АТ.

В первом варианте (связи показаны сплошными линиями) к каждому вычислительному комплексу (ВК) через интерфейс типа «Общая шина» подключается контроллер одной стойки абонентских терминалов с модулями системы информационного обмена (СИО) и используемым бортовым оборудованием. К каждой стойке абонентских терминалов бортовое оборудование подключается с учетом задач, решаемых вычислительным комплексом. Например, во втором вычислительном комплексе решаются задачи динамики полета, системы дистанционного управления, автоматической системы штурвального управления.

К стойке абонентских терминалов, взаимодействующих с вычислительным комплексом два (ВК 2), подключаются органы управления полетом, бортовое оборудование систем, связанных с вычислительной системой управления полетом, вычислительной системой управления тягой, системой предупреждения критических режимов, системой предупреждения о приближении к земле, а также модуль имитатора акселерационных эффектов. Для уменьшения запаздывания к общей шине ВК 2 подключается модуль имитации визуальной обстановки. Здесь возможно создание тренажера из функционально-законченных модулей:

- пилотажного модуля (включает ВК 2, ВК 3, две стойки абонентских терминалов и соответствующее оборудование);
- модуля бортинженера (включает ВК 4, стойку абонентских терминалов, приборы, органы управления, сигнализаторы силовой установки, бортовых систем контроля двигателей, топливной системы, системы регулирования давления, комплексную информационную систему сигнализации);
- модуля самолетовождения (включает ВК 5, стойку абонентских терминалов, бортовое оборудование системы самолетовождения и радиотехнических систем; к общей шине ВК 5 подключается имитатор метеонавигационной радиолокационной станции);
- модуля инструктора (включает ВК 1, стойку абонентских терминалов, оборудование рабочего места инструктора (РМИ)).

Такое построение позволяет разрабатывать и специализированные тренажеры (естественно, с объемом решаемых задач, меньшим, чем в комплексном АТ). Запаздывание от управляющего воздействия от штурвала до изменения изображения в имитаторе визуальной обстановки определяется запаздыванием собственно имитатора визуальной обстановки и 1–2 циклами решения задач.

Во втором варианте (связи показаны пунктирными линиями на рис. 1) контроллеры стоек абонентских терминалов подключаются не к интерфейсу «Общая шина» ВК 1 – ВК 5, а к магистрали модуля передачи данных (подключается к одному из входов общего поля памяти оперативного запоминающего устройства (ОЗУ)).

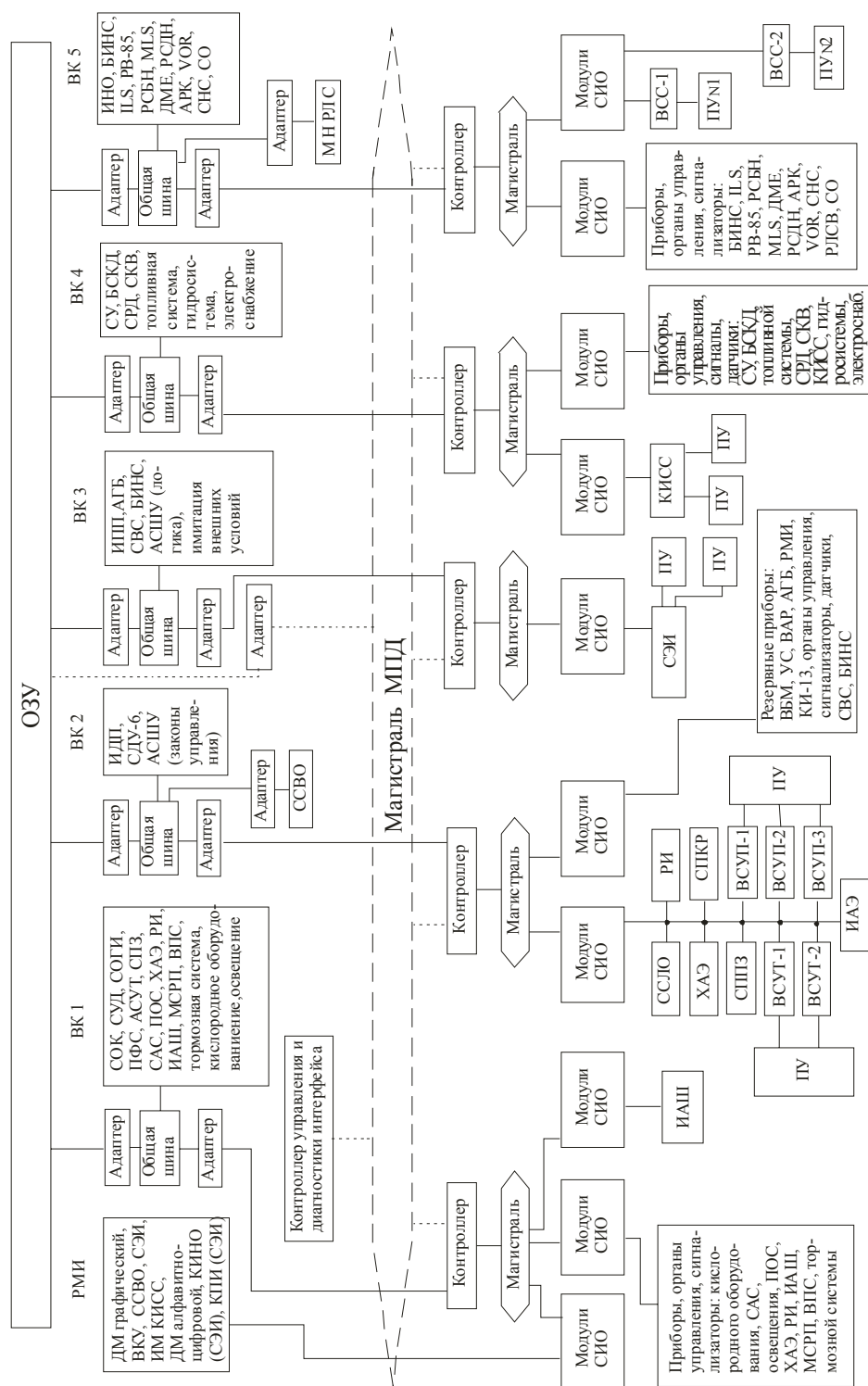


Рис. 1. Структурная схема комплексного тренажера самолета

Здесь устройство комплексирования ОЗУ, кроме функций обмена данными между ВК 1 – ВК 5, выполняет функции приема и передачи данных от оборудования кабины и рабочего места инструктора. Это позволяет создать гибкую структуру вычислительной системы тренажера, обеспечивающую перераспределение потоков информации (позволяет уменьшить количество коммутирующих средств). Эксплуатационная надежность тренажера во многом определяется надежностью устройства комплексирования (отказ устройства приведет к отказу тренажера в целом).

С учетом изложенного выше очевидна целесообразность построения тренажера по первому варианту, что практически и реализовано.

Ограничимся рассмотрением основных модулей.

Модуль имитатора динамики полета. Структурная схема с указанием взаимосвязей с другими имитаторами приводится на рис. 2.

Воспроизводится движение самолета в пространстве (решение замкнутой системы нелинейных дифференциальных уравнений; входные параметры – управляющие воздействия экипажа, выходные – параметры полета).

Моделируются:

- движение по взлетно-посадочной полосе (ВПП) и рулежным дорожкам;

- взлет и набор высоты;

- полет по маршруту, снижение и заход на посадку;

- уход на второй круг (с использованием средств комплекса стандартного пилотажно-навигационного оборудования);

- экстренное снижение;

- полет по кругу, заход на посадку и посадка;

- пробег по ВПП с использованием всех средств торможения;

- полет при опасных внешних воздействиях.

Учитываются:

- температура воздуха и атмосферное давление;

- высота (уровень местности) аэродрома;

- горизонтальная составляющая скорости ветра;

- сдвиг и порывы ветра (вертикальный и горизонтальный);

- обледенение (изменение аэродинамических характеристик);

- состояние ВПП;

- сила тяжести и центровка самолета;

- режимы работы силовой установки (включая реверс);

- положение управляющих поверхностей, механизация крыла, шасси (при движении по земле учитываются характеристики устойчивости и управляемости; коэффициенты увода и трения колес с учетом состояния ВПП, торможения колес и работы антиюзového устройства);

- аэроупругость.

Для обеспечения подобия реального и моделируемого полетов самолета непрерывно вычисляются параметры полета в реальном масштабе времени. Экипажу с соответствующих модулей предоставляется визуальная, акустическая, акселерационная информация, показания приборов пилотажно-навигационного комплекса, положения и усилия на органах управления (формируемых в соответствии с параметрами, получаемыми в имитаторе динамики полета (ИДП)).

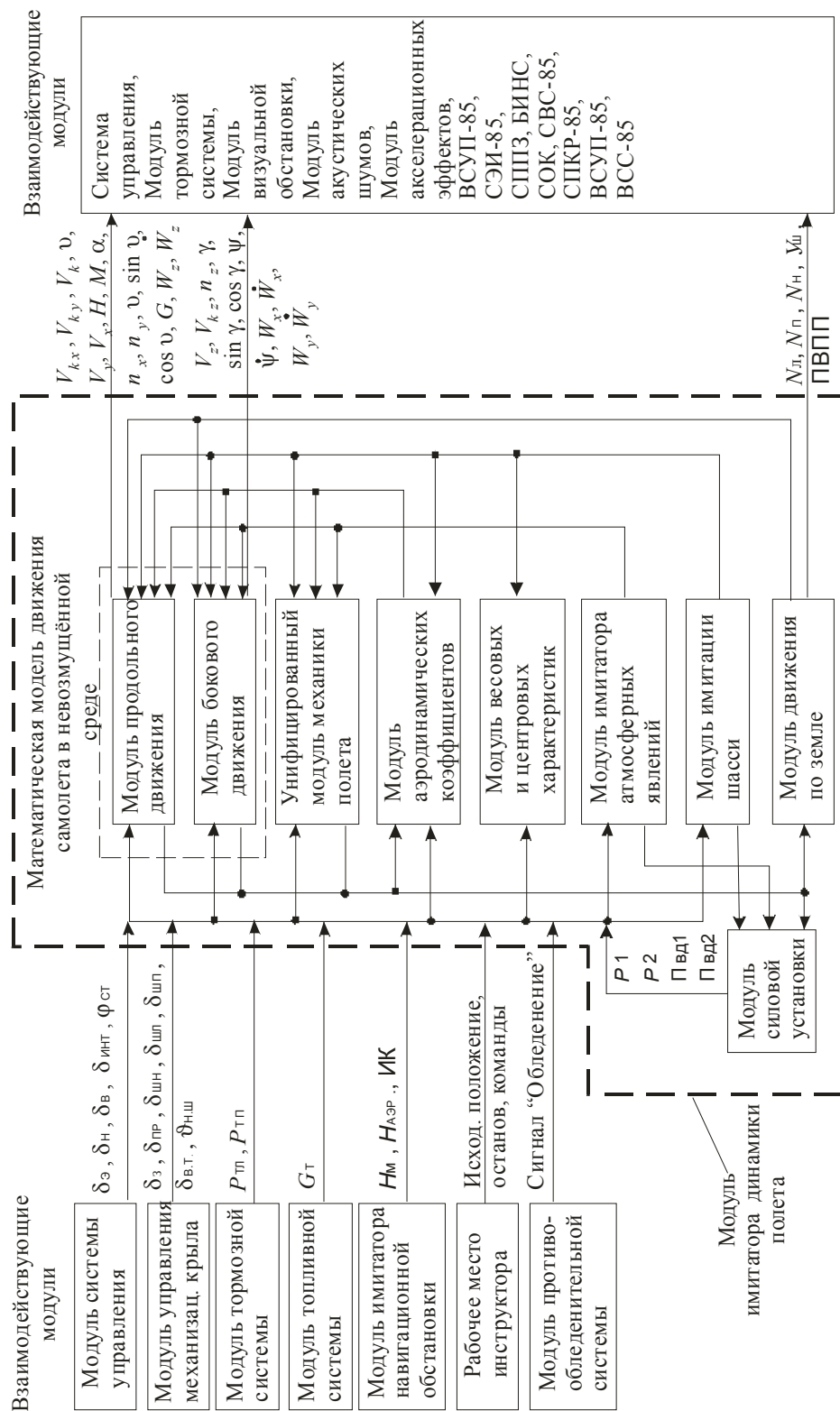


Рис. 2. Структурная схема модуля имитатора динамики полета

Производится декомпозиция системы уравнений движения (продольное и боковое движения, движение по земле, модуль аэродинамических коэффициентов).

Шаг интегрирования задается из условий устойчивости и мощности ЭВМ.

Предусматривается возможность программного ввода и вывода значений физических параметров на дисплей.

Метод моделирования позволяет обеспечить воспроизведение динамических характеристик в полном соответствии с «Нормами годности авиационных тренажеров для подготовки авиаперсонала воздушного транспорта».

Модуль имитатора акселерационных эффектов предназначается для воспроизведения составляющих акселерационного поля информации, связанных с изменениями линейных перегрузок, угловых ускорений и перемещений, характерных для различных режимов полета, а также воспроизведения эффектов:

- постоянного ненаправленного «фонового» движения в процессе всего полета;
- вибрации от работающих двигателей;
- шумов от выпуска и уборки шасси;
- обжатия стоек шасси при торможении, растормаживании, а также эффектов, вызванных их касанием ВПП, включая раздельное касание носовой, левой и правой стойками;
- тряски при движении по ВПП;
- сваливания.

Модуль также позволяет одновременно имитировать пространственное маневрирование и движение, вызванное внешним воздействием.

Разработка динамических стендов (включая системы управления ими) осуществляется на основе компромисса между возможно более точным моделированием (в наземных условиях) акселерационных ощущений летчика и жесткими техническими ограничениями на максимальные линейные (в меньшей степени на угловые) перемещения кабины стенда (во много раз меньшие реальных перемещений самолета). Точное моделирование акселерационных ощущений летчика во всем диапазоне частот по линейным степеням свободы принципиально невозможно, а по угловым – затруднительно. Противоречие устраняется с использованием особенностей восприятия перегрузок летчиком. В имитаторе акселерационных эффектов воспроизводятся частоты 0,3–3 Гц. Задаваемые перемещения определяются с учетом их допустимого диапазона (независимо от способа формирования управляющих сигналов).

Для того чтобы при моделировании низких частот в изменении перегрузки кабина тренажера не только не выходила за ограничения по перемещению, но и незаметно для летчика стремилась вернуться в среднее, нейтральное положение, используются фильтры верхних частот. Воспроизведение на стенде длительно действующих линейных, боковых и продольных перегрузок (в диапазоне не более $\pm 0,5g$) обеспечивается возможностью наклона кабины на соответствующий угол. При разработке имитатора используются основы формирования ощущений наклонов у человека по реакциям отолитов, тактильных рецепторов. При разгоне и торможении самолета результирующая массовых сил, действующая на рецепторы, вызывает ощущение наклона по тангажу. Кабина тренажера наклоняется на соответствующую

ший угол тангажа (линия горизонта на экране перед летчиком не изменяет своего положения). Аналогично строится имитатор для боковых перегрузок.

При имитации длительно действующих перегрузок учитывается взаимное расположение отолитового аппарата (в районе головы летчика) и центра вращения самолета (отолиты воспринимают перегрузки от возникающих при вращении центробежных сил). Одновременно с вращением подвижной платформы задаются линейные перемещения (табл. 1). Как в канале крена, так и тангажа используются фильтры.

Таблица 1

Рабочие характеристики динамического стенда

Параметр движения	Диапазон перемещения	Максимальная скорость	Максимальное ускорение
Продольное	$\pm 0,6$ м	$\pm 0,8$ м/с	± 8 м/с ²
Вертикальное	$\pm 0,5$ м	$+ 0,7...0,8$ м/с	± 8 м/с ²
Боковое	$\pm 0,6$ м	$\pm 0,8$ м/с	± 8 м/с ²
Крен	$\pm 24^\circ$	± 20 °/с	± 100 °/с ²
Тангаж	$\pm 21^\circ$	± 20 °/с	± 100 °/с ²
Рыскание	$\pm 23^\circ$	± 20 °/с	± 100 °/с ²

Заключение

1. Рассматриваются методологические принципы создания авиационных тренажеров модульной архитектуры с использованием системного подхода.

2. Приводятся структура, состав, а также указываются взаимосвязи информационно-вычислительных систем тренажеров, предусматривающих распараллеливание вычислительных процессов.

3. Предлагаются варианты модульного построения тренажеров.

4. Результаты исследований использованы при разработке динамических стендов и комплексных тренажеров самолетов, широко используемых для подготовки летного состава в учебно-тренировочных центрах.

5. Приведенные принципы разработки авиационных тренажеров могут использоваться при проектировании и других управляемых в пространстве динамических систем.

Список литературы

1. **Анисимов, А. Н.** Методологические принципы проектирования сложных управляемых в пространстве динамических систем в приложении к разработке авиационных тренажеров / А. Н. Анисимов, А. М. Данилов, И. А. Гарькина, Б. В. Ключев, Э. В. Лапшин // Идентификация систем и задачи управления SICPRO'04 : труды III Международной конференции (Москва, 28–30 января 2004 г.). – М. : ИПУ РАН им. В. А.Трапезникова, 2004. – С. 279–311.
2. **Данилов, А. М.** Динамика и управление внеатмосферными астрономическими обсерваториями / А. М. Данилов, Л. З. Дулькин, А. С. Земляков, В. М. Матросов, В. А. Стрежнев // Труды VI Международного симпозиума ИФАК по управлению в пространстве : в 2 т. Т. 1. Управление в пространстве. – М. : Наука, 1976. – С. 153–171.
3. **Данилов, А. М.** Модульная архитектура пилотажных авиационных тренажеров на базе ПЭВМ / А. М. Данилов, Э. В. Лапшин, В. Н. Шихалеев // Тренажеры и

компьютеризация профессиональной подготовки : труды III Всесоюзной научно-технической конференции. – Калининград, 1991. – С. 212–213.

4. **Гарькина, И. А.** Системная методология идентификации сложных динамических систем. Аэрокосмические приложения / И. А. Гарькина, А. М. Данилов, А. П. Прошин // Аэрокосмические приборные технологии : сборник материалов III Международного симпозиума (Санкт-Петербург, 2–4 июня 2004 г.). – СПб., 2004. – С. 39–43.
5. **Данилов, А. М.** Теория и практика имитационного моделирования и создания тренажеров / А. М. Данилов, Э. В. Лапшин // Приборы и системы управления. – 1989. – № 8. – С. 11–12.

Данилов Александр Максимович

доктор технических наук, профессор,
заведующий кафедрой высшей
математики, Пензенский
государственный университет
архитектуры и строительства

E-mail: fmatem@pguas

Danilov Aleksander Maksimovich

Doctor of engineering sciences, professor,
head of sub-department of higher
mathematics, Penza State University
of Architecture and Construction

Лапшин Эдуард Владимирович

доктор технических наук, профессор,
кафедра конструирования
и производства радиоаппаратуры,
Пензенский государственный
университет

E-mail: edlapshin@mail.ru

Lapshin Eduard Vladimirovich

Doctor of engineering sciences, professor,
sub-department of radio devices engineering
and production, Penza State University

Беликов Георгий Георгиевич

доцент, кафедра конструирования
и производства радиоаппаратуры,
Пензенский государственный
университет

E-mail: edlapshin@mail.ru

Belikov Georgy Georgievich

associate professor, sub-department
of radio devices engineering
and production, Penza State University

Лебедев Виктор Борисович

доктор технических наук, профессор,
кафедра информационного обеспечения
управления и производства, Пензенский
государственный университет

E-mail: inoup@pnzgu.ru

Lebedev Viktor Borisovich

Doctor of engineering sciences, professor,
sub-department of dataware, information
management and production,
Penza State University

УДК 629.73.08

Данилов, А. М.

Методологические принципы организации многопоточковой обработки данных с распараллеливанием вычислительных процессов / А. М. Данилов, Э. В. Лапшин, Г. Г. Беликов, В. Б. Лебедев // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 26–34.

А. А. Архангельский, Е. И. Гурин, Б. А. Савельев

ПОДКЛЮЧЕНИЕ СПЕЦИАЛИЗИРОВАННЫХ ПРОЦЕССОРОВ К ПЕРСОНАЛЬНОМУ КОМПЬЮТЕРУ С ПРИМЕНЕНИЕМ ИНТЕРФЕЙСА USB

Аннотация. Рассматриваются вопросы подключения специализированных вычислительных устройств, реализованных на программируемых логических интегральных схемах, к персональной ЭВМ по интерфейсу USB. Приводятся примеры реализации на языке VHDL и результаты экспериментального исследования.

Ключевые слова: интерфейс USB, программируемые логические интегральные схемы, язык VHDL, персональный компьютер.

Abstract. The article considers the problems of connection of specialized computers, realized on the basis of FPGA, to a personal computer via USB. The authors give examples of the realization in VHDL language and results of the experimental research.

Key words: USB interface, FPGA, VHDL, personal computer.

Введение

Для связи внешних устройств с ЭВМ широко применяются последовательные интерфейсы. Они интересны тем, что позволяют объединить множество устройств, используя небольшое количество проводов. До недавнего времени последовательные интерфейсы персонального компьютера были представлены коммуникационным портом, работающим согласно спецификации RS-232. Интерфейс RS-232 сохраняет все преимущества последовательной связи, однако он имеет ряд недостатков, в том числе низкое быстродействие, плохую помехозащищенность, отсутствие гальванической развязки. Поэтому в настоящее время все большее распространение получает интерфейс USB. Применение интерфейса USB требует разработки устройств для реализации достаточно сложных алгоритмов обмена. Упростить процедуры обмена позволяют микросхемы преобразования интерфейсов, которые берут на себя реализацию этих алгоритмов. Такие микросхемы широко применяются для подключения устройств пользователя [1–3]. Кроме того, на их основе выпускаются готовые модули, которые еще более упрощают использование интерфейса USB. В работе рассматривается сопряжение устройств с компьютером по USB с использованием модулей USB.

1. Разработка аппаратной части

На рис. 1 показана схема с типовым модулем М, обеспечивающим сопряжение процессора, реализованного на ПЛИС, с персональной ЭВМ типа IBM PC обычным USB-кабелем. Основу модуля М может составлять микросхема FT245 фирмы FTDI; описание этой микросхемы [4] можно найти на сайте www.ftdichip.com. Микросхема обеспечивает преобразование данных из интерфейса USB в параллельный интерфейс. Синхронизация работы модуля производится от тактового генератора, установленного на модуле. Для составления программ обмена используются драйверы, которые можно бесплатно скачать с сайта www.ftdichip.com.

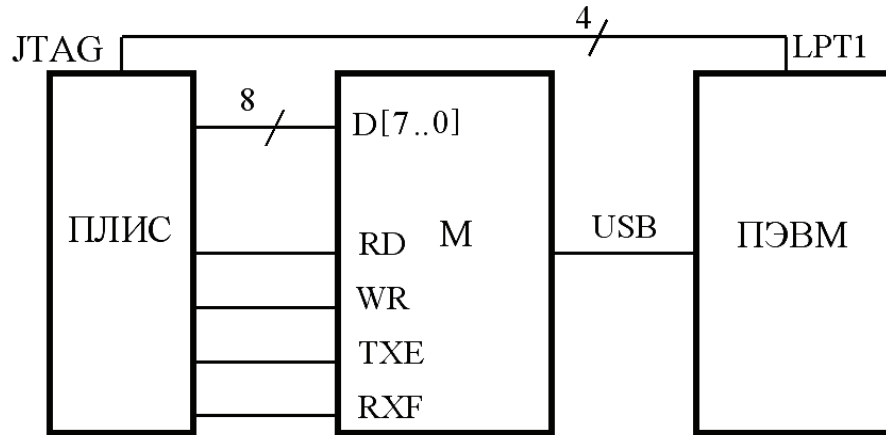


Рис. 1. Схема подключения типового модуля USB

Обмен данными между ПЛИС и модулем M производится по двунаправленной шине D[7..0], управление обменом производится сигналами RD, WR, TXE и RXF. При чтении данных из модуля USB в ПЛИС согласно техническому описанию микросхемы FT245 должна быть реализована следующая диаграмма (рис. 2,а).

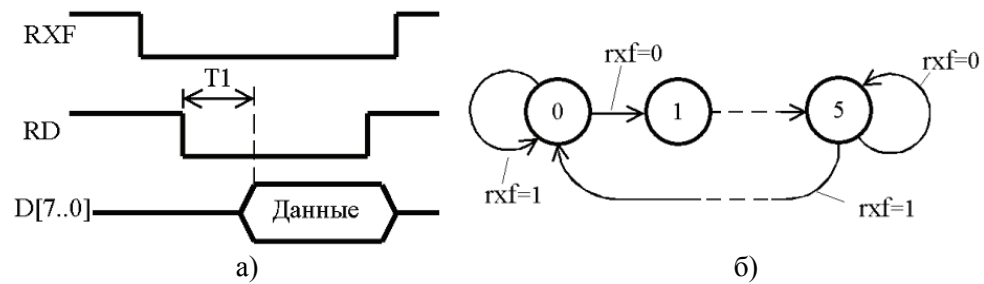


Рис. 2. Диаграммы обмена данными

Микросхема FT245 формирует сигнал RXF (активное значение ноль) в момент готовности данных, передаваемых в устройство пользователя. В ответ устройство пользователя формирует сигнал чтения RD (активное значение – ноль), по которому микросхема FT245 формирует передаваемые данные D[7..0]. Согласно техническому описанию временной интервал T1 от начала RD до выдачи данных составляет от 20 до 50 нс. Таким образом, от момента формирования RD до принятия данных при частоте 50 МГц надо выполнить ожидание не менее трех тактов синхросигнала.

На рис. 2,б показан граф переходов данной схемы. Начальным состоянием автомата записи является состояние 0, схема стартует при получении сигнала готовности $gxf=0$. Следующие четыре состояния автомат проходит при любых значениях сигналов, при прохождении этих состояний вырабатывается активное значение сигнала чтения rd . На рис. 2,б не показаны состояния 2–4, последовательность этих состояний обеспечивает подсчет необходимого числа тактов синхросигнала частотой 50 МГц для выполнения временных соотношений обмена, указанных в техническом описании. В состоянии 4 данные принимаются на внутренний регистр `data_rd`, этим обеспечива-

ется задержка приема данных на три такта (интервал T1 на рис. 2,а), т.е. не менее чем на 50 нс. В состоянии 5 ожидается окончание сигнала gxf.

Аналогичные последовательности реализуются при записи данных из ПЛИС в модуль USB. Если микросхема FT245 готова принять от ПЛИС данные и передать их в компьютер, то сигнал TXE устанавливается в ноль. Для записи данных ПЛИС формирует сигнал WR (активное значение – единица), а затем производит выдачу данных D[7..0]. Согласно техническому описанию длительность сигнала WR должна составлять не менее 50 нс, т.е. не менее трех тактов синхросигнала частотой 50 МГц.

Разработка схемы, выполняющей обмен и записываемой в ПЛИС, осуществлялась на языке VHDL. Схема работает в двух режимах: чтения и записи. Первоначально схема находится в режиме чтения, в котором ПЛИС принимает данные из компьютера. При получении соответствующих команд схема переходит в режим записи, в котором в компьютер передается заданное количество байтов, а затем схема снова переходит в режим чтения. Рассмотрим описание схемы чтения 1 байта, которая работает в соответствии с временной диаграммой, показанной на рис. 2,а:

```

if avt_rd="000" and rxfz='0' and wr_mode='0'    -- 1
then avt_rd<="001";                             -- 2
elsif avt_rd="001" or avt_rd="010"              -- 3
or avt_rd="011" or avt_rd="100" then            -- 4
    avt_rd<=avt_rd+1;                             -- 5
elsif avt_rd="101" and rxfz='1' then            -- 6
    avt_rd<="000";                                 -- 7
end if;                                           -- 8

```

Приведенный выше фрагмент находится внутри процесса, задающего работу схем по синхросигналу clk. Начальным состоянием является состояние 0 ("000"), схема стартует при получении сигнала готовности gxf=0, это условие указано в первой строке. Следующие четыре состояния (строки 3–5) схема проходит при любых значениях сигналов, при прохождении этих состояний вырабатывается активное значение сигнала чтения rd. В состоянии 5 (шестая строка) ожидается окончание сигнала gxf, после чего происходит переход из состояния 5 в состояние 0.

При передаче заданного количества чисел в компьютер схема ожидает команду из компьютера и при поступлении этой команды формирует сигнал режима записи на время передачи в компьютер заданного количества байтов. Так, например, по команде, имеющей код 0Dh, инициируется передача массива 3000 байт. Для проверки временных параметров передачи данных в рассматриваемую схему введен аппаратный таймер, который работает по традиционному принципу и содержит счетчик и два регистра, которые имеют разрядность 32. В нужные моменты состояние счетчика фиксируется в этих регистрах, а затем передается в компьютер для вычисления измеряемых интервалов по разности поступивших чисел.

В настоящей работе использовались модули DLP-USB245M и MMUSB245RL. Для реализации схем на ПЛИС использовались модули Spartan-3E Sample Pack с установленной на ней ПЛИС XC3S100E-TQ144 фирмы Xilinx и LDM-ACEX1K10-T144 с ПЛИС ACEX EP1K10TC144-3N фирмы Altera. Конфигурирование ПЛИС производилось через интерфейс JTAG с помощью параллельного кабеля.

2. Разработка программного обеспечения

Программа управления составлена на языке C. Она посылает в ПЛИС по интерфейсу USB команду 0Dh и другие команды и данные, а также принимает из канала данные, которые при необходимости можно просмотреть на экране дисплея. В начале исходного файла программы нужно вставить строку `#include "ftd2xx.h"`, к проекту нужно подключить библиотеки FTBUSUI.dll, FTD2XX.dll, FTD2XX.lib [5]. Рассмотрим отдельные фрагменты программы, которые обеспечивают связь с устройством и представляют интерес в контексте данной статьи.

Открытие канала USB производится следующими операторами [5]:

```
ftStatus = FT_Open(0, &ftHandle);  
if (ftStatus == FT_OK)  
    printf(" FT Open OK = %x  \n",ftStatus);}  
else // ERROR
```

При успешном открытии функция FT_Open возвращает значение FT_OK, при неуспешном работа прерывается. Переменная ftStatus имеет тип FT_STATUS. Запись одного значения производится следующими командами:

```
ftStatus=FT_Write (ftHandle, &b1, 1, &q1);  
if (ftStatus != FT_OK) // ERROR
```

В приведенном фрагменте b1 – записываемые данные, q1 – число реально записанных байтов, это число возвращает функция FT_Write. В данном примере записывается один байт, однако можно записывать массив байтов, в этом случае вместо &b1 должен присутствовать указатель на массив, а вместо 1 нужно указать количество выводимых байтов. Перед чтением данных желательно проверить содержимое очереди чтения:

```
ftStatus=FT_GetQueueStatus (ftHandle, &n);
```

После выполнения этой функции переменная n примет значение количества готовых к чтению данных. После этого производится чтение:

```
ftStatus=FT_Read (ftHandle, bt, n, &q2);
```

В этой функции q2 – число реально считанных байтов, это число возвращает функция FT_Read, массив bt может быть определен как unsigned char. Если данные не готовы, т.е. n=0, то операцию чтения надо пропустить, либо повторить функцию FT_GetQueueStatus. Если выполнить чтение данных без предварительной проверки, то в случае отсутствия данных программа может зависнуть. После окончания работы с каналом USB надо выполнить функцию FT_Close. Возможен также другой вариант чтения данных, рекомендуемый в [5], который использует функцию задания таймаута FT_SetTimeouts.

3. Экспериментальное исследование

Схема, приведенная на рис. 1, подключалась к компьютеру с помощью обычного кабеля USB. При получении кода 0Dh схема, записанная в ПЛИС, выдает в модуль DLP-USB245M три тысячи байтов данных, причем первое передаваемое число равно 04h, а далее значения уменьшаются на единицу, после 00h выдается FFh, и цикл повторяется несколько раз. На рис. 3 показан

момент перехода схемы, записанной в ПЛИС, из режима чтения в режим записи; момент старта, от которого производится отсчет тактов, отмечен маркером Т.

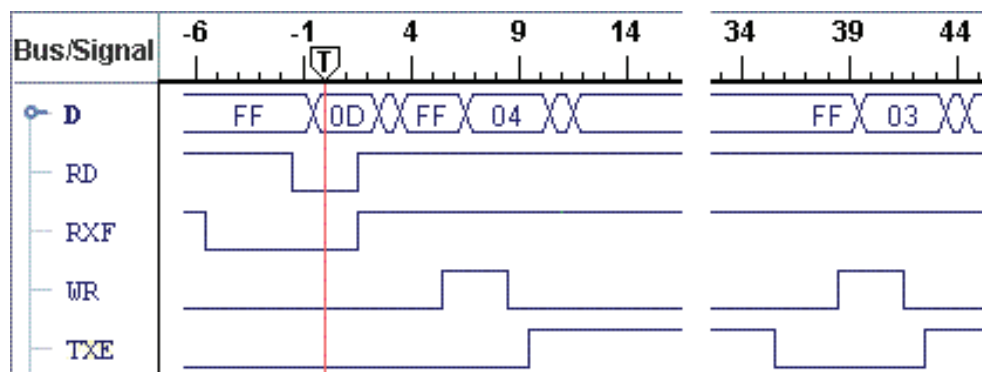


Рис. 3. Временные диаграммы обмена информацией

Диаграмма, приведенная на рис. 3, получена с помощью логического анализатора ChipScope Pro фирмы Xilinx, аппаратная часть которого встраивается в ПЛИС вместе с разрабатываемой схемой. Следует заметить, что на рис. 3 показаны диаграммы внутренних сигналов ПЛИС, сигналы на контактах микросхемы FT245 могут несколько отличаться. Маркером Т отмечен прием из микросхемы FT245BL в ПЛИС числа 0Dh, в тактах с 7 по 9 осуществляется запись числа 04h из ПЛИС в микросхему FT245BL. Следующее число 03h передается в тактах с 40 по 42. В таком же темпе данные передаются и в дальнейшем, темп передачи определяется микросхемой FT245BL.

Максимальная скорость передачи данных в микросхеме FT245BL согласно технической документации составляет 1 Мбайт/с. Из диаграммы на рис. 3 видно, что период передачи данных составляет чуть более 30 тактов, что при частоте 50 МГц равно примерно 0,7 мкс. Это соответствует скорости передачи чуть большей, чем 1 Мбайт/с. Анализ временных диаграмм показывает, что период передачи данных составляет от 0,66 до 0,82 мкс. Кроме того, в процессе передачи возникают задержки от 3 до 4 мкс, вызванные заполнением внутреннего буфера микросхемы FT245BL.

Усредненные временные параметры передачи данных измерялись с помощью аппаратного таймера, встроенного в ПЛИС. Как следует из результатов эксперимента, время передачи 3000 байтов из ПЛИС в ЭВМ в разных экспериментах составляло от 2,38 до 2,44 мс, среднее время передачи одного байта – около 0,8 мкс. Были также получены и проанализированы временные диаграммы приема данных из микросхемы FT245BL в ПЛИС. Анализ диаграмм показывает, что период чтения составляет в среднем 21 такт, что соответствует 0,42 мкс, т.е. данные при чтении поступают в более высоком темпе, чем при записи. Однако при чтении периоды ожидания возникают чаще и сами эти периоды дольше, они могут составлять 30 мкс и более. Были проведены измерения временных параметров операции чтения. В ПЛИС из компьютера передавались 3000 байтов, время передачи измерялось аппаратным таймером и составило 2907,56 мкс, что дает в среднем 0,97 мкс на передачу одного числа, что соответствует техническим данным микросхемы FT245BL.

Для измерения времени работы отдельных участков программы обработки использовался таймер компьютера, данные из которого считывались с использованием драйвера WinIO, который разрешает прямое обращение к портам. Считывание одного отсчета таймера производилось следующими операторами:

```
_outp( 0x43, 0xd2 ); // запись в порт 43h числа d2h
ml=_inp( 0x40 );      // чтение младшего байта
st=_inp( 0x40 );      // чтение старшего байта
t1=( (st<<8) & 0xff00) + (ml & 0x00ff) );
```

Вычисление временных интервалов производится путем вычитания соседних отсчетов. Период работы таймера равен 1/18,2 Гц (примерно 54,9 мс); квант времени, определяющий точность измерения, равен $(1/18,2) \text{ мс}/2^{16}$, что составляет примерно 0,8 мкс.

Как показали результаты измерений, на выполнение программных операций, включающих в себя попеременно передачу и прием небольших порций данных (по одному или двум байтам), требуется значительно больше времени, чем необходимо для работы аппаратуры. Операция передачи одного числа из компьютера в ПЛИС с помощью функции `ftStatus=FT_Write(ftHandle, &b1, 1, &q1)` занимает 1 мс, хотя, как отмечалось выше, в аппаратуре это действие выполняется в среднем за 1 мкс. Время, необходимое на передачу из ЭВМ в ПЛИС одного байта и прием после этого из ПЛИС в ЭВМ двух байт, в разных экспериментах составляло от 2 до 4 мс, а в отдельных случаях – до 15 мс. Время от момента передачи команды 0Dh, по которой ПЛИС начинает передавать массив данных, до момента получения этого массива программой составляет около 18 мс. Такие задержки вызваны особенностями операционной системы Windows и должны учитываться разработчиком. Для сравнения следует отметить, что функции OS Windows, предназначенные для обмена данными с COM портом, работают еще медленнее. Приведенные выше данные получены на персональном компьютере типа IBM PC с процессором AMD Athlon 64 (3000 МГц). Полученные результаты временных параметров могут отличаться от подобных данных, полученных на другом компьютере и при другой загрузке процессора.

На рис. 4 показана схема, содержащая отладочную плату LDM-ACEX1K10-T144 с установленным на ней модулем MMUSB245RL.

Цифрой 1 отмечен модуль MMusb245RL, цифрой 2 – плата LDM-ACEX1K10-T144. Программирование ПЛИС производится кабелем JTAG (отмечен цифрой 3). Основу модуля MMusb245RL составляет микросхема FT245RL фирмы FTDI, на модуле находятся перемычки, которые отвечают за его питание и напряжения его выходов. В данном случае модуль питался от шины USB, напряжение микросхемы составляло 3,3 В. На модуле 1 имеется разъем типа B для подключения кабеля USB экспериментальные данные формируются в блоке ПЛИС.

Заключение

Использование рассмотренных модулей позволяет обеспечить скорость обмена до 1 Мбайт в секунду, при этом время до получения готового образца средней сложности может занимать 1–2 дня. Если скорость, обеспечиваемая

модулями недостаточна, то можно использовать более скоростные микросхемы, например CY7C68013 фирмы Cypress или ISP1581 фирмы Philips, однако для их применения необходимо создание более сложных схем и соответственно требуется больше времени на разработку.

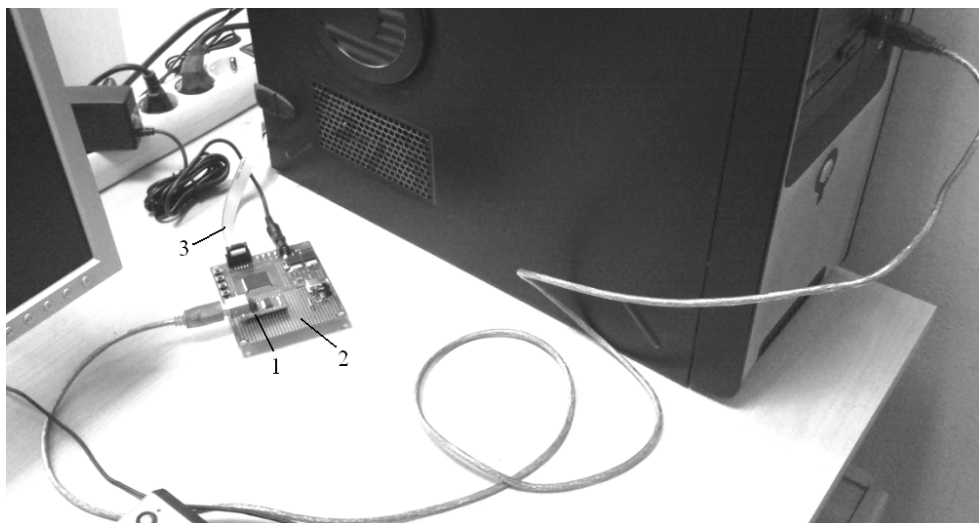


Рис. 4. Отладочная плата с модулем USB

Таким образом, применение модулей USB позволяет достаточно быстро и относительно просто получить работающий макет устройства на ПЛИС, имеющего связь с компьютером по каналу USB со средними характеристиками по скорости обмена, которые достаточны для многих практических приложений.

Список литературы

1. **Агуров, П. В.** Интерфейсы USB. Практика использования и программирования / П. В. Агуров. – СПб. : БХВ-Петербург, 2004. – 567 с.
2. **Шестопалов, А.** Адаптер USB – RS-232 / А. Шестопалов // Схемотехника. – 2006. – № 3. – С. 45–46.
3. **Лысенко, А.** Преобразователи интерфейса USB на микросхемах FT8U232AM, FT8U245AM / А. Лысенко, Р. Назмутдинов, И. Малыгин // Радио. – 2002. – № 6. – С. 36–38.
4. FT245BL USB FIFO (USB-Parallel) // Future Technology Devices Intl. Ltd, DS245BL Version 1.7. – 2005. – URL: www.ftdichip.com
5. D2XX Programmer's Guide / Future Technology Devices International Ltd. – 2006. – 110 P.

Архангельский Алексей Александрович
аспирант, Пензенский
государственный университет

Arkhangelsky Aleksey Alexandrovich
Postgraduate student,
Penza State University

E-mail: gurin2@yandex.ru

Гурин Евгений Иванович

доктор технических наук, профессор,
кафедра вычислительной техники,
Пензенский государственный
университет

E-mail: gurin2@yandex.ru

Gurin Evgeny Ivanovich

Doctor of engineering sciences, professor,
sub-department of computer engineering,
Penza State University

Савельев Борис Александрович

доктор технических наук,
профессор, кафедра информационно-
вычислительных систем, Пензенский
государственный университет

E-mail: gurin2@yandex.ru

Savelyev Boris Alexandrovich

Doctor of engineering sciences, professor,
sub-department of data-computing
systems, Penza State University

УДК 681.327

Архангельский, А. А.

Подключение специализированных процессоров к персональному компьютеру с применением интерфейса USB / А. А. Архангельский, Е. И. Гурин, Б. А. Савельев // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 35–42.

ИТЕРАЦИОННЫЙ МЕТОД ОПТИМАЛЬНОЙ НЕЛИНЕЙНОЙ ФИЛЬТРАЦИИ ИЗОБРАЖЕНИЙ

Аннотация. Рассматривается задача построения оптимальных нелинейных фильтров в классе полиномиальных фильтров, характеризуемых многомерными функциональными рядами Вольтерры. На основе использования матричного представления полиномиальных фильтров формулируется задача оптимальной фильтрации, которая сводится к минимизации квадратичной функции на линейном подпространстве при наличии ограничений. Предлагается итерационный метод последовательного приближения, позволяющий осуществить декомпозицию исходной задачи большой размерности на ряд задач меньшей размерности, связанных с нахождением коэффициентов отдельных составляющих полиномиального фильтра. Показана сходимость итерационного процесса к искомому решению задачи нахождения оптимальных коэффициентов полиномиального фильтра.

Ключевые слова: нелинейная фильтрация изображений, ряды Вольтерры, полиномиальные фильтры.

Abstract. The article considers construction of optimal nonlinear filter in the category of polynomial filters, based on Volterra multidimensional functional rows. On the basis of applying matrix representation of polynomial filters the author formulates a problem of the optimal filtration, which is reduced to minimization of quadratic function on the linear subspace with limitations given. The researcher offers an iteration method of progressive approximation, allowing to decompose the initial problem of high dimensionality to a set of problems of smaller dimensionality, dealing with finding separate component coefficients of a polynomial filter. The article also shows the convergence of the iteration process to the sought solution of the problem of finding optimal coefficients of a polynomial filter.

Key words: image nonlinear filtration, Volterra rows, polynomial filters.

Введение

Методы линейной фильтрации стали уже классическими и с успехом используются в различных системах обработки изображений. С целью расширения спектра задач, решаемых методами линейной фильтрации, и преодоления присущих им ограничений были предложены различные классы нелинейных фильтров [1, 2], каждый из которых имеет свои преимущества и область применения. Некоторые направления такие, например, как медианная и гомоморфная фильтрация, имеют достаточно долгую историю. Другие направления, в частности полиномиальная фильтрация [3, 4], появились сравнительно недавно и активно развиваются в настоящее время.

В общем случае цифровой полиномиальный фильтр размерности r и порядка M определяется конечным дискретным рядом Вольтерры (функциональным полиномом) вида

$$y_m(\mathbf{n}) = H_m[x(\mathbf{n})] = \sum_{\mathbf{n}_1, \dots, \mathbf{n}_m \in \mathcal{R}_r} \dots \sum h_m(\mathbf{n}_1, \dots, \mathbf{n}_m) \prod_{j=1}^m x(\mathbf{n} - \mathbf{n}_j), \quad (1)$$

здесь суммирование по векторным аргументам $\mathbf{n}_j = [n_{j1} \dots n_{jr}]$ распространяется на некоторую опорную область, представляющую собой r -мерную решетку вида

$$\mathfrak{R}_r = \{(n_{j1}, \dots, n_{jr}) : 0 \leq n_{ji} \leq N_i - 1; i = 1, \dots, r\};$$

$h_m(\mathbf{n}_1, \dots, \mathbf{n}_m)$ – многомерные импульсные характеристики (ядра) фильтра, зависящие от векторных аргументов $\mathbf{n}_j$.

Фильтры вида (1) часто называются также фильтрами (процессорами) Вольтерра. Выходной сигнал $y(\mathbf{n})$ таких фильтров представляет собой сумму составляющих, характеризующих нелинейности различного порядка, причем составляющая $y_m(\mathbf{n})$ фильтра, определяемая сверткой m -го порядка, является нелинейной относительно отсчетов входного сигнала, оставаясь линейной по отношению к коэффициентам фильтра. Полиномиальные фильтры имеют ряд полезных аналогий с многомерными линейными фильтрами и являются их естественным обобщением [4]. В задачах обработки изображений используются двухмерные фильтры Вольтерра ($r = 2$).

1. Матричное представление многомерных полиномиальных фильтров

Наряду с аналитическим представлением (1) полиномиальные фильтры могут быть представлены в эквивалентной матричной форме, которая может быть получена путем лексикографического упорядочения элементов опорной области $\mathfrak{R}_r$ фильтра. Для этого воспользуемся системой счисления со смешанным основанием, в которой число i записывается в виде

$$i = \sum_{i=1}^r n_i S_i,$$

здесь $S_i = N_{i+1}, N_{i+2}, \dots, N_r$ ($i < r$), $S_r = 1$. Данное выражение позволяет нам установить взаимно однозначное соответствие между элементами вектора $\mathbf{n} = [n_1 \dots n_r]$, определенного на r -мерной решетке $\mathfrak{R}_r$, и значениями числа i , лежащими в диапазоне от нуля до $N - 1$, где $N = N_1, \dots, N_r$.

Отображая каждый векторный аргумент $\mathbf{n}_j, j = 1, \dots, m$ в индекс i_j , представим ядро $h_m(\mathbf{n}_1, \dots, \mathbf{n}_m)$ как функцию скалярных аргументов $h_m(i_1, \dots, i_m)$, а входной сигнал $x(\mathbf{n} - \mathbf{n}_j)$ – в виде $x_{\mathbf{n}}(i_j)$. В результате двухмерный нелинейный фильтр (1) с импульсной характеристикой от векторных аргументов преобразуется в эквивалентный одномерный нелинейный фильтр с импульсной характеристикой $h_m(i_1, \dots, i_m)$ от скалярных аргументов:

$$H_m[x(\mathbf{n})] = \sum_{i_1=0}^{N-1} \dots \sum_{i_m=0}^{N-1} h_m(i_1, \dots, i_m) \prod_{j=1}^m x_{\mathbf{n}}(i_j).$$

Теперь областью задания импульсной характеристики $h_m(i_1, \dots, i_m)$ является m -мерная решетка вида

$$\mathfrak{R}_m = \{(i_1, \dots, i_m) : 0 \leq i_j \leq N - 1; j = 1, \dots, m\}.$$

Элементы данной многомерной области также могут быть упорядочены в одномерную последовательность с помощью отображения

$$i = \sum_{j=1}^m i_j N^{m-j}.$$

Таким образом, путем последовательного применения данных преобразований элементы ядра $h_m(\mathbf{n}_1, \dots, \mathbf{n}_m)$ могут быть упорядочены в вектор $\mathbf{h}_m$. При этом зависимость номера элемента вектора $\mathbf{h}_m$ от аргументов многомерного ядра $h_m(\mathbf{n}_1, \dots, \mathbf{n}_m)$ имеет вид

$$i = \sum_{j=1}^m N^{m-j} \sum_{i=1}^r n_{ji} S_i,$$

Например, значение $h_2(0, 1, 2, 1)$ ядра второго порядка, определенного в области размером 3×2 , сначала будет отображаться по пространственным координатам в $h_2(1, 5)$, а затем по оставшимся двум координатам (определяющим нелинейное взаимодействие) в значение $h_2(11)$.

Для перехода к матричной форме записи выражения (1) введем в рассмотрение вектор входного сигнала

$$\mathbf{x}_n^T = [x_n(0) \quad x_n(1) \quad \dots \quad x_n(N-1)],$$

содержащий упорядоченные согласно ранее введенному отображению отсчеты r -мерного входного сигнала $x(\mathbf{n})$ в опорной области $\mathcal{R}_r$. Тогда, используя свойства кронекеровской степени матриц, определяемой как $\mathbf{x}_n^{(k)} = \mathbf{x}_n \otimes \mathbf{x}_n^{(k-1)}$, векторное представление составляющей $y_m(\mathbf{n})$ m -го порядка может быть записано в виде

$$y_m(\mathbf{n}) = \mathbf{h}_m^T \mathbf{x}_n^{(m)}, \quad (2)$$

где $\mathbf{h}_m$ – вектор коэффициентов фильтра, содержащий лексикографически упорядоченные значения нелинейной импульсной характеристики $h_m(\mathbf{n}_1, \dots, \mathbf{n}_m)$.

Без ограничения общности можно считать ядра $h_m(\mathbf{n}_1, \dots, \mathbf{n}_m)$, $m > 1$, соответствующие коммутативным произведениям входных отсчетов, симметричными функциями своих аргументов $\mathbf{n}_1, \dots, \mathbf{n}_m$. Следовательно, мы можем оставить в (1) лишь члены, соответствующие различным сочетаниям $(\mathbf{n}_1, \dots, \mathbf{n}_m)$. При этом общее количество уникальных членов для полиномиального нелинейного фильтра порядка M составит

$$L_M = \sum_{m=0}^M C_{N+m-1}^m = C_{N+M}^M, \quad (3)$$

где C_n^m – число сочетаний из n по m .

При синтезе цифровых фильтров для обработки изображений очень часто требуется обеспечить условие изотропности оператора F фильтрации, состоящее в том, чтобы его реакция была инвариантна к ориентации входного изображения. Математически это требование можно записать в виде

$$F[T[\mathbf{P}]] = T[F[\mathbf{P}]],$$

где $T[\mathbf{P}]$ – оператор изменения ориентации изображения $\mathbf{P}$.

В задачах фильтрации изображений это условие, как правило, сводится к инвариантности результата относительно вращения изображения на углы, кратные 90° , и его зеркального отражения относительно вертикальной оси.

Наложение условия изотропности фильтра позволяет дополнительно сократить количество уникальных коэффициентов. В частности, можно показать [5], что для изотропного квадратичного фильтра с маской 3×3 количество коэффициентов $L_2 = 15$ вместо $C_{11}^2 = 55$, необходимых в общем случае.

С учетом сказанного многомерный полиномиальный фильтр вида (1) будет определяться $L_M \times 1$ вектором коэффициентов

$$\mathbf{h}^T = [h_0 \mid \mathbf{h}_1^T \mid \mathbf{h}_2^T \mid \dots \mid \mathbf{h}_M^T], \quad (4)$$

составленным из векторов $\mathbf{h}_m^T$, соответствующих различным составляющим, содержащим лишь уникальные коэффициенты фильтра.

Аналогично формируется вектор произведений отсчетов входного сигнала:

$$\chi_{\mathbf{n}}^T = \left[1 \mid \mathbf{x}_{\mathbf{n}}^T \mid \left(\mathbf{x}_{\mathbf{n}}^{(2)} \right)^T \mid \dots \mid \left(\mathbf{x}_{\mathbf{n}}^{(M)} \right)^T \right]. \quad (5)$$

Таким образом, многомерный полиномиальный фильтр (1) может быть представлен в следующей простой векторной форме:

$$y(\mathbf{n}) = \mathbf{h}^T \chi_{\mathbf{n}}, \quad (6)$$

линейной относительно вектора $\mathbf{h}$, содержащего L_M коэффициентов фильтра.

2. Решение задачи оптимальной фильтрации

Используя матричное представление (6), задачу синтеза оптимального полиномиального фильтра в общем случае можно сформулировать как задачу нахождения $n \times 1$ вектора коэффициентов $\mathbf{h}_{\text{opt}}$, минимизирующего некоторый функционал качества:

$$F(\mathbf{h}_{\text{opt}}) = \min_{\mathbf{h} \in G} F(\mathbf{h}); \quad (7)$$

$$\mathbf{h}_{\text{opt}} \in G = \{\mathbf{h} \in R^n : g(\mathbf{h}) = 0\}, \quad (8)$$

где $F(\mathbf{h})$ – целевая функция; $g(\mathbf{h})$ – векторная функция ограничений.

Свойства проектируемого фильтра определяются выбором целевой функции (7). Наиболее распространенным является использование средне-квадратической ошибки, характеризующей разность выходных сигналов идеального $d(\mathbf{n})$ и синтезируемого $y(\mathbf{n})$ фильтров в некоторой области Ψ_r изменения $\mathbf{n}$. Такой выбор может быть оправдан тем, что выход полиномиального фильтра (2) линеен по отношению к его коэффициентам, что позволяет использовать эффективные алгоритмы минимизации квадратичных функций. В этом случае целевая функция (7) определяется суммой квадратов:

$$F(\mathbf{h}) = \frac{1}{|\Psi_r|} \sum_{\mathbf{n} \in \Psi_r} \left(d(\mathbf{n}) - \mathbf{h}^T \boldsymbol{\chi}_{\mathbf{n}} \right)^2, \quad (9)$$

где Ψ_r – область определения выходной реализации r -мерного сигнала,

$$\Psi_r = \{\mathbf{n} = (n_1, \dots, n_r) : 0 \leq n_i \leq K-1; \ i = 1, \dots, r\};$$

$|\Psi_r|$ – количество элементов в ней.

Согласно (3) размерность вектора $\mathbf{h}$ для полиномиального фильтра порядка M составит $n = L_M = C_{N+M}^M$.

Функция ограничений (8) определяется спецификой проектируемого фильтра. Например, для квадратичной фильтрации требование сохранения постоянного уровня яркости в однородной зоне изображения приводит к следующим линейным ограничениям:

$$h_0 = 0, \quad \sum_{i_1} h_1(i_1) = 1, \quad \sum_{i_1} \sum_{i_2} h_2(i_1, i_2) = 0,$$

которые могут быть представлены в матричной форме:

$$\mathbf{A}\mathbf{h} = \mathbf{b}. \quad (10)$$

Целевая функция (7) может быть записана также в матричной форме:

$$F(\mathbf{h}) = \mathbf{h}^T \mathbf{R}_{\chi} \mathbf{h} - 2\mathbf{h}^T \mathbf{r}_{d\chi} + r_d, \quad (11)$$

где $\mathbf{R}_{\chi}$ – автокорреляционная матрица произведений входных отсчетов,

$$\mathbf{R}_{\chi} = \frac{1}{|\Psi_r|} \sum_{\mathbf{n} \in \Psi_r} \boldsymbol{\chi}_{\mathbf{n}} \boldsymbol{\chi}_{\mathbf{n}}^T;$$

$\mathbf{r}_{d\chi}$ – вектор взаимных корреляций между заданным выходным сигналом и произведениями отсчетов входного сигнала,

$$\mathbf{r}_{d\chi} = \frac{1}{|\Psi_r|} \sum_{\mathbf{n} \in \Psi_r} y(\mathbf{n}) \boldsymbol{\chi}_{\mathbf{n}};$$

r_d – среднеквадратическое значение заданного сигнала,

$$r_d = \frac{1}{|\Psi_r|} \sum_{\mathbf{n} \in \Psi_r} d^2(\mathbf{n}).$$

Таким образом, задача синтеза оптимального полиномиального фильтра сводится к минимизации квадратичной функции (11) на линейном подпространстве, определяемом условием (10). Решение этой задачи хорошо известно [6] и достигается из любой начальной точки $\mathbf{h}^{[0]}$ за один шаг в соответствии с выражением

$$\mathbf{h}_{\text{opt}} = \mathbf{h}^{[0]} + \mathbf{Z}(\mathbf{Z}^T \mathbf{R}_{\chi} \mathbf{Z})^{-1} \mathbf{Z}^T (\mathbf{r}_{d\chi} - \mathbf{R}_{\chi} \mathbf{h}^{[0]}), \quad (12)$$

где $\mathbf{Z}$ – матрица со столбцами, являющимися базисом нуль-пространства, образованного строками матрицы $\mathbf{A}$ ограничений, т.е. удовлетворяющая условию $\mathbf{AZ} = 0$.

В качестве примера рассмотрим задачу синтеза квадратичного фильтра для обнаружения границ деталей изображения, устойчивого к воздействию шумов. Для этого воспользуемся синтезированным изображением размером 64×64 , составленным из светлых и темных треугольников с уровнями яркости, равными соответственно 170 и 80. Данное изображение было искажено гауссовым шумом с дисперсией $\sigma^2 = 400$ и использовалось в качестве входного сигнала (рис. 1,а). Эталонное изображение, состоящее из выделенных перепадов, показано на рис. 1,б. Для его получения к исходному (неискаженному) изображению был применен известный оператор Собела [7] с последующим сравнением с порогом.

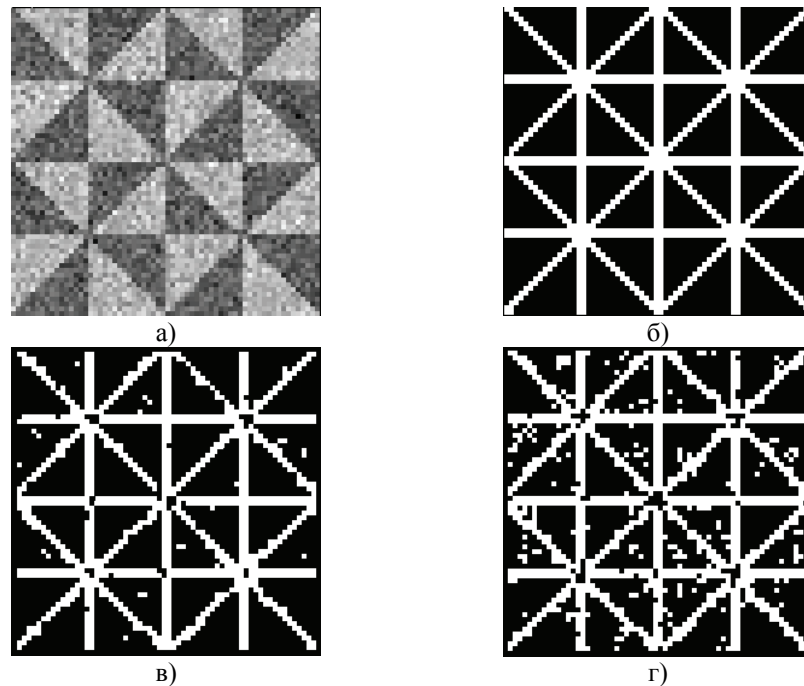


Рис. 1. Результаты обнаружения перепадов в зашумленном изображении:

- а – искаженное изображение, используемое в качестве входного;
- б – эталонное изображение, содержащее идеальные перепады;
- в – результат обработки с помощью синтезированного оптимального фильтра;
- г – результат обработки с помощью оператора Собела

При проектировании фильтров для обнаружения перепадов условия сохранения уровня яркости не накладываются. Оптимальный вектор $\mathbf{h}$ коэффициентов определялся при $\mathbf{Z} = \mathbf{I}$ и представлен в табл. 1. На рис. 1,в,г приведены результаты обработки зашумленного изображения с помощью рассчитанного оптимального фильтра и оператора Собела, являющегося по данным, приведенным в [7], одним из наиболее эффективных среди известных детекторов перепада. Из сравнения полученных результатов видно, что синтезированный фильтр обладает большей устойчивостью к шумам по сравнению

с оператором Собела, что проявляется в существенно меньшем количестве ложных обнаружений.

Таблица 1

m	i	$\mathbf{h}$	Значение
1	1	$h_1(0)$	-0,36673
	2	$h_1(1)$	0,60851
	3	$h_1(4)$	-1,00000
2	4	$h_2(0, 0)$	0,01002
	5	$h_2(0, 1)$	0,00309
	6	$h_2(0, 2)$	-0,00029
	7	$h_2(0, 4)$	-0,00198
	8	$h_2(0, 5)$	-0,00607
	9	$h_2(0, 8)$	-0,00022
	10	$h_2(1, 1)$	0,01600
	11	$h_2(1, 3)$	-0,00467
	12	$h_2(1, 4)$	-0,00054
	13	$h_2(1, 7)$	-0,00267
	14	$h_2(4, 4)$	0,01357

3. Итерационный метод решения задачи оптимальной фильтрации

При малом числе коэффициентов фильтра решение задачи оптимальной нелинейной фильтрации в классе полиномиальных фильтров не составляет труда и достигается из любой начальной точки за один шаг с помощью выражения (12). Если количество весовых коэффициентов фильтра велико, то обращение корреляционной матрицы $\mathbf{R}_\chi$ будет представлять собой достаточно трудоемкую операцию. Кроме того, в случае плохой обусловленности данной матрицы решение может оказаться неустойчивым и содержать большие погрешности, связанные с неточностью исходных данных.

Для преодоления указанных трудностей рассмотрим задачу построения оптимального полиномиального фильтра M -го порядка более детально с учетом его структуры, определяемой матричным представлением вида (6), где векторы $\mathbf{h}$ и χ_n размером $L_M \times 1$ согласно выражениям (4) и (5) составлены из блоков, определяющих отдельные составляющие фильтра.

Нахождение вектора $\mathbf{h}_{\text{opt}}$ оптимальных коэффициентов полиномиального фильтра сводится к минимизации среднеквадратической ошибки

$$\varepsilon = M\{d(\mathbf{n}) - y(\mathbf{n})\}^2, \quad (13)$$

где $d(\mathbf{n})$ – заданный (опорный) сигнал на выходе фильтра.

Дифференцируя (13) по $\mathbf{h}$ с учетом (6) и приравнявая результат к нулю, получаем матричное уравнение Винера – Хопфа, определяющее оптимальный фильтр:

$$\mathbf{R}_\chi \mathbf{h} = \mathbf{r}_{d\chi}. \quad (14)$$

где $\mathbf{R}_\chi = M\{\chi_n \chi_n^T\}$ – автокорреляционная матрица; $\mathbf{r}_{d\chi} = M\{d(\mathbf{n}) \chi_n\}$ – вектор взаимных корреляций.

Заметим, что в силу линейности полиномиального фильтра относительно весовых коэффициентов решение (14) выглядит так же, как и в линейном случае. Отличие состоит в том, что матрицы $\mathbf{R}_\chi$ и $\mathbf{r}_{d\chi}$ определяются через корреляции высших порядков. С учетом структуры (5) входного вектора χ_n данные матрицы могут быть представлены в блочном виде следующим образом:

$$\mathbf{R}_\chi = \begin{bmatrix} \mathbf{R}_{0,0} & \mathbf{R}_{0,1} & \dots & \mathbf{R}_{0,M} \\ \mathbf{R}_{1,0} & \mathbf{R}_{1,1} & \dots & \mathbf{R}_{1,M} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{M,0} & \mathbf{R}_{M,1} & \dots & \mathbf{R}_{M,M} \end{bmatrix}, \quad \mathbf{r}_{d\chi} = \begin{bmatrix} \mathbf{r}_0 \\ \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_M \end{bmatrix}. \quad (15)$$

Здесь блоки $\mathbf{R}_{i,j} = M\{\mathbf{x}_n^{(i)} \mathbf{x}_n^{(j)T}\}$ и $\mathbf{r}_i = M\{d(n) \mathbf{x}_n^{(i)}\}$ являются соответственно матрицей автокорреляций порядка $i+j$ и вектором взаимных корреляций порядка i . В линейном случае матрицы (15) состоят лишь из блоков $\mathbf{R}_{0,0}$, $\mathbf{R}_{0,1}$, $\mathbf{R}_{1,0}$, $\mathbf{R}_{1,1}$ и $\mathbf{r}_0$, $\mathbf{r}_1$.

На практике при вычислении среднеквадратической ошибки (13) длительность используемых реализаций всегда конечна. Определим векторы $\mathbf{d}$ и $\mathbf{y}$ размером $I \times 1$, содержащие соответственно I отсчетов заданного и выходного сигналов фильтра. Как отмечалось ранее, для многомерных сигналов это всегда можно сделать путем лексикографического упорядочения многомерных массивов в одномерные. На основании (6) вектор $\mathbf{y}$ реакции фильтра может быть представлен в матричной форме:

$$\mathbf{y} = \mathbf{X}\mathbf{h}, \quad (16)$$

где $\mathbf{h}$ – вектор коэффициентов фильтра, равный

$$\mathbf{h}^T = [h_0 \mid \mathbf{h}_1^T \mid \mathbf{h}_2^T \mid \dots \mid \mathbf{h}_M^T]; \quad (17)$$

$\mathbf{X}$ – матрица входного сигнала фильтра, строками которой являются χ_i^T , $i = 1, \dots, I$, определяемая в развернутом виде согласно (5) следующим выражением:

$$\mathbf{X} = \begin{bmatrix} \chi_1^T \\ \vdots \\ \chi_I^T \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{x}_1^T & (\mathbf{x}_1^{(2)})^T & \dots & (\mathbf{x}_1^{(M)})^T \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \mathbf{x}_I^T & (\mathbf{x}_I^{(2)})^T & \dots & (\mathbf{x}_I^{(M)})^T \end{bmatrix}. \quad (18)$$

В качестве критерия при конечном объеме выборки вместо (13) обычно используется квадрат евклидовой нормы вектора ошибок:

$$\|\boldsymbol{\varepsilon}\|^2 = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\mathbf{d} - \mathbf{y})^T (\mathbf{d} - \mathbf{y}). \quad (19)$$

Как известно [6], минимум (19) определяется решением системы нормальных уравнений вида

$$\mathbf{X}^T \mathbf{X} \mathbf{h} = \mathbf{X}^T \mathbf{d} \quad (20)$$

и достигается в точке

$$\mathbf{h}_{\text{opt}} = \mathbf{X}^+ \mathbf{d}, \quad (21)$$

где $\mathbf{X}^+$ является обобщенной обратной (псевдообратной) матрицей для $\mathbf{X}$.

Если столбцы матрицы $\mathbf{X}$ линейно независимы, что бывает далеко не всегда, то псевдообратная матрица определяется выражением

$$\mathbf{X}^+ = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T.$$

В частности, для квадратной матрицы $\mathbf{X}^+ = \mathbf{X}^{-1}$.

При неполном ранге для определения $\mathbf{X}^+$ можно воспользоваться сингулярным представлением матрицы $\mathbf{X} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}$. Здесь $\mathbf{U}$ и $\mathbf{V}$ являются унитарными матрицами со столбцами, равными собственным векторам соответственно матриц $\mathbf{X} \mathbf{X}^T$ и $\mathbf{X}^T \mathbf{X}$, а $\mathbf{\Lambda}$ – диагональной матрицей вида

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

где $\mathbf{D} = \text{diag}(\sigma_1, \dots, \sigma_k)$, σ_i – отличные от нуля сингулярные числа $\mathbf{X}$.

Псевдообратная матрица $\mathbf{X}^+$, соответствующая матрице $\mathbf{X}$ ранга k , равна

$$\mathbf{X}^+ = \mathbf{V} \mathbf{D}^{-1} \mathbf{U}. \quad (22)$$

В соответствии с выражением (16) и структурой матрицы (18) выходной вектор $\mathbf{y}$ формируется путем взвешенного суммирования компонент y_i , $i = 0, \dots, M$, обусловленных отдельными столбцами матрицы $\mathbf{X}$. Первый столбец определяет постоянную составляющую y_0 , столбец из блоков $\mathbf{x}_i$ – линейную составляющую y_1 , $\mathbf{x}_i^{(2)} = \mathbf{x}_i \otimes \mathbf{x}_i$ – нелинейную составляющую y_2 второго порядка и т.д. до M -го порядка включительно. Непосредственное использование (21) предусматривает одновременное вычисление всех компонент фильтра заданного порядка M , что связано с решением задачи большой размерности. При этом могут затрачиваться неоправданные усилия на вычисление нелинейных составляющих, влиянием которых из-за их незначительности можно пренебречь. Более оправданным является последовательное формирование компонент фильтра, начиная с линейного члена, постепенно увеличивая степень нелинейности.

Представим выходной вектор $\mathbf{y}$ полиномиального фильтра в виде суммы компонент

$$\mathbf{y} = \sum_{m=0}^M \mathbf{X}_m \mathbf{h}_m, \quad (23)$$

где вектор $\mathbf{X}_0 = \mathbf{1}$ целиком состоит из единиц; $\mathbf{X}_m^T = \begin{bmatrix} \mathbf{x}_1^{(m)} & \dots & \mathbf{x}_I^{(m)} \end{bmatrix}$ – входная матрица для составляющей фильтра m -го порядка.

Формирование коэффициентов фильтра осуществляется путем последовательной оценки векторов $\mathbf{h}_m$, начиная с составляющих низших порядков. Полагая $m = 0$ и минимизируя $\|\mathbf{d} - \mathbf{X}_0 \mathbf{h}_0\|^2$, получим

$$\mathbf{h}_0 = \mathbf{X}_0^+ \mathbf{d} = \mathbf{1}^+ \mathbf{d} = \frac{1}{I} \sum_{i=1}^I d(i),$$

т.е. оптимальным выбором постоянной составляющей фильтра является среднее значение заданного вектора $\mathbf{d}$. Вычислим далее выходной вектор $\mathbf{y}_0 = \mathbf{X}_0^+ \mathbf{h}_0$ и вектор ошибки $\boldsymbol{\varepsilon}_0 = \mathbf{d} - \mathbf{y}_0$.

Вектор $\mathbf{h}_1$ линейной составляющей фильтра определим из условия минимизации $\|\boldsymbol{\varepsilon}_0 - \mathbf{X}_1 \mathbf{h}_1\|^2$, в результате чего получим

$$\mathbf{h}_1 = \mathbf{X}_1^+ \boldsymbol{\varepsilon}_0.$$

Выходной сигнал фильтра будет теперь содержать две составляющие:

$$\mathbf{y}_1 = \mathbf{y}_0 + \mathbf{X}_1^+ \mathbf{h}_1,$$

а ошибка станет равной $\boldsymbol{\varepsilon}_1 = \mathbf{d} - \mathbf{y}_1$. Минимизируя далее $\|\boldsymbol{\varepsilon}_1 - \mathbf{X}_2 \mathbf{h}_2\|^2$, получим вектор $\mathbf{h}_2$ квадратичной составляющей, равный

$$\mathbf{h}_2 = \mathbf{X}_2^+ \boldsymbol{\varepsilon}_1.$$

Продолжая процесс вычисления аналогичным образом, последовательно получим векторы $\mathbf{h}_m$ коэффициентов нелинейных составляющих любого порядка.

Ошибка $\boldsymbol{\varepsilon}_m$ для фильтра m -го порядка может быть записана в виде

$$\boldsymbol{\varepsilon}_m = \mathbf{d} - \mathbf{y}_m = \mathbf{d} - \mathbf{y}_{m-1} - \mathbf{X}_m \mathbf{X}_m^+ \boldsymbol{\varepsilon}_{m-1} = (\mathbf{I} - \mathbf{X}_m \mathbf{X}_m^+) \boldsymbol{\varepsilon}_{m-1}. \quad (24)$$

На основании определения (22) псевдообратной матрицы нетрудно доказать свойство $\mathbf{X}^T \mathbf{X} \mathbf{X}^+ = \mathbf{X}^T$, используя которое, можно показать, что вектор ошибки $\boldsymbol{\varepsilon}_m$ на каждом этапе вычислений будет ортогонален входной матрице $\mathbf{X}_m$, т.е.

$$\mathbf{X}_m^T \boldsymbol{\varepsilon}_m = 0, \quad m = 0, 1, \dots \quad (25)$$

Последовательное увеличение порядка m нелинейности фильтра происходит до тех пор, пока норма $\|\boldsymbol{\varepsilon}_m\|$ не уменьшится до заданного предела. Условие, при котором решение, полученное на m -м шаге, является точным и не может быть далее улучшено, следует из (24) и выглядит следующим образом:

$$\mathbf{X}_m \mathbf{X}_m^+ = \mathbf{I}. \quad (26)$$

С учетом тождества $\mathbf{X}_m^+ = \mathbf{X}_m^T (\mathbf{X}_m \mathbf{X}_m^T)^+$, в справедливости которого нетрудно убедиться, преобразуем (26) к виду

$$\mathbf{X}_m \mathbf{X}_m^T (\mathbf{X}_m \mathbf{X}_m^T)^+ = \mathbf{I}.$$

Данное условие выполняется, если строки матрицы $\mathbf{X}_m \mathbf{X}_m^T$ линейно независимы и, следовательно, существует матрица, обратная $\mathbf{X}_m \mathbf{X}_m^T$. В частности, если существует $(\mathbf{X}_1 \mathbf{X}_1^T)^{-1}$, то оптимальное решение достигается в классе линейных фильтров.

Используя свойства кронекеровского произведения матриц, можно показать, что (i, j) -й элемент матрицы $\mathbf{X}_m \mathbf{X}_m^T$ равен

$$\left(\mathbf{x}_i^{(m)} \right)^T \mathbf{x}_j^{(m)} = \left(\mathbf{x}_i^T \mathbf{x}_j \right)^m. \quad (27)$$

Обозначим через $\mathbf{X}^{<m>}$ операцию возведения в степень m всех элементов матрицы $\mathbf{X}$. Тогда на основании (27) можно записать равенство

$$\mathbf{X}_m \mathbf{X}_m^T = \left(\mathbf{X}_1 \mathbf{X}_1^T \right)^{<m>}.$$

Таким образом, решение, получаемое полиномиальным фильтром m -го порядка, будет являться точным, если существует матрица, обратная $(\mathbf{X}_1 \mathbf{X}_1^T)^{<m>}$.

В процессе последовательного приближения может возникнуть ситуация, когда при добавлении очередной i -й составляющей фильтра вновь вычисленная ошибка $\|\mathbf{e}_m\|$ практически не будет отличаться от предыдущей $\|\mathbf{e}_{m-1}\|$. В этом случае включение данной составляющей не имеет смысла и осуществляется переход к следующей.

Рассмотрим условие, при котором величина ошибки, связанной с добавлением члена m -го порядка к уже построенному фильтру k -го порядка, не изменяется. Связь выходных векторов $\mathbf{y}_m$ и $\mathbf{y}_k$ данных фильтров определяется выражением

$$\mathbf{y}_m = \mathbf{y}_k + \mathbf{X}_m \mathbf{X}_m^+ \mathbf{e}_k. \quad (28)$$

Если фильтр k -го порядка в свою очередь был сформирован вслед за фильтром порядка s , то ошибка будет равна $\mathbf{e}_k = (\mathbf{I} - \mathbf{X}_k \mathbf{X}_k^+) \mathbf{e}_s$ и величина коррекции $\mathbf{X}_m \mathbf{X}_m^+ \mathbf{e}_k$ в (28) может быть записана в виде

$$\mathbf{X}_m \mathbf{X}_m^T (\mathbf{I} - \mathbf{X}_k \mathbf{X}_k^T) \mathbf{e}_s.$$

Данная величина будет равна нулю при произвольном $\mathbf{e}_s$, если выполняется следующее условие:

$$\mathbf{X}_m \mathbf{X}_m^T \mathbf{X}_k \mathbf{X}_k^T = \mathbf{X}_k \mathbf{X}_k^T. \quad (29)$$

Из свойства $\mathbf{X} = \mathbf{X} \mathbf{X}^+ \mathbf{X}$, в частности, следует, что равенство (29) справедливо при условии $\mathbf{X}_k \mathbf{X}_k^+ = \mathbf{X}_m \mathbf{X}_m^+$.

Пусть в результате рассмотренной процедуры получен набор векторов $\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_M$, характеризующих M составляющих фильтра. Объединим данные векторы согласно (17) для получения вектора $\mathbf{h}$ коэффициентов полиномиального фильтра M -го порядка. Очевидно, что в общем случае нельзя гарантировать совпадение данного вектора с оптимальным $\mathbf{h}_{\text{opt}}$, определяемым выражением (21), так как вместо суммарной ошибки $\|\mathbf{e}\|^2$ вида (19) минимизировались частичные ошибки $\|\mathbf{e}_k\|^2$, $k = 0, 1, \dots, M$, по отдельным компонентам фильтра.

Последовательность $\|\mathbf{e}_k\|$ является убывающей. Действительно, справедливо неравенство

$$\|\boldsymbol{\varepsilon}_k\| = \|(\mathbf{I} - \mathbf{X}_k \mathbf{X}_k^+) \boldsymbol{\varepsilon}_{k-1}\| \leq \|\mathbf{I} - \mathbf{X}_k \mathbf{X}_k^+\|_2 \|\boldsymbol{\varepsilon}_{k-1}\|,$$

где $\|\mathbf{A}\|_2$ – спектральная норма матрицы $\mathbf{A}$, равная максимальному сингулярному числу $\mathbf{A}$. На основании определения (22) псевдообратной матрицы можно показать, что $\|\mathbf{I} - \mathbf{X}_k \mathbf{X}_k^+\| = \|\mathbf{I} - \mathbf{U} \mathbf{\Lambda} \mathbf{D}^{-1} \mathbf{U}^T\| = 1$, поэтому $\|\boldsymbol{\varepsilon}_0\| \geq \|\boldsymbol{\varepsilon}_1\| \geq \dots \geq \|\boldsymbol{\varepsilon}_M\| \geq \|\boldsymbol{\varepsilon}\|$. Конечная ошибка $\|\boldsymbol{\varepsilon}_M\|$, в свою очередь, может быть уменьшена, если повторить процедуру коррекции вновь, начиная с минимизации $\|\boldsymbol{\varepsilon}_M - \mathbf{X}_0 \mathbf{h}_0\|^2$. В результате будет сформирована новая последовательность векторов $\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_M$, объединение которых даст более точное приближение к искомому вектору $\mathbf{h}_{\text{opt}}$.

Структура полиномиального фильтра, реализующего такой итерационный процесс последовательного приближения, показана на рис. 2.

Обозначим через $\mathbf{h}_m^{[0]}$ и $\boldsymbol{\varepsilon}_m^{[0]}$, $m = 0, 1, \dots, M$, векторы коэффициентов и ошибок, полученные на начальном этапе приближения. Тогда процесс настройки коэффициентов фильтра может быть представлен в виде следующей итерационной схемы:

$$\begin{aligned} \mathbf{h}_0^{[n]} &= \mathbf{h}_0^{[n-1]} + \mathbf{X}_0^+ \boldsymbol{\varepsilon}_M^{[n-1]}, \quad \mathbf{y}_0^{[n]} = \mathbf{y}_M^{[n]} + \mathbf{X}_0 \mathbf{h}_0^{[n]}, \\ &\dots\dots\dots \\ \mathbf{h}_m^{[n]} &= \mathbf{h}_m^{[n-1]} + \mathbf{X}_m^+ \boldsymbol{\varepsilon}_{m-1}^{[n]}, \quad \mathbf{y}_m^{[n]} = \mathbf{y}_{m-1}^{[n]} + \mathbf{X}_m \mathbf{h}_m^{[n]}, \\ \boldsymbol{\varepsilon}_m^{[n]} &= \mathbf{d} - \mathbf{y}_m^{[n]}, \quad n = 1, 2, \dots \end{aligned} \quad (30)$$

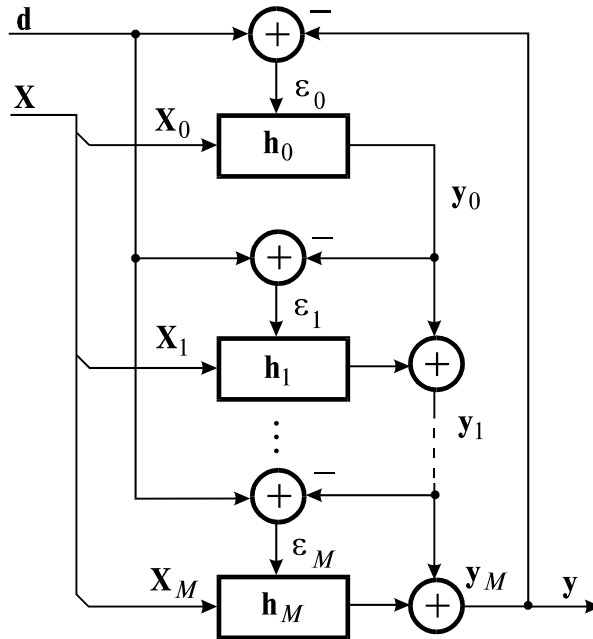


Рис. 2. Структура полиномиального фильтра с последовательной настройкой коэффициентов

4. Доказательство сходимости метода

Покажем, что данный процесс сходится, причем

$$\left\| \boldsymbol{\varepsilon}_M^{[n]} \right\| \xrightarrow{n \rightarrow \infty} \left\| \boldsymbol{\varepsilon} \right\|, \quad \left\| \mathbf{h}^{[n]} \right\| \xrightarrow{n \rightarrow \infty} \mathbf{h}_{\text{opt}}.$$

Так как последовательность $\left\| \boldsymbol{\varepsilon}_M^{[n]} \right\|$ является убывающей и ограничена снизу, она имеет предел $\left\| \tilde{\boldsymbol{\varepsilon}} \right\|$. Пусть соответствующие предельные векторы коэффициентов равны $\tilde{\mathbf{h}}_0, \tilde{\mathbf{h}}_1, \dots, \tilde{\mathbf{h}}_M$. Тогда вектор $\tilde{\boldsymbol{\varepsilon}}$ ошибок можно представить в виде

$$\tilde{\boldsymbol{\varepsilon}} = \mathbf{d} - \mathbf{X}_0 \tilde{\mathbf{h}}_0 - \mathbf{X}_1 \tilde{\mathbf{h}}_1 - \dots - \mathbf{X}_M \tilde{\mathbf{h}}_M.$$

Согласно (25) этот вектор будет ортогонален всем входным матрицам $\mathbf{X}_m$, $m = 0, 1, \dots, M$, т.е. векторы $\tilde{\mathbf{h}}_0, \tilde{\mathbf{h}}_1, \dots, \tilde{\mathbf{h}}_M$ удовлетворяют следующей системе уравнений:

$$\mathbf{X}_m^T \mathbf{X}_0 \tilde{\mathbf{h}}_0 + \mathbf{X}_m^T \mathbf{X}_1 \tilde{\mathbf{h}}_1 + \dots + \mathbf{X}_m^T \mathbf{X}_M \tilde{\mathbf{h}}_M = \mathbf{X}_m^T \mathbf{d}, \quad m = 0, 1, \dots, M.$$

Данная система уравнений полностью тождественна нормальным уравнениям (20), определяющим оптимальный вектор $\mathbf{h}_{\text{opt}}$ коэффициентов полиномиального фильтра. Поэтому $\tilde{\mathbf{h}} = \mathbf{h}_{\text{opt}}$ и $\tilde{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}$.

Заключение

Класс полиномиальных фильтров, характеризуемый дискретным функциональным рядом Вольтерра, является естественным обобщением класса линейных фильтров, сохраняя свойство линейности относительно параметров фильтра. Это обстоятельство и использование матричного представления полиномиальных фильтров позволяет сформулировать задачу оптимальной нелинейной фильтрации в виде решения известного матричного уравнения Винера – Хопфа, отличающегося от линейного аналога лишь тем, что входящие в его состав матрицы содержат корреляции высших порядков. Для решения данного уравнения предложен итерационный метод последовательного приближения, состоящий в постепенном увеличении степени нелинейности фильтра путем добавления новых компонент до тех пор, пока ошибка фильтрации не станет меньше некоторого заданного предела. Получены условия, гарантирующие уменьшение ошибки; доказана сходимость итерационного процесса к искомому решению задачи оптимальной фильтрации. Предложенный метод может быть положен в основу реализации полиномиальных фильтров для обработки сигналов и изображений с последовательной настройкой коэффициентов.

Список литературы

1. **Pitas, I.** Nonlinear digital filters: principles and applications / I. Pitas, A. N. Venetsanopoulos. – Kluwer Academic Publishers, 1990. – 391 p.
2. **Ling, W. K.** Nonlinear digital filters: analysis and applications / W. K. Ling. – Academic Press, 2007. – 216 p.

3. **Mathews, V. J.** Polynomial signal processing / V. J. Mathews, G. L. Sicuranza. – A Wiley-Interscience publication, 2000. – 452 p.
4. **Щербаков, М. А.** Цифровая полиномиальная фильтрация: теория и приложение / М. А. Щербаков. – Пенза : Изд-во Пенз. гос. техн. ун-та, 1997. – 246 с.
5. **Щербаков, М. А.** Свойства симметрии изотропных нелинейных операторов для обработки изображений / М. А. Щербаков // Синтез и сложность управляющих систем : материалы XIII Международной школы-семинара. Ч. II / под. ред. О. Б. Лупанова – М. : МГУ, 2002 – С. 208–215.
6. **Гилл, Ф.** Практическая оптимизация / Ф. Гилл, У. Мюррей, М. Райт. – М. : Мир, 1985. – 510 с.
7. **Прэтт, У.** Цифровая обработка изображений / У. Прэтт. – М. : Мир, 1982. – 312 с.

Щербаков Михаил Александрович
доктор технических наук, профессор,
заведующий кафедрой автоматизации
и телемеханики, Пензенский
государственный университет

E-mail: mush@sura.ru

Shcherbakov Mikhail Alexandrovich
Doctor of engineering sciences, professor,
head of sub-department of automation
and remote control, Penza State University

УДК 681.5.015.52+621.372.542

Щербаков, М. А.

Итерационный метод оптимальной нелинейной фильтрации изображений / М. А. Щербаков // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 43–56.

В. И. Горбаченко, Н. О. Матвеева, Е. И. Гурин

РАСПАРАЛЛЕЛИВАНИЕ НЕЙРОСЕТЕВОГО АЛГОРИТМА РЕШЕНИЯ КРАЕВЫХ ЗАДАЧ МАТЕМАТИЧЕСКОЙ ФИЗИКИ НА МНОГОЯДЕРНЫХ ПРОЦЕССОРАХ NVIDIA

Аннотация. Рассматривается возможность эффективной реализации нейросетевого алгоритма решения краевых задач математической физики на многоядерном графическом процессоре NVIDIA GeForce 8800 GTX. Предлагаются методы распараллеливания основных вычислений алгоритма обучения радиально-базисной нейронной сети (RBFNN), решающей уравнение Пуассона, с использованием технологии CUDA. Проводится сравнение времени обучения RBFNN на центральном и массивно-параллельном графическом процессорах.

Ключевые слова: массивно-параллельная архитектура, графический процессор, параллельность, CUDA, радиально-базисная нейронная сеть, дифференциальное уравнение.

Abstract. The article describes the possibility of effective realization of the algorithm for solving boundary value problems of mathematical physics on multicore Graphics Processing Unit (GPU) NVIDIA GeForce 8800 GTX. The authors propose parallel realization of the training algorithm of the radial basis neural network (RBFNN), using CUDA. The proceeding compares execution time using CPU versus GPU.

Key words: massive parallel processing, Graphics Processing Unit, parallelism, CUDA, radial basis neural network, differential equation.

Введение

Математическое моделирование физических процессов в настоящее время применяется в самых различных областях деятельности человека. Одной из важных проблем, возникающих при решении задач математической физики, является большие вычислительные затраты, а следовательно, длительное время вычислений. Но, с другой стороны, многим методам решения задач математической физики присущ внутренний параллелизм вычислений, что позволяет использовать параллельные системы для их решения.

Вычислительные мощности современных суперкомпьютеров на данный момент доступны далеко не всем, но созданные относительно недавно и активно развивающиеся многопроцессорные системы, первоначально появившиеся как видеокарты новой архитектуры, предоставляют реальную возможность ускорения вычислений путем их распараллеливания. Так как рост тактовой частоты одного процессора уже достиг своего максимума, дальнейшее повышение производительности связано с увеличением площадей кристалла. Графическим процессорам первоначально был свойственен параллелизм, поэтому неудивительно, что видеокарты стали многопроцессорными. Еще до появления графических процессоров (GPU) нового поколения были предприняты попытки использовать вычислительные ресурсы видеокарты для решения задач общего назначения, а после выхода удобной технологии программирования новых GPU nVidia CUDA (Compute Unified Device Architecture) [1] они стали использоваться для решения самых разных задач, требующих больших вычислительных затрат.

Для численного решения краевых задач математической физики, описываемых дифференциальными уравнениями в частных производных, наибольшее распространение получили методы конечных разностей и конечных элементов. Эти методы требуют построения расчетных сеток. Сеточные методы позволяют получить решение только в узлах сетки.

Альтернативным подходом является использование различных вариантов метода взвешенных невязок, когда в качестве базисных функций применяются радиальные базисные функции (RBF – radial basis function) [2–6]. Использование RBF-функций рассматривается как бессеточный метод (meshless, meshfree). Применение радиально-базисных функций позволяет обойтись без трудоемкого построения сетки и позволяет получить приближенное дифференцируемое решение в произвольных точках области. Недостатком метода является сложность определения параметров радиально-базисных функций.

Бессеточные методы эффективно реализуются на радиально-базисных нейронных сетях (RBFNN) [6–9]. RBFNN отличаются простотой, так как содержат только один скрытый слой, поэтому исключается неформализуемый подбор структуры сети, необходимый для большинства типов нейронных сетей. RBFNN могут быть реализованы на параллельных системах. Главное достоинство RBFNN состоит в использовании принципов обучения для формирования оптимальных параметров радиально-базисных функций.

Целью данной работы является разработка параллельной реализации алгоритма обучения радиально-базисной нейронной сети для решения краевых задач математической физики на графическом процессоре с использованием технологии CUDA и исследование эффективности распараллеливания путем сравнения времени решения краевой задачи на GPU и на центральном процессоре.

1. Структура радиально-базисной нейронной сети

Радиально-базисная нейронная сеть (рис. 1) представляет собой сеть с двумя слоями [6–8].

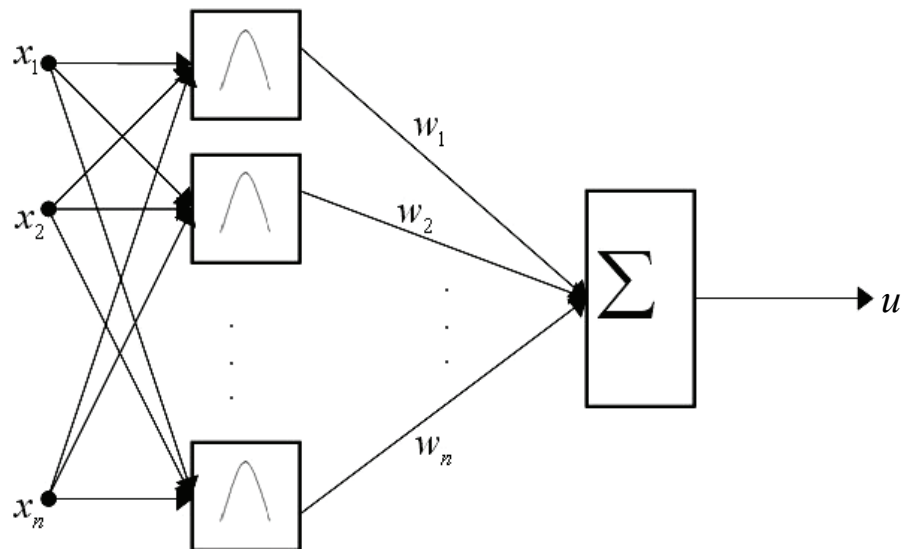


Рис. 1. Радиально-базисная нейронная сеть

Первый слой осуществляет преобразование входного вектора $\mathbf{x}$ с использованием радиально-базисных функций (RBF). Практически используются различные радиально-базисные функции. В дальнейшем будем использовать наиболее часто употребляемую функцию – Гауссиан, имеющий вид для k -го нейрона:

$$\varphi_k(\mathbf{x}) = \exp\left(-r_k^2 / a_k^2\right), \quad (1)$$

где $\mathbf{x}$ – входной вектор; $r_k = \|\mathbf{x} - \mathbf{c}_k\|$ – радиус k -го нейрона; a_k – ширина k -го нейрона; $\mathbf{c}_k$ – центр k -го нейрона.

Выход сети описывается выражением

$$u = \sum_{k=1}^m w_k \varphi_k(\mathbf{x}), \quad (2)$$

где w_k – вес, связывающий выходной нейрон с k -м нейроном первого слоя; m – число нейронов первого слоя.

2. Алгоритм обучения радиально-базисной нейронной сети

Рассмотрим градиентный алгоритм обучения радиально-базисной нейронной сети на примере решения двумерного уравнения Пуассона:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y), \quad (x, y) \in \Omega; \quad (3)$$

$$u = p(x, y), \quad (x, y) \in \partial\Omega, \quad (4)$$

где $\partial\Omega$ – граница области; f и p – известные функции (x, y) .

Выбирая в качестве радиально-базисной функции Гауссиан (1), определяемый как $\varphi_k(x, y) = e^{-\frac{r_k^2}{a_k^2}}$, рассмотрим RBF-сеть как аппроксиматор функции решения:

$$u(x) = \sum_{k=1}^m w_k \varphi_k(r_k), \quad (5)$$

где m – число радиально-базисных функций (скрытых нейронов), $r_k = \sqrt{(x - c_{xk})^2 + (y - c_{yk})^2}$, (c_{xk}, c_{yk}) – координаты центра нейрона k , a_k – ширина нейрона k .

Обучение сети сводится к настройке весов, расположения центров и ширины нейронов, минимизирующих функционал качества (функционал ошибки), представляющий собой сумму квадратов невязок в контрольных точках

$$I(w, c, a) = \frac{1}{2} \sum_{i=1}^N \left[\frac{\partial^2 u(x_i, y_i)}{\partial x^2} + \frac{\partial^2 u(x_i, y_i)}{\partial y^2} - f(x_i, y_i) \right]^2 +$$

$$+\frac{\lambda}{2}\sum_{j=1}^K\left[u\left(x_j,y_j\right)-p_j\right]^2, \quad (6)$$

где λ – штрафной множитель; p_j – значение граничных условий первого рода в точке j границы; N и K – количество внутренних и граничных контрольных точек.

Функционал (6) необходим для обучения сети. Используется пакетный режим обучения, т.е. ошибка вычисляется усредненная по всем контрольным точкам.

Очень важно соблюдать при обучении соотношение между оптимальным количеством нейронов m и количеством контрольных точек [7, с. 389]:

$$m \propto \sqrt[3]{N+K}, \quad (7)$$

где $\propto$ – знак пропорциональности; N – число внутренних контрольных точек; K – число контрольных точек на границе.

Для выполнения соотношения необходимо большое количество контрольных точек, но это ведет к увеличению времени решения задачи. Эксперименты показали, что многократная случайная генерация небольшого числа контрольных точек компенсирует нарушение пропорции.

Градиентный алгоритм спуска, одновременно оптимизирующий веса, центры и ширину может быть построен в виде последовательности двух шагов [9].

Шаг 1. Зафиксировав $w_k^{(n)}$, находим центры и ширину, минимизирующие функционал ошибки (6), используя метод градиентного спуска

$$c_k^{(n)} = c_k^{(n-1)} - \beta^{(n-1)} \frac{\partial I\left(c_k^{(n-1)}, a_k^{(n-1)}, w_k^{(n)}\right)}{\partial c_k^{(n-1)}},$$

$$a_k^{(n)} = a_k^{(n-1)} - \alpha^{(n-1)} \frac{\partial I\left(c_k^{(n)}, a_k^{(n-1)}, w_k^{(n)}\right)}{\partial a_k^{(n-1)}},$$

где

$$\frac{\partial I}{\partial a_k} = 2 \sum_{i=1}^N (\Delta u_i - f_i) w_k e^{-\frac{r_{ik}^2}{a_k^2}} \cdot \left[\frac{r_{ik}^4 - 3a_k^2 r_{ik}^2 + a_k^4}{a_k^7} \right] +$$

$$+ \lambda \cdot 2 \sum_{j=1}^K (u_j - p_j) \cdot w_k e^{-\frac{r_{jk}^2}{a_k^2}} \cdot \frac{r_{jk}^2}{a_k^3};$$

$$\frac{\partial I}{\partial c_{xk}} = 2 \sum_{i=1}^N (\Delta u_i - f_i) w_k e^{-\frac{r_{ik}^2}{a_k^2}} \cdot (x_i - c_{xk}) \cdot \frac{r_{ik}^2 - 2a_k^2}{a_k^6} +$$

$$\begin{aligned}
 & + \lambda \cdot 2 \sum_{j=1}^K (u_j - p_j) \cdot e^{-\frac{r_{jk}^2}{a_k^2}} \cdot \frac{(x_j - c_{xk})}{a_k^2}; \\
 \frac{\partial I}{\partial c_{yk}} &= 2 \sum_{i=1}^N (\Delta u_i - f_i) w_k e^{-\frac{r_{ik}^2}{a_k^2}} \cdot (y_i - c_{yk}) \cdot \frac{r_{ik}^2 - 2a_k^2}{a_k^6} + \\
 & + \lambda \cdot 2 \sum_{j=1}^K (u_j - p_j) \cdot e^{-\frac{r_{jk}^2}{a_k^2}} \cdot \frac{(y_j - c_{yk})}{a_k^2}.
 \end{aligned}$$

Направление изменения параметров сети выбирается противоположным градиенту функционала I .

Шаг 2. Зафиксировав центры и ширину, находим веса, минимизирующие функционал ошибки (6).

Для обучения весов сети использовался алгоритм сопряженных градиентов для минимизации квадратичного функционала, предложенный в [6]. Данный алгоритм эквивалентен при зафиксированных центрах и ширине решению системы линейных алгебраических уравнений $\mathbf{A}\mathbf{w} = \mathbf{s}$, где $\mathbf{A} = \frac{1}{2}(\mathbf{M}^T \mathbf{M} + \lambda \mathbf{N}^T \mathbf{N})$ – симметричная положительно определенная матрица;

$\mathbf{s} = \frac{1}{2}(\mathbf{M}^T \mathbf{f} + \lambda \mathbf{N}^T \mathbf{p})$; $\mathbf{M}$ – матрица $N \times m$ с элементами $m_{ik} = 4e^{-\frac{r_{ik}^2}{a_k^2}} \frac{r_{ik}^2 - a_k^2}{a_k^4}$;

$\mathbf{N}$ – матрица $K \times m$ с элементами $n_{ik} = e^{-r_{ik}^2/a_k^2}$; $\mathbf{f}$ – вектор значений функции правой части во внутренних точках; $\mathbf{p}$ – вектор граничных условий в граничных контрольных точках. Алгоритм метода сопряженных градиентов для решения СЛАУ $\mathbf{A}\mathbf{w} = \mathbf{s}$ представлен на рис. 2.

$\mathbf{r}_0 = \mathbf{s} - \mathbf{A}\mathbf{w}_0$
 $\mathbf{p}_0 = \mathbf{r}_0$
 Для $j = 1, 2, \dots$
 $\alpha_j = (\mathbf{r}_j, \mathbf{r}_j) / (\mathbf{A}\mathbf{p}_j, \mathbf{p}_j)$
 $\mathbf{w}_{j+1} = \mathbf{w}_j + \alpha_j \mathbf{p}_j$
 $\mathbf{r}_{j+1} = \mathbf{r}_j - \alpha_j \mathbf{A}\mathbf{p}_j$
 $\beta_j = (\mathbf{r}_{j+1}, \mathbf{r}_{j+1}) / (\mathbf{r}_j, \mathbf{r}_j)$
 Если $\|\mathbf{r}_{j+1}\| / \|\mathbf{s}\| < \varepsilon$, то КОНЕЦ
 $\mathbf{p}_{j+1} = \mathbf{r}_{j+1} + \beta_j \mathbf{p}_j$
 увеличить j

Рис. 2. Метод сопряженных градиентов для решения СЛАУ

Центры и ширина обучались по формулам шага 1 алгоритма градиентного спуска.

3. Алгоритмы распараллеливания вычислений на графическом процессоре в технологии CUDA

Как было описано выше, для минимизации весов сети использовался алгоритм сопряженных градиентов. Распараллеливание данного алгоритма заключается в основном в параллельной реализации матрично-векторных операций, как описано в [10]. Но кроме коррекции весов большая доля времени приходится на вычисление матриц **M** и **N**, расстояний r_{ik} , коррекцию векторов ширины и центров. При этом элементы данных матриц и векторов находятся не с помощью матрично-векторных операций, а в циклах, что делает невозможным использование стандартной библиотеки CUDA CUBLAS. Например, функция вычисления матрицы **M** и **r₁** в MATLAB представлена на рис. 3.

```
function r1m
global w a nv m r2_v f R1 M
R1=zeros(1,nv);
M=zeros(nv,m);
for i=1:nv
    for k=1:m
        r2=r2_v(i,k);
        M(i,k)=4*exp(-r2/(a(k)^2))*((r2-
a(k)^2)/(a(k)^4));
        R1(i)=R1(i)+M(i,k)*w(k);
    end
    R1(i)=R1(i)-f(i);
end
end
```

Рис. 3. Функция вычисления матрицы **M** и вектора **r₁** в MATLAB

Для вычисления одного элемента матрицы **M** требуется прочитать из памяти r_{ik} и a_k и выполнить все операции по формуле $m_{ik} = 4e^{-\frac{r_{ik}^2}{a_k^2}} \frac{r_{ik}^2 - a_k^2}{a_k^4}$,

т.е. необходимо небольшое количество чтений из памяти и большое количество вычислений. Такие задачи прекрасно реализуются на массивно-параллельной архитектуре графического процессора, так как слабое место GPU – относительно медленное чтение из памяти – компенсируется за счет большого количества параллельно выполняемых вычислений.

На графическом процессоре каждый элемент матрицы **M** вычисляется независимо от других в отдельном потоке. Вычисление вектора

$r_{1i} = \sum_{k=1}^m m_{ik} w_k - f_i$ реализуется в том же ядре, что и матрицы **M**, для уменьшения чтений из памяти, так как найденные элементы матрицы **M** сразу ис-

пользуются для вычисления вектора r_1 через разделяемую память блока размером $1 \times m$, что можно видеть на рис. 4.

```
__global__ void r1m(float* M, float* R1, float* r2_v,
float* a, float* w, float* f, int n)
{
    const unsigned int tx = threadIdx.x;
    const unsigned int by = blockIdx.y;
    int i = tx;
    int j = by*n + i;
    extern __shared__ float sdata[];

    float r = r2_v[j];
    float ak2 = powf(a[i], 2);
    float sd = 4*expf(-r/ak2)*(r-ak2)/powf(ak2, 2);
    M[j] = sd;
    sdata[i] = sd*w[i];

    for(int stride = BLOCK_SIZE/2; stride > 0; stride >>= 1)
    {
        __syncthreads();
        if(tx < stride)
            sdata[tx] += sdata[tx + stride];
    }
    __syncthreads();

    if(tx == 0)
        R1[by] = sdata[0] - f[by];
}
```

Рис. 4. Функция вычисления матрицы $\mathbf{M}$ и вектора r_1 в CUDA

Другие вычисления алгоритма обучения радиально-базисной нейронной сети, такие как вычисление матрицы $\mathbf{N}$, коррекция центров и ширины методом градиентного спуска, метод сопряженных градиентов для обучения весов, тоже эффективно распараллеливаются на графическом процессоре, так как сама структура нейронной сети, представленной на рис. 1, подразумевает параллелизм.

4. Анализ результатов экспериментов

Алгоритм обучения радиально-базисной нейронной сети реализовывался с помощью технологии CUDA [1, 11] на графическом процессоре nVidia GeForce 8800 GTX и на центральном процессоре Intel Pentium 4 3 GHz в системе MATLAB 7.7.0. Экспериментальное исследование проводилось на примере модельной задачи (3)–(4) для $f(x, y) = \sin(\pi x) \cdot \sin(\pi y)$, $p(x, y) = 0$. Анализ результатов экспериментов показал, что часть алгоритма, не включающая обучение весов сети, распараллеливается очень эффективно, значительно уменьшая время вычислений. Метод сопряженных градиентов для обучения весов становится более эффективным при увеличении числа нейронов. Для решаемой задачи увеличивать количество нейронов не имело смысла, поэтому ускорения представленные в табл. 1, были получены в основном за счет распараллеливания шага 1, вычисления матриц $\mathbf{M}$, $\mathbf{N}$ и др. При решении более сложного уравнения эффективность распараллеливания на гра-

фическом процессоре будет значительней, но и для решаемой задачи было получено существенное ускорение. Ускорение рассчитывалось как

$$S = \frac{T_{CPU}}{T_{GPU}}, \text{ где } T_{CPU} - \text{ время решения задачи на центральном процессоре,}$$

T_{GPU} – на графическом процессоре.

На рис. 5 приведены результаты экспериментов, из которых видно, что уже при небольшом числе нейронов и контрольных точек реализация алгоритма обучения на графическом процессоре дает существенный выигрыш во времени. При увеличении числа контрольных точек было достигнуто ускорение в 83 раза.

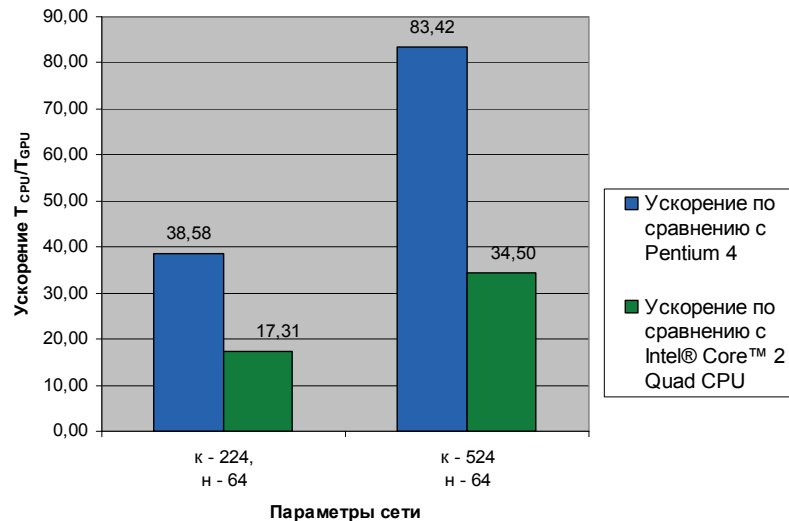


Рис. 5. Результаты экспериментов
(н – число нейронов, к – число контрольных точек)

Заключение

Проанализирован алгоритм обучения радиально-базисной нейронной сети с точки зрения его эффективного распараллеливания на графическом процессоре.

В результате проведенной работы была разработана программа, реализующая в технологии CUDA алгоритм обучения радиально-базисной нейронной сети.

В результате распараллеливания алгоритма было достигнуто существенное уменьшение времени вычислений, например, при числе нейронов 64 и числе контрольных точек 524 ускорение приблизительно равно 83.

Список литературы

1. NVIDIA CUDA Homepage. – URL: <http://www.nvidia.ru/object/cuda.html>.
2. **Buhmann, M. D.** Radial Basis Functions: Theory and Implementations / M. D. Buhmann. – Cambridge : Cambridge University Press, 2004. – 259 p.
3. **Liu, G. R.** An Introduction to Meshfree Methods and Their Programming / G. R. Liu, Y. T. Gu. – Springer, 2005. – 479 p.
4. Meshfree Methods for Partial Differential Equations / ed.: M. Griebel, Marc. A. Schweitzer. – Springer, 2008. – 412 p.

5. **Толстых, А. И.** Бессеточный метод на основе радиальных базисных функций / А. И. Толстых, Д. А. Широбоков // Журнал вычислительной математики и математической физики. – 2005. – Т. 45, № 8. – С. 1498–1505.
6. **Горбаченко, В. И.** Исследование градиентных алгоритмов обучения весов радиально-базисных нейронных сетей для решения краевых задач математической физики / В. И. Горбаченко, Е. В. Артюхина // Известия ПГПУ им. В. Г. Белинского. – 2009. – № 13 (17). – С. 134–143.
7. **Хайкин, С.** Нейронные сети: полный курс / С. Хайкин. – М. : Вильямс, 2006. – 1104 с.
8. **Осовский, С.** Нейронные сети для обработки информации / С. Осовский. – М. : Финансы и статистика, 2002. – 344 с.
9. **Jianyu, L.** Numerical solution of elliptic partial differential equation using radial basis function neural networks / L. Jianyu, L. Siwei, Q. Yingjiana, H. Yapinga // Neural Networks. – 2003. – № 16 (5/6). – P. 729–734.
10. **Горбаченко, В. И.** Реализация итерационных алгоритмов решения систем линейных алгебраических уравнений на графических процессорах в технологии CUDA / В. И. Горбаченко, Н. О. Матвеева // Вопросы радиоэлектроники. Серия «Электронная вычислительная техника». – 2008. – Вып. 5. – С. 65–75.
11. NVIDIA CUDA Compute Unified Device Architecture. Programming Guide. Version 2.0. 06.07.2008.

Горбаченко Владимир Иванович

доктор технических наук, профессор,
заведующий кафедрой информатики
и вычислительных систем, Пензенский
государственный педагогический
университет им. В. Г. Белинского

E-mail: gorvi@mail.ru

Gorbachenko Vladimir Ivanovich

Doctor of engineering sciences, professor,
head of sub-department of informatics
and computing systems, Penza State
Pedagogical University named
after V. G. Belinsky

Матвеева Наталья Олеговна

программист, отдел автоматизации,
Пензенский государственный
педагогический университет
им. В. Г. Белинского

E-mail: gorvi@mail.ru

Matveeva Natalya Olegovna

Programmer, automation department,
Penza State Pedagogical University
named after V. G. Belinsky

Гурин Евгений Иванович

доктор технических наук, профессор,
кафедра вычислительной техники,
Пензенский государственный университет

E-mail: gurin2@yandex.ru

Gurin Evgeny Ivanovich

Doctor of engineering sciences, professor,
sub-department of computer engineering,
Penza State University

УДК 004.272:004.032.26

Горбаченко, В. И.

Распараллеливание нейросетевого алгоритма решения краевых задач математической физики на многоядерных процессорах NVIDIA / В. И. Горбаченко, Н. О. Матвеева, Е. И. Гурин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 57–65.

СИТУАЦИОННАЯ АДАПТАЦИЯ ПРОСТРАНСТВЕННЫХ ДАННЫХ ДЛЯ РЕШЕНИЯ БОРТОВЫХ ГЕОИНФОРМАЦИОННЫХ ЗАДАЧ

Аннотация. Решение бортовых геоинформационных задач построено на обработке пространственных данных. Ресурсы бортовых вычислительных комплексов довольно ограничены. Для обеспечения решения ресурсоемких геоинформационных задач за приемлемое время применяются различные средства ситуационной адаптации цифровой информации о местности. Описываемый метод основан на модели, в которой пространственные данные представлены в виде совокупности последовательных приближений, что позволяет адаптировать их за счет выборки на требуемом уровне детализации.

Ключевые слова: геоинформационные задачи, пространственные данные, ситуационная адаптация, уровни детализации, вейвлет-преобразование.

Abstract. Execution of the on-board geographic information applications is based on the use of spatial data. Airborne computer systems have quite limited computational power. To provide for the execution of resource-intensive on-board geographic information applications in an acceptable time, different means of situational adaptation of digital spatial data are used. The described method is based on a model representing spatial data as a set of step-by-step approximations, which makes it possible to provide their situational adaptation by fetching data at the required level of detail.

Key words: on-board geographic information applications, spatial data, situational adaptation, levels of detail, wavelet transformation.

Введение

Большинство новых целевых задач, выполняемых комплексами бортового радиоэлектронного оборудования современных, модернизируемых и перспективных летательных аппаратов, требует для успешного решения наличия на борту баз разнообразных априорных данных, в том числе и пространственных. Некоторые из таких задач полностью построены на обработке пространственных данных и без них в принципе не решаются. Другие задачи лишь на определенных этапах выполнения требуют геоинформационной поддержки, и хотя могут быть решены без нее, но, вероятно, с худшим качеством или более низкой эффективностью. В любом случае такие бортовые задачи называются геоинформационными.

Практика показывает, что само обеспечение бортовых потребителей пространственными данными является задачей нетривиальной. Даже при условии применения для создания бортовой аппаратуры самой современной элементной базы разработчики сталкиваются с целым рядом проблем. Во-первых, с возрастанием требований к точности решения геоинформационных задач растут и требования к разрешению (уровню детализации) исходных пространственных данных, что, в свою очередь, приводит к резкому скачку объемов хранимой и обрабатываемой информации и критической нагрузке на вычислители, каналы передачи и систему отображения. Во-вторых, для бортовых комплексов характерно параллельное выполнение сразу нескольких

ресурсоемких геоинформационных задач, что требует непрерывного обеспечения в реальном времени пространственными данными одновременно нескольких потребителей. Кроме того, требования к качеству и скорости решения целевых задач всегда опережают темп роста мощностей бортовых вычислительных средств.

Таким образом, эффективное обеспечение бортовых потребителей пространственными данными требует, помимо модернизации аппаратуры, применения оригинальных архитектурных решений при проектировании бортовых систем баз пространственных данных.

Как показывает опыт разработки авиационных бортовых цифровых картографических систем, ключевое значение имеет способ организации баз данных, так как скорость, эффективность, а в некоторых случаях и возможность решения геоинформационных задач напрямую зависят от времени выборки и предобработки исходной информации.

Именно по этой причине в структуры бортовых баз пространственных данных «закладывают» средства, позволяющие адаптировать выборки соответственно решаемым задачам, снижая при этом нагрузку на вычислитель и каналы передачи данных и обеспечивая приемлемую скорость решения задач.

Способы обеспечения ситуационной адаптации в бортовых картографических системах

Под *ситуационной адаптацией* цифровой информации о местности понимается процесс настройки состава и уровня детальности пространственных данных при их выборке из соответствующих баз и передаче потребителям в зависимости от сложившихся условий, в том числе от текущей скорости, высоты, осуществляемого этапа, реализуемого режима полета, приоритета бортового потребителя, требуемых параметров запрашиваемых данных, а также оценки доступных вычислительных ресурсов [1].

В связи с тем, что информационные потребности систем-потребителей и членов экипажей на различных этапах, в различных режимах и условиях полета, а также и при выполнении различных задач существенно различаются, необходимо обеспечить адекватные сложившейся обстановке выборки данных.

Простейшие средства ситуационной адаптации уже применяются в разрабатываемых картографических системах. Например, в устройстве, описанном в [2], для каждого режима и подрежима назначается свой, соответствующий потребностям и напряженности деятельности экипажа набор параметров, определяющих содержание картографических мнемокадров. Для этого применяется *формализованный отбор* объектов, подлежащих отображению. Отбор осуществляется по ряду критериев, среди которых основные: принадлежность к определенной категории пространственной информации (элементы топогеодезической или аэронавигационной информации, элементы плана полета, результаты пространственных расчетов и т.п.); принадлежность к определенному слою пространственной информации (рельеф, гидрография, населенные пункты, дорожная сеть, растительный покров, промышленные и социально-культурные объекты, заданный маршрут, радиотехнические средства обеспечения полетов и т.п.); значимость объектов (крупные, средние, мелкие); характер влияния объектов на полет (препятствия, визуальные ориентиры, радиолокационные ориентиры и т.п.); связь характера влияния объектов на полет с условиями естественного освещения (круглосуточные, днев-

ные, ночные) и т.п. Таким образом, ситуационная адаптация осуществляется за счет управления составом выборки. Для реализации возможности осуществления ситуационной адаптации за счет настройки уровня детализации пространственных данных предлагается использовать оригинальный подход к организации баз пространственных данных, который позволил бы в значительной степени уменьшить время доставки данных до потребителя за счет внедрения средств ситуационной адаптации в саму структуру базы, позволяя задействовать эти средства уже на этапе выборки.

Разложение сигнала по базису вейвлетов

Предлагаемый подход основан на представлении сигнала в виде совокупности его последовательных приближений, а именно в виде грубой составляющей на самом низком уровне детализации и набора уточняющих коэффициентов для более высоких уровней. Такое представление исходного сигнала получается посредством разложения его по базису специальных математических функций. Например, в устройстве [3] для этой цели применялись тригонометрические функции. В предлагаемом подходе исходная информация (одномерный дискретный сигнал – в случае векторной формы представления данных, двумерный – в случае растровой формы) раскладывается по вейвлет-базису с помощью дискретного вейвлетного преобразования. Грубая составляющая сигнала, полученная таким образом на некотором минимальном уровне детализации, и уточняющие коэффициенты, полученные для более высоких уровней детализации, записываются в бортовую базу данных. При запросе записанной таким способом пространственной информации осуществляется передача сначала грубой составляющей на минимальном уровне детализации, а затем необходимое количество уточняющих коэффициентов. По полученным грубой составляющей и уточняющим коэффициентам выполняется восстановление сигнала посредством обратного дискретного вейвлетного преобразования до требуемого уровня детализации.

Разложение сигнала на грубую составляющую и уточняющие коэффициенты с помощью дискретного вейвлетного преобразования может быть представлено как дерево низкочастотных и высокочастотных фильтров [4] (рис. 1,а).

Низкочастотные фильтры, обозначенные оператором $\{L_n\}$, уменьшают количество информации исходного сигнала. Высокочастотные фильтры $\{H_n\}$ предоставляют недостающую информацию. Размерность результата преобразования низкочастотным и высокочастотным фильтрами вдвое меньше размерности входного сигнала, т.е. $L_n, H_n: R^n \rightarrow R^{n/2}$. На рис. 1 показан пример вейвлет-преобразования сигнала X из пространства R^8 . Вейвлет-преобразование этого сигнала есть элемент пространства R^8 вида $\{L_2L_4L_8X, H_2H_4L_8X, H_4L_8X, H_8X\}$, где $L_2L_4L_8X$ – грубое приближение сигнала на минимальном уровне разрешения, а остальные члены – векторы деталей, сохраняемых на каждом шаге преобразования. Операторы могут быть представлены в виде матриц, элементы $L_n[i, j]$ и $H_n[i, j]$ которых задаются следующим образом:

– пусть $k = j - 2i + 1$, тогда:

$$L_n[i, j] = \begin{cases} c_k, & \text{если } 0 \leq k \leq N - 1, \\ c_{k+n}, & \text{если } k < 0 \text{ и } k + n \leq N - 1, \\ 0, & \text{если } k < 0 \text{ и } k + n > N - 1; \end{cases} \quad (1)$$

– пусть $k = 2i - j$, тогда:

$$\mathbf{H}_n[i, j] = \begin{cases} (-1)^{j+1} c_k, & \text{если } 0 \leq k \leq N-1, \\ (-1)^{j+1} c_{k+n}, & \text{если } k < 0 \text{ и } k+n \leq N-1, \\ 0, & \text{если } k < 0 \text{ и } k+n > N-1, \end{cases} \quad (2)$$

где i – номер строки матрицы ($i = 1, 2, \dots, n/2$); j – номер столбца матрицы ($j = 1, 2, \dots, n$); n – длина сигнала; c_k – коэффициенты фильтра, зависящие от вейвлета, используемого в преобразовании ($k = 0, 1, 2, \dots, N-1$); N – количество коэффициентов вейвлета, используемого в преобразовании.

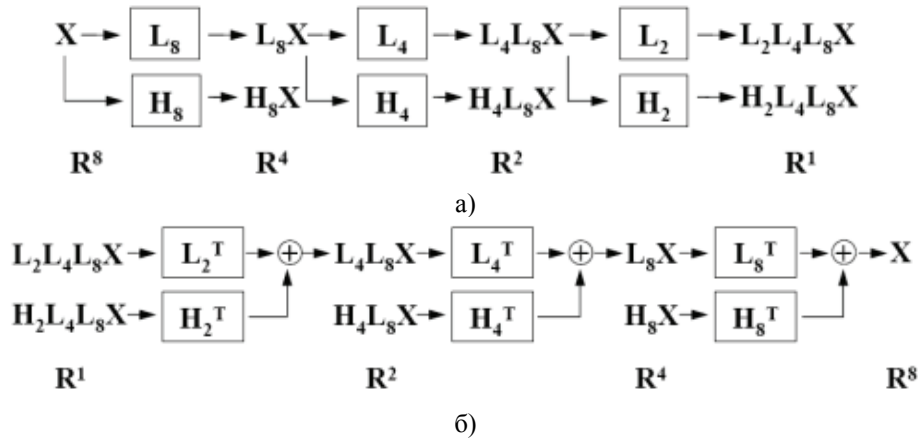


Рис. 1. Схема прямого (а) и обратного (б) дискретного вейвлет-преобразования

Если, например, вейвлет задается шестью коэффициентами $\{c_0, c_1, c_2, c_3, c_4, c_5\}$, а длина исходного сигнала $\mathbf{X}$ равна 8, то матрицы $\mathbf{L}_8$ и $\mathbf{H}_8$ примут вид

$$\mathbf{L}_8 = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 & c_4 & c_5 & 0 & 0 \\ 0 & 0 & c_0 & c_1 & c_2 & c_3 & c_4 & c_5 \\ c_4 & c_5 & 0 & 0 & c_0 & c_1 & c_2 & c_3 \\ c_2 & c_3 & c_4 & c_5 & 0 & 0 & c_0 & c_1 \end{bmatrix}; \quad (3)$$

$$\mathbf{H}_8 = \begin{bmatrix} c_1 & -c_0 & 0 & 0 & c_5 & -c_4 & c_3 & -c_2 \\ c_3 & -c_2 & c_1 & -c_0 & 0 & 0 & c_5 & -c_4 \\ c_5 & -c_4 & c_3 & -c_2 & c_1 & -c_0 & 0 & 0 \\ 0 & 0 & c_5 & -c_4 & c_3 & -c_2 & c_1 & -c_0 \end{bmatrix}. \quad (4)$$

Таким образом, вейвлетное преобразование может быть представлено в следующем виде:

$$\begin{bmatrix} \mathbf{a}_{m-1} \\ \mathbf{d}_{m-1} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_m \\ \mathbf{H}_m \end{bmatrix} \mathbf{X}, \quad (5)$$

здесь

$$\mathbf{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_{2^m} \end{bmatrix}, \quad \mathbf{a}_{m-1} = \begin{bmatrix} a_{m-1,0} \\ \vdots \\ a_{m-1,2^{m-1}-1} \end{bmatrix}, \quad \mathbf{d}_{m-1} = \begin{bmatrix} d_{m-1,0} \\ \vdots \\ d_{m-1,2^{m-1}-1} \end{bmatrix}, \quad (6)$$

где $\mathbf{X}$ – вектор-столбец исходного сигнала с 2^m элементами; $\mathbf{L}_m, \mathbf{H}_m$ – матрицы фильтров, каждый из которых имеет размер $2^{m-1} \times 2^m$; $\mathbf{a}_{m-1}$ и $\mathbf{d}_{m-1}$ – векторы-столбцы с 2^{m-1} элементами, представляющие приближение сигнала на более низком уровне разрешения, и коэффициенты детализации соответственно. Продолжая процесс, будем применять низкочастотные и высокочастотные фильтры к $\mathbf{a}_k$ до достижения требуемого уровня разрешения. Результирующее вейвлетное преобразование $\{\mathbf{a}_k, \mathbf{d}_k, \mathbf{d}_{k+1}, \dots, \mathbf{d}_{m-1}\}$, состоящее из 2^m элементов, заносится в базу данных. Коэффициенты могут быть специальным образом закодированы, что дает возможность сжать информацию перед записью ее в бортовую базу данных.

Процесс реконструкции пространственной информации показан на рис. 1,б. При осуществлении обратного дискретного вейвлетного преобразования выполняется следующая операция:

$$\mathbf{a}_k = \begin{bmatrix} \mathbf{L}_k \\ \mathbf{H}_k \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{a}_{k-1} \\ \mathbf{d}_{k-1} \end{bmatrix}. \quad (7)$$

Так как базисные функции (масштабирующая и вейвлет-функции) ортонормированны, обратную матрицу можно записать в следующем виде:

$$\begin{bmatrix} \mathbf{L}_k \\ \mathbf{H}_k \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{L}_k^T & \mathbf{H}_k^T \end{bmatrix}, \quad (8)$$

тогда шаг обратного дискретного вейвлетного преобразования можно записать в виде

$$\mathbf{a}_k = \begin{bmatrix} \mathbf{L}_k \\ \mathbf{H}_k \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{a}_{k-1} \\ \mathbf{d}_{k-1} \end{bmatrix} = \begin{bmatrix} \mathbf{L}_k^T & \mathbf{H}_k^T \end{bmatrix} \begin{bmatrix} \mathbf{a}_{k-1} \\ \mathbf{d}_{k-1} \end{bmatrix} = \mathbf{L}_k^T \mathbf{a}_{k-1} + \mathbf{H}_k^T \mathbf{d}_{k-1}. \quad (9)$$

Упаковка и упорядочивание информации в базе данных

Полученный набор коэффициентов, записывается в базу данных (рис. 2). На рис. 2 : $\mathbf{a}_3$ – коэффициенты аппроксимации (фактически, приближение исходного сигнала) на 3-м уровне детализации; $\mathbf{d}_3, \mathbf{d}_2, \mathbf{d}_1$ – коэффициенты детализации соответственно на 3-м, 2-м и 1-м уровнях детализации.

Самым общим принципом, лежащим в основе организации баз пространственных данных, является то, что данные в файлах упорядочиваются на основе пространственной кластеризации [5]. Целью кластеризации является сокращение времени поиска и времени задержки при формировании результатов часто встречающихся запросов. По отношению к пространственным базам данных это означает, что объекты, являющиеся смежными в про-

пространстве и нередко запрашиваемые совместно, должны на самом деле храниться вместе и во вторичной памяти. В многомерном пространстве (области определения пространственных данных) отсутствует естественный порядок. Накопитель данных, как правило, является одномерным логическим устройством. Поэтому требуется выполнить такое отображение пространства высокой размерности на одномерное пространство, которое сохраняет расстояние, иначе говоря, пространственно близкие элементы отображаются на близкие точки прямой, причем однозначно: никакие две точки пространства не отображаются на одну и ту же точку прямой.



Рис. 2. Порядок размещения коэффициентов в базе данных

Для обеспечения эффективного поточного считывания информации (подкачки) вся территория, содержащаяся в базе данных, разбивается на фрагменты определенного размера, что позволяет считывать информацию из энергонезависимого накопителя и обновлять данные в оперативной памяти небольшими блоками.

На рис. 3 изображено вейвлет-преобразование, построенное для четырех фрагментов рельефа в районе горы Эльбрус. Здесь n – размерность «грубого» приближения a_3 матрицы рельефа на уровне минимальной детализации; d_3 , d_2 и d_1 – уточняющие коэффициенты для уровней более высокой детализации.

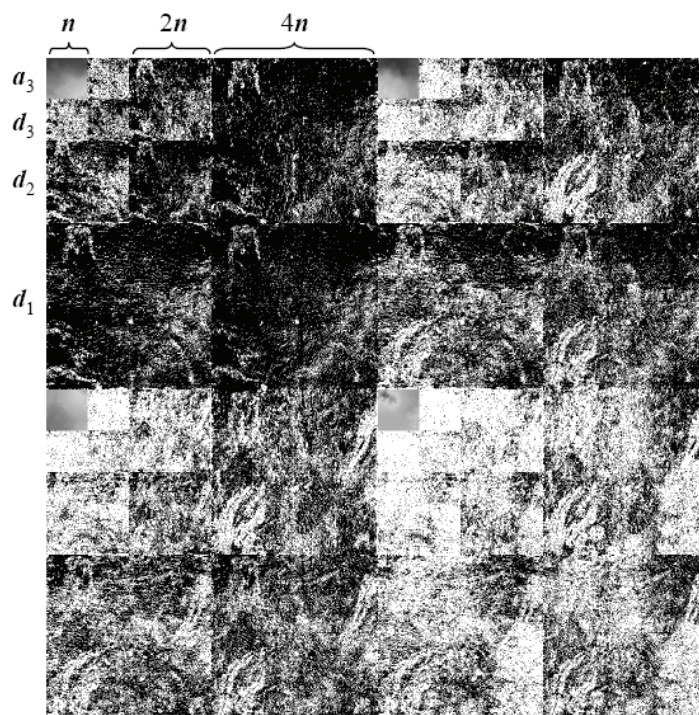


Рис. 3. Двумерное вейвлет-преобразование четырех фрагментов рельефа местности (район горы Эльбрус)

Информация для нескольких соседних фрагментов должна запрашиваться совместно, поэтому данные перегруппируются таким образом, чтобы можно было получить информацию по рельефу на определенном уровне детализации за меньшее число обращений к накопителю. Таким образом, в накопителе размещаются сначала коэффициенты аппроксимации для всех фрагментов (в соответствии со схемой упорядочивания), затем коэффициенты детализации уровня 3, уровня 2 и уровня 1 (рис. 4). На рис. 4: $a_{3,1}...a_{3,4}$ – коэффициенты аппроксимации для фрагментов 1–4; $d_{3,1}...d_{3,4}$ – коэффициенты детализации на уровне 3, $d_{2,1}...d_{2,4}$ – коэффициенты детализации на уровне 2; $d_{1,1}...d_{1,4}$ – коэффициенты детализации на уровне 1 для этих же фрагментов.



Рис. 4. Упорядочивание коэффициентов на основе пространственной кластеризации

Показатели отклонения рельефа на различных уровнях детализации

В табл. 1 представлены показатели, характеризующие отклонение рельефа на уровне максимальной детализации от его аппроксимации на 1-м, 2-м и 3-м уровнях детализации.

Таблица 1

Расхождения между рельефом на уровнях максимальной и минимальной детализации (для 16 фрагментов района горы Эльбрус)

Фрагмент	$\Delta^1_{\max}$, м	σ^1 , м	$\Delta^2_{\max}$, м	σ^2 , м	$\Delta^3_{\max}$, м	σ^3 , м
1	13,15	0,69	16,21	1,52	23,95	3,20
2	10,31	0,63	11,42	1,19	18,28	2,44
3	13,72	0,94	22,10	1,88	33,02	3,67
4	9,98	0,66	14,66	1,24	25,61	2,57
5	34,23	0,93	40,37	1,91	42,95	3,68
6	12,36	1,02	14,70	1,96	21,21	3,99
7	35,06	2,78	41,00	4,64	53,68	8,19
8	14,24	1,11	20,14	1,94	23,66	3,95
9	21,93	0,96	25,33	2,01	33,83	4,02
10	28,52	1,86	40,27	3,18	47,86	5,95
11	34,10	2,77	52,46	4,82	73,35	9,25
12	24,17	1,56	26,82	2,65	34,77	5,38
13	12,11	0,52	18,80	1,14	20,71	2,25
14	17,81	0,67	19,31	1,40	28,69	3,19
15	14,39	1,03	23,36	2,10	34,60	4,42
16	20,38	3,48	27,80	5,35	41,43	9,26

Район был разбит на 16 фрагментов (аналогично разбиению на 4 фрагмента, представленному на рис. 3) и для каждого из фрагментов были вычислены максимальное отклонение ($\Delta_{\max}$) и среднеквадратическое отклонение (σ).

Полученные отклонения должны быть также занесены в базу данных рельефа и могут быть использованы для оценки точности геоинформационной поддержки при решении целевых задач в каждом фрагменте.

Для целевых задач с жесткими требованиями к исходным пространственным данным восстановление выполняется до максимального уровня

разрешения. Для менее требовательных задач можно выполнять восстановление данных до уровня разрешения ниже максимального (или восстановление только части данных до максимального уровня). Например, для задач визуализации, требующих значительных ресурсов, но не критичных к точности исходной информации, можно использовать данные на уровнях более низкой детализации.

На рис. 5 показан режим формирования плоского изображения рельефа методом светотеневой пластики.

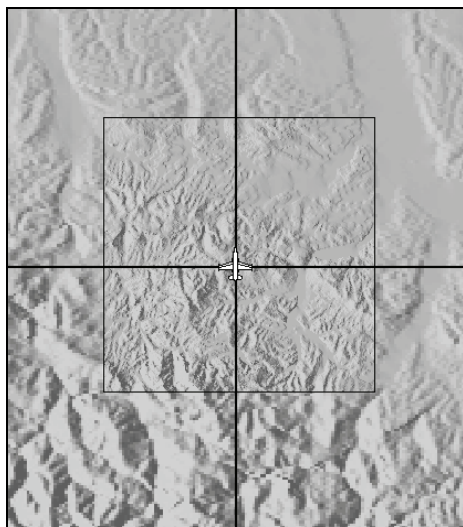


Рис. 5. Адаптация пространственного разрешения рельефа земной поверхности при формировании картографического изображения на бортовом индикаторе

На рис. 5 на всю область экрана индикатора отображаются данные по рельефу земной поверхности в районе полета низкого разрешения, а на небольшой фрагмент этого района в непосредственной близости от летательного аппарата отображается рельеф, восстановленный до максимального разрешения. Детальность участков с низким разрешением на 2 уровня ниже, чем участков с высоким разрешением. Значит, для хранения значений превышений на участок с низким разрешением понадобится приблизительно в 16 раз меньше памяти, чем на участок с высоким разрешением. Кроме того, такой режим позволит экипажу сосредоточить внимание на более четкой зоне высокого разрешения, а по всей остальной территории иметь только тенденцию пространственного распределения превышений высот.

На рис. 6 представлены трехмерные изображения рельефа, построенные по данным на нулевом (слева) и третьем (справа) уровнях детализации. Тенденции распределения поля рельефа переданы довольно точно в обоих случаях и визуально практически неразличимы.

Заключение

Применение описанного метода ситуационной адаптации и соответствующего подхода к организации баз пространственных данных может обеспечить снижение нагрузки на бортовые системы, прежде всего на систему управления базами данных и систему визуализации за счет допустимого

снижении уровня детальности запрашиваемых и отображаемых данных. Это, в свою очередь, может способствовать обеспечению непрерывности и «живучести» процессов выборки, обработки и выдачи информации бортовым потребителям и, в итоге, повышению эффективности решения целевых задач.

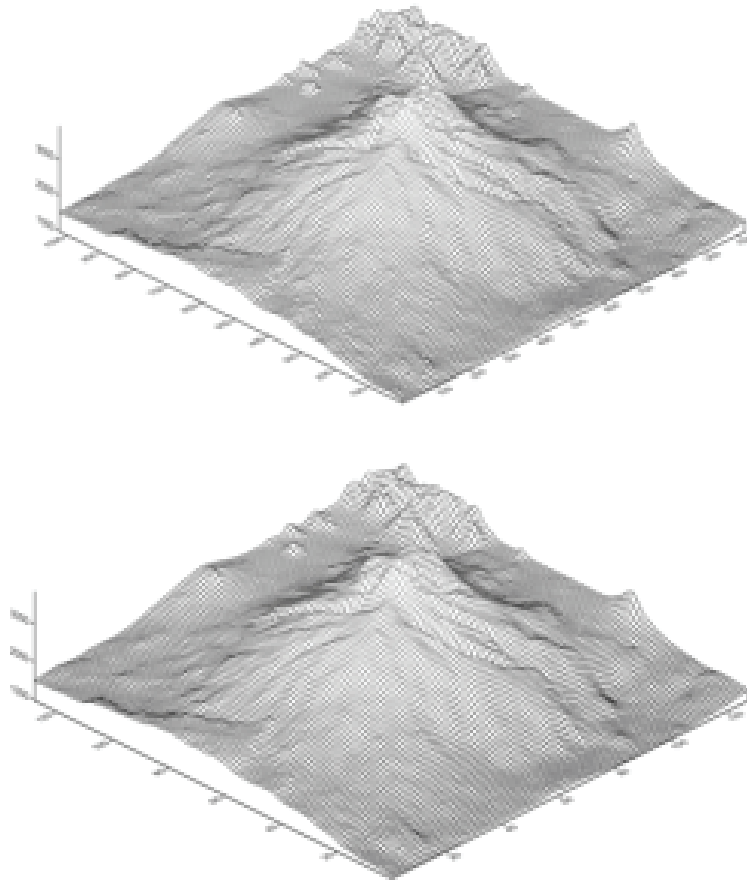


Рис. 6. Трехмерные изображения рельефа местности

Список литературы

1. **Малынкин, К. В.** Ситуационная адаптация цифровой информации о местности для бортовых геоинформационных задач / К. В. Малынкин // Вопросы оборонной техники. Серия 9. – 2011. – Вып. 1–2. – С. 246–247.
2. Пат. 2 250 182 Российская Федерация. Устройство синтеза картографических изображений / Бабак В. П., Быков В. Н., Виноградов Ю. Н., Ильченко Ю. А., Парамонов П. П., Суслов В. Д., Сухомлинов Д. В., Уткин Б. В., Юшинский Ю. Т. ; заявители и патентообладатели Бабак В. П., Быков В. Н., Виноградов Ю. Н., Ильченко Ю. А., Парамонов П. П., Суслов В. Д., Сухомлинов Д. В., Уткин Б. В., Юшинский Ю. Т. – № 2004121191/11 ; заявл. 13.07.2004 ; опубл. 20.04.2005, Бюл. № 11.
3. **Уэлстид, С.** Фракталы и вейвлеты для сжатия изображений в действии : учебное пособие / С. Уэлстид. – М. : Триумф, 2003 – 320 с.
4. United States Patent № 5.140.532. DIGITAL MAP GENERATOR AND DISPLAY SYSTEM / Paul B. Beckwith, Jr., Indialantic; Kent P. Bascle, Melbourne; Luen C.

Chan, Indian Harbour Beach; Wayne E. Basta, Palm Bay, all of Fla., Harris Corporation. – Melbourne, Fla., 1992.

5. **Шекхар, Ш.** Основы пространственных баз данных : пер. с англ. / Ш. Шекхар, С. Чаула. – М. : КУДИЦ-ОБРАЗ, 2004. – 336 с.

Малынкин Константин Вадимович

заместитель начальника, отдел разработки бортовых и наземных информационных систем, Научно-производственное объединение «Мобильные информационные системы» (г. Москва)

E-mail: kmalynkin@npomis.ru

Malynkin Konstantin Vadimovich

Deputy director, Airborne and Ground-based Information Systems Department, JSC SPA «Mobile Information Systems» (Moscow)

Мухин Иван Борисович

старший инженер-программист, Научно-производственное объединение «Мобильные информационные системы» (г. Москва)

E-mail: muhin.ivan@gmail.com

Mukhin Ivan Borisovich

Senior Programming Engineer, JSC SPA «Mobile Information Systems» (Moscow)

УДК 629.7.05

Малынкин, К. В.

Ситуационная адаптация пространственных данных для решения бортовых геоинформационных задач / К. В. Малынкин, И. Б. Мухин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 66–75.

**ОБ УЛУЧШЕНИИ КАЧЕСТВА ПРОГНОЗИРОВАНИЯ
РАСПРОСТРАНЕНИЯ РАСТИТЕЛЬНОСТИ НА ТЕРРИТОРИИ
РОССИИ ПО ДАННЫМ КОСМИЧЕСКОГО ЗОНДИРОВАНИЯ
(на примере модели SEVER)¹**

Аннотация. Предложен метод оценки качества глобальных моделей динамики растительности (на примере модели SEVER) с помощью данных дистанционного зондирования Земли из космоса, а также методы улучшения этого качества. В алгоритмы модели внесены изменения, а также проведена оптимизация ряда параметров модели, целью которой являлось улучшение качества работы модели для территории России. Полученная в результате модель позволяет значительно точнее воспроизводить распределение видов растительности на территории России.

Ключевые слова: моделирование, растительность, SEVER, оптимизация.

Abstract. The authors propose a method of quality estimation for Dynamic Global Vegetation Model, based on comparison with remote sensing products, and methods to improve the quality. In order to improve the model quality for the territory of Russia, the researchers have changed the algorithm and optimized the model parameters. The final model more precisely reproduces distribution of vegetation types for the territory of Russia.

Key words: modeling, vegetation, DGVM, SEVER, optimization.

Введение

В последние несколько десятилетий внимание исследователей все больше привлекают изменения, происходящие в климате Земли. Эти изменения существенно сказываются на растительном покрове земли, в свою очередь растительность земли также влияет на изменения в климате. В связи с этими процессами возросла потребность в методах оценки изменений в биомассе и продуктивности растительных экосистем, а также в пространственном распределении различных видов растительности и их влиянии на глобальный цикл углерода. Особенно эти оценки важны для России в связи с тем, что она занимает 1/6 всей площади суши и содержит 40 % хвойных лесов Земли.

Существует множество теорий, которые объясняют эти изменения как естественными, так и антропогенными факторами. Для оценки этих изменений можно использовать глобальные модели растительности. Пример такого их применения приведен в [1]. Однако возможность применения этих моделей на континентальном уровне еще не исследована. Не ясно, каким будет качество построенных прогнозов применительно к территории России.

В работе исследуется поведение глобальной модели динамики растительного покрова SEVER [2] при ее применении к территории России. Мо-

¹ Работа выполнена при поддержке гранта Правительства РФ для государственной поддержки научных исследований, проводимых под руководством ведущих ученых в российских образовательных учреждениях высшего профессионального образования по направлению «Космические исследования и технологии» 2011–2013 гг.

дель растительности SEVER основана на модели LPJ (Lund-Postdam-Jena) [3]. Эта модель имитирует основные процессы, происходящие в растительности, такие как фотосинтез, соревнование за ресурсы, изменения в плотности популяции и в запасе органических остатков в почве и лесной подстилке и пожары. Данная модель использует 10 типов растительности обобщенных по морфологическим, фенологическим и биоклиматическим признакам. Моделирование происходит для некоторого среднего индивида на уровне клетки $0,5 \times 0,5$.

Для оценки качества работы модели используется карта земных покровов TerraNorte RLC, созданная в Институте космических исследований РАН [4]. Эта карта была построена по данным спектрорадиометра MODIS и не имеет в настоящее время аналогов по сочетанию таких характеристик, как пространственное разрешение (250 м), тематическая детальность (22 тематических класса) и уровень достоверности при охвате всей территории страны.

1. Оценка качества модели

Для оценки качества модели растительности SEVER на территории России использовалась карта растительных покровов России TerraNorte, построенная по спутниковым данным. Качество оценивалось по двум критериям: визуальное сравнение распределения растительности по карте и в модели и формальное сравнение, основанное на корреляциях.

По визуальному сравнению данных модели и карты земных покровов (рис. 1, а, г) можно сказать, что модель показывает адекватное распределение растительности при сравнении с картой TerraNorte, различные виды растительности располагаются близко к своим настоящим ареалам распространения. Однако, хотя ареалы различных видов растительности в соответствии с моделью близки к реальным, между ними наблюдаются и отличия, не столь значительные для глобальной модели, но существенные для модели, работающей на территории только России.

Для построения формальной оценки качества оптимизации воспользуемся корреляцией между процентным содержанием растительности по выходным данным модели и данным TerraNorte для каждой ячейки модели попарно. Так как площадь ячеек неодинакова (сторона клетки всегда равна 0,5 градусов, но значение градуса долготы в километрах изменяется с изменением широты), то использовалась корреляция с весами, где в качестве весов бралась площадь данной ячейки. Формула для этого вида корреляции имеет следующий вид:

$$\bar{X}_\omega = \frac{\sum_i \omega_i x_i}{\sum_i \omega_i},$$

$$\text{cov}(X, Y)_\omega = \frac{\sum_i \omega_i (x_i - \bar{X}_\omega)(y_i - \bar{Y}_\omega)}{\sum_i \omega_i},$$

$$\text{corr}(X, Y)_\omega = \frac{\text{cov}(X, Y)_\omega}{\sqrt{\text{cov}(X, X)_\omega \text{cov}(Y, Y)_\omega}},$$

где ω_i – вес каждого наблюдения, индексом ω обозначаются среднее, ковариация и корреляция по весам; x_i и y_i – процент содержания растительности в данной клетке по данным карты TerraNorte и по данным модели соответственно.

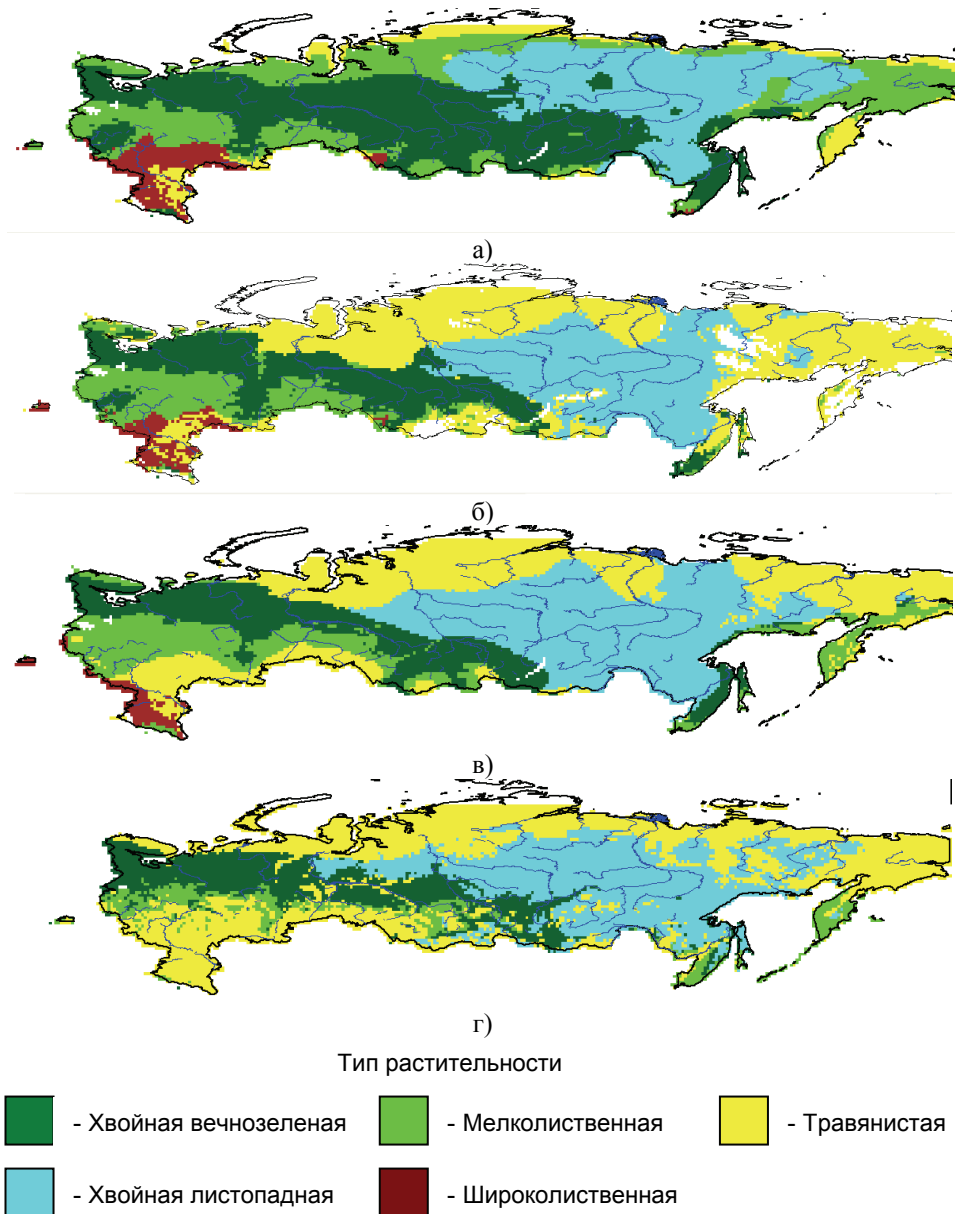


Рис. 1. Состояние модели SEVER на разных этапах работы (доминирующий тип растительности в каждой клетке): а – исходное состояние модели; б – модель после изменений; в – модель после оптимизации; г – карта TerraNorte RLC

Таким образом, получались оценки качества моделирования для каждого типа растительности отдельно. Для получения общей оценки качества работы модели бралась сумма квадратов корреляций по всем видам растительности.

Полученные для нашей исходной модели корреляции оказались довольно низкими. В частности, для широколиственных видов они близки к нулю (табл. 1). Таким образом, анализ качества работы модели в исходном состоянии указывает на необходимость ее улучшения.

Таблица 1

Качество модели на разных этапах работы

Виды моделей	Корреляции по				Общее качество (сумма квадратов)
	широколиственным лесам	хвойным вечнозеленым лесам	хвойным листопадным лесам	травянистой растительности	
Исходная модель	0,04	0,499	0,38	0,369	0,53
Модель после изменений	0,35	0,47	0,62	0,33	0,83
Модель после оптимизации	0,42	0,55	0,61	0,5	1,1

2. Корректировка модели

Для повышения качества работы модели внесены поправки как в саму модель, так и в ее параметры. В данном случае изменены алгоритмы моделирования растительности и в модель добавлено несколько процессов, неучтенных в ней ранее и важных для территории России. Во-первых, была введена дополнительная коррекция температуры по высоте рельефа над уровнем моря (так как температурные данные обладали менее точным разрешением, чем модель). Во-вторых, доработан механизм конкуренции между различными видами растительности за свет, теперь более высокие виды получают больший прирост биомассы по сравнению с более низкими. Степень увеличения прироста зависит от разницы высот видов и коэффициента соревнования по высоте, значение которого далее определено в ходе оптимизации. В-третьих, в модель введена дополнительная имитация тундры в виде травянистых видов растительности (ранее этот тип растительности не был представлен в модели). Результат всех этих улучшений представлен на рис. 1,б, и в табл. 1. Ареалы всех видов растительности значительно больше соответствуют спутниковым данным; формальная оценка качества повысилась на 55 % по сравнению с исходным значением.

В модели для воспроизведения динамики различных типов растительности используется множество параметров. Исходные значения этих параметров были заданы экспертами, создававшими модель, в соответствии с их представлениями о свойствах конкретного типа растительности. Очевидно, что эти параметры могут не совпадать со свойствами реальных типов растительности, представленных на территории России. Для улучшения воспроизводимого моделью распределения растительности оптимизированы параметры, которые могут существенно влиять на распределение типов растительности, такие как: биоклиматические ограничения на выживание и рост, индекс континентальности для лиственных, количество накопленной общей первичной продукции, необходимой для полного восстановления кроны летом и ряд других. Затем посредством их вариации и визуальной оценки изменения качества модели был выявлен набор из 12 параметров, которые могут наиболее

существенно сказаться на качестве модели. Список этих параметров приведен в табл. 2.

Таблица 2

Список параметров модели, использованных для оптимизации

Параметр	Исходное значение	Формальная оптимизация
Индекс континентальности лиственницы	43	38,75
Минимальная температура самого холодного месяца для широколиственных лесов	–17	–12,5
Минимальная температура самого холодного месяца для мелколиственных лесов	–50	–27,56
Минимальная температура самого холодного месяца для лиственницы	–50	–50
Минимальная температура самого холодного месяца для хвойных вечнозеленых лесов	–32,5	–28,6
Максимальная температура самого холодного месяца для хвойных вечнозеленых лесов	–2	3,2
Максимальная температура самого холодного месяца для мелколиственных лесов	–2	–11,4
Максимальная температура самого теплого месяца для хвойных вечнозеленых лесов	23	22,43
Максимальная температура самого теплого месяца для мелколиственных лесов	23	21,48
Дополнительное уменьшение температуры для каждые 100 м изменения высоты	0	0,00093
Коэффициент конкуренции по высоте	1	5315
Множитель для удельной листовой поверхности лиственницы	1	1,1

Целью данной оптимизации была максимизация общего качества модели посредством изменения указанных 12 параметров, т.е., как было сказано ранее, увеличение суммы квадратов корреляции по каждому типу растительности, представленному в модели.

Следует отметить, что выполнение расчетов по полной модели для всех видов растительности и всей территории России требует порядка семи часов для одноядерного компьютера или около двух часов для четырехъядерного. Низкая скорость выполнения расчетов существенно затрудняет оптимизацию параметров и увеличение качества модели.

В связи с этим для оценки качества исходной модели использовалась «огрубленная» версия, в которой осуществляется расчет только на малой части всей площади России (около 5 %). Точки, для которых осуществляется моделирование, равномерно распределены по территории России. «Огрубленная» модель для этих точек должна соответствовать исходной модели при тех же значениях параметров. Качество для приближенной модели оценивалось так же, как и для полной, на основе суммы квадратов корреляций. Время выполнения «огрубленной» версии модели составляет порядка 5 мин.

Оптимизация параметров модели проведена в два этапа. Сначала на «огрубленной» модели использован метод эффективной глобальной оптимизацией (Efficient Global Optimization, EGO). Это позволило относительно

быстро найти для нее оптимум, но при этом гарантировано, что эта точка будет являться оптимумом и для точной модели. В связи с этим был проведен второй этап оптимизации исходной модели в окрестности глобального оптимума грубой модели. Для этого использовался один из квазиньютоновских методов оптимизации BFGS (Broyden – Fletcher – Goldfarb – Shanno method). В результате этих двух шагов получены новые значения параметров, значительно улучшающие качество модели, гарантированно близкие к локальному оптимуму и при этом, возможно, близкие к глобальному оптимуму модели.

3. Аппроксимация модели методом EGO (Efficient Global Optimization)

Метод оптимизации EGO появился и стал популярен в последние два десятилетия. Впервые он был предложен в работе [5]. Этот метод основан на идее совмещения поиска наилучшего значения функции и одновременного исследования всего пространства параметров. Эти две цели совмещены благодаря использованию метода аппроксимации, называемого «кригинг». Идея кригинга заключается в представлении функции в виде суммы некоторого полинома (в нашем случае константы) и случайной величины. Первая часть обычно аппроксимирует всю функцию в среднем на пространстве, в то время как вторая часть дает приближение к каждой из точек, по которым строится эта функция. Основной особенностью кригинга является естественное определение возможной ошибки аппроксимации в каждой точке (как стандартного отклонения случайной величины) и точная интерполяция данных (как среднего от этой случайной величины).

Алгоритм EGO стартует с некоторого набора точек, по которым он строит аппроксимацию с помощью кригинга. Затем алгоритм уже для этой аппроксимации ищет наилучшие значения параметров, используя особый критерий качества, который совмещает оптимальность собственно значения функции и наибольшую неопределенность аппроксимации (наибольшую ее ошибку). Этот критерий называется «ожидаемое улучшение» (Expected Improvement, EI) и выглядит следующим образом:

$$EI(x) = (f_{\min} - \hat{y})\Phi\left(\frac{f_{\min} - \hat{y}}{s}\right) + s\phi\left(\frac{f_{\min} - \hat{y}}{s}\right),$$

где $f_{\min}$ – наименьшее найденное значение функции; $\hat{y}$ – значение функции в x ; s – стандартное отклонение аппроксимации в этой точке; Φ и ϕ – функция нормального распределения и его плотность соответственно.

Как можно видеть из приведенной формулы, значение критерия тем больше, чем меньше значение функции (по сравнению с наименьшим найденным на данный момент значением) и чем больше неопределенность аппроксимации в этой точке.

В точке, на которой значение критерия ожидаемого улучшения максимально для текущей аппроксимации, выполняется подсчет соответствующей функции и аппроксимация пересчитывается заново (с добавлением этой точки). Этот алгоритм выполняется до тех пор, пока значение EI не окажется ниже заданного порога. После этого в качестве оптимального значения берется лучшая из точек, пройденных алгоритмом.

Для выполнения всех описанных операций использовался язык для статистических вычислений R [6]. На этом языке реализован пакет DiceOptim

[7], в котором заложен алгоритм EGO и аппроксимация методом кригинга. В качестве критерия останова было выбрано значение ожидаемого улучшения, равное 0,00001 (подобрано эмпирически).

Этот этап оптимизации потребовал значительного количества времени для своего завершения. Всего было осуществлено 1253 вычисления по грубой модели. В результате были найдены значения параметров, дающие качество модели, равное 1,06 (как сумму квадратов корреляции по четырем видам растительности). Таким образом, удалось улучшить качество работы модели на 28 % по сравнению с предыдущим состоянием модели (с 0,83 до 1,06). Тем не менее, так как оптимизация проводилась на грубой модели, мы не можем утверждать, что найденные значения параметров дают даже локальный оптимум для исходной модели. Для достижения локального оптимума необходимо применить второй метод оптимизации.

4. Оптимизация модели квазиньютоновским методом (BFGS)

Квазиньютоновский метод BFGS [8] – один из наиболее часто употребляемых методов оптимизации, был почти одновременно предложен четырьмя учеными (Broyden, Fletcher, Goldfarb, Shanno). Он выполняет оптимизацию функции с помощью градиента (подсчитываемого напрямую) и гессиана (который аппроксимируется по пройденным оптимизацией точкам). Несмотря на отсутствие аналитической формы рассматриваемой модели и, как следствие, увеличенное время счета, было решено воспользоваться именно этим методом ввиду его высокой популярности среди расчетчиков. Для аппроксимации градиента была применена разностная схема. Использована версия алгоритма BFGS, имплементированная в язык R [7]. В качестве критерия останова установлено значение 0,005. Для достижения результата методу потребовался 161 цикл по вычислению модели. В результате было получено малое улучшение качества модели (с 1,062 до 1,1), т.е. несколько больше 3 %, что говорит о том, что точка, полученная на первом этапе оптимизации, уже была близка к локальному оптимуму.

В целом можно сказать, что качество работы модели удалось улучшить в основном за счет первого метода оптимизации. Использование двух методов глобальной и локальной оптимизации последовательно дает возможность утверждать, что найденный нами оптимум близок к глобальному оптимуму и с большой вероятностью близок к локальному оптимуму.

Заключение

В результате проведенных исследований было впервые оценено качество работы глобальной модели растительной динамики (SEVER) на континентальном уровне (для территории России). В целом качество работы модели было найдено удовлетворительным, но модель потребовала существенного улучшения. После внесения изменений в модель и проведения оптимизации ее параметров качество прогнозов по модели на территории России было значительно улучшено.

Впервые осуществлена формальная оптимизация параметров модели с помощью карты типов земных покровов, полученной по спутниковым данным. В целом было показано, что модель может достаточно адекватно отражать распределение растительности по данным дистанционного зондирова-

ния. Предложенная методика оптимизации параметров дала существенное увеличение качества работы модели в целом, несмотря на некоторые трудности, связанные, например, с моделированием горных массивов и территорий, измененных человеком (в результате которых затруднительно добиться полного сходства между моделью и картой).

Список литературы

1. **Cox, P. M.** Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model / Peter M. Cox et al. // Nature. – 2000. – № 408. – P. 184–187.
2. **Venevsky, S.** SEVER: A modification of the LPJ global dynamic vegetation model for daily time step and parallel computation / S. Venevsky, S. Maksyutov // Environmental Modelling & Software. – 2007. – № 22. – P. 104–109.
3. **Sitch, S.** Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model / S. Sitch et al // Global Change Biology. – 2003. – № 9. – P. 161–185.
4. **Барталев, С. А.** Новая карта растительного покрова России / С. А. Барталев, В. А. Егоров, Д. В. Ершов, А. С. Исаев и др. // Современные проблемы дистанционного зондирования Земли из космоса : тезисы докладов восьмой Всероссийской открытой конференции (Москва, ИКИ РАН, 15–19 ноября). – М., 2010.
5. **Donald, R. J.** Welch Efficient Global Optimization of Expensive Black-Box Functions / R. Jones Donald, Matthias Schonlau, J. William // Journal of Global Optimization. – 1998. – № 13. – С. 455–492.
6. R Development Core Team. R: A language and environment for statistical computing // R Foundation for Statistical Computing. – Vienna, Austria, 2010.
7. **Roustant, O.** DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization / O. Roustant, D. Ginsbourger, Y. Deville. – Preprint, 2010.
8. **Nocedal, J.** Numerical Optimization / Jorge Nocedal // Springer Science + Business Media. – 2nd ed. LLC, 2006. – 664 p.

Чумаченко Евгений Николаевич

доктор технических наук, профессор,
заведующий кафедрой математического
моделирования, Московский
государственный институт электроники
и математики (технический университет)

E-mail: mmkaf@miem.edu.ru

Chumachenko Evgeny Nikolaevich

Doctor of engineering sciences, professor,
head of sub-department of mathematical
modeling, Moscow State Institute
of Electronics and Mathematics
(technical university)

Кулагин Владимир Петрович

доктор технических наук, профессор,
ректор Московского государственного
института электроники и математики
(технический университет)

E-mail: kvp@miem.edu.ru

Kulagin Vladimir Petrovich

Doctor of engineering sciences, professor,
rector of Moscow State Institute
of Electronics and Mathematics
(technical university)

Хвостиков Сергей Антонович

аспирант, Московский государственный
институт электроники и математики
(технический университет)

E-mail: mmkaf@miem.edu.ru

Khvostikov Sergey Antonovich

Postgraduate student, Moscow State
Institute of Electronics and Mathematics
(technical university)

УДК 629.7.05

Чумаченко, Е. Н.

Об улучшении качества прогнозирования распространения растительности на территории России по данным космического зондирования (на примере модели SEVER) / Е. Н. Чумаченко, В. П. Кулагин, С. А. Хвостиков // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 76–84.

МЕТОДИКА ВЫДЕЛЕНИЯ В ЗВУКОВОМ СИГНАЛЕ УЧАСТКОВ РЕЧЕВОЙ АКТИВНОСТИ

Аннотация. Рассматривается алгоритм выделения пауз в речевом сигнале. Разработанный алгоритм основан на использовании теории активного восприятия, адаптированной к анализу речевых сигналов. Результаты проведенных экспериментов подтверждают возможность использования предложенного алгоритма для решения поставленной задачи.

Ключевые слова: цифровая обработка сигналов, теория активного восприятия, анализ речевых сигналов, выделение пауз в речевом сигнале.

Abstract. The article considers an algorithm of isolating the pauses in a speech signal. The developed algorithm is based on the use of the theory of active perception adapted for the analysis of speech signals. Results of the experiments confirm the possibility of the offered algorithm application for solving the problem.

Key words: digital signal processing, theory of active perception, speech signal analysis, pauses isolation in a speech signal.

Введение

Сегментация является одной из задач цифровой обработки и анализа речевых сигналов с целью выделения в сигнале участков речевой активности и пауз. В данной работе под паузой понимается сегмент речевого сигнала, на котором отсутствует речевая активность.

Сегментация является одним из первых этапов решения различных задач цифровой обработки сигналов, например задачи улучшения речевого сигнала, распознавания речи и т.п. В телекоммуникационных системах сегментация сигнала применяется для уменьшения объема передаваемых данных за счет исключения пауз.

Длительность паузы обычно определяется причиной ее образования. В разговорной речи паузы возникают, когда органы артикуляции находятся в сомкнутом состоянии, что связано с произношением смычных согласных. Длительность таких пауз составляет 0,1 с [1]. В процессе чтения длительность лингвистических пауз на границах синтагм (последовательности слов или морфем) не превышает 0,75 с, а между предложениями изменяется от 0,5 до 1,5 с [2, 3].

Таким образом, в данной работе будет использоваться предположение о том, что длительность паузы может изменяться от 0,1 до 1,5 с.

Одна из проблем, которая возникает при выделении пауз – зашумленность исходного сигнала. В связи с этим алгоритм сегментации сигнала должен обладать устойчивостью к наличию различного рода шумов в сигнале. Необходимо отметить, что в слуховой системе человека также реализованы механизмы помехоустойчивого восприятия [4]:

1) эфферентная обратная связь снижает чувствительность улитки при наличии шума и снижает риск перегрузки, защищает улитку от повреждения громким звуком [5];

2) бинауральное взаимодействие правого и левого каналов слуховой системы позволяет повысить разборчивость речи;

3) стремленная мышца среднего уха позволяет увеличить жесткость цепи слуховых косточек, что приводит к уменьшению проводимой ими энергии и позволяет компенсировать помехи большой интенсивности.

Предлагаемые в работе алгоритмы основаны на теории активного восприятия [6, 7].

1. Обзор существующих алгоритмов

Рассмотрим алгоритмы, применяемые при выделении пауз. В работе [8] предложен метод сегментации сигналов, основанный на теории распознавания речи и нечеткой логике. Данный метод может применяться для обработки зашумленных сигналов.

Алгоритм определения активности голоса на основе максимума апостериорной информации и критерия Неймана – Пирсона рассматривается в [9].

Алгоритм выделения пауз, описанный в [10, 11], основан на предположении о том, что дискретное преобразование Фурье речи и шума представляет собой асимптотически независимые гауссовы случайные переменные.

В работе [12] рассматривается алгоритм определения активности речи, основанный на использовании дискретного косинусного преобразования. Алгоритм разработан с учетом того, что закон распределения речи и шума соответствует гауссовой и лапласовой моделям.

Алгоритм сегментации сигнала, основанный на вычислении автокорреляционной функции сигнала, рассматривается в [13]. Данный алгоритм разработан для систем реального времени и реализован на DSP-платформе для улучшения речи в слуховых аппаратах.

Алгоритм выделения пауз, основанный на адаптивном преобразовании Карунена – Лозва, рассматривается в [14].

Необходимо отметить, что существующие методы выделения пауз обычно пытаются оптимизировать один из следующих параметров: точность выделения пауз, временную задержку, вычислительную сложность.

Для оценки качества сигнала используем отношение сигнал/шум (signal-to-noise ration, SNR) [15]:

$$SNR = 10 \log_{10} \left(\frac{\sum_{n=1}^{N_s} s^2(n)}{\sum_{n=1}^{N_v} v^2(n)} \right),$$

где s – исходный сигнал; v – шум; N_s и N_v – количество отсчетов в исходном и шумовом сигналах.

2. Предлагаемые методы выделения пауз

2.1. Алгоритм выделения пауз в неискаженном сигнале

Пусть $f(t)$ – исходный (анализируемый) сигнал. Предлагаемый алгоритм состоит в последовательном «грубо-точном» анализе сигнала:

- 1) разбить анализируемый сигнал $f(t)$ на сегменты длиной L отсчетов;
- 2) вычислить первую, вторую и третью производные для каждого сегмента;
- 3) если значения производных i -го сегмента равны нулю, то отметить рассматриваемый сегмент как относящийся к паузе и исключить его из дальнейшей обработки;

4) если значения производных i -го сегмента не равны нулю, выполнить дальнейшее разбиение рассматриваемого сегмента, после чего для каждого его подсегмента выполнить шаги 2–3.

Алгоритм завершает работу, когда будет выполнена классификация каждого сегмента сигнала или достигнут определенный уровень разрешения. Если достигнут определенный уровень разрешения и существуют неклассифицированные сегменты, производные для которых не равны нулю, то такие сегменты помечаются как относящиеся к участкам речевой активности.

С учетом того, что принятая минимальная длина паузы в работе составляет 0,1 с, можно рассчитать минимальную длину анализируемого сегмента сигнала при известной частоте дискретизации:

$$L = 0,1F, \quad (1)$$

где F – частота дискретизации сигнала.

2.2. Алгоритм выделения пауз в зашумленном сигнале

Описанный выше алгоритм позволяет сегментировать незашумленный сигнал. В реальных условиях такие сигналы вследствие наличия искажений в линиях связи, аппаратном обеспечении и наложении шума окружающей среды практически не встречаются. Следовательно, алгоритм выделения пауз должен учитывать присутствие шума. Алгоритм, предлагаемый для решения этой задачи, состоит из следующих шагов:

- 1) разбить анализируемый сигнал $f(t)$ на сегменты длиной L ;
- 2) вычислить первую, вторую, третью производные и среднее

$$M = \frac{1}{L} \sum_{i=1}^L f(t) \text{ для каждого из сегментов;}$$

- 3) вычислить суперпозицию полученных значений параметров для каждого сегмента, если результат вычисления суперпозиции для i -го сегмента превышает некоторое пороговое значение, можно предположить, что данный сегмент содержит участок речевой активности;

- 4) определить участки сигнала, в которых сосредоточена речевая активность;

- 5) реализовать механизм расширения/сжатия областей, в которых сосредоточена речевая активность с целью уточнения границ областей сигнала, относящихся к областям речевой активности (данный механизм подобен алгоритму расширения/сжатия области зрительного поля [16]).

Алгоритм завершает работу, когда будет выполнена классификация каждого сегмента сигнала.

3. Вычислительный эксперимент

Реализация предложенных алгоритмов выполнена в системе Octave. Тестирование выполнялось на ЭВМ с процессором AMD Turion 2 Dual-Core Mobile M500, 2,20 ГГц, 4 Гб оперативной памяти.

Тестовый сигнал показан на рис. 1. Частота дискретизации сигнала – 16 кГц, глубина кодирования – 16 бит, длительность – 8,192 с. Результат ручной сегментации сигнала приведен на рис. 2. Минимальный размер сегмента L при указанных параметрах сигнала составляет 1024 отсчета (64 мс). При вычислении длины сегмента по формуле (1) (рис. 3) в качестве результирующей

щей длины выбирается число, являющееся степенью двойки. Минимальная длина сегмента составляет 128 отсчетов (8 мс). Необходимо отметить, что слуховая система оценивает изменения в звуковом импульсе, происходящие на интервале времени меньше 1 мс [16].

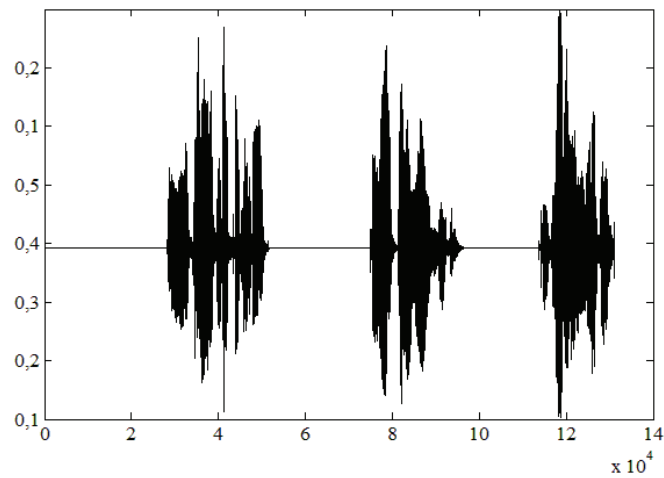


Рис. 1. Исследуемый сигнал

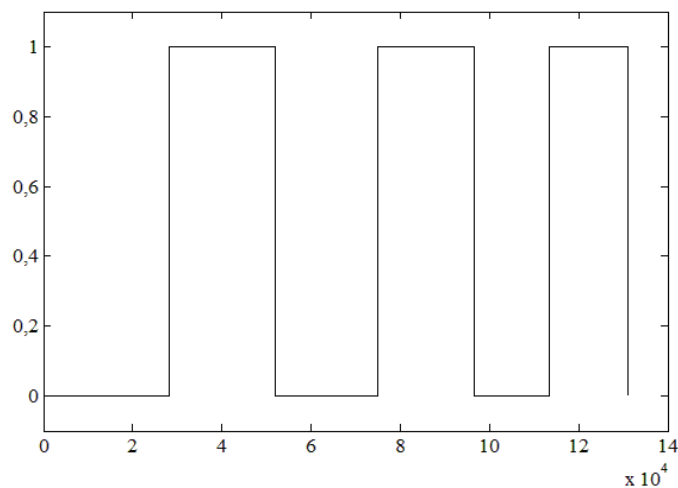


Рис. 2. Ручная сегментация сигнала

Результаты исследования предложенного алгоритма на тестовом сигнале (рис. 4) показали, что ошибка первого рода (ошибочное выделение паузы) составляет 0,02, ошибка второго рода (пропуск паузы) – 0,017. Время обработки сигнала составляет 0,18 с.

При тестировании алгоритма выделения пауз в искаженном сигнале использованы следующие аддитивные шумы:

1) теоретические (модельные):

- нормальный;
- равномерный;

2) реальные:

- шум автомагистрали;
- шум улицы;

- звук двигателя поезда;
- звук горелки аэростата.

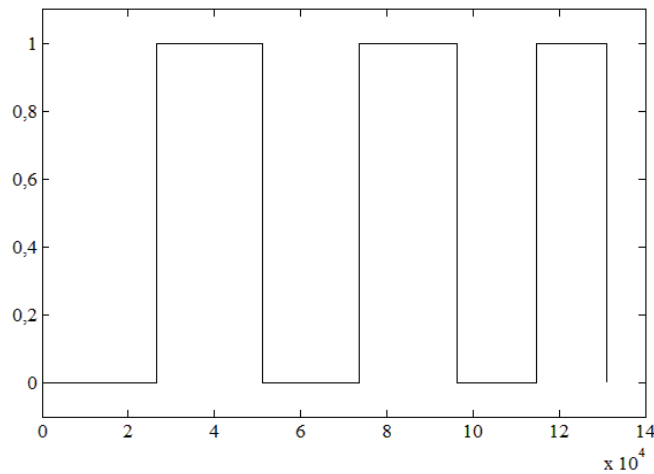


Рис. 3. Результат выделения пауз

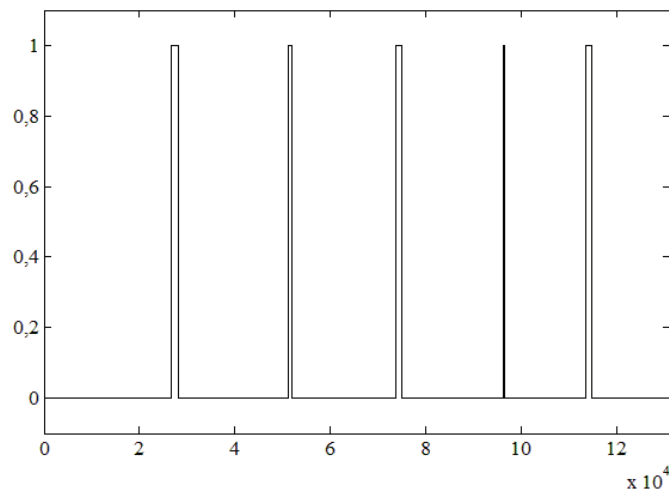


Рис. 4. Разница между ручной и автоматической сегментацией

Шум автомагистрали субъективно похож на нормальный шум. Отличие между ними заключается в том, что шум автомагистрали содержит участки нарастания интенсивности, появляющиеся в момент приближения автомобиля к устройству регистрации звука.

Длина обрабатываемого сегмента L на основе экспериментов, проведенных с незашумленным сигналом, выбрана равной 128 отсчетам. Время обработки сигнала не зависит от типа шума и составляет 0,14 с.

Результаты применения алгоритма к исходному сигналу, искаженному различными шумами, показаны в табл. 1. В столбце с заголовком **1** приведена ошибка первого рода (количество ложных выделений пауз, в процентах), в столбце **2** – ошибка второго рода (количество пропущенных пауз, в процентах).

Результаты оценки среднеквадратического отклонения (СКО) шума, искажающего сигнал, приведены в табл. 2.

Таблица 1

Результаты работы алгоритма

ОСШ	Шум											
	нормальный		равномерный		аэростата		автомагистрали		поезда		улицы	
	1	2	1	2	1	2	1	2	1	2	1	2
30 дБ	0	6	0	6	0	0	5	5,6	0	5	0	5,8
20 дБ	0	6	0	6,7	3,1	3,3	4,3	5,9	0	6	0	4,2
10 дБ	0	7,1	10	3,8	30,6	14,6	3,7	3,9	1,5	3,5	5,3	3,3
5 дБ	1,5	7,8	32,4	1,9	35,4	40,2	3,8	4,4	14,8	3,5	20,1	1,5
0 дБ	31,4	4	51,8	0	35,5	46,7	4	2,4	40,1	3,7	20,7	3,6

Таблица 2

Характеристики шума (СКО шума)

ОСШ	Шум					
	нормальный	равномерный	аэростата	автомагистрали	поезда	улицы
30 дБ	0,00171	0,00405	0,00260	0,00091	0,00089	0,00083
20 дБ	0,00559	0,01327	0,00813	0,00454	0,00445	0,00417
10 дБ	0,02021	0,04034	0,01951	0,01362	0,01423	0,00835
5 дБ	0,03495	0,05762	0,03903	0,02088	0,02313	0,01669
0 дБ	0,05013	0,11537	0,06992	0,84024	0,03503	0,03004

На рис. 5–7 приведен тестовый сигнал, искаженный различными шумами, а также результат его сегментации.

Результаты исследования помехоустойчивого алгоритма сегментации речевого сигнала:

1) при увеличении длительности обрабатываемого сигнала снижается точность сегментации, одно из возможных решений данной проблемы – отдельная обработка участков сигнала длительностью 2–3 с;

2) если сравнить точность выделения границ участков S_1 и S_2 (A_1 – максимальная амплитуда участка S_1 ; A_2 – участка S_2 ; $A_1 \ll A_2$), то точность выделения границ участка S_1 ниже, чем участка S_2 , это можно объяснить различной степенью влияния накладываемого шума на участки речевого сигнала с различной амплитудой;

3) если сравнить точность выделения границ участков S_1 и S_2 (L_1 – длина S_1 ; L_2 – участка S_2 ; $L_1 \ll L_2$), то точность выделения границ участка S_1 ниже, чем участка S_2 .

Можно отметить, что предложенный алгоритм по параметру «точность сегментации» не уступает существующим алгоритмам сегментации речевых сигналов [15], а в ряде случаев, показывает лучшие результаты.

Заключение

В работе рассмотрены подходы к сегментации речевого сигнала на участки речевой активности и паузы. Предложены два алгоритма, один из которых предназначен для сегментации речевых сигналов, искаженных шумом, второй – для сегментации неискаженных сигналов. При разработке алгоритмов выделения пауз использовались методы, основанные на аналогичных механизмах, функционирующих в системе зрительного восприятия.

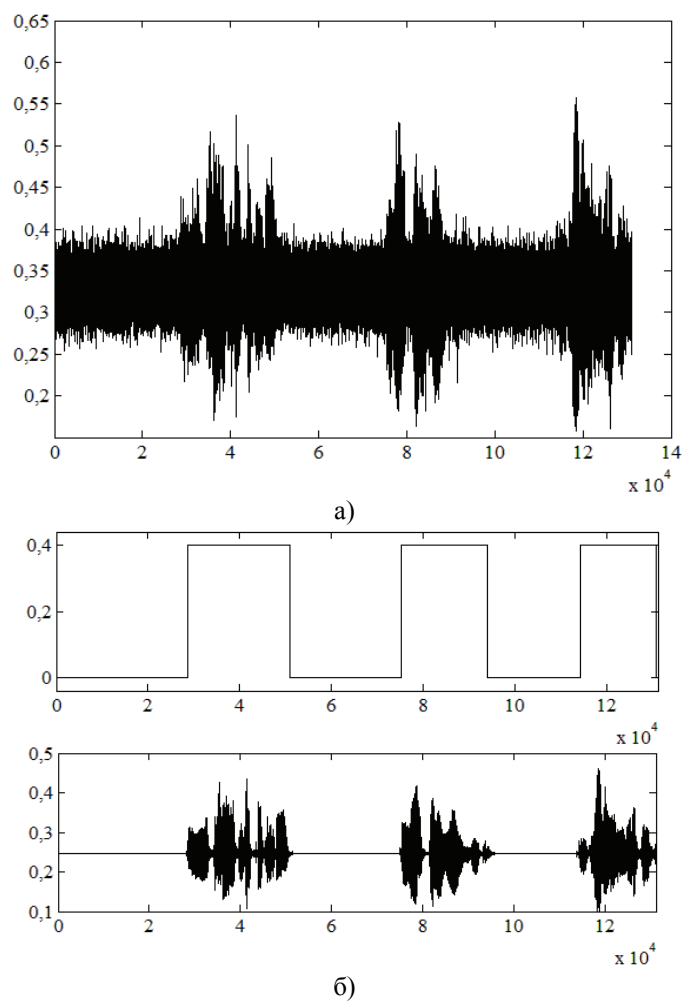


Рис. 5. Сегментация сигнала, зашумленного нормальным шумом (ОСШ = 10 дБ): а – сигнал; б – результат сегментации

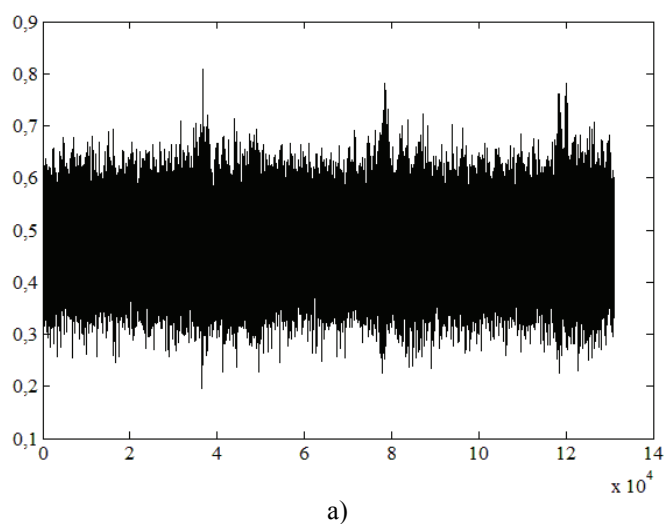


Рис. 6. Сегментация сигнала, зашумленного нормальным шумом (ОСШ = 0 дБ): а – сигнал; б – результат сегментации (см. также с. 92)

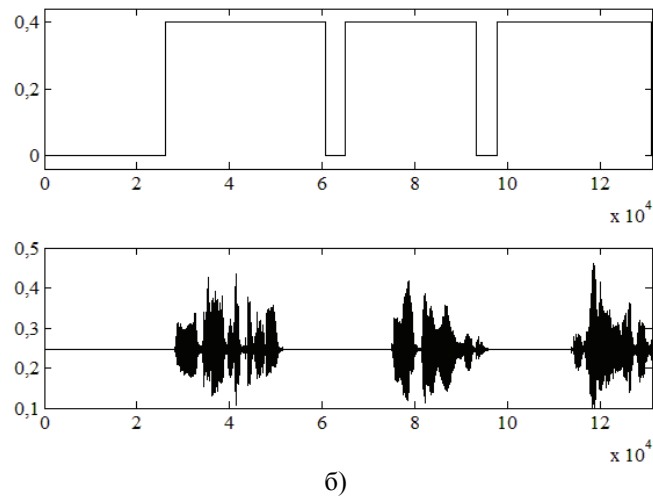


Рис. 6. Окончание

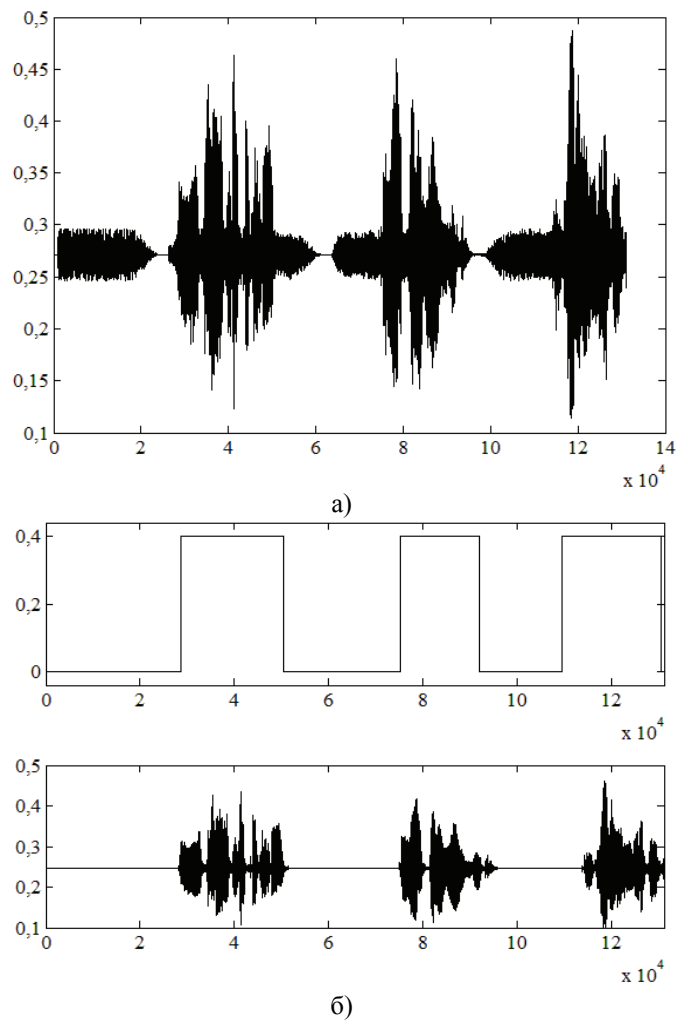


Рис. 7. Сегментация сигнала, искаженного звуком горелки аэростата (ОСШ = 20 дБ): а – сигнал; б – результат сегментации

В работах отечественных и зарубежных авторов также отмечается аналогия процессов обработки информации в зрительной и слуховой системах.

Можно отметить следующие преимущества предложенного алгоритма перед существующими: низкая вычислительная сложность, отсутствие настраиваемых параметров, простота реализации и точность выделения пауз.

Дальнейшие исследования будут направлены на анализ возможности применения предложенного подхода к сегментации речевых сигналов, искаженных мультипликативным шумом.

Список литературы

1. **Варшавский, Л. А.** Исследование формантного состава и некоторых других характеристик звуков русской речи / Л. А. Варшавский, И. М. Литвак // Проблемы физиологической акустики. – 1955. – Т. 3. – С. 5–17.
2. **Шейкин, Р. Л.** К анализу механизмов возникновения пауз в речи / Р. Л. Шейкин // Механизмы речеобразования и восприятия сложных звуков. – М.-Л. : Наука, 1966. – С. 31–44.
3. **Goldman-Eister, F.** Pauses, clauses, sentences / F. Goldman-Eister // Language and Speech. – 1972. – V. 15, № 3. – P. 103–113.
4. **Альтман, Я. А.** Тенденции развития физиологии слуха / Я. А. Альтман, С. Ф. Вайтулевич, И. А. Вартанян, И. Г. Андреева, Е. С. Малинина // Успехи физиологических наук. – 2005. – Т. 36, № 1. – С. 3–23.
5. **Николлс, Д.** От нейрона у мозга / Д. Николлс, Р. Мартин, Б. Валлас, П. Фукс ; пер. с англ. П. М. Балабана, А. В. Галкина, Р. А. Гиниатуллина, Р. Н. Хазипова, Л. С. Хируга. – М. : Едиториал УРСС, 2003. – 672 с.
6. **Утробин, В. А.** Информационные модели системы зрительного восприятия для задач компьютерной обработки изображений : учебное пособие / В. А. Утробин. – Н. Новгород : Изд-во НГТУ, 2001. – 234 с.
7. **Утробин, В. А.** Компьютерная обработка изображений. Принятие решений в пространстве эталонов : учебное пособие / В. А. Утробин. – Н. Новгород : Изд-во НГТУ, 2004. – 221 с.
8. **Beritelli, F.** A robust voice activity detector for wireless communications using soft computing / F. Beritelli, S. Casale, A. Cavallaro // IEEE Journal on Selected Areas in Communications. – 1998. – V. 16, № 9. – P. 1818–1829.
9. **McKinley, B. L.** Model based speech pause detection / B. L. McKinley, G. H. Whipple // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '97). – Munich, Germany, 1997. – V. 2. – P. 1179–1182.
10. **Sohn, J.** A statistical model-based voice activity detection / J. Sohn, N. S. Kim, W. Song // IEEE Signal Processing Letters. – 1999. – V. 6, № 1. – P. 1–3.
11. **Cho, Y. D.** Analysis and improvement of a statistical model-based voice activity detector / Y. D. Cho, A. Kondoz // IEEE Signal Processing Letters. – 2001. – V. 8, № 10. – P. 276–278.
12. **Gazor, S.** A soft voice activity detector based on a Laplacian-Gaussian model / S. Gazor, W. Zhang // IEEE Transactions on Speech and Audio Processing. – 2003. – V. 11, № 5. – P. 498–505.
13. **Sheikhzadeh, H.** Real-time implementation of HMM-based MMSE algorithm for speech enhancement in hearing aid applications / H. Sheikhzadeh, R. L. Brennan, H. Sameti // Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '95). – Detroit, Mich, USA, 1995. – V. 1. – P. 808–811.
14. **Rezayee, A.** An adaptive KLT approach for speech enhancement / A. Rezayee, S. Gazor // IEEE Transactions on Speech and Audio Processing. – 2001. – V. 9, № 2. – P. 87–95.

15. **Pwint, M.** Speech / Nonspeech Detection Using Minimal Walsh Basis Functions / M. Pwint, F. Sattar // EURASIP Journal on Audio, Speech, and Music Processing. – 2006. – V. 2007. – P. 3–12.
16. Элементы теории биологических анализаторов / под ред. Н. В. Позина. – М. : Наука. Главная редакция физико-математической литературы, 1978. – 360 с.

Гай Василий Евгеньевич

кандидат технических наук, доцент,
кафедра вычислительных систем
и технологий, Нижегородский
государственный университет
им. Р. Е. Алексеева

E-mail: iamuser@inbox.ru

Gai Vasily Evgenyevich

Candidate of engineering sciences, associate
professor, sub-department of computing
systems and technologies, Nizhny
Novgorod State University named
after R. E. Alekseev

Утробин Владимир Александрович

доктор технических наук, профессор,
кафедра вычислительных систем
и технологий, Нижегородский
государственный университет
им. Р. Е. Алексеева

E-mail: utrobin-va@yandex.ru

Utrobin Vladimir Alexandrovich

Doctor of engineering sciences, professor,
sub-department of computing systems
and technologies, Nizhny Novgorod State
University named after R. E. Alekseev

УДК 534.87

Гай, В. Е.

Методика выделения в звуковом сигнале участков речевой активности / В. Е. Гай, В. А. Утробин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 85–94.

АНАЛИТИКО-ИМИТАЦИОННЫЙ МЕТОД ОЦЕНКИ ВРЕМЕННЫХ ХАРАКТЕРИСТИК ДЛЯ РАЗЛИЧНЫХ ВАРИАНТОВ ПРЕДСТАВЛЕНИЯ ОТНОШЕНИЙ КАТЕГОРИЗАЦИИ В РЕЛЯЦИОННЫХ БАЗАХ ДАННЫХ

Аннотация. Проанализированы процедуры доступа к информации для различных вариантов представления отношений категоризации. Разработана совокупность аналитических моделей для оценки производительности фрагментов баз с категоризованными данными.

Ключевые слова: базы данных, отношения категоризации, время доступа, аналитическая модель.

Abstract. The article analyzes information access procedures for various variants of the supertype-subtype relations representation. The author has developed a set of analytical models to estimate the fragment access time in databases with categorized data.

Key words: databases, supertype-subtype relations, access time, analytical model.

Введение

Распоряжением Правительства Российской Федерации от 20 октября 2010 г. утверждена Государственная программа Российской Федерации «Информационное общество (2011–2020 годы)» [1]. В Программе указано: «повышение эффективности государственного управления предусматривает: ...создание и развитие государственных межведомственных информационных систем, предназначенных для принятия решений в реальном времени...».

Для решения поставленных задач требуется интенсификация работ по созданию и развитию средств проектирования и поддержки эксплуатации информационных систем.

Оценка производительности создаваемых баз данных (БД) на проектных стадиях обеспечивает своевременное определение эффективности принимаемых решений.

Значительная часть данных, хранящихся в БД информационных систем, связана отношениями категоризации. В статье выполнен анализ зависимости времени исполнения запроса (времени доступа) от варианта представления категоризованных данных.

1. Способы представления категоризованных отношений

Наиболее развиты способы реализации категоризованных отношений в СУБД Oracle. Варианты СУБД Oracle охватывают практически все возможные варианты других СУБД, поэтому они приняты за основу при разработке аналитического метода оценки объемов памяти, необходимых для реализации категоризованных отношений в реляционных БД. В многотомном описании методологии проектирования Oracle [2] приведены только словесные описания вариантов реализации с кратким перечнем достоинств и недостатков каждого.

На рис. 1 показана модель отношения категоризации по нотации Баркера («супертип» – «тип»).

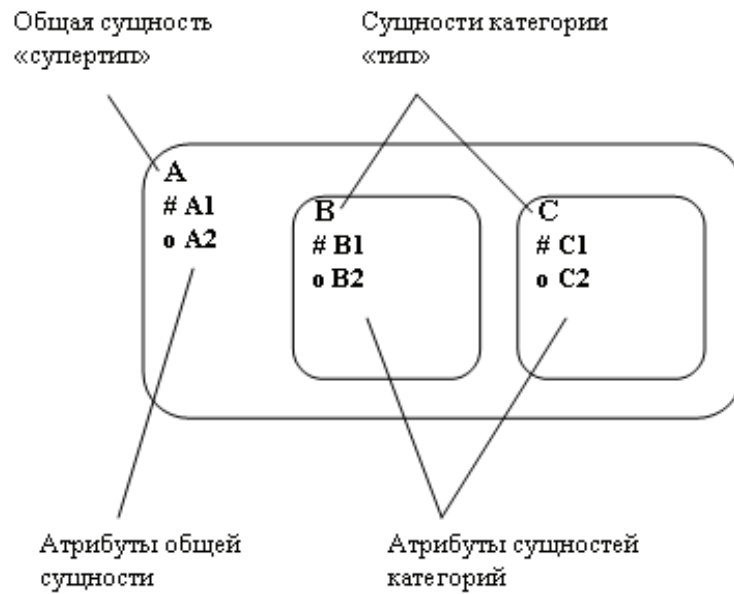


Рис. 1. Базовая ER-диаграмма

Категорированные данные могут быть представлены (рис. 2):

- в одной таблице (вариант а);
- в нескольких (по количеству категорий) таблицах (вариант б);
- в нескольких (по количеству категорий + одна) таблицах (варианты в, г).

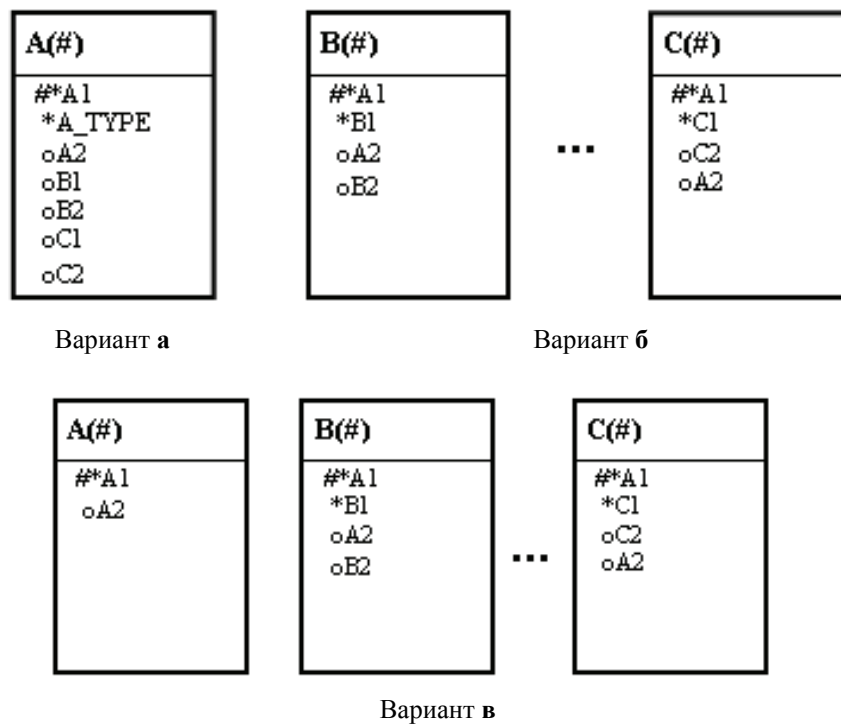
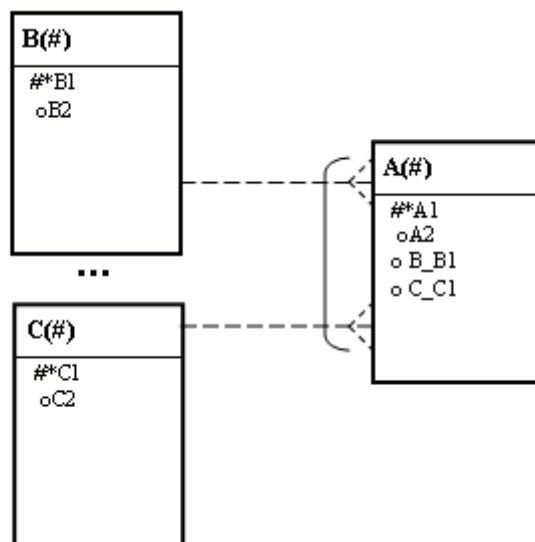


Рис. 2. Варианты реализации категорированных отношений



Вариант г

Рис. 2. Окончание

Для проектирования фрагментов баз с категоризованными данными предлагается использовать оценки времени исполнения запроса (T) к таким фрагментам.

2. Проверка гипотезы о зависимости времени доступа от числа колонок и числа записей таблицы

Сложные процессы, реализуемые в СУБД, операционных системах и в телекоммуникациях системы, постоянно меняющееся количество записей в таблицах БД приводят к тому, что время исполнения запроса имеет характер сложной функции от системы случайных величин [3, 4].

С другой стороны, время исполнения запроса к фрагментам базы с категоризованными данными, очевидно, является функцией нескольких переменных и зависит от числа таблиц, с помощью которых реализованы фрагменты, от числа записей и от числа колонок таблиц.

Гипотезу о зависимости времени доступа от числа колонок и числа записей таблицы необходимо подтвердить экспериментальными исследованиями. Для проверки была использована имитационная SQL-процедура, формирующая в тестовой таблице (TEST_TBL) различное количество колонок, записей и измеряющей время доступа к таблице.

Под временем доступа здесь понимается астрономическое время выполнения запроса типа

SELECT * FROM TEST_TBL. (1)

В созданной имитационной программе время рассчитывалось при помощи функции пакета DBMS_UTILITY – GET_TIME, позволяющей делать оценки времени с точностью до 0,01 с.

Имитационная программа использовалась для определения зависимости времени доступа к таблице данных СУБД Oracle 8. Для получения зави-

симостей времени доступа от числа колонок и числа записей тестовых таблиц использовались запросы вида (1) к таблице TEST_TBL вида

CREATE TABLE TEST_TBL (nm number (38),
nm1 number (38), nm2 number (38) ...),

(2)

причем число записей и число колонок варьировалось. В одном из вариантов в колонки записывались целые числа, равные номеру записи в таблице.

По второму варианту для получения зависимостей использовались запросы вида (1) к таблице TEST_TBL вида

CREATE TABLE TEST_TBL (nm varchar2(100)),

(3)

причем в колонку *nm* записывались строки случайной длины, длина которых равномерно распределялась в диапазоне 0...100.

Таблицы (2), (3) не индексировались и не входили в состав кластера, так что для выполнения запроса типа (1) оптимизатор ORACLE выбирал всегда режим полного поиска (FULL) в таблицах (2), (3).

Измерения проводились путем запуска файлов-процедур в SQL*Plus, причем вывод результатов запроса на экран подавлялся установкой переменной среды *termout* в FALSE, что позволяло исключить время вывода на экран из результатов измерений.

При построении графиков зависимостей вычисленное время усреднялось по нескольким (до десяти) значениям для получения каждой точки графиков.

Так как статистические данные, получаемые с помощью имитационной программы, представляют функцию от системы случайных величин, то была проведена оценка данных. Проводился расчет стандартного отклонения в каждой точке зависимостей, вычисление относительной оценки отклонения и определение арифметического среднего этих относительных оценок по всему массиву экспериментальных данных. Отклонения получаемых оценок не превышали нескольких процентов от средних значений.

Полученные зависимости (для СУБД Oracle) представлены на рис. 3.

Экспериментальные данные характеризуются линейными зависимостями времени доступа к таблице:

$$T = f_1(N);$$

(4)

$$T = f_2(k),$$

(5)

где N – число записей в БД; k – число полей в записях.

Оценки достоверности аппроксимации с помощью R^2 (для семейства зависимостей (4): 0,9986–0,9996; для семейства (5): 0,9897–0,9925) достаточны для выявления зависимости времени доступа от числа записей и от числа колонок таблиц баз данных.

Одним из эффективных методов повышения производительности баз данных является индексирование записей. При индексировании записей увеличиваются требуемые ресурсы памяти: средствами СУБД создаются дополнительные колонки, с помощью которых записи таблиц упорядочиваются. Время доступа к упорядоченным записям существенно сокращается. Зависимость времени доступа к индексированным записям также достаточно точно аппроксимируется линейным уравнением.

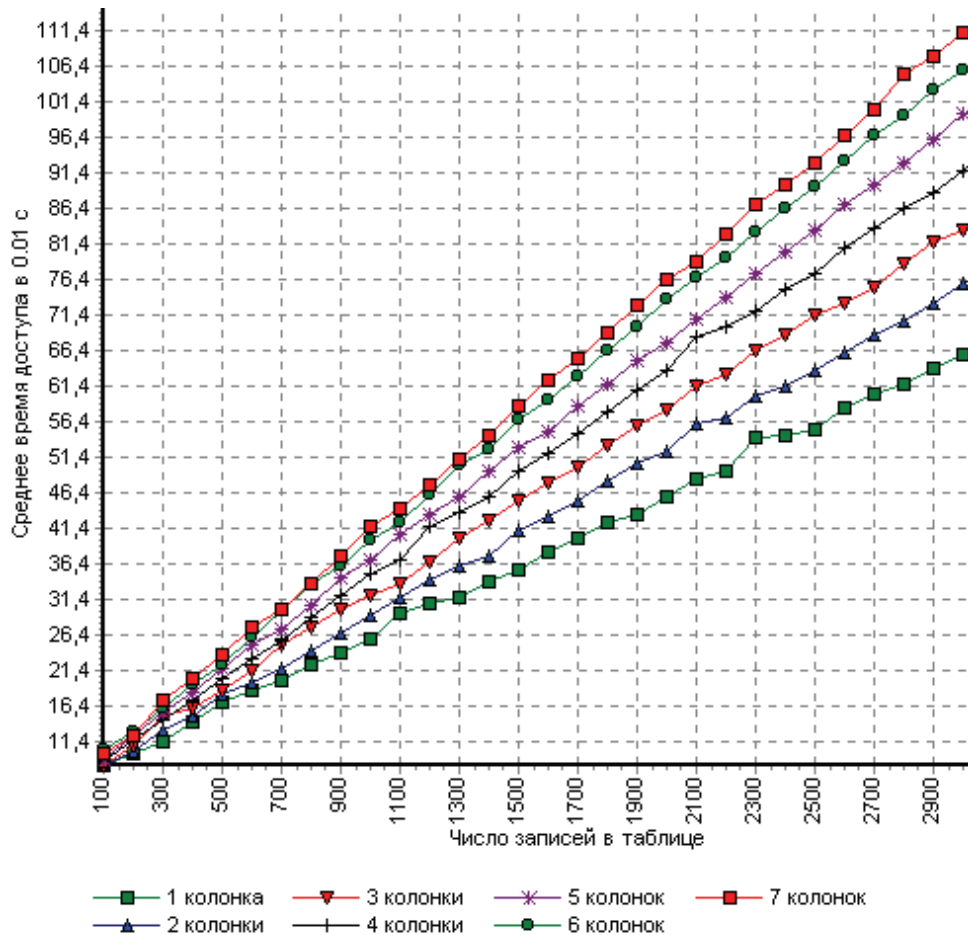


Рис. 3. Зависимости времени доступа к таблице от числа колонок и числа записей

3. Анализ данных имитационной программы

Зависимости коэффициентов m и b линейных уравнений ($y = mx + b$) семейства регрессий (4) от числа колонок таблицы представлены на рис. 4. На полученные зависимости $m = f_3(k)$ и $b = f_4(k)$ также построены линейные регрессии. Коэффициенты последних двух регрессий вида $y = mx + b$ обозначим соответственно m_m , b_m ; m_b , b_b .

4. Модель времени доступа к таблице

Предлагается использовать в качестве аналитической модели времени доступа к таблице данных функцию двух переменных N и k (число записей и число колонок таблицы):

$$T = (m_m k + b_m)N + (m_b k + b_b). \quad (6)$$

Адекватность линейной модели регрессии экспериментальным результатам подтверждена с помощью критерия отношения дисперсий.

Построенная на данных имитационной программы модель может также трактоваться как модель производительности СУБД и технических средств.

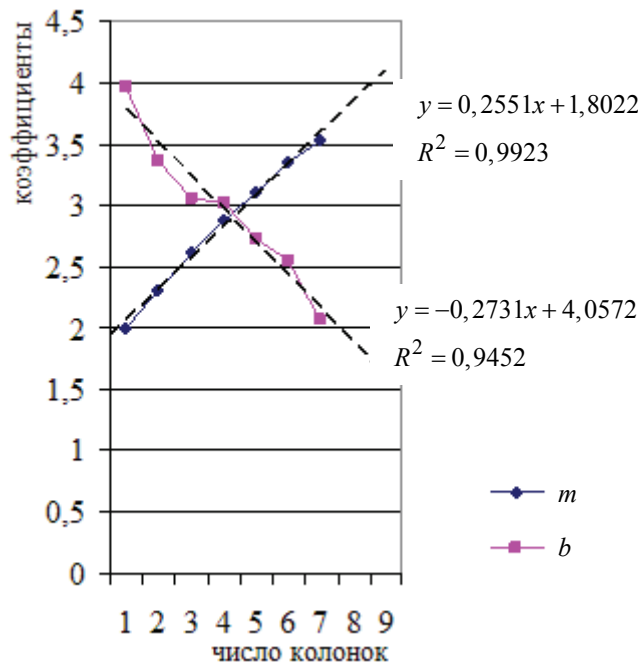


Рис. 4. Зависимости коэффициентов m и b семейства регрессий на экспериментальные данные $T = f_k(N)$ от числа колонок таблиц баз данных

Для инженерной практики необходимо использовать две модели: дополнительная модель должна быть сформирована для варианта индексирования записей.

Такие оценки позволяют использовать модель (6) в проектных работах по созданию баз данных информационных систем. Модель (6) может быть использована не только при выборе вариантов построения фрагментов баз с категоризованными данными, но и при оценке производительности других фрагментов.

5. Процедуры доступа к информации для различных вариантов представления категоризованных отношений

Введем обозначения:

N – число записей в БД;

N_i – число записей категории i ;

k_k – число категорий;

k_{oa} – число общих атрибутов;

k_{ia} – число частных атрибутов (для категории i).

Вариант а. Для выборки данных по конкретной записи достаточно выполнения одного запроса к единственной таблице.

Вариант б. Для выборки данных по конкретной записи достаточно выполнения одного запроса к одной из k_k таблиц.

Замечание. На практике в некоторых случаях возможны дополнительные затраты времени на определение таблицы, в которой находится искомая запись. В данной статье такие варианты не рассматриваются; считается, что

таблица с искомыми записями определяется реализацией отношений внешних таблиц с таблицами категорий.

Вариант в. Для выборки данных по конкретной записи достаточно выполнения одного запроса к одной из k_k таблиц. Так как общие атрибуты записей в таблице супертипа и в таблицах категорий дублируются, выборка только общих атрибутов конкретной записи может выполняться и с помощью одного запроса к таблице супертипа (по сравнению с вариантом **а** в таблице супертипа меньше колонок).

Справедливо и замечание к варианту **б**.

Вариант г. Для выборки данных по конкретной записи в полном объеме необходимо выполнение двух запросов: к одной из k_k таблиц категорий и к таблице супертипа.

Для выборки данных по конкретной записи в объеме частных атрибутов необходимо выполнение запроса: к одной из k_k таблиц категорий.

Для выборки данных по конкретной записи, но только по общим атрибутам (без выборки частных атрибутов), достаточно выполнения одного запроса к таблице супертипа; по сравнению с вариантом **в** в таблице супертипа дополнительные колонки предназначены для хранения внешних ключей, указывающих на конкретную таблицу категорий и ее запись с частными атрибутами.

В данном варианте решается и проблема поиска таблицы категории; если задача не решается на уровне структуры базы данных, то дополнительные колонки в таблице супертипа позволяют определить таблицы категорий и первичные ключи записей с частными атрибутами.

6. Аналитические модели времени доступа для различных вариантов представления категоризованных отношений

Введем обозначения для оценок времени доступа к фрагменту базы с категоризованными данными:

T_a, T_b, T_v, T_r – арифметическое среднее время доступа (АСВД) для вариантов организации **а, б, в, г** соответственно;

$T_{aoc}, T_{bos}, T_{voc}, T_{roc}$ – АСВД при условии полного выбора данных (общие и частные атрибуты записей);

$T_{ao}, T_{bo}, T_{vo}, T_{ro}$ – АСВД при выборе общих атрибутов записей;

$T_{ac}, T_{bc}, T_{vc}, T_{rc}$ – АСВД при выборе частных атрибутов;

$T_{ai}, T_{bi}, T_{vi}, T_{ri}$ – АСВД к данным категории i ;

$T_{aioc}, T_{bioc}, T_{vioc}, T_{rioc}$ – время доступа к данным категории i (ВДК) при условии выборе общих и частных атрибутов;

$T_{aio}, T_{bio}, T_{vio}, T_{rio}$ – ВДК при выборе общих атрибутов;

$T_{aic}, T_{bic}, T_{vic}, T_{ric}$ – ВДК при выборе частных атрибутов.

Вариант а:

$$T_{ai} = T_{aioc} = T_{aio} = T_{aic} = \left[m_m \left(k_{oa} + \sum_{i=1}^{k_k} k_{ia} \right) + b_m \right] N + m_b \left(k_{oa} + \sum_{i=1}^{k_k} k_{ia} \right) + b_b \quad (7)$$

Замечание. В СУБД возможна несущественная зависимость времени доступа от числа выбираемых атрибутов (время доступа к таблице зависит не

только от числа колонок в таблице, но и от числа колонок, данные из которых должны быть представлены в запросе). В статье данный вопрос детально не рассматривается.

Вариант б:

$$T_{\bar{b}i} = T_{\bar{b}i\bar{o}c} = T_{\bar{b}i\bar{o}} = T_{\bar{b}ic} = [m_m(k_{oa} + k_{ia}) + b_m]N_i + m_b(k_{oa} + k_{ia}) + b_b. \quad (8)$$

Вариант в:

$$T_{\bar{b}i\bar{o}c} = T_{\bar{b}ic} = [m_m(k_{oa} + k_{ia}) + b_m]N_i + m_b(k_{oa} + k_{ia}) + b_b; \quad (9)$$

$$T_{\bar{b}i\bar{o}} = (m_mk_{oa} + b_m)N + m_bk_{oa} + b_b. \quad (10)$$

Вариант г:

$$T_{\bar{g}i\bar{o}c} = (m_mk_{ia} + b_m)N_i + m_bk_{ia} + b_b + \\ + [m_m(k_{oa} + k_k) + b_m]N + m_b(k_{oa} + k_k) + b_b; \quad (11)$$

$$T_{\bar{g}i\bar{o}} = [m_m(k_{oa} + k_k) + b_m]N + m_b(k_{oa} + k_k) + b_b; \quad (12)$$

$$T_{\bar{g}ic} = (m_mk_{ia} + b_m)N_i + m_bk_{ia} + b_b. \quad (13)$$

Для определения оставшихся моделей необходимо учесть неравнозначность запросов к отдельным категориям данных, отразить влияние распределений обращений к категориям данных и типа обращений.

Введем обозначения: P_{O_i} – вероятности обращения к категории i ;

$|P_{B_p^i}, P_{B_o^i}, P_{B_c^i}|$ – распределение вероятностей типов обращений для категории i , где B_p^i – обращение к категории i с запросом полных данных; B_o^i – обращение к категории i с запросом только общих атрибутов; B_c^i – обращение к категории i с запросом только частных атрибутов;

$$T_a = T_{a\bar{o}c} = T_{a\bar{o}} = T_{ac} = \sum_{i=1}^{k_k} P_{O_i} T_{a\bar{i}o\bar{c}}; \quad (14)$$

$$T_{\bar{b}} = T_{\bar{b}o\bar{c}} = T_{\bar{b}o} = T_{\bar{b}c} = \sum_{i=1}^{k_k} P_{O_i} T_{\bar{b}i\bar{o}c}; \quad (15)$$

$$T_b = \sum_{i=1}^{k_k} P_{O_i} \left[(P_{B_p^i} + P_{B_c^i}) T_{\bar{b}i\bar{o}c} + P_{B_o^i} T_{\bar{b}i\bar{o}} \right]. \quad (16)$$

С учетом теоремы гипотез (формулы Байеса) [3] получили:

$$T_{\bar{b}o\bar{c}} = \frac{\sum_{i=1}^{k_k} P_{O_i} P_{B_p^i} T_{\bar{b}i\bar{o}c}}{\sum_{i=1}^{k_k} P_{O_i} P_{B_p^i}}; \quad (17)$$

$$T_{\text{BC}} = \frac{\sum_{i=1}^{k_k} P_{O_i} P_{B_c^i} T_{\text{Bic}}}{\sum_{i=1}^{k_k} P_{O_i} P_{B_c^i}} = \frac{\sum_{i=1}^{k_k} P_{O_i} P_{B_c^i} T_{\text{BIOC}}}{\sum_{i=1}^{k_k} P_{O_i} P_{B_c^i}}; \quad (18)$$

$$T_{\text{BO}} = \frac{\sum_{i=1}^{k_k} P_{O_i} P_{B_o^i} T_{\text{BIO}}}{\sum_{i=1}^{k_k} P_{O_i} P_{B_o^i}}; \quad (19)$$

$$T_{\text{Bi}} = (P_{B_p^i} + P_{B_c^i}) T_{\text{BIOC}} + P_{B_o^i} T_{\text{BIO}}. \quad (20)$$

Для варианта **в** возможны два способа доступа к общим атрибутам. Оценка T_{BO} при выполнении запроса к таблицам категорий должна выполняться по формуле (17), к таблице супертипа – по формуле (19):

$$T_{\Gamma} = \sum_{i=1}^{i=k_k} P_{O_i} (P_{B_p^i} T_{\text{BIOC}} + P_{B_c^i} T_{\text{GIC}} + P_{B_o^i} T_{\text{GIO}}); \quad (21)$$

$$T_{\text{ГOC}} = \frac{\sum_{i=1}^{k_k} P_{O_i} P_{B_p^i} T_{\text{GIOC}}}{\sum_{i=1}^{k_k} P_{O_i} P_{B_p^i}}; \quad (22)$$

$$T_{\text{ГC}} = \frac{\sum_{i=1}^{k_k} P_{O_i} P_{B_c^i} T_{\text{GIC}}}{\sum_{i=1}^{k_k} P_{O_i} P_{B_c^i}}; \quad (23)$$

$$T_{\text{ГО}} = \frac{\sum_{i=1}^{k_k} P_{O_i} P_{B_o^i} T_{\text{GIO}}}{\sum_{i=1}^{k_k} P_{O_i} P_{B_o^i}}; \quad (24)$$

$$T_{\text{Гi}} = P_{B_p^i} T_{\text{GIOC}} + P_{B_c^i} T_{\text{GIC}} + P_{B_o^i} T_{\text{GIO}}. \quad (25)$$

Формулы (7)–(25) составляют совокупность аналитических моделей для оценки различных аспектов производительности фрагментов баз данных с категоризованными отношениями.

7. Анализ результатов моделирования

Зависимости, полученные с помощью аналитических моделей, позволяют выполнить оценку времени доступа к данным. Одна из зависимостей

приведена на рис. 5. Варианты организации могут отличаться по времени доступа более чем в 5 раз.

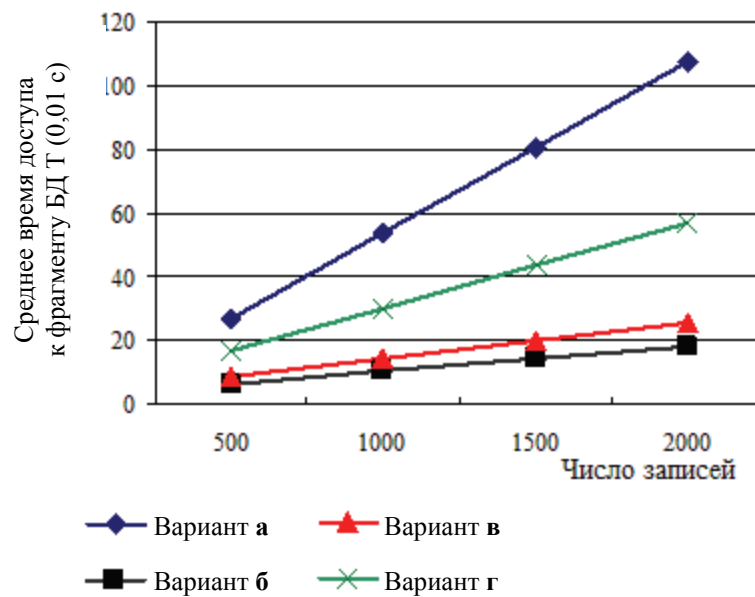


Рис. 5. Зависимости среднего времени доступа от числа записей фрагментов с различными вариантами организации (равномерное распределение вероятностей типа обращений и числа записей по категориям; $k_{oa} = 2$; $k_{ia} = 3$; $k_k = 4$)

Заключение

Оценки времени доступа к фрагментам баз с категоризованными данными показывают их зависимость от большого числа факторов:

- параметров процессов выборки данных (распределения вероятностей обращений к категориям данных и типов запросов);
- структуры категоризованных данных (числа общих и частных атрибутов, числа категорий и т.д.);
- вариантов организации категоризованных данных;
- принимаемых мер по повышению производительности БД (например, индексирования записей).

Существенный разброс времени доступа к категоризованным данным для различных вариантов их представления подтверждает актуальность разработки моделей и необходимость детального обоснования выбора варианта.

Список литературы

1. Государственная программа Российской Федерации «Информационное общество (2011–2020 годы)», утверждена распоряжением Правительства Российской Федерации от 20 октября 2010 г. № 1815-р.
2. CDM – метод разработки информационных систем фирмы Oracle // Oracle Magazine. Russian Edition. – 1997. – № 2. – С. 31–49.
3. **Вентцель, Е. С.** Теория вероятностей / Е. С. Вентцель. – М.: Гос. изд-во физ.-мат. лит-ры, 1962. – 576 с.
4. **Феллер, В.** Введение в теорию вероятностей и ее приложения : в 2 т. ; пер. с англ. / В. Феллер. – М.: Мир, 1984. – Т. 2. – 738 с.

Бистерфельд Ольга Александровна

кандидат технических наук, доцент,
кафедра информатики и вычислительной
техники, Рязанский государственный
радиотехнический университет

E-mail: bist19@yandex.ru

Bisterfeld Olga Alexandrovna

Candidate of engineering sciences, associate
professor, sub-department of informatics
and computer engineering, Ryazan State
Radio Engineering University

УДК 004.652

Бистерфельд, О. А.

Аналитико-имитационный метод оценки временных характеристик для различных вариантов представления отношений категоризации в реляционных базах данных / О. А. Бистерфельд // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 95–105.

ЧЕЛОВЕКО-МАШИННОЕ ВЗАИМОДЕЙСТВИЕ В СИСТЕМЕ МОНИТОРИНГА И ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В ОБЛАСТИ ТОРГОВЛИ КРЕДИТНЫМИ ДЕРИВАТИВАМИ

Аннотация. Представлено описание решений, моделей и методов для создания инструментов доступа к многомерным данным в области торговли кредитными деривативами и последующего их анализа для людей, не являющихся техническими специалистами в области баз данных, благодаря интуитивно понятным пользовательским интерфейсам и алгоритмам, автоматизирующим низкоуровневые операции с данными, позволяя абстрагироваться от них.

Ключевые слова: мониторинг, система поддержки принятия решений, пользовательский интерфейс, хеджирование, кредитный дериватив.

Abstract. The article describes solutions, models and algorithms providing access and analysis of multidimensional data structures in the scope of credit derivatives trading, intended for people without any technical skills in databases. The concept is based on implementation of intuitive measuring interfaces and algorithms, automating low-level data operations and allowing to abstract away from them.

Key words: monitoring, decision support system, user interface, hedging, credit derivative.

Введение

Кредитные операции – самая доходная статья банковского бизнеса. Управление рисками при торговле производными кредитными продуктами позволяет банку защититься от неблагоприятных изменений цен на рынке акций, процентных ставок и т.п. Для быстрого и верного принятия управленческого решения требуется проанализировать достаточный объем информации, адаптированной для понимания, т.е. правильно обработанной специализированной системой мониторинга и поддержки принятия решений [1]. Для некоторых областей, например бюджетирования и планирования, имеется довольно много хорошо проработанных программных решений от таких производителей, как SAP, Oracle, PeopleSoft [2, 3]. Однако все еще есть области, в которых автоматизация процесса анализа и принятия управленческих решений требует усовершенствований. Одной из таких областей и является управление рисками при торговле производными кредитными продуктами – кредитными деривативами.

Кредитные деривативы позволяют отделить кредитный риск от всех других рисков, присущих конкретному инструменту, и перенести такой риск от продавца риска (приобретателя кредитной защиты) к покупателю риска (продавцу кредитной защиты). Основной набор таких инструментов – это кредитные дефолтные свопы (CDS), опционы, фьючерсы и т.д.

Обычно целью покупки дериватива является не получение базового актива, а хеджирование (страхование, снижение риска от потерь) ценового или валютного риска во времени или получение спекулятивной прибыли от изменения цены базового актива. Операции на рынке осуществляются биржевыми торговцами – трейдерами, функции которых заключаются в анализе текущей ситуации на рынке и заключении торговых сделок.

В данной работе приводится описание разрабатываемой специализированной системы анализа рыночных сделок (CAPC), которая позволяет пользователям, не имеющим достаточной квалификации, выполнять операции с большими объемами многомерных данных и использовать инструментарий интеллектуального анализа данных. Система обеспечивает поддержку принятия решений, наглядность, гибкость и высокую производительность работы специалистам в области биржевого трейдинга без участия технических специалистов.

1. Структура распределенной системы анализа трейдерских сделок

Система представляет собой Интернет портал, посредством которого приложения, выполняемые на стороне клиента, организуют взаимодействие с серверной частью. Функциональная (структурная) схема системы на высоком уровне представлена на рис. 1.

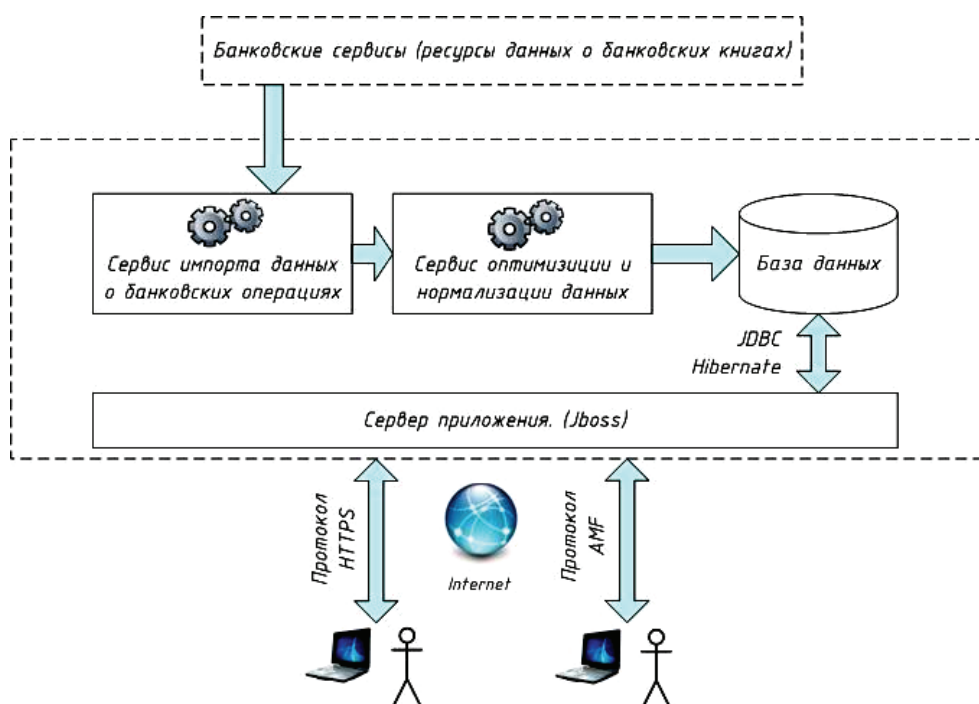


Рис. 1. Структурная схема компонентов CAPC

Зарегистрированный трейдер работает с порталом через клиентскую прикладную программу, пользовательский интерфейс которой включает ряд инструментов (описываемых ниже) для удобной и наглядной работы с данными.

Серверная часть, получая запросы от программ-клиентов, производит необходимые операции с данными и передает требуемую информацию. Серверная логика разработана на платформе Java Enterprise Edition (язык программирования Java) с использованием технологий Spring и ORM Hibernate. Hibernate – библиотека для языка программирования Java, предназначенная для решения задач объектно-реляционного проецирования (object-relational mapping – ORM). Hibernate решает задачу связи классов Java с таблицами ба-

зы данных, осуществляет автоматическое обновление структуры таблиц, построение запросов и обработки полученных данных [4].

В качестве серверной платформы был выбран JBoss Application server – Java Enterprise Edition сервер приложений с открытым исходным кодом, разработанный одноименной компанией. Достаточно хорошая реализация делает его конкурентом для аналогичных проприетарных программных решений, таких как IBM WebSphere или Oracle WebLogic.

Клиентское приложение обменивается информацией с сервером по протоколам HTTPS (шифрованный HTTP) и AMF (Adobe Media Format). AMF – это специальный протокол, используемый для обмена данными и вызова удаленных серверных процедур с клиентского приложения, разработанного на платформа Adobe Flex [5]. Flex является высокоэффективной средой разработки с открытым кодом для создания динамичных мобильных, настольных и веб-приложений. Среда Flex позволяет разрабатывать мобильные и веб-приложения с использованием базы типовых кодов, что значительно сокращает время разработки приложений.

Данный подход позволяет без лишних затрат вычислительных ресурсов обрабатывать большие объемы данных, используя многофункциональный и «дружественный» пользовательский интерфейс.

2. Процесс обработки информации

Понятие «ситуация на рынке» выражается совокупностью данных из различных источников информации, таких как курсы валют, котировки акций на фондовых рынках, открытые позиции других трейдеров и т.д. За временную единицу мониторинга обычно принимается один банковский день. Сделки представляют собой покупку/продажу дериватива определенного типа (облигаций, кредитных дефолтных свопов и т.п.). В процессе сделок в зависимости от типа продукта требуется вести мониторинг изменений около 300 свойств, некоторые из которых представлены сложными структурами данных. В конце каждого рабочего банковского дня на корпоративных серверах банков появляются официальные данные обо всех сделках, произведенных в этот день. Сервис импорта собирает предоставленную всеми банками-партнерами информацию с их серверов, обрабатывает, сортирует и передает во внутреннее временное хранилище. Процесс обычно проходит в течение нескольких часов в ночное время, так как разные банки делают нужную информацию доступной на своих серверах в разное время. В некоторых банках информация предварительно проверяется специальными людьми.

Процесс уменьшения избыточности информации в базе данных называется нормализацией. Приведение данных в нормализованный вид позволяет: обеспечить быстрый доступ к данным в таблицах, исключить ненужное повторение данных, которое может являться причиной ошибок при вводе и нерационального использования памяти сервера [6].

Утром до начала рабочего дня запускается сервис оптимизации, который преобразует собранную предыдущим сервисом денормализованную информацию из временного хранилища в структурированный нормализованный вид и пополняет внутреннюю базу данных. Таким образом, к началу рабочего дня внутренняя база данных системы содержит проиндексированную информацию обо всех торговых сделках за предыдущий интервал времени, например за рабочий день, в оптимизированном и готовом к работе виде.

Рассмотрим основные инструменты разработанной системы, позволяющие трейдерам принимать решения о сделках, анализируя наиболее важную информацию, выбранную из огромного массива данных и представленную в удобном для восприятия и анализа виде. Приложение состоит из нескольких модулей: модуль визуализации отчета о совершенных сделках, модуль создания и редактирования режимов агрегации, модуль определения подмножества данных.

3. Модуль визуализации отчета о совершенных сделках

Данный модуль является ключевым для трейдера, он позволяет пользователю просматривать информацию о нужных сделках в структурированном и удобном для восприятия виде, в виде гипертаблицы. Гипертаблица представляет собой новый нестандартный элемент пользовательского интерфейса. Сочетает в себе функциональность таблицы с древовидной структурой и включает элементы управления для просмотра изменений значений, отображаемых в ячейках, во времени. Гипертаблица является способом наглядной визуализации гиперкуба данных, в котором данные сгруппированы по заданным параметрам и уровням агрегации. Она предоставляет возможность ориентироваться в многоуровневой и многомерной древовидной структуре данных в реальном времени. Пример фрагмента таблицы представлен на рис. 2.

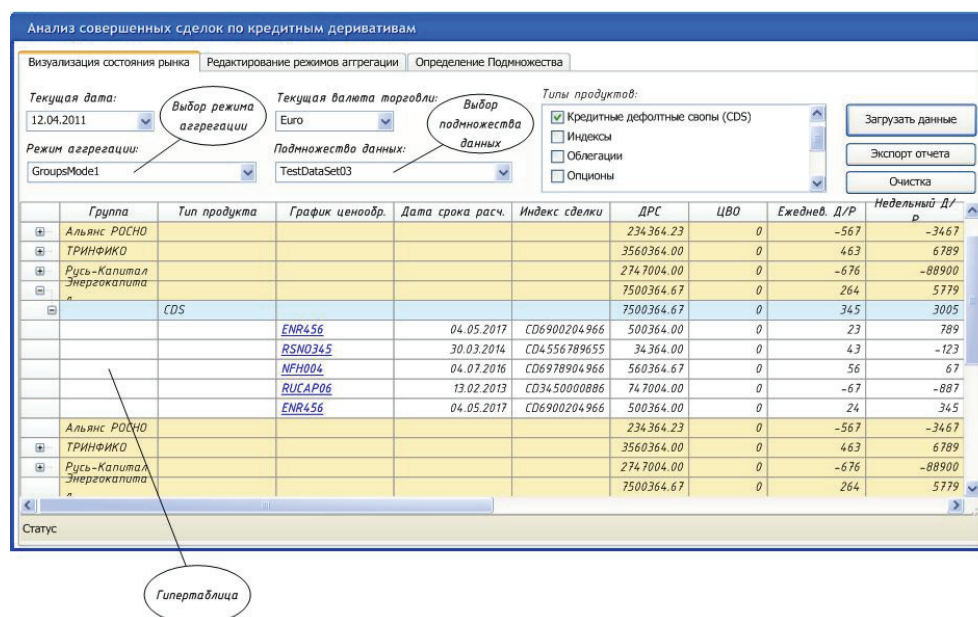


Рис. 2. Экран пользовательского интерфейса модуля визуализации отчета о совершенных сделках

Отличительной особенностью гипертаблицы является то, что число строк в таблице – величина не статичная, сами строки по характеру и функциональности неравнозначны, так как некоторые из них являются группирующими. Группирующие или агрегационные строки являются узловыми и показывают суммарную информацию по соответствующим колонкам. В свою

Гипертаблица позволяет в реальном времени посмотреть график изменения любой выбранной величины за нужный период времени (рис. 3). Перемещая ползунок на временной шкале, можно видеть изменения значений всех остальных ячеек таблицы, тем самым анализируя интересные области согласно графику.



110

4. Модуль создания и редактирования режимов агрегации

Для того чтобы добавить новый режим агрегации, которым в дальнейшем будет определяться вид и содержание таблицы данных о биржевых сделках, пользователю необходимо воспользоваться инструментами, предоставленными этим модулем, определить набор параметров режима агрегации и сохранить его под определенным названием, чтобы в дальнейшем можно было воспользоваться им для отображения гипертаблицы (рис. 4).

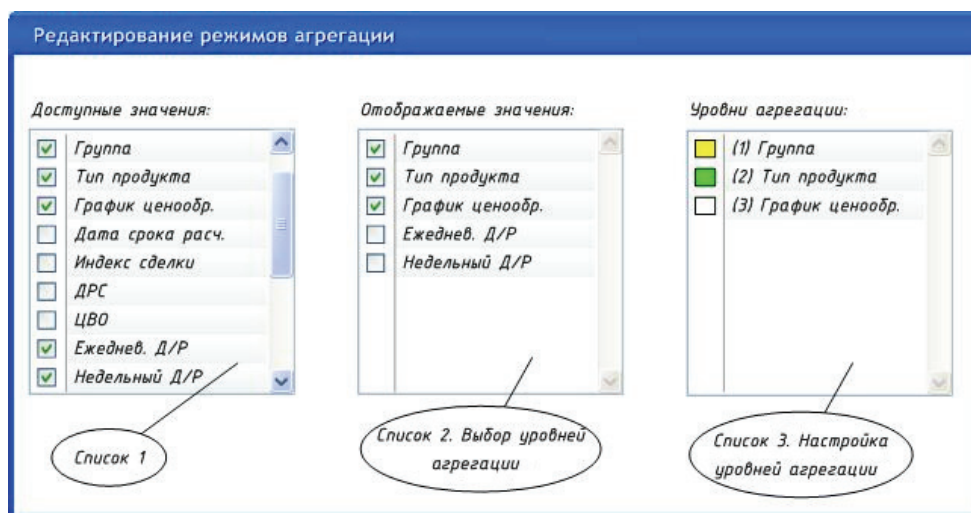


Рис. 4. Экран пользовательского интерфейса модуля создания и редактирования режима агрегации

При конфигурировании режима отображения данных пользователю в начале работы следует определить колонки значений, которые будут показаны в гипертаблице. В первом списке имеются все доступные для добавления колонки. Этот список определяется характеристиками, хранящимися в базе данных по каждой торговой сделке из банковских книг, а также значениями, вычисляемыми на основе этих характеристик. Во втором списке отображаются выбранные пользователем колонки, значения которых пользователь счел нужными для просмотра и анализа. Именно эти значения будут отображаться при просмотре гипертаблицы, и именно по ним будет производиться группирование записей.

Полезной является возможность многоуровневой группировки значений и возможность дальнейшего отображения древовидной структуры этих данных с помощью гипертаблицы.

Для задания уровня агрегации необходимо выбрать колонку значений (отметив ее во втором списке), по которой будут группироваться строки. Эти строки могут содержать подгруппы, сгруппированные по другим, нужным пользователю критериям, если будет добавлен следующий уровень агрегации, который в свою очередь может также содержать подгруппы, если будет добавлен еще один, низлежащий уровень агрегации и т.д.

Все вложенные уровни в гипертаблице отображаются в виде дополнительных группирующих строк, содержащих суммарные значения соответ-

ствующих колонок строк нижних уровней агрегации. Для каждого уровня агрегации имеется возможность задать цветовой индикатор для еще большей наглядности и облегчения работы с таблицей. Строки таблицы, соответствующие определенному уровню, будут окрашены в заданные цвета.

Все эти инструменты дают дополнительные возможности и удобство просмотра и анализа информации. При этом пользователь, не являясь специалистом в области баз данных и информационных технологий, может без особого труда выполнять все необходимые действия с большими массивами данными, оперируя понятиями только той области, специалистом которой он является.

5. Модуль определения подмножества данных

При всем многообразии доступной для просмотра информации необходимым инструментом, позволяющим пользователю выбирать именно те данные, которые необходимы для анализа конкретного случая, чтобы не переводить внимание на нерелевантные в данный момент записи.

Условия, по которым определяется подмножество данных, интересующих пользователя, могут быть достаточно сложными. В реляционных базах данных для этих целей используют многоуровневые запросы и срезы данных (это фильтрующие элементы управления, которые ограничивают данные, отображаемые в сводных таблицах и сводных диаграммах).

Данный модуль дает возможность пользователям с экономическим образованием, не являющимся специалистами в области баз данных и информационных технологий, создавать запросы на выбор данных, оперируя только понятиями предметной области – биржевого трейдинга.

Модуль представляет собой конструктор, позволяющий задавать неограниченное количество гибких запросов выборки данных с разными условиями и критериями. Форма пользовательского интерфейса модуля (рис. 5) состоит из конструктора условий выборки данных, наглядной интерпретации результата и выбора временного диапазона.

Конструктор условий выборки данных содержит компоненты, определяющие условия, по которым во время загрузки данных будут проверяться все строки заданного временного диапазона на предмет выборки для анализа или исключения из рассмотрения. Компонент пользовательского интерфейса, определяющий каждое условие, включает: флаг «NOT» для возможности отрицания условия; поля названия колонки, значение которой будет проверяться; оператор проверки условия (доступны все операторы сравнения и некоторые дополнительные, например «В списке»); поле со значением, с которым будет сравниваться значение колонки при помощи оператора сравнения.

Значение колонки может быть непосредственным, введенным в текстовое поле, либо ссылкой на значение другой колонки. Дополнительно имеются кнопки удаления условия – «×», добавление нового условия – «+», и группировки условий аналогично взятию выражения в скобки – кнопка «()». При объединении условий появляется компонент для установки логического оператора «И», «ИЛИ», «НЕ ИЛИ».

Составленное условие выборки вместе с условием выбора временного диапазона передается с клиента на сервер в виде сериализованной (кодиро-

ванной последовательности байт) структуры объектов Action Script 3 по протоколу AMF и затем конвертируется в Java-объекты. На сервере по структуре объектов строится сложный запрос на HQL – Hibernate Query Language (язык запросов библиотеки Hibernate). HQL-запрос является более универсальным, в отличие от SQL-запроса, не зависит от используемой базы данных и оперирует объектами, а не реляционными сущностями. Затем с использованием библиотеки Hibernate и при помощи драйвера базы данных и драйвера SQL-диалекта генерируется многоуровневый запрос к базе данных на совместимом с ней языке. Такая абстрагированность позволяет скрыть от пользователя все низкоуровневые операции с данными и работать только гибким и понятным конструктором.

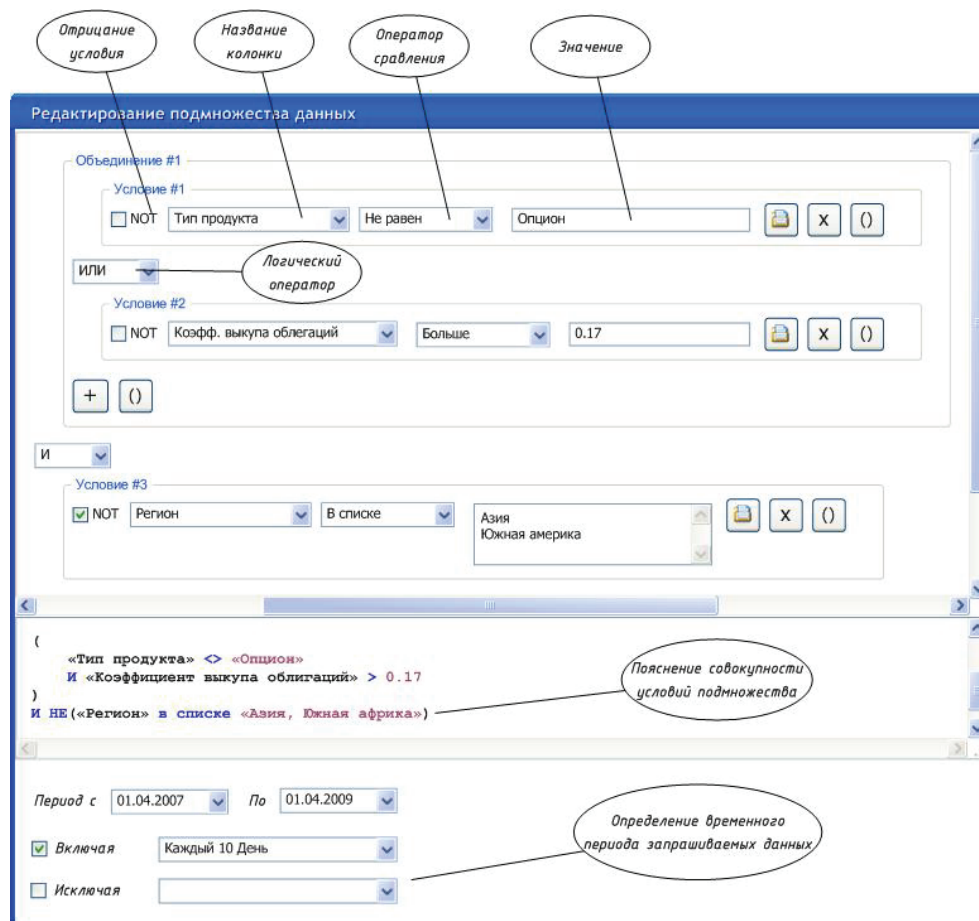


Рис. 5. Экран пользовательского интерфейса модуля определения подмножества данных

Заключение

Разрабатываемая система позиционируется как наиболее специализированная для области торговли кредитными деривативами. Она предоставляет инструменты, облегчающие доступ к данным и последующий анализ и ориентирована на самих трейдеров, давая им возможность работать с системой без помощи дополнительных технических специалистов, благодаря ин-

туитивно понятным «дружественным» пользовательским интерфейсам и алгоритмам, автоматизирующим низкоуровневые операции с данными, позволяя абстрагироваться от них.

В системе были разработаны и кратко описаны в статье: модуль визуализации отчета о совершенных сделках, модуль создания и редактирования режимов агрегации, модуль определения подмножества данных. Далее планируется разработка модуля конструирования сценариев поведения рынка и серверного приложения симуляции поведения рынка по заданным сценариям. Эти инструменты позволят трейдерам проводить эксперименты, проверять собственные идеи и гипотезы без каких-либо материальных затрат. Система представляет собой веб-портал и может использоваться как частными трейдерами-одиночками, так и профессиональными аналитиками, работающими в интересах крупных коммерческих банков.

Список литературы

1. **Шишкин, А. И.** Сущность, задачи и принципы мониторинга / А. И. Шишкин // Экономический рост в регионах России : материалы рабочего совещания 9 февраля 2003 г. – URL: <http://www.aspe.spb.ru/Workshop/Shishkin.pdf>
2. Information Technology for Management, Transforming Organizations in the Digital Economy / Turban et al. – Massachusetts : John Wiley & Sons, Inc., 2008. – P. 300–343.
3. **Лаврушин, О. И.** Банковское дело: современная система кредитования : учебное пособие / О. И. Лаврушин, О. Н. Афанасьева, С. Л. Корниенко ; под ред. засл. деят. науки РФ, д-ра экон. наук, проф. О. И. Лаврушина. – 3-е изд., доп. – М. : КНОРУС, 2007. – 264 с.
4. **Christian Bauer.** Hibernate in Action / Christian Bauer, Gavin King Manning. – Publications Co, 2004.
5. Al Hilwa, Randy Perry Adobe Flex in the Enterprise: The Case for More Usable Software Copyright 2010 IDC. – URL: <http://www.adobe.com/enterprise/pdf/idc.pdf>.
6. **Когаловский, М. Р.** Энциклопедия технологий баз данных / М. Р. Когаловский. – М. : Финансы и статистика, 2002. – 800 с.

Финогеев Алексей Германович

доктор технических наук, профессор,
кафедра систем автоматизированного
проектирования, Пензенский
государственный университет

E-mail: bam@pnzgu.ru

Finogeev Aleksey Germanovich

Doctor of engineering sciences, professor,
sub-department of CAD systems,
Penza State University

Овечкин Роман Михайлович

аспирант, Пензенский
государственный университет

E-mail: bam@pnzgu.ru

Ovechkin Roman Mikhaylovich

Postgraduate student,
Penza State University

УДК 004.5

Финогеев, А. Г.

Человеко-машинное взаимодействие в системе мониторинга и поддержки принятия решений в области торговли кредитными деривативами / А. Г. Финогеев, Р. М. Овечкин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 106–115.

СИСТЕМНЫЙ АНАЛИЗ МЕТОДОВ ПРОЕКТИРОВАНИЯ АВТОМОБИЛЬНЫХ ДОРОГ¹

Аннотация. Рассмотрены основные методы автоматизированного проектирования, используемые в САПР автомобильных дорог на всех этапах разработки проекта. Проанализированы основные функциональные возможности ведущих САПР автомобильных дорог, выявлены их достоинства и недостатки.

Ключевые слова: автомобильные дороги, системный анализ, системы автоматизированного проектирования.

Abstract. The article considers basic methods used at all stages of computer-aided design of roads. The authors analyze key features of the top CAD-systems and reveal their benefits and weaknesses.

Key words: automobile roads, systems analysis, computer-aided design.

Введение

Сеть автомобильных дорог – важнейший элемент экономики и наиболее крупная составляющая транспортной инфраструктуры любой страны. Поддержание высоких эксплуатационных характеристик и устойчивое развитие дорожной сети – необходимые условия экономического роста, обеспечения целостности и национальной безопасности государства, повышения уровня и улучшения условий жизни населения. К современным автомобильным дорогам предъявляются высокие требования по осуществлению бесперебойной работы автомобильного транспорта, обеспечению безопасности движения и архитектурно-эстетические требования. Выполнение их может быть обеспечено только при использовании системного подхода на всех этапах проектирования, эксплуатации и содержания автомобильных дорог.

На сегодня системный подход при проектировании автомобильных дорог реализуется через применение:

- современных технологий и методов производства изысканий, основанных на использовании высокопроизводительных методов сбора информации о местности (использование ГИС-технологий);
- современных систем автоматизированного проектирования автомобильных дорог, предоставляющих проектировщику инструменты для решения задач по формированию цифровых моделей местности зоны проектирования, трассированию проектируемой автомобильной дороги, проектированию продольного профиля дороги, проектированию поперечных профилей и дорожных одежд, проектированию искусственных сооружений и инженерно-сервисного обустройства дороги, оценке проектных решений.

Место геоинформационных систем при проектировании автомобильных дорог

Широкое распространение и развитие геоинформационных систем (ГИС) в последние годы позволило интегрировать их в процесс автоматизированного проектирования автомобильных дорог.

¹ Работа проводилась при финансовой поддержке Министерства образования и науки РФ, номер государственной регистрации № 16.513.11.3128 от 12.10.2011 г.

Из-за особенностей ГИС при проектировании автомобильных дорог на них возлагается ряд задач, несвойственных системам автоматизированного проектирования дорог.

Например, в ГИС, в отличие от САПР, используется небольшое количество графических примитивов: точки, линии, полигоны, поверхности и растры. Данная особенность подобных систем позволяет строго определить такие пространственные операции, как поиск объектов в заданном регионе, поиск смежных или пересекаемых объектов, построение объединений, пересечений и разностей полигонов, построение буферных зон.

В ГИС широко применяются алгоритмические методы для хранения больших объемов данных, быстрого поиска объектов, упрощения данных для быстрого вывода на экран, что позволяет применять ГИС для представления сети дорог на электронных мелкомасштабных картах, анализа транспортного обеспечения районов, получения оперативной информации об объектах дорожной сети, выбора наиболее приемлемого коридора варьирования проектируемой трассы для соответствующей цифровой модели местности (ЦММ).

В ГИС графические объекты, имеющие одинаковый тип, представляются в одном слое графических данных и имеют одинаковый набор атрибутов. Данная особенность ГИС позволяет представить слой графических данных с наборами атрибутов объектов в виде таблицы реляционной базы данных, а следовательно, использовать соответствующий аппарат баз данных для анализа атрибутов графических объектов. Наличие атрибутивной поддержки в ГИС позволяет применять их при решении задач диагностики, паспортизации, инвентаризации, кадастра дорог.

Перечисленные особенности ГИС позволяют интегрировать их с САПР дорог, такие системы находят в последнее время все большее распространение.

Обзор методов автоматизированного проектирования, используемых в САПР автомобильных дорог

Развитие вычислительной техники позволило применять при проектировании автомобильных дорог различные методы оптимизации и моделирования, что полностью видоизменило процесс проектирования. Рассмотрим наиболее распространенные методы, применяемые в САПР автодорог.

Методы автоматизированного проектирования плана дороги

Традиционным принципом трассирования дорог, на основе которого строятся современные методы, используемые в большинстве САПР автодорог, является принцип тангенциального трассирования. Данный принцип заключается в назначении тангенциального хода, в изломы которого вписывают закругления с требуемыми параметрами, рассчитываемыми по определенным алгоритмам.

Принцип тангенциального трассирования является базовым для методов «однозначно определенной оси». При тангенциальном трассировании трасса характеризуется ломаной линией с вписанными кривыми в ее изломы так, что отрезки ломаной представляют собой касательные к кривым, поэтому ломаную называют тангенциальным ходом. Наиболее применимы для проектирования закруглений трассы в составе тангенциального хода кривые Безье, которые являются пространственными функциями и способны обеспечивать трехмерное трассирование автомобильных дорог.

Разновидностью «методов однозначно определенной оси» является метод «опорных элементов», который состоит в том, что при помощи шаблонов кривых и линейки устанавливают положение опорных элементов, оптимально аппроксимирующих эскизную трассу дороги с последующей аналитической их увязкой.

Одним из альтернативных методов трассирования дорог является метод «сглаживания эскизной линии трассы», который заключается в аналитической аппроксимации полиномами высоких степеней массива точек эскизного варианта трассы. В методе «сглаживания эскизной линии трассы» задается избыточное количество точек эскизной линии, и их рассматривают как «приближенные», вблизи которых должна пройти трасса, так как в общем случае через все заданные точки не может быть проведена удовлетворительная трасса.

Наиболее подходящими функциями для использования в методе «сглаживания эскизной линии трассы» являются сплайны, являющиеся универсальным математическим аппаратом для описания, хранения, преобразования, анализа и геометрического представления трасс проектируемых автомобильных дорог.

Таким образом, при пространственном трассировании дорог используются как сплайны, так и кривые Безье, которые позволяют осуществлять математически корректную запись трассы автомобильной дороги в пространстве [1].

Методы автоматизированного проектирования продольного профиля автомобильных дорог

Автоматизированное проектирование земляного полотна является составной частью задачи построения трассы в пространстве.

При проектировании земляного полотна учитывают множество факторов: категорий проектируемой дороги, рельеф местности, данные геологических и гидрогеологических изысканий.

При проектировании продольного профиля используется два подхода: проработка множества вариантов метода «опорных точек» или использование одного из методов оптимизации.

В методе «опорных точек» расчет проектируемой линии продольного профиля осуществляется по заданным опорным точкам и радиусам вертикальных кривых в классе функций вертикальных выпуклых и вогнутых параболических кривых. Данный метод является наиболее применимым при проектировании продольного профиля автомобильных дорог в равнинной местности и в стесненных условиях (при большом количестве опорных точек).

Одним из методов оптимизации является метод проекции градиента. Метод состоит из трех этапов. На первом этапе проектируемая линия представляется в виде «цепочечной линии» с узлами, совпадающими с переломными точками черного профиля земли и удовлетворяющей всем техническим условиям и ограничениям. Поиск оптимального решения осуществляется итерационным методом проекции градиента, на каждом шаге которого получается новая проектная линия с меньшим значением целевой функции, чем на предыдущем шаге.

На втором этапе цепочечная линия аппроксимируется последовательностью традиционных элементов: квадратных парабол и прямых. Параметры

парабол определяются таким образом, чтобы не нарушались ограничения по уклону и кривизне во всех точках предполагаемого элемента, а также выполнялись граничные условия в точках примыкания и отмыкания.

На третьем этапе оптимизируются коэффициенты кусочно-параболических кривых проектной линии по строительной стоимости.

Для решения поставленной задачи в методе граничных итераций определен класс функций, в котором отыскивается оптимальное решение, – это ломаная линия со строительным шагом, принимаемым обычно равным 20 м. Ломаная в этом случае представляет собой дискретный аналог кривой переменного радиуса. Проектирование линии продольного профиля в классе ломаных чаще всего применимо при ремонте и реконструкции автомобильных дорог, когда математическая модель проектной линии в виде последовательности прямых и квадратных парабол оказывается недостаточно гибкой для получения приемлемого проектного решения.

Для возможности автоматизированного проектирования в классе ломаных функций описание проектной линии необходимо математически формализовать. Наиболее просто и эффективно эта задача формализуется в терминах математического аппарата сплайн-функций.

Методы проектирования продольного профиля, основанные на применении сплайнов высоких степеней (например, кубических), используют для описания проектной линии как интерполяционные, так и сглаживающие сплайны.

Сглаживающие сплайны необходимы для поиска оптимального положения проектной линии в заданной полосе варьирования. А интерполяционные сплайны позволяют обеспечить процесс корректировки проектной линии в интерактивном режиме работы инженера-проектировщика при эвристическом проектировании [1].

Автоматизированное проектирование оптимальных нежестких дорожных одежд

Расчет нежестких дорожных одежд при кратковременном действии нагрузки выполняется по критериям прочности: упругому прогибу, сопротивлению сдвигу в грунте и в слабосвязных слоях одежды, растяжению при изгибе слоев одежды из грунтов и каменных материалов.

Расчет дорожных одежд нежесткого типа предусматривает, как правило, расчетную схему, включающую в себя несколько конструктивных слоев с учетом подстилающего грунта. Слои конструкции располагаются по убыванию (сверху вниз) модулей упругости материалов. Толщина отдельных слоев может быть задана по конструктивным соображениям, и эти слои в процессе поиска оптимального решения остаются неизменными.

Оптимальное решение ищется в диапазоне между минимальным и максимальным значением толщины каждого конструктивного слоя при определенном шаге перебора толщин. Минимальная толщина задается не менее значений регламентированных ОДН 218.046-01, а шаг перебора задается из условия производства и технологии выполнения работ по устройству каждого конструктивного слоя, точности устройства каждого слоя, имеющегося в наличии оборудования, возможности обеспечения требуемого уплотнения каждого слоя и требуемой точности при приемке работ [1].

Системы автоматизированного проектирования автомобильных дорог

Существующие в настоящее время САПР автомобильных дорог имеют идентичные структуры. Они состоят из различных модулей (подсистем, технологических линий проектирования, пакетов прикладных программ), отвечающих за выполнение определенных проектных операций: переработку исходной информации и формирование цифровой модели местности, трассирование автомобильных дорог, проектирование продольного профиля, проектирование земляного полотна и дорожных одежд, проектирование искусственных сооружений, оценку проектных решений, подготовку проектно-сметной документации.

На сегодня наибольшее распространение получили следующие системы автоматизированного проектирования автомобильных дорог ведущих отечественных и зарубежных разработчиков: IndorCAD/Road (разработчик – Indorsoft, Россия), ROBUR (разработчик – Topomatic, Россия), GIP (разработчик – ГИПРОДОРНИИ, Россия), AutoCAD Civil 3D (разработчик – Autodesk, США), CARD/1 (разработчик – IB&T Group, Германия), MXROAD (разработчик – Bentley Systems, США), Plateia (разработчик – CGS plus d.o.o., Словения), CREDO ДОРОГИ (разработчик – Кредо-Диалог, Беларусь).

Перечисленные программные комплексы предназначены для проектирования строительства, реконструкции и ремонта автомобильных дорог. Исходными данными для этих систем являются данные геодезических изысканий и карты местности. Выходными данными являются трехмерная модель дороги и проектная документация.

Данные программные продукты позволяют реализовать комплексный подход к разработке проектов на всех этапах создания и эксплуатации автомобильных дорог, предложить несколько вариантов проектных решений и выбрать наиболее приемлемый вариант в рамках одного проекта, наглядно оценить преимущества и недостатки различных вариантов проектных решений.

Работа в системе IndorCAD/Road [2] начинается с подготовки цифровой модели местности. Программа предоставляет полный набор инструментов для ее создания. Возможна как обработка данных инженерно-геодезических изысканий с помощью специального модуля, так и импорт уже обработанных данных через обменные форматы файлов. Программа поддерживает форматы различных геодезических приборов, файлы GPS-съемки, обменный формат CREDO, множество растровых форматов, DXF-файлы, шейп-файлы ESRI, текстовые файлы с координатами точек и пр. К проекту возможно подключение растровых картографических изображений с точной координатной привязкой. На основе геодезических данных формируется ЦММ. В системе IndorCAD/Road имеется возможность ее ручного редактирования.

Трассирование автомобильных дорог в плане может выполняться различными геометрическими элементами прямыми, дугами, клотоидами, кривыми Безье и сплайнами.

Проектирование продольного профиля можно выполнять на основе тангенциального хода или сплайновым методом.

При проектировании верха земляного полотна возможно автоматическое создание геометрии проезжей части, обочин, разделительных полос, бордюров.

Для формирования конструкции поперечного профиля и дорожной одежды в IndorCAD/Road используются типовые поперечные профили и конструкции дорожной одежды, возможно создание собственных решений произвольной сложности. После проектирования рассчитываются объемы земляных работ, дорожной одежды, площади откосов.

После выполнения основных шагов проектирования в IndorCAD/Road формируется комплект чертежей и ведомостей в формате системы подготовки чертежей в соответствии с требованиями ГОСТ Р 21.1701–97 «Правила выполнения рабочей документации автомобильных дорог».

Программный комплекс Топоматик Robur [3] состоит из нескольких модулей.

Топоматик Robur «Геодезия» предназначен для подготовки исходных данных для программ Топоматик Robur «Автомобильные дороги» и Топоматик Robur «Железные дороги», выноса проекта в натуру и контрольно-исполнительных съемок.

Топоматик Robur «Геодезия» обладает следующими функциональными возможностями: чтение данных с приборов, уравнивание теодолитных ходов, расчет тахеометрии, расчет нивелирования, построение цифровой модели рельефа (ЦМР), проложение оси трассы, создание продольного и поперечных профилей по ЦМР, подсчет объемов между поверхностями, оформление планшетов.

Топоматик Robur «Автомобильные дороги» предназначен для проектирования загородных дорог и городских улиц. Помимо традиционных компонентов для работы с планом, профилем и поперечниками, модуль содержит блок задач по выравниванию покрытия, расчет дорожной одежды, оценки проектных решений и визуализации проекта. Программный комплекс Топоматик Robur «Автомобильные дороги» является многофункциональным программным продуктом, обеспечивающим сквозной технологический процесс от обработки данных изысканий до выноса проекта в натуру и его инженерного сопровождения.

Топоматик Robur «Дорожная одежда» предназначен для расчета нежестких и жестких дорожных одежд автомобильных дорог общей сети и городских дорог и улиц. В алгоритм расчета заложены рекомендации следующих нормативных документов:

ОДН 218.046–01 «Проектирование нежестких дорожных одежд».

ОДН 218.1.052–2002 «Оценка прочности нежестких дорожных одежд»; «Методические рекомендации по проектированию жестких дорожных одежд (взамен ВСН 197–91)», утв. расп. Минтранса России № ОС-1066-р от 03.12.2003).

ГОСТ Р 52748–2007 «Дороги автомобильные общего пользования. Нормативные нагрузки, расчетные схемы нагружения и габариты приближения».

ОДМ 218.5.003–2010 «Рекомендации по применению геосинтетических материалов при строительстве и ремонте автомобильных дорог».

ВСН 21–83 «Указания по определению экономической эффективности капитальных вложений в строительство и реконструкцию автомобильных дорог»;

ОДМ 218.5.001–2009 «Методические рекомендации по применению геосеток и плоских георешеток для армирования асфальтобетонных слоев

усовершенствованных видов покрытий при капитальном ремонте и ремонте автомобильных дорог».

В программе реализованы следующие расчеты:

- на прочность по упругому прогибу, по условию сдвигоустойчивости подстилающего грунта и малосвязных конструктивных слоев;
- на сопротивление монолитных слоев усталостному разрушению на растяжение при изгибе;
- на статическую нагрузку;
- на морозоустойчивость;
- дренающего слоя.

В качестве исходных данных для расчета задаются интенсивность движения по видам транспорта на любой год эксплуатации дороги и коэффициент ежегодного прироста интенсивности движения. Требуемый модуль упругости и допускаемое напряжение на изгиб определяются с учетом суммарного количества проходов транспорта по одному следу за срок службы дорожной одежды либо суточной интенсивности в обоих направлениях.

Расчетные характеристики конструктивных слоев дорожной одежды выбираются из стандартной базы данных. Имеется возможность вводить в базу данных новые материалы и задавать их характеристики. Программа позволяет автоматически выполнять перебор толщин конструктивных слоев дорожной одежды в заданных пределах и с заданным шагом. Варианты конструкций, отвечающих условиям прочности, упорядочиваются по критерию стоимости.

САПР GIP [4] обеспечивает процесс создания цифровой модели местности, проектирование планов, продольных и поперечных профилей, дорожной одежды, транспортных развязок в одном уровне. Все проектные решения оцениваются с точки зрения пропускной способности, скорости и безопасности движения, воздействия на окружающую воздушную, водную, акустическую среду, почвы в придорожной полосе. Решаются задачи землеотвода, распределения земляных масс, выравнивания покрытия, назначения ограждений. Отличительной особенностью данного комплекса является возможность экологической оценки проектных решений.

AutoCAD Civil 3D [5, 6] – программа, базирующаяся на платформе AutoCAD.

AutoCAD Civil 3D соответствует требованиям следующих нормативных документов: СНиП 2.05.02–85, ГОСТ 21.101–97, ГОСТ 21.204–93, ГОСТ 21.508–93, ГОСТ 21.604–82, ГОСТ 21.610–85, ГОСТ Р 21.1207–97, ГОСТ Р 21.1701–97.

Возможности обработки геодезических данных полностью встроены в AutoCAD Civil 3D. Поддерживается широкий круг задач – импорт полевых журналов, обработка методом наименьших квадратов, задание и преобразование системы координат, редактирование результатов съемки, автоматическое создание фигур съемки и поверхностей.

Участки можно создавать путем преобразования полилиний AutoCAD или с помощью гибких топологических инструментов, автоматизирующих работу. При управлении участками AutoCAD Civil 3D опирается на топологию, поэтому с изменением какого-либо участка автоматически перестраиваются соседние.

Трассы содержат прямые участки, повороты и спиральные переходные кривые. В окончательных чертежах к трассам добавляются метки. При редактировании трассы с помощью графического интерфейса или табличного редактора метки обновляются автоматически. Поддерживается совместное использование трасс и профилей в нескольких чертежах.

Продольные профили извлекаются из множества поверхностей на основе геометрии трассы.

В AutoCAD Civil 3D реализована функция формирования поперечных профилей поверхностей и коридоров. Их можно создавать на отдельных станциях с интервалом вдоль трассы или в отдельных точках. Программа позволяет компоновать целые чертежные листы из групп профилей.

Программный комплекс Plateia [7] включает инструменты и средства для проектирования новых и реконструкции существующих дорог всех технических категорий. Кроме того, существует возможность анализировать рельеф местности, рассчитывать объемы работ, моделировать процессы движения транспортных средств. Программа работает на платформе AutoCAD Civil 3D, AutoCAD Map 3D и AutoCAD.

Программный комплекс Plateia состоит из пяти модулей: «Местность», «Оси», «Продольные профили», «Поперечные сечения» и «Транспорт».

Модуль «Местность» предназначен для работы с координатной геометрией и цифровой моделью местности, в том числе с картами. Модуль позволяет импортировать данные из электронных геодезических приборов, а также из файлов различных форматов, автоматически или интерактивно устанавливать связи между точками (топологию), сохранять атрибутивную информацию точек/связей, осуществлять отбор точек/связей по заданным критериям. Модуль включает набор функций для «классической» геодезии, а также инструменты для специалистов-дорожников.

Модуль «Оси» предназначен для трассировки осевых линий на основе комбинации различных элементов (прямых, круговых кривых, спиральных участков, составных спиралей), а также для оформления планов дорог в соответствии с местными стандартами. Запроектированные в плане продольные и поперечные оси проецируются на трехмерную модель рельефа. Данные записываются в текстовых файлах, которые в дальнейшем используются при проектировании земляного полотна.

Модуль «Продольный профиль» предназначен для построения продольных профилей и вписывания вертикальных кривых. Вписывание кривых осуществляется как в ручном, так и в автоматическом режимах. Модуль позволяет формировать собственные типы таблиц-сеток («шапки») с любым содержанием граф, которые автоматически заполняются при отрисовке соответствующих элементов профиля. На стадии создания продольного профиля производится быстрый расчет приблизительных объемов земляных масс, который позволяет определить более экономичный вариант трассы. При проектировании трассы осуществляется контроль соответствия принятым нормам. Предусмотрены функции для проектирования реконструируемых участков дорог.

Модуль «Поперечные сечения» предлагает инструменты для расстановки поперечных уклонов, длин и отметок, определения точки пересечения выемки и насыпи, подготовки данных для отображения измерений на местности, а также для быстрого изменения масштаба изображения.

Модуль «Транспорт» предназначен для проектирования пересечений и примыканий дорог, разметки и расстановки дорожных знаков, щитов. Специально разработанный инструмент позволяет осуществлять проверку любых транспортных средств (легковые автомобили, автобусы, грузовые автомобили, в том числе с прицепами, и т.д.) на вписывание в проектируемое пересечение или примыкание дорог. При этом рассчитываются заносы как по передней, так и по задней части автотранспорта, производится отрисовка опасных зон.

Plateia работает со стандартами следующих стран: Россия, Германия, Австрия, Словения, Хорватия, Босния и Герцеговина, Сербия, Македония, Польша, Чехия, Венгрия, Болгария.

В программном комплексе MXROAD [8] выполняются проектные работы от создания трехмерной модели местности и проектируемого объекта до формирования проектной документации – чертежей и ведомостей.

Проектирование оси трассы в плане возможно тремя методами:

- метод проектирования с помощью вершин углов поворота;
- метод проектирования с помощью элементов (прямая, круговая кривая, переходная кривая);
- метод проектирования с помощью сплайнов.

Программа позволяет производить редактирования оси, строить сложные геометрические элементы.

Проектирование продольного профиля возможно тремя методами:

- метод проектирования с помощью точек перелома уклонов с вписанием вогнутых и выпуклых кривых или создание пилообразного профиля без вписывания кривых;
- метод проектирования с помощью элементов (прямая, круговая кривая);
- метод проектирования с помощью сплайнов.

Программа позволяет редактировать профиль, производить автоматическую привязку начальной и конечной точек красной линии относительно начала и конца трассы, подгружать сечения дополнительных поверхностей, которые дают возможность точно проектировать пересечения, примыкания и транспортные развязки.

При проектировании проезжей части используется редактируемая библиотека шаблонов проезжей части с различными вариантами: без разделительной полосы, с разделительной полосой и т.д. При необходимости можно создавать индивидуальные шаблоны.

Конструкция дорожной одежды задается как по проезжей части, так и по тротуарам и обочинам. Варианты конструкции дорожной одежды можно сохранять в библиотеке и использовать в дальнейшем для проектирования других объектов.

Верх земляного полотна задается по различным параметрам, в том числе по уклонам, в зависимости от геометрии проезжей части и т.д.

Программный комплекс CARD/1 «Проектирование автомобильных дорог» [9] позволяет решать задачи проектирования новых, реконструкции и ремонта существующих дорог. Реализовано не только проектирование городских и загородных дорог, но и детальное, многовариантное проектирование примыканий и пересечений в одном и в нескольких уровнях. На основе дан-

ных проекта производится оценка видимости, расчет объемов, создание отчетной документации, а также трехмерная визуализация проекта.

Базовый пакет включает в себя:

- подготовку топографических планов;
- создание и трехмерное представление цифровых моделей рельефа, геологических слоев;
- задание собственных атрибутов различным объектам;
- трассирование в плане с использованием классических элементов проектирования (прямая, переходная кривая, круговая кривая);
- конструирование примыканий, пересечений и кольцевого движения в автоматизированном режиме с возможностью автоматического обновления при изменении взаимного расположения осей дорог;
- мощный инструментарий для проектирования транспортных развязок, дающий неограниченную свободу для создания проектов, готовых к строительству;
- проектирование проектной линии в продольном профиле классическими элементами. Проектирование продольного профиля с учетом плана, поперечных уклонов, ширин. Узвязка продольных профилей различных осей на транспортных развязках;
- автоматический перерасчет данных пикетов (например: ширины, поперечные уклоны, продольный профиль) при изменении пикетажа трассы;
- ввод неправильных (резанных) пикетов;
- проектирование типовых и индивидуальных поперечных профилей с учетом всех данных проекта;
- расчет объемов земляных работ, слоев дорожной одежды и т.д.;
- создание чертежей плана, продольного и поперечных профилей; вывод всевозможных ведомостей;
- обмен данными в формате DXF и других текстовых форматах.

В состав системы Ctedo [10] входят четыре основные подсистемы: «Топоплан», «Линейные изыскания», «Генеральный план», «Дороги» и ряд других пакетов прикладных программ. Подсистема «Дороги» предназначена для проектирования загородных автомобильных дорог всех категорий, городских улиц и магистралей и транспортных развязок.

Проектирование ведется по цифровой модели местности площадки или полосы участка проектирования. В качестве ЦММ используются данные изысканий, обработанных в системе «Топоплан» или «Линейные изыскания». Система имеет инструментарий самостоятельного формирования ЦММ с использованием существующих картографических материалов, представленных в виде растровых или векторных данных.

Система позволяет создавать и редактировать трассы дороги с использованием различных стилей трассирования: от «жестких» (прямыми и круговыми кривыми) до очень плавных и эстетичных трасс (прямыми, круговыми кривыми, клотоидами и сплайнами), с удовлетворением архитектурно-ландшафтных требований.

Проектирование продольного профиля включает в себя:

- простое и оптимизационное проектирование профиля дороги;
- использование при интерактивном проектировании профилей всего многообразия геометрических элементов (прямых, квадратных парабол, сплайнов) в любых необходимых комбинациях;

– применение при проектировании профиля методов, аналогичных созданию и редактированию трасс в плане: аппроксимация фрагментов реконструируемого профиля, создание и последующее сопряжение элементов.

– проектирование продольного профиля наиболее подходящими методами, в том числе:

- полное автоматизированное проектирование;
- по секущим с вписыванием кривых;
- методом «опорных точек» и «элементов» с учетом ограничений,

накладываемых технологией производства работ или особенностей реконструкции дороги.

Подсистема «Дороги» позволяет проектировать земляное полотно и проезжую часть с необходимым количеством и параметрами их элементов: полос движения, обочин, разделительной полосы, тротуаров, переходно-скоростных полос, карманов автобусных остановок и пр. Имеется возможность конструирования поперечного профиля как в интерактивном режиме по поперечным уклонам, так и с использованием шаблонов поперечников.

Общим недостатком существующих в настоящее время систем автоматизированного проектирования автомобильных дорог является использование при расчете дорожных одежд в качестве параметра механических свойств грунтов модуля упругости только в статике, что не позволяет адекватно на этапе проектирования оценить прочность и ресурс многослойных структур.

Заключение

Из приведенного обзора методов проектирования автомобильных дорог, в том числе заложенных в современные системы автоматизированного проектирования и геоинформационные технологии, следует основной вывод: реальный ресурс дорог может быть существенно увеличен за счет разработки и введения в нормативную документацию расчетных методик динамики многослойных гетерогенных структур дорог при различных формах внешнего воздействия, при различных ситуациях анализа и проектирования дорог. В расчет жестких дорожных одежд заложены закономерности накопления деформаций при реальном (т.е. с учетом выходных данных динамических воздействий) многократном нагружении вследствие упруго-пластических деформаций и усталости (деструкции материала).

Список литературы

1. Справочная энциклопедия дорожника. V том Проектирование автомобильных дорог / под ред. Г. А. Федотова, П.И. Пospelова. – М., 2007.
2. ИндорСофт. Разработка программного обеспечения для проектирования, строительства, эксплуатации автомобильных дорог и электрических сетей. – URL: <http://www.indorsoft.ru/>
3. Научно-производственная фирма «Топоматик». – URL: <http://topomatic.ru>.
4. ГИПРОДОРИИ. – URL: <http://dortver.ru/programs/?go=www.giprodor.ru>.
5. AutoCAD Civil 3D-проектирование объектов инфраструктуры – Autodesk. – URL: <http://www.autodesk.ru/adsk/servlet/pc/index?siteID=871736&id=14634283>.
6. AutoCAD Civil 3D 2010. – URL: <http://www.csoft.ru/catalog/soft/autocad-civil-3d/autocad-civil-3d-2010.html>.
7. PLATEIA. – URL: <http://oneroads.ru/po/plateia/plateia.htm>.
8. EMT/ Программные Продукты MX Series / MXROAD. – URL: <http://www.emt.ru/products.php?product=104>.

9. Официальный российский сайт CARD/1 – Computer Aided Road/Rail Design. – URL: <http://www.card-1.ru/xslt.php>.
10. Credo-Dialogue – Программные продукты комплекса CREDO. – URL: http://www.credo-dialogue.com/software/list_programs/credo_dorogi.aspx.
-

Смогунов Владимир Васильевич

доктор технических наук, профессор,
заведующий кафедрой теоретической
и прикладной механики, Пензенский
государственный университет

E-mail: penzgu.tmt@inbox.ru

Smogunov Vladimir Vasilyevich

Doctor of engineering sciences, professor,
head of sub-department of theoretical
and applied mechanics, Penza
State University

Митрохина Наталья Юрьевна

кандидат технических наук, доцент,
кафедра теоретической и прикладной
механики, Пензенский государственный
университет

E-mail: penzgu.tmt@inbox.ru

Mitrokhina Natalya Yuryevna

Candidate of engineering sciences, associate
professor, sub-department of theoretical
and applied mechanics, Penza
State University

УДК 534.1 + 625.717

Смогунов, В. В.

Системный анализ методов проектирования автомобильных дорог /
В. В. Смогунов, Н. Ю. Митрохина // Известия высших учебных заведений.
Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 116–127.

ЭЛЕКТРОНИКА, ИЗМЕРИТЕЛЬНАЯ ТЕХНИКА И РАДИОТЕХНИКА

УДК 621.3: 681.3

С. Ю. Байдаров, В. В. Комаров, М. Ю. Паршуков, А. В. Светлов

МЕТОДИКА ОБРАБОТКИ РЕЗУЛЬТАТОВ ИЗМЕРЕНИЙ ПАРАМЕТРОВ ОПЕРАЦИОННОГО УСИЛИТЕЛЯ

Аннотация. Рассмотрена методика автоматизированной обработки результатов измерений параметров операционного усилителя, включая расчет его макро-модели.

Ключевые слова: операционный усилитель, измерение параметров, макромодель.

Abstract. The article considers a technique of automated processing of measurement results of the operational amplifier parameters, including calculation of its macro-model.

Key words: operational amplifier, measurement of parameters, macromodel.

Введение

При разработке радиоэлектронной аппаратуры специального назначения одной из важных и актуальных задач является получение достоверной информации о фактических параметрах элементной базы, используемой в проектируемых изделиях. При производстве такой аппаратуры необходим 100 % входной контроль параметров элементов. По степени сложности измерения и контроля параметров электронных устройств одно из первых мест среди электрорадиоэлементов занимают интегральные операционные усилители (ОУ). Методы измерений статических и динамических параметров ОУ регламентированы в 17 нормативных документах (ГОСТ 23089.1–83 – 23089.17–83). Практическая реализация этих методов с использованием комплекта отдельных измерительных приборов постоянного и переменного тока низких и высоких частот (генераторов сигналов, вольтметров, осциллографов) требует больших затрат времени на коммутацию, настройку приборов и считывание их показаний. Такая технология применима только для выборочного исследования отдельных экземпляров микросхем и не может быть использована для массового входного контроля и отбора микросхем с заданными значениями параметров для наиболее критичных применений.

Для автоматизации измерения параметров ОУ необходимо создание аппаратно-программного комплекса с применением технологии виртуальных приборов [1]. За счет программного управления аппаратной частью такого комплекса обеспечивается его многофункциональность и гибкость. Программное обеспечение комплекса, кроме автоматизации процессов подготовки и проведения измерительного эксперимента, предназначено для математической обработки и представления его результатов.

Авторами предложена методика автоматизированной обработки результатов измерений параметров ОУ, предусматривающая следующие этапы:

1. Ввод измеренных значений параметров ОУ и исходных данных, задаваемых пользователем.

2. Принятие решения о соответствии (или несоответствии) исследуемого экземпляра ОУ требованиям технических условий с индикацией соответствующего сообщения на экране монитора.

3. Выдача сообщения о принадлежности данного ОУ к той или иной группе рассортировки по измеренным значениям параметров.

4. Создание паспорта исследуемого экземпляра ОУ в виде выходного компьютерного файла в текстовом формате и в распечатанном на бумаге виде.

5. Расчет макромодели исследуемого экземпляра ОУ и создание ее текстового описания в виде выходного компьютерного файла в формате, принятом в Spice-совместимых программах схемотехнического моделирования (PSpice, OrCAD, Multisim и др.).

Для реализации всех этапов данной методики разработаны алгоритмы и виртуальные приборы, созданные в среде LabVIEW. Ниже рассмотрены наиболее важные элементы предложенной методики.

1. Ввод измеренных значений параметров ОУ и исходных данных, задаваемых пользователем

На подготовительном этапе измерительного эксперимента пользователем указывается тип исследуемого ОУ. Программой, управляющей работой аппаратно-программного комплекса, из базы данных вызываются исходные данные ОУ, предварительно подготовленные пользователем с помощью соответствующего шаблона:

- расположение и назначение выводов микросхемы;
- номинальные значения электрических параметров;
- предельные допустимые электрические режимы эксплуатации;
- номинальные значения положительного V_{pwrp} и отрицательного

V_{pwrm} напряжений питания;

– емкость коррекции C_c – значение внутренней емкости частотной коррекции, взятое из технического описания микросхемы; при наличии внешней коррекции рассчитывается результирующее значение корректирующей емкости (например, для ОУ К544УД2 – результат последовательного соединения внешней и внутренней емкостей частотной коррекции);

– значения параметров, которые не могут быть определены экспериментально, в частности: максимальный ток короткого замыкания I_{os} – типовое значение из справочника (не измеряется из-за опасности выхода микросхемы из строя); входной ток смещения I_b и разность входных токов смещения I_{bos} для ОУ с входным каскадом на полевых транзисторах.

По завершении измерительного эксперимента управляющей программой вводятся измеренные значения параметров ОУ:

– максимальные выходные напряжения положительной V_{outp} и отрицательной V_{outm} полярности;

– максимальные скорости нарастания SR_p и спада SR_m выходного напряжения; при создании макромодели Бойла со стандартной топологией различие скоростей нарастания и спада достоверно не воспроизводится, по-

этому выбирается меньшее значение и принимается $SRp = SRm$, вследствие чего емкость, учитывающая различие этих скоростей, $C_{ss} = 0$;

- потребляемая мощность в статическом режиме Pd ;
- входной ток смещения Ib и разность входных токов смещения $Ibos$ для ОУ с входным каскадом на биполярных транзисторах;
- коэффициент усиления постоянного напряжения $Avdc$;
- частота единичного усиления f_{0dB} ;
- коэффициент ослабления синфазного сигнала $CMRR$;
- запас устойчивости по фазе на частоте единичного усиления Phi ;
- выходное сопротивление на низких R_{odc} и высоких R_{oac} частотах.

2. Постановка задачи оперативного создания макромодели ОУ

Моделирование устройств на ОУ с использованием Spice-совместимых программ схемотехнического моделирования позволяет снизить материальные и временные затраты при разработке и оптимизации устройств по сравнению с традиционным макетированием [2]. Зарубежными фирмами-производителями ОУ распространяются макромодели, достаточно достоверно воспроизводящие усредненные характеристики ОУ определенного типа. Применение разработанного авторами аппаратно-программного комплекса позволяет оперативно создавать макромодели конкретных экземпляров ОУ по измеренным значениям их параметров, что обеспечивает:

- улучшение сходимости результатов моделирования схем на ОУ и экспериментальных данных, что, в свою очередь, позволяет вместо макетирования схем проводить их моделирование, значительно расширив программу исследования, в том числе в критических режимах работы схем с угрозой выхода микросхем из строя;
- моделирование схем на ОУ в заданном диапазоне температур с предварительным измерением параметров микросхем при разных температурах и созданием серии соответствующих макромоделей ОУ для этих температур;
- моделирование схем на отечественных ОУ, для которых отсутствуют готовые макромодели, а применение макромоделей импортных аналогов невозможно из-за большого разброса параметров отечественных микросхем разных партий.

При выборе топологии макромодели для моделирования широкого класса ОУ общего применения, использующихся для построения большинства узлов современной аппаратуры, предпочтение следует отдать макромодели Бойла. Эта макромодель ОУ, предложенная в 1974 г. [3] и впоследствии дополненная цепями автоматического регулирования коэффициента передачи для ограничения максимальных выходного тока и выходного напряжения [4], фактически стала стандартной макромоделью ОУ и используется всеми фирмами-производителями аналоговых микросхем. Появившаяся позднее макромодель с многополюсной амплитудно-частотной характеристикой [5] используется для моделирования быстродействующих и высокочастотных ОУ.

Топология макромодели Бойла (рис. 1) применяется в современных Spice-совместимых программах схемотехнического моделирования для создания пользователем собственных макромоделей ОУ (например, в разделе Parts программ PSpice, Design Center, DesignLab или в разделе Model Editor программы OrCAD и т.п.). Однако эти программы предусматривают только

последовательный «ручной» ввод исходных данных. В известных публикациях отсутствуют методики составления текстового описания макромодели ОУ, лишь в [3] имеются расчетные соотношения для исходной версии макромодели Бойла.

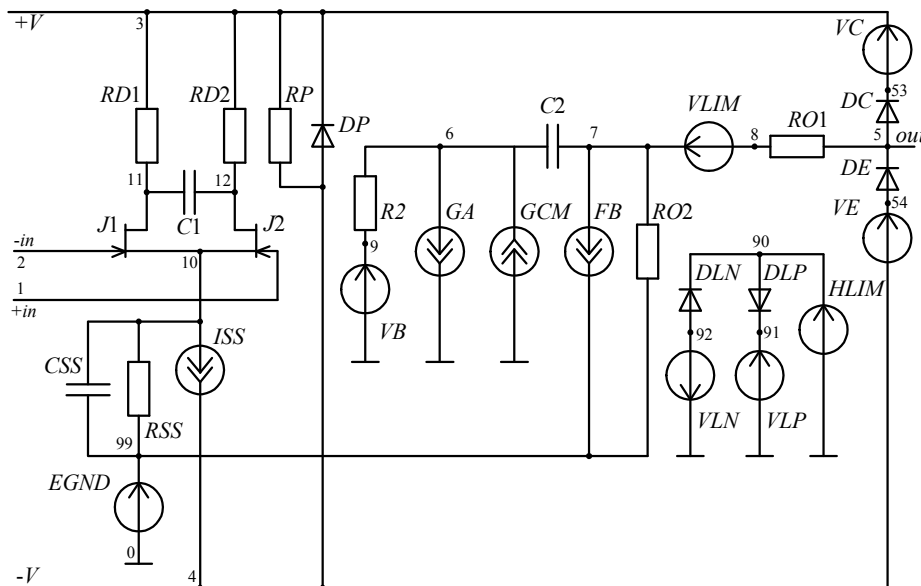


Рис. 1. Стандартная топология макромодели Бойла

Для обеспечения возможности оперативного создания макромодели ОУ по измеренным значениям его параметров необходимо:

- получить расчетные соотношения, связывающие параметры макромодели непосредственно с измеренными значениями параметров ОУ;
- разработать программные средства автоматизированного ввода исходных данных и измеренных параметров ОУ, расчета макромодели и создания выходного файла с ее текстовым описанием.

3. Расчетные соотношения для макромодели ОУ

3.1. Независимый источник постоянного тока I_{SS} с выходным током:

$$I_{SS} = \frac{(SRp \cdot Cc + 2 \cdot Ib + I_{bos}) \cdot SRm + (SRm \cdot Cc + 2 \cdot Ib - I_{bos}) \cdot SRp}{2 \cdot SRm}.$$

3.2. Независимые источники постоянного напряжения V_B и V_{LIM} с нулевым выходным напряжением: $V_b = 0$, $V_{lim} = 0$.

Токи I_{VB} и I_{VLIM} , протекающие через эти источники, выступают в качестве управляющих токов для управляемых источников FB и $HLIM$ соответственно.

3.3. Независимые источники постоянного напряжения V_C , V_E , V_{LP} и V_{LN} с выходными напряжениями:

$$V_C = V_{pwrp} - V_{outp} + U_t \cdot \ln\left(\frac{I_{os}}{I_s}\right);$$

$$V_e = -(V_{pwr m} - V_{out m}) + U_t \cdot \ln\left(\frac{I_{os}}{I_s}\right);$$

$$V_{lp} = V_{ln} = I_{os} \cdot h_{lim},$$

где $I_s = 8 \cdot 10^{-16}$ А – значение тока насыщения полупроводникового диода;
 $U_t = 25,85 \cdot 10^{-3}$ В – температурный потенциал при температуре 27 °С [3].

3.4. Нелинейный источник напряжения, управляемый напряжением, *EGND* с выходным напряжением

$$V_{EGND} = egnd1 \cdot V_{pwr p} + egnd2 \cdot V_{pwr m};$$

$$egnd1 = egnd2 = 0,5,$$

где *egnd1*, *egnd2* – коэффициенты передачи.

3.5. Линейные источники тока, управляемые напряжением, *GA* и *GCM* с выходными токами:

$$I_{GA} = ga \cdot (V(11) - V(12)); I_{GCM} = gcm \cdot (V(10) - V(99));$$

$$ga = 2\pi \cdot f_{0dB} \cdot Cc; gcm = \frac{2\pi \cdot f_{0dB} \cdot Cc}{CMRR},$$

где *ga*, *gcm* – коэффициенты передачи, 1/Ом; *V*(10), *V*(11), *V*(12), *V*(99) – напряжения в узлах 10, 11, 12 и 99 соответственно.

3.6. Линейный источник напряжения, управляемый током, *HLIM* с выходным напряжением:

$$V_{HLIM} = h_{lim} \cdot I_{VLIM},$$

где $h_{lim} = 10^3$ Ом – коэффициент передачи; *I_{VLIM}* – ток, протекающий через источник напряжения *VLIM*.

3.7. Нелинейный источник тока, управляемый током, *FB* с выходным током:

$$I_{FB} = fb1 \cdot I_{VB} - fb2 \cdot I_{VC} + fb3 \cdot I_{VE} + fb4 \cdot I_{VLP} - fb5 \cdot I_{VLN};$$

$$fb1 = \frac{10^5 \cdot Avdc}{ga \cdot R2 \cdot Ro2}; fb2 = fb3 = 10^3; fb4 = fb5 = 4,4 \cdot 10^5,$$

где *fb1*, *fb2*, *fb3*, *fb4*, *fb5* – коэффициенты передачи; *I_{VB}*, *I_{VC}*, *I_{VE}*, *I_{VLP}*, *I_{VLN}* – токи, протекающие через управляющие источники напряжений *VB*, *VC*, *VE*, *VLP* и *VLN* соответственно.

Источник тока *FB* задает коэффициент усиления промежуточного каскада макромодели ОУ при моделировании работы в линейном режиме и уменьшает коэффициент усиления при моделировании ограничения выходного напряжения и выходного тока ОУ. При превышении заданного уровня выходного напряжения открываются диоды *DC* или *DE* в выходном каскаде макромодели. Токи *I_{VC}* или *I_{VE}*, протекающие через эти диоды и включен-

ные последовательно с ними источники напряжения VC или VE , выступают для источника тока, управляемого током (ИТУТ), FB как управляющие сигналы, уменьшающие его коэффициент передачи. Полученная цепь автоматического регулирования ограничивает выходное напряжение макромодели ОУ на заданном уровне.

При превышении заданного максимального выходного тока открываются диоды DLP и DLN . Токи I_{VLP} или I_{VLN} , протекающие через эти диоды, и включенные последовательно с ними источники напряжения VLP или VLN также выступают для ИТУТ FB как управляющие сигналы, создавая еще один контур системы автоматического регулирования коэффициента усиления промежуточного каскада макромодели ОУ.

3.8. Пассивные элементы макромодели ОУ:

$$C1 = \frac{Cc}{2} \cdot \operatorname{tg} \left(\frac{\pi \cdot (90 - \Phi)}{180} \right); C2 = Cc; C_{ss} = 1 \cdot 10^{-30},$$

так как не может принимать нулевое значение;

$$R2 = 100 \cdot 10^3 \text{ Ом}; R_{d1} = R_{d2} = \frac{1}{2\pi \cdot f_{0dB} \cdot Cc};$$

$$R_{o1} = R_{oac}; R_{o2} = R_{odc} - R_{oac};$$

$$R_p = \frac{(V_{pwrp} - V_{pwrn})^2}{P_d - V_{pwrp} \cdot S_{Rp} \cdot Cc - V_{pwrn} \cdot I_{ss}}; R_{ss} = \frac{U_A}{I_{ss}},$$

где $U_A = 200 \text{ В}$ – напряжение Эрли.

3.9. Модели диодов и транзисторов, входящие в состав макромодели.

Используемые в макромодели ОУ диоды DLP , DLN и DP представляются моделью dx , а диоды DC и DE – моделью dy . Эти модели имеют параметры, принятые по умолчанию для моделирования полупроводникового диода в программе PSpice, кроме значения тока насыщения при температуре 27°C , которое принимается равным $I_s = 8 \cdot 10^{-16} \text{ А}$. Модель dy также характеризуется заданными значениями объемного сопротивления $R_s = 10^{-3} \text{ Ом}$ и барьерной емкости $C_{jo} = 10^{-11} \text{ Ф}$.

Транзисторы $J1$ и $J2$ представляются моделями $jx1$ и $jx2$ с параметрами, принятыми по умолчанию для моделирования полевого транзистора в программе PSpice, кроме значений трех параметров:

- пороговое напряжение принимается равным $V_{to} = -1 \text{ В}$;
- коэффициент пропорциональности рассчитывается по формуле

$$\text{Beta} = \frac{(2\pi \cdot f_{0dB} \cdot Cc)^2}{I_{ss}};$$

- ток насыщения управляющего p – n –перехода:

$$I_{s1} = \frac{2 \cdot I_b + I_{bos}}{4}; I_{s2} = \frac{2 \cdot I_b - I_{bos}}{4}.$$

4. Создание выходного файла с текстовым описанием макромоделей

Результаты расчета параметров макромоделей ОУ заносятся в бланк-файл «Spice-модель ОУ» (см. далее). Курсивом обозначены параметры, вставляемые программой.

Тестирование разработанной методики и ее программной реализации в среде LabVIEW проводилось путем сравнения полученных текстовых описаний макромоделей ОУ с аналогичными описаниями макромоделей, созданных программой OrCAD 9.1, раздел Model Editor. Установлено полное совпадение результатов, что свидетельствует о корректности решения поставленной задачи.

Бланк-файл:

```
* Spice-модель ОУ name Экземпляр # Дата __ __ ____
*
*Исходные данные: Vpwrp = Vpwrp; Vpwrn = Vpwrn; Ios = Ios; Cc = Cc;
* (Ib = Ib; Ibos = Ibos)
*Измеренные параметры: Voutp = Voutp; Voutm = Voutm; SRp = SRp;
*SRm = SRm; Pd = Pd; (Ib = Ib; Ibos = Ibos); Avdc = Avdc; f-0dB = f0dB;
*CMRR = CMRR; Phi = Phi; Rodc = Rodc; Roac = Roac
*
* connections: non-inverting input
*               | inverting input
*               | | positive power supply
*               | | | negative power supply
*               | | | | output
*               | | | |
.subckt name # 1 2 3 4 5
*
c1 11 12 C1
c2 6 7 C2
css 10 99 Css
dc 5 53 dy
de 54 5 dy
dlp 90 91 dx
dln 92 90 dx
dp 4 3 dx
egnd 99 0 poly(2) (3,0) (4,0) 0 egnd1 egnd2
fb 7 99 poly(5) vb vc ve vlp vln 0 fb1 fb2 fb3 fb4 fb5
ga 6 0 11 12 ga
gcm 0 6 10 99 gcm
iss 10 4 dc Iss
hlim 90 0 vlim hlim
j1 11 2 10 jx1
j2 12 1 10 jx2
r2 6 9 R2
rd1 3 11 Rd1
rd2 3 12 Rd2
ro1 8 5 Ro1
ro2 7 99 Ro2
rp 3 4 Rp
```

```

rss 10 99 Rss
vb 9 0 dc 0
vc 3 53 dc Vc
ve 54 4 dc Ve
vlim 7 8 dc 0
vlp 91 0 dc Vlp
vln 0 92 dc Vln
.model dx D(Is=800.00E-18)
.model dy D(Is=800.00E-18 Rs=1m Cjo=10p)
.model jx1 NJF(Is = Is1 Beta Vto=-1)
.model jx2 NJF(Is = Is2 Beta Vto=-1)
.ends

```

5. Модифицированная макромодель ОУ

Как было отмечено выше, стандартная топология макромодели Бойла не позволяет достоверно воспроизвести асимметрию реальных значений максимальных скорости нарастания и спада выходного напряжения ОУ. Если специфика области применения ОУ на первый план выдвигает требование достоверного воспроизведения формы импульсов, в том числе длительности их фронтов и времени задержки, то при моделировании схем целесообразно использовать предложенную авторами модифицированную макромодель Бойла (рис. 2) [6, 7], имеющую следующие отличия от макромодели со стандартной топологией:

1. Для воспроизведения различных скоростей нарастания и спада выходного напряжения ОУ дополнительно введены: источник тока, управляемый напряжением GSS ; ключ $S1$, управляемый напряжением; резистор $RS1$. В качестве управляющего напряжения для GSS и $S1$ используется дифференциальное напряжение $[V(12) - V(11)]$, снимаемое с узлов 12 и 11.

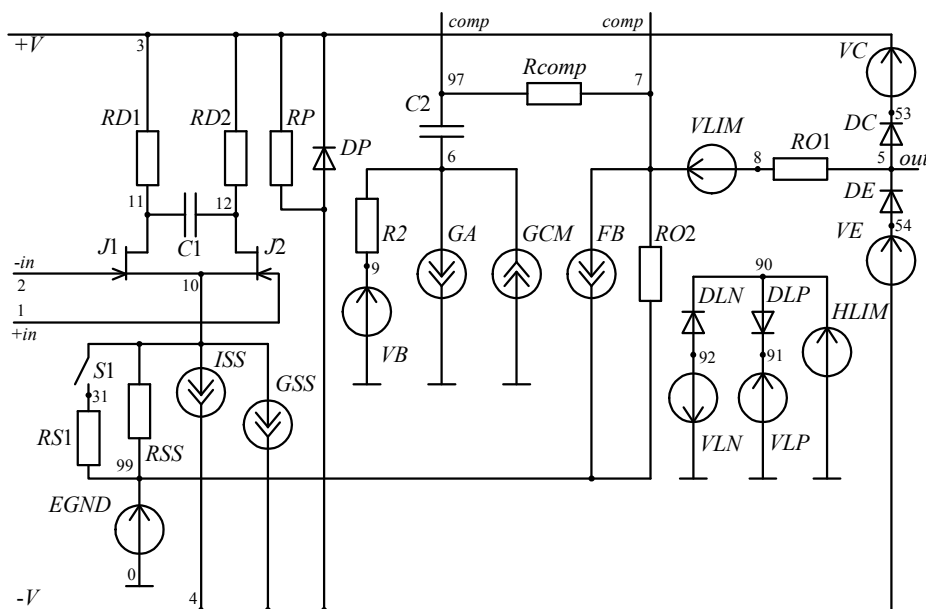


Рис. 2. Топология модифицированной макромодели Бойла

2. Изменена схема подключения цепи частотной коррекции: выводы «*comp*» соединяются между собой при включении внутренней коррекции или к ним подключается внешняя корректирующая емкость (т.е. так же, как у реальной микросхемы). Этим облегчается автоматизация моделирования с вариацией глубины частотной коррекции, поскольку одна и та же макро модель работает и при внутренней, и при внешней частотной коррекции, тогда как стандартная топология требует создания двух разных макромоделей. Для сохранения возможности расчета по постоянному току в состав макро модели дополнительно введен резистор R_{comp} с большим сопротивлением (10^{10} Ом).

Методика создания модифицированной макро модели ОУ предусматривает следующие этапы:

1. С помощью описанной выше программы рассчитывается макро модель ОУ со стандартной топологией с одинаковыми значениями скоростей $(SRp)_1$ и $(SRm)_1$, равными измеренному значению SRp максимальной скорости нарастания выходного напряжения ОУ. Определяются значения $(Iss)_1$ и $(Rss)_1$ – тока источника ISS и сопротивления резистора RSS . Выполняется моделирование повторителя напряжения с ОУ, представляемым полученной макро моделью. На вход ОУ подается скачок напряжения $+U$ с амплитудой, обеспечивающей насыщение входного каскада ОУ (например, на 2–3 В меньше V_{outp}). Определяется разность напряжений на внутренних узлах макро модели $[V(12) - V(11)]_1$ в момент времени, соответствующий максимальной скорости нарастания SRp (например, при достижении выходным напряжением ОУ значения, равного половине напряжения питания V_{pwrp}).

2. Рассчитывается макро модель ОУ с одинаковыми значениями скоростей $(SRp)_2$ и $(SRm)_2$, равными измеренному значению SRm максимальной скорости спада выходного напряжения ОУ. Определяются значения $(Iss)_2$ и $(Rss)_2$. В результате моделирования повторителя напряжения определяется разность напряжений $[V(12) - V(11)]_2$ в момент времени, соответствующий максимальной скорости спада SRm , при воздействии скачком напряжения $-U$.

3. Составляется система уравнений

$$\begin{cases} Iss + gss[U(12) - U(11)]_1 = (Iss)_1, \\ Iss + gss[U(12) - U(11)]_2 = (Iss)_2, \end{cases}$$

в результате решения которой определяются значение Iss – тока источника ISS модифицированной макро модели, а также значение gss – коэффициента передачи источника тока, управляемого напряжением GSS .

4. Сопротивление резистора RSS принимается равным $(Rss)_1$, а сопротивление резистора $RS1$ рассчитывается по формуле

$$RS1 = (Rss)_1 \cdot (Rss)_2 / [(Rss)_1 - (Rss)_2].$$

5. Выбираются параметры модели ключа $S1$, управляемого напряжением: V_{on} – напряжение замыкания ключа, например, 2 мВ; V_{off} – напряжение размыкания ключа, например, 1 мВ; R_{on} – сопротивление замкнутого ключа, например, 1 мОм; R_{off} – сопротивление разомкнутого ключа, например, 10^{10} Ом.

6. Редактируется описание макромодели, полученной в п. 1:

6.1. Макромодели присваивается новое имя, добавляются два новых вывода (97 и 7) в списке внешних выводов:

```
* connections:      non-inverting input
*                  | inverting input
*                  | | positive power supply
*                  | | | negative power supply
*                  | | | | output
*                  | | | | | comp
*                  | | | | | |
*                  | | | | | |
.subckt new name # 1 2 3 4 5 97 7
```

6.2. Изменяется нумерация вывода конденсатора C2.

6.3. Добавляются новые строки:

```
Rcomp 97 7 1E10
GSS 10 4 12 11 gss
RS1 31 99 Rs1
S1 10 31 12 11 SX
.model SX VSWITCH (Von Voff Ron Roff).
```

6.4. Значение тока $(I_{ss})_1$ заменяется на новое значение I_{ss} .

В качестве примера применим предложенную методику для создания макромодели ОУ К544УД2А с внутренней частотной коррекцией по исходным данным и измеренным параметрам, приведенным в табл. 1.

Таблица 1

Параметр	Значение	Параметр	Значение
V_{pwrp} , В	+15	SRp , В/мкс	32
V_{pwrn} , В	-15	SRn , В/мкс	45
I_{os} , мА	20	$Avdc$	20000
C_c , пФ	15	$CMRR$	30000
I_b , пА	100	f_{0dB} , МГц	8,0
I_{bos} , пА	100	Φ_i , град	60
V_{outp} , В	+13	R_{odc} , Ом	120
V_{outn} , В	-13	R_{oac} , Ом	60
P_d , Вт	0,21		

Результаты расчетов приведены в табл. 2.

Таблица 2

Параметр	Значение	Параметр	Значение
$(I_{ss})_1$, А	$480,0 \cdot 10^{-6}$	C_1 , пФ	4,33
$(I_{ss})_2$, А	$675,0 \cdot 10^{-6}$	C_2 , пФ	15,0
V_c , В	2,7979	C_{ss} , пФ	$1 \cdot 10^{-30}$
V_b , В	0	R_2 , Ом	$100 \cdot 10^3$
V_{lim} , В	0	R_{d1} , Ом	$1,3263 \cdot 10^3$
V_e , В	2,7979	R_{d2} , Ом	$1,3263 \cdot 10^3$
V_{lp} , В	20,0	R_{o1} , Ом	60

V_{ln}, B	20,0	$Ro2, \text{Ом}$	60
$egnd1$	0,5	$Rp, \text{Ом}$	$4,2857 \cdot 10^3$
$egnd2$	0,5	$(Rss)_1, \text{Ом}$	$416,67 \cdot 10^3$
$ga, 1/\text{Ом}$	$753,98 \cdot 10^{-6}$	$(Rss)_2, \text{Ом}$	$296,3 \cdot 10^3$
$gcm, 1/\text{Ом}$	$25,133 \cdot 10^{-9}$	$Beta, A/B^2$	$1,1844 \cdot 10^{-3}$
$hlim, \text{Ом}$	$1 \cdot 10^3$	$Is1, A$	$75,0 \cdot 10^{-12}$
$fb1$	$442 \cdot 10^3$	$Is2, A$	$25,0 \cdot 10^{-12}$
$fb2$	$-10 \cdot 10^3$	$gss, 1/\text{Ом}$	$126,54 \cdot 10^{-6}$
$fb3$	$10 \cdot 10^3$	I_{ss}, A	$564,27 \cdot 10^{-6}$
$fb4$	$440 \cdot 10^3$	$Rs1, \text{Ом}$	$1,026 \cdot 10^6$
$fb5$	$-440 \cdot 10^3$		

На рис. 3 представлены временные диаграммы выходного напряжения повторителя, полученные в результате моделирования с использованием рассмотренной макромоделю ОУ K544УД2А с внутренней частотной коррекцией, а также с включением внешних корректирующих емкостей. Данные временные диаграммы с приемлемой для практики точностью совпадают с осциллограммами напряжений на выходе повторителя с ОУ, измеренные значения параметров которого использовались при создании макромоделю.

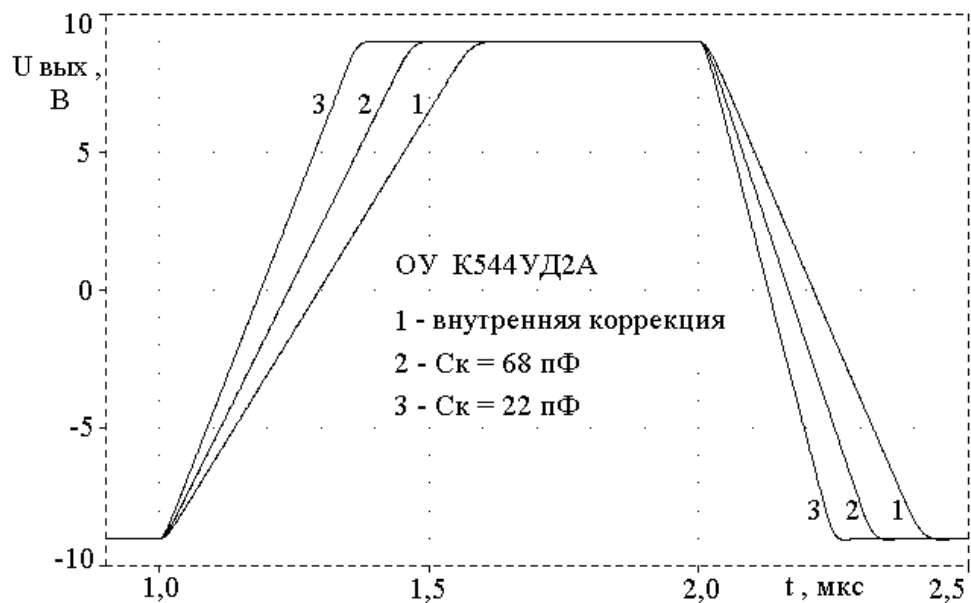


Рис. 3. Результаты моделирования повторителя напряжения

Заключение

В результате решения поставленной задачи авторами разработана методика обработки результатов измерений параметров ОУ, позволяющая автома-

тизировать процессы создания его паспорта, расчета и составления полного текстового описания макромодели, отличающаяся тем, что используемые в ней расчетные соотношения связывают параметры макромодели непосредственно с измеренными значениями параметров ОУ.

Список литературы

1. Расширение функциональных возможностей платы сбора данных для автоматизированного измерения параметров операционных усилителей и создания по измеренным параметрам его макромодели / А. В. Светлов, С. Ю. Байдаров, М. Ю. Паршуков, И. В. Ханин // Образовательные, научные и инженерные приложения в среде LabVIEW и технологии National Instruments : труды IX Международной научно-практической конференции. – М. : Изд-во РУДН, 2010. – С. 229–231.
2. **Казаков, В. А.** Исследование средств измерений параметров активных элементов с помощью схемотехнического моделирования / В. А. Казаков, А. В. Светлов // Датчики и системы. – 2005. – № 5. – С. 33–35.
3. **Boyle, G.** Macromodeling of integrated circuit operational amplifiers / G. Boyle, B. Cohn, D. Pederson, J. Solomon // IEEE journal of solid state circuits. – 1974. – V. SC-9, № 6. – P. 353–364.
4. **Alexander Mark.** SPICE-Compatible Op Amp Macro-Models / Alexander Mark, Derek F. Bowers // Application Note AN-138, Analog Devices. – 1990. – February. – P. 155–170.
5. **Лозицкий, С.** Эволюция SPICE-совместимых макромоделей операционных усилителей / С. Лозицкий // Современная электроника. – 2005. – № 4–6. – Ч. 1–3.
6. **Андреев, А. Н.** Макромодель операционного усилителя для моделирования измерительных цепей с импульсными сигналами / А. Н. Андреев, В. А. Казаков, А. В. Светлов // Измерительная техника. – 1999. – № 9. – С. 26–29.
7. **Светлов, А. В.** Особенности моделирования измерительных схем на операционных усилителях / А. В. Светлов, А. В. Трундов // Информационно-измерительная техника : межвузовский сборник научных трудов. – Пенза : Изд-во Пенз. гос. ун-та, 2005. – Вып. 29. – С. 91–97.

Байдаров Сергей Юрьевич
генеральный директор, ФНПЦ
«Производственное объединение
«Старт» им. М. В. Проценко»
(г. Заречный)

E-mail: rtech@pnzgu.ru

Комаров Всеволод Владимирович
аспирант, Пензенский
государственный университет

E-mail: v.komarov09@rambler.ru

Паршуков Максим Юрьевич
аспирант, Пензенский
государственный университет

E-mail: parshucow@bk.ru

Baydarov Sergey Yuryevich
General manager, FSPC “Production
association “Start” named
after M. V. Protsenko” (Zarechny)

Komarov Vsevolod Vladimirovich
Postgraduate student,
Penza State University

Parshukov Maxim Yuryevich
Postgraduate student,
Penza State University

Светлов Анатолий Вильевич

доктор технических наук, профессор,
заведующий кафедрой радиотехники
и радиоэлектронных систем,
Пензенский государственный
университет

E-mail: rtech@pnzgu.ru

Svetlov Anatoly Vilyevich

Doctor of engineering sciences, professor,
head of sub-department of radio
engineering and radio electronic systems,
Penza State University

УДК 621.3: 681.3

Байдаров, С. Ю.

Методика обработки результатов измерений параметров операционного усилителя / С. Ю. Байдаров, В. В. Комаров, М. Ю. Паршуков, А. В. Светлов // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 128–140.

УЛЬТРАЗВУКОВАЯ ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНАЯ СИСТЕМА ОБЪЕМНОГО РАСХОДА ЧИСТОЙ НЕФТИ

Аннотация. Рассматривается ультразвуковая информационно-измерительная система объемного расхода нефти с возможностью оперативного измерения количества твердых частиц, присутствующих в потоке.

Ключевые слова: ультразвуковой измерительный преобразователь, нефтепровод, измерительный канал, зондирующий сигнал.

Abstract. The article considers an ultrasound information and measurement system for volumetric oil discharge with possibility of effective measurement of an amount of solid particles existing in the oil stream.

Key words: ultrasonic transducer; oil pipeline; measurement channel; probing signal.

Введение

Транспортируемая по нефтепроводу нефть представляет собой сложную смесь самой нефти, растворенного и свободного газа, пластовой воды, твердых веществ (парафина, песка, глины и др.).

При измерении расхода товарной нефти [1], т.е. подготовленной к поставке потребителю, с помощью информационно-измерительных систем (ИИС) определяют такие параметры, как объемный расход нефти и температуру. По объединенной пробе нефти в лаборатории находят другие параметры, очень важным из которых является количественное содержание твердых частиц. В связи со сложностью и трудоемкостью операций по определению указанного параметра его определяют периодически (один раз в месяц) [2] и принимают постоянным на установленный период времени.

Однако это приводит к погрешностям при определении объема чистой нефти (без учета примесей), так как количество частиц в потоке может значительно меняться в зависимости от качества нефти, поступающей в магистральный нефтепровод от разных скважин и даже месторождений.

1. Назначение и структура ИИС объемного расхода чистой нефти

В статье рассматривается ИИС, которая позволяет в процессе измерения общего расхода нефти оперативно (в реальном масштабе времени) определять количество твердых частиц в потоке, а также в дальнейшем осуществлять коррекцию результата измерения объемного расхода товарной нефти путем вычитания из общего результата измерения объема твердых частиц. Таким образом, с помощью данной ИИС возможно определять и качество товарной нефти в зависимости от количества присутствующих в ней твердых частиц.

Блок-схема ИИС представлена на рис. 1. ИИС содержит: два измерительных канала (ИК) – ИК 1 объемного расхода нефти и ИК 2 количественного определения твердых частиц в потоке нефти; датчик температуры (ДТ), блок обработки информации (БОИ) – общий для двух ИК и ДТ, запоминающее устройство (ЗУ) и средство отображения информации (СОИ).

В качестве первого измерительного канала ИК 1 может быть использована ИИС, основанная на любом из известных принципов действия (турбин-

ная, ультразвуковая, роторная, вихревая, электромагнитная ИИС и т.п.), которая должна быть выбрана в зависимости от требований потребителя к точности измерения расхода.

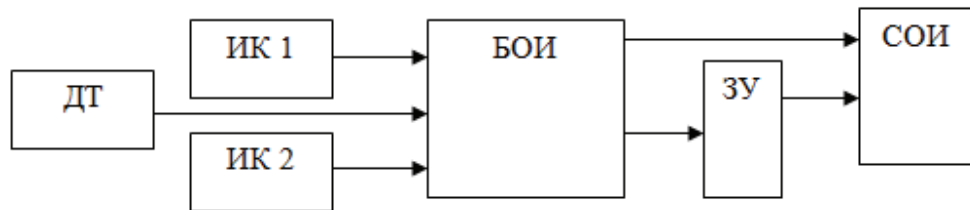


Рис. 1. Блок-схема ИИС объемного расхода чистой нефти

2. Структура и принцип действия второго измерительного канала

Структурная схема второго измерительного канала представлена на рис. 2.

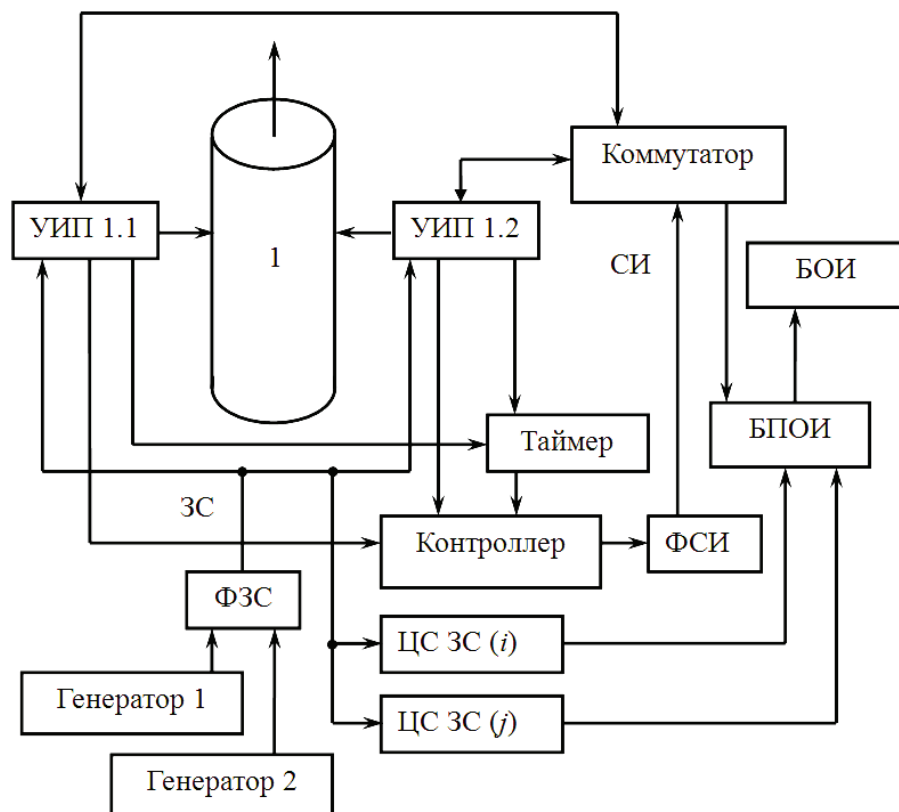


Рис. 2. Структурная схема второго измерительного канала

В качестве измерительных преобразователей (ИП) в исследуемой ИИС используются бесконтактные ультразвуковые измерительные преобразователи (УИП), которые обладают высокой надежностью, простотой обслуживания и длительным межповерочным интервалом [3]. Эти преобразователи являются обратимыми, принцип их действия основан на прямом и обратном

пьезоэлектрическом эффекте. Данные УИП преобразуют акустическую энергию в электрическую и электрическую энергию в акустическую [4].

На рис. 2 приняты следующие обозначения: 1 – нефтепровод, направление потока указано стрелкой; ФСИ – формирователь стробирующих импульсов (СИ), которые переключают коммутатор; ФЗС – формирователь зондирующих сигналов (ЗС), которые подаются на УИП (причем длительность ЗС отлична от длительности тактовых импульсов, задаваемых генератором 1); БПОИ – блок предварительной обработки информации, который вычисляет значения v_i , S_i , φ ; ЦС ЗС – циклические счетчики зондирующих сигналов, которые формируют значения i (номер текущего положения ЗС) и j (номер сканируемой площадки площадью S), значения i и j формируются для обеспечения обработки принятого ЗС; генератор 2 формирует высокочастотное заполнение ЗС; генератор 1 задает период следования для импульсов высокой частоты генератора 2. Многоканальный коммутатор, управляемый контроллером, осуществляет: поочередную активизацию УИП 1.1 и УИП 1.2; перевод УИП из режима передачи ЗС в режим приема и наоборот. Таймер отсчитывает время τ_a прохождения акустической волны на расстоянии D (вдоль диаметра) от одного УИП к другому.

УИП 1.1 и УИП 1.2 создают высокочастотные акустические импульсные колебания (ЗС) в потоке транспортируемой нефти попеременно, чтобы исключить взаимное влияние сканирующих лучей.

Для определения количества твердых частиц, находящихся в транспортируемой нефти, подсчитываются ЗС, отраженные от них. Таким образом, по количеству отраженных сигналов ИИС определяет количество частиц и их объем.

3. Определение количества твердых частиц с помощью исследуемой ИИС

Контролируемый объем перемещается через выбранное сечение « b », которое находится на расстоянии L от места приложения акустического сигнала (рис. 3). Сечение « b » образовано секущей плоскостью « p », расположенной перпендикулярно к нефтепроводу, и представляет собой круг.

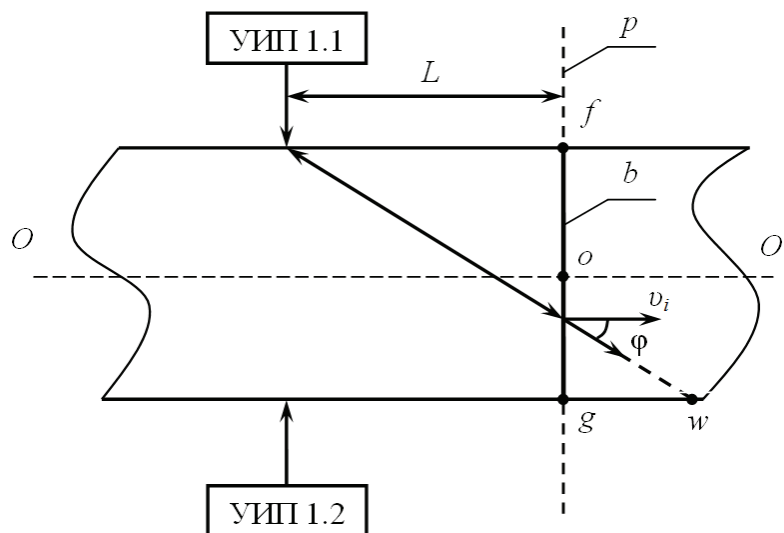


Рис. 3. Схема расположения УИП 1.1 и УИП 1.2 на нефтепроводе

УИП 1.1 обеспечивает зондирование половины сечения «b»: от точки «o» до точки «g»; вторую половину сечения «b» зондирует УИП 1.2 (от точки «o» до точки «f»). ЗС ориентирован под углом φ к вектору скорости v_i перемещения парциального потока нефти, на который в данный момент времени воздействует один из УИП.

При сканировании ЗС сечение «b» разделяется на элементарные площадки, так называемые «ячейки», не перекрывающиеся и вплотную примыкающие друг к другу. В пределах каждой «ячейки» скорость $v_i = \text{const}$, а площадь каждой «ячейки» S_i определяется шириной диаграммы направленности $\Delta\alpha$ измерительного преобразователя.

Площадь «ячейки» S_i должна быть не более среднего статистического размера твердой частицы. Каждому значению φ можно поставить в соответствие значение S_i . Количество «ячеек» зависит от S_i , а длина L (удаление выбранного сечения «b» от места установки УИП) выбирается равной $2D$.

Количество секторов для сканирования сечения нефтепровода составляет

$$N = \text{Int} \left(\frac{2\pi}{\Delta\alpha} \right), \quad (1)$$

а количество «ячеек», размещаемых в пределах данного сектора, составляет

$$M = \text{Int} \left(\frac{D}{2L \sin(\Delta\alpha)} \right). \quad (2)$$

Общее количество зондирований составляет $M \cdot N$, сигналы, получившие отклик, т.е. отразившиеся от частиц, подсчитываются в БОИ (см. рис. 1), их количество составляет K .

4. Определение общего объема твердых частиц с помощью исследуемой ИИС

С помощью ИИС возможно также вычислить объем частиц, проходящих через выбранное контролируемое сечение. Общий объем частиц $V_{\text{общ}}$ равен их количеству K , умноженному на объем средней частицы $V_{\text{ср.ч}}$ (исходя из выбранной модели для средней частицы):

$$V_{\text{общ}} = K \cdot V_{\text{ср.ч}}. \quad (3)$$

За модель средней статистической твердой частицы примем куб с ребром a (в среднем составляет 2 мм), которое соответствует ширине диаграммы направленности $\Delta\alpha$ (рис. 4).

Площадь боковой поверхности куба S_6 равна площади ячейки S_i :

$$S_6 = S_i = a^2. \quad (4)$$

Площадь полной поверхности средней частицы равна

$$S_{\text{п}} = 6a^2 = 6S_i, \quad (5)$$

откуда следует

$$a = \sqrt{\frac{S_n}{6}}. \quad (6)$$

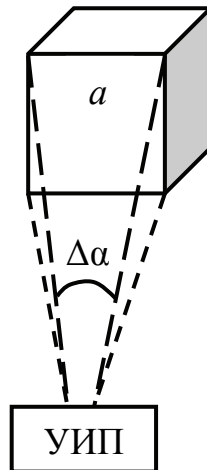


Рис. 4. Модель средней статистической твердой частицы в виде куба с ребром a

Исходя из выбранной модели для средней частицы, ее объем с учетом (5) и (6) рассчитываем следующим образом:

$$V_{\text{ср.ч}} = a^3 = \left(\sqrt{\frac{S_{\text{п}}}{6}} \right)^3 = (\sqrt{S_i})^3. \quad (7)$$

Значение S_i (площади одной грани) рассчитывается в соответствии с [5]. С учетом (7) выражение (3) примет следующий вид:

$$V_{\text{общ}} = K \cdot \left(\frac{\pi}{NM^2} \left(i^2 \frac{D^2}{4} - (i-1)^2 \left(\frac{D}{2} - L \sin(\Delta\alpha) \right)^2 \right) \right)^3. \quad (8)$$

Если реальная частица по размеру больше, чем средняя для выбранной модели, то она будет учтена с помощью двух или более ЗС.

Таким образом формируется значение общего объема твердых частиц, присутствующих в потоке нефти.

5. Влияние скорости движения частиц на результат измерения

Так как частицы движутся со скоростью потока нефти через выбранное сечение p (см. рис. 3), то просканировать все сечение в двух направлениях от точки 0 необходимо за очень короткий промежуток времени $2t_{\text{ск}}$ ($t_{\text{ск}}$ – время сканирования половины сечения). Половина сечения должна быть просканирована за время смещения потока на величину грани куба a (см. рис. 4). Поток будем считать неподвижным в течение времени $2t_{\text{ск}}$.

При средней скорости движения потока нефти в трубопроводе $v_{\text{ср.пот}}$ время сканирования рассчитывается следующим образом:

$$t_{\text{ск}} = \frac{a}{v_{\text{ср.пот}}}. \quad (9)$$

При этом время движения ЗС $t_{\text{ЗС}}$ от УИП до выбранного сечения должно быть не более $t_{\text{ск}}$. Время $t_{\text{ЗС}}$ определяется экспериментально, когда ИИС

установлена на трубопроводе (калибровка ИИС) и при необходимости корректируется расстояние L (см. рис. 3).

Сигналы от УИП, зондирующие половину сечения, должны следовать один за другим, учитывая выражения (1) и (2), с изменяющимся углом ввода сигналов φ_i с временной задержкой:

$$\tau = \frac{2t_{\text{ск}}}{M \cdot N}. \quad (10)$$

Расстояние от УИП до противоположной стенки трубопровода (точка w на рис. 3) для каждого угла φ_i является калиброванным расстоянием (измеряемым при калибровке ИИС на трубопроводе). Время прохождения ЗС калиброванного расстояния в двух направлениях также известно – это калиброванное время. Если время прохождения ЗС до стенки и обратно будет соответствовать калиброванному времени с учетом доверительного интервала, то считаем, что на пути ЗС твердые частицы отсутствуют и данный сигнал не будет учтен ИИС. В противном случае отраженные от частиц сигналы будут учтены ИИС при соответствии калиброванного времени прохождения сигнала до выбранного сечения и обратно.

6. Основные факторы, влияющие на возникновение погрешности измерения ИИС

Погрешности измерения исследуемой ИИС обусловлены следующими основными обстоятельствами:

- вследствие пульсаций и турбулентности потока движение частиц в трубопроводе имеет случайный характер, поэтому одна и та же частица может быть учтена ИИС не один раз;
- частицы в потоке могут быть разных размеров, частица меньше средней (выбранной модели) будет соответствовать завышенному результату измерения;
- при последовательном сканировании частицы большего размера, чем средняя (за несколько сканирований), ее часть может быть не учтена из-за резкого изменения траектории;
- невозможно полностью устранить взаимовлияние ЗС одного УИП на отраженные от частиц ЗС другого УИП;
- несовершенство выбранной модели для твердой частицы, а также допущение неподвижности потока на время сканирования.

7. Экспериментальное исследование ИИС

Испытания ИИС были реализованы с помощью объемной трубопоршневой установки (ТПУ), которая используется для поверки и калибровки различных расходомеров и счетчиков [6]. Структурная схема ТПУ представлена на рис. 5.

Исследуемая ИИС устанавливалась на прямолинейном участке трубопровода 2 ТПУ (масса нефти в резервуаре ТПУ – 170 кг) сразу после первого сигнального детектора 1, т.е. в начале калиброванной секции. Калиброванной секцией считается участок трубопровода, расположенный между двумя сигнальными детекторами 1.

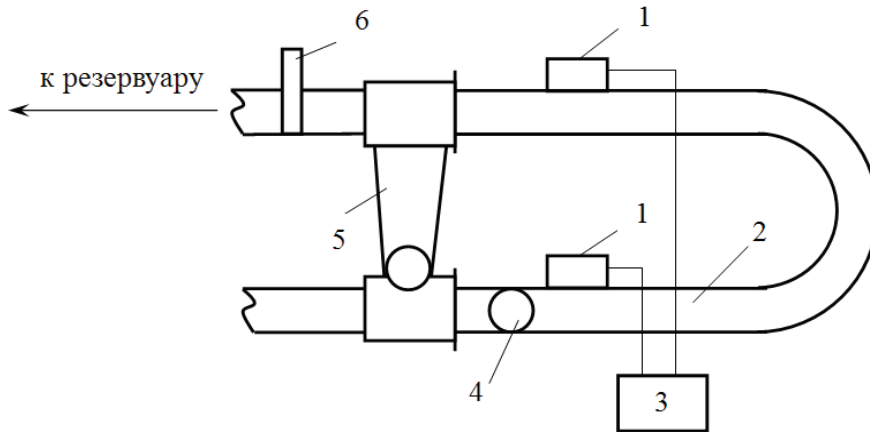


Рис. 5. Основные структурные элементы испытательной ТПУ

Испытание проводилось следующим образом. Через клапан-манипулятор 5 в «петлю» 2 вводятся два шаровых разделителя 4. Один из них выполняет функции запорного устройства (шар-клапан), второй – шарового поршня, который под действием напора рабочей жидкости движется по «петле». В момент прохождения шарового поршня мимо первого детектора 1 одновременно включаются счетное устройство 3, измеряющее количество импульсов фотоэлектрического датчика, и исследуемая ИИС. В момент прохождения шарового поршня мимо второго детектора счетное устройство и ИИС выключаются, тем самым завершается один цикл работы ТПУ. Далее цикл повторяется еще n раз (в данном случае, $n = 5$). Результаты эксперимента сведены в табл. 1.

Таблица 1

Результаты измерений количества твердых частиц в контролируемом потоке с помощью исследуемой ИИС

Цикл, n	Результат измерения исследуемой ИИС в соответствии с (8), м^3	Масса твердых частиц в контролируемом потоке m_i (плотность контролируется ареометром и составляет 850 кг/м^3 , $m_i = V_{\text{общ}} \cdot \rho$), г	Массовая доля твердых частиц в контролируемом потоке, %
1	0,000088	75	0,044
2	0,000097	82	0,049
3	0,000071	60	0,035
4	0,000076	65	0,038
5	0,000082	70	0,041

Среднее значение массы твердых частиц, измеренной с помощью ИИС:

$$m_{\text{ср.ИИС}} = \frac{1}{n} \sum_{i=1}^n m_i = 41 \text{ г.} \quad (11)$$

В результате измерения количество твердых частиц в контролируемом потоке с помощью исследуемой ИИС составило 0,02 % от общего объема, ка-

чество нефти соответствует требованиям [7], так как измеренное значение не превысило наибольшего допустимого значения массовой доли твердых частиц в нефти.

На практике объем твердых частиц в контролируемом потоке нефти определяют по отобранной пробе, т.е. массовую долю частиц в пробе принимают равной массовой доле частиц во всем контролируемом объеме. Поэтому после пяти циклов эксперимента в пробоотборник 6 ТПУ поступили последовательно, через равные промежутки времени, h проб одинакового объема $V_{\text{пробы}}$ (масса пробы 0,2 кг) для проведения их анализа и сравнения результатов лабораторных исследований с результатами измерений ИИС. Отбор проб осуществлялся в соответствии с [8], а массу твердых частиц в отобранных пробах определяли в соответствии с методикой, приведенной в [9]. Данные лабораторных исследований сведены в табл. 2.

Таблица 2

Результаты лабораторных исследований

Проба, h	Масса твердых частиц в отобранной пробе $m_{\text{ч.пр.}i}$, г	Массовая доля твердых частиц во всем контролируемом объеме, определяемая по отобранной пробе, %
1	0,094	0,047
2	0,064	0,032
3	0,070	0,035
4	0,054	0,027
5	0,128	0,064

Далее в соответствии с методикой, представленной в [8], была определена масса твердых частиц в оставшемся объеме $V_{\text{ост}}$. Масса твердых частиц во всем контролируемом объеме определяется как сумма масс частиц в пяти пробах и оставшемся объеме:

$$m_{\text{ч.д}} = m_{\text{ч.ост}} + \sum_{i=1}^h m_{\text{ч.пр.}i} = 46 \text{ г} . \quad (12)$$

По данным экспериментальных исследований из всего контролируемого объема нефти (170 кг) масса твердых частиц составила 46 г, т.е. 0,03 % общего объема, что соответствует требованиям [7].

Полученный результат – действительное значение массы твердых частиц в контролируемом объеме, $m_{\text{ч.д}}$, с которым сравнивались результаты измерений лабораторным методом по отобранной пробе и с помощью ИИС.

Сравнив результат измерения с помощью ИИС с действительной массой частиц в контролируемом объеме, получили абсолютную погрешность, г:

$$\Delta_{\text{ИИС}} = m_{\text{ч.д}} - m_{\text{ср.ИИС}} = 46 - 37 = 9 . \quad (13)$$

Относительная погрешность измерения ИИС, %:

$$\delta_{\text{ИИС}} = \frac{\Delta_{\text{ИИС}}}{m_{\text{ч.д}}} \cdot 100 = 19,6 . \quad (14)$$

Далее результаты лабораторных исследований отобранных проб (табл. 2) сравнили с действительным значением $m_{ч,д}$ и для каждой пробы рассчитали относительную погрешность измерения (табл. 3).

Таблица 3

Относительные погрешности лабораторного метода

Проба, h	Масса твердых частиц во всем объеме (по отобранной пробе), $m_{ч,г}$	Относительная погрешность $\delta_{\text{лаб.}i}$, %
1	54	17,0
2	54	17,4
3	55	20,3
4	55	19,7
5	64	40,3

Из результатов лабораторных исследований видно, что массовая доля твердых частиц неодинакова во всех пробах, так как частицы распределены неравномерно по всему объему нефти. В последней пробе концентрация твердых частиц превысила максимальное допустимое значение [7], и ошибочно должно быть принято решение о дополнительной фильтрации нефти, в то время как процентное содержание твердых частиц всего объема максимального допустимого значения не превысило.

Таким образом, с учетом данных лабораторных исследований, принимаемая периодичность определения количества твердых частиц – один раз в месяц, поставщик рискует сбывать товарную нефть по заниженной стоимости. Поэтому выгоднее исследовать весь объем поставляемого продукта, а не его часть. Такой возможностью обладает исследуемая ИИС. Кроме того, измерение количества твердых частиц с помощью ИИС точнее, чем при лабораторных исследованиях, хотя погрешность измерения не исключается, соответствует выражению (14) и обусловлена обстоятельствами, описываемыми в шестом разделе настоящей статьи.

Заключение

Применение исследуемой ИИС в качестве дополнительного канала дает возможность измерения расхода чистой нефти, сокращения объема лабораторных исследований и оперативной оценки качества товарной нефти в зависимости от содержания в ней твердых частиц точнее, чем при лабораторных исследованиях.

Список литературы

1. ГОСТ 8.595–2004. Государственная система обеспечения единства измерений. Масса нефти и нефтепродуктов. Общие требования к методикам выполнения измерений. – М. : Стандартинформ, 2005. – 8 с.
2. Автоматизированный учет нефти и нефтепродуктов при добыче, транспорте и переработке / А. Ш. Фатхутдинов и др. – М. : Недра, 2002. – 417 с.
3. **Кремлевский, П. П.** Расходомеры и счетчики количества : справочник / П. П. Кремлевский. – Л. : Машиностроение. Ленингр. отд-ние, 1989. – 701 с.
4. **Петров, П. Н.** Акустика. Электроакустические преобразователи : учебное пособие / П. Н. Петров. – СПб. : СПбГУАП, 2003. – 80 с.

5. Патент № 2319933 Российская Федерация. Измеритель объема жидкости, транспортируемой по нефтепроводу / Куликовский К. Л., Еремин И. Ю. – № 2006110181 ; заявл. 20.03.2008.
 6. **Хансуваров, К. И.** Техника измерения давления, расхода, количества и уровня жидкости, газа и пара : учебное пособие для техникумов / К. И. Хансуваров, В. Г. Цейтлин. – М. : Издательство стандартов, 1990. – 287 с.
 7. ГОСТ Р 51858–2002. Нефть. Общие технические условия. – М. : Издательство стандартов, 2006. – 12 с.
 8. ГОСТ 2517–85. Нефть и нефтепродукты. Методы отбора проб. – М. : Издательство стандартов, 2003. – 27 с.
 9. ГОСТ 6370–83. Нефть, нефтепродукты и присадки. Метод определения механических примесей. – М. : Стандартинформ, 2007. – 7 с.
-

Куликовский Константин Лонгинович
доктор технических наук, профессор,
кафедра информационно-измерительной
техники, Самарский государственный
технический университет

E-mail: kl195lm@mail.ru

Kulikovskiy Konstantin Longinovich
Doctor of engineering sciences, professor,
sub-department of information-measuring
equipment, Samara State
Technical University

Якунина Виктория Валерьевна
аспирант, Самарский государственный
технический университет

E-mail: kl195lm@mail.ru

Yakunina Viktoriya Valeryevna
Postgraduate student,
Samara State Technical University

УДК 621.6-52

Куликовский, К. Л.

Ультразвуковая информационно-измерительная система объемного расхода чистой нефти / К. Л. Куликовский, В. В. Якунина // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2012. – № 1 (21). – С. 141–150.

ИССЛЕДОВАНИЕ ЭЛЕКТРОМАГНИТНЫХ И ТЕПЛОВЫХ ПОЛЕЙ В БЛИЖНЕЙ ЗОНЕ КОНТАКТНОГО ВОЛНОВОДНОГО АППЛИКАТОРА¹

Аннотация. Рассмотрена конструкция контактного волноводного аппликатора для микроволновой терапии злокачественных опухолей. Разработана математическая модель аппликатора с фокусирующей линзой, вспомогательным поглощающим элементом и прилегающей двухслойной биологической средой. Методом конечных элементов решается связанная задача электродинамики и теплопроводности. Приведены результаты численного анализа тепловых полей в области взаимодействия для различных режимов работы данного устройства.

Ключевые слова: электромагнитное поле, биологическая ткань, температура, численное моделирование.

Abstract. The article considers a design of contact type waveguide applicator for microwave therapy of tumor. The authors have developed a mathematical model of the applicator with a focusing lens, an auxiliary absorbing component and adjacent double layered biological media. Coupled electromagnetic and heat conduction problem is solved using the finite element method. The article adduces the results of numerical analysis of thermal fields in the interaction domain for various operating regimes of the given device.

Key words: electromagnetic field, biological tissue, temperature, numerical modeling.

Введение

Способность электромагнитных (ЭМ) волн микроволнового диапазона проникать в глубь диэлектрических материалов широко применяется в установках СВЧ-нагрева различного назначения. Для этих целей выделены так называемые ISM (Industrial Scientific Medicine) частоты: 433, 915, 2450, 5800 МГц, название которых определяет основные сферы их использования: промышленная, научная и медицинская.

В современной медицине СВЧ-излучение является эффективным инструментом как диагностики, так и лечения многих заболеваний, в том числе онкологических. В частности, при СВЧ-гипертермии с помощью ЭМ поля удается повысить температуру пораженного органа на 6–8 °С, а при проведении СВЧ-абляции – осуществить полную деструкцию раковых клеток в зоне коагуляции.

Современные установки для локальной СВЧ-гипертермии представляют собой целый комплекс устройств, в том числе: генератор микроволнового излучения, фидеры, аппликаторы для подведения излучения к пораженной ткани, системы управления и термоконтроля, приборы снятия и обработки информации. Важнейшим элементом такой установки является волноводный, микрополосковый или коаксиальный аппликатор для облучения опухолей разного размера и типа. При этом волноводные излучатели относятся к кате-

¹ Работа выполнена при поддержке Министерства образования и науки РФ (грант РНП.2.1.1.575).

гории неинвазивных [1], микрополосковые могут быть как неинвазивными [2], так и инвазивными [3], а коаксиальные, как правило, служат для интерстициального СВЧ-воздействия на биоткани [4].

Волноводные и рупорные аппликаторы ориентированы на облучение опухолей, находящихся на небольшой глубине от поверхности тела. С целью увеличения глубины проникновения ЭМ поля в многослойную диэлектрическую среду (биологическую ткань) применяют рабочие частоты 433 и 915 МГц, что приводит к увеличению габаритных размеров волноводных элементов. Для уменьшения размеров и улучшения фокусирующих свойств такие аппликаторы частично или полностью заполняются жидкими [5] или твердыми [6] диэлектриками. В ряде случаев излучатель выполняется в виде металлопластинчатой линзы [5] или волновода сложной формы [7].

В работе [8] была предложена модель микроволнового аппликатора сложной конфигурации, в конструкции которого для фокусировки ЭМ поля в биологических средах предусмотрена выпуклая диэлектрическая линза. В данной работе анализируется влияние еще одного элемента, часто используемого в системах СВЧ-гипертермии и абляции для предотвращения перегрева кожных покровов и получившего название болюса [9], на распределение ЭМ и теплового поля в области взаимодействия.

1. Математическая модель аппликатора

На рис. 1 показана конфигурация цилиндрического аппликатора с металлическими Т-ребрами и фокусирующим элементом [8]. Диаметр фокусирующего элемента совпадает с диаметром аппликатора. Он включает в себя диэлектрическую линзу ($R_{\text{л}} = 45$ мм; $H_{\text{л}} = 40$ мм) с комплексной диэлектрической проницаемостью $\epsilon_{\text{л}}$, остальное пространство имеет воздушное заполнение.

Здесь же приведена двухслойная модель жировой и мышечной биотканей, прилегающих к открытому концу волновода, диэлектрические и теплофизические свойства которых можно найти в литературе [4, 10]. Между аппликатором высотой $L = 210$ мм и облучаемым объектом размещается болюс, представляющий собой пластиковый контейнер, заполненный водопроводной водой, нагретой до некоторой температуры $T = \text{const}$.

С помощью коаксиально-волноводного перехода в аппликаторе возбуждается основная волна с рабочей частотой 915 МГц. При этом максимум электрического поля формируется между металлическими ребрами в центре апертуры ($a = 92$ мм; $d = 7$ мм; $t = 50$ мм; $w = 10$ мм; $l = 4$ мм; $u = 1,5$ мм).

Для описания процессов распространения и поглощения ЭМ волн в такой системе применяются неоднородные уравнения Гельмгольца для комплексных амплитуд электрического ($\dot{E}$) и магнитного ($\dot{H}$) полей, а также сторонних источников ($\dot{J}_{\text{ст}}$) с учетом диэлектрических и магнитных ($\mu' = 1$; $\mu'' = 0$) свойств биологических тканей:

$$\nabla^2 \dot{E} + k_0^2 (\epsilon' - j\epsilon'') \dot{E} = j\omega\mu_0 \dot{J}_{\text{ст}} + \frac{1}{j\omega\epsilon_0} \text{grad}(\text{div} \dot{J}_{\text{ст}}); \quad (1)$$

$$\nabla^2 \dot{H} + k_0^2 \dot{H} = -\text{rot} \dot{J}_{\text{ст}}, \quad (2)$$

где $k_0^2 = \omega^2 \epsilon_0 \mu_0$ – волновое число свободного пространства; ω – круговая частота; ϵ_0, μ_0 – диэлектрическая и магнитная постоянные; ϵ', μ' – диэлектрическая и магнитная проницаемость среды; ϵ'', μ'' – коэффициенты диэлектрических и магнитных потерь.

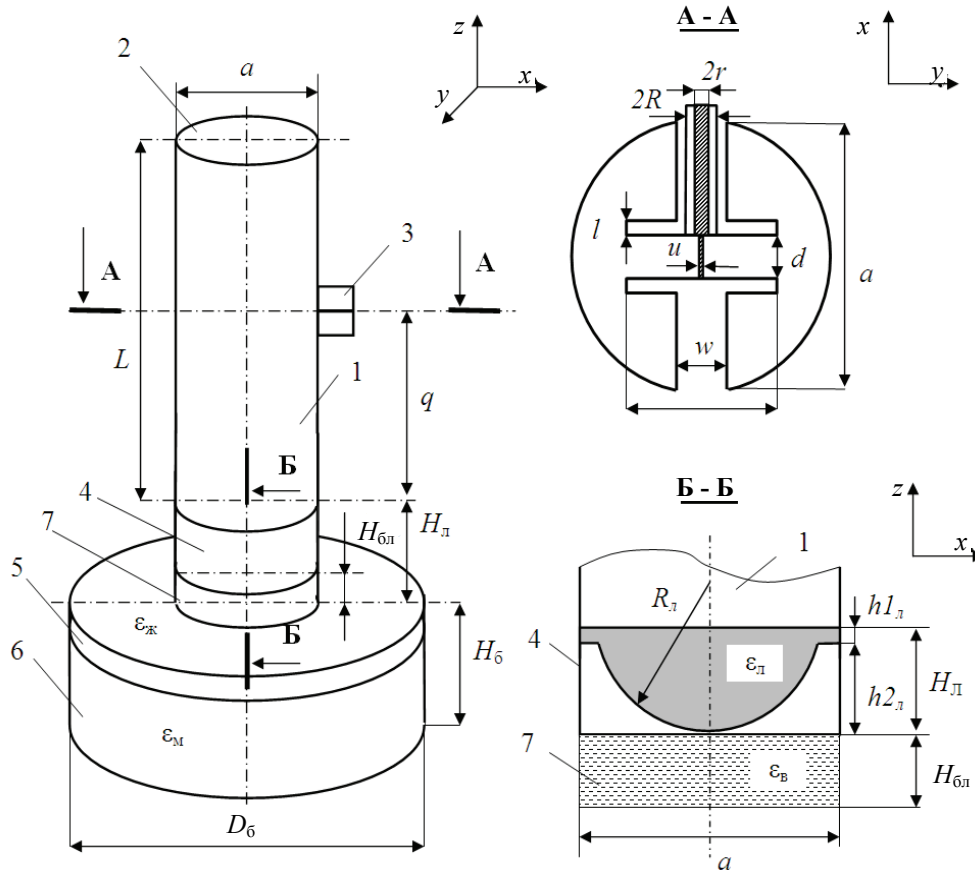


Рис. 1. Модель линзового аппликатора с прилегающей диссипативной средой:
 1 – отрезок цилиндрического волновода; 2 – короткозамкнутый конец волновода;
 3 – коаксиальный элемент возбуждения; 4 – фокусирующий элемент;
 5 – слой жировой ткани; 6 – слой мышечной ткани; 7 – болюс

Решения (1) и (2) должны удовлетворять граничным условиям на металлических стенках и условиям непрерывности тангенциальных составляющих поля на границе раздела сред. Кроме того, на входе аппликатора необходимо задать гармонический сигнал заданной амплитуды и фазы, а также распределение собственных функций поля *ТЕМ*-волны. На границе области, занимаемой биотканью, задаются условия рассеяния ЭМ поля [4].

Уровень СВЧ-мощности, поглощенной многослойной диэлектрической средой, можно оценить, зная удельный коэффициент поглощения:

$$P_{уд} = 0,5 \omega \epsilon_0 \epsilon'' \dot{E}^2 \rho_T^{-1}, \quad (3)$$

где ρ_T – плотность среды с потерями.

Тепловые процессы внутри биоткани, нагреваемой СВЧ-излучением, определяются из решения биотеплового уравнения [11]:

$$\rho_T C_T \frac{\partial T}{\partial \tau} = \lambda_T \nabla^2 T - \rho_T \rho_K C_K \eta (T - T_K) + Q_M + Q_V, \quad (4)$$

где ρ_T , ρ_K – плотность ткани и крови; C_T , C_K – теплоемкость ткани и крови; λ_T – теплопроводность ткани; η – интенсивность кровотока в зоне нагрева ($4,2 \cdot 10^{-7} \leq \eta$, $\text{м}^3/\text{кг} \cdot \text{с} \leq 1,67 \cdot 10^{-5}$); Q_M – удельная плотность энергии, обусловленной комплексом биохимических и связанных с ними энергетических процессов; T_K – температура крови; Q_V – плотность СВЧ-энергии.

Температуру болюса можно установить из решения классического уравнения теплопроводности:

$$\rho_B(T) C_B(T) \frac{\partial T}{\partial \tau} = \lambda_B(T) \nabla^2 T + Q_V, \quad (5)$$

где ρ_B , C_B , λ_B – плотность, теплоемкость и теплопроводность воды.

Причем [12]

$$\rho_B(T) = -4 \cdot 10^{-5} T^3 + 0,0023 T^2 - 0,3599 T + 1005,1, R^2 = 0,9991; \quad (6)$$

$$C_B(T) = -10^{-4} T^3 + 0,0309 T^2 - 2,0378 T + 4210,8, R^2 = 0,9988; \quad (7)$$

$$\lambda_B(T) = -10^{-5} T^2 + 0,0024 T + 0,5565, R^2 = 0,9994, \quad (8)$$

где R^2 – коэффициент корреляции.

Соотношения (6)–(8) справедливы в температурном интервале $10 \leq T \leq 100$ °C.

2. Результаты численного расчета

Для дискретизации уравнений (1), (2), (4) и (5) в данной работе был использован трехмерный метод конечных элементов (МКЭ), реализованный в многофункциональном пакете программ COMSOL¹, который позволяет проводить численный расчет связанных ЭМ и тепловых полей с учетом вариаций физических свойств нагреваемой диэлектрической среды. Так как рассматриваемый аппликатор (см. рис. 1) предназначен для СВЧ-гипертермии биологических тканей, что подразумевает незначительное повышение температуры в зоне облучения, решалась упрощенная сопряженная задача электродинамики и теплопроводности. При этом значения $\epsilon'(T) - j\epsilon''(T)$ для болюса можно найти в [13]. На электродинамическом этапе решения задачи применялись векторные тетраэдрические элементы второго порядка, а на этапе решения тепловой задачи – скалярные функции формы с кубической полиномиальной аппроксимацией.

Результаты численного моделирования плотности энергии в области взаимодействия СВЧ-излучения с биотканью в двух взаимно ортогональных плоскостях показаны на рис. 2.

Расчеты показали, что максимум выделяемой СВЧ-мощности формируется в области мышечной ткани на глубине от 10 до 15 мм. Значение диэлектрической проницаемости материала, из которого изготовлена линза, составляет $\epsilon_n = 16$ (керамика).

¹ www.comsol.com.

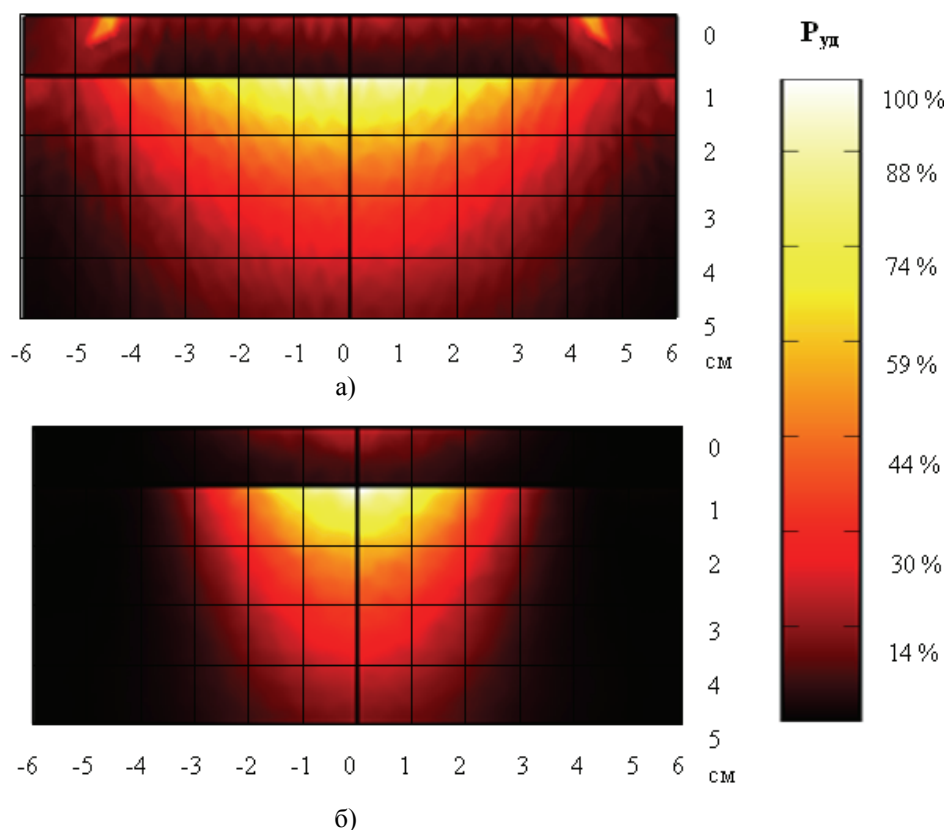


Рис. 2. Распределение удельного коэффициента поглощения в биологической ткани 0–1 см – жировая ткань, 1–5 см – мышечная ткань; а – плоскость XZ ; б – плоскость YZ

Далее в работе были получены тепловые поля в зоне нагрева при фиксированной входной мощности аппликатора 30 Вт и построены зависимости температуры от продольной координаты Z (рис. 3). Для сравнения были выбраны два варианта конструкции аппликатора: с диэлектрической линзой (но без болюса) и без нее. Исследование термодинамических процессов показало, что температура 42 °С в биоткани достигается через 2 мин после начала облучения. При этом в первом случае (рис. 3,а) перегрева жировой ткани не возникает, в отличие от ситуации, когда используется аппликатор без диэлектрической линзы (рис. 3,б).

Как видно из рис. 3, поверхностные ткани даже при использовании фокусирующего элемента все равно подвергаются паразитному нагреву. Для устранения этой проблемы конструкция аппликатора дополняется водяным болюсом с постоянной температурой.

Водяной болюс в микроволновой гипертермии выполняет две важные функции: остужает область тела для устранения перегрева поверхностных тканей и обеспечивает согласование аппликатора с нагрузкой (биологической средой). Выбор оптимальной температуры болюса является отдельной задачей, так как современные системы СВЧ-гипертермии позволяют динамически изменять его температуру, обычно в диапазоне от 5 до 40 °С для достижения наилучших параметров процесса нагрева.

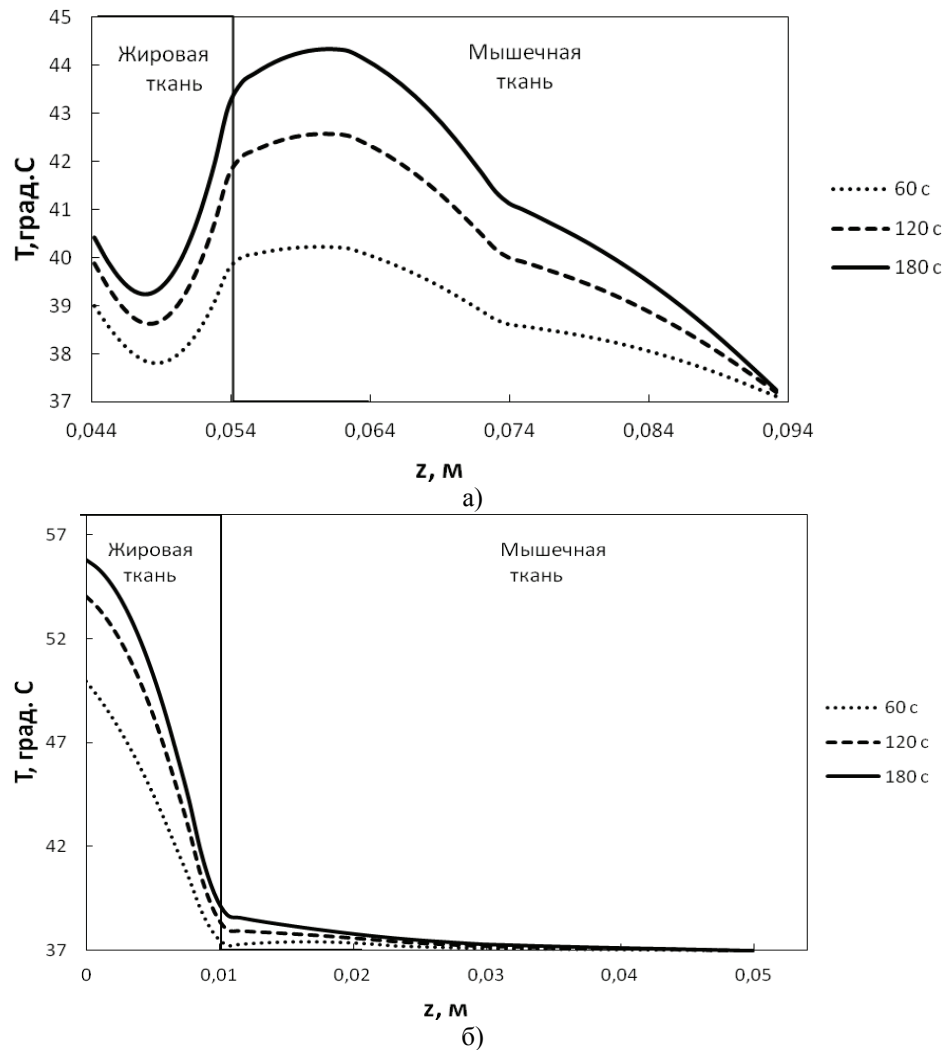


Рис. 3. Распределение температуры вдоль центральной оси модели для аппликатора с фокусирующей линзой (а) и без фокусирующей линзы (б)

Для данной конструкции аппликатора использовался болюс высотой 2 см, диаметром, равным диаметру аппликатора. Температура болюса оставалась постоянной. На рис. 4 приведены зависимости температуры биологической ткани вдоль основной оси на разной глубине для фиксированного значения подводимой мощности и для болюса с температурой 30 °C.

Численная модель продемонстрировала, что применение болюса несколько замедляет процесс нагрева ткани: температура 42 °C достигается за 3 мин, но при этом удается добиться заданного распределения температуры внутри двухслойной биологической среды. Нагрев поверхностных тканей компенсируется температурой водяного болюса, которая поддерживается постоянной.

Заключение

С помощью трехмерной конечно-элементной модели в данной работе были установлены распределения плотности энергии и температуры в ближ-

ней зоне линзового волноводного аппликатора для микроволновой гипертермии биологических тканей. Проведенные исследования показали, что специальный поглощающий элемент аппликатора, болюс, позволяет не только избежать перегрева кожных покровов, но и обеспечить формирование максимума ЭМ поля в мышечной ткани на глубине 1–3 см от поверхности тела. Температура болюса при этом должна поддерживаться $T = 30^\circ\text{C}$.

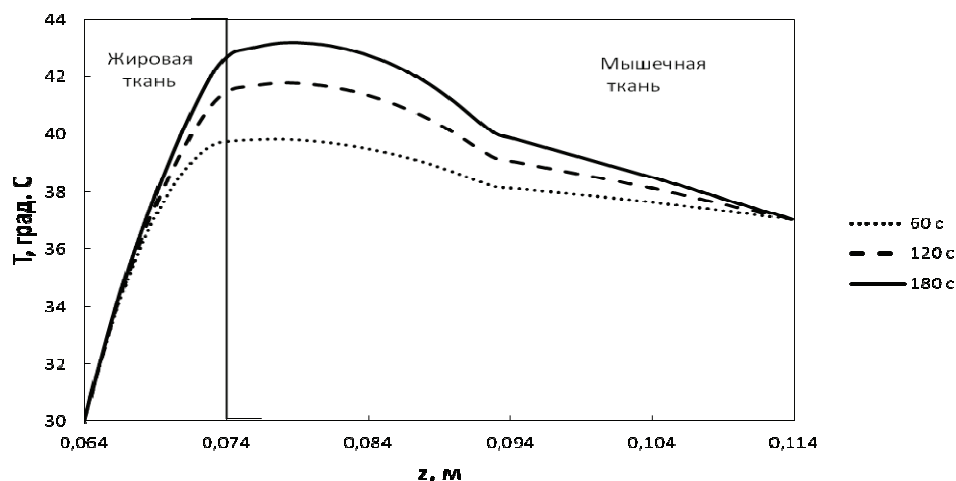


Рис. 4. Распределение температуры вдоль центральной оси модели для аппликатора с фокусирующей линзой и водяным болюсом

Список литературы

1. **Давидович, М. В.** Нагрев биологических тканей аппликатором типа открытый конец волновода / М. В. Давидович // Биомедицинские технологии и радиоэлектроника. – 2007. – № 1. – С. 51–55.
2. **Johnson, J. E.** Evaluation of a dual-arm Archimedean spiral array for microwave hyperthermia / J. E. Johnson, D. G. Neuman, P. F. Maccarini, T. Juang, P. R. Stauffer, P. Turner // International Journal of Hyperthermia. – 2006. – V. 22, № 6. – P. 475–490.
3. **Dissanayake, T.** Dielectric load impedance matching for wideband implanted antennas / T. Dissanayake, K. P. Esselle, M. R. Yuce // IEEE Trans. on Microwave Theory and Techniques. – 2009. – V. 57, № 10. – P. 2480–2487.
4. **Макаров, В. Н.** Сравнительный анализ микроволнового и радиочастотного нагрева при тепловой абляции опухолей / В. Н. Макаров, Г. В. Ющенко // Биомедицинская радиоэлектроника. – 2009. – № 2. – С. 3–10.
5. **Maruoka, S.** Focusing applicator for microwave heating / S. Maruoka, Y. Nikawa // Proceedings of the First Global Congress on Microwave Energy Applications. – Otsu, Japan, – 2008. – P. 795–798.
6. **Van Rhoon G. C.** A 433 MHz Lucite cone waveguide applicator for superficial hyperthermia / G. C. Van Rhoon, P. J. M. Rietveld, Van Der Zee J. // International Journal of Hyperthermia. – 1998. – V. 14, № 1. – P. 13–27.
7. **Kantor, G.** The performance of a new 915-MHz direct contact applicator with reduced leakage / G. Kantor, D. M. Witters // Journal of Microwave Power and Electromagnetic Energy. – 1983. – V. 18, № 2. – P. 133–142.
8. **Новрузов, И. И.** Волноводный аппликатор для локального нагрева биологических тканей СВЧ-излучением / И. И. Новрузов, В. В. Комаров // Актуальные проблемы электронного приборостроения : материалы Международной научно-технической конференции. – Саратов : Изд-во СГТУ, 2010. – С. 199–202.

9. **Togni, P.** Water bolus influence on temperature distribution of applicator for small melanoma tumors / P. Togni, J. Vrba, L. Vannucci // Proceedings of the 39th European Microwave Conference. – Rome, Italy, 2009. – P. 874–877.
10. **Kumar, S. B.** Dielectric properties of certain biological materials at microwave frequencies / S. B. Kumar, K. T. Mathew, U. Raveendranath, P. Augustine // International Journal of Microwave Power and Electromagnetic Energy. – 2001. – V. 36, № 2. – P. 67–75.
11. **Pennes, H. H.** Analysis of tissue and arterial blood temperature in the resting forearm / H. H. Pennes // Journal of Applied Physiology. – 1948. – V. 4. – P. 93–122.
12. **Луканин, В. Н.** Теплотехника / В. Н. Луканин, М. Г. Шатров, Г. М. Камфер ; под ред. В. Н. Луканина. – М. : Высшая школа, 2002. – 671 с.
13. **Eves, E.** Analysis of operational regimes of a high power water load / E. Eves, V. V. Yakovlev // Journal of Microwave Power and Electromagnetic Energy. – 2002. – V. 37, № 3. – P. 127–144.

Новрузов Илья Игоревич

аспирант, Саратовский государственный
технический университет

E-mail: novruzov@bk.ru

Novruzov Ilya Igorevich

Postgraduate student, Saratov
State Technical University

Комаров Вячеслав Вячеславович

доктор технических наук, профессор,
кафедра радиотехники, Саратовский
государственный технический
университет

E-mail: vyacheslav.komarov@gmail.com

Komarov Vyacheslav Vyacheslavovich

Doctor of engineering sciences, professor,
sub-department of radio engineering,
Saratov State Technical University

УДК 615.471/621.372.822

Новрузов, И. И.

Исследование электромагнитных и тепловых полей в ближней зоне контактного волноводного аппликатора / И. И. Новрузов, В. В. Комаров // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 151–158.

ПРИНЦИПЫ ПОСТРОЕНИЯ И АППАРАТУРНАЯ РЕАЛИЗАЦИЯ МНОГОФУНКЦИОНАЛЬНЫХ ФОТОТЕРАПЕВТИЧЕСКИХ УСТРОЙСТВ

Аннотация. Разработаны принципы построения многоцветного фототерапевтического аппарата широкого спектра лечебного действия, обеспечивающего возможность проведения светодиодного облучения по методикам фототерапии и фоторефлексотерапии. Многоцветность и широкий терапевтический спектр аппарата обеспечены техническими решениями авторов: применением трехцветных RGB-светодиодов, генерирующих красный, зеленый, синий цвета, и модулятора светодиодного излучения, генерирующего пространственно-модулированное светодиодное излучение, обладающее новыми биотропными характеристиками. Предложен вариант лечебно-диагностического комплекса, включающего фототерапевтический аппарат, для реализации принципа обратной связи с пациентом.

Ключевые слова: RGB-светодиоды, пространственная модуляция излучения, аппарат для фототерапии, лечебно-диагностический комплекс.

Abstract. The article is devoted to the multi-color phototherapeutic device with a wide range of medical effects and the principles of its construction. This device provides the possibility of carrying out light-emitting diode irradiation, using phototherapy and photoreflexotherapy. Polychromy and wide therapeutic application of the device under consideration are provided by the following technical decisions of the authors: the use of three-colored RGB-light-emitting diodes, together with a modulator of light-emitting diode radiation, which generates spatially modulated light-emitting diode radiation and possesses new biotropic characteristics. A variant of the medical-diagnostic complex is offered, including the phototherapeutic device to provide the feedback with patients.

Key words: RGB-light-emitting diodes, spatial modulation, device for phototherapy, medical-diagnostic complex.

Введение

В современной фототерапевтической практике наряду с лазерной терапией все чаще используются методы лечения с использованием некогерентных источников излучения, среди которых перспективным является применение излучающих светодиодов. Теоретическим обоснованием возможности замены излучения лазеров на излучение светодиодов явились работы [1–3] отечественных и зарубежных исследователей, где показана схожесть эффектов данных видов фотовоздействия на человеческий организм. Практическим обоснованием для использования светодиодов для терапии различных патологических состояний стал прорыв в производстве нитридных многокомпонентных гетероструктур и приборов на их основе. Это позволило решить и проблему низкого светового выхода, ограниченного диапазона цветов, что ранее препятствовало применению светодиодов для хромотерапии.

При разработке светодиодных терапевтических аппаратов основной проблемой является выбор вида светодиода для получения наибольшего терапевтического эффекта. Можно выделить несколько видов излучающих свето-

диодов: моргающие светодиоды, содержащие интегрированный кругооборот мультивибратора внутри, который заставляет светодиод вспыхивать с типичным периодом одной секунды; цветные мигающие светодиоды, состоящие из двух работающих навстречу светодиодов (электрический ток в одном направлении производит один цвет, ток в противоположном направлении производит другой цвет); трехцветные светодиоды (два светодиода в одном, имеющие общий анод или катод). В свою очередь среди трехцветных выделяют RGB-светодиоды, содержащие красный, зеленый и синий эмитенты. При соответствующем управлении RGB-светодиодом биологический объект можно облучать как красным, зеленым, синим лучами, обладающими наиболее выраженными биотропными свойствами, так и, при необходимости, лучами нескольких цветов, используя их смешивание.

Для повышения эффективности фототерапии необходимо генерировать светодиодное излучение с более выраженными биотропными параметрами. Кроме того, от качества и надежности модулятора излучения очень сильно зависят стабильность и долговечность работы фототерапевтического аппарата. Поэтому первоочередной задачей при разработке аппаратуры для светотерапии является поиск нового вида модуляции светодиодного облучения, разработка оптимальной конструкции модулятора излучения и блока управления цветовым спектром, разработка системы обратной связи с пациентом, оптимальный выбор точек и зон приложения лучей при различных заболеваниях.

Целью данной статьи является рассмотрение проблем, возникающих при разработке и построении фототерапевтических аппаратов на основе RGB-светодиодов, путей решения этих проблем и нахождения оптимальной конструкции модулятора излучения, блока управления цветовым спектром светодиодного аппарата и подключение аппарата к диагностической системе.

1. Основные проблемы, возникающие при модуляции светодиодного излучения

Модуляция излучения – это процесс изменения во времени мощности излучения (амплитудная модуляция), частоты (частотная) или фазы (фазовая). На практике используется только амплитудная модуляция. При этом режиме средняя мощность излучения меньше максимальной. В традиционных аппаратах частота модуляции излучения меняется обычно переключателем на панели блока управления и визуализируется цифровым индикатором. В некоторых случаях дискретный переключатель фиксируется в нужном положении с конкретным значением и размерностью параметра. Возможно управление длительностью импульсов в миллисекундах или микросекундах для модулированного режима.

При разработке узлов аппаратуры, обеспечивающих получение модулированного излучения в видимом диапазоне спектра, необходимо учитывать ряд особенностей, касающихся как особенностей функционирования организма при заболевании, так и характера взаимодействия излучения с биологическими тканями. В настоящее время многие производители фототерапевтической аппаратуры выпускают установки с неоправданно высокой выходной мощностью излучения, недостаточно подтвержденной экспериментально, хотя многими исследователями показано преимущество в эффективности модулированного излучения перед непрерывным [1, 4].

Следующая проблема касается частотной модуляции излучения. Предпосылкой для этого послужил тот факт, что модуляции луча определенной частоты лучше усваиваются теми или иными тканями организма. Поэтому подбор необходимых частот при светодиодной терапии является важным компонентом в лечебном процессе. Однако до сих пор еще не найдены оптимальные частоты модуляции при той или иной патологии организма. Часто предлагаемые режимы модуляции при ряде заболеваний экспериментально до конца не обоснованы.

Среди физических факторов воздействия давление света является одним из эффективных [1–3, 5]. Разработанный нами метод модуляции света базируется на усилении его пондеромоторного действия и реализуется при пространственном изменении плотности потока мощности излучения.

2. Особенности конструкции модулятора светодиодного аппарата

Модулятор светодиодного излучения должен обеспечивать: стабильные частоты пространственной модуляции излучения; варьирование частотного диапазона; оперативное управление величинами и диапазоном колебаний плотности падающей мощности излучения на объекте облучения.

В результате изучения литературы и существующих модуляторов излучения и сканирующих устройств авторы пришли к выводу, что для генерации пространственно модулированного луча наилучшим является модулятор лазерного излучения гинекологического аппарата «АГИН-01» (регистрационное удостоверение Росздравнадзора № ФСР 2009/04330), разработанного по патентам на изобретения и полезные модели профессора В. Н. Баранова [4, 6]. В аппарате «АГИН-01» световолокна, передающие излучение лазеров, под действием электродвигателя через сложную систему шестерен и направляющих стержней модулятора [4], совершают поступательные перемещения относительно объекта облучения. Перемещения световолокон происходят с частотой 1–3 Гц. Такое решение оказывается не очень удовлетворительным по следующим причинам: недостаточная частота модуляций; повышенный шум и вибрация при работе аппарата на частотах более 3 Гц; ненадежная фиксация световолокна в области направляющих стержней. Принимая во внимание эти недостатки, мы по согласованию с производителем аппарата «АГИН-01» световолокна прикрепили подвижно к периферии диска, который приводится в движение электродвигателем без использования шестерен. Диск при вращении перемещает световолокна в возвратно-поступательном направлении и не создает шума.

В результате модификации «АГИН-01» позволил получить частоты модуляции от 0,5 до 10,0 Гц. Расширившийся функциональный диапазон «АГИН-01» и высокая клиническая эффективность способствовали его распространению в клинических учреждениях г. Тюмени [4].

Следующим важным аспектом явилась разработка модулятора к светодиодному аппарату. В данном аппарате диск был закреплен непосредственно на оси электродвигателя, а на дистальном конце толкателя помещен RGB-светодиод типа ARL-5213RGB/4C. В данном модуляторе, так же как в «АГИН-01», используется механическое возвратно-поступательное перемещение излучателя. Такой подход открывает следующие возможности: регулирование диапазона перемещения светодиода относительно биологиче-

ского объекта; обеспечение высокой частоты модуляции излучения, так как источник излучения не связан с корпусом световолокном, связь между ними осуществляется только посредством тонкого сетевого провода; оптимального варьирования плотности падающей мощности излучения и спектра излучения светодиодов.

3. Особенности конструирования светодиодного аппарата

Сущность нового конструктивного решения светодиодного аппарата, обеспечивающего возможность воздействия на очаги поражения пространственно модулированным светодиодным излучением с длинами волн 0,62–0,64; 0,52–0,53; 0,46–0,47 мкм, поясняет рис. 1.

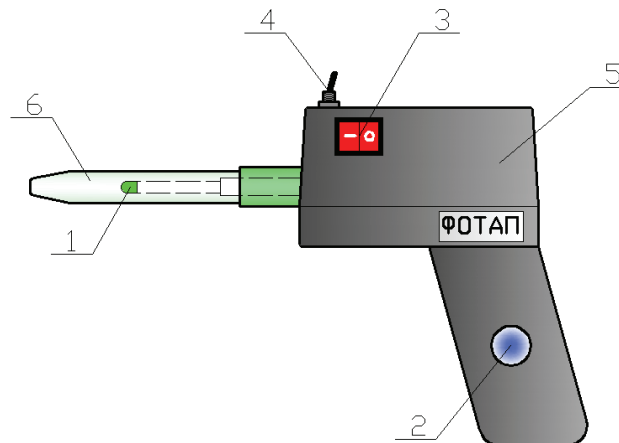


Рис. 1. Схема устройства для фототерапии: 1 – RGB-светодиод; 2 – ручка регулировки частоты модуляции излучения; 3 – выключатель аппарата; 4 – тумблер переключения цвета излучения; 5 – корпус аппарата; 6 – одноразовый наконечник

Как следует из рис. 1, аппарат состоит из трех основных частей: корпуса, в котором размещены электронный блок и светодиодный излучатель; одноразовой насадки; источника постоянного тока.

Светодиод (СД) генерирует оптическое излучение с длинами волн в диапазоне 0,4–0,6 мкм, которое воздействует на биологический объект (рис. 2). Поддержание определенного спектра излучения обеспечивается блоком управления цветовым спектром (БУЦС). В качестве блока питания (БП) применяется источник постоянного тока (9 В, 300 мА). Плата стабилизатора напряжения (ПСН) позволяет мотору аппарата формировать соответствующие клинической ситуации режимы пространственной модуляции светодиодного излучения.

Излучатель состоит из одного RGB-светодиода. Светодиод устанавливается на дистальном конце толкателя, который, в свою очередь, проксимальным концом на шарнире подвижно закреплен на периферии плоского диска, совершающего вращение вокруг своей вертикальной оси с регулируемой частотой. Светодиод с толкателем совершают поступательные движения в пластмассовой одноразовой наконечнике, который навинчивается основанием на резьбу корпуса.

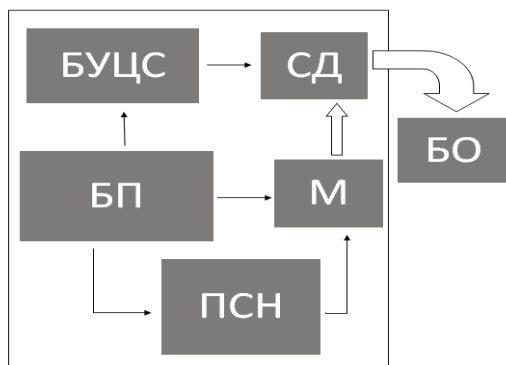


Рис. 2. Структурная схема аппарата: М – мотор; БО – биологический объект

При разработке опытного образца аппарата учтены требования устойчивости к температурным воздействиям, влажности и механической прочности. Кроме того, аппарат малогабаритен, имеет малую массу, простую и технологическую конструкцию, отвечающую требованиям электробезопасности.

Проведенное исследование показало, что при проведении процедур фототерапии мощность излучения светодиода меняется от 0,1 до 2,0 мВт, диаметр светодиодного пятна на объекте облучения – от 1,0 до 5,0 см, соответственно плотность падающей мощности – в среднем от 1,0 до 2,5 мВт на см², что не превышает рекомендуемые для фототерапии параметры воздействия. В результате аппаратной реализации модулятора аппарата удалось получить частоты пространственной модуляции светодиодного излучения от 0,5 до 15,0 Гц.

4. Аппаратурная реализация лечебно-диагностического комплекса

В работе предусмотрено подключение аппарата для фототерапии к диагностическому комплексу, состоящему из компьютера с автоматическим рабочим местом врача и диагностического блока с определителем локализации и функции биологически активных точек на основе величины их электрического сопротивления.

На рис. 3 представлена схема лечебно-диагностической системы, в которую включен светодиодный аппарат.

В задачу диагностического комплекса входит обработка информации, поступающей с приемников электронного зонда с биологически активных точек. В компьютерном блоке сигнал подвергается аналоговой обработке, в частности согласованию, усилению и фильтрации с последующей передачей на микроконтроллер. Аналоговая часть электронного блока состоит из пяти идентичных измерительных каналов, каждый из которых включает в себя такие основные узлы, как повторитель, усилитель и фильтр низких частот.

Для формирования полезного сигнала и устранения помех отраженного сигнала устанавливается фильтр нижних частот. Далее преобразование осуществляется с помощью схемы выборки и хранения, мультиплексора и АЦП, которые преобразуют исходные данные от одного или нескольких измерительных датчиков в выходной сигнал, пригодный для хранения в компьютере и цифровой обработки. Поскольку информация, поступающая на вход системы, преобразуется в цифровую форму и в дальнейшем подвергается вычислительной обработке, то во главе всей структурной схемы ставится централь-

ный микропроцессор, управляющий работой отдельных узлов системы и всеми процедурами, касающимися сбора, хранения и обработки информации. Затем информация поступает на персональный компьютер и отображается в удобной для пользователя форме. Разработка программного обеспечения проводилась с помощью компилятора IAR Embedded Workbench и программного симулятора AVR Studio 3.51, для программирования микроконтроллеров использовалась утилита PonyProg 2000.

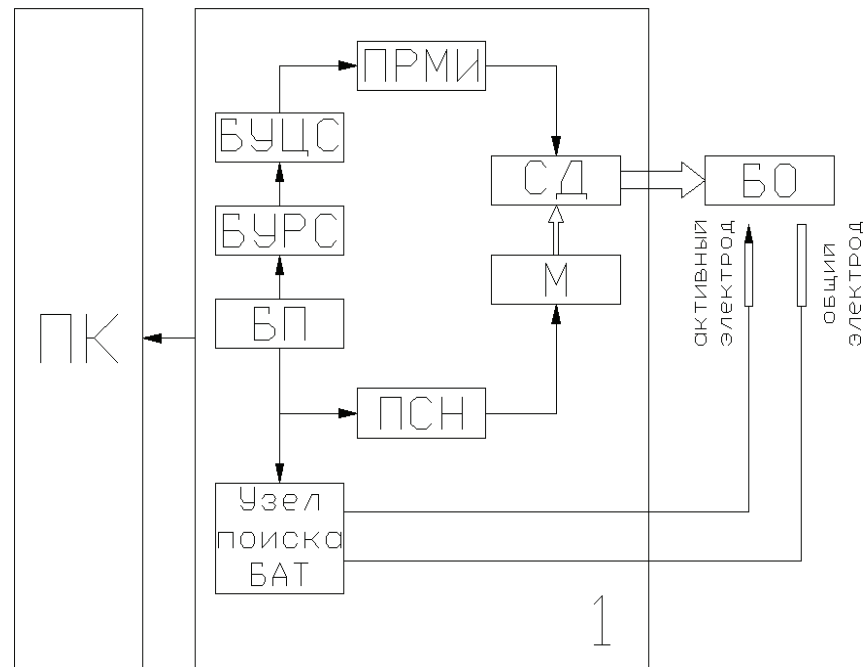


Рис. 3. Блок-схема аппарата с узлом поиска биологически активных точек:
1 – светодиодный аппарат; БУРС – блок управления режимом сканирования;
ПРМИ – плата регулировки мощности излучения; узел поиска БАТ – узел
поиска биологических активных точек; ПК – персональный компьютер

В результате мы пришли к схеме лечебно-диагностической системы, которая показана на рис. 4. Благодаря графической системе программирования на уровне функциональных блок-диаграмм LabView, данные визуализируются на мониторе ПК в более удобной форме для пользователя.

Заключение

Из проведенных теоретических, экспериментальных и клинических исследований можно заключить, что при построении аппаратов для фототерапии необходимо предусмотреть разработку эффективного модулятора излучения, обеспечивающего изменение интенсивности излучения во времени и пространстве для получения колебаний плотности мощности в биологических тканях ударного типа. Подобные колебания являются в определенной степени механостимулом, ведущим к активации механизмов саногенеза. Включение в конструкцию аппаратов светодиодов, излучающих на разных длинах волн, расширит функциональные возможности фототерапевтической аппаратуры.

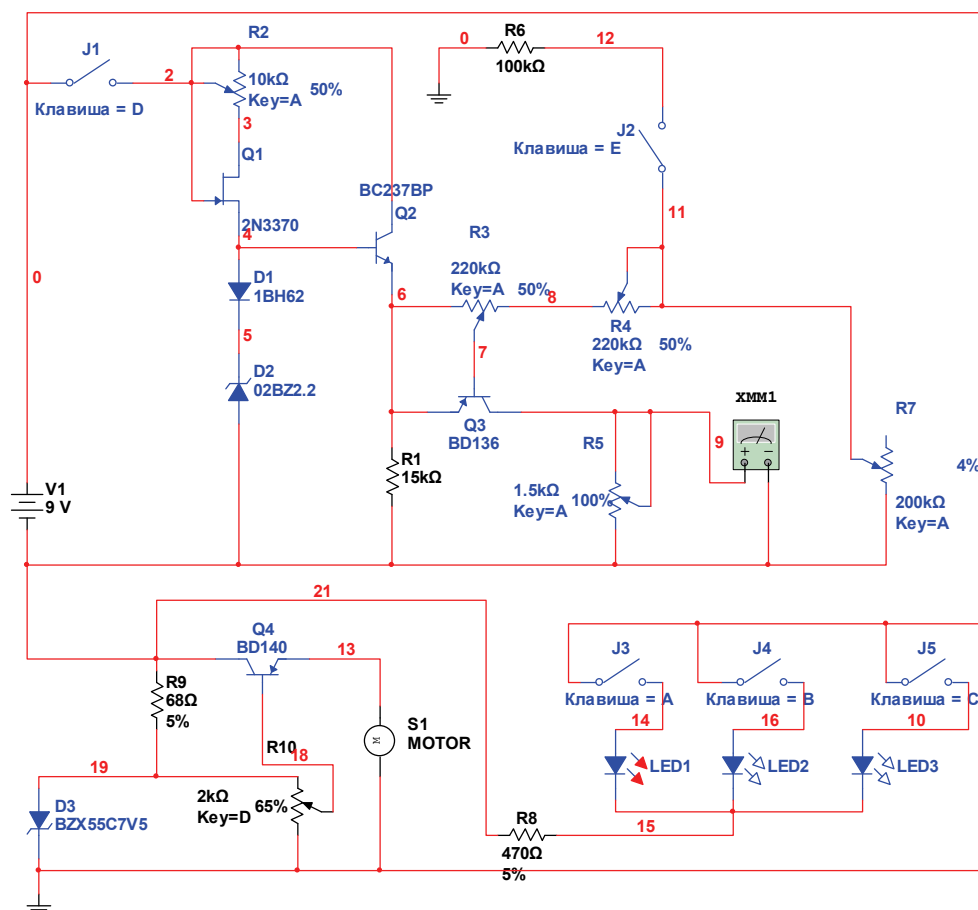


Рис. 4. Схема электрическая принципиальная разработанной лечебно-диагностической системы

Данному требованию отвечает пространственно модулированное излучение светодиодного аппарата, разработанного на основе RGB-светодиодов. Экспериментальные исследования макетного образца аппарата показали, что он обеспечивает широкий диапазон длин волн излучения: 0,62–0,64; 0,52–0,53; 0,46–0,47 мкм и управляемую пространственную модуляцию светодиодного излучения от 0,5 до 15,0 Гц. Подключение аппаратуры для фототерапии к диагностическому комплексу ускорит работу врача.

Список литературы

1. **Москвин, С. В.** Лазерная хромо- и светотерапия / С. В. Москвин, В. Г. Купеев. – М. ; Тверь : Триада, 2007. – 95 с.
2. **Romberg, H. A.** Properties of real phototherapy LED devices / H. A. Romberg, I. M. Heidelberg // Laser Helsinki 2010 : 15th International Congress of EMLA. – Helsinki, Finland, 2010. – S. 28–29.
3. **Makela, A. M.** Use of blue light and laser in the treatment of Alzheimer's dementia and Parkinson's disease / A. M. Makela // Laser Helsinki 2010 : 15th International Congress of EMLA. – Helsinki, Finland, 2010. – S. 26.
4. **Баранов, В. Н.** Повышение эффективности применения лазерного гинекологического аппарата «АГИН-01» в гинекологии с использованием метода пальце-

- вой фотоплетизмографии / В. Н. Баранов, Е. Л. Малиновский, В. А. Новиков, Т. В. Баимова, Р. Н. Хизбуллин // Казанский медицинский журнал. – 2010. – № 4. – С. 555–560.
5. **Горис, А. П.** Исследование деформируемости мембран эритроцитов методом «Лазерный пинцет» в условиях лактоацидоза / А. П. Горис, С. В. Москвин, Е. Г. Зарубина // Инновационные технологии в лазерной медицине : материалы научно-практической конференции. – М., 2011. – С. 116–117.
6. **Серов, В. Н.** Лазеротерапия в ранней реабилитации родильниц / В. Н. Серов, Ю. В. Кубицкая // Материалы IV съезда акушеров-гинекологов России. – М., 2008. – С. 237.
-

Хлынов Михаил Анатольевич

аспирант, Тюменский государственный
Нефтегазовый университет

E-mail: khlinov@mail.ru

Khlynov Mikhail Anatolyevich

Postgraduate student,
Tyumen State Oil and Gas University

Баранов Владимир Николаевич

доктор медицинских наук, профессор,
кафедра кибернетических систем,
Институт кибернетики, информатики
и связи, Тюменский государственный
Нефтегазовый университет

E-mail: kafedra-bmt@mail.ru

Baranov Vladimir Nikolaevich

Doctor of medical sciences, professor,
sub-department of cybernetic systems,
Institute of cybernetics, informatics
and communication, Tyumen State Oil
and Gas University

Рашев Ренат Рамазанович

аспирант, Тюменский государственный
Нефтегазовый университет

E-mail: renatik_72@mail.ru

Rashev Renat Ramazanovich

Postgraduate student,
Tyumen State Oil and Gas University

УДК 621.382.2+615.8-7

Хлынов, М. А.

Принципы построения и аппаратурная реализация многофункциональных фототерапевтических устройств / М. А. Хлынов, В. Н. Баранов, Р. Р. Рашев // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 159–166.

В. Н. Ашанин, Б. В. Чувькин, Р. Г. Тер-Аракелян

ОПТИМИЗАЦИЯ ПРОЕКТИРОВАНИЯ $\Sigma\Delta$ -МОДУЛЯТОРОВ ВЫСОКИХ ПОРЯДКОВ ПО КРИТЕРИЮ УСТОЙЧИВОСТИ

Аннотация. Описывается методика проектирования $\Sigma\Delta$ -модуляторов высоких порядков, основанная на аналитическом определении области значений коэффициентов обратной связи модуляторов для обеспечения устойчивого режима работы, расчета их значений с учетом допусковых значений и критерия минимума среднеквадратического отклонения шума квантования. Приводится пример проектирования $\Sigma\Delta$ -модулятора третьего порядка.

Ключевые слова: $\Sigma\Delta$ -АЦП, $\Sigma\Delta$ -модулятор, непрерывно-дискретная система, аналого-цифровой преобразователь, устойчивость, шум квантования.

Abstract. The article describes a designing method of the $\Sigma\Delta$ -modulators of high degrees, based on the analytical determination of the modulator feedback coefficients' values for sustainable operation mode. The authors have calculated coefficients' value, taking into account the tolerance values and the criterion of minimum standard deviation of the quantization noise. The researcher also provides an example of the $\Sigma\Delta$ -modulator design of the third degree.

Key words: $\Sigma\Delta$ -ADC, $\Sigma\Delta$ -modulator, analog-digital system, analog-to-digital converter, pulse-width modulator, sustainability, quantization.

Введение

Степень снижения погрешности от краевых эффектов $\Sigma\Delta$ -АЦП напрямую зависит от порядка (количества интеграторов) однобитного $\Sigma\Delta$ -модулятора и от коэффициента прореживания фильтра-дециматора, т.е. от его ширины полосы пропускания [1]. Однако увеличение порядка однобитного $\Sigma\Delta$ -модулятора, построенного на основе структур непрерывно-дискретных систем с финитной длительностью переходного процесса, приводит к снижению их устойчивости [2], а это приводит к уменьшению диапазона входного сигнала, а значит, и к снижению точности АЦП.

Устойчивое функционирование $\Sigma\Delta$ -модулятора обеспечивается при условии ограничения амплитуды входного сигнала тем большего, чем выше порядок $\Sigma\Delta$ -модулятора. Причем граница устойчивой работы $\Sigma\Delta$ -АЦП зависит от отношения шага дискретизации к постоянной времени интеграторов [2]. По мере уменьшения шага дискретизации по отношению к постоянной времени граница устойчивости приближается к пределу диапазона входного сигнала. Однако при этом длительность переходного процесса в $\Sigma\Delta$ -модуляторе увеличивается и он становится не финитным, а экспоненциально затухающим.

С целью нахождения предельных значений коэффициентов обратной связи $\Sigma\Delta$ -модуляторов высоких порядков, обеспечивающих максимально-возможный диапазон преобразуемого сигнала, целесообразно определить область их значений, соответствующую устойчивому режиму автоколебаний при нулевом входном сигнале.

1. Методика проектирования $\Sigma \Delta$ -модуляторов высоких порядков по критерию устойчивости

Сложность описания динамических процессов в $\Sigma \Delta$ -модуляторах связана с нелинейным преобразованием информации в однобитном квантователе. На практике чаще всего данная проблема решается методом имитационного моделирования с использованием известных алгоритмов оптимизации динамических процессов в многомерном пространстве дискретного времени. Формально задача сводится к решению систем нелинейных уравнений высокого порядка. На этом этапе необходимо аналитически записать уравнения состояний инерционных звеньев-интеграторов $\Sigma \Delta$ -модулятора, относящихся к преобразователям информации (ПИ) замкнутой структуры. Для решения этой задачи целесообразно использовать графо-аналитический метод [3].

Рассмотрим основные элементы и процедуры моделирования, необходимые для построения графа и перехода к системе нелинейных уравнений, описывающих динамические процессы в $\Sigma \Delta$ -модуляторе.

Если предположить, что в каждом цикле преобразования входная величина X_0 не изменяется, то в дискретной модели состояние интеграторов может быть задано в дискретные моменты времени, совпадающие с шагом дискретизации импульсного элемента (АЦП) и описывается разностным уравнением:

$$U_{k[n]} = U_{k[n-1]} + \varepsilon X_{[n-1]}, \quad (1)$$

где $U_{k[n]}$ – напряжение на выходе k -го интегратора в момент времени $t_n = nT_d$

(T_d – шаг дискретизации, n – номер шага); $\varepsilon = \frac{T_d}{\tau}$ – относительная постоянная времени интегратора; $X_{[n-1]}$ – напряжение на входе интегратора на предыдущем шаге дискретизации.

Переходя от итерационного уравнения к операторной форме, получим

$$U_k = U_k z^{-1} + \varepsilon X z^{-1}. \quad (2)$$

Это позволяет каждый интегратор заменить сумматором, который можно представить в виде графа, изображенного на рис. 1.

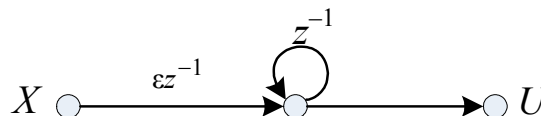


Рис. 1

В этом случае математическая модель $\Sigma \Delta$ -модулятора первого порядка в виде графа будет иметь вид, приведенный на рис. 2, где операция **fd** – дискретизация и формирование импульсов обратной связи, которая описывается выражением

$$y(t) = \begin{cases} U_{[n]} & t \in (T_{d(n-1)}; T_{d(n)}) \end{cases} \quad (3)$$

Для $\Sigma \Delta$ -модулятора с однобитной обратной связью выражение (3) примет следующий вид:

$$y(t) = \begin{cases} U_0 \text{sign}(U_{l[n]}) \end{cases} \quad (4)$$

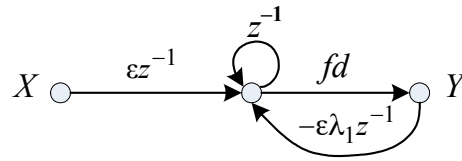


Рис. 2

Такой подход позволяет достаточно просто перейти от функциональной схемы ПИ с нелинейной обратной связью к соответствующему графу и системе разностных уравнений, описывающих динамические процессы в дискретном пространстве времени. Соответствующий переход от непрерывной модели ПИ к дискретной представлен в табл. 1.

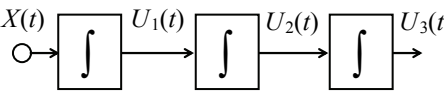
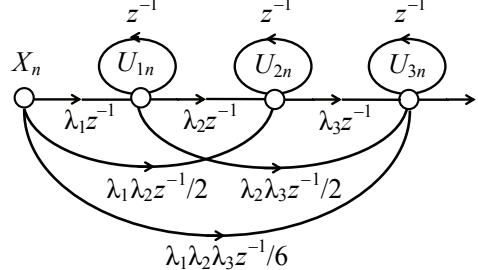
Особенностью этих структур является то, что реакция на входной сигнал «распространяется» по всей цепочке инерционных звеньев в прямой цепи преобразования. Это относится и к сигналам в цепи обратной связи.

Поясним на конкретном примере. Цепочка из k интеграторов, имеющих одинаковые постоянные времени τ , реагирует на входное ступенчатое воздействие X_0 в виде монотонно нарастающей функции, которая описывается параболой k -го порядка:

$$U_k(t) = \frac{t^k X_0}{k! \tau^k} \quad (5)$$

Таблица 1

Непрерывная модель	Дискретная модель
1	2
$U_{1(T_d n)} = U_{1(T_d(n-1))} + \frac{1}{\tau_1} \int_{T_d(n-1)}^{T_d n} X(t) dt$	$U_{1n} = U_{1n-1} + \lambda_1 \bar{X}_{n-1}, \text{ где } \lambda_1 = T_d / \tau_1$
$U_{2(T_d n)} = U_{2(T_d(n-1))} + \frac{1}{\tau_1 \tau_2} \times \int_{T_d(n-1)}^{T_d n} \left(\int X(t) dt \right) dt + \frac{T_d}{\tau_2} U_{1(T_d(n-1))}$	$U_{2n} = U_{2n-1} + \lambda_2 U_{1n-1} + \frac{\lambda_1 \lambda_2}{2} \bar{X}_{n-1},$ где $\lambda_2 = T_d / \tau_2$

1	2
 $U_{3(T_d n)} = U_{3(T_d(n-1))} +$ $+ \frac{1}{\tau_1 \tau_2 \tau_3} \int_{T_d(n-1)}^{T_d n} \int \int X(t) d^3 t +$ $+ \frac{(T_d)^2}{2 \tau_2 \tau_3} U_{1(T_d(n-1))} + \frac{T_d}{\tau_3} U_{2(T_d(n-1))}$	 $U_{3n} = U_{3n-1} + \lambda_3 U_{2n-1} + \frac{\lambda_2 \lambda_3}{2} U_{1n-1} +$ $+ \frac{\lambda_1 \lambda_2 \lambda_3}{6} \bar{X}_{n-1}, \text{ где } \lambda_3 = T_d / \tau_3$

Следовательно, в конце цикла преобразования $t = T_d$, и каждый из интеграторов находится в состоянии

$$U_{k[1]} = \frac{\varepsilon^k X_0}{k!}. \quad (6)$$

Используя принцип суперпозиции, для следующего такта преобразования можно воспользоваться уравнением (6), если допустить, что каждый из интеграторов имеет ненулевое начальное состояние, равное конечному состоянию предыдущего такта. Рассматривая X_0 как входной сигнал, приложенный к первому интегратору или сигнал обратной связи, приложенный к соответствующему интегратору, и учитывая масштабные коэффициенты обратной связи $\lambda_1, \lambda_2, \dots, \lambda_k$, можно получить систему разностных уравнений в общем виде:

$$\left\{ \begin{array}{l} U_{1[n]} = U_{1[n-1]} + \varepsilon X_{[n-1]} - \varepsilon \lambda_1 U_{OC[n-1]}, \\ U_{2[n]} = U_{2[n-1]} + \varepsilon U_{1[n-1]} - \varepsilon \lambda_2 U_{OC[n-1]} - \frac{\varepsilon^2}{2!} \lambda_1 U_{OC[n-1]} + \frac{\varepsilon^2}{2!} X_{[n-1]}, \\ \vdots \\ U_{k[n]} = U_{k[n-1]} + \varepsilon U_{k-1[n-1]} + \frac{\varepsilon^2}{2!} U_{k-2[n-1]} + \dots + \frac{\varepsilon^{k-1}}{(k-1)!} U_{1[n-1]} - \\ - \varepsilon \lambda_k U_{OC[n-1]} - \frac{\varepsilon^{k-1}}{(k-1)!} \lambda_{k-1} U_{OC[n-1]} - \dots - \lambda_1 \frac{\varepsilon^k}{k!} U_{OC[n-1]} + \frac{\varepsilon^k}{k!} X_{[n-1]}. \end{array} \right. \quad (7)$$

Поскольку $\Sigma \Delta$ -модулятор является замкнутой нелинейной системой, функциональная связь между сигналом обратной связи и сигналом с выхода k -го интегратора определяется в зависимости от вида нелинейности, следова-

тельно, система из k -уравнений (7) должна быть дополнена следующим уравнением:

$$U_{OC[n-1]} = U_0 \cdot \text{sign}(U_{k[n-1]}). \quad (8)$$

В работе [4] показано, что при нулевом входном сигнале $\Sigma \Delta$ -модулятор k -го порядка может находиться в устойчивом автоколебательном состоянии с периодом колебаний, равном 2^k тактам. Особенностью этих периодических колебаний является то, что знаки в однобитной последовательности выходного сигнала $Y[n]$ в течение периода чередуются аналогично знакам коэффициентов полинома вида

$$(1-z) \cdot (1-z^2) \cdot \dots \cdot (1-z^{2^{k-1}}) = 1 - z - z^2 + z^3 - z^4 + z^5 + z^6 - z^7 - z^8 + \dots (-1)^k z^{2^k-1}. \quad (9)$$

Закономерность (9) позволяет для нулевого входного сигнала определить область значений коэффициентов λ , приведенных в табл. 2, которые соответствуют устойчивому режиму работы $\Sigma \Delta$ -модулятора.

Таблица 2

Знаки коэффициентов a_i полинома: $P_k(z) = (1-z) \cdot (1-z^2) \cdot \dots \cdot (1-z^{2^{k-1}}) = 1 - z^1 - z^2 + \dots + a_i z^i \dots$																
k	$Y_i = \text{sign}(a_i)$															
	$i=0$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$	$i=9$	$i=10$	$i=11$	$i=12$	$i=13$	$i=14$	$i=15$
1	+	-														
2	+	-	-	+												
3	+	-	-	+	-	+	+	-								
4	+	-	-	+	-	+	+	-	-	+	+	-	+	-	-	+

При этом выходной сигнал $Y_k[n]$ однобитного $\Sigma \Delta$ -модулятора k -го порядка в режиме преобразования сигналов малого уровня описывается уравнением

$$Y_k[n+1] = \text{sign}(e_1 B_1[n] + e_2 B_2[n] + \dots + e_k B_k[n]),$$

где $B_1[n] = \frac{U_0}{\tau} \int_0^{T_d n} Y(t) dt$, $B_2[n] = \frac{U_0}{\tau^2} \int_0^{T_d n} \int Y(t) d^2 t$, ..., $B_k[n] = \frac{U_0}{\tau^k} \int_0^{T_d n} \dots \int_k Y(t) d^k t$ –

реакции интеграторов по k -й цепи обратной связи; U_0 – амплитуда импульсов обратной связи; τ – постоянная времени интеграторов; $e_1, e_2, \dots, e_k$ – коэффициенты обратной связи.

Данный метод может быть использован и в более сложных случаях нелинейной функции обратной связи, однако нужно учитывать, что если информативный параметр сигнала обратной связи не амплитуда, а время (ШИМ-сигнал), частота, фаза, то реакция интеграторов на сигнал обратной связи будет описываться нелинейной функцией, одним из параметров кото-

рой будет информативная составляющая сигнала (время, фаза). Например, если информативным параметром сигнала обратной связи является длительность импульса при неизменной амплитуде, то реакция интеграторов на такой сигнал будет нелинейной функцией (начиная со второго интегратора). В этом случае каждый коэффициент обратной связи $\lambda_1, \lambda_2, \dots, \lambda_k$ будет функционально связан с амплитудным значением выходного сигнала [5, 6]. Это количественно усложняет граф и систему уравнений, однако качественно не меняет подход, позволяя моделировать любые нелинейности, а также любые прямые и обратные связи в цепи преобразования [1].

2. Проектирование $\Sigma \Delta$ -модулятора третьего порядка

Рассмотрим предлагаемую методику проектирования по критерию устойчивости на примере $\Sigma \Delta$ -модулятора третьего порядка.

На основе графо-аналитического метода составим систему уравнений, описывающую состояния интеграторов $\Sigma \Delta$ -модулятора. Математическая модель $\Sigma \Delta$ -модулятора третьего порядка в виде графа приведена на рис. 3.

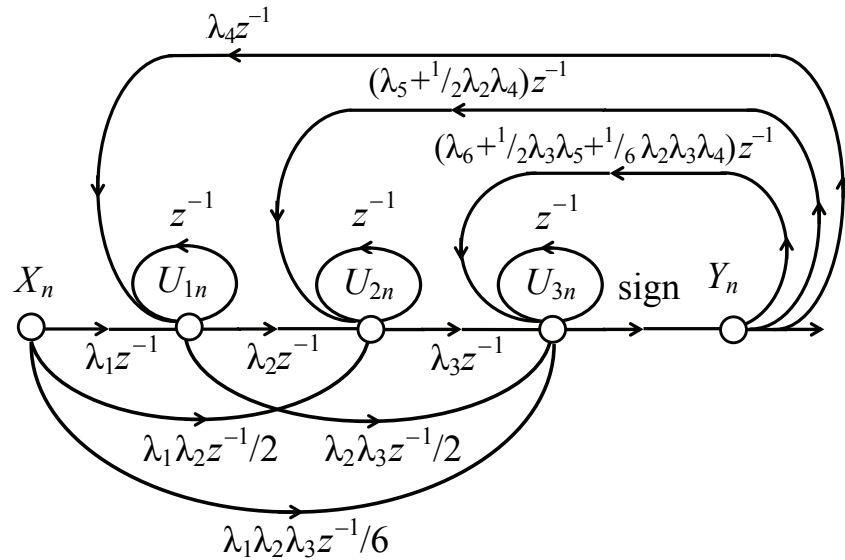


Рис. 3

В соответствии с графом система уравнений для выходных напряжений интеграторов имеет вид

$$\begin{cases} U_{1n} = U_{1n-1} + \lambda_1 X_{n-1} + \lambda_4 Y_{n-1}, \\ U_{2n} = U_{2n-1} + \lambda_2 U_{1n-1} + \frac{\lambda_1 \lambda_2}{2} X_{n-1} + \left(\lambda_5 + \frac{\lambda_1 \lambda_2}{2} \right) Y_{n-1}, \\ U_{3n} = U_{3n-1} + \lambda_3 U_{2n-1} + \frac{\lambda_2 \lambda_3}{2} U_{1n-1} + \frac{\lambda_1 \lambda_2 \lambda_3}{6} X_{n-1} + \\ + \left(\lambda_3 + \frac{\lambda_2 \lambda_3}{2} + \frac{\lambda_1 \lambda_2 \lambda_3}{6} \right) Y_{n-1}, \\ Y_n = \text{sign}(U_{3n}). \end{cases} \quad (10)$$

Масштаб преобразования $\Sigma\Delta$ -модулятора определяется отношением $Y/X = \lambda_4/\lambda_1 = \tau_1/\tau_4$. Зададим значение масштаба равным единице, тогда $\tau_1 = \tau_4$. Масштаб преобразования для первой петли обратной связи равен $e_1 = \lambda_2\lambda_3\lambda_4$, для второй $e_2 = \lambda_3\lambda_5$ и для третьей $e_3 = \lambda_6$. С учетом этого система уравнений модулятора принимает вид

$$\begin{cases} U_{1n} = U_{1n-1} + e_1(X_{n-1} + Y_{n-1}), \\ U_{2n} = U_{2n-1} + U_{1n-1} + e_2Y_{n-1} + \frac{e_1}{2}(X_{n-1} + Y_{n-1}), \\ U_{3n} = U_{3n-1} + U_{2n-1} + e_3Y_{n-1} + \frac{1}{2}(U_{1n-1} + e_2Y_{n-1}) + \frac{e_1}{6}(X_{n-1} + Y_{n-1}), \\ Y_n = \text{sign}(U_{3n}). \end{cases} \quad (11)$$

Соответствующая *Simulink*-модель $\Sigma\Delta$ -модулятора представлена на рис. 4.

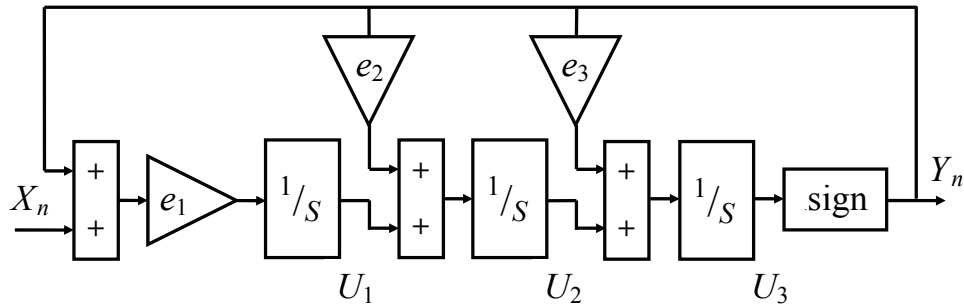


Рис. 4

Выходное напряжение $\Sigma\Delta$ -модулятора третьего порядка описывается выражением

$$U_{3n} = A_n e_3 + B_n e_2 + C_n e_1 + U_0,$$

а рассчитанные значения реакций интеграторов по k -й цепи обратной связи при условии $\tau = T_d$ приведены в табл. 3.

Таблица 3

	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 5$	$n = 6$	$n = 7$	$n = 8$
A_n	1	0	-1	0	1	0	-1	0
B_n	1/2	1	1/2	0	-1/2	-1	-1/2	0
C_n	1/6	1	11/6	2	11/6	1	1/6	0

Для расчета устойчивости $\Sigma\Delta$ -модулятора определим область значений e_1, e_2, e_3 . В соответствии с (11) составим систему неравенств состояния интегратора U_{3n} при условии, что его начальное состояние $U_{30} > 0$:

$$U_{31} < 0; U_{32} < 0, U_{33} > 0, U_{34} < 0, U_{35} > 0, U_{36} > 0, U_{37} < 0. \quad (12)$$

Примем, что $U_{10} = 0, U_{20} = 0, U_{30} = c$, где $c \rightarrow 0$, и рассчитаем значения $U_{31}, U_{32}, \dots, U_{37}$ согласно системе уравнений (11). Сделав подстановку в (12), получаем систему неравенств относительно коэффициентов e_1, e_2, e_3 :

$$\begin{cases} U_0 > 0, \\ 1/6e_1 + 1/2e_2 + e_3 + U_0 < 0, \\ e_1 + e_2 + U_0 < 0, \\ 11/6e_1 + 1/2e_2 + e_3 + U_0 > 0, \\ 2e_1 + U_0 < 0, \\ 11/6e_1 - 1/2e_2 - e_3 + U_0 > 0, \\ e_1 - e_3 + U_0 > 0, \\ 1/6e_1 - 1/2e_2 + e_3 + U_0 < 0. \end{cases} \quad (13)$$

Упрощая систему неравенств (13), получим

$$\begin{cases} 11/6e_1 + 1/2e_2 > e_3, \\ e_1 > e_3. \end{cases} \quad (14)$$

Система неравенств (13) позволяет ограничить область поиска оптимальных значений коэффициентов e_1, e_2, e_3 по критерию минимума средне-квадратического отклонения (СКО) шума квантования с учетом интервалов допусковых отклонений и тем самым существенно уменьшить объем численного моделирования.

Зададим диапазон изменения входного сигнала в интервале $\pm 0,5U_{OC}$ и рассчитаем значения e_i из области значений, ограниченных системой неравенств (14), которые соответствуют устойчивому режиму работы модулятора. На рис. 5 приведены результаты численного расчета коэффициентов e_1 и e_2 при условии, что $e_3 = 1$. Двумя прямыми ограничена область допустимых значений коэффициентов e согласно полученным неравенствам (12). За- темненная область соответствует неустойчивому режиму работы модулятора.

Как видно, подмножество значений коэффициентов e_1, e_2, e_3 , соответствующих устойчивой динамике переходных процессов для диапазона входного сигнала $\pm 0,5U_0$, совпадает с множеством значений e_1, e_2, e_3 для нулевого входного сигнала.

Зададим интервалы отклонений значений e_1, e_2, e_3 ($\pm 0,01$) и найдем оптимальные значения коэффициентов e_1, e_2, e_3 по критерию минимума СКО шума квантования. В качестве цифрового фильтра-дециматора будем использовать цифровой фильтр со сплайновой весовой функцией третьего порядка [7]. Результат расчета приведен на рис. 6.

Последним этапом проектирования является решение задачи оптимизации динамического диапазона выходных сигналов интеграторов путем изменения значений постоянных времени $\tau_1, \tau_2, \tau_3, \tau_4, \tau_5, \tau_6$. Критерием оптимизации будет условие, что динамические диапазоны изменения выходных сигналов интеграторов равны друг другу. Эта задача решается путем анализа работы *Simulink*-модели для заданного диапазона входных сигналов из условия,

что максимальные напряжения на выходах интеграторов равны друг другу и не превышают напряжение питания.

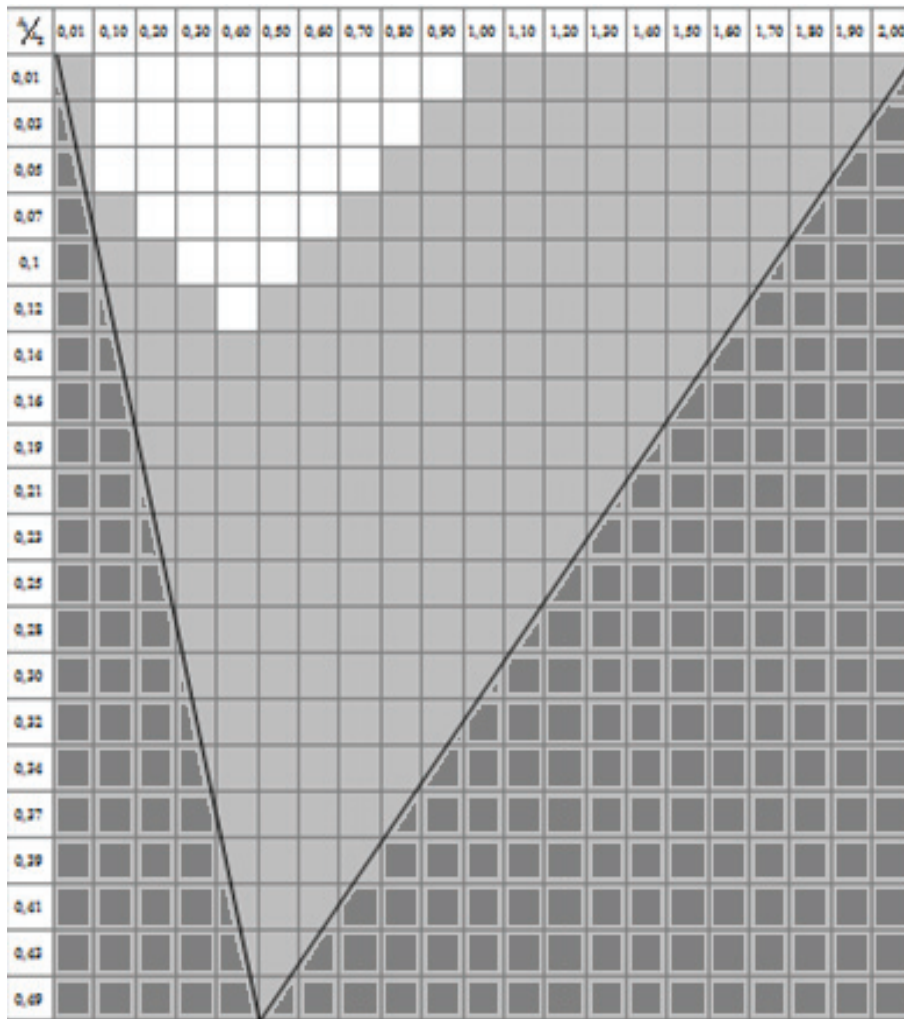


Рис. 5

Для описания динамики состояний интеграторов по функциональной схеме (рис. 4) составим *Simulink*-модель модулятора (рис. 7).

Упрощенная принципиальная схема модулятора (рис. 8) содержит цепочку инвертирующих *RC*-интеграторов, компаратор (К), генератор импульсов (G), *D*-триггер, аналоговые ключи Кл₁, Кл₂ и источник опорного напряжения ($\pm U$).

RC-параметры схемы модулятора и параметры τ , λ *Simulink*-модели связаны следующими соотношениями:

$$R_1 C_1 = \tau_1, R_2 C_1 = \tau_2, R_3 C_2 = \tau_3, R_4 C_2 = \tau_4, R_5 C_3 = \tau_5, R_6 C_3 = \tau_6;$$

$$\lambda_1 = \frac{T_d}{R_1 C_1}; \lambda_2 = \frac{T_d}{R_2 C_2}; \lambda_3 = \frac{T_d}{R_3 C_3}; \lambda_4 = \frac{T_d}{R_4 C_1}; \lambda_5 = \frac{T_d}{R_5 C_2}; \lambda_6 = \frac{T_d}{R_6 C_3}.$$

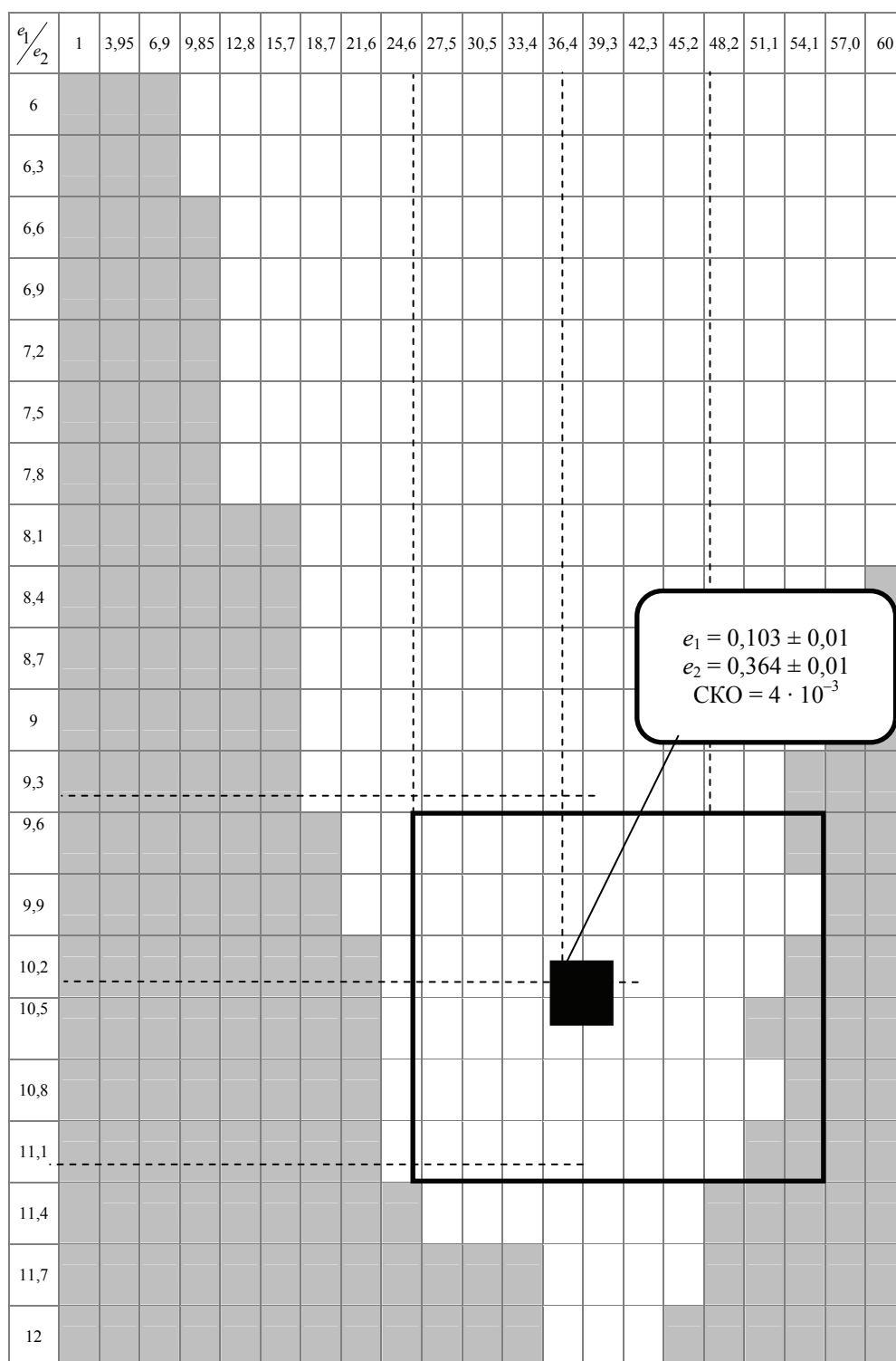


Рис. 6

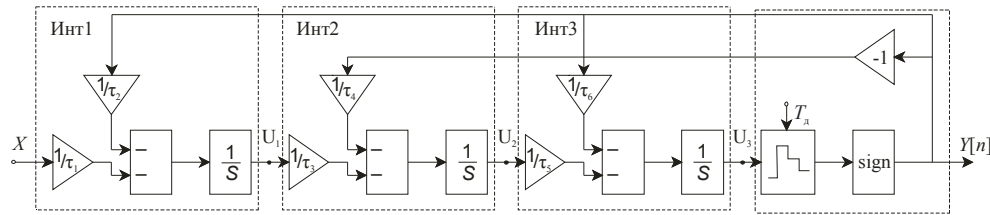


Рис. 7

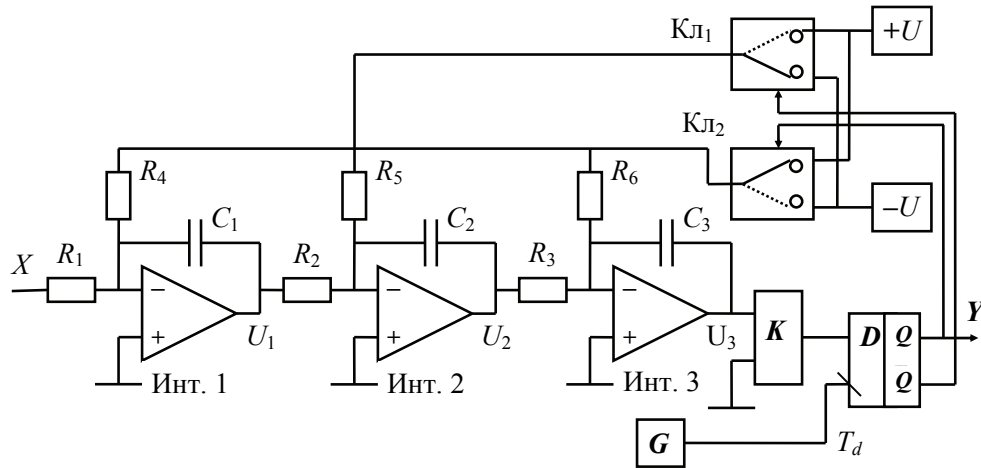


Рис. 8

Система уравнений для выходных напряжений интеграторов модулятора принимает вид

$$\begin{cases} U_{1n} = U_{1n-1} - \frac{T_d}{R_1 C_1} X_{n-1} - \frac{T_d}{R_4 C_1} Y_{n-1} U, \\ U_{2n} = U_{2n-1} - \frac{T_d}{R_2 C_2} U_{1n-1} + \frac{(T_d)^2}{2R_1 R_2 C_1 C_2} X_{n-1} - \left(\frac{T_d}{R_5 C_2} - \frac{(T_d)^2}{2R_2 R_4 C_1 C_2} \right) Y_{n-1} U, \\ U_{3n} = U_{3n-1} - \frac{T_d}{R_3 C_3} U_{2n-1} + \frac{(T_d)^2}{2R_2 R_3 C_2 C_3} U_{1n-1} - \frac{(T_d)^3}{6R_1 R_2 R_3 C_1 C_2 C_3} X_{n-1} - \\ - \left(\frac{T_d}{R_6 C_3} - \frac{(T_d)^2}{2R_3 R_5 C_2 C_3} + \frac{(T_d)^3}{6R_2 R_3 R_4 C_1 C_2 C_3} \right) Y_{n-1} U, \\ Y_n = \text{sign}(U_{3n}). \end{cases}$$

Заключение

Таким образом, предлагаемая методика проектирования $\Sigma\Delta$ -модуляторов высоких порядков по критерию устойчивости предполагает выполнение следующих этапов:

1. Запись системы нелинейных уравнений на основе графо-аналитического метода, описывающих состояние интеграторов $\Sigma\Delta$ -модуляторов в дискретные моменты времени.
2. Задание начальных условий интеграторов, входных тестовых сигналов, количества циклов преобразования.
3. Аналитическое решение систем неравенств для нулевого входного сигнала и определение границы области допустимых значений коэффициентов обратной связи.
4. Проведение численных расчетов коэффициентов обратной связи в области их допустимых значений для заданных тестовых входных сигналов с целью исключения значения коэффициентов, при которых на заданном множестве циклов преобразования наблюдается неустойчивая динамика модулятора.
5. Нахождение области оптимальных значений коэффициентов обратной связи по минимуму СКО шума квантования с учетом интервалов их отклонений.
6. Анализ работы *Simulink*-модели с целью оптимизации постоянных времени интеграторов по критерию обеспечения максимального динамического диапазона их выходных сигналов.
7. Разработка *Simulink*-модели и функциональной схемы $\Sigma\Delta$ -модулятора.

Список литературы

1. **Ашанин, В. Н.** $\Sigma\Delta$ -аналого-цифровые преобразователи: основы теории и проектирование / В. Н. Ашанин, Б. В. Чувькин, Э. К. Шахов. – Пенза : Информационно-издательский центр ПГУ, 2009. – 188 с.
2. **Ашанин, В. Н.** $\Sigma\Delta$ -АЦП: анализ погрешности от краевых эффектов / В. Н. Ашанин, Б. В. Чувькин, Э. К. Шахов // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2007. – № 3. – С. 80–90.
3. **Чувькин, Б. В.** Графо-аналитический метод моделирования $\Sigma\Delta$ -преобразователей информации замкнутой структуры / Б. В. Чувькин, А. В. Селезнев, Р. Г. Тер-Аракелян // Перспективные информационные технологии для авиации и космоса (ПИТ-2010) : труды Международной конференции. – Самара, 2010. – С. 60–65.
4. **Сидорова, И. А.** Анализ флуктуационных шумов квантования $\Sigma\Delta$ -АЦП / И. А. Сидорова, И. А. Долгова, Р. Г. Тер-Аракелян // Наука и современность – 2011 : труды XI Международной научно-практической конференции. – Новосибирск, 2011. – С. 132–135.
5. **Ашанин, В. Н.** Синтез $\Sigma\Delta$ -модулятора с многоуровневым ЦАП на основе широкоимпульсной модуляции сигнала / В. Н. Ашанин, Б. В. Чувькин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2010. – № 4 (16). – С. 97–105.
6. **Ашанин, В. Н.** Синтез $\Sigma\Delta$ -АЦП с многоуровневым ЦАП на основе широкоимпульсного модулятора / В. Н. Ашанин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 1 (15). – С. 132–137.
7. **Schreier, R.** An Empirical Study of High-Order Single-Bit Delta-Sigma Modulators / R. Schreier // IEEE Transaction on Circuits and Systems-II. – 1993. – V. 40, № 8. – P. 461–466.

Ашанин Василий Николаевич

кандидат технических наук, профессор,
заведующий кафедрой электротехники
и транспортного электрооборудования,
Пензенский государственный
университет

E-mail: eltech@pnzgu.ru

Ashanin Vasily Nikolaevich

Candidate of engineering sciences, professor,
head of sub-department of electrical
engineering and transport electrical
equipment, Penza State University

Чувькин Борис Викторович

доктор технических наук, профессор,
кафедра информационных
вычислительных систем, Пензенский
государственный университет

E-mail: Chuvykin_BV@mail.ru

Chuvykin Boris Viktorovich

Doctor of engineering sciences, professor,
sub-department of information-computing
systems, Penza State University

Тер-Аракелян Руслан Геворкович

аспирант, Пензенский
государственный университет

E-mail: rt-a@mail.ru

Ter-Arakelyan Ruslan Gevorkovich

Postgraduate student,
Penza State University

УДК 621.3.087.92

Ашанин, В. Н.

Оптимизация проектирования $\Sigma\Delta$ -модуляторов высоких порядков по критерию устойчивости / В. Н. Ашанин, Б. В. Чувькин, Р. Г. Тер-Аракелян // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 167–179.

МАШИНОСТРОЕНИЕ И МАШИНОВЕДЕНИЕ

УДК 532.66:532.528

В. Д. Кревчик, Н. П. Симонов, А. В. Соколов

КАВИТАЦИОННЫЙ МЕХАНИЗМ УПРАВЛЕНИЯ КАЧЕСТВОМ ПОВЕРХНОСТНОГО СЛОЯ ДЕТАЛЕЙ ПРИ ИСПОЛЬЗОВАНИИ СМАЗЫВАЮЩЕ-ОХЛАЖДАЮЩИХ ЖИДКОСТЕЙ С НАНОДОБАВКАМИ В УСЛОВИЯХ ЗВУКОКАПИЛЛЯРНОГО ЭФФЕКТА

Аннотация. Теоретически исследован кавитационный механизм управления эффективным модулем Юнга поверхностного слоя металла с цепочкой микротрещин при наличии смазывающе-охлаждающих жидкостей с нанодобавками. Проведена модификация формулы Кенига для случая парного взаимодействия наночастиц с учетом дисперсии их характерного размера. Предложена одномерная модель поверхностного слоя металла с цепочкой микротрещин. Показано, что в случае достаточно плотной регулярной цепочки микротрещин эффективный модуль Юнга поверхностного слоя определяется в основном модулем Юнга материала наночастиц.

Ключевые слова: кавитационный механизм, поверхностный слой металла, микротрещина, эффективный модуль Юнга поверхностного слоя, наночастицы, звукокапиллярный эффект.

Abstract. The authors have theoretically investigated a cavitation mechanism of the effective Young's modulus control for the surface layer of metal with the chain of microcracks under the influence of lubricating – cooling liquids with nano-additives. The researchers have modified the Koenig formula for the case of pair interaction of nanoparticles with account of their characteristic size dispersion. The article suggests a 1D – model of the metal surface layer with the chain of microcracks. It is shown that in case of sufficiently dense regular chain of microcracks the effective Young's modulus for surface layer is mainly determined by the Young's modulus for the nanoparticles material.

Key words: cavitation mechanism, surface layer of metal, microcrack, effective Young's modulus of the surface layer, nanoparticles, acoustocapillary effect.

Введение

Качество поверхностного слоя деталей машин во многом определяет их эксплуатационные характеристики и в большинстве случаев формируется при шлифовании. Хорошо известно [1, 2], что данный процесс характеризуется высокой тепловой напряженностью и большой вероятностью появления в поверхностных слоях дефектов в виде микротрещин (МТ), снижающих их эксплуатационные свойства. В настоящее время находят широкое применение кавитационные технологии для интенсификации шлифования [3] и нанотехнологии в приготовлении смазывающе-охлаждающих жидкостей (СОЖ) [4].

Однако сами по себе данные подходы не решают более общей проблемы – проблемы управления качеством поверхностного слоя деталей. В настоящей работе предпринята попытка теоретического обоснования возможности управления эффективным модулем Юнга поверхностного слоя металла с цепочкой МТ посредством кавитационного механизма звукокапиллярного эффекта с участием наночастиц (НЧ) металла, модуль Юнга которого можно варьировать в широких пределах, выбирая соответствующий тип металла. В разд. 1 статьи теоретически исследуется воздействие ультразвука на СОЖ в МТ с участием НЧ металла в режиме кавитации. Кавитация представляет собой средство локальной концентрации энергии низкой плотности в высокую плотность энергии, связанную с пульсациями и захлопыванием кавитационных пузырьков.

Кавитация играет важную роль в эффекте проникновения нанодопированной СОЖ в каналы МТ. Возникновение кавитационного облачка в объеме СОЖ обусловлено в основном следующими причинами: наличием в поверхностном слое МТ, в которых заземлен воздух и которые являются источниками зародышей кавитации; увеличением концентрации зародышей кавитации вследствие того, что пузырьки при пульсациях и захлопывании у твердой стенки, т.е. у торца МТ, движутся к данному торцу. Кроме того, при отражении звука от препятствия с большим волновым сопротивлением, чем волновое сопротивление СОЖ, давление не изменяет фазу, и в результате суперпозиции падающей и отраженной волн на поверхности такого препятствия образуется максимум звукового давления [5]. На рис. 1 приведена качественная картина попадания кумулятивной струи в канал МТ.

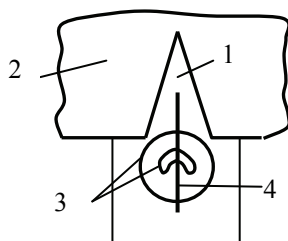


Рис. 1. Качественная картина попадания кумулятивной струи в канал МТ:
1 – МТ; 2 – поверхностный слой металла; 3 – поверхность кавитационного пузырька в различные моменты времени; 4 – кумулятивная струя

Теоретическое рассмотрение звукокапиллярного эффекта с кавитационным механизмом в нанодопированной СОЖ позволило получить аналитическую формулу для зависимости числа НЧ от параметров МТ, кавитационной области и СОЖ. В разд. 2 статьи рассмотрено взаимодействие НЧ в кластере в поле ультразвуковой волны. Проведена модификация формулы Кенига [6] на случай парного приближения с учетом дисперсии характерного размера НЧ.

Исследовано влияние дисперсии и расстояния между НЧ на силу их взаимодействия в ультразвуковом поле. В разд. 3 данной статьи предложена одномерная модель поверхностного слоя металла с цепочкой МТ. В рамках данной модели получено аналитическое выражение для эффективного модуля Юнга поверхностного слоя металла и исследована его зависимость от числа НЧ в МТ.

1. Воздействие ультразвука на СОЖ в микротрещинах при наличии кавитации

Пусть МТ достаточно удалена от излучателя акустических волн, т.е. $\delta \gg D_{\max}$ (δ – длина зазора между МТ и излучателем, $D_{\max}$ – максимальный диаметр кавитационного пузырька), так что влиянием поверхности излучателя на захлопывание пузырька у торца МТ можно пренебречь и считать, что пузырек захлопывается с образованием кумулятивной струи, направленной к торцу МТ. При попадании струи в канал МТ высота подъема СОЖ изменяется на величину ΔH_n . При этом энергия струи сообщается столбику СОЖ в МТ и расходуется на увеличение потенциальной энергии ΔU столбика СОЖ с находящимися в нем НЧ вследствие увеличения высоты и работу против сил трения $\Delta A_{\text{тр}}$:

$$\Delta E_C = \Delta A_{\text{тр}} + \Delta U. \quad (1)$$

Пусть ΔP_0 – избыточное давление над мениском, которым мениск удерживается в среднем на неизменном уровне H_0 . Тогда приращение потенциальной энергии в результате увеличения подъема имеет вид

$$\Delta U = \pi r_k^2 \cdot \left[\int_0^{H_0 + \Delta H_n} dH \left(\Delta P_0 + \rho_C gH + \rho_H gH + N_H \frac{F_{\text{int}}}{\pi r_k^2} \right) - \int_0^{H_0} dH \left(\Delta P_0 + \rho_C gH + N_H \frac{F_{\text{int}}}{\pi r_k^2} \right) \right], \quad (2)$$

где r_k – эффективный радиус МТ у ее торца; ρ_C – плотность СОЖ; ρ_H – плотность материала НЧ; F_{int} – сила взаимодействия НЧ в поле ультразвуковой волны; N_H – число НЧ в объеме МТ; P_0 – атмосферное давление.

Пренебрегая в (2) слагаемыми, пропорциональными $(\Delta H_n)^2$ и H_0^2 , после интегрирования получим

$$\Delta U \approx \pi r_k^2 \Delta H_n \left[\Delta P_0 + (\rho_C + \rho_H) g H_0 + N_H \frac{F_{\text{int}}}{\pi r_k^2} \right]. \quad (3)$$

Работа сил трения при подъеме СОЖ с НЧ в МТ со средней скоростью V_n , определяется как

$$\Delta A_{\text{тр}} = \left(\frac{\rho_C}{4} \cdot \pi r_k \lambda H_0 \cdot V_n^2 + \mu g \frac{4}{3} \pi R_0^3 \rho_H \cdot N_H \right) \Delta H_n, \quad (4)$$

где λ – постоянный коэффициент, зависящий от числа Рейнольдса: $\lambda = 64 / \text{Re}$; R_0 – радиус НЧ; μ – коэффициент трения, возникающего при подъеме НЧ по каналу МТ.

Подставляя (3) и (4) в (1) и учитывая, что ΔE_C имеет порядок потенциальной энергии кавитационного пузырька $\Delta W = 4\pi R_{\max}^3 \cdot P_0 / 3$, а также, что вытекание СОЖ из МТ происходит в течение периода ультразвуковых колебаний, получим квадратное уравнение для определения ΔP_0 : $ax^2 + bx - 4\pi / 3 = 0$, где $x = \Delta P_0 / P_0$,

$$a = \frac{3\pi^2}{16} \gamma_2^4 \cdot \frac{R_{\max} \cdot P_0 \cdot \gamma_1}{\omega \cdot N_H \gamma_5 \cdot \gamma_3 \rho_H \cdot L \cdot \nu}; \quad (5)$$

$$b = \frac{1}{8} \pi^2 \gamma_2^4 \gamma_1 \frac{R_{\max}}{\omega \cdot N_H \gamma_5 \cdot \gamma_3 \rho_H \cdot L \cdot \nu} \left[(\rho_C + \rho_H) g H_0 + N_H \frac{F_{\text{int}}}{\pi \gamma_2^2 R_{\max}^2} \right] + \\ + \frac{1}{6} \pi^2 \gamma_6^3 \mu g \gamma_2^2 R_{\max}^2 \cdot \frac{\gamma_1}{\gamma_2 \gamma_5 L \cdot \nu \omega}, \quad (6)$$

здесь L – длина МТ; ω – частота ультразвука; ν – кинематическая вязкость СОЖ; $\gamma_1 = N_H / (V_C \cdot N)$; N – концентрация кавитационных пузырьков; V_C – минимальный объем, в котором один кавитационный пузырек появляется в среднем за $1 / (V_C \cdot N)$ период колебаний; $\gamma_2 = r_k / R_{\max}$; $\gamma_3 = H_0 / L$; $\gamma_5 = \rho_C / \rho_H$. В результате для $\Delta P_0 / P_0$ будем иметь

$$\frac{\Delta P_0}{P_0} = \frac{1}{2a} \left(\sqrt{b^2 + \frac{16}{3} \pi a} - b \right). \quad (7)$$

Далее, учитывая, что $H_0 + \Delta H_n \approx L$ и

$$\gamma_4 = \frac{\Delta H_n}{L} = \frac{\gamma_2^2 R_{\max}^2 2\pi \gamma_1 \Delta P_0}{\omega \cdot N_H \cdot 16 \gamma_5 \gamma_3 \rho_H L^2 \nu}, \quad (8)$$

$$\gamma_3 + \gamma_4 = 1,$$

получим иррациональное уравнение для определения N_H :

$$\gamma_3 + \frac{\gamma_2^2}{V_C N \gamma_3 \gamma_5} \cdot \left(\frac{R_{\max}}{L} \right)^2 \cdot \frac{\pi P_0}{8 \omega \rho_H \cdot \nu} \cdot \frac{1}{2a} \left(\sqrt{b^2 + \frac{16}{3} \pi a} - b \right) = 1. \quad (9)$$

Применяя обычную процедуру решения иррациональных уравнений, находим

$$N_H = \frac{8}{3} \cdot \frac{\pi a G}{B(1 - \gamma_3)} - \frac{(1 - \gamma_3)}{2G \cdot B} - \frac{A}{B}, \quad (10)$$

где

$$A = \frac{\pi^2}{8} \gamma_2^4 \cdot \frac{1}{V_C N} \cdot \frac{R_{\max}}{\omega \gamma_5 \nu} g (\gamma_5 + 1); \quad (11)$$

$$B = \frac{1}{8} \pi \gamma_2^2 \frac{1}{V_C N} \frac{F_{\text{int}} \gamma_7}{\omega \gamma_3 \rho_C R_{\text{max}}^2} + \frac{1}{6} \pi^2 \gamma_6^3 \mu g \gamma_2^2 \frac{R_{\text{max}} \gamma_7}{\gamma_5 \cdot \gamma_3 v \omega V_C N}; \quad (12)$$

$$G = \frac{\gamma_2^2}{\gamma_3 V_C N} \gamma_7^2 \frac{\pi P_0}{16 \omega \rho_C v a}, \quad (13)$$

$$\gamma_6 = R_0 / R_{\text{max}}; \quad \gamma_7 = R_{\text{max}} / L.$$

Из формул (10)–(13) видно, что число N_H является функцией параметров кавитационной области (N, R_{max}, V_C) , СОЖ (ρ_C, v, H_0) , НЧ $(\rho_H, R_0, F_{\text{int}})$ и МТ (L, r_k) . Для исследования вклада указанных параметров в формирование наноструктурированного поверхностного слоя необходимо получить выражение для силы взаимодействия НЧ в поле ультразвуковой волны F_{int} .

2. Взаимодействие наночастиц в поле ультразвуковой волны

В данном разделе проведено обобщение известной формулы Кенига [6] для взаимодействия сферических частиц в звуковом поле на случай парного приближения с учетом дисперсии радиуса частиц $u = R_0 / \bar{R}_0$, где $\bar{R}_0$ – средний радиус НЧ. Кратко рассмотрим общую последовательность решения данной задачи [6]. Пусть исходная НЧ радиуса R_0 имеет координату $\vec{r}_1$, а на расстоянии $L_H = |\vec{r}_2 - \vec{r}_1|$ от нее имеется другая НЧ. Падающее поле $\varphi_i(\vec{r})$ в окрестности первой НЧ можно представить как суперпозицию исходного поля $\varphi_i^{(0)}(\vec{r})$ без учета второй НЧ и поля, рассеянного от второй НЧ: $\varphi_i = \varphi_i^{(0)} + \varphi_s(\vec{r} - \vec{r}_2)$. Рассеянное поле от второй НЧ состоит из суммы сферических волн, образованных в результате рассеяния на второй НЧ как исходной волны $\varphi_i^{(0)}$, так и многократно перерассеянных волн на первой НЧ. Ограничиваясь однократным рассеянием, можно записать $\varphi_s(\vec{r} - \vec{r}_2)$ в виде

$$\begin{aligned} \varphi_i^{(1)}(\vec{r}) = \varphi_s(\vec{r} - \vec{r}_2) = & A \cdot h_0(k |\vec{r} - \vec{r}_2|) + A_\alpha \cdot n_\alpha h_1(k |\vec{r} - \vec{r}_2|) + \\ & + \frac{A_{\alpha\beta}}{3} \left(\frac{3n_\alpha \cdot n_\beta - \delta_{\alpha\beta}}{2} \right) h_2(k |\vec{r} - \vec{r}_2|), \end{aligned} \quad (14)$$

где $h_n(z) = (-z)^n (z^{-1} \partial / \partial z)^n \exp(iz) / z$ – сферические функции Ханкеля [7]; коэффициенты A , A_α , $A_{\alpha\beta}$ определяются из граничных условий для $\varphi_i(0)$, взятых на поверхности сферической НЧ с центром в точке $\vec{r}_2$.

Далее, разложив потенциал (14) в ряд Тэйлора в окрестности первой НЧ в точке $\vec{r}_1$, можно найти (через производные от $\varphi_i^{(1)}(\vec{r}_1)$) значения коэффициентов B , B_α , $B_{\alpha\beta}$, входящих в выражение для силы радиационного давления звука на малую сферическую частицу [6]:

$$\bar{F}_r = \frac{2}{9} \pi R_0 \rho_C \left[k_H^2 \left(1 + 2 \frac{\rho_H}{\rho_C} - \frac{3 \rho_H^2 \cdot C_H^2}{\rho_C^2 \cdot C_C^2} \right) \text{Re}(B \cdot B_\alpha) + \right.$$

$$+ \frac{1}{5} \left(3 \frac{\rho_H}{\rho_C} + 2 \right) \left(\frac{\rho_H}{\rho_C} - 1 \right) \operatorname{Re} (3 B_{\alpha\gamma}^* B_\gamma - B_\alpha \cdot B_{\gamma\gamma}^*) \Big], \quad (15)$$

где k_H – волновое число для НЧ; C_H и C_C – скорость звука в материале НЧ и в жидкой среде соответственно; Re означает взятие действительной части соответствующих выражений.

Эти коэффициенты в приближении однократного рассеяния имеют вид

$$B = B^{(0)} + B^{(1)}, \quad B_\alpha = B_\alpha^{(0)} + B_\alpha^{(1)}, \quad B_{\alpha\beta} = B_{\alpha\beta}^{(0)} + B_{\alpha\beta}^{(1)},$$

где $B^{(0)} \approx \varphi_i^{(0)}(\vec{r}_1)$, $B^{(1)} \approx \varphi_i^{(1)}(\vec{r}_1)$ (индексы (0) означают величины коэффициентов, соответствующих падающему полю, а индексы (1) – рассеянному полю).

Из выражения (15) видно, что сила радиационного давления $\bar{F}_r$ представляет собой сумму двух составляющих. Первая, в которую входят произведения вида $B^{(0)} B_\alpha^{(0)}$, $B_\beta^{(0)} B_{\alpha\beta}^{(0)}$, представляет собой радиационное давление невозмущенного звука, падающего на НЧ. Вторая составляющая, определяемая произведением коэффициентов $B^{(0)} B_\alpha^{(1)}$, $B^{(1)} B_\alpha^{(0)}$, представляет собой силу взаимодействия между двумя НЧ F_{int} [6]:

$$F_{\text{int}} = \frac{3\pi\rho_C}{2} \frac{R_1^3 \cdot R_2^3}{L_H^4} u_s^2 (1 + 3 \cos \alpha), \quad (16)$$

где α – угол, который образован направлением скоростей осциллирующих сферических частиц с линией, соединяющей их центры; $u_s = u_{0C} \cos \omega t$; u_0 – амплитуда скорости колеблющихся частиц; R_1 и R_2 – радиусы взаимодействующих сферических частиц.

Полагая в (16) $R_1 = R_2 = R_0$ и проведя усреднение по всем возможным значениям α , а также по периоду звуковой волны и дисперсии u радиуса НЧ с весовой функцией $P(u)$, получим

$$\begin{aligned} \langle F_{\text{int}} \rangle &= \frac{3}{4} \pi^2 \rho_H \frac{\bar{R}_0^6}{L_H^4} u_{0C}^2 \times \\ &\times \left[\int_0^{u_0 - u_{\min}} dt (u_0 - t)^6 \cdot e^{-t^2} + \int_0^{u_{\max} - u_0} dt (t + u_0)^6 e^{-t^2} \right]. \end{aligned} \quad (17)$$

Выражение (17) было получено с учетом того, что весовая функция $P(u)$ – гауссова вида

$$P(u) = \frac{2}{\sqrt{\pi}} [\Phi(u_{\max} - u_0) + \Phi(u_0 - u_{\min})]^{-1} \cdot e^{-(u - u_0)^2}, \quad (18)$$

где $\Phi(z)$ – интеграл ошибок [8]; $u_0 = (u_{\max} + u_{\min}) / 2$; $u_{\max}$ и $u_{\min}$ – максимальная и минимальная дисперсии радиуса НЧ.

После вычисления интегралов в (17) для $\langle F_{\text{int}} \rangle$ окончательно получим

$$\begin{aligned} \frac{\langle F_{\text{int}} \rangle}{F_0} = & \frac{\rho_H}{\rho_C} \left(\frac{\bar{R}_0}{R_{\text{max}}} \right)^6 \left(\frac{R_{\text{max}}}{L_H} \right)^4 [\Phi(u_{\text{max}} - u_0) + \Phi(u_0 - u_{\text{min}})]^{-1} \times \\ & \times \left[\frac{1}{2} u_0^6 \sqrt{\pi} \Phi(u_0 - u_{\text{min}}) - 3u_0^5 (1 - e^{-(u_0 - u_{\text{min}})^2}) + \frac{15}{2} u_0^4 \gamma\left(\frac{3}{2}, (u_0 - u_{\text{min}})^2\right) - \right. \\ & - 10u_0^3 \gamma\left(2, (u_0 - u_{\text{min}})^2\right) + 3u_0^2 \gamma\left(3, (u_0 - u_{\text{min}})^2\right) + \frac{9}{2} u_0^2 \gamma\left(\frac{5}{2}, (u_0 - u_{\text{min}})^2\right) - \\ & - 3u_0 \gamma\left(\frac{7}{2}, (u_0 - u_{\text{min}})^2\right) + \frac{1}{2} \gamma\left(4, (u_0 - u_{\text{min}})^2\right) + \frac{1}{2} \gamma\left(\frac{7}{2}, (u_{\text{max}} - u_0)^2\right) + \\ & + 3u_0 \gamma\left(3, (u_{\text{max}} - u_0)^2\right) + \frac{15}{2} u_0^2 \gamma\left(\frac{5}{2}, (u_{\text{max}} - u_0)^2\right) + u_0^3 \gamma\left(2, (u_{\text{max}} - u_0)^2\right) + \\ & + 9u_0^2 \gamma\left(2, (u_{\text{max}} - u_0)^2\right) + \frac{15}{2} u_0^4 \gamma\left(\frac{3}{2}, (u_{\text{max}} - u_0)^2\right) + \\ & \left. + 3u_0^5 \cdot (1 - e^{-(u_{\text{max}} - u_0)^2}) + \frac{u_0^6}{2} \sqrt{\pi} \Phi(u_{\text{max}} - u_0) \right], \end{aligned} \quad (19)$$

где $F_0 = 3\pi^2 \rho_C \cdot R_{\text{max}}^2 \cdot u_{0C}^2 / (2\sqrt{\pi})$.

При вычислении в (17) появлялись интегралы вида

$$\int_0^u x^{\nu-1} \cdot e^{-\mu x} dx = \mu^{-\nu} \cdot \gamma(\nu, \mu u), \quad (20)$$

где $\gamma(\alpha, \beta)$ – неполная гамма-функция [8].

На рис. 2 представлена зависимость силы взаимодействия (в относительных единицах) от нормированного расстояния L / R_{max} между НЧ и параметра $u_0 = (u_{\text{max}} + u_{\text{min}}) / 2$. Можно видеть, что с уменьшением расстояния между НЧ величина $\langle F_{\text{int}} \rangle$ существенно возрастает, что является следствием формулы Кенига (16). Видно также, что основной вклад в $\langle F_{\text{int}} \rangle$ вносят системы НЧ с большой дисперсией их радиуса.

3. Моделирование модуля Юнга поверхностного слоя металла с цепочкой микротрещин

Модель фрагмента рассматриваемого поверхностного слоя металла с периодическим изменением величины локального модуля Юнга представлена на рис. 3.

Согласно рис. 3 эффективный модуль Юнга поверхностного слоя металла можно представить в виде

$$E = E_2 + (E_1 - E_2) \cdot \left| \sin \frac{\pi x}{a^*} \right|. \quad (21)$$

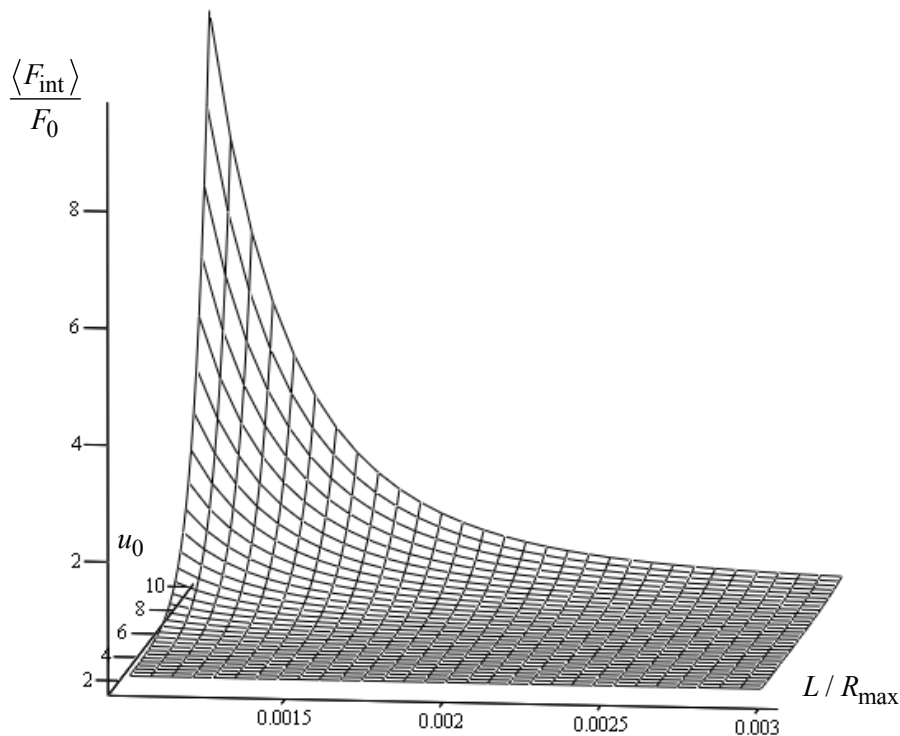


Рис. 2. Зависимость силы взаимодействия НЧ $\langle F_{\text{int}} \rangle$ (в относительных единицах) от нормированного расстояния L / R_{max} между НЧ и параметра $u_0 = (u_{\text{max}} + u_{\text{min}}) / 2$ для $\rho_H / \rho_C = 0,15$; $\bar{R}_0 / R_{\text{max}} = 2 \cdot 10^{-3}$

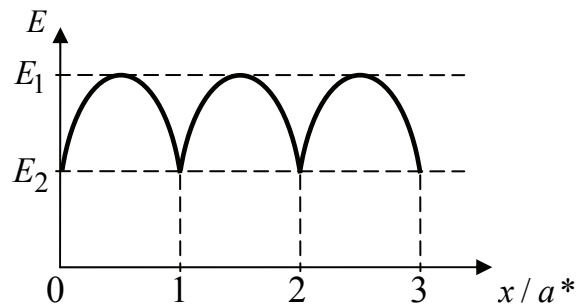


Рис. 3. Модель изменения величины локального модуля Юнга вдоль фрагмента поверхностного слоя металла: E_1 – модуль Юнга матрицы; E_2 – модуль Юнга материала НЧ; a^* – расстояние между МТ на поверхности металла (предполагается, что $r_k \ll a^*$)

После усреднения (21) по дисперсии $V = a^* / \bar{a}^*$, где $\bar{a}^*$ – среднее расстояние между МТ, с весовой функцией гауссова типа получим

$$\langle E \rangle_{a^*} = E_2 + (E_1 - E_2) \frac{16\bar{a}^*}{\pi l} \cdot V_0, \quad (22)$$

где l – длина поверхностного слоя металла; $V_0 = (V_{\max} + V_{\min}) / 2$.

Рассмотрим частные случаи полученного выражения для $\langle E \rangle_{a^*}$. Пусть $\bar{a}^* / l \ll 1$, т.е. поверхностный слой металла содержит достаточно густую цепочку МТ (такая ситуация может иметь место на поверхности, например, феррита). Тогда, как следует из (22), $\langle E \rangle_{a^*} \approx E_2$ и эффективный модуль Юнга будет определяться в основном модулем Юнга материала НЧ. Если $16\bar{a}^* / (\pi l) V_0 \approx 1$ (редкая цепочка МТ и небольшая дисперсия величины a^*), то $\langle E \rangle_{a^*} \approx E_1$, т.е. эффективный модуль Юнга в этом случае определяется модулем Юнга матрицы. Таким образом, появляется возможность путем варьирования вида материала НЧ изменять эффективный модуль Юнга поверхностного слоя металла в достаточно широких пределах и тем самым целенаправленно изменять качество поверхностного слоя деталей машин. На рис. 4 представлена зависимость величины $\langle E \rangle_{a^*} / E_2$ для предельного случая эффективно заполненных МТ наночастицами металла с модулем Юнга E_2 от $\bar{a}^* / l$ и параметра $V_0 = (V_{\max} + V_{\min}) / 2$.

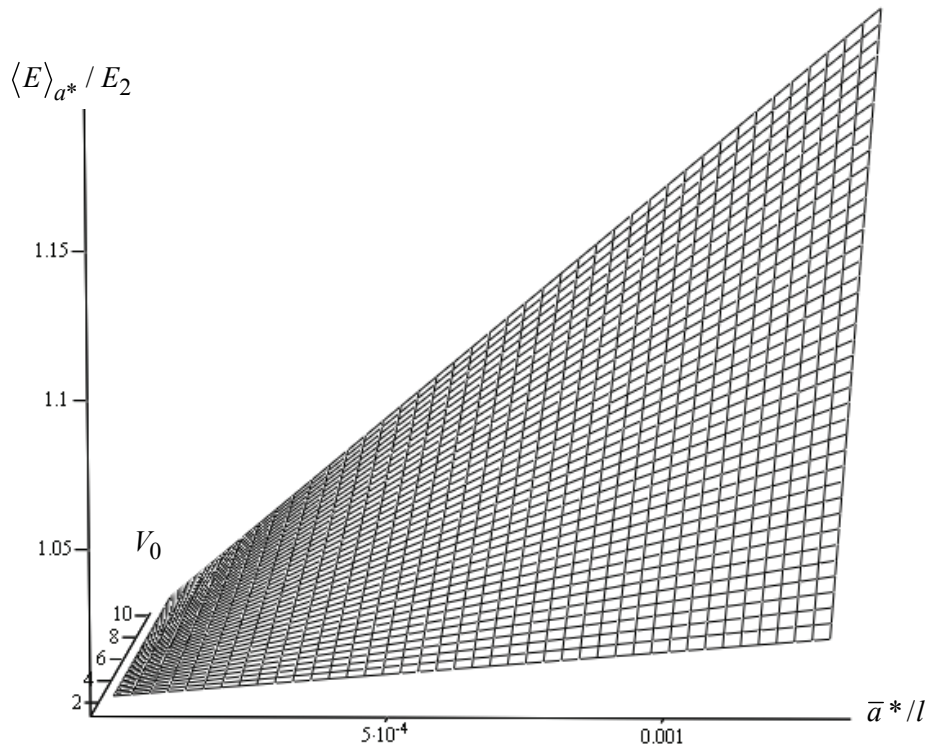


Рис. 4. Зависимость величины $\langle E \rangle_{a^*} / E_2$ от $\bar{a}^* / l$ и параметра $V_0 = (V_{\max} + V_{\min}) / 2$ для $E_1 / E_2 = 4$

Из рис. 4 видна существенная роль дисперсии координат МТ в формировании эффективного модуля Юнга поверхностного слоя металла: более упорядоченное расположение МТ (малая величина дисперсии a^*) способствует формированию поверхностного слоя с заранее заданной величиной модуля Юнга E_2 (модуль Юнга материала НЧ). При хаотическом расположении МТ становится ощутимым вклад в величину $\langle E \rangle_{a^*}$ модуля Юнга материала матрицы (рис. 4).

Необходимо отметить, что E_2 является, вообще говоря, функцией числа N_H НЧ, заполняющих МТ в процессе звукокапиллярного эффекта. Воспользовавшись формулой для модуля Юнга пористого материала [9], можно получить качественную зависимость $E_2(N)$: $E_2 = (1 - b \cdot \Theta) E_{20}$, где Θ – коэффициент пористости: $\Theta = 1 - V_H \cdot N_H / V_0$, V_H – объем наночастицы; V_0 – объем МТ; E_{20} – модуль Юнга материала в отсутствие пор; b – постоянный коэффициент, порядка единицы. В результате для E_2 получим $\frac{E_2}{E_{20}} \approx \frac{V_H}{V_0} \cdot N_H$, где N_H определяется формулой (10). На рис. 5 приведена зависимость величины E_2 / E_{20} от эффективного радиуса устья МТ, нормированного на максимальный радиус кавитационного пузырька $r_k / R_{\max}$ для разных частот ультразвуковых колебаний ω (с учетом того, что $V_H / V_0 \approx 10^{-5}$).

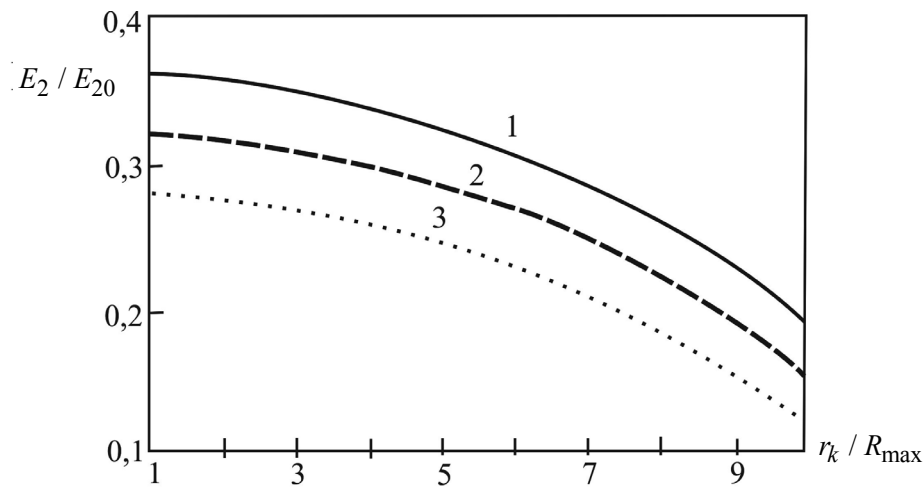


Рис. 5. Зависимость величины E_2 / E_{20} от эффективного радиуса устья МТ $r_k / R_{\max}$ для разных частот ультразвуковых колебаний ω ; кривые: 1 – $3 \cdot 10^4 \text{ с}^{-1}$; 2 – $3,5 \cdot 10^4 \text{ с}^{-1}$; 3 – $4 \cdot 10^4 \text{ с}^{-1}$

Как видно из рис. 5, величина E_2 / E_{20} уменьшается с ростом эффективного радиуса МТ, что связано с соответствующим уменьшением избыточного давления ΔP_0 , необходимого для удержания мениска на высоте обычного капиллярного подъема H_0 . Видно также, что с ростом частоты ультразву-

ковых колебаний величина E_2 / E_{20} уменьшается за счет уменьшения числа образующихся кавитационных пузырьков.

Заключение

Таким образом, в работе теоретически исследован кавитационный механизм управления качеством поверхностного слоя деталей машин с цепочкой МТ, основанный на звукокапиллярном эффекте с участием НЧ. Выявлена возможность управления эффективным модулем Юнга поверхностного слоя путем соответствующего подбора материала НЧ. Модифицирована формула Кенига для силы взаимодействия сферических частиц в поле ультразвуковой волны для случая парного взаимодействия НЧ в кластере с учетом дисперсии их характерного размера. Показано, что основной вклад в силу взаимодействия НЧ вносят кластеры НЧ с большой дисперсией их радиуса. Предложена одномерная модель поверхностного слоя металла с цепочкой МТ. Найдено, что в случае достаточно плотной регулярной цепочки МТ эффективный модуль Юнга поверхностного слоя определяется в основном модулем Юнга материала НЧ. Полученные теоретические результаты могут составить основу технологии повышения эффективности шлифования, в том числе и достаточно хрупких материалов.

Список литературы

1. **Сипайлов, В. А.** Тепловые процессы при шлифовании и управление качеством поверхности / В. А. Сипайлов. – М. : Машиностроение, 1978. – 167 с.
2. **Фесенко, А. В.** Повышение эффективности шлифования при активации и рациональном использовании СОЖ / А. В. Фесенко, Ю. Н. Любимый // Вестник Сумского государственного университета. Серия «Технические науки». – 2010. – № 3. – С. 71–100.
3. **Промтов, М. А.** Перспективы применения кавитационных технологий для интенсификации химико-технологических процессов / М. А. Промтов // Вестник ТГТУ. – 2008. – Т. 14, № 4. – С. 861–869.
4. **Shen, B.** Performance of novel Mo S2 nanoparticles based grinding fluids in minimum quantity lubrication grinding / Bin Shen, Ajay P. Malshe, Parash Kalita, Albert J. Shih // Transaction of NAMRI/SME. – 2008. – V. 36. – P. 357–364.
5. **Кувшинов, Г. И.** Акустическая кавитация у твердых поверхностей / Г. И. Кувшинов, П. П. Прохоренко. – Минск : Наука и техника, 1989. – 112 с.
6. **Красильников, В. А.** Введение в физическую акустику / В. А. Красильников, В. В. Крылов. – М. : Наука, 1984. – 400 с.
7. **Бейтмен, Г.** Высшие трансцендентные функции / Г. Бейтмен, А. Эрдейи. – М. : Наука, 1974. – Т. 1, 2.
8. **Янке, Е.** Специальные функции / Е. Янке, Ф. Эмде, Ф. Леш. – М. : Наука, 1977. – 343 с.
9. **Андриевский, Р. А.** Размерные эффекты в нанокристаллических материалах / Р. А. Андриевский, А. М. Глезер // Механические и физические свойства // Физика металлов и металловедение. – 2000. – Т. 89, № 1. – С. 91–112.

Кревчик Владимир Дмитриевич

доктор физико-математических наук,
профессор, заведующий кафедрой
физики, Пензенский государственный
университет

E-mail: physics@pnzgu.ru

Krevchik Vladimir Dmitrievich

Doctor of physical and mathematical
sciences, professor, head of sub-department
of physics, Penza State University

Симонов Николай Петрович

генеральный директор,
ОАО «Пензадизельмаш»

E-mail: physics@pnzgu.ru

Simonov Nikolay Petrovich

General manager, public
corporation "Penzdizelmash"

Соколов Алексей Владимирович

аспирант, Пензенский
государственный университет

E-mail: physics@pnzgu.ru

Sokolov Aleksey Vladimirovich

Postgraduate student,
Penza State University

УДК 532.66:532.528

Кревчик, В. Д.

Кавитационный механизм управления качеством поверхностного слоя деталей при использовании смазывающе-охлаждающих жидкостей с нанодобавками в условиях звукокапиллярного эффекта / В. Д. Кревчик, Н. П. Симонов, А. В. Соколов // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 180–191.

КВАНТОВЫЙ МЕХАНИЗМ ОБРАЗОВАНИЯ НАНОКЛАСТЕРОВ В ПОВЕРХНОСТНОМ СЛОЕ МЕТАЛЛА С ЦЕПОЧКОЙ МИКРОТРЕЩИН В УСЛОВИЯХ КАВИТАЦИОННОГО РЕЖИМА ЗВУКОКАПИЛЛЯРНОГО ЭФФЕКТА

Аннотация. Предложен механизм образования нанокластеров в объеме микротрещины поверхностного слоя металла в условиях кавитационного режима звукокапиллярного эффекта. Основу механизма составляет представление об электронном обмене между наночастицами посредством диссипативного туннелирования с последующим образованием диполей. Показано, что эффективный модуль Юнга наноструктурированной части поверхностного слоя определяется соотношением между темпом туннельных переходов и частотой ультразвуковой волны.

Ключевые слова: наночастицы, кавитационный режим, звукокапиллярный эффект, диссипативное туннелирование, микротрещина, наноструктурированный слой.

Abstract. The authors suggest a mechanism of nano-clusters formation in the volume of microcracks on a surface layer of metal under conditions of the cavitation regime of acoustocapillary effect. The basis of the mechanism is the concept of electronic exchange between nano-particles by means of dissipative tunneling with subsequent dipoles formation. It is shown that the effective Young's modulus of a nano-structured part of the surface layer is determined by the ratio between the rate of tunneling transitions and the frequency of ultrasonic waves.

Key words: nano particles, cavitation mode, acoustocapillary effect, dissipative tunneling, microcrack, nano-structured layer.

Введение

Теоретически исследован один из возможных механизмов образования нанокластеров в объеме микротрещин (МТ) поверхностного слоя металла в условиях кавитационного режима звукокапиллярного эффекта. Физическая сущность рассматриваемого механизма состоит в следующем. На первом этапе возникает кавитационное облачко в объеме смазывающе-охлаждающей жидкости (СОЖ) с нанодобавками с последующим захлопыванием кавитационных пузырьков у оснований МТ и образованием кумулятивных струй. Последние, увлекая наночастицы (НЧ), попадают в канал МТ. Данная ситуация повторяется с периодом ультразвуковой волны, в течение которого формируется кластер из НЧ в объеме МТ. На этом этапе за счет температурной активации возникает процесс обмена электронами между НЧ посредством диссипативного туннелирования. В результате происходит образование диполей и включается механизм диполь-дипольного взаимодействия НЧ между собой и с оборванными связями внутренней полости МТ за счет сил электростатического изображения, что приводит к образованию нанокластера. При этом эффективный модуль Юнга наноструктурированной части поверхностного слоя E^* является функцией частоты ω_S ультразвуковой волны и вероятности диссипативного туннелирования Γ_0 :

$$E^* = E_0 \frac{N_{\text{НЧ}}^{(\text{МТ})}}{N_{\text{НЧ}}} \cdot \frac{2\pi}{\omega_S} \cdot \Gamma_0, \quad (1)$$

где E_0 – модуль Юнга материала НЧ; $N_{\text{НЧ}}^{(\text{МТ})}$ – число НЧ в объеме МТ; $N_{\text{НЧ}}$ – число НЧ в объеме СОЖ.

Из формулы (1) видно, что величина E^* существенно зависит от соотношения между темпом туннельных переходов и частотой ультразвуковой волны.

Для расчета Γ_0 использовался формализм квантового туннелирования с диссипацией в адиабатическом, квазиклассическом инстантонном приближении [1–11] применительно к туннельной химической динамике.

1. Расчет вероятности диссипативного туннелирования

Рассматриваются адиабатические туннельные реакции, у которых параметр Ландау – Зинера велик:

$$\frac{\Delta^2}{\hbar u |F_2 - F_1|} \gg 1,$$

где Δ – электронный матричный элемент взаимодействия между начальным и конечным состояниями; u – скорость переносимой частицы; $F_{1,2}$ – силы в точке пересечения термов.

Состояние реакционной системы в среде характеризуется многомерной потенциальной поверхностью (по аналогии с постановкой задачи в химической туннельной кинетике [2–4])

$$U_i = \sum_{i=1}^N \frac{1}{2} \omega_{0i}^2 (x_i + x_{0i})^2, \quad U_f = \sum_{i=1}^N \frac{1}{2} \omega_{0i}^2 (x_i - x_{0i})^2 - \Delta I. \quad (2)$$

Гамильтониан системы (вдоль координаты туннелирования):

$$\hat{H} = \frac{p_1^2}{2} + v_1(y_1) + y_1 \sum_{\alpha=2}^N C_{\alpha} y_{\alpha} + \frac{1}{2} \sum_{\alpha=2}^N (p_{\alpha}^2 + \omega_{\alpha}^2 y_{\alpha}^2), \quad (3)$$

где C_{α} – коэффициенты взаимодействия с осцилляторными модами среды.

Вероятность туннелирования в единицу времени определяется как [5, 6]

$$\Gamma = 2T \frac{\text{Im} Z}{\text{Re} Z}, \quad (4)$$

где $Z = \prod_{\alpha} \int Dy_1 \int Dy_{\alpha} \exp[-S\{y_1; y_{\alpha}\}]$ – статистическая сумма системы,

представимая в виде континуального интеграла; $y_{\alpha}(-\beta/2) = y_{\alpha}(\beta/2)$ (здесь

$\beta \equiv T^{-1}$, или $\beta \equiv \frac{\hbar}{kT}$, где $\hbar$ и k полагаются равными единице). Квазиклассическое (инстантонное) действие [2, 5, 6]:

$$S\{y_1\} = \int_{-\beta/2}^{\beta/2} d\tau \left[\frac{1}{2} \dot{y}_1^2 + v(y_1) + \frac{1}{2} \int_{-\beta/2}^{\beta/2} d\tau' K(\tau - \tau') y_1(\tau) y_1(\tau') \right]; \quad (5)$$

$$v(y_1) = v_1(y_1) - \frac{1}{2} \sum_{\alpha=2}^N \frac{C_\alpha^2}{\omega_\alpha^2} y_1^2, \quad (6)$$

где $\nu_n \equiv 2\pi n T$ – мацубаровская частота.

Перенормированная потенциальная энергия частицы вдоль координаты туннелирования в случае двухъямной модели (рис. 1) принимает вид

$$v(q) = \frac{1}{2} \omega_0^2 (q + q_0)^2 \theta(-q) + \left[\frac{1}{2} \omega_0^2 (q - q_1)^2 - \Delta I \right] \theta(q); \quad (7)$$

$$q = y_1 + \frac{\Delta I}{2\lambda}, \quad \omega_0^2 = \omega_1^2 - \sum_{\alpha=2}^N \frac{C_\alpha^2}{\omega_\alpha^2},$$

$$q_0 = \frac{\lambda}{\omega_0^2} - \frac{\Delta I}{2\lambda}, \quad q_1 = \frac{\lambda}{\omega_0^2} + \frac{\Delta I}{2\lambda}, \quad \lambda^2 = \sum_{i=1}^N \omega_{0i}^4 x_{0i}^2,$$

где $\theta(q)$ – единичная функция Хевисайда.

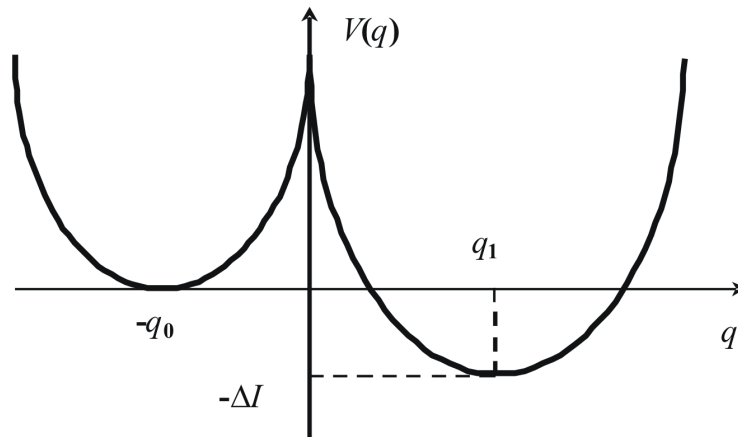


Рис. 1. Потенциальная энергия частицы вдоль координаты туннелирования (двухъямная модель)

Предполагается, что в действие $S(q)$ основной вклад вносит траектория (инстантон), минимизирующая функционал (5) и подчиняющаяся уравнению Эйлера – Лагранжа:

$$-\ddot{q}_B(\tau) + \frac{\partial v(q_B)}{\partial q_B} + \int_{-\beta/2}^{\beta/2} d\tau' K(\tau - \tau') q_B(\tau') = 0, \quad (8)$$

$$q_B(\tau) = q_B(\tau + \beta).$$

Решение (8) ищем в виде

$$q_B(\tau) = \beta^{-1} \sum_{n=-\infty}^{\infty} q_n \exp(i v_n \tau); \quad (9)$$

$$q_B(\tau) = -q_0 + \frac{2(q_0 + q_1)\tau_0}{\beta} + \frac{2\omega_0^2(q_1 + q_0)}{\beta} \sum_{n=1}^{\infty} \frac{\sin v_n \tau_0 \cdot \cos v_n \tau}{v_n (\nu_n^2 + \omega_0^2 + \zeta_n)}, \quad (10)$$

где

$$\zeta_n = \nu_n^2 \sum_{\alpha=2}^N \frac{C_{\alpha}^2}{\omega_{\alpha}^2 (\omega_{\alpha}^2 + \nu_n^2)}, \quad (11)$$

тогда

$$S_B = 2\omega_0^2 (q_0 + q_1) q_0 \tau_0 - \frac{2\omega_0^2 (q_0 + q_1)^2 \tau_0^2}{\beta} - \frac{4\omega_0^4 (q_0 + q_1)^2}{\beta} \sum_{n=1}^{\infty} \frac{\sin^2 v_n \tau_0}{\nu_n^2 (\nu_n^2 + \omega_0^2 + \zeta_n)}. \quad (12)$$

В случае взаимодействия с выделенной локальной модой выражение (12) запишется в виде

$$2S = (q_1 + q_0)(3q_0 - q_1)\omega^2 \tau_0 - \frac{4\omega^2 (q_0 + q_1)^2 (\tau_0)^2}{\beta} - \frac{\omega^2 (q_0 + q_1)^2}{2\tilde{\gamma}} \left\{ \frac{(\omega^2 - \tilde{x}_2)}{\sqrt{\tilde{x}_1}} \left(\operatorname{cth} \left(\frac{\beta}{2} \sqrt{\tilde{x}_1} \right) - \frac{1}{\operatorname{sh} \left(\frac{\beta}{2} \sqrt{\tilde{x}_1} \right)} \left(\operatorname{ch} \left[\left(\frac{\beta}{2} - 2\tau_0 \right) \sqrt{\tilde{x}_1} \right] - \right. \right. \right. \right. \\ \left. \left. \left. - \operatorname{ch} \left[\left(\frac{\beta}{2} \right) \sqrt{\tilde{x}_1} \right] + \operatorname{ch} \left[\left(\frac{\beta}{2} - 2\tau_0 \right) \sqrt{\tilde{x}_1} \right] \right) \right) - \right. \\ \left. - \frac{(\omega^2 - \tilde{x}_1)}{\sqrt{\tilde{x}_2}} \left(\operatorname{cth} \left(\frac{\beta}{2} \sqrt{\tilde{x}_2} \right) - \frac{1}{\operatorname{sh} \left(\frac{\beta}{2} \sqrt{\tilde{x}_2} \right)} \left(\operatorname{ch} \left[\left(\frac{\beta}{2} - 2\tau_0 \right) \sqrt{\tilde{x}_2} \right] - \operatorname{ch} \left[\left(\frac{\beta}{2} \right) \sqrt{\tilde{x}_2} \right] + \right. \right. \right. \right. \right.$$

$$\left. + \operatorname{ch} \left[\left(\frac{\beta}{2} - 2\tau_0 \right) \sqrt{\tilde{x}_2} \right] \right] \right\}, \quad (13)$$

где

$$\tilde{x}_{1,2} = \frac{1}{2} \left(\omega^2 + \omega_L^2 + \frac{C^2}{\omega_L^2} \right) \mp \frac{1}{2} \sqrt{\left(\omega^2 + \omega_L^2 + \frac{C^2}{\omega_L^2} \right)^2 - 4\omega^2 \omega_L^2},$$

$$\tilde{\gamma} = \sqrt{\left(\omega^2 + \omega_L^2 + \frac{C^2}{\omega_L^2} \right)^2 - 4\omega^2 \omega_L^2}.$$

Та же формула (13) в боровских единицах принимает вид

$$S = \frac{1}{2} \frac{E_d}{\hbar} a_d^2 \epsilon_0^{*2} l_1^2 \left(\frac{l_2}{2l_1} \tau_0^* - \tau_0^{*2} \epsilon_T^* - \frac{1}{2\gamma^*} \left\{ \frac{(\epsilon_0^{*2} - x_2^*)}{\sqrt{x_1^*}} \times \right. \right.$$

$$\times \left(\operatorname{cth} \left(\frac{\sqrt{x_1^*}}{2\epsilon_T^*} \right) - \frac{1}{\operatorname{sh} \left(\frac{\sqrt{x_1^*}}{2\epsilon_T^*} \right)} \left(2\operatorname{ch} \left[\left(\frac{1}{\epsilon_T^*} - 2\tau_0^* \right) \frac{\sqrt{x_1^*}}{2} \right] - \operatorname{ch} \left[\frac{\sqrt{x_1^*}}{2\epsilon_T^*} \right] \right) \right) - \frac{(\epsilon_0^{*2} - x_1^*)}{\sqrt{x_2^*}} \times$$

$$\left. \times \left(\operatorname{cth} \left(\frac{\sqrt{\tilde{x}_2}}{2\epsilon_T^*} \right) - \frac{1}{\operatorname{sh} \left(\frac{\sqrt{\tilde{x}_2}}{2\epsilon_T^*} \right)} \left(2\operatorname{ch} \left[\left(\frac{1}{\epsilon_T^*} - 2\tau_0^* \right) \frac{\sqrt{\tilde{x}_2}}{2} \right] - \operatorname{ch} \left[\frac{\sqrt{\tilde{x}_2}}{2\epsilon_T^*} \right] \right) \right) \right\} \right\}, \quad (14)$$

где

$$x_{1,2}^* = \frac{1}{2} \left(\epsilon_0^{*2} + \epsilon_L^{*2} + \frac{\gamma_0^*}{\epsilon_L^{*2}} \right) \mp \frac{1}{2} \sqrt{\left(\epsilon_0^{*2} + \epsilon_L^{*2} + \frac{\gamma_0^*}{\epsilon_L^{*2}} \right)^2 - 4\epsilon_0^{*2} \epsilon_L^{*2}};$$

$$\tilde{\gamma} = \sqrt{\left(\epsilon_0^{*2} + \epsilon_L^{*2} + \frac{\gamma_0^*}{\epsilon_L^{*2}} \right)^2 - 4\epsilon_0^{*2} \epsilon_L^{*2}};$$

$$\tau_0^* = \frac{1}{\epsilon_0^*} \operatorname{Arcsh} \left[\frac{a^* - b^*}{a^* + b^*} \operatorname{sh} \frac{\epsilon_0^*}{2\epsilon_T^*} \right] + \frac{1}{2\epsilon_T^*};$$

$$\epsilon_T^* = \frac{\hbar}{\beta E_d}, \quad \epsilon_L^* = \frac{\hbar \omega_L}{E_d}, \quad \beta = \frac{\hbar}{\epsilon_T^* E_d}, \quad \epsilon_T^{*2} = \frac{4U_0^*}{q_0^{*2}}, \quad U_0^* = \frac{U_0}{E_d};$$

$$l_1 = a^* + b^*, \quad l_2 = 3a^* - b^*, \quad a^* = \frac{q_0}{a_d}, \quad b^* = \frac{q_1}{a_d}, \quad \gamma_0^* = \frac{\hbar^4 C^2}{E_d^4}, \quad \frac{q_1}{q_0} = \frac{b^*}{a^*} = \frac{b}{a}. \quad (15)$$

С экспоненциальной точностью вероятность туннелирования Γ оценивается как $\Gamma_0 \sim \exp(-S)$.

2. Зависимость вероятности диссипативного туннелирования от температуры, частоты фононной моды и константы взаимодействия с контактной средой

Исследования температурной зависимости вероятности туннелирования Γ_0 для НЧ в объеме МТ представлены на рис. 2. Вероятность туннелирования чувствительна к частоте фононной моды и к константе взаимодействия с контактной средой (рис. 3, 4). С точки зрения физики процесса результаты вполне ожидаемы: с ростом частоты фононной моды увеличивается эффективность электрон-фононного взаимодействия, что сопровождается соответствующим ростом энергии туннелирующего электрона и приводит к росту вероятности туннельного переноса (переход от кривых 1 к кривым 2 на рис. 3 и 4); возрастание константы взаимодействия приводит к увеличению вязкости контактной среды, т.е. к росту ее «степени диссипативности» и соответствующему уменьшению вероятности туннельного переноса (переход от кривых 1 к кривым 3 на рис. 3 и 4).

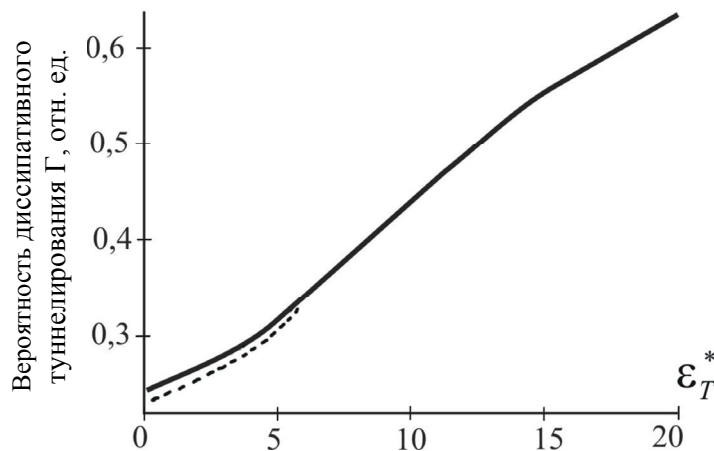


Рис. 2. Зависимость величины Γ от величины ϵ_T^* для НЧ в объеме МТ

Рисунок 3 демонстрирует ряд интересных особенностей туннелирования в рассматриваемой системе. Во-первых, при совпадении радиусов НЧ имеет место эффект блокировки одноэлектронной волновой функции (характерный минимум на рис. 3). Интерес к такому эффекту существенно возрос в последнее время в связи с изучением динамического контроля электронных состояний в двойной квантовой точке в условиях слабой диссипации [12].

Кроме того, на рис. 3 представлены характерные температурно управляемые максимумы в вероятности туннельного переноса при следующих условиях: а) радиус левой НЧ больше радиуса правой НЧ (левый максимум); б) радиус левой НЧ меньше радиуса правой НЧ (правый максимум). Линейная динамика термоуправляемости правого максимума и нелинейная левого представлены на рис. 5. Рисунок 4 демонстрирует пороговый характер термоуправляемого туннелирования в рассматриваемой системе, когда радиус левой НЧ больше радиуса правой НЧ.

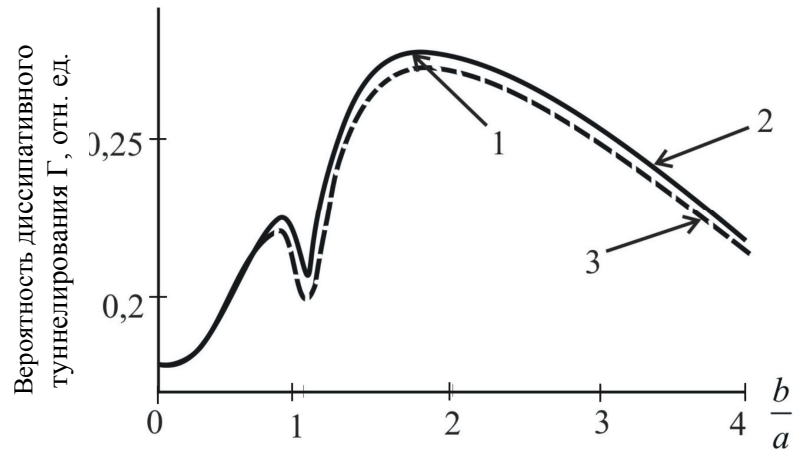


Рис. 3. Зависимость Γ от величины параметра асимметрии b/a :
 1 – $U_0^* = 200$, $\epsilon_T^* = 3$, $\epsilon_L^* = 1$, $\gamma_0^* = 10$; 2 – $U_0^* = 200$, $\epsilon_T^* = 3$, $\epsilon_L^* = 10$, $\gamma_0^* = 10$;
 3 – $U_0^* = 200$, $\epsilon_T^* = 3$, $\epsilon_L^* = 1$, $\gamma_0^* = 50$

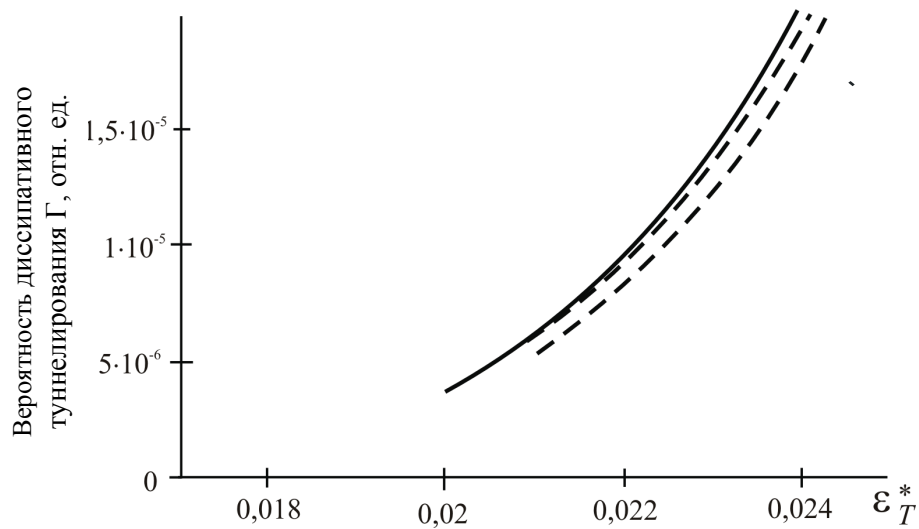
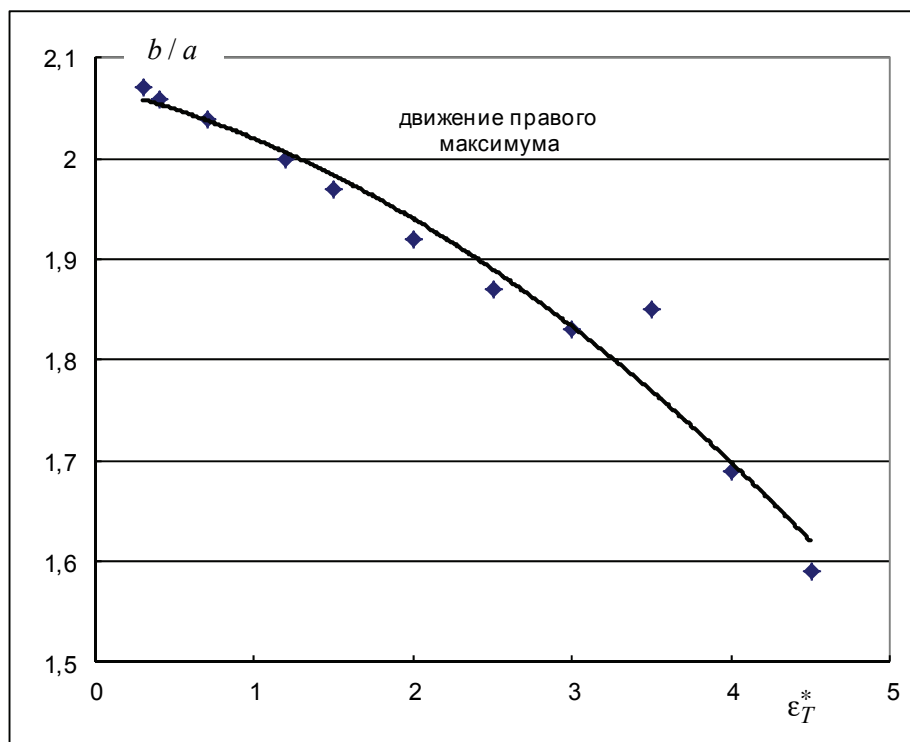
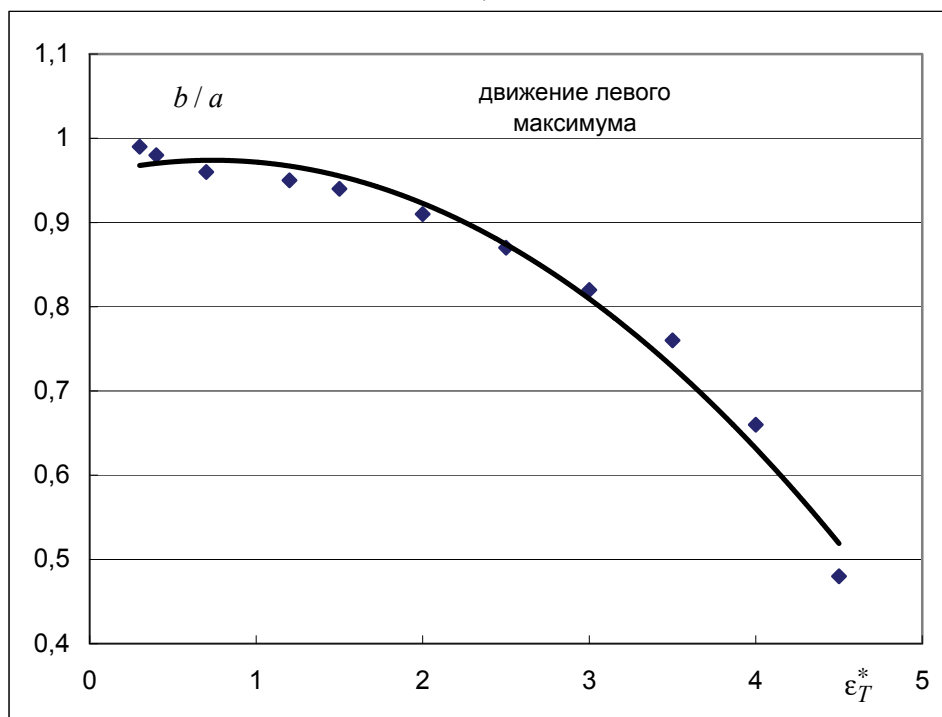


Рис. 4. Зависимость Γ от величины ϵ_T^* при $b/a < 1$: 1 – $U_0^* = 200$, $b/a = 0,98$,
 $\epsilon_L^* = 1$, $\gamma_0^* = 10$; 2 – $U_0^* = 200$, $b/a = 0,98$, $\epsilon_L^* = 10$, $\gamma_0^* = 10$;
 3 – $U_0^* = 200$, $b/a = 0,98$, $\epsilon_L^* = 1$, $\gamma_0^* = 50$



а)



б)

Рис. 5. Зависимость экстремального значения b/a от ϵ_T^* при:
а – $b/a > 1$; б – $b/a < 1$

3. Модель распадного потенциала

В случае, когда одна из НЧ находится достаточно близко к поверхности внутренней полости МТ, можно перейти к модели распадного потенциала в пределе $\tau_0 = 0$ (рис. 6). Тогда выражение для инстантонного квазиклассического действия принимает вид

$$S_B = \frac{q_0^2 \beta}{2} \left[\sum_{n=-\infty}^{\infty} \frac{1}{v_n^2 + \omega_0^2 + \zeta_n} \right]^{-1}. \quad (16)$$

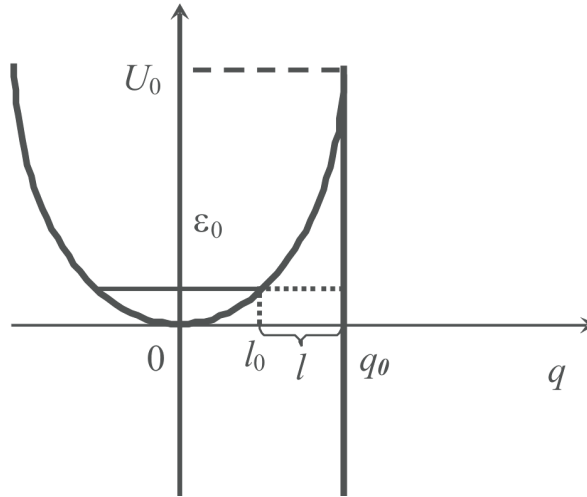


Рис. 6. Потенциальная энергия электрона на границе «НЧ – объемный контакт»: ϵ_0 – энергия основного состояния НЧ; l – ширина барьера; $q_0 \equiv R_0$ – радиус НЧ; $l_0 = q_0 \sqrt{\epsilon_0 / U_0}$; U_0 – амплитуда потенциала конфайнмента НЧ

Рассмотрим (16) с учетом взаимодействия с одной локальной фононной модой (ω_L), $\frac{C}{\omega_0^2} \ll 1$ и $\frac{C}{\omega_L^2} \ll 1$; $\zeta_n = \frac{C^2 v_n^2}{\omega_L^2 (\omega_L^2 + v_n^2)}$:

$$S_B = q_0^2 \left[\frac{A}{\sqrt{\tilde{\lambda}_1}} \operatorname{cth} \left(\frac{\sqrt{\tilde{\lambda}_1} \beta}{2} \right) + \frac{B}{\sqrt{\tilde{\lambda}_2}} \operatorname{cth} \left(\frac{\sqrt{\tilde{\lambda}_2} \beta}{2} \right) \right]^{-1}, \quad (17)$$

где

$$A = \frac{\omega_L^2 - \tilde{\lambda}_1}{\tilde{\lambda}_2 - \tilde{\lambda}_1}, \quad B = \frac{\omega_L^2 - \tilde{\lambda}_2}{\tilde{\lambda}_1 - \tilde{\lambda}_2}; \quad \tilde{\lambda}_{1,2} = \frac{1}{2} \left[(\omega_0^2 + \omega_L^2) + \frac{C^2}{\omega_L^2} \right] \pm \frac{1}{2} \sqrt{(\omega_0^2 - \omega_L^2)^2 + \frac{2C^2}{\omega_L^2} (\omega_0^2 + \omega_L^2) + \frac{C^4}{\omega_L^4}}.$$

Тогда выражение для Γ_0 (с экспоненциальной точностью) принимает вид

$$\Gamma_0 = \exp \left\{ -\frac{Mq_0^2}{\hbar} \left[\frac{A}{\sqrt{\tilde{\lambda}_1}} \operatorname{cth} \left(\frac{\sqrt{\tilde{\lambda}_1} \beta}{2} \right) + \frac{B}{\sqrt{\tilde{\lambda}_2}} \operatorname{cth} \left(\frac{\sqrt{\tilde{\lambda}_2} \beta}{2} \right) \right] \right\}^{-1}. \quad (18)$$

Для оценки вероятности туннелирования в рассматриваемой системе «наночастица – объемный контакт» величину Γ_0 из (18) целесообразно переписать в боровских единицах:

$$\Gamma_0 \sim \exp \left\{ -\frac{q_0^{*2}}{2} \left[\frac{A^*}{\sqrt{\lambda_1^*}} \operatorname{cth} \left(\frac{\sqrt{\lambda_1^*}}{2 \epsilon_T^*} \right) + \frac{B^*}{\sqrt{\lambda_2^*}} \operatorname{cth} \left(\frac{\sqrt{\lambda_2^*}}{2 \epsilon_T^*} \right) \right] \right\}^{-1}, \quad (19)$$

где

$$q_0^* = q_0 / a_d; \quad \epsilon_L^* = \hbar \omega_L / E_d; \quad \epsilon_T^* = k_B T / E_d;$$

$$A^* = (\epsilon_L^{*2} - \lambda_1^*) / (\lambda_2^* - \lambda_1^*); \quad B^* = (\epsilon_L^{*2} - \lambda_2^*) / (\lambda_1^* - \lambda_2^*).$$

Здесь $\lambda_{1,2}^*$ определяется как

$$\lambda_{1,2}^* = \frac{1}{2} \left[(\epsilon_0^{*2} + \epsilon_L^{*2}) + \frac{\gamma_0^*}{\epsilon_L^{*2}} \right] \pm \frac{1}{2} \sqrt{(\epsilon_0^{*2} - \epsilon_L^{*2})^2 + \frac{2\gamma_0^*}{\epsilon_L^{*2}} (\epsilon_0^{*2} + \epsilon_L^{*2}) + \frac{\gamma_0^{*2}}{\epsilon_L^{*4}}}, \quad (20)$$

где $\gamma_0^* = \hbar^4 c^2 / E_d^4$; $\epsilon_0^{*2} = 4U_0^* / q_0^{*2}$.

Следует отметить, что квазиклассическое приближение применимо только для достаточно «широких» барьеров, когда

$$\chi \cdot l \gg 1, \quad \chi = \sqrt{U_0^* - \epsilon_0^*} / a_d, \quad l = q_0 - l_0,$$

где $U_0^* = U_0 / E_d$.

Зависимость вероятности туннелирования от температуры представлена на рис. 7 (распадный потенциал). Как и в случае двухъямной модели, вероятность туннелирования чувствительна к частоте фононной моды и к константе взаимодействия с контактной средой (рис. 7 и 8). В случае распадного потенциала конфайнмента имеет место монотонное уменьшение вероятности туннелирования с ростом размера НЧ (рис. 8).

Заключение

Предложен механизм образования нанокластеров в объеме МТ поверхностного слоя металла в условиях кавитационного режима звукокапиллярного эффекта. Основу механизма составляет представление об электронном обмене между НЧ посредством диссипативного туннелирования с последующим образованием диполей. Показано, что эффективный модуль Юнга нано-

структурированной части поверхностного слоя определяется соотношением между темпом туннельных переходов и частотой ультразвуковой волны.

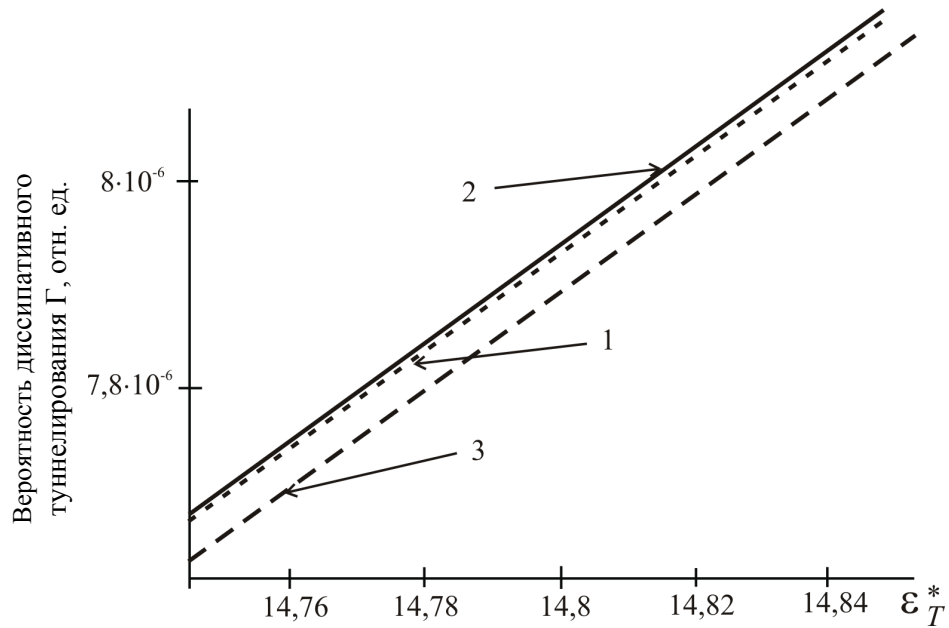


Рис. 7. Зависимость Γ (в относительных единицах) от величины ε_T^* : 1 – $q_0^* = 1,4$, $\varepsilon_L^* = 1$, $\gamma_0^* = 10$; 2 – $q_0^* = 1,4$, $\varepsilon_L^* = 10$, $\gamma_0^* = 10$; 3 – $q_0^* = 1,4$, $\varepsilon_L^* = 1$, $\gamma_0^* = 50$

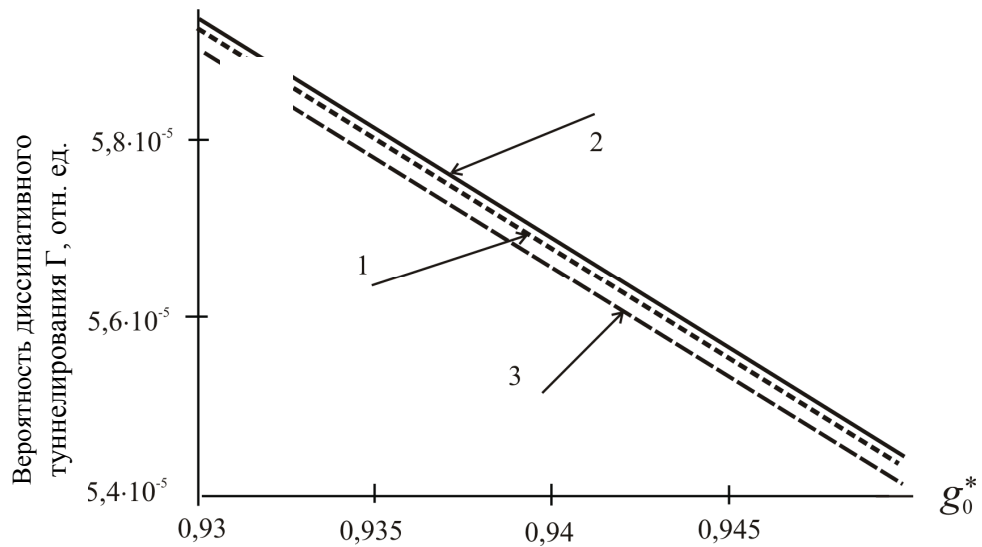


Рис. 8. Зависимость Γ (в относительных единицах) от величины q_0^* :
1 – $U_0^* = 200$, $\varepsilon_T^* = 16$, $\varepsilon_L^* = 1$, $\gamma_0^* = 10$; 2 – $U_0^* = 200$, $\varepsilon_T^* = 16$, $\varepsilon_L^* = 10$, $\gamma_0^* = 10$;
3 – $U_0^* = 200$, $\varepsilon_T^* = 16$, $\varepsilon_L^* = 1$, $\gamma_0^* = 50$

В рамках науки о квантовом туннелировании с диссипацией в инстан-
тонном квазиклассическом приближении исследована управляемость тунне-

лирования в системах «НЧ – внутренняя полость» В качестве управляющих параметров рассматривались температура системы и соотношение размеров радиуса НЧ. Найдено условие блокировки одноэлектронной волновой функции в пределах НЧ. Численно исследованы условия, при которых туннельный ток в системе взаимодействующих НЧ достигает максимального значения. Получены аналитические результаты для вероятности туннелирования в рассматриваемых системах с учетом взаимодействия с локальными осцилляторными модами среды.

Таким образом, продемонстрирована возможность использования науки о квантовом туннелировании с диссипацией к изучению термоуправляемости туннелирования в МТ с НЧ, что открывает новые возможности для получения поверхностного слоя деталей с заранее заданными свойствами.

Список литературы

1. **Ландау, Л. Д.** Квантовая механика (нерелятивистская теория) / Л. Д. Ландау, Е. М. Лифшиц. – М. : Наука, 1989. – Т. 3. – 766 с.
2. **Дахновский, Ю. И.** Низкотемпературные химические реакции как туннельные системы с диссипацией / Ю. И. Дахновский, А. А. Овчинников, М. Б. Семенов // ЖЭТФ. – 1987. – Т. 92, № 3. – С. 955–967.
3. **Aringazin, A. K.** Two-dimensional tunnel correlations with dissipation / A. K. Aringazin, Yu. I. Dahnovsky, V. D. Krevchik et al. // Physical Review B. – 2003. – V. 68. – P. 155426-1–155426-12.
4. **Овчинников, А. А.** Принципы управляемой модуляции низкоразмерных структур : монография / А. А. Овчинников, Ю. И. Дахновский, В. Д. Кревчик и др. – М. : Изд-во УНЦ ДО, 2003. – 510 с.
5. **Caldeira, A. O.** Influence of dissipation on quantum tunneling in macroscopic systems / A. O. Caldeira, A. J. Leggett // Phys. Rev. Lett. – 1981. – V. 46, № 4. – P. 211–214.
6. **Ларкин, А. И.** Квантовое туннелирование с диссипацией / А. И. Ларкин, Ю. Н. Овчинников // Письма в ЖЭТФ. – 1983. – Т. 37, № 7. – С. 322–325.
7. **Gorokhov, D. A.** Ultrasharp crossover from quantum to classical decay in a quantum dot flanked by a double – barrier tunneling structure / D. A. Gorokhov, A. da Silveira Rava. – URL: <http://arXiv.org/abs/cond-mat/0308023>.
8. **Foa Torres, L. E. F.** Coherent versus sequential electron tunneling in quantum dots / L. E. F. Foa Torres, C. H. Lewenkopf, H. M. Pastawski. – URL: <http://arXiv.org/abs/cond-mat/0306148>.
9. **Thielmann, A.** Shot noise in tunneling transport through molecules and quantum dots / A. Thielmann, M. H. Hettler, J. Konig, G. Schon // Phys. Rev. B. – 2003. – V. 68. – P. 115105. – URL: <http://arXiv.org/abs/cond-mat/0302621>.
10. **Ханин, Ю. Н.** Резонансное Г–Х-туннелирование в однопобарьерных гетероструктурах GaAs/AlAs/GaAs / Ю. Н. Ханин, Е. Е. Вдовин, Ю. В. Дубровский // Физика и техника полупроводников. – 2004. – Т. 38, № 4. – С. 436–447.
11. **Тернов, И. М.** Квантовая механика и макроскопические эффекты / И. М. Тернов, В. Ч. Жуковский, А. В. Борисов. – М. : Изд-во МГУ, 1993. – 198 с.
12. **Бурдов, В. А.** Динамический контроль электронных состояний в двойной квантовой точке в условиях слабой диссипации / В. А. Бурдов, Д. С. Соленов // ЖЭТФ. – 2004. – Т. 125, № 3. – С. 684–692.

Артемов Игорь Иосифович

доктор технических наук, профессор,
проректор по научной работе
и инновационной деятельности,
Пензенский государственный
университет

E-mail: rectorat@pnzgu.ru

Artyomov Igor Iosifovich

Doctor of engineering sciences,
professor, vice-rector for research
and innovation, Penza State University

Кревчик Владимир Дмитриевич

доктор физико-математических наук,
профессор, заведующий кафедрой
физики, Пензенский государственный
университет

E-mail: physics@pnzgu.ru

Krevchik Vladimir Dmitrievich

Doctor of physical and mathematical
sciences, professor, head of sub-department
of physics, Penza State University

Симонов Николай Петрович

генеральный директор,
ОАО «Пензадизельмаш»

E-mail: physics@pnzgu.ru

Simonov Nikolay Petrovich

General manager, public
corporation "Penzadizelmash"

УДК 532.66; 532.528

Артемов, И. И.

Квантовый механизм образования нанокластеров в поверхностном слое металла с цепочкой микротрещин в условиях кавитационного режима звукокапиллярного эффекта / И. И. Артемов, В. Д. Кревчик, Н. П. Симонов // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 192–204.

ТЕПЛОФИЗИЧЕСКИЕ И ДИНАМИЧЕСКИЕ ХАРАКТЕРИСТИКИ НАЧАЛЬНОГО УЧАСТКА ВЕРТИКАЛЬНОЙ НИЗКОСКОРОСТНОЙ СТРУИ В ПОПЕРЕЧНОМ ПОТОКЕ

Аннотация. Экспериментально исследовано действие турбулентного потока на вдуваемую нормально низкоскоростную осесимметричную струю. Получены эмпирические зависимости, описывающие геометрию оси струи в поперечном воздушном потоке. Показано, что линии максимальной температуры и максимальной скорости не совпадают, при этом тепловая ось лежит ниже динамической. Получены эмпирические зависимости, описывающие геометрию динамического и теплового пограничного слоев струи в поперечном потоке.

Ключевые слова: струя, поток, геометрия струи, ось струи, пограничный слой.

Abstract. The author has tested an influence of a turbulent flow on a normally injected low-speed axially symmetric jet. The article shows the obtained empirical dependences describing the geometry of the jet axis in a transversal airflow. The researcher states that the lines of maximum temperature and velocity do not coincide; the thermal axis lies lower than the dynamic axis. The obtained empirical dependences describe the geometry of dynamic and thermal boundary layers of a jet in a cross flow.

Key words: jet, flow, jet axis, jet geometry, boundary layer.

Введение

Круглая струя, развивающаяся в сносящем поперечном потоке, представляет интерес не только потому, что это сложная газодинамическая задача, а, прежде всего, из-за широкого ее использования в различных технических устройствах. Основные параметры струи: траектория и дальность, падение скорости и температуры вдоль оси струи, границы динамического и пограничного слоев.

Основная доля работ, посвященных течению при нормальном вдуве струи, относится к условиям вдува в канал струй со сравнительно большим скоростным напором по отношению к напору сносящего потока. В некоторых задачах имеет место небольшое соотношение скоростных напоров струи и потока. Такие течения изучены недостаточно.

1. Постановка задачи

Физическая модель изучаемого нами течения представляет собой воздушную осесимметричную струю, которая выбрасывается из трубки диаметром $d = 0,022$ м под прямым углом к направлению движения воздушного потока. На рис. 1 представлена полученная автором задымленная струя. Для исследования геометрии струи в сносящем потоке применялся аэродинамический канал [1] длиной 5,5 м и поперечным сечением $0,8 \times 0,5$ м. Значения чисел Рейнольдса для воздушных струй, формирующихся в вертикальных трубках, составляли $Re_0 = 2,7 \cdot 10^3 - 6,9 \cdot 10^3$ м/с, где $Re_0 = u_0 d / \nu$ (ν – кинематическая вязкость воздуха).

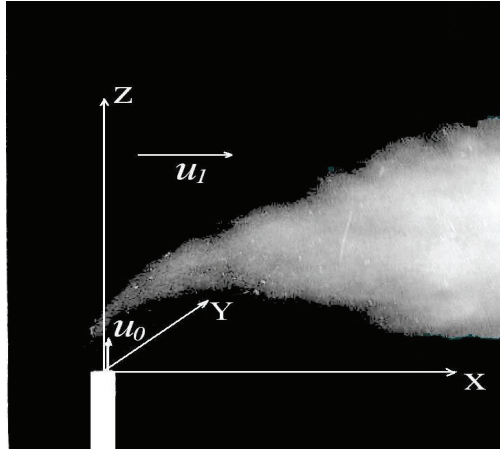


Рис. 1. Постановка задачи

В центральной части канала был сформирован равномерный поток с числом $Re_1 \approx 8,5 \cdot 10^4$, где $Re_1 = u_1 D / \nu$ (D – условный диаметр канала, который определялся из равенства площадей поперечного сечения канала и круглой трубы: $a \times b = \pi D^2 / 4$; отсюда $D = \sqrt{(4ab) / \pi} \approx 0,71$ м). Исследования проводились для относительных динамических напоров $q = 0,85 - 6,12$ ($q = q_0 / q_1 = u_0^2 / u_1^2$) с изотермическими и слабонеизотермическими (критерий Архимеда $Ar = (gd \cdot \Delta T_0) / (2u_0^2 \cdot T_1) = 2,8 \cdot 10^{-4} - 1,0 \cdot 10^{-3}$, где g – ускорение свободного падения; ΔT_0 – разность абсолютных температур в струе и сносящем потоке; T_1 – абсолютная температура потока) струями с использованием термоанемометра ПТ-22, приведенная погрешность измерений которого по скорости воздушных потоков составляет не более 10 %, а по температуре – не более 1,5 °С.

По экспериментальным данным были построены профили скорости и температуры изотермических и слабонеизотермических струй в поперечных вертикальных сечениях на различных расстояниях от трубки.

2. Геометрия оси струи

Геометрия сносимой потоком струи достаточно сложна, что связано с различными динамическими условиями с наветренной и подветренной ее сторон. Под геометрией струи обычно подразумевается положение ее оси, которая представляет собой линию, соединяющую точки с максимальной скоростью течения или температурой.

Известны различные подходы к решению задачи по определению геометрии оси струи в сносящем потоке.

Анализ экспериментальных и расчетных работ по сносу вертикальных струй поперечным потоком [2–5] показывает, что геометрия осевой линии сносимой струи достаточно хорошо может быть описана зависимостью

$$\frac{z_0}{d} = \alpha \left(\frac{\rho_0 u_0^2}{\rho_1 u_1^2} \right)^\beta \left(\frac{x}{d} \right)^\gamma,$$

где α, β, γ – эмпирические коэффициенты ($\alpha \approx 1, \beta \approx 0,4 - 0,5, \gamma \approx 0,33 - 0,5$); ρ_0, ρ_1 – плотности струи и потока соответственно; u_0, u_1 – скорости струи на выходе из сопла и потока соответственно.

Существуют и теоретические решения задачи определения геометрии оси струи. Однако эти решения можно считать грубо приближенными, пригодными только для небольшого интервала параметров струи и потока. Сложная картина движения частиц струи, а следовательно, и сложность образования полей скорости и температуры не позволяют точно описать траекторию струи в сносимом потоке теоретическими методами, тем более в широком диапазоне относительных скоростных напоров струи и потока. Приходится ограничиваться эмпирическим описанием опытов, что позволяет иметь достаточно надежные данные в той области граничных условий, в которых проводились эксперименты.

Для точек максимальной скорости струи воздуха, истекающего в поперечный поток, автором была получена эмпирическая формула

$$\bar{z}_0 = q^{0,51} \bar{x}^{0,41}, \quad (1)$$

где $\bar{x} = x/d$, $\bar{z}_0 = z_0/d$ – безразмерные координаты точек оси струи, соответственно продольная и вертикальная, отсчитываемые от центра среза трубки (x, z_0 – координаты декартовой системы координат).

Для относительных скоростных напоров $q = 2,00; 3,10; 5,55$ струи было получено эмпирическое выражение для точек максимальной температуры [1]:

$$\bar{z}_0 = q^{0,51} \bar{x}^{0,33}. \quad (2)$$

На рис. 2 показана обобщенная геометрия оси струи в поперечном потоке: сплошная линия – поднятие струи над устьем источника, вычисленное по (2), штриховая линия – ось струи, вычисленная по (1).

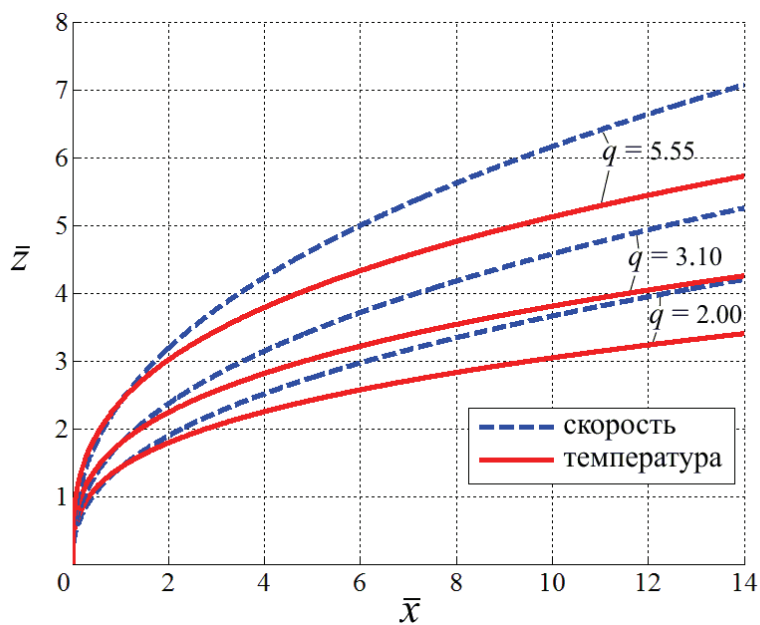


Рис. 2. Геометрия оси струи

Измерения скоростного и температурного полей, визуальные наблюдения и фотосъемка задымленных струй показали, что наиболее достоверно ось струи описывается точками максимальной температуры в поперечных сечениях неизотермической струи. Положения точек максимальной скорости располагаются ближе к верхней границе струи и с уменьшением q могут выйти за ее пределы.

3. Изменение скорости и температуры вдоль оси

С помощью экспериментальных данных профилей скорости в изотермической струе ($q = 2,5; 6,12$) была получена эмпирическая зависимость для значений осевой относительной безразмерной скорости [6]:

$$\frac{u_m - u_1}{u_0 - u_1} = \frac{0,8}{\bar{x}^{0,9}}, \quad (3)$$

где u_m – максимальная скорость в рассматриваемом поперечном сечении струи.

На рис. 3 представлены экспериментальные данные избыточных безразмерных осевых значений скорости воздушных струй. Там же для сравнения приведены данные для спутной и затопленной струй [7], которые показывают менее интенсивное уменьшение осевой скорости по сравнению с нашими опытами.

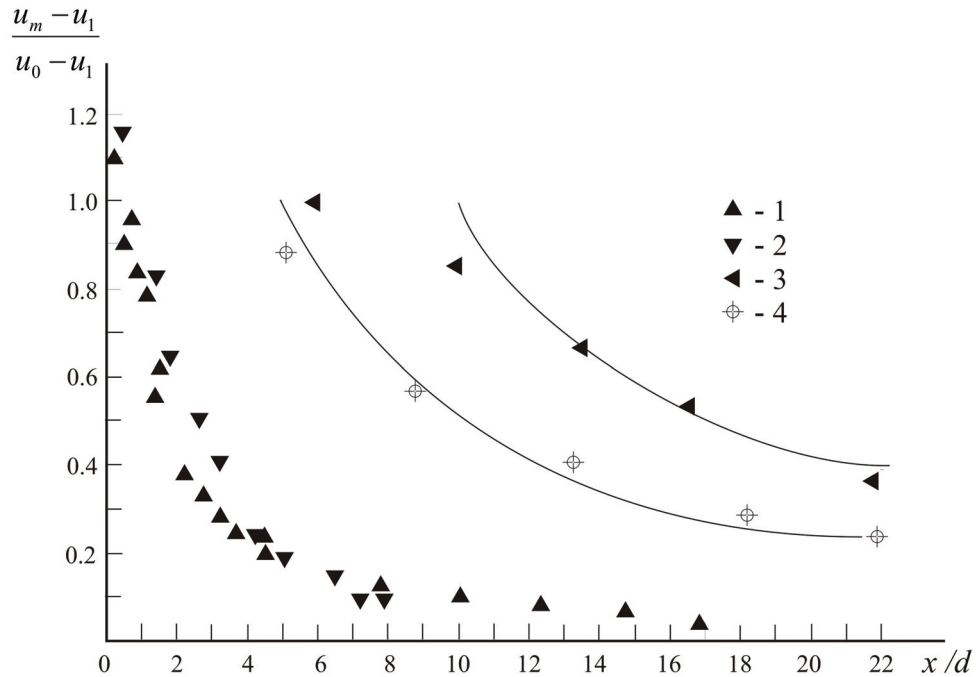


Рис. 3. Изменение осевых относительных избыточных значений скорости в сносимой поперечным потоком струе с расстоянием от устья источника при $q = 6, 12$ (1); $2, 5$ (2) и аналогичные зависимости [7] для спутной при $q = 3, 17$ (3) и затопленной струй (4) при $\rho_1/\rho_0 = 1,34$

На основе опытных данных для слабонеизотермических струй с $q = 2,0$; $3,1$; $5,55$ для значений осевой относительной безразмерной температуры была получена зависимость, которая может быть представлена в виде [6]

$$\frac{t_m - t_1}{t_0 - t_1} = \frac{1}{\bar{x}^{0,75}}, \quad (4)$$

где t_0 – абсолютная температура в начальном сечении струи; t_1 – абсолютная температура в сносящем потоке; t_m – максимальная температура в рассматриваемом поперечном сечении струи.

На рис. 4 представлены экспериментальные данные изменения относительной осевой температуры в струе. Там же для сравнения приведены данные для спутной и затопленной струй [7].

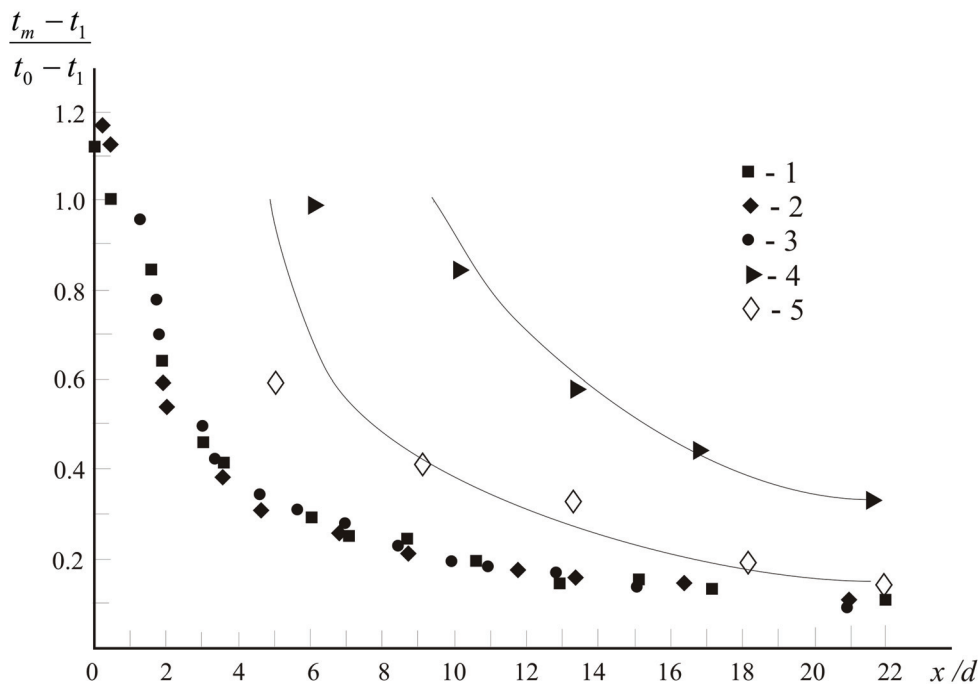


Рис. 4. Изменение осевых относительных избыточных значений температуры в сносимой поперечным потоком струе с расстоянием от устья источника при $q = 5,55$, $\Delta t_0 = 12,5$ °C (1); $q = 3,1$, $\Delta t_0 = 15$ °C (2); $q = 2,0$, $\Delta t_0 = 16$ °C (3) и аналогичные зависимости [7] для спутной при $q = 3,17$, $\Delta t_0 = 230$ °C (4) и затопленной струй при $\Delta t_0 = 105$ °C (5)

В работе показано, что затухание осевых значений скорости и температуры в сносимой поперечным потоком струе происходит более интенсивно по сравнению с затопленной и спутной.

4. Профили скорости и температуры

Для струи в поперечном потоке относительные безразмерные, избыточные значения скорости могут быть представлены в виде

$$\frac{u - u_1}{u_m - u_1} = f\left(\frac{z - z_0}{|z_0 - z_{0,5}|}\right), \quad (5)$$

где u – скорость в измеряемой точке; z – вертикальная координата измеряемой точки; z_0 – та же величина оси струи (точки максимальной скорости) в рассматриваемом поперечном сечении; $z_{0,5}$ – расстояние от оси струи до точки, в которой скорость составляет половину избыточного максимального значения в рассматриваемом сечении (рис. 5).

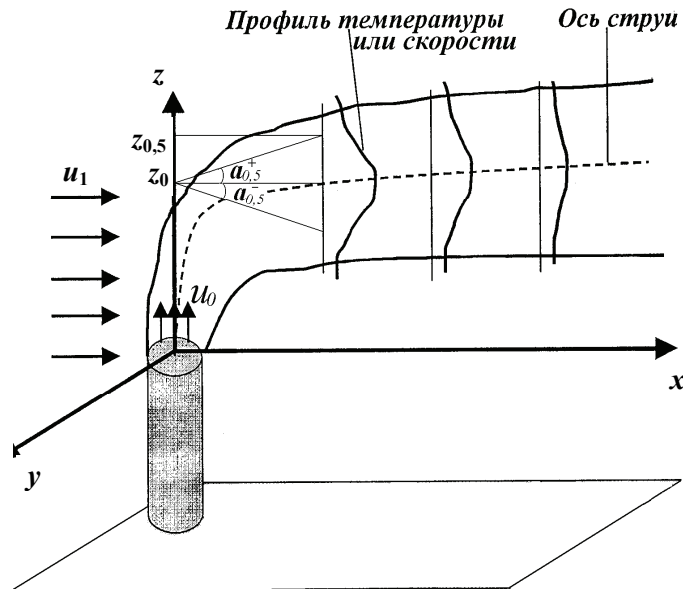


Рис. 5. Геометрические характеристики границ пограничного слоя в струе

На основе опытных данных для изотермических струй ($q = 2,5; 6,12$) был получен безразмерный профиль избыточной скорости, который может быть представлен в виде

$$\frac{u - u_1}{u_m - u_1} = \exp\left\{-\frac{0,7}{\text{tg}^2 \alpha_{0,5u}} \left(\frac{z - z_0}{x}\right)^2\right\}, \quad (6)$$

где $\text{tg} \alpha_{0,5u}$ – тангенс угла, образованного точкой половинной избыточной максимальной скорости и осью X в рассматриваемом поперечном сечении.

Была построена зависимость тангенса угла $\alpha_{0,5u}$ от расстояния до трубки и относительного скоростного напора (для искривленной части струи) [8]:

$$\text{tg} \alpha_{0,5u}^+ = \frac{0,17q^{0,51}}{\bar{x}^{0,75}}, \quad \text{tg} \alpha_{0,5u}^- = \frac{0,15q^{0,51}}{\bar{x}^{0,75}}; \quad (7)$$

знак «+» над $\text{tg} \alpha_{0,5u}$ соответствует значениям тангенса выше оси струи (наветренная сторона); знак «-» – соответственно ниже оси (подветренная сторона).

Используя (3) и (6), можно получить выражение для относительной скорости:

$$\frac{u - u_1}{u_0 - u_1} = \frac{0.8}{\bar{x}^{0.9}} \cdot \exp \left\{ -\frac{0.7}{\operatorname{tg}^2 \alpha_{0,5u}} \left(\frac{z - z_0}{x} \right)^2 \right\}. \quad (8)$$

На рис. 6 представлены профили скорости, построенные по (8) с использованием (1) и (7).

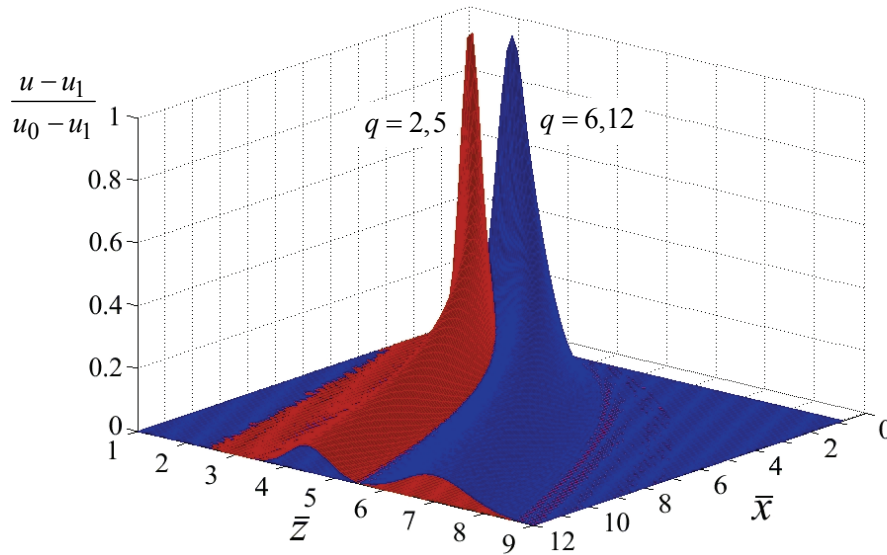


Рис. 6. Безразмерные профили скорости в плоскости XZ

По аналогии с профилем скорости изотермической струи профиль безразмерных избыточных температур слабонеизотермических струй в вертикальной плоскости строился в виде

$$\frac{t - t_1}{t_m - t_1} = \varphi \left(\frac{z - z_0}{|z_0 - z_{0,5}|} \right), \quad (9)$$

где t – температура в измеряемой точке; t_m – максимальная температура в рассматриваемом поперечном сечении струи; z – вертикальная координата измеряемой точки; z_0 – вертикальная координата оси струи (точки максимальной температуры [8]) в рассматриваемом поперечном сечении; $z_{0,5}$ – расстояние от оси струи до точки, в которой температура составляет половину избыточного максимального значения в рассматриваемом сечении (см. рис. 5).

На основе опытных данных для слабонеизотермических струй с $q = 2,0$; $3,1$; $5,55$ был получен безразмерный профиль избыточной температуры, который может быть представлен в виде

$$\frac{t - t_1}{t_m - t_1} = \exp \left\{ -\frac{0.7}{\operatorname{tg}^2 \alpha_{0,5t}} \left(\frac{z - z_0}{x} \right)^2 \right\}, \quad (10)$$

где $\operatorname{tg} \alpha_{0,5t}$ – тангенс угла, образованного точкой половинной избыточной максимальной температуры и осью X в рассматриваемом поперечном сечении струи.

Для некоторых значений координаты ($\bar{x} = 1,36 - 10,45$) были найдены значения тангенсов отдельно выше ($\operatorname{tg} \alpha_{0,5t}^+$) и ниже ($\operatorname{tg} \alpha_{0,5t}^-$) оси в поперечных сечениях струи:

$$\operatorname{tg} \alpha_{0,5t}^+ = \frac{0,39q^{0,43}}{\bar{x}^{0,75}}, \quad \operatorname{tg} \alpha_{0,5t}^- = \frac{0,54q^{0,28}}{\bar{x}^{0,75}}. \quad (11)$$

Используя (4) и (10), можно получить выражение для профиля относительной температуры:

$$\frac{t - t_1}{t_0 - t_1} = \frac{1}{\bar{x}^{0,75}} \exp \left\{ -\frac{0,7}{\operatorname{tg}^2 \alpha_{0,5t}} \left(\frac{z - z_0}{x} \right)^2 \right\}. \quad (12)$$

На рис. 7 представлены профили температуры, построенные по (12) с использованием (4) и (11).

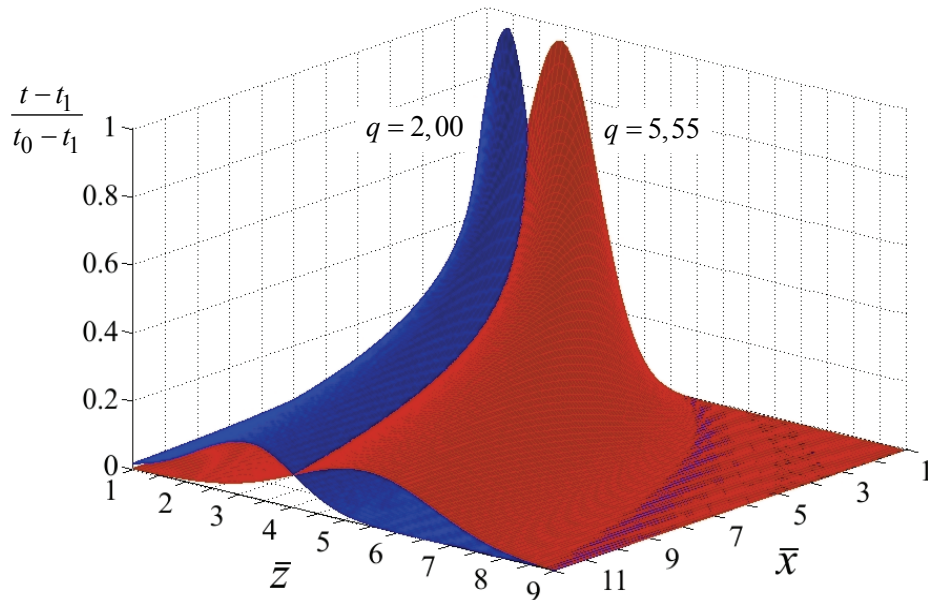


Рис. 7. Безразмерные профили температуры в плоскости XZ

Заключение

Эксперименты показали, что в низкоскоростных нагретых струях, сносимых поперечным потоком, положение максимумов температуры отличается от положения максимумов скорости. При малых относительных скоростных напорах q в некоторых задачах, связанных с полями температуры или концентрации, за ось струи предпочтительнее принимать точки максимальной температуры. Показано, что затухание осевых безразмерных параметров в сносимой поперечным потоком струе происходит быстрее по сравнению с затопленной и спутной.

Начальный участок струи условно можно разбить на две части: вертикальную и искривленную. На вертикальной части (на выходе из трубки)

струи в поперечном сечении имеет практически круглую форму, которая быстро деформируется и приобретает подковообразную форму. На оси струи образуется область перегиба. Скорость и температура в этой области увеличиваются по сравнению с величинами в устье источника. Этот рост составляет примерно 6–7 %. Далее на искривленной части возрастает влияние динамических характеристик сносящего потока, происходит большее вовлечение внешней массы воздуха в струю. Тепловой и динамический пограничный слои струи приобретают асимметричную форму в вертикальной плоскости относительно оси.

Список литературы

1. **Добросельский, К. Г.** Геометрия слаборапорной струи в поперечном воздушном потоке / К. Г. Добросельский // Теплофизика и аэромеханика. – 2001. – Т. 8, № 2. – С. 199–203.
2. **Аксенов, А. А.** Смешение газов при вдуве низконапорной струи в поперечный поток / А. А. Аксенов, А. В. Гудзовский, А. А. Дядькин, А. П. Тишин // Механика жидкости и газа. – 1996. – № 3. – С. 67–74.
3. **Абрамович, Г. Н.** Теория турбулентных струй / Г. Н. Абрамович, Т. А. Гиршович, С. Ю. Крашенинников и др. – М. : Наука, 1984. – 716 с.
4. **Гузеев, А. С.** Визуальное исследование взаимодействия круглой струи жидкости и сносящего потока / А. С. Гузеев, В. А. Тюшкевич, И. С. Христич // Сибирский физико-технический журнал. – 1993. – Вып. 1. – С. 142–146.
5. **Бузуков, А. А.** Действие сносящего потока газа на струю распыливаемой жидкости / А. А. Бузуков // Теплофизика и аэромеханика. – 1996. – Т. 3, № 1. – С. 65–72.
6. **Добросельский, К. Г.** Динамические и тепловые характеристики струи, вдуваемой по нормали к поперечному потоку / К. Г. Добросельский // Теплофизика и аэромеханика. – 2004. – Т. 11, № 1. – С. 79–86.
7. **Абрамович, Г. Н.** Турбулентное смешение газовых струй / Г. Н. Абрамович, С. Ю. Крашенинников, А. Н. Секундов, И. П. Смирнова. – М. : Наука, 1974. – 272 с.
8. **Добросельский, К. Г.** Вертикальная слаборапорная струя в поперечном потоке / К. Г. Добросельский // Инженерно-физический журнал. – 2005. – Т. 78, № 6. – С. 143–147.

Добросельский Константин Геннадьевич
старший научный сотрудник, Институт
теплофизики им. С. С. Кутателадзе
Сибирского отделения РАН
(г. Новосибирск)

Dobroselsky Konstantin Gennadyevich
Senior staff scientist, Institute of thermal
physics named after S.S. Kutateladze
of the Siberian department of the Russian
Academy of Sciences (Novosibirsk)

E-mail: dobroselsky@mail.ru

УДК 533.6.071.4

Добросельский, К. Г.

Теплофизические и динамические характеристики начального участка вертикальной низкоскоростной струи в поперечном потоке / К. Г. Добросельский // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 205–213.

АНАЛИЗ УРАВНОВЕШЕННОСТИ ДВИГАТЕЛЕЙ ВНУТРЕННЕГО СГОРАНИЯ С КРИВОШИПНО-ШАТУННЫМ МЕХАНИЗМОМ СО СДВОЕННЫМИ КИНЕМАТИЧЕСКИМИ СВЯЗЯМИ

Аннотация. Представлен анализ уравновешенности двигателей внутреннего сгорания, созданных на основе кривошипно-шатунного механизма со сдвоенными кинематическими связями, проведена оценка уравновешенности рядных двигателей внутреннего сгорания с двумя коленчатыми валами.

Ключевые слова: двигатель внутреннего сгорания, кривошипно-шатунный механизм, уравновешивание, крутящий момент, силы инерции.

Abstract. The article introduces a balance analysis of internal combustion engines constructed on the basis of a connecting rod gear with dual kinematic connections. The author estimates the balance of in-line internal combustion engines with two bent shafts.

Key words: internal combustion engine, connecting rod gear, balance, twisting moment, forces of inertia.

Введение

При работе двигателя внутреннего сгорания (ДВС) возникают силы, которые разделяются на уравновешенные и неуравновешенные [1, 2]. Силы, сумма которых равна нулю и которые при суммировании не дают момента, называют уравновешенными. К таким силам относят силы трения и силы давления газов в цилиндрах.

К неуравновешенным силам относят силы, которые передаются на опоры двигателя, а именно: вес двигателя; силы реакции выхлопных газов и движущихся жидкостей; сила тяги вентилятора; центробежные силы инерции вращающихся масс двигателя; касательные силы инерции вращающихся масс, возникающие при непостоянной угловой скорости вращения коленчатого вала; силы инерции возвратно-поступательно движущихся масс двигателя.

Силы давления газов P_g и вообще рабочего тела на тактах впуска, сжатия, расширения и выпуска совместно с силами инерции P_j формируют суммарную силу P_Σ , которую при переносе на ось поршневого пальца делят на две составляющие: силу P , действующую по оси шатуна, и нормальную силу N_Σ , поджимающую поршень к стенке цилиндра. Обе эти составляющие имеют знакопеременный характер. Сила N_Σ обуславливает повышенный неравномерный износ деталей цилиндропоршневой группы и, кроме того, создает опрокидывающий момент, который воспринимается опорами двигателя.

Реакция опор двигателя определяется с использованием принципа д'Аламбера, согласно которому действующие на систему внешние силы вместе с силами инерции движущихся масс системы уравновешиваются сопротивлением связей системы:

$$\begin{aligned}\sum R_{\text{оп}} &= \sum R_{\text{внеш}} - \sum R_{\text{и}}; \\ \sum M_{\text{оп}} &= \sum M_{\text{внеш}} - \sum M_{\text{и}},\end{aligned}$$

где $\sum R_{\text{оп}}$ – главный вектор системы сил реакции опор двигателя; $\sum M_{\text{оп}}$ – сумма моментов сил реакции опор двигателя; $\sum R_{\text{внеш}}$ – сумма внешних сил, действующих на двигатель; $\sum M_{\text{внеш}}$ – сумма моментов внешних сил, действующих на двигатель; $\sum R_{\text{и}}$ – сумма сил инерции движущихся деталей двигателя; $\sum M_{\text{и}}$ – сумма моментов сил инерции движущихся деталей двигателя.

Неуравновешенные силы, постоянные по величине и направлению, вибраций двигателя не вызывают. Сила тяги вентилятора, а также силы реакции выхлопных газов и движущихся жидкостей изменяются по величине очень мало и на уравновешенность двигателя практически не влияют. Неуравновешенные силы, переменные по величине и направлению, могут вызывать вибрации как двигателя, так и всего транспортного средства (агрегата), причем наибольшие вибрации вызываются силами инерции движущихся возвратно-поступательно и вращающихся масс двигателя. Вибрации двигателя при недостаточной жесткости его деталей могут возникнуть также под действием переменных сил давления газов. Эти вибрации устраняются увеличением жесткости деталей двигателя.

Для устранения отрицательных последствий, связанных с наличием вибраций (перегрузка двигателя, увеличенный износ его деталей, ослабление резьбовых соединений и пр.) ДВС должен быть динамически уравновешен. Динамическое уравновешивание двигателя заключается в создании такой системы сил, в которой действующие силы и моменты этих сил постоянны по величине и направлению или равны нулю.

Уравновешивание двигателей с классическим кривошипно-шатунным механизмом (КШМ) возможно двумя способами:

- 1) расположением определенным образом цилиндров и выбором такой кривошипной схемы коленчатого вала (порядка работы цилиндров), при которой уравновешивались бы силы инерции и их моменты;
- 2) созданием с помощью дополнительных масс (противовесов) новых сил, противоположных уравновешиваемым силам.

Как правило, оба эти способа применяются вместе.

1. Теоретическое обоснование способа уравновешивания

В цилиндрах ДВС с КШМ со сдвоенными кинематическими связями [3, 4] (рис. 1) силы давления газов P_g совместно с силами инерции P_i формируют суммарную силу P_{Σ} , которая прикладывается к осям левого A_l и правого A_r пальцев и передается на верхние головки левого ($P_{\Sigma}/2$) и правого ($P_{\Sigma}/2$) шатунов. На оси каждого пальца силы $P_{\Sigma}/2$ раскладывают на две составляющие силы: P_l и P_r , действующие по осям соответственно левого и правого шатунов, и две нормальные силы $N_{\Sigma l}$ и $N_{\Sigma r}$, направленные друг против друга и уравновешивающие друг друга. При этом исключается соударение стенок юбки поршня и цилиндра при перекладке поршня и их трение при движении поршня от одной мертвой точки к другой.

Кривошипно-шатунный механизм со сдвоенными кинематическими связями представляет собой динамическую систему с заложенным в ней третьим способом динамического уравновешивания. Так, касательные силы инерции вращающихся масс двигателя, возникающие при непостоянной угловой скорости вращения коленчатых валов с маховиками, кинематически связанных между собой посредством зубчатой передачи, будут взаимно

уравновешиваться или увеличивать нагрузку на опоры двигателя, не вызывая колебательных процессов.

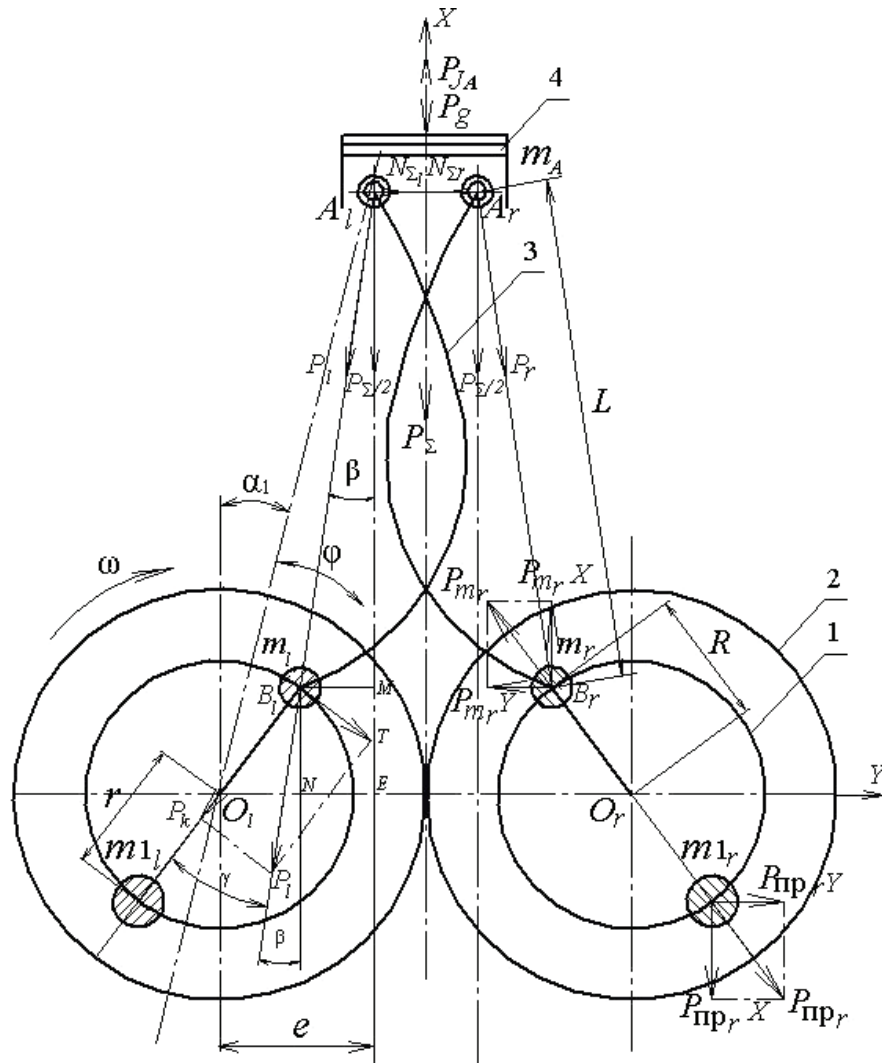


Рис. 1. Расчетная схема сил, действующих в ДВС с КШМ со сдвоенными кинематическими связями:
1 – кривошип коленчатого вала; 2 – маховик; 3 – шатун; 4 – поршень

Кроме того, во всех классических двигателях возникает переменный реактивный момент M_R при любом положении коленчатого вала, равный по величине, но противоположный по направлению крутящему моменту $M_{кр}$. В обычных двигателях этот момент уравновесить невозможно и он всегда передается на опоры двигателя, вызывая его значительные механические колебания.

В КШМ со сдвоенными кинематическими связями крутящий момент формируется встречно вращающимися коленчатыми валами с маховиками, связанными между собой посредством зубчатой передачи. Следовательно, реактивный момент будет оказывать значительно меньшее влияние на опоры

двигателя. В классических ДВС проблема снижения негативного воздействия переменного реактивного момента M_R решается путем увеличения числа цилиндров и обеспечения в них одинакового протекания рабочих процессов и равномерного чередования вспышек.

К числу наиболее значимых сил и моментов, требующих уравнивания, относятся:

– P_{jA}^I – гармонически изменяющаяся сила инерции первого порядка возвратно-поступательно движущихся масс, сосредоточенных на оси (осях) поршневого пальца (пальцев);

– P_{jA}^{II} – гармонически изменяющаяся сила инерции второго порядка возвратно-поступательно движущихся масс;

– $P_{jMA} = P_{jA}^I + P_{jA}^{II}$ – гармонически изменяющаяся сила инерции возвратно-поступательно движущихся масс;

– P_{jB} – центробежная сила инерции вращающихся масс двигателя;

– M^I – момент от сил инерции первого порядка;

– M^{II} – момент от сил инерции второго порядка;

– M_B – момент от сил инерции вращающихся масс.

Особенно значительные вибрации могут вызываться неравномерным реактивным моментом M_R и гармонически изменяющимися силами инерции и их моментами при резонансе.

При учете приведенных выше сил и моментов условия уравновешенности двигателя имеют следующий вид:

$$\sum P_{jA}^I = 0; \sum P_{jA}^{II} = 0; \sum P_{jB} = 0;$$

$$\sum M^I = 0; \sum M^{II} = 0; \sum M_B = 0.$$

2. Расчет сил инерции и уравнивающих сил в рядных ДВС

Уравнивание сил инерции вращающихся масс КШМ со сдвоенными кинематическими связями, аналогично уравниванию сил инерции вращающихся масс классического КШМ, достигается установкой одинарных или парных противовесов $m'l = m'l + m''l$ на продолжении правой и левой шеек кривошипов коленчатых валов на расстоянии r от осей их вращения (рис. 2): для левого коленчатого вала: $m'l_l = m'l_l + m''l_l$; для правого: $m'l_r = m'l_r + m''l_r$.

Для обеспечения динамической уравновешенности:

– равнодействующая центробежных сил инерции масс $m'l_l$ и $m'l_r$ и центробежные силы инерции масс m_{bl} и m_{br} , сосредоточенных на осях шатунных шеек правого B_r и левого B_l кривошипов, должны быть равными по величине и противоположными по направлению:

$$m'l_l \cdot r \cdot \omega^2 = m_{bl} \cdot R \cdot \omega^2,$$

$$m'l_r \cdot r \cdot \omega^2 = m_{br} \cdot R \cdot \omega^2; \quad (1)$$

– должна быть равна нулю сумма моментов этих сил относительно любой точки, расположенной на оси коленчатого вала; для точки А:

$$m'l_l \cdot r \cdot \omega^2 \cdot a = m''l_l \cdot r \cdot \omega^2 \cdot b,$$

$$m'l_r \cdot r \cdot \omega^2 \cdot a = m''l_r \cdot r \cdot \omega^2 \cdot b, \quad (2)$$

где ω – угловая скорость вращения коленчатых валов.

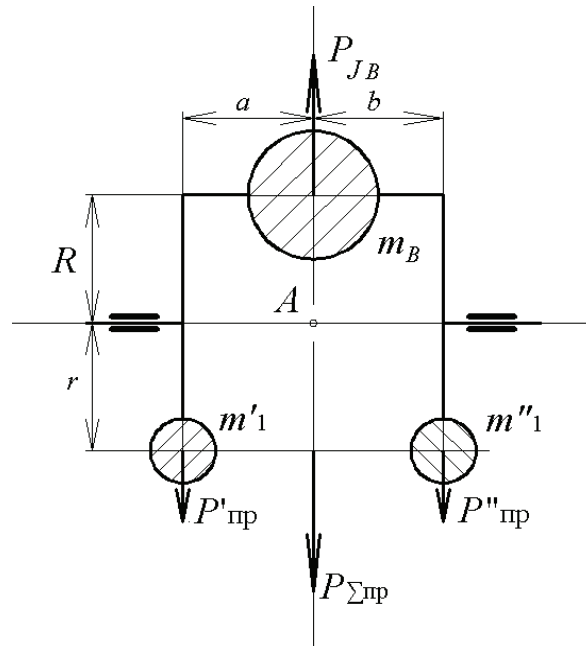


Рис. 2. Расчетная схема динамического уравнивания вращающейся массы кривошипа коленчатого вала

Из вышеприведенных уравнений следует, что

$$m_l = m_B \frac{R}{r},$$

$$m'_1 = m''_1.$$

Уравнивание сил инерции возвратно-поступательно движущихся масс может быть полным или частичным. При полном уравнивании одноцилиндрового двигателя используют метод Ланчестера. Основной идеей данного метода является установка двух дополнительных масс на двух валиках, вращающихся в противоположные стороны с угловой скоростью, равной угловой скорости вращения коленчатого вала. Сумма сил инерции по оси Y (направление движения возвратно-поступательно движущихся масс) от вращения дополнительных масс должна быть равной значению силы инерции первого порядка от возвратно-поступательно движущихся масс P_{jA}^I .

Сила инерции второго порядка от возвратно-поступательно движущихся масс P_{jA}^{II} уравнивается аналогично – установкой двух дополнительных масс на двух валиках, вращающихся в противоположные стороны с удвоенной угловой скоростью вращения коленчатого вала. Горизонтальные составляющие от сил инерции дополнительных вращающихся масс взаимно уравниваются.

Однако, учитывая сложность его реализации на практике, данный метод широкого распространения не получил.

В многоцилиндровых рядных двигателях силы P_{jA}^I и P_{jA}^{II} действуют в одной плоскости, параллельны друг другу и могут быть приведены к главной силе и главному моменту. В зависимости от числа цилиндров и формы коленчатого вала эти силы могут быть полностью или частично уравновешены.

В многоцилиндровом двигателе с КШМ со сдвоенными кинематическими связями каждый цилиндр можно рассматривать как i -й одноцилиндровый двигатель (рис. 1).

Силы инерции первого и второго порядков будут определяться аналогично классическому одноцилиндровому двигателю [1, 2, 4]:

$$P_{JA}^I = -m_A R \omega^2 \cos(\varphi + \arcsin k_3);$$

$$P_{JA}^{II} = -m_A \cdot \left(1 - (k_2 - \lambda \sin(\varphi + \arcsin k_3))^2\right)^{-3/2} \left[L \omega^2 (k_2 \lambda \sin(\varphi + \arcsin k_3) + \right.$$

$$+ \lambda^2 \cos 2(\varphi + \arcsin k_3))^2 \times \left(1 - (k_2 - \lambda \sin(\varphi + \arcsin k_3))^2\right) +$$

$$\left. + L \omega^2 \left(k_2 \lambda \cos(\varphi + \arcsin k_3) - \frac{\lambda^2}{2} \sin(\varphi + \arcsin k_3) \right)^2 \right],$$

где $k_2 = e/L$ – второй дополнительный коэффициент кинематического подобия КШМ; $k_3 = e/(R + L)$ – третий дополнительный коэффициент кинематического подобия КШМ; ω – угловая скорость вращения коленчатых валов.

Графики этих сил для случая, когда $n = 1500 \text{ мин}^{-1}$, $m_A = 60 \text{ кг}$, будут иметь вид (рис. 3, 4). При этом по аналогии с классическим ДВС силу $P_{JA}^{II}(\varphi)$ будем называть силой инерции второго порядка.

Особенностью графиков сил инерции является их сдвиг относительно нулевого положения на угол α_1 . Кроме этого, чередование величин амплитуды силы инерции второго порядка $P_{JA}^{II}(\varphi)$ обусловлено несимметричностью кривой ускорения поршня J_{Π} .

Суммарная сила инерции возвратно-поступательно движущихся масс будет определяться как

$$Pj_{m_A} = -m_A \left\{ \left(1 - (k_2 - \lambda \sin(\varphi + \arcsin k_3))^2\right)^{-3/2} \left[L \omega^2 (k_2 \lambda \sin(\varphi + \arcsin k_3) + \right.\right.$$

$$+ \lambda^2 \cos 2(\varphi + \arcsin k_3))^2 \cdot \left(1 - (k_2 - \lambda \sin(\varphi + \arcsin k_3))^2\right) +$$

$$\left. + L \omega^2 \left(k_2 \lambda \cos(\varphi + \arcsin k_3) - \frac{\lambda^2}{2} \sin(\varphi + \arcsin k_3) \right)^2 \right] + R \omega^2 \cos(\varphi + \arcsin k_3) \right\}.$$

График суммарной силы инерции возвратно-поступательно движущихся масс представлен на рис. 5.

Сила инерции возвратно-поступательно движущихся масс может быть частично уравновешена за счет применения дополнительных масс противовесов $m_{\Pi P}$ на продолжении шек кривошипа:

$$Pj_{\Pi P} = m_{\Pi P} \cdot R \cdot \omega^2 \cos(\varphi + \alpha_1).$$

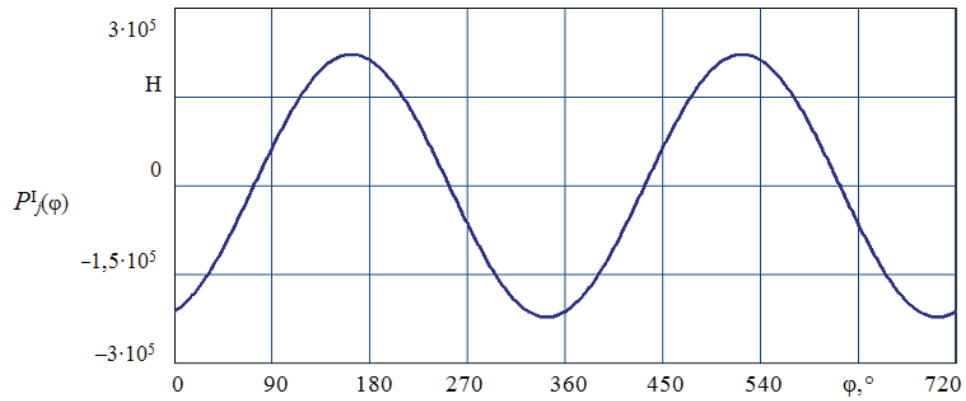


Рис. 3. Силы инерции первого порядка возвратно-поступательно движущихся масс

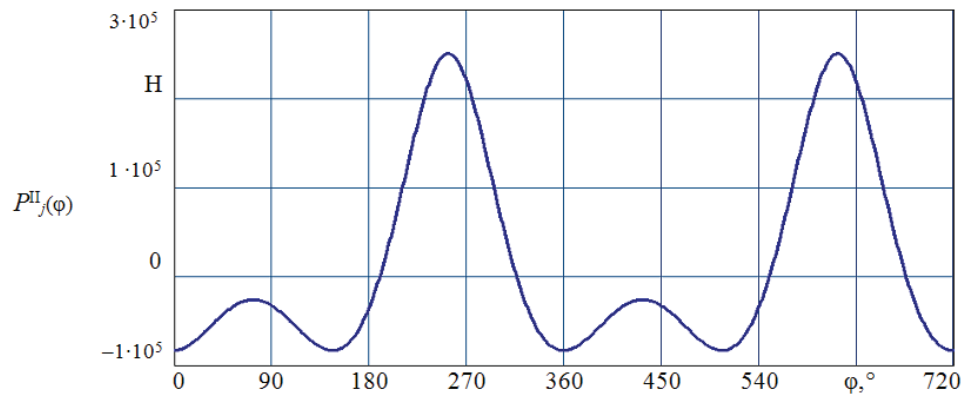


Рис. 4. Силы инерции второго порядка возвратно-поступательно движущихся масс

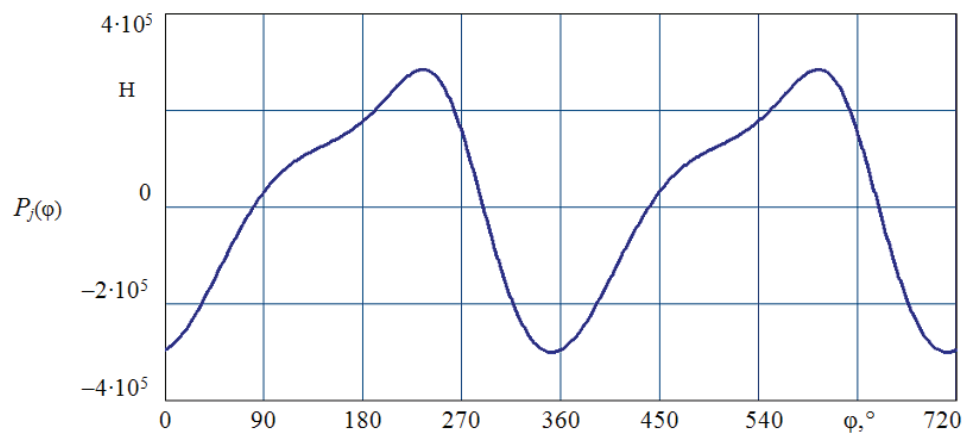


Рис. 5. График суммарной силы инерции возвратно-поступательно движущихся масс в одноцилиндровом ДВС

График уравнивающей силы представлен на рис. 6.

С учетом уравнивающей силы $P_{j_{\text{пр}}}$ график суммарной силы инерции возвратно-поступательно движущихся масс в одноцилиндровом ДВС примет вид (рис. 7)

$$PJ_{ur1} = PJ + PJ_{пр.}$$

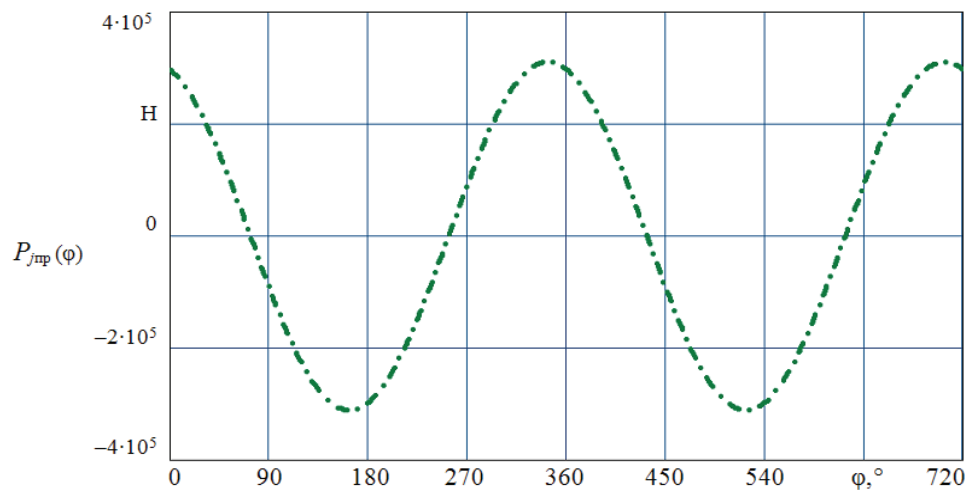


Рис. 6. Уравновешивающая сила от инерции вращающихся масс дополнительных противовесов

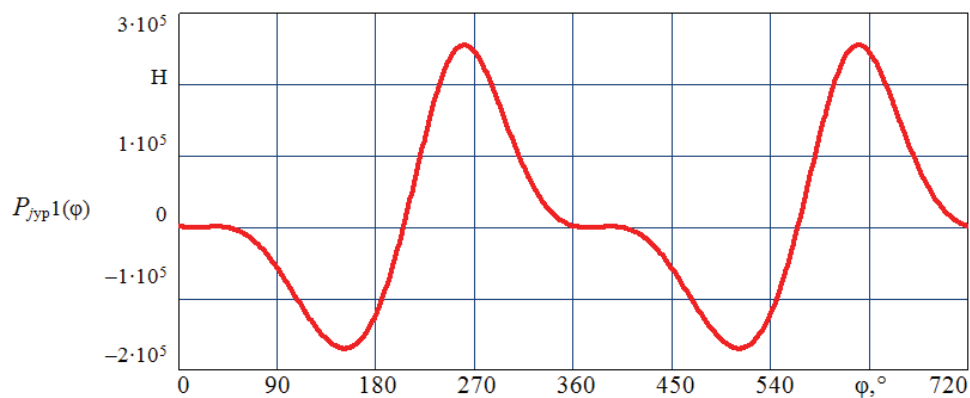


Рис. 7. График суммарной силы инерции возвратно-поступательно движущихся масс с учетом уравновешивания

Из анализа кривой суммарной силы инерции возвратно-поступательно движущихся масс одноцилиндрового ДВС с учетом уравновешивания следует, что при движении поршня от верхней мертвой точки (ВМТ) к нижней мертвой точке (НМТ) на тактах впуска и расширения при значениях угла поворота кривошипа φ от 0 до 50° сила инерции возвратно-поступательно движущихся масс полностью уравновешена. Далее начинается ее нарастание в отрицательную область, при положении кривошипа $\varphi = 145^\circ$ она достигает максимума и начинается ее уменьшение. При достижении поршнем НМТ (угол $\varphi = 197^\circ$) суммарная сила инерции имеет еще отрицательное значение и ее вектор совпадает с направлением движения поршня к ВМТ.

При движении поршня к ВМТ примерно в течение 45° поворота кривошипа коленчатого вала наблюдается возрастание суммарной силы инерции

возвратно-поступательно движущихся масс. При значении $\varphi = 240^\circ$ достигается ее максимальное значение и начинается ее резкий спад.

При подходе поршня к ВМТ на такте сжатия суммарная сила инерции возвратно-поступательно движущихся масс принимает нулевое значение, которое сохраняется на $1/8$ поворота коленчатого вала на такте расширения. При дальнейшем повороте кривошипа начинается плавное увеличение суммарной силы инерции возвратно-поступательно движущихся масс и при значении $\varphi = 510^\circ$ достигается ее максимальное значение. Но в этом случае угол отклонения шатуна β приближается к максимальному значению и вектор силы, действующей по оси шатуна, приближается к линии действия суммарных сил, что способствует росту величины тангенциальной силы T , а следовательно, и крутящего момента.

В двух- и четырехцилиндровых ДВС с КШМ со сдвоенными кинематическими связями силы инерции первого и второго (первого и второго, третьего и четвертого для четырехцилиндрового) цилиндров смещены относительно нулевого положения на 180° . Это обуславливает возникновение суммарных сил инерции, имеющих гармонические колебания с периодом повторения 180° , аналогичных силам инерции второго порядка, возникающим в двигателе с центральным КШМ. График этих сил представлен на рис. 8.

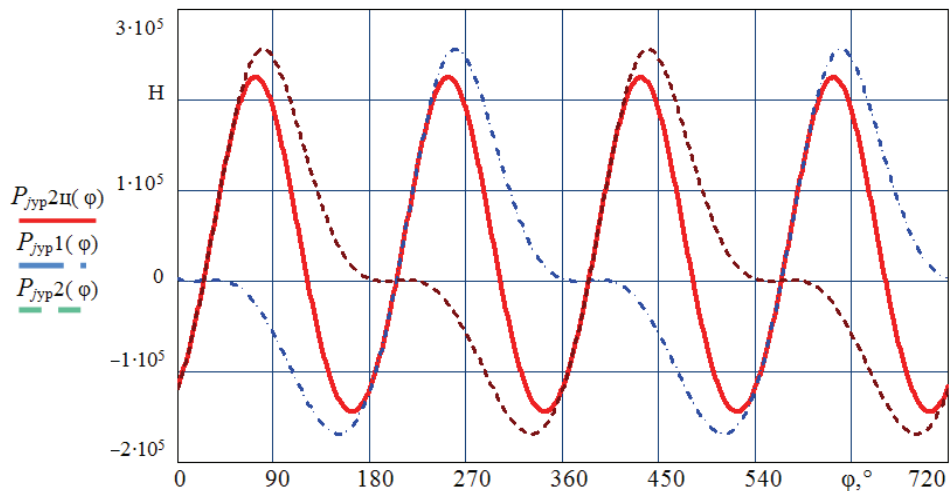


Рис. 8. График суммарных сил инерции возвратно-поступательно движущихся масс двухцилиндрового ДВС

Как показывает опыт, уравновесить эти силы возможно посредством двух дополнительных масс, вращающихся с удвоенной угловой скоростью.

Для трех- и шестицилиндровых ДВС силы инерции первого, второго и третьего (первого и шестого, второго и пятого, третьего и четвертого) цилиндров смещены относительно нулевого положения на 120° . Это обуславливает возникновение суммарных сил инерции возвратно-поступательно движущихся масс, имеющих гармонические колебания с периодом повторения 120° . График этих сил представлен на рис. 9.

Следует обратить внимание на то, что кривая суммарных сил находится в положительной области графика и вектор ее действия совпадает с вектором перемещения поршня в начале его пути к НМТ на тактах впуска и расширения, что способствует увеличению суммарных сил P_Σ , действующих на поршень.

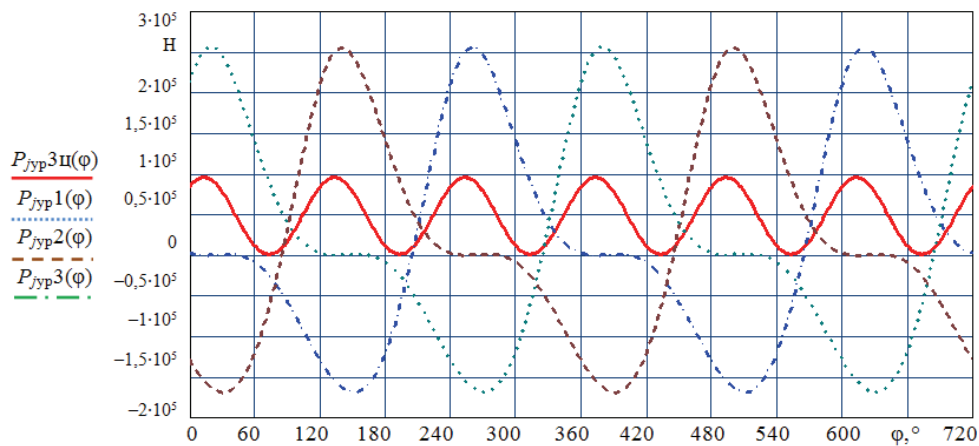


Рис. 9. График суммарных сил инерции возвратно-поступательно движущихся масс трехцилиндрового ДВС

Для пятицилиндрового рядного ДВС кривая суммарных сил (рис. 10) также представляет собой синусоиду в положительной области графика с периодом повторения 72° .

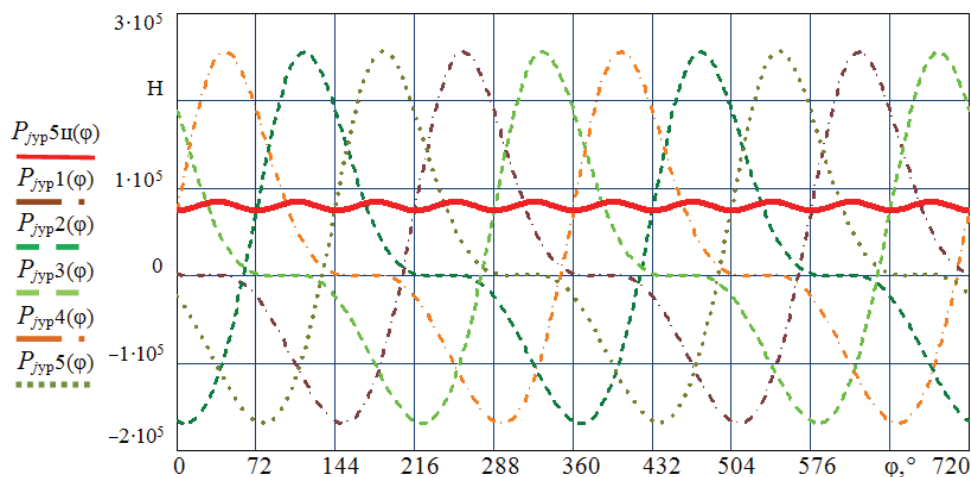


Рис. 10. Кривая суммарных сил инерции возвратно-поступательно движущихся масс пятицилиндрового рядного ДВС

Вектор направления действия суммарной силы совпадает с вектором направления движения одного из пяти поршней при его перемещении от ВМТ к НМТ на тактах впуска и расширения. При необходимости снижения влияния этих сил на формирование колебательных процессов для ДВС с КШМ со сдвоенными кинематическими связями целесообразно применение дополнительного груза, установленного на свободно вращающемся валу.

Заключение

Таким образом, применение КШМ со сдвоенными кинематическими связями не решает весь комплекс проблем с уравниванием рядных ДВС. Однако оно позволяет уравновесить:

– нормальную составляющую N_{Σ} , обуславливающую величину сил трения поршня о цилиндр;

– касательные силы инерции вращающихся масс двигателя, возникающие при непостоянной угловой скорости вращения коленчатых валов с маховиками;

– силы инерции первого порядка P'_j возвратно-поступательно движущихся масс.

Вместе с тем оно позволяет значительно снизить влияние неравномерного реактивного момента M_R .

Кроме этого в предлагаемом КШМ происходит сдвиг суммарных сил инерции возвратно-поступательно движущихся масс со смещением направления вектора их действия по отношению к направлению вектора перемещения поршня, что позволяет снизить негативное воздействие сил инерции возвратно-поступательно движущихся масс на силу давления рабочего тела на поршень.

Список литературы

1. Автомобильные двигатели / под ред. М. С. Ховаха. – М. : Машиностроение, 1977. – 591 с.
2. **Колчин, А. И.** Расчет автомобильных и тракторных двигателей : учебное пособие для вузов / А. И. Колчин В. П., Демидов. – 2-е изд., перераб. и доп. – М. : Высшая школа, 1980. – 400 с.
3. Патент 2382891 Российская Федерация. Кривошипно-шатунный механизм со сдвоенными кинематическими связями / Грабовский А. А., Грабовский А. А. – № 2008110199/06 ; заявл. 17.03.2008 ; опубл. 27.02.2010.
4. **Грабовский, А. А.** Кинематика кривошипно-шатунного механизма с двумя коленчатыми валами / А. А. Грабовский, Е. С. Аверьянова / Известия МГТУ «МАМИ». – 2009. – № 2 (8). – С. 317.

Грабовский Александр Андреевич
кандидат технических наук, доцент,
кафедра транспортных машин,
Пензенский государственный
университет

E-mail: algra888@yandex.ru

Grabovsky Alexander Andreevich
Candidate of engineering sciences,
associate professor, sub-department
of transport machines, Penza
State University

УДК 621.33; 621.43

Грабовский, А. А.

Анализ уравниваемости двигателей внутреннего сгорания с кривошипно-шатунным механизмом со сдвоенными кинематическими связями / А. А. Грабовский // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 214–224.

ДВУХКООРДИНАТНЫЙ ДАТЧИК ПОЛОЖЕНИЯ

Аннотация. Рассмотрено аппаратное и алгоритмическое обеспечение двухкоординатного датчика положения, работающего на принципе определения абсолютных координат расположения объекта при отсутствии механического контакта частей первичного измерительного преобразователя. Проведена оценка причин возникновения погрешностей.

Ключевые слова: измерение перемещения в абсолютных координатах, бесконтактное измерение, двумерные измерения, измерения перемещений и углов, датчик положения.

Abstract. The article considers hardware and knoware of a two-dimensional sensor of displacement. The sensor works on a principle of measurement of absolute coordinates of an object without mechanical contact of the primary measuring transducer parts. The authors estimate the causes of errors.

Key words: measurement of motion in absolute coordinates, noncontact sensing, two-dimensional measurements, measurements of displacement and angles, position sensor.

Введение

В современной промышленности часто требуется измерение перемещений на прямой и в плоскости с минимизацией воздействия на измеряемый объект. Иногда необходимо измерять не непосредственно значение перемещения, а отклонение от какого-то заданного значения [1].

Известен способ измерения координат точки на плоскости, при котором определяется приращение каждой координаты по отметкам, нанесенным на плоскость измерений [2]. В 1999 г. в фирме Agilent Technologies этот способ был модифицирован и в настоящее время он состоит в том, что при движении оптический сенсор последовательно считывает изображения поверхности и затем сравнивает их и по разнице между этими изображениями определяет измерение координаты. Данный способ относится к инкрементным, т.е. измеряющим относительную, а не абсолютную координату.

Однако во многих случаях (например, в системах инерционной навигации, следящих системах наведения и др.) существует необходимость оперативного определения именно абсолютных координат. Например, в процессе разработки балансирующего устройства, работающего на принципе измерения отклонения маятника от вертикали, возникающего из-за наличия дисбаланса на балансируемом роторе, закрепленном на маятнике, возникла практическая задача – определить положение точки пересечения оси маятника и плоскости датчика одновременно по двум координатам в радиусе 5 мм от центра с максимальной погрешностью d , не более 0,015 мм (рис. 1). Измерение должно происходить без взаимного воздействия перемещающихся частей датчика. Дополнительным условием является необходимость измерений с постоянной точностью в большом диапазоне (до 10 мм).

1. Возможные варианты решения задачи

Рассматривались варианты измерения с использованием:

– индуктивного датчика положения;

- емкостного датчика положения;
- оптического триангуляционного датчика положения;
- оптического инкрементного датчика положения.

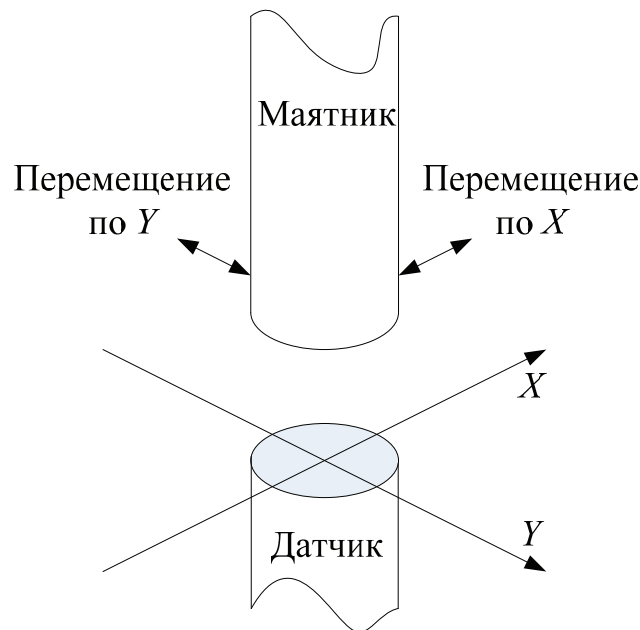


Рис. 1. Определение положения оси маятника

Индуктивный датчик положения имеет ряд недостатков. Общепринято, что рабочий диапазон изменения в таком датчике составляет не более 10 % от общей величины магнитного зазора. Это приводит к увеличению значения магнитного (рабочего) зазора. Дополнительным минусом является нелинейность такого преобразователя и необходимость обеспечения процедуры разделения координат. Это значит, что сигнал должен зависеть от перемещения в перпендикулярном зазору направлении, но не должен зависеть от перемещения в параллельном зазору направлении, что трудно обеспечить при использовании такого принципа. Также практически любой индуктивный датчик создает дополнительную, пусть и небольшую, силу, действующую на объект измерения [3].

Емкостной датчик имеет аналогично индуктивному такие недостатки, как большую величину рабочего зазора и сложность обеспечения независимого определения каждой координаты [4].

Оптический триангуляционный датчик обеспечивает достаточную точность измерения каждой отдельной координаты, но, как и индуктивный, и емкостной датчики, не обеспечивает однозначного и просто реализуемого независимого определения координат на плоскости. Так, в частности, координаты на плоскости достаточно точно можно определить с помощью двух датчиков только в ситуации, если деталь в горизонтальном сечении имеет форму окружности, либо если деталь имеет известную форму и движется без изменения угловой координаты относительно оси, перпендикулярной плоскости измерения датчика. Дополнительным аргументом против применения триангуляционных датчиков является высокая сложность их конструкции [5].

Оптический инкрементный датчик положения, основанный на описанном выше способе, достаточно хорошо известен и применяется, например, в манипуляторах типа «компьютерная мышь». Его достоинством является простота конструкции и, соответственно, дешевизна, однако такой датчик не исключает «проскальзываний», т.е. накопления в процессе перемещения ошибки. Кроме того, датчик не дает однозначной точки на плоскости, он лишь показывает текущее смещение от ранее сохраненного положения, что требует постоянной установки нуля.

Таким образом, применение всех вышеперечисленных датчиков не решает задачи измерений перемещений на плоскости с минимизацией воздействия на измеряемый объект.

2. Предлагаемое решение задачи

Для решения задачи измерений перемещений на плоскости с минимизацией воздействия на измеряемый объект предлагается установить в качестве датчика видеокамеру, направленную на маркированную определенным образом плоскость. Обработка картинки с видеокамеры позволит «привязать» видимую область изображения ко всему рисунку маркированной плоскости, и таким образом определить координаты центра картинки на плоскости. Так как перемещения малы, камера должна обеспечивать достаточное увеличение картинки, поэтому в каждый момент времени на матрицу видеокамеры будет попадать лишь малая часть изображения маркированной плоскости.

Оценим требования к камере и плоскости. Эти требования зависят друг от друга, поэтому необходимо задаться некоторыми значениями, приняв их как данность.

На настоящий момент самым дешевым и самым распространенным типоразмером матрицы для простых видеокамер является 640×480 пикселей (0,3 мегапикселя). Предположим, что алгоритм позволит определять положение центра видимой картинки на плоскости с точностью один пиксел. Можно оценить требуемый размер видимой картинки по наименьшей стороне:

$$P_{st} \leq Y \cdot d_t, \quad (1)$$

где P_{st} – размер видимой картинки по вертикали; Y – количество пикселей по наименьшей стороне матрицы; d_t – необходимая погрешность измерения.

Для указанных выше параметров получается высота видимой картинки не более 7,2 мм, это даже несколько больше размера картинки имеющихся в широкой продаже дешевых USB-микроскопов, один из которых как раз и предлагается использовать в качестве камеры.

3. Маркировка плоскости

При маркировке плоскости обеспечивались следующие условия:

- поверхность содержит минимум оптического «мусора»;
- поверхность имеет равномерное освещение.

Рассматривались различные варианты маркировки плоскости, в том числе разбивка плоскости на квадраты линиями и маркировка различными фигурами, установленными в пересечениях сетки, а также использование цветных изображений. Однако после анализа доступных алгоритмов распознавания маркеров было решено, что наиболее легко распознаваемое изобра-

жение – ряд закрашенных несовпадающих кругов разного диаметра с цветом, максимально контрастирующим с общим фоном. Использование цветовой составляющей особого смысла не имеет, к тому же, как правило, видеокамеры имеют урезанные возможности по восприятию цветовой информации (например, кодируется один цвет на 4 пиксела).

На рис. 2 приведено изображение, полученное с видеокамеры. Для привязки к масштабу и положению видимой картинке на плоскости вся рабочая область поверхности покрыта с равномерным шагом H_c по обеим осям базовыми метками – закрашенными темными кругами диаметра D_{c1} . Внутри каждого квадрата, образованного базовыми метками, расположены закрашенные круги меньшего диаметра D_{c2} (маркировочные метки), наличие или отсутствие которых в заданных местах кодирует начальную координату X_{sq}, Y_{sq} по горизонтали и по вертикали. Код координат X_{sq}, Y_{sq} защищен контрольной суммой CS .

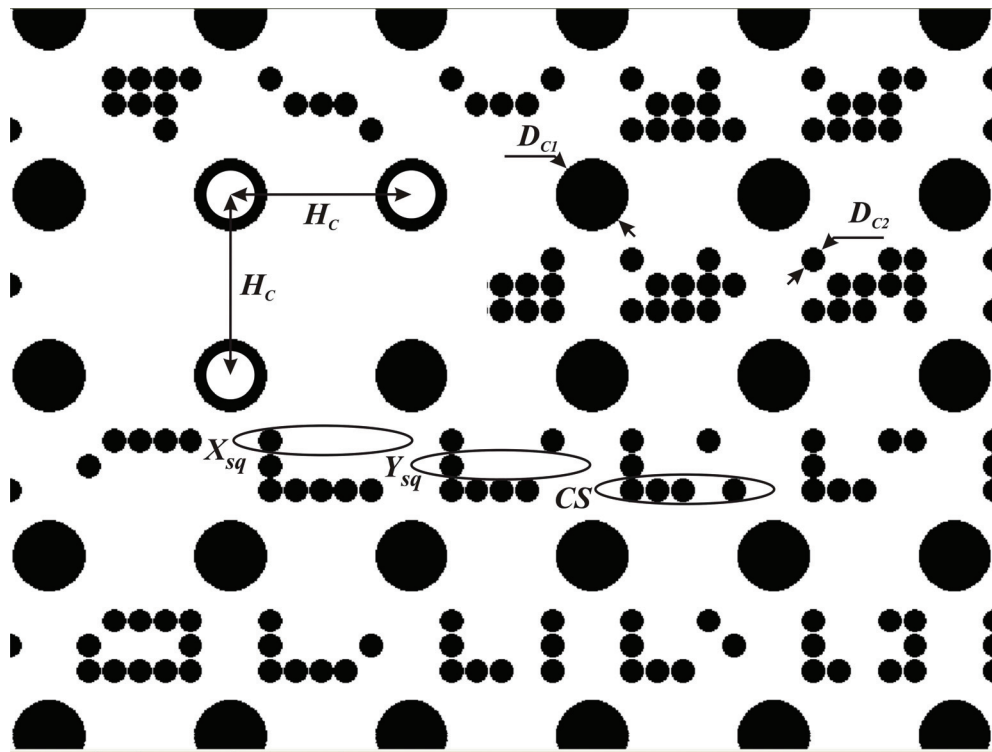


Рис. 2. Вид маркированной плоскости (вырезка)

Параметр H_c выбирается таким образом, чтобы при любом положении видимого кадра на плоскости по вертикали в него попало как минимум две базовых метки:

$$H_c \leq \frac{P_{st}}{3}.$$

Код (номер) квадрата задается в двоичном формате, для чего все поле квадрата разбито условными линиями с одинаковым шагом по вертикали и по горизонтали. Шаг выбран таким образом, чтобы получаемое число верти-

кальных линий было бы больше минимум на две, чем количество требуемых для кодировки номера квадрата двоичных разрядов. Линии проводятся таким образом, чтобы центр квадрата обязательно совпадал с пересечением линий.

Пример маркировки квадрата приведен на рис. 3. Крайние линии со всех сторон не используются, шаг разбиения равен $h_m = \frac{H_c}{7}$. Три центральными линиями, как указано на рисунке, кодируется X_{sq} , Y_{sq} и контрольная сумма CS . При такой маркировке полный размер плоскости не превышает 32 квадратов по вертикали и 32 квадратов по горизонтали, что ограничено числом двоичных разрядов в значениях X_{sq} , Y_{sq} . Контрольная сумма рассчитывается в двоичной системе счисления по формуле

$$CS = \overline{(X_{sq} + Y_{sq})}. \quad (2)$$

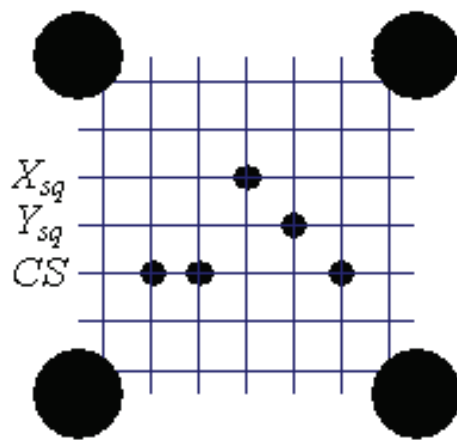


Рис. 3. Разбиение квадрата для маркировки

К размерам меток предъявляются следующие требования:

- диаметр базовых меток должен быть значительно (не менее чем в два раза) больше диаметра меток маркировки;
- базовые метки не должны соприкасаться с маркировочными, даже при наличии оптического «мусора» на картинке. Слияние маркировочных меток в небольших пределах допустимо, поскольку разделения меток не производится. Для считывания кода используется анализ средней яркости в центре пересечения линий и сравнение с заданным пороговым значением.

4. Алгоритм расчета положения

На вход алгоритма расчета поступает изображение с камеры. Под изображением здесь и далее следует понимать фиксируемую камерой часть маркированной плоскости. Наблюдаемое изображение может иметь несколько неоднородную засветку, а также ряд искажений (например, ориентация изображения может не совпадать с осями координат). Ряд искажений картинки зависит от способа создания маркированной плоскости. В работе использовалась печать с помощью струйного принтера на качественной фотобумаге. Даже для качественной бумаги при таком разрешении видна фактура бумаги,

также видны отдельные пиксеты, обусловленные способом печати принтера. Пример входной картинке с увеличением в 25 раз приведен на рис. 4.

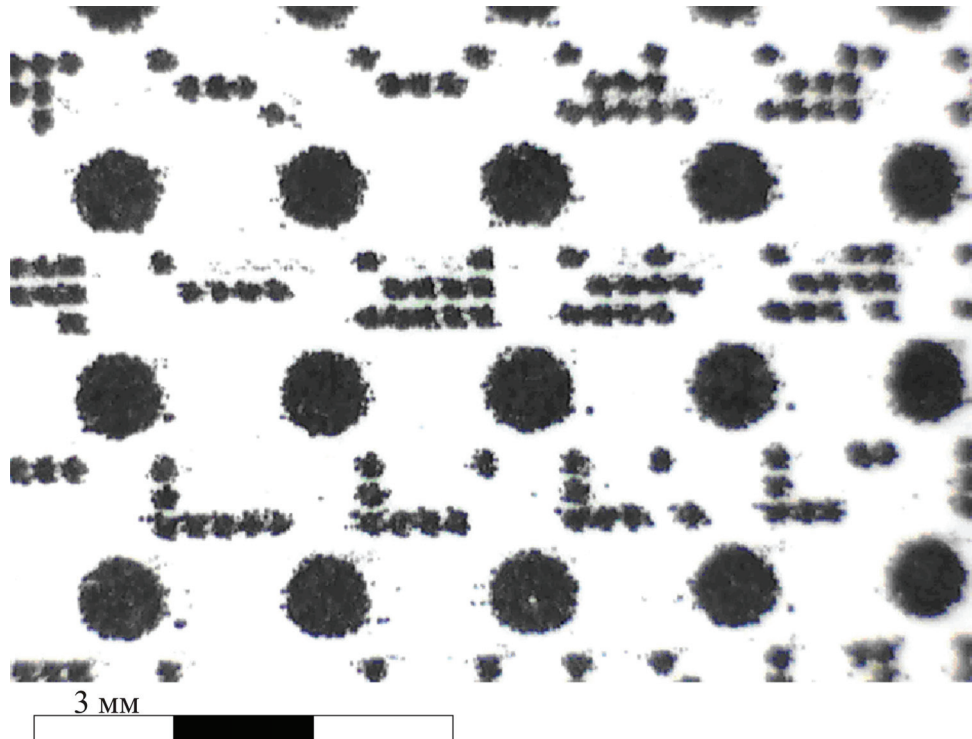


Рис. 4. Изображение маркированной плоскости

- Алгоритм определения положения включает в себя следующие стадии:
1. Разделение изображения на «фон» и «метки».
 2. Поиск всех базовых меток на изображении.
 3. Поиск четырех наиболее близких к центру изображения базовых меток.
 4. Определение координат центра квадрата, вектора вертикали и размера стороны квадрата, образованного данными метками.
 5. По данным квадрата определение шага линий маркировки и считывание кода положения квадрата на плоскости.
 6. Расчет положения центра видимого изображения на плоскости с использованием полученных данных.

5. Реализация алгоритма

Разделение изображения на «фон» и «метки» осуществляется наиболее простым способом – по гистограмме яркостей изображения. Используется только одна цветовая составляющая, так как изображение одноцветное. Строится гистограмма яркостей, после чего производится группировка всех яркостей в два кластера. Начальными точками при кластеризации являются самый яркий и самый темный набор точек на изображении. После разбиения на два кластера используется средняя яркость кластеров для вычисления граничной яркости, для которой все более яркие точки являются фоном, а менее яркие – метками.

Результат работы алгоритма удовлетворителен при достаточно равномерном освещении и при любой яркости. Как показала практика, алгоритм не требует настройки.

Поиск базовых меток на картинке осуществляется последовательной заливкой всех областей, имеющих цвет «метка». Для всех точек, отнесенных к очередной заливкой области, определяется среднее арифметическое координат, что дает центр масс области X_{o_i}, Y_{o_i} . Количество точек дает площадь области S_{o_i} . По полученной площади области производится первоначальная фильтрация – небольшие области отбрасываются. Если площадь достаточна, по ее значению определяется условный диаметр окружности D_i , затем производится подсчет точек, попавших в условный диаметр. Если число попавших точек больше 80 % от общего числа точек области, считается, что область имеет форму окружности, и данная область является базовой меткой. Параметры метки – координаты X_{o_i}, Y_{o_i} – заносятся в массив меток.

На рис. 5 приведен пример результата работы алгоритма с изображением.

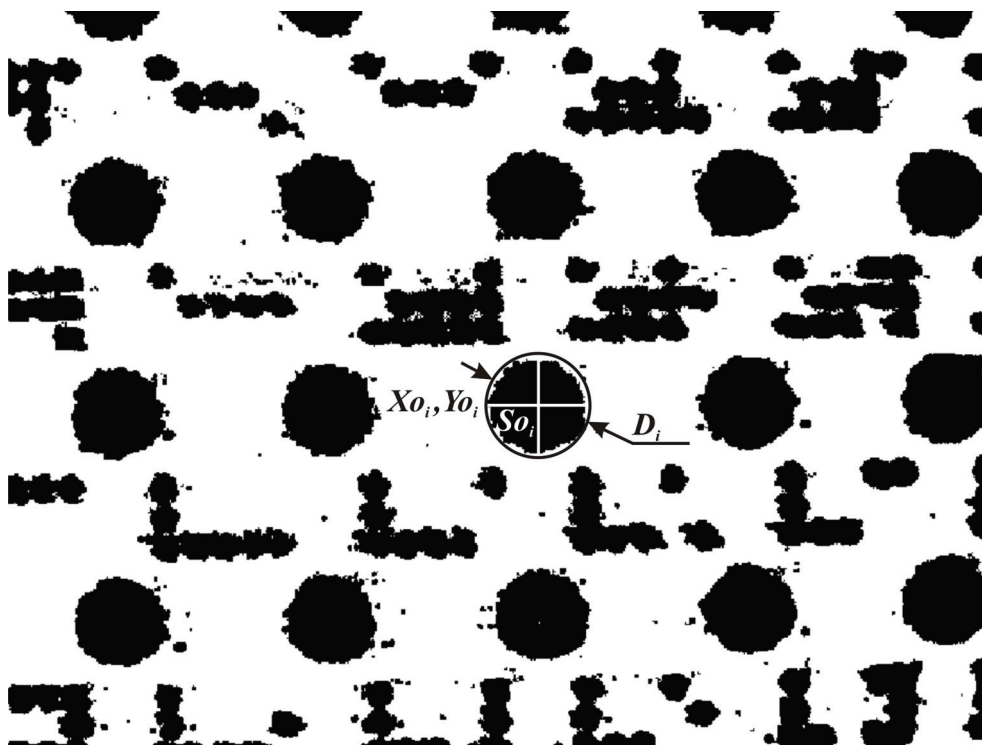


Рис. 5. Результат кластеризации и поиска меток

Темные области на рисунке принадлежат меткам, светлые – фону. Обозначенная метка имеет координаты X_{o_i}, Y_{o_i} и диаметр D_i , а количество принадлежащих этой метке пикселей черного цвета равно площади окружности S_{o_i} .

Поиск четырех наиболее близких к центру изображения меток, составляющих квадрат, осуществляется достаточно сложным образом. Дело в том, что просто четыре наиболее близких к центру изображения меток могут принадлежать разным квадратам, а требуется найти именно углы одного квадра-

та. Поэтому первоначально определяются три наиболее близкие метки к центру изображения. Далее из этих трех точек выбираются две такие, расстояние между которыми максимально – это будет диагональ квадрата. Третья метка отбрасывается.

Для диагонали квадрата рассчитываются прогнозируемые координаты двух точек, составляющих вторую диагональ квадрата, и для обеих расчетных точек производится поиск ближайшей метки. В результате определяются индексы, следовательно, координаты четырех точек – x_1, y_1, x_3, y_3 – как одна диагональ квадрата; x_2, y_2, x_4, y_4 – как вторая диагональ квадрата. Так как определение координат точек происходит с ошибкой, данные диагонали рассматриваются как диагонали двух различных квадратов. Описанный ниже расчет производится для обоих квадратов.

Все вычисления проводятся в системе координат, в которой точка (0;0) находится в левом верхнем углу экрана, координата X увеличивается при движении вправо, а Y – при движении вниз.

Исходными данными являются координаты x_1, y_1, x_3, y_3 диагонали квадрата. Остальные данные вычисляются по приведенным ниже формулам.

Центр квадрата cx, cy вычисляется как среднее арифметическое между координатами меток диагонали:

$$cx = \frac{x_1 + x_3}{2}, \quad cy = \frac{y_1 + y_3}{2}.$$

Недостающие координаты вершин квадрата x_2, y_2, x_4, y_4 определяются по формулам

$$x_2 = cx + (y_1 - cy), \quad y_2 = cy - (x_1 - cx),$$

$$x_4 = cx - (y_1 - cy), \quad y_4 = cy + (x_1 - cx).$$

Направление вертикали квадрата:

$$vx = \frac{x_1 + x_2}{2} - cx, \quad vy = \frac{y_1 + y_2}{2} - cy.$$

Под направлением вертикали в данном случае понимается отрезок, проведенный от центра квадрата к середине одной из сторон. После расчета производится его корректировка последовательным поворотом 90 градусов до тех пор, пока не выполнится условие

$$|vx| \leq |vy|, \quad vy < 0.$$

Данная корректировка обеспечивает независимость направления вертикали квадрата от порядка определения меток на изображении.

Далее определяется масштаб квадрата – длина отрезка направления, и производится расчет нормированного направления, совпадающего с направлением вертикали квадрата, но имеющего единичную длину:

$$m = \sqrt{vx^2 + vy^2},$$

$$nx = \frac{vx}{m}, \quad ny = \frac{vy}{m}.$$

Полученные на выходе параметры квадрата:

- координаты углов квадрата $x_1, y_1 \dots x_4, y_4$;
- координаты центра квадрата cx, cy ;
- нормированное направление вертикали квадрата nx, ny ;
- половина размера стороны квадрата m .

Аналогичный расчет проводится для второго найденного квадрата.

Данные о вершинах квадратов игнорируются, а для определения базовых параметров квадрата для дальнейших вычислений берется среднее арифметическое между параметрами cx, cy, nx, ny, m обоих квадратов. При этом производится предварительное сравнение параметров квадратов – расхождение больше определенного говорит о неверном распознавании картинки и невозможности вычислить результат.

Расположение меток определяется по данным положения центра квадрата cx, cy , направления квадрата nx, ny и масштаба квадрата m . Метки считываются построчно. После считывания меток получается три числа – Mx, My, CS . Первые два числа – номер квадрата на маркированной плоскости, третье число – контрольная сумма.

Если контрольная сумма не совпадает со значением, полученным при расчете по формуле (2), то одно из чисел считано неправильно и дальнейший расчет координат невозможен.

Далее вычисляется расположение центра текущего квадрата относительно центра крайнего верхнего левого квадрата маркированной плоскости:

$$X_0 = Hc \cdot Mx,$$

$$Y_0 = Hc \cdot My.$$

Затем определяется нормированное смещение центра изображения относительно центра квадрата:

$$Y_{otn} = \frac{nx \cdot \left(\frac{Width}{2} - cx \right) + ny \cdot \left(\frac{Height}{2} - cy \right)}{m \cdot (ny^2 + nx^2)},$$

$$X_{otn} = \frac{\left(\frac{Height}{2} - cy \right) - ny \cdot Y_{otn}}{nx},$$

где $Width, Height$ – ширина и высота картинки с камеры; Y_{otn}, X_{otn} – смещение центра экрана относительно центра квадрата, нормированное таким образом, что координаты углов квадрата заданы точками $(-1, 1); (1, 1); (1, -1); (-1, -1)$.

И наконец, определяется координата центра изображения на маркированной плоскости:

$$X = X_0 + \frac{Hc \cdot (X_{otn} + 1)}{2},$$

$$Y = Y_0 + \frac{Hc \cdot (1 - Y_{otn})}{2}.$$

Результаты расчета приведены на рис. 6.

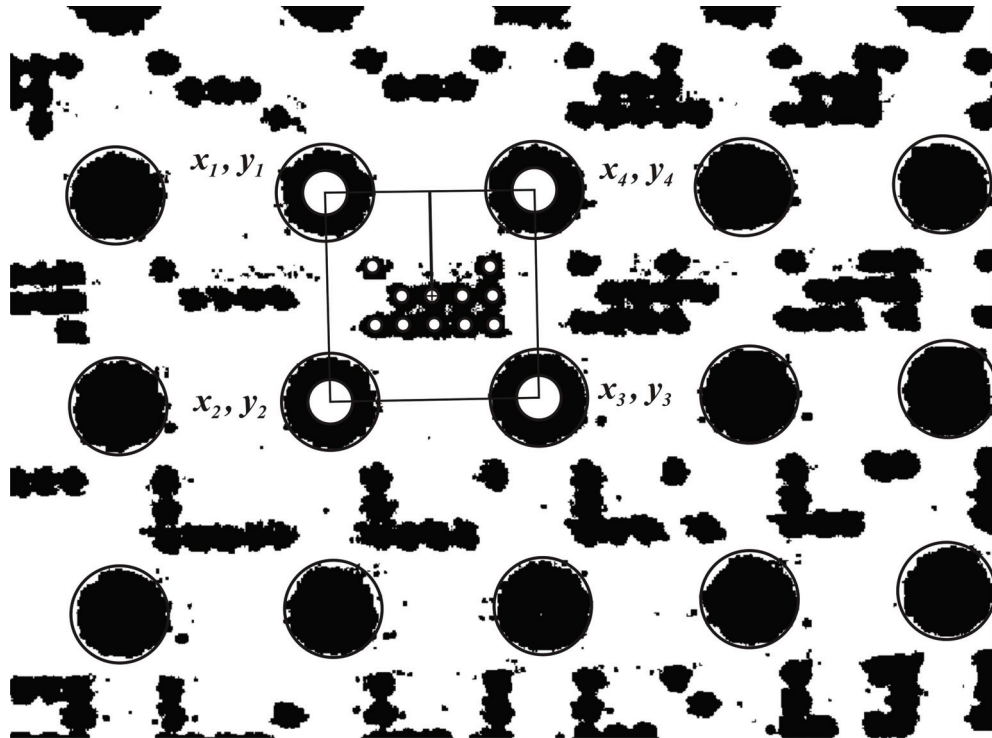


Рис. 6. Результат работы алгоритма

6. Погрешности метода

Строгий аналитический анализ погрешностей не проводился. Можно заметить, что основных источников погрешности два – неточность изготовления маркированной плоскости и неточность определения координат базовых меток на рисунке. Погрешность остальных влияющих факторов несущественна.

Неточность изготовления маркированной плоскости сильно зависит от технологии изготовления, однако можно с уверенностью сказать, что эта задача может быть решена с достаточно высокой точностью, как это делается, например, при изготовлении фотошаблонов для производства микросхем.

Неточность определения координат базовых меток сильно зависит от контраста изображения, линейности камеры и других факторов.

Оценка погрешности, вносимой камерой, проводилась с использованием изображения, аналогичного приведенному на рис. 4. Как видно из данного рисунка, используемая камера не слишком хороша – изображение размыто по правому краю, кроме того, на глаз заметна «выпуклость» изображения. Проводился анализ точности расчета координат всех базовых меток по строкам и столбцам. Расхождение среди всех точек, кроме крайних левого и правого рядов составило не более 1,5 пиксела, крайние ряды расходятся на 3 пиксела. Таким образом, можно посоветовать использовать меньшее расстояние между базовыми метками для уменьшения влияния искажений камеры.

Дальнейшая обработка включает в себя перевод изображения в монохромное и закрашивание области. При переводе могут быть потеряны краевые точки базовой метки, не полностью совпавшие с пикселями камеры.

И наоборот, к метке может добавиться случайный мусор в несколько пикселей размером. Для уменьшения влияния данных факторов рекомендуется делать размер базовых меток максимальным, поскольку при закраске берется среднее арифметическое всех координат входящих точек. Ряд экспериментов с различными яркостями подсветки картинки показал, что разница между определяемыми центрами координат метки не превышает 1 пикселя, даже при менее контрастном изображении, чем приведенное на рис. 4.

Общая погрешность определения координат на рисунке оценивается в 2 пикселя. Это заставляет пересмотреть выражение (1) и выбрать меньший размер P_{st} видимой области для достижения требуемой точности.

Заключение

Предложенный способ проведения измерений относительно маркированной плоскости с использованием оптических методов, который был положен в основу разработанного двухкоординатного датчика положения, позволил с достаточно высокой точностью проводить измерения положения детали (абсолютных координат). Датчик был апробирован в балансировочном устройстве для определения значений и углов статического дисбаланса абразивных шлифовальных кругов, работающем на принципе измерения отклонения маятника от вертикали, возникающего из-за наличия дисбаланса на роторе, закрепленном на маятнике. Опытный образец балансировочного устройства изготовлен и испытан в ООО «Измеритель» г. Пенза. При испытаниях датчик показал следующие результаты:

- измерения координат на плоскости;
- рабочий диапазон не менее ± 24 мм по каждой координате;
- линейность на всем рабочем диапазоне;
- погрешность по каждой координате не более 15 мкм.

Список литературы

1. **Туричин, А. М.** Электрические измерения неэлектрических величин / А. М. Туричин, П. В. Новицкий, Е. С. Левшина и др. ; под ред. П. В. Новицкого. – Л. : Энергия, 1975. – 559 с.
2. US Patent. Publication No. US 4022969 published on 10-May-1977, Application № US 5/616000 filed on 23-Sep-1975.
3. Измерения электрических и неэлектрических величин : учебное пособие для вузов / Н. Н. Евтихеев, Я. А. Купершмидт, В. Ф. Папуловский и др. ; под ред. Н. Н. Евтихеева. – М. : Энергоатомиздат, 1990. – 353 с.
4. **Агейкин, Д. И.** Датчики контроля и регулирования / Д. И. Агейкин, Е. Н. Костина, Н. Н. Кузнецова. – М. : Машиностроение, 1965. – 928 с.
5. **Форсайт, Д. А.** Компьютерное зрение. Современный подход : пер. с англ. / Дэвид А. Форсайт, Жан Понс. – М. : Вильямс, 2004. – 928 с.

Брякин Алексей Леонидович

кандидат технических наук, начальник
бюро программирования, научно-
производственное предприятие «Сенсор»
(г. Заречный)

E-mail: alexalb@crowd.ru

Bryakin Aleksey Leonidovich

Candidate of engineering sciences, head
of programming department, research
and production enterprise "Sensor"
(Zarechny)

Кочкин Сергей Вячеславович

кандидат технических наук, начальник
инновационно-аналитического отдела,
Пензенский государственный
университет

E-mail: ksvp@rambler.ru

Kochkin Sergey Vyacheslavovich

Candidate of engineering sciences, head
of sub-department of innovation analytics,
Penza State University

Кожевников Вячеслав Владимирович

кандидат технических наук, ассистент,
кафедра металлообрабатывающих
станков и комплексов, Пензенский
государственный университет

E-mail: redans@yandex.ru

Kozhevnikov Vyacheslav Vladimirovich

Candidate of engineering sciences, assistant,
sub-department of metalworking machine
tools and complexes, Penza State University

УДК 631.7, 53.082.14

Брякин, А. Л.

Двухкоординатный датчик положения / А. Л. Брякин, В. В. Кожевников, С. В. Кочкин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2011. – № 4 (20). – С. 225–236.

Вниманию авторов!

Редакция журнала «Известия высших учебных заведений. Поволжский регион. Технические науки» приглашает специалистов опубликовать на его страницах оригинальные статьи, содержащие новые научные результаты в области информатики, вычислительной техники, управления, электроники, измерительной техники, радиотехники, машиностроения, машиноведения, а также обзорные статьи по тематике журнала.

Статьи, ранее опубликованные, а также принятые к опубликованию в других журналах, редколлегией не рассматриваются.

Редакция принимает к рассмотрению статьи, подготовленные с использованием текстового редактора Microsoft Word for Windows версий не выше 2003.

Необходимо представить статью в электронном виде (VolgaVuz@mail.ru, дискета 3,5", CD-диск) и дополнительно на бумажном носителе в двух экземплярах.

Оптимальный объем рукописи 10–14 страниц формата А4. Основной шрифт статьи – Times New Roman, 14 pt через полуторный интервал. Тип файла в электронном виде – RTF.

Статья **обязательно** должна сопровождаться индексом УДК, краткой аннотацией и ключевыми словами **на русском и английском языках**.

Рисунки и таблицы должны быть размещены в тексте статьи и представлены в виде отдельных файлов (растровые рисунки в формате TIFF, BMP с разрешением 300 dpi, векторные рисунки в формате Corel Draw с минимальной толщиной линии 0,75 pt). Рисунки должны сопровождаться подрисуночными подписями.

Формулы в тексте статьи выполняются в редакторе формул Microsoft Word Equation, версия 3.0 и ниже. Символы греческого и русского алфавита должны быть набраны прямо, нежирно; латинского – курсивом, нежирно; обозначения векторов и матриц прямо, жирно; цифры – прямо, нежирно. Наименования химических элементов набираются прямо, нежирно. Эти же требования **необходимо** соблюдать и в рисунках. Допускается вставка в текст специальных символов (с использованием шрифтов Symbol).

В списке литературы **нумерация источников** должна соответствовать **очередности ссылок** на них в тексте ([1], [2], ...). Номер источника указывается в квадратных скобках. В списке указывается:

- для книг – фамилия и инициалы автора, название, город, издательство, год издания, том, количество страниц;
- для журнальных статей, сборников трудов – фамилия и инициалы автора, название статьи, полное название журнала или сборника, серия, год, том, номер, выпуск, страницы;
- для материалов конференций – фамилия и инициалы автора, название статьи, название конференции, время и место проведения конференции, город, издательство, год, страницы.

В конце статьи допускается указание наименования программы, в рамках которой выполнена работа, или наименование фонда поддержки.

К материалам статьи **должна** прилагаться информация для заполнения учетного листа автора: фамилия, имя, отчество, место работы и должность, ученая степень, ученое звание, адрес, контактные телефоны (желательно сотовые), e-mail.

Плата с аспирантов за публикацию рукописей не взимается.

Рукопись, полученная редакцией, не возвращается.

Редакция оставляет за собой право проводить редакторскую и допечатную правку текстов статей, не изменяющую их основного смысла, без согласования с автором.

Статьи, оформленные без соблюдения приведенных выше требований, к рассмотрению не принимаются.

Уважаемые читатели!

Для гарантированного и своевременного получения журнала «**Известия высших учебных заведений. Поволжский регион. Технические науки**» рекомендуем вам оформить подписку.

Журнал выходит 4 раза в год по тематике:

- ***информатика***
- ***вычислительная техника***
- ***управление***
- ***электроника***
- ***измерительная техника***
- ***радиотехника***
- ***машиностроение***
- ***машиноведение***

Стоимость одного номера журнала – 500 руб. 00 коп.

Для оформления подписки через редакцию необходимо заполнить и отправить заявку в редакцию журнала: факс (841-2) 36-84-87, тел.: 36-84-87, 56-47-33; E-mail: VolgaVuz@mail.ru

Подписку на второе полугодие 2012 г. можно также оформить по каталогу агентства «РОСПЕЧАТЬ» «Газеты. Журналы», тематический раздел «Известия высших учебных заведений». Подписной индекс – 36966

ЗАЯВКА

Прошу оформить подписку на журнал «Известия высших учебных заведений. Поволжский регион. Технические науки» на 2012 г.

№ 1 – _____ шт., № 2 – _____ шт., № 3 – _____ шт., № 4 – _____ шт.

Наименование организации (полное) _____

ИНН _____ КПП _____

Почтовый индекс _____

Республика, край, область _____

Город (населенный пункт) _____

Улица _____ Дом _____

Корпус _____ Офис _____

ФИО ответственного _____

Должность _____

Тел. _____ Факс _____ E-mail _____

Руководитель предприятия _____

(подпись)

(ФИО)

Дата «____» _____ 2012 г.