

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО ОБРАЗОВАНИЮ

**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ**

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК

Выпуск 29

**I СЕССИЯ НАУЧНОЙ ШКОЛЫ
«Информационная безопасность,
проектирование, технология элементов и
узлов компьютерных систем»**



**САНКТ-ПЕТЕРБУРГ
2006**

Выпуск содержит материалы **I сессии научной школы «Информационная безопасность, проектирование, технология элементов и узлов компьютерных систем»**, проведенной в рамках **III межвузовской конференции молодых ученых**.

Конференция была организована 10–13 апреля 2006 года Санкт-Петербургским государственным университетом информационных технологий, механики и оптики в сотрудничестве с

- Санкт-Петербургским государственным университетом
- Санкт-Петербургским государственным инженерно-экономическим университетом
- Казанским государственным университетом
- Санкт-Петербургским государственным технологическим институтом (техническим университетом)
- Санкт-Петербургским государственным политехническим университетом
- Санкт-Петербургским государственным университетом кино и телевидения
- Санкт-Петербургским государственным морским техническим университетом
- Самарским государственным архитектурно-строительным университетом
- Российским государственным гуманитарным университетом
- Мурманским государственным техническим университетом
- Ленинградским государственным университетом им. Пушкина
- ФГУП «НИТИОМ ВНИЦ «ГОИ им. С.И. Вавилова»
- ОАО «Российским институтом радионавигации и времени»
- Российской академией художеств.

В выпуске представлены работы, поддержанные финансированием в рамках:

- Федеральной целевой научно-технической программы «Исследования и разработки по приоритетным направлениям развития науки и техники на 2002–2006 гг.» (Федеральное агентство по науке и инновациям);
- аналитической ведомственной целевой программы «Развитие научного потенциала высшей школы (2006–2008 гг.)» (Федеральное агентство по образованию);
- Федеральной целевой программы развития образования на 2006–2010 гг. (Федеральное агентство по образованию);
- Российского фонда фундаментальных исследований,

а также инициативные разработки.

ПРОГРАММНЫЙ КОМИТЕТ КОНФЕРЕНЦИИ

Председатель – ректор СПбГУ ИТМО, д.т.н., профессор **В.Н. Васильев**

Сопредседатели – проректор по развитию, д.т.н., профессор **В.О. Никифоров**,
проректор по УО и АР, д.ф.-м.н., профессор **Ю.Л. Колесников**,
проректор по УМР, к.т.н., профессор **А.А. Шехонин**,
декан факультета ППО, д.т.н., профессор **В.Л. Ткалич**

Члены программного комитета – д.т.н., профессор **Ю.А. Гатчин**, д.т.н., профессор **В.М. Мусалимов**, д.т.н., профессор **С.Б. Смирнов**, д.т.н., профессор **В.А. Тарлыков**, д.т.н., профессор **Е.Б. Яковлев**, к.т.н. **Т.В. Точилина**

ОРГАНИЗАЦИОННЫЙ КОМИТЕТ КОНФЕРЕНЦИИ

Председатель – начальник НИЧ **Л.М. Студеникин**

Зам. председателя – к.т.н. **Т.В. Точилина**

Члены организационного комитета – **П.А. Борисов**, **Н.Н. Валентик**, **И.Н. Жданов**, **С.Ю. Керпелева**, **Н.В. Когай**, **А.В. Козаченко**, **И.М. Кудрявцева**, **Д.В. Лукичѳв**, **А.А. Малинин**, **Л.В. Можжухина**, **Ю.С. Монахов**, **Н.Б. Нечаева**, **М.В. Никитина**, **М.С. Петрищев**, **С.С. Резников**, **В.Н. Фролков**

ISSN 1819-222X

© Санкт-Петербургский государственный университет
информационных технологий, механики и оптики, 2006

**МЕТОДЫ И АЛГОРИТМЫ СОЗДАНИЯ СИСТЕМЫ ВЕРХНЕГО
(БЛОЧНОГО) УРОВНЯ АСУ ТП АЭС****А.Е. Пантелеев****Научный руководитель – доктор технических наук, профессор Ю.А. Гатчин**

Рассматриваются вопросы проектирования автоматизированной системы верхнего (блочного) уровня АСУ ТП АЭС (СВБУ). Приведены основные решения по созданию СВБУ в области структуры, комплекса технических средств, алгоритма функционирования и программного обеспечения.

Введение

На современном этапе развития промышленности невозможно обойтись без широкого применения автоматизированных систем управления технологическими процессами (АСУ ТП). Важные с точки зрения безопасности объекты атомной промышленности не являются исключением. АСУ ТП, применяемые в атомной энергетике, состоят из двух уровней: нижнего и верхнего. Нижний уровень – это уровень датчиков и контроллеров. Верхний уровень – это уровень серверов и рабочих станций.

Система верхнего (блочного) уровня АСУ ТП АЭС (СВБУ) – это система автоматического сбора, хранения, представления информации о текущем состоянии технологического объекта управления (ТОУ) и автоматизированного дистанционного формирования команд управления механизмами ТОУ алгоритмами АСУ ТП.

Применение методов искусственного интеллекта повышает надежность работы операторов по контролю состояния оборудования и диагностике технических процессов. Применение в системе «оператор» системы автоматизированного проектирования на всех уровнях разработки прикладных программ обеспечивает минимальный срок разработки готовых систем и высокую надежность программ.

Важной функцией, отличающей СВБУ от других известных отечественных систем верхнего уровня АСУ ТП АЭС, является дистанционное управление различным оборудованием энергоблока с экрана дисплея рабочей станции с помощью курсора вместо управления через ключи на панелях, что удешевляет технические средства и увеличивает скорость ввода управляющих воздействий. Применение компьютерной мультипликации для отображения технологической информации, технологических инструкций повышает наглядность представления информации, удобство работы оператора и пр. [4, стр. 1].

Основная часть

Общий алгоритм функционирования СВБУ представляет собой сумму алгоритмов своих подсистем. При этом особенностью взаимодействия элементов подсистем является применение технологии «клиент-сервер», в результате чего алгоритм функционирования каждой подсистемы разбивается на совокупность алгоритмов функционирования шлюзов, серверов и РС, решающих соответствующие им задачи внутри себя и обменивающихся между собой сетевыми сообщениями [3, стр. 2–3].

Шлюзы, являясь частью ПТК смежных систем АСУ ТП, включают интерфейсное ПО (ИПО) и решают следующие задачи:

- получение от низовых систем АСУ ТП дискретных сигналов с приписанными им метками времени и формирования признака обновления сигнала;

- получение от низовых систем АСУ ТП аналоговых сигналов, сравнение с апертурами, хранящимися в базе данных шлюза, приписывание метки времени и формирование признака обновления сигнала;
- получение от низовых систем АСУ ТП диагностических сигналов;
- подготовка и буферизация сообщений серверам;
- получение от серверов и передача низовым системам АСУ ТП команд дистанционного управления;
- синхронизация собственных часов по протоколу NTP и передача значения времени низовым системам АСУ ТП.

Общий алгоритм функционирования шлюзов (упрощенный) состоит из серии шагов, выполняемых циклически:

- опрос смежных систем АСУ ТП с целью получения от них информации перечисленных выше типов и запись ее в буфера обмена;
- прием от серверов запросов на получение информации и команд дистанционного управления;
- передача информации серверам, от которых получены запросы;
- передача команд дистанционного управления в ПТК смежных систем АСУ ТП;
- прием от АТПС синхронизирующих время сообщений, коррекция собственных часов и передача синхронизирующих время сообщений в ПТК смежных систем.

Данный алгоритм начинает выполняться сразу после запуска и инициализации ИПО шлюзов. Он является самодостаточным и не требует запуска никаких других элементов ПО СВБУ. Алгоритм предусматривает возможность функционирования как при наличии запущенных серверов, так и автономно. При этом предусматривается возможность работы шлюзов с серверами различных систем СВБУ одновременно.

Шлюзы отвечают на запросы тех серверов, от которых поступила команда на инициализацию обменов. Данная команда передает в шлюз список параметров, необходимых для работы конкретного сервера.

Серверы подсистем СВБУ выполняют следующий однотипный набор задач:

- прием и обработка аналоговых и дискретных сигналов от шлюзовых процессов и их занесение в локальную базу данных сервера;
- ведение локальной базы данных;
- хранение временных трендов поведения аналоговых параметров;
- ведение архивов по своим технологическим подсистемам и задачам;
- подготовка и передача рабочим станциям требуемой информации;
- прием и обработка серверами АТПС диагностической информации, получаемой от элементов СВБУ и от шлюзовых процессов, связанной с работой внешних к СВБУ систем, к которым данные шлюзы подключены;
- решение задач диагностики технологических процессов в объеме обслуживаемого сервером набора технологических подсистем;
- хранение инструкций, технологической и др. документации в машинно-ориентированном виде;
- решение общесистемных задач СВБУ, возложенных на данный сервер (например, расчет технико-экономических показателей в сервере ИУРО);
- прием команд дистанционного управления от рабочих станций и соответствующих этим командам сигналов от шлюзовых процессов.

Общий алгоритм функционирования серверов (упрощенный) состоит в серии шагов, выполняемых циклически:

- выдача запросов на получение информации шлюзам;
- прием информации от шлюзов и ее обработка;

- прием от РС запросов на получение информации и команд дистанционного управления;
- передача информации РС, от которых получены запросы;
- передача команд дистанционного управления в шлюзы;
- поддержание единого времени.

Данный алгоритм начинает выполняться сразу после запуска и инициализации ППО серверов. Он не требует запуска никаких других серверов и РС, но требует, чтобы были запущены шлюзы смежных систем АСУ ТП соответствующей подсистемы СВБУ. Алгоритм предусматривает функционирование как при наличии запущенных РС, так без них. При этом имеется возможность работы серверов с РС различных систем СВБУ одновременно.

Основные и резервные серверы функционируют по единому алгоритму независимо друг от друга. Оба они принимают информацию от одних и тех же шлюзов в одинаковом объеме и темпе. Отличие основного от резервного сервера состоит в том, что к первому подключаются РС, а второй работает автономно на прием, обработку и накопление информации.

РС поддерживают функции человеко-машинного интерфейса на соответствующих постах управления. Общий алгоритм функционирования РС состоит в серии шагов, выполняемых циклически:

- выдача запроса на получение информации серверу (основному либо резервному);
- прием информации от сервера, ее обработка и отображение на экране дисплеев;
- прием команд дистанционного управления, введенных оператором;
- передача команд дистанционного управления серверу;
- прием от АТПС синхронизирующих сообщений, коррекция собственных часов.

Данный алгоритм начинает выполняться сразу после запуска и инициализации ППО РС. Он требует, чтобы предварительно были запущены серверы соответствующих подсистем СВБУ.

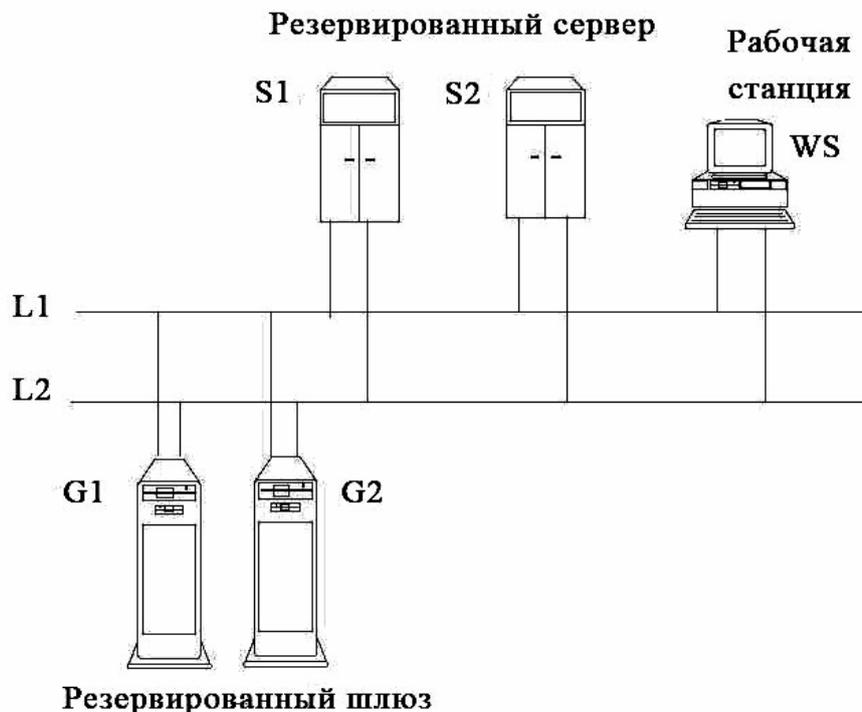


Рис. 1. Типовой фрагмент сети СВБУ

При взаимодействии абонентов сети СВБУ использован следующий основной принцип: все посылки в адрес абонента осуществляются только по его запросу/разрешению; при этом на каждый запрос осуществляется только один акт передачи информации. Такой подход устраняет возможность стохастического накопления посылок в буферах системы и их переполнения, исключает клинчевые ситуации в сети [1, стр. 2–4].

На рис. 1 приведен типовой фрагмент сети СВБУ, включающий резервированный шлюз G1/G2, резервированный сервер S1/S2 и одну рабочую станцию WS. Каждый из перечисленных элементов является одновременно абонентом сетей L1 и L2.

На рис. 2 представлена техническая структура фрагмента сети, приведенного на рис. 1. В этой структуре появляются дополнительные элементы – концентраторы Н1 и Н2 в сети L1 и Н2 в сети L2, обеспечивающие электрооптическое соединение элементов сети.

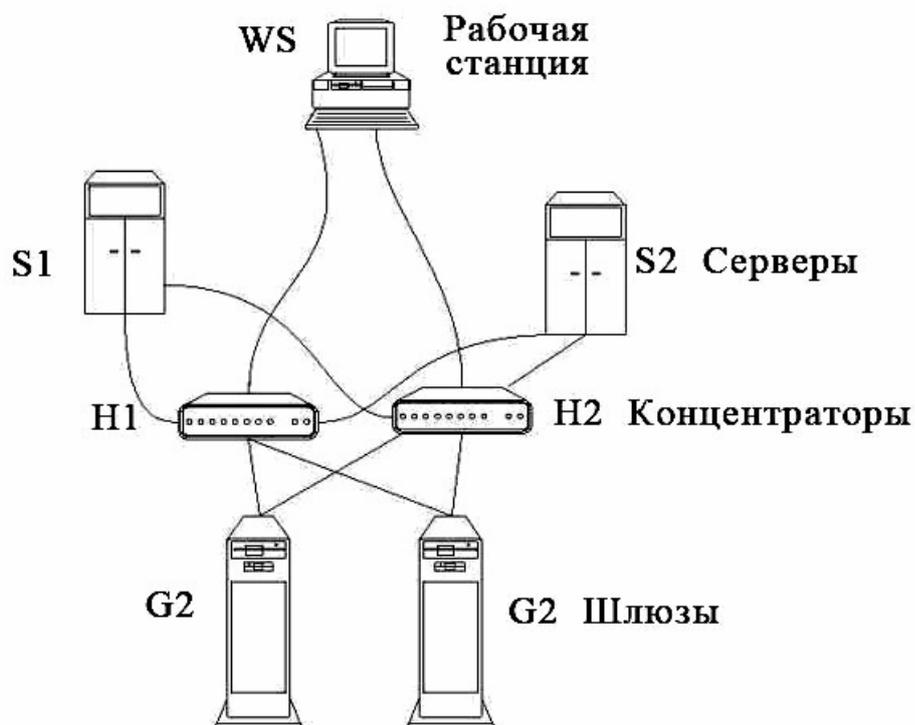


Рис. 2. Базовый элемент СВБУ (технологическая структура)

Принятыми техническими и алгоритмическими решениями обеспечивается принцип раздельного резервирования элементов сети, что позволяет любому из серверов S1 и S2 взаимодействовать с любым из шлюзов G1 и G2 через любой из концентраторов Н1 и Н2. Аналогичным образом с любым из абонентов сети может взаимодействовать рабочая станция WS. Реализация принципа раздельного резервирования позволяет достигать более высоких значений показателей надежности, чем при канальном общем резервировании, и обеспечивает большую отказоустойчивость при отказах элементов сети. В частности, при раздельном резервировании отказы и/или сбои неодновременных элементов не приведут к отказу сети, что не обеспечивается общим резервированием каналов. При необходимости в отдельных фрагментах сети алгоритмическими средствами можно обеспечить реализацию принципа общего резервирования.

При включении абонента в сеть СВБУ он устанавливает каналы связи со всеми уже действующими абонентами сети, с которыми ему предписано обмениваться информацией. При этом с каждым из абонентов устанавливаются два канала, оба из которых действуют на прием и передачу данных, но один из них реализован в сети L1, а другой – L2. В даль-

нейшем эти каналы называются логическими. Каждый абонент сети через фиксированный интервал времени проверяет наличие логических каналов связи со смежными абонентами.

Посылка сообщений абонентом 1 абоненту 2 осуществляется по следующему алгоритму. Один из двух логических каналов является основным, второй – резервным. При исправности основного логического канала посылка осуществляется по нему. При этом надежность передачи обеспечивается протоколом ТСР/ІР. При неисправности основного канала посылка осуществляется по резервному. После этого резервный канал назначается основным, но абоненты с заданной периодичностью пытаются установить связь по неисправному логическому каналу. Если связь по этому каналу удается восстановить, то данный канал опять становится основным.

Переход сервера S1 на взаимодействие с резервным для него шлюзом G2 происходит в следующих ситуациях:

- при отсутствии в сервере S1 обоих логических каналов связи со шлюзом G1;
- при получении сервером S1 диагностического сообщения от шлюза G1 о возникшем сбое;
- при формировании в сервере S1 диагностического сообщения о возникшем сбое в шлюзе G1;
- по команде администратора сети.

При этом основным логическим каналом между S1 и G2 назначается канал в сети L1. Шлюз G2 получает от сервера S1 разрешение на последующую передачу сообщений. Шлюз G2 высылает серверу S1 информацию из буфера, гарантирующего компенсацию возможного временного рассогласования в работе шлюзов G1 и G2. Данная посылка снабжается признаком возможности дублирования уже ранее полученной сервером S1 от шлюза G1 информации.

По аналогичной схеме осуществляется переход сервером S2 на работу с резервным для него шлюзом G1. Основным логическим каналом при этом назначается в сети L2.

Переход рабочей станцией WS на взаимодействие с резервным для нее сервером S2 происходит в следующих ситуациях:

- при отсутствии в рабочей станции WS обоих логических каналов связи с сервером S1;
- при получении рабочей станцией WS диагностического сообщения от сервера S1 о возникшем сбое;
- при формировании в рабочей станции WS диагностического сообщения о возникшем сбое в сервере S1;
- по команде администратора сети.

При этом основным логическим каналом между WS и S2 назначается канал в сети L1. Сервер S2 получает от рабочей станции WS разрешение на последующую передачу сообщений. Сервер S2 высылает рабочей станции WS информацию из буфера, гарантирующего компенсацию возможного временного рассогласования в работе серверов S1 и S2. Данная посылка снабжается признаком возможности дублирования уже ранее полученной рабочей станцией WS от сервера S1 информации [2, стр. 2–4].

Заключение

СВБУ, основные решения по реализации которой изложены в статье, – это первая отечественная интегрированная система верхнего уровня для атомных реакторов ВВЭР-1000. В настоящее время СВБУ внедряется в АСУ ТП зарубежных (Иран, Индия) и отечественных АЭС (Калининская АЭС, Балаковская АЭС).

Литература

1. Пантелеев А.Е., Проников Р.В. Динамическое моделирование с целью оценки временных характеристик выполнения задач управления. // «Информационные техно-

- логии в профессиональной деятельности и научной работе». Сборник материалов региональной научно-практической конференции. Йошкар-Ола, 2005.
2. Пантелеев А.Е., Проников Р.В. Аспекты проектирования пульта дистанционного управления подвижными объектами. // Там же.
 3. Пантелеев А.Е. Программный комплекс автоматизированной системы управления технологическим процессом перегрузки ядерного топлива на атомных электростанциях. // IX Международная научная конференция, посвященная 45-летию Сибирского государственного аэрокосмического университета имени академика М.Ф. Решетнева. Красноярск, 2005.
 4. Пантелеев А.Е. Создание программно-технических средств и комплексов ВБУ управления АСУ ТП АЭС нового поколения. // «Информационно-телекоммуникационные системы». Всероссийский конкурсный отбор инновационных проектов аспирантов и студентов.

БАЗА ДАННЫХ САПР ТП ГАЛЬВАНИЧЕСКОЙ ОБРАБОТКИ

Е.Г. Чернокнижная

Научный руководитель – доктор технических наук, профессор Ю.А. Гатчин

Работа посвящена разработке базы данных гальванического производства как основы САПР технологическими процессами гальванической обработки. Рассматриваются этапы проектирования, предлагается состав программного обеспечения, описываются структуры разработанных реляционных таблиц баз данных.

Введение

Системы автоматизированного проектирования технологических процессов (САПР ТП) играют главную роль в ускорении технологической подготовки производства и, как следствие, в сокращении сроков освоения новых изделий. Конечной целью САПР ТП является разработка комплекта технологической документации.

Повышение производительности труда технологов и нормировщиков и повышение качества проектных работ обеспечивается за счет автоматизации работ поискового и расчетного характера, работ по оформлению технических документов, ведения архива техпроцессов. Это должно решаться, в первую очередь, за счет рационального построения и ведения информационной базы данных (БД).

Основная часть

Современные САПР ТП – это сложные системы, состоящие из многочисленных взаимосвязанных звеньев с последовательными, параллельными и обратными связями. Поведение отдельных звеньев таких систем и функционирование системы в целом может иметь различный, в том числе детерминированно-стохастический, характер.

САПР ТП и ее компоненты являются сложными системами, и при их проектировании целесообразно использовать нисходящий стиль блочно-иерархического проектирования, включающего ряд уровней и этапов.

Проектирование базы данных состоит из двух этапов: логическое проектирование и физическое проектирование. На первом этапе определяется, какая информация, необходимая для процесса проектирования, должна храниться в БД; какие атрибуты данных для каждого объекта должны храниться в БД и как они должны быть идентифицированы; какие связи (отношения) должны существовать между объектами. На втором этапе происходит разработка базы данных с использованием конкретной СУБД на основании модели, полученной при логическом проектировании.

Поскольку разрабатываемая САПР ТП гальванического производства строится как открытая система, большое значение для нее имеет унификация и стандартизация средств межпрограммного интерфейса, т.е. профилей САПР ТП для информационного взаимодействия программ, входящих в САПР. Профилем открытой системы называют совокупность стандартов и нормативных документов, обеспечивающих выполнение системой заданных функций. В профилях САПР ТП могут использоваться язык EXPRESS стандарта STEP [1] и унифицированный язык SQL обмена данными между различными системами управления базами данных.

Разработка технологического процесса получения покрытия производится путем выбора конкретных вариантов операций и конкретных величин параметров обработки в соответствии с указаниями, приведенными в ГОСТ 9.305-84 «Покрyтия металлические и неметаллические неорганические. Операции технологических процессов получения покрытий». На основании анализа всех объектов, необходимых для автоматизированного проектирования технологических процессов, определено, что информационное обеспечение системы состоит из совокупности БД (рис. 1).

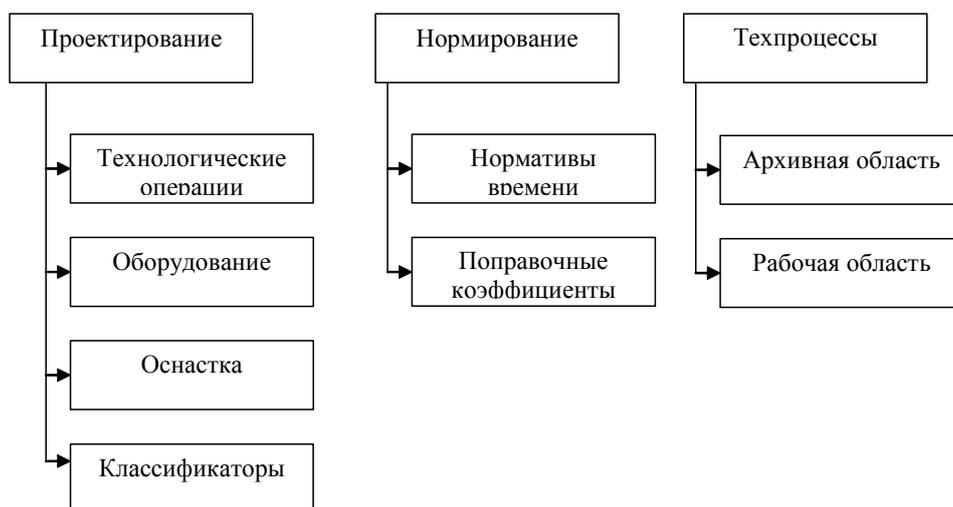


Рис. 1. Структура базы данных САПР ТП гальванообработки

БД «Проектирование» содержит сведения о производственной системе, используемые для реализации функции проектирования типовых технологических процессов (ТП) и единичных технологических процессов (ЕТП), расчета норм расхода материала на ТП. Логическая структура БД «Проектирование» построена в виде таблиц: «Технологические операции» (ГОСТ 9.305-84), «Оборудование», «Оснастка», «Классификаторы и справочники». БД «Нормирование» содержит данные по нормированию ЕТП: нормативы времени; дополнительные данные, которые идентифицируют нормативы времени по каждой операции ЕТП. БД «Техпроцессы» содержит техпроцессы (ТП, ЕТП) получения гальванических покрытий и должна быть использована для формирования и хранения технологических документов.

Логическое проектирование БД «Проектирование» начинается с системного анализа и содержательного описания предметной области и реальных связей, которые присутствуют между ее объектами. Каждая технологическая операция гальванического производства характеризуется следующими параметрами: состав электролита (тип операции), материал детали, плотность тока, время выдержки, температура электролита, дополнительные условия. Оборудование характеризуется инвентарным номером, кодом оборудования, наименованием. Аналогично определены свойства остальных информационных объектов: гальванической ванны, гальванического автомата, автооператора, выпрямительного агрегата и т.д.

В процессе проектирования технологического процесса система решает следующие задачи:

- выбирает (если возможно) подходящий технологический процесс из архива типовых ТП;
- создает новый ТП на основании требований, предъявляемых к детали в конструкторской документации;
- осуществляет адаптацию технологической схемы применительно к конкретным производственным условиям;
- рассчитывает оптимальную загрузку гальванических автоматов в зависимости от суточной производственной программы;
- рассчитывает площадь покрываемой поверхности деталей новой номенклатуры;
- рассчитывает токовые режимы в зависимости от типов деталей, ванн и приспособлений;
- формирует комплект необходимой сопроводительной документации.

Программное обеспечение системы включает в себя общесистемное программное обеспечение, состоящее из операционной системы WINDOWS, и специальное про-

граммное обеспечение, написанное на языке SQL. Основой системы управления базами данных SQL Server является реляционная модель. Реляционная база данных состоит из сущностей (таблиц), находящихся в определенном отношении друг с другом.

Сущности в данном случае – это технология гальванической обработки деталей, реализованная на конкретном предприятии. Отношения – это виртуальные связи, основанные на реальном взаимодействии персонала, оборудования, регламентов, направленные на реализацию технологий. Все данные представляются в виде двумерных таблиц, и все операции над данными сводятся к операциям над таблицами. Строка реляционной таблицы (запись, кортеж) представляет собой один экземпляр информационного объекта (сущности). Столбец (поле, атрибут) отражает какое-либо свойство этого объекта, т.е. совокупность значений для данного атрибута. Элементы данных кортежа связаны между собой отношениями. Таким образом, вся таблица называется отношением. Для однозначного определения каждого атрибута используется ключ, состоящий из одного или нескольких полей записи. Ключ должен быть уникальным и однозначно определять запись.

В соответствии с теорией реляционных баз данных [2, 3] таблицы должны отвечать следующим требованиям:

- в отношении нет одинаковых кортежей;
- каждый атрибут имеет уникальное имя;
- элементы данных атрибута однородны;
- все значения атрибутов атомарны (один элемент данных – одно значение);
- порядок следования строк и столбцов произвольный.

Реляционные базы данных позволяют объединить информацию, принадлежащую разным таблицам базы данных. Для этого в процессе проектирования определены связи между таблицами, в соответствии с которыми должно выполняться объединение таблиц. Инфологическая модель предметной области представляет собой описание базы данных и отражает предметную область в виде совокупности информационных объектов и их структурных связей. Основным инструментом разработки инфологической модели реляционной базы данных являются диаграммы «сущность–связь» (entity relationship diagrams), так называемые ER-диаграммы.

На рис. 2 показан основной фрагмент рассматриваемой базы данных в виде диаграммы.

Подготовка запросов к базе данных на сервере (на языке SQL) со стороны клиентской части может выполняться с помощью некоторой утилиты, например «Query Analyzer». Для предоставления пользователю больших возможностей и удобства в подготовке и выполнении запросов создаются клиентские приложения.

Для организации запросов к серверной базе данных непосредственно на языке SQL или с помощью клиентского приложения возможны различные способы взаимодействия, заметно влияющие на его эффективность. К числу основных способов такого взаимодействия можно отнести следующие способы, основанные на использовании:

- интерфейса DB-Library или DB-LIB (библиотек баз данных);
- технологии ODBC (совместимость открытых баз данных);
- интерфейса OLE DB (связывания и встраивания объектов баз данных);
- технология DAO (объект доступа к данным);
- технология ADO (объект данных ActiveX).

В нашем случае для организации запросов использована технология ODBC. Технология ODBC была разработана фирмой Microsoft для обеспечения возможности взаимодействия между различными СУБД. Она предусматривает создание дополнительного уровня между приложением и используемой СУБД. Службы ODBC обеспечивают получение от приложения запросов на выборку информации, перевод их на язык ядра адресуемой базы данных доступа к хранимой в ней информации. Основное назна-

чение ODBC состоит в абстрагировании приложения от особенностей ядра серверной базы данных, с которой оно осуществляет взаимодействие, поэтому серверная база данных становится как бы прозрачной для любого клиентского приложения.



Рис. 2. Модель базы данных «Проектирование» САПР ТП гальванического производства

Рассмотрим организацию обработки в базе данных на сервере SQL Server 2000 [4] с помощью технологии ODBC и использование Access в качестве клиентского приложения.

При использовании в клиентском приложении средств ODBC осуществляется обращение к определенному источнику данных, а через него – к СУБД, которую он представляет. При установке средств ODBC устанавливается общая подсистема ODBC и определяются пары «драйвер – база данных», которым задаются имена, используемые при соединении с базой данных. Соответствующие пары называются DSN (Data Source Names – имена источников данных)

При обработке данных на сервере с использованием технологии ODBC и применением клиентского приложения можно выделить следующие этапы: задание источника данных – создание и настройка соединений; собственно обработка данных с помощью запросов.

Создание источника данных выполняется с помощью утилиты «Администратор источников данных ODBC». С ее помощью создается новое соединение, и настраиваются параметры имеющихся соединений.

Access поддерживает следующие три типа источников данных ODBC:

- пользовательские, специфичные для каждого пользователя компьютера-клиента, можно создавать с помощью утилиты «Администратор ODBC» и с помощью Access;
- системные также можно создавать с помощью утилиты «Администратор ODBC» и с помощью Access;
- файловые являются индивидуальными текстовыми файлами, определяющими источник данных.

После задания базы данных SQL Server в качестве источника данных выполняется подключение таблиц базы данных SQL Server в базу данных Access.

Создание запроса к базе данных SQL Server включает следующие шаги:

- создание нового запроса и добавление в него требуемых таблиц из базы данных SQL Server;
- перетаскивания требуемых полей из добавленных таблиц в строку Field (поле) бланка запроса;
- нажатие на панели инструментов кнопки «Запуск» для ввода результатов выполнения запроса;
- закрытие и сохранения запроса.

Приведенная выше процедура подключения таблиц используется в большинстве случаев для баз данных SQL Server при наличии драйверов ODBC.

В SQL Server можно писать хранимые процедуры и выполнять их вместо отправки серверу отдельных команд. Выполнение запроса в виде хранимой процедуры осуществляется быстрее, чем с помощью ядра базы данных Jet. Это обусловлено уменьшением объема передаваемых данных и тем, что хранимые процедуры выполняются быстрее, поскольку предварительно компилируются. Кроме того, запрос к SQL Server применяется при необходимости использовать средства Transact-SQL, отсутствующие в SQL Access.

Заключение

Предложенный подход к созданию базы данных, а также выбор и использование СУБД, позволяют создать информационный банк данных, который является основой САПР ТП гальванического производства. Такая система может использоваться как самостоятельно в сетевой клиент-серверной среде вычислений, так и в интеграции с информационно-управляющими системами различных уровней для обеспечения непрерывного жизненного цикла промышленных изделий.

Литература

1. Норенков И.П. Основы автоматизированного проектирования. М: МГТУ им. Баумана, 2002. 336 с.
2. Базы данных. Под ред. А.Д. Хомоненко. СПб: КОРОНА-принт, 2003. 672 с.
3. Кренке Д. Теория и практика построения баз данных. СПб: Питер, 2005. 859 с.
4. Тихомиров Ю. SQL Server 2000: разработка приложений. БХВ-Петербург, 2000. 368 с.

МОДЕЛИ ПРЕДСТАВЛЕНИЯ ЗНАНИЙ ЭКСПЕРТНЫХ СИСТЕМ

М.В. Барышев, И.Ю. Гатчин

Научный руководитель – доктор технических наук, профессор Ю.А. Гатчин

В статье рассматриваются типичные модели представления знаний экспертных систем. При рассмотрении их положительных и отрицательных сторон, а также их возможного применения на практике, делается вывод, что в дальнейшем будет отдаваться предпочтение экспертным системам с комбинированным способом представления знаний.

Введение

В начале 80-х гг. в рамках искусственного интеллекта окончательно сформировалось новое направление – «инженерия знаний», целью которого является разработка и исследование экспертных систем. Отличительной особенностью экспертных систем по сравнению с ранее существовавшими системами искусственного интеллекта является то, что экспертные системы способны решать не модельные, а вполне реальные задачи, при этом получая результаты, сравнимые с теми, которые способен получить эксперт. Кроме того, экспертные системы применяются в неформализованных или слабоформализованных областях знаний, которые до этого были мало доступны средствам вычислительной техники. И, наконец, экспертные системы ориентируются на общение с пользователями, мало знакомыми с программированием, что позволяет резко расширить круг людей, способных решать свои прикладные задачи с помощью компьютера на достаточно высоком профессиональном уровне. Все это вызывает на протяжении последнего десятилетия повышенный интерес к проблематике экспертных систем как со стороны специалистов в области вычислительной техники и искусственного интеллекта, так и со стороны конечных пользователей.

Но, несмотря на огромное количество фактов, подталкивающих к разработкам новых экспертных систем, одной из наиболее важных проблем, характерных для систем, основанных на знаниях, является проблема представления знаний. Это объясняется тем, что форма представления знаний оказывает существенное влияние на характеристики и свойства системы. Чтобы манипулировать всевозможными знаниями из реального мира с помощью компьютера, необходимо осуществлять их моделирование. В таких случаях необходимо отличать знания, предназначенные для обработки компьютером, от знаний, используемых человеком. Кроме того, при большом объеме знаний желательно упростить последовательное управление отдельными элементами знаний.

Целью статьи является рассмотрение уже созданных моделей представления знаний, констатация их положительных и отрицательных сторон, и сравнительная оценка применимости той или иной модели знаний для разработок экспертных систем.

Модели представления знаний

При проектировании модели представления знаний следует учитывать такие факторы, как однородность представления и простота понимания. Однородное представление приводит к упрощению механизма управления логическим выводом и упрощению управления знаниями. Представление знаний должно быть понятным экспертам и пользователям системы. В противном случае затрудняются приобретение знаний и их оценка. Однако выполнить это требование в равной степени для простых и для сложных задач трудно. Обычно для несложных задач останавливаются на некотором среднем (компромиссном) представлении, но для решения сложных и больших задач необходимы структурирование и модульное представление.

Типичными моделями представления знаний являются:

- логическая модель;

- продукционная модель;
- фреймовая модель;
- семантическая модель.

Во всех разработанных в прошлом системах с базами знаний, помимо этих моделей, использовались специализированные средства, поэтому представление знаний получалось сложным. Тем не менее, классификация моделей оставалась неизменной. Язык, используемый для разработки систем, спроектированных на основе этих моделей, называется языком представления знаний [1].

Логическая модель представления знаний

Логическая модель используется для представления знаний в системе логики предикатов первого порядка и логики высказываний.

Формальная дедуктивная система, лежащая в основе логической модели представления знаний, может быть представлена в виде четверки: $\phi = \langle B, S, A, P \rangle$, где B – множество базовых элементов (алфавит); S – множество синтаксических правил, на основе которых из B строятся правильно построенные формулы; A – множество аксиом (правильно построенных формул); P – множество правил вывода (семантических правил), которые из множества аксиом позволят получать новые правильно построенные формулы – теоремы [2].

Основные преимущества использования логики предикатов для представления знаний заключаются в том, что обладающий хорошо понятными математическими свойствами мощный механизм вывода может быть непосредственно запрограммирован; имеется возможность контроля логической целостности базы знаний, т.е. ее непротиворечивости и полноты; весьма просто реализуется запись фактов, следствием чего является простота и компактность базы знаний. Главным недостатком логической модели является отсутствие четких принципов организации фактов в базе знаний. В связи с этим достаточно большие базы трудно поддаются анализу и обработке.

Таким образом, логические модели представления знаний целесообразно использовать в тех предметных областях, где база знаний невелика по объему и относительно проста по структуре.

Продукционная модель представления знаний

Продукционная модель наиболее адекватно соответствует процедурному характеру знаний. Эта модель описывает знания с помощью так называемых правил продукции, которые, по существу, являются программами, состоящими из одного оператора вида «ЕСЛИ <условие>, ТО<действие>».

Системы с базами знаний, основанные на этой модели, называются продукционными системами. Эти системы бывают двух диаметрально противоположных типов – с прямыми и обратными выводами.

В системе продукции с обратными выводами с помощью правил строится дерево И/ИЛИ, связывающее в единое целое факты и заключения; оценка этого дерева на основании фактов, имеющихся в базе данных, и есть логический вывод. Логические выводы бывают прямыми, обратными и двунаправленными. При прямом выводе отправной точкой служат предоставленные данные, процесс оценки приостанавливается в узлах с отрицанием, причем в качестве заключения (если не все дерево пройдено) используется гипотеза, соответствующая самому верхнему уровню дерева (корню). Однако для такого вывода характерно большое количество данных, а также оценок дерева, не имеющих прямого отношения к заключению, что излишне. Преимущество обратных выводов – в том, что оцениваются только те части дерева, которые имеют отношение к заключению,

однако, если отрицание или утверждение невозможны, то порождение дерева лишено смысла. В двунаправленных выводах сначала оценивается небольшой объем полученных данных и выбирается гипотеза (по примеру прямых выводов), а затем запрашиваются данные, необходимые для принятия решения о пригодности данной гипотезы. На основе этих выводов можно реализовать более мощную и гибкую систему.

Системы продукций с прямыми выводами среди систем, основанных на использовании знаний, имеют наиболее давнюю историю, поэтому они являются в некотором смысле основополагающими. Эти системы включают три компонента: базу правил, состоящую из набора продукций (правил вывода), базу данных, содержащую множество фактов, и интерпретатор для получения логического вывода на основании этих знаний. База правил и база данных образуют базу знаний, а интерпретатор соответствует механизму логического вывода. Вывод выполняется в виде цикла «понимание – выполнение», причем в каждом цикле выполняемая часть выбранного правила обновляет базу данных. В результате содержимое базы данных преобразуется от первоначального к целевому, т.е. целевая система синтезируется в базе данных. Иначе говоря, для системы продукций характерен простой цикл выбора и выполнения (или оценки) правил, однако из-за необходимости периодического сопоставления с образцом в базе правил (отождествлением) с увеличением числа последних (правил) существенно замедляется скорость вывода. Следовательно, такие системы не годятся для решения крупномасштабных задач. Упорядочим слабые и сильные стороны хорошо известных систем продукций.

Сильные стороны:

- простота создания и понимания, отдельных правил;
- простота модификаций и расширения базы знаний;
- простота механизма логического вывода;
- естественная модульность организации баз знаний.

Слабые стороны:

- неясность взаимных отношений правил;
- сложность оценки целостного образа знаний;
- крайне низкая эффективность обработки;
- отличие от человеческой структуры знаний;
- отсутствие гибкости в логическом выводе.

Фреймовая модель представления знаний

Фреймовый подход к представлению знаний является частным случаем моделей представления знаний с помощью семантических сетей. Здесь единицей представления является объект, называемый фреймом. Формально под фреймом понимают обычно следующую структуру: $f[\langle v_1, g_1 \rangle, \langle v_2, g_2 \rangle, \dots, \langle v_k, g_k \rangle]$, где f – имя фрейма; $\langle v_i, g_i \rangle$ – i -й слот; v_i – имя слота; g_i – его значение.

Таким образом, фрейм представляет собой некоторую законченную структуру, которая при заполнении слотов значениями описывает вполне конкретный факт, событие, явление или процесс.

Фреймы можно разделить на две группы: фреймы-описания; ролевые фреймы.

Фрейм-описание:

[<программное обеспечение>, <программа 1С бухгалтерия, версия 7.7>, <программа 1С торговля, версия 7.7>, <правовая программа «Консультант +» проф.>].

Ролевой фрейм:

[<заявка на продажу>, <что, установка и покупка программы 1С торговля, версия 7.7>, <откуда, фирма Рим>, <куда, фирма «Эл Ти Хелскеа»>, <кто, курьер Сидоров>, <когда, 18 февраля 2006г.>].

Во фрейме-описании в качестве имен слотов задан вид программного обеспечения, а значение слота характеризует массу и производителя конкретного вида продукции. В ролевом фрейме в качестве имен слотов выступают вопросительные слова, ответы на которые являются значениями слотов. Для данного примера представлены уже описания конкретных фреймов, которые могут называться либо фреймами-примерами, либо фреймами-экземплярами. Если в приведенном примере убрать значения слотов, оставив только имена, то получим так называемый фрейм-прототип.

Достоинство фрейма-представления во многом основывается на включении в него предположений и ожиданий. Это достигается за счет присвоения по умолчанию слотам фрейма стандартных ситуаций. В процессе поиска решений эти значения могут быть заменены более достоверными. Некоторые переменные выделены таким образом, что об их значениях система должна спросить пользователя. Часть переменных определяется посредством встроенных процедур, называемых внутренними. По мере присвоения переменным определенных значений осуществляется вызов других процедур. Этот тип представления комбинирует декларативные и процедурные знания.

Фреймовые модели обеспечивают требования структурированности и связанности. Это достигается за счет свойств наследования и вложенности, которыми обладают фреймы, т.е. в качестве слотов может выступать система имен слотов более низкого уровня, а также слоты могут быть использованы как вызовы каких-либо процедур для выполнения.

Для многих предметных областей фреймовые модели являются основным способом формализации знаний.

Семантическая модель представления знаний

Основная идея данного подхода к представлению знаний состоит в том, что предметная область рассматривается как совокупность объектов и связей между ними. Семантическая сеть, представляющая некоторую предметную область, есть граф, в котором множеству вершин сопоставимы объекты (понятия, события, процессы), а множеству дуг – отношения на множестве объектов. Вершины и дуги семантической сети помечаются именами соответствующих объектов и отношений предметной области.

Базовым функциональным элементом семантической сети служит структура из двух компонентов: дуги и инцидентных ей пары вершин. Можно считать, что каждый такой элемент представляет собой простой факт или высказывание предметной области. Например, высказывание: «Сидоров учится в школе» можно представить в виде элемента семантической сети, изображенного на рис.



Рис. Пример базового функционального элемента семантической сети

Как видно из рисунка, дуга в сети имеет направленность, благодаря чему между понятиями в рамках определенного факта выражается отношение «субъект/объект». Очевидно, что каждая вершина графа может быть соединена с любым числом других вершин. Таким образом, осуществляется формирование сети фактов.

Базовую структуру семантической сети, с точки зрения логики, можно рассматривать как эквивалент бинарного предиката. Тогда предикат соответствует направленной

дуге, а два его аргумента – двум инцидентным этой дуге вершинам. Пример на рисунке можно описать предикатом «учиться (Сидоров, школа)».

В семантических представлениях используются три основных типа объектов: понятия, события и свойства. Понятия обычно соответствуют предметам или абстракциям. Под событием понимают действия, приводящие к изменению ситуации. Свойства используются для уточнения или модификации понятий, событий или других свойств.

При представлении знаний семантическими сетями особый практический интерес имеют связи следующих видов: «является», «имеет», «есть». Связь вида «является» отражает принадлежность к некоторому классу объектов. Связь вида «имеет» и «есть» указывают, соответственно, на то, что одно понятие служит атрибутом другого. Используя эти отношения, можно достаточно просто представить сложные совокупности фактов [3].

Достоинствами моделей представления знаний с помощью семантических сетей являются большие выразительные возможности, наглядность, близость структуры сети семантической структуре предметной области.

Основным недостатком подобных моделей является слабое развитие теоретических основ логического вывода, в связи с чем усиливается элемент произвольности, вводимый разработчиком конкретной системы обработки знаний на основе семантической сети. Поэтому процедуры вывода в таких системах часто таят в себе угрозу противоречий, а, следовательно, процедурам анализа противоречивости в семантических сетях необходимо уделять больше внимания, чем, например, в продукционных системах. Очень часто эта функция в семантических моделях возлагается на человека. Таким образом, при достаточно большом объеме знаний решение этой задачи резко усложняется, что ограничивает круг предметных областей, описываемых семантическими сетями, сравнительно небольшими проблемами.

Заключение

При рассмотрении типичных моделей знаний были выявлены достоинства и недостатки каждой из перечисленных моделей представления. Этот факт объясняет, почему в последнее время разработчики интеллектуальных систем все шире используют комбинированные способы представления знаний, что позволяет компенсировать недостатки одного представления достоинствами другого. Например, такой недостаток продукционной системы, как отсутствие явного структурирования правил предметной области, можно устранить с помощью сети фреймов, обеспечивающих группирование правил по ситуациям. Отмечается, что использование комбинированных подходов предоставляет новые возможности для создания интеллектуальных систем, обладающих уникальными свойствами [4].

Литература

1. Убейко В.Н. Экспертные системы. М.: Май, 1992. 246 с.
2. Поспелова Д.А. Логика рассуждений и ее моделирование. М.: Науч. совет по комплекс. проб. «Кибернетика», 1983. 180 с.
3. Элти Дж., Кумбс М. Экспертные системы: концепции и примеры. М.: Финансы и статистика, 1987. 191 с.
4. Попов Э.В. Экспертные системы: Решение неформализованных задач в диалоге с ЭВМ. М.: Наука, 1987. 288 с.

АЛГОРИТМ ФУНКЦИОНИРОВАНИЯ СИСТЕМЫ АВТОМАТИЗИРОВАННОГО ПРОЕКТИРОВАНИЯ РЕКУРСИВНЫХ КОДОВЫХ ШКАЛ

В.А. Климанов

Научный руководитель – доктор технических наук, профессор А.Г. Коробейников

В работе рассматриваются различные варианты построения преобразователей линейных и угловых перемещений на основе рекурсивных кодовых шкал. Предложен алгоритм функционирования автоматизированной системы, предназначенной для проектирования шкал данного вида.

Введение

В информационно-измерительных и автоматизированных системах управления с использованием ЭВМ широко применяются устройства аналого-цифрового преобразования, одним из видов которых являются преобразователи перемещения в код. В [1] приведена классификация таких преобразователей. В последнее время в связи с интенсивным развитием интегральной технологии, наряду с традиционными требованиями по разрешающей способности, быстродействию и помехоустойчивости, существенное значение приобрели предъявляемые к устройствам кодирования требования по технологичности, массо-габаритным характеристикам и надежности.

В большей степени требованиям по разрешающей способности, быстродействию и помехоустойчивости удовлетворяют преобразователи с непосредственным преобразованием перемещения в код на основе считывания с использованием пространственного кодирования, основным элементом которых является кодовая шкала (КШ).

Классические шкалы обычно выполняются с использованием обыкновенного двоичного кода (ОДК), циклических кодов и на основе кодов специальной структуры. По структуре применяемого кода в [1] представлена классификация КШ. Известен ряд работ, посвященных вопросам синтеза кодовых шкал, трудоемкость изготовления которых в 2–8 раз, а габариты приблизительно в два раза меньше по сравнению со шкалами, имеющими кодовую маску, выполненную в обыкновенном двоичном или в традиционном коде Грея. Значительный вклад в решение данных вопросов внесли В.Г. Домрачев, Б.С. Мейко, В.В. Кривенков, Ю.С. Шарин и другие отечественные и зарубежные ученые.

В [2] рассмотрены рекурсивные кодовые шкалы (РКШ), которые позволяют создавать преобразователи перемещений с улучшенными массо-габаритными, технологическими и одновременно надежностными характеристиками.

Традиционные шкалы, выполненные в ОДК или в коде Грея, практически всегда имеют один вариант построения. В отличие от них РКШ, как показано в [2], могут быть представлены множеством вариантов выполнения кодовой маски, а также различным размещением считывающих элементов (СЭ). Поэтому представляется актуальной задача разработки системы, которая позволила бы автоматизировать некоторые этапы построения РКШ для преобразователей перемещения.

Преобразователи перемещений на основе РКШ

Исходя из многообразия вариантов построения РКШ, проведем классификацию преобразователей перемещения на их основе (рис. 1).

В качестве первого классификационного признака используется вид преобразования и конструктивного исполнения, в соответствии с которым преобразователи перемещений на основе РКШ делятся на круговые и линейные.

Далее по принципу их функционирования преобразователи могут быть абсолютными или квазиабсолютными. Абсолютные преобразователи – устройства, построенные по принципу непосредственного преобразования перемещения в код по методу считывания своего положения с кодовой шкалы. Квазиабсолютные преобразователи занимают промежуточное положение между названными и устройствами, построенными по методу последовательного счета.



Рис. 1. Преобразователи перемещений на основе РКШ

Среди круговых квазиабсолютных преобразователей можно выделить реверсивные и нереверсивные. В реверсивных преобразователях информация об угловом положении кодируемого вала последовательно снимается с информационной дорожки шкалы посредством блока информационных СЭ и в зависимости от направления перемещения шкалы поступает на соответствующие входы схемы обработки. Нереверсивные круговые преобразователи в силу своих конструктивных особенностей обеспечивают кодирование углового положения при вращении вала только в одном направлении.

Линейные квазиабсолютные преобразователи выполняют только в реверсивном исполнении.

Типы рекурсивных кодовых шкал

На рис. 2 приведена классификация рекурсивных кодовых шкал (РКШ), в основу построения кодовых дорожек которых положены двоичные последовательности. Классификация относится к шкалам преобразователей углового и линейного перемещения.

В качестве первого классификационного признака используется свойство линейности рекурсивных последовательностей по отношению к оператору суммирования по модулю два. По этому признаку РКШ делятся на две группы: кодовые шкалы на основе линейных рекурсивных последовательностей (РП) и шкалы на основе нелинейных РП.

В свою очередь, КШ на основе линейных РП подразделяются на три группы: КШ на основе псевдослучайных последовательностей максимальной длины (ПСПМД), КШ на основе последовательностей неполных циклов (ПНЦ) и КШ на основе композиционных последовательностей (КП).

При получении рекурсивного соотношения, используемого для генерации символов ПСПМД (М-последовательности), берутся только примитивные полиномы. Кодовые шкалы на основе ПСПМД называются псевдослучайными кодовыми шкалами (ПСКШ).

Последовательности неполных циклов (ПНЦ) строятся на основе неприводимых (над полем Галуа $GF(2)$) полиномов. Исследования показали, что такие КШ будут иметь значительно меньшую, по сравнению с ПСКШ, разрешающую способность.

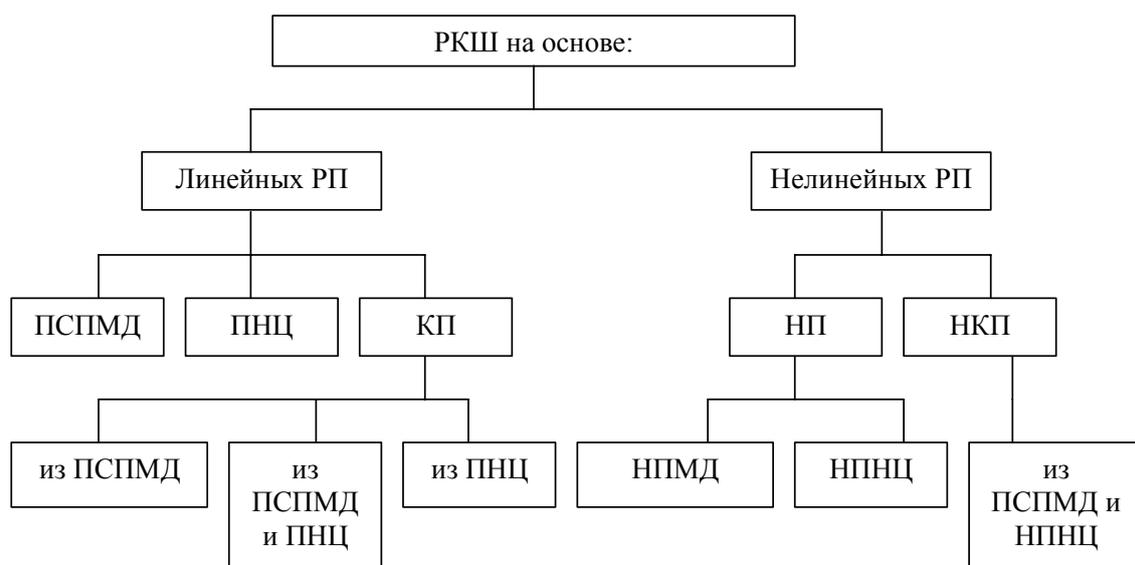


Рис. 2. Классификация рекурсивных кодовых шкал

Кодовые шкалы на основе композиционных последовательностей могут быть разделены на три группы: КШ на основе композиционных последовательностей, где КП получаются из различных сочетаний М-последовательностей; КШ на основе композиционных последовательностей, где КП получаются из сочетаний М-последовательностей и ПНЦ; КШ на основе композиционных последовательностей, где КП получаются из различных сочетаний ПНЦ.

Принципы построения КШ на основе КП, где последовательности получаются из различных сочетаний М-последовательностей, рассмотрены в [2]. Именно такие кодовые шкалы названы ККШ.

Принципы построения КШ второй и третьей групп на основе КП в данной работе не рассмотрены. Нетрудно видеть, что наименьшей разрешающей способностью обладают КШ на основе РП третьей группы, наибольшей – ККШ, а КШ второй группы занимают по этому параметру промежуточное положение. Кодовые шкалы на основе нелинейных РП подразделяются на две группы: КШ на основе нелинейных последовательностей (НП) и КШ на основе нелинейных КП (НКП).

В свою очередь, КШ на основе НП также делятся на две группы: КШ на основе НП максимальной длины (НПМД) и КШ на основе НП неполных циклов (НПНЦ).

Принципы построения КШ на основе НПМД рассмотрены в [2]. При получении рекурсивного соотношения, необходимого для генерации символов нелинейной последовательности максимальной длины, берутся примитивные полиномы с введением в рекурсивное соотношение слагаемого, обеспечивающего нелинейность синтезируемой последовательности. Кодовые шкалы на основе НПМД получили название нелинейных кодовых шкал (НКШ).

Принципы построения КШ на основе НПНЦ в данной работе не рассмотрены. Получение НПНЦ подробно рассмотрено в работе [3]. Последовательности неполных циклов, также как и НПМД, позволяют строить на своей основе КШ, однако их разрешающая способность будет заведомо меньше, чем разрешающая способность КШ на основе НПМД.

Кодовые шкалы на основе НКП могут быть разделены на две группы: КШ на основе НКП, где НКП получаются из сочетаний М-последовательностей и НПМД, и КШ на основе НКП, где НКП получаются из различных сочетаний линейных и нелинейных последовательностей как максимальных, так и неполных циклов.

Нетрудно видеть, что наилучшей разрешающей способностью будут обладать КШ на основе НКП, где НКП получаются из сочетаний М-последовательностей и НПМД, а КШ на основе НКП, где НКП получаются из других сочетаний линейных и нелинейных последовательностей, обладают, по сравнению с первыми, меньшей разрешающей способностью.

Автоматизированная система проектирования РКШ

Автоматизированная система проектирования РКШ должна обеспечивать решение следующих задач:

- синтез РКШ для преобразователей значительной (до 20 разрядов) разрешающей способности, связанный с трудоемкой процедурой получения кодовой последовательности длиной до 2^{20} ;
- корректность размещения считывающих элементов. РКШ выполняются с одной информационной кодовой дорожкой, вдоль которой размещаются информационные и корректирующие считывающие элементы;
- многообразие вариантов выбора РКШ, которые удовлетворяют требуемым параметрам преобразователя и являются оптимальными с точки зрения простоты его конструкции.

Таким образом, автоматизированная система проектирования РКШ должна обеспечивать:

- выбор типа преобразователя перемещений (рис. 1);
- выбор разрядности преобразователя;
- выбор типа рекурсивной кодовой шкалы (рис. 2);
- получение рисунка кодовой маски шкалы;
- размещение вдоль кодовой дорожки шкалы информационных считывающих элементов;
- размещение вдоль кодовой дорожки шкалы корректирующих считывающих элементов.

Алгоритм функционирования автоматизированной системы проектирования РКШ

Обобщенный алгоритм функционирования автоматизированной системы проектирования рекурсивных кодовых шкал представлен на рис. 3. Здесь цифрами 1–4 показаны основные этапы работы системы.

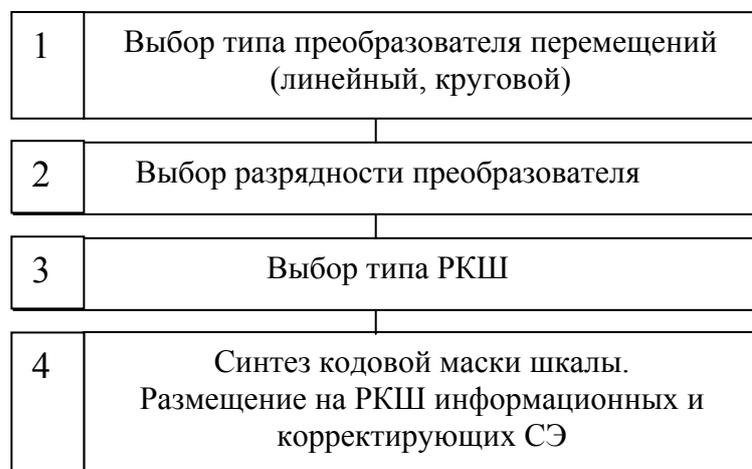


Рис. 3. Обобщенный алгоритм функционирования автоматизированной системы проектирования РКШ

Очевидно, что работа системы на этапах 1, 2 и 3 обеспечивается при активном участии разработчика, так как именно он определяет требования к проектированию РКШ соответствующего вида, выбирает тип преобразователя и вводит необходимые исходные данные.

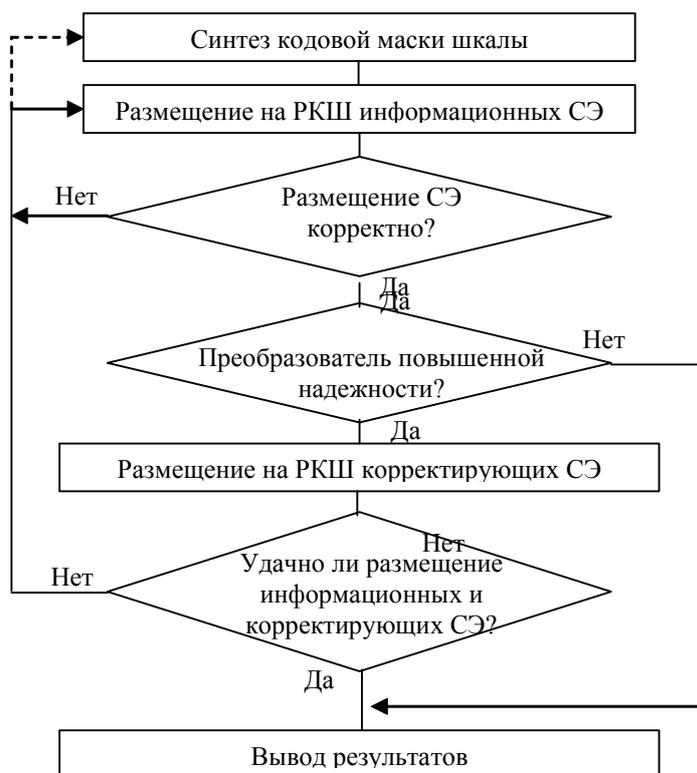


Рис. 4. Этап 4 обобщенного алгоритма функционирования автоматизированной системы проектирования РКШ

Рассмотрим более подробно порядок функционирования автоматизированной системы на этапе 4 (рис. 4).

1. Обеспечивается синтез РКШ.
2. Для преобразователей любых типов осуществляется расчет положения информационных СЭ.
3. Проверяется корректность расположения информационных СЭ. Их расположение будет корректным, если будут получены неповторяющиеся кодовые комбинации с РКШ на всем диапазоне перемещения шкалы при данном размещении информационных СЭ.
4. Если проектируется преобразователь повышенной надежности, то возможно формирование корректирующих кодов (КК). В соответствии с выбранным методом формирования КК (код Хемминга, циклические коды, коды BCH) осуществляется определение числа корректирующих СЭ и их размещение вдоль информационной дорожки РКШ.
5. Осуществляется проверка, является ли удачным совместное расположение информационных и корректирующих СЭ. Расположение считывающих элементов является удачным, если выполняются технологические требования к размещению СЭ на кодовой шкале.

Также следует учитывать, что наиболее оптимальным считается следующее расположение считывающих элементов:

- для круговых кодовых шкал СЭ располагаются наиболее равномерно по окружности шкалы;

- для линейных шкал СЭ располагаются наиболее компактно (чтобы избежать значительного увеличения длины кодовой шкалы).

В случае неудачного совместного размещения информационных и корректирующих СЭ могут выполняться повторно:

- расчет другого варианта расположения информационных СЭ на кодовой шкале;
- синтез другого варианта кодовой маски шкалы.

Соответственно, если было принято решение осуществить синтез иной кодовой маски, то размещение всех СЭ производится повторно.

Заключение

В работе предложена классификация преобразователей перемещений на основе рекурсивных кодовых шкал. Сформулированы основные задачи, решаемые автоматизированной системой при проектировании РКШ. Разработан обобщенный алгоритм функционирования системы.

Литература

1. Домрачев В.Г., Мейко Б.С. Цифровые преобразователи угла: принципы построения, теория точности, методы контроля. М.: Энергоатомиздат, 1984. 328 с.
2. Азов А.К., Ожиганов А.А., Тарасюк М.В. Рекурсивные кодовые шкалы. // Информационные технологии. 1998. № 6. С. 39–43.
3. Агульник А.Р., Мусаелян С.С. Построение нелинейных двоичных последовательностей. // Радиоэлектроника. 1983. № 4. С. 19–28.

ПРАКТИЧЕСКИЕ ВОПРОСЫ СОЗДАНИЯ АВТОМАТИЗИРОВАННЫХ СИСТЕМ В ЗАЩИЩЕННОМ ИСПОЛНЕНИИ

Н.Ю. Дрюков

Научный руководитель – доктор технических наук, профессор Ю.А. Гатчин

В статье рассматриваются вопросы, связанные с организацией разработки и внедрения автоматизированных систем в защищенном исполнении. Основное внимание уделено стадиям создания систем, перечню и содержанию разрабатываемых документов. Данная статья может быть использована как руководство к действию при построении защищенных автоматизированных систем.

Введение

В настоящее время на книжных полках можно найти массу изданий, посвященных защите информации. Однако их содержание, в основном, отражает только теоретические аспекты безопасности данных. Из практических вопросов, как правило, освещаются только функциональные и технические характеристики отдельных технических средств защиты информации. В результате проектирование комплексных систем становится довольно сложной задачей, а создаваемые системы порой только осложняют работу пользователя, в то время как злоумышленник без особого труда преодолевает защитные механизмы.

Рассматриваемые в статье практические вопросы позволяют правильно организовать работы по проектированию систем для получения наиболее оптимального результата.

Основная часть

Большинство локальных вычислительных сетей, которые используют организации, невозможно назвать автоматизированными системами. Поэтому при решении вопросов, связанных с защитой информации, уместнее будет создавать не отдельную систему защиты информации, а автоматизированную систему в защищенном исполнении со встроенными механизмами защиты. Это позволит в рамках решения одной задачи устранить весь комплекс проблем, связанных как с информационной безопасностью, так и с организацией бесперебойного производственного процесса.

Автоматизированная система в защищенном исполнении – система, состоящая из персонала и комплекса средств автоматизации его деятельности, реализующая информационную технологию выполнения установленных функций в соответствии с требованиями стандартов и нормативных документов по защите информации [1, 2].

Создание автоматизированной системы в защищенном исполнении проводится в несколько стадий (рис.).

Предпроектная стадия включает в себя следующие работы:

- проведение предпроектного обследования;
- разработка аналитического обоснования создания автоматизированной системы в защищенном исполнении;
- разработка технического задания на создание автоматизированной системы в защищенном исполнении.

По результатам предпроектного обследования определяются защищаемая информация на основе документально оформленных перечней защищаемых сведений, угрозы безопасности информации и модель вероятного нарушителя, состав используемых технических средств и связи между ними, состав разработанной организационно-распорядительной документации, класс защищенности автоматизированной системы в защи-

щенном исполнении. Принимаются решения, касающиеся состава технических средств и систем, предполагаемых к использованию в разрабатываемой системе, и мероприятий по обеспечению конфиденциальности информации на этапе проектирования системы.

Стоит отметить, что чем более тщательно было проведено и обследовано и чем больше информации об организации было получено, тем более надежная система будет построена.

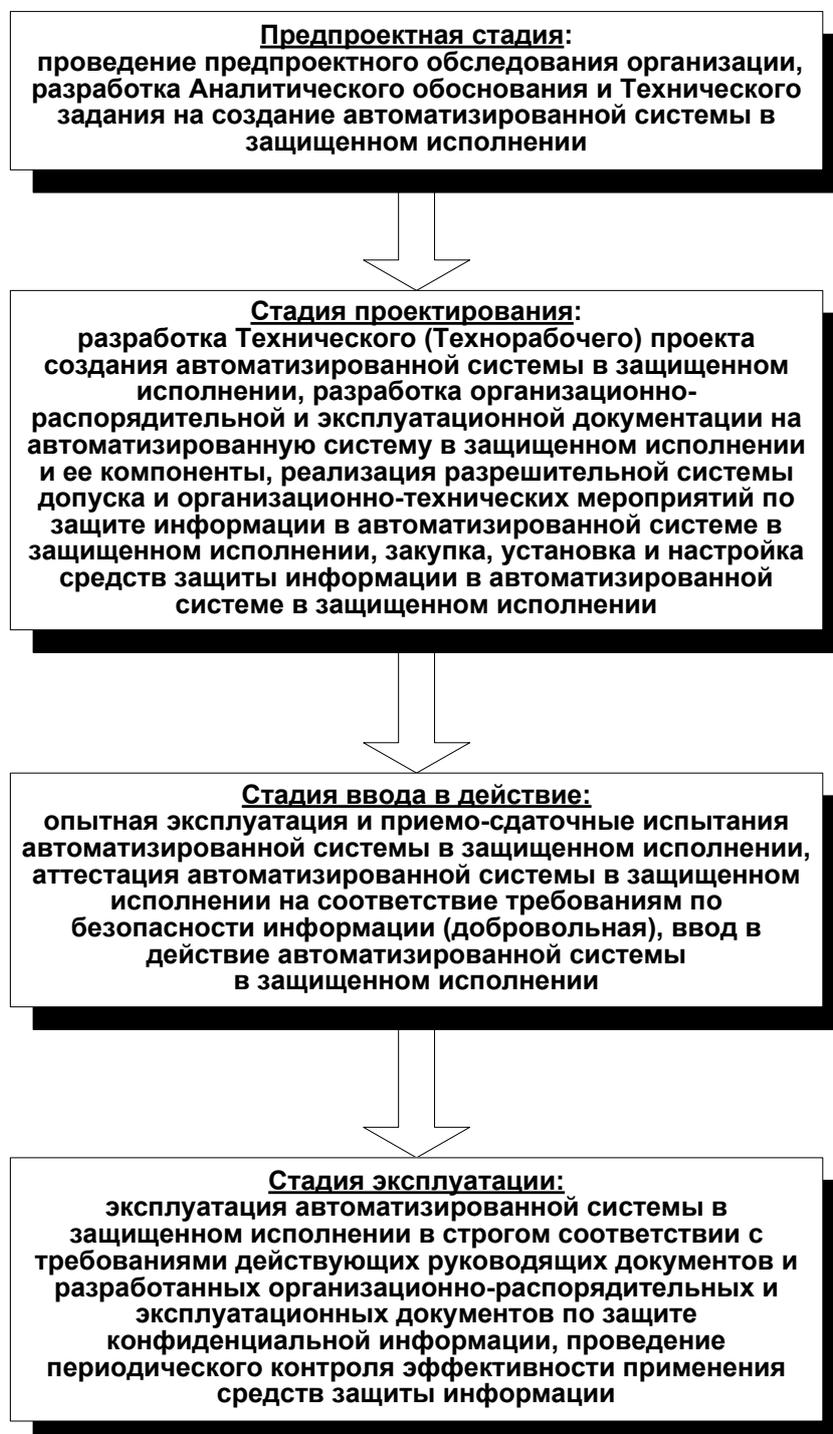


Рис. Стадии создания автоматизированной системы в защищенном исполнении

На основе полученных данных разрабатывается аналитическое обоснование необходимости создания автоматизированной системы в защищенном исполнении, кото-

рое должно содержать информационную характеристику организации, описание используемых технических средств, каналов утечки информации. Это позволяет оценить степень автоматизации деятельности сотрудников и информационную безопасность в организации и, в случае неудовлетворительного положения вещей в этих областях, принять решение о разработке автоматизированной системы в защищенном исполнении. Также в обосновании приводится оценка затрат и ориентировочные сроки на проектирование и внедрение автоматизированной системы в защищенном исполнении.

Следующим шагом будет разработка технического задания на создание автоматизированной системы в защищенном исполнении, которое должно содержать сведения об организации в техническом, программном, информационном и организационном аспектах, требования к автоматизированной системе в защищенном исполнении, перечень предполагаемых к использованию сертифицированных средств защиты информации, состав проведения работ по созданию системы.

Техническое задание является основным документом, на основании которого будет производиться проектирование системы.

Стадия проектирования включает в себя следующие работы:

- разработка технического проекта создания автоматизированной системы в защищенном исполнении;
- разработка организационно-распорядительной и эксплуатационной документации на автоматизированную систему в защищенном исполнении и ее компоненты;
- реализация разрешительной системы допуска и организационно-технических мероприятий по защите информации в автоматизированной системе в защищенном исполнении;
- закупка, установка и настройка средств защиты информации в автоматизированной системе в защищенном исполнении.

Проект должен включать в себя разделы, посвященные обеспечению автоматизации деятельности сотрудников организации, информационного обмена по высокоскоростным линиям связи, защиты информации. При необходимости разрабатываются задания и проекты на другие работы, например строительно-монтажные. На основе этого документа будет создаваться автоматизированная система в защищенном исполнении, поэтому он должен включать в себя описание всех внедряемых технических средств, их настройки, а также все необходимые организационные решения.

В состав разрабатываемой организационно-технической и эксплуатационной документации входят технический паспорт на автоматизированную систему в защищенном исполнении, инструкции и руководства по эксплуатации технических средств для пользователей, администраторов системы, приказы, акты и распоряжения, связанные с проектированием, внедрением, испытаниями и вводом в эксплуатацию системы.

Под реализацией разрешительной системы допуска и организационно-технических мероприятий по защите информации понимаются проведение следующих действий:

- определение состава субъектов доступа к защищаемой информации (сотрудников организации);
- ознакомление сотрудников с правилами обработки защищаемой информации и обеспечения информационной безопасности, наложение на всех пользователей персональной ответственности за разглашение вверенной им защищаемой информации путем подписи ими соответствующих документов;
- разработка и утверждение перечня защищаемых информационных ресурсов, матрицы доступа сотрудников к защищаемым ресурсам;
- определение состава носителей защищаемой информации и их маркировка;
- формирование журналов учета паролей, журналов учета персональных идентификаторов, журналов учета носителей защищаемой информации.

В конце этой стадии производится закупка, установка и настройка технических средств согласно требованиям, приведенным в проекте на автоматизированную систему в защищенном исполнении.

На стадии ввода в действие производятся:

- опытная эксплуатация и приемо-сдаточные испытания автоматизированной системы в защищенном исполнении;
- аттестация автоматизированной системы в защищенном исполнении на соответствие требованиям руководящих документов по безопасности информации (добровольная);
- ввод в действие автоматизированной системы в защищенном исполнении.

Опытная эксплуатация автоматизированной системы в защищенном исполнении производится в целях проверки ее функционирования в соответствии с установленными требованиями, а так же выявления недостатков и их устранения. Она осуществляется в соответствии с программой, в которой указываются условия и порядок функционирования автоматизированной системы в защищенном исполнении, продолжительность опытной эксплуатации, порядок устранения выявленных недостатков [3].

Приемо-сдаточные испытания производятся с целью проверки полноты и качества реализации системой своих функций в разных условиях функционирования. Они осуществляются в соответствии с программой, в которой указываются перечень испытываемых объектов и предъявляемых к ним требований, критерии приемки системы и их частей, условия и сроки проведения испытаний, методику испытаний и обработки результатов [3].

Проведение аттестации автоматизированной системы в защищенном исполнении является добровольным, за исключением случая, когда в организации хранятся и обрабатываются сведения, собственником которых является государство. Аттестация производится в целях проверки соответствия системы защиты информации требованиям, установленным Руководящими документами Федеральной службы по техническому и экспортному контролю (Гостехкомиссии). Проведение аттестации автоматизированной системы в защищенном исполнении осуществляется в соответствии с программой и методикой аттестационных испытаний. По результатам испытаний составляется заключение и, в случае положительного решения аттестационной комиссии, выдается аттестат соответствия объекта информатизации требованиям по безопасности информации.

Ввод в действие автоматизированной системы оформляется соответствующим актом. С этого момента в автоматизированной системе в защищенном исполнении можно обрабатывать информацию, подлежащую защите.

Последняя стадия создания системы – это эксплуатация. Здесь следует выделить две особенности.

Во-первых, чтобы в автоматизированной системе в защищенном исполнении информация находилась в безопасности, необходимо осуществлять эксплуатацию в соответствии с требованиями разработанных организационно-распорядительных и эксплуатационных документов, а также руководящих документов по защите информации.

Во-вторых, обязательно должен проводиться ежегодный контроль эффективности применения средств защиты информации с целью выявления новых угроз безопасности данных и проверки адекватности применяемых в организации защитных мер.

Заключение

Создание систем защиты информации – сложная и ответственная задача. Предложенный в статье план разработки и внедрения защитных механизмов позволяет повысить эффективность работ по созданию защищенных систем за счет структурированной

последовательности действий и возможности согласовывать результаты каждой из стадий с заказчиком системы.

Литература

1. ГОСТ 34.003-90. Информационная технология. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Термины и определения. М., 1990.
2. ГОСТ Р 51583-2000. Защита информации. Порядок создания автоматизированных систем в защищенном исполнении. Общие положения. М., 2000.
3. ГОСТ 34.603-92. Информационная технология. Виды испытаний автоматизированных систем. М., 1993.
4. ГОСТ 34.201-89. Информационная технология. Комплекс стандартов на автоматизированные системы. Виды, комплектность и обозначение документов при создании автоматизированных систем. М., 1989.
5. ГОСТ 34.601-90. Информационная технология. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Стадии создания. М., 1990.
6. Домарев В.В. Безопасность информационных технологий. Системный подход. Киев, 2004.

ГЕНЕТИЧЕСКИЙ АЛГОРИТМ РАЗМЕЩЕНИЯ

К.В. Касев

Научный руководитель – доцент Н.Ю. Иванова

В статье рассмотрены основные положения и принципы работы генетического алгоритма размещения и программа на языке C++, иллюстрирующая работу алгоритма.

Введение

Природа всегда поражала многообразием и богатством своих проявлений. Среди примеров можно назвать сложные социальные системы, иммунные и нейронные системы, взаимосвязи между видами. Они – всего лишь некоторые из чудес, которые стали более очевидны, когда человек стал глубже исследовать себя самого и мир вокруг. Многие из того, что мы видим и наблюдаем, можно объяснить единой теорией – теорией эволюции через наследственность, изменчивость и отбор.

Поэтому неудивительно, что ученые, занимающиеся компьютерными исследованиями, обратились к теории эволюции в поисках вдохновения. Возможность того, что вычислительная система, наделенная простыми механизмами изменчивости и отбора, могла бы функционировать по аналогии с законами эволюции в природных системах, была очень привлекательна. Эта надежда стала причиной появления ряда вычислительных систем и программных комплексов, построенных на принципах естественного отбора.

Есть и иные исследовательские проекты, такие как аспектно-ориентированное программирование, интенциональное программирование, самонастраивающиеся вычислительные системы и т.д. Они не имеют тесной связи друг с другом и, как правило, не идут так далеко, как генетическое программирование, имитирующее биологические процессы эволюции компьютерных технологий. Существует утверждение, что вся биология есть не что иное, как сложная серия вычислений. Но все эти проекты строятся на принципах эволюции и процедуре постижения сложных явлений путем структурирования новыми методами.

Основные понятия, термины и описание работы алгоритма

Нельзя описывать генетический алгоритм, не напомнив элементарные понятия из курса биологии, поэтому начнем с них:

- популяция – случайное множество случайных решений;
- особь – одно из решений множества;
- хромосома – кодированное представление одного варианта топологии;
- ген – элемент хромосомы, задающий некоторый фрагмент топологии.

Определяя размещение как определение координат установки элементов низшего иерархического уровня в монтажном пространстве элементов высшего иерархического уровня, можно сформулировать задачу алгоритма размещения: есть какое-то монтажное пространство, на котором требуется разместить элементы, связанные между собой, оптимальным образом с использованием целевой функции. В качестве целевой функции в предлагаемом алгоритме используется совокупность двух важных показателей: количество пересечений между связями и суммарная длина всех связей. Для получения оптимального размещения целевая функция должна стремиться к минимуму.

$$F = k_1 F_1 + k_2 F_2 \Rightarrow \min,$$

где F_1 – количество пересечений между связями, F_2 – суммарная длина всех связей, k_1 , k_2 – коэффициенты важности показателей, причем $k_1 \ll k_2$.

Так как программа иллюстрирует работу алгоритма с помощью графов, то проведем аналогию между графами и биологическими терминами: граф – хромосома, вершина графа – ген хромосомы, координаты вершины – характеристики гена.

Степень приспособленности каждой особи определяем количеством пересечений между ребрами графа и суммарной длиной всех ребер графа. Наиболее приспособленной считается особь с наименьшим количеством пересечений, если таковых особей несколько, то выигрывает особь с наименьшей суммарной длиной.

Опишем весь алгоритм, используя введенные понятия.

Шаг 1. Определение количества генов в одной хромосоме.

Шаг 2. Задание первой родительской особи.

Шаг 3. Генерация второй родительской особи со случайными параметрами генов.

Шаг 4. Парно сравниваем соответственные гены родительских особей и случайным образом обмениваем их характеристики.

Шаг 5. Из получившихся двух хромосом создаем две особи потомков.

Шаг 6. Вычисляем степени приспособленности каждой особи.

Шаг 7. Сравниваем степени приспособленности особей и выбираем лучшую из них.

Шаг 8. Если получившийся результат удовлетворяет пользователя то конец, иначе Шаг 2.

Пример генерации потомков.

Пусть хромосома каждой особи состоит из четырех генов, каждый ген, в свою очередь хранит данные **X** и **Y**, хромосомы родительских особей приведены в табл. 1. Один из многочисленных вариантов хромосом потомков показан в табл. 2.

		Ген № 1	Ген № 2	Ген № 3	Ген № 4
Родительская Особь	X	1	3	5	7
	Y	2	4	6	8
Новая особь	X	9	11	13	15
	Y	10	12	14	16

Таблица 1. Хромосомы родительских особей

		Ген № 1	Ген № 2	Ген № 3	Ген № 4
Потомок № 1	X	1	11	5	7
	Y	2	12	6	16
Потомок №2	X	9	3	13	15
	Y	10	4	14	8

Таблица 2. Хромосомы потомков

В данном примере у нас появились два потомка (см. табл. 2) вследствие полной замены информации гена № 2 и частичной замены у гена № 4.

Описание программы

Программа была написана на объектно-ориентированном языке C++, что позволило подойти ближе к решению поставленной задачи, используя более простой и удобный синтаксис языка.

Исходные данные. Массив множества вершин (элементов), на основании которого заполняется матрица связности $R = // r_{ij} //$, где r_{ij} – количество связей между вершина-

ми (элементами) графа. Начальный граф является самой первой родительской особью, матрица связности данного графа является хромосомой данной особи, вершины графа – генами хромосомы.

Вывод результатов. При генерации новой популяции происходит выбор наиболее приспособленной особи, которая в дальнейшем станет родительской особью. В результате итерационного процесса работы алгоритма данная родительская особь и будет являться наилучшим или оптимальным вариантом размещения.

Заключение

В данной статье рассмотрены основные положения и принципы работы генетического алгоритма размещения, а также программа, иллюстрирующая работу алгоритма.

Литература

1. Курейчик В.М. Генетические алгоритмы. Таганрог: ТРТУ, 1998. 242 с.
2. <http://www.comprice.ru/logovo/2003-21.phtml>

ГЕНЕТИЧЕСКИЙ АЛГОРИТМ ТРАССИРОВКИ

В.С. Полежаев, Д.В. Пудов

Научный руководитель – доцент Н.Ю. Иванова

В статье рассмотрены основные положения и принцип работы генетического алгоритма трассировки, использующего волновую процедуру для выбора оптимального варианта трассировки печатных плат.

Введение

Задаваясь вопросом, почему хамелеон запросто меняет цвет кожи под окружающую обстановку, почему у пустынных растений (верблюжья колючка) не листья, а иголки, чтобы не испарять лишней влаги, у снежного барса и уссурийского тигра густой мех, мы приходим к выводу, что эти особенности животные приобретают в процессе эволюции, оставляя необходимые качества, чтобы наилучшим образом приспособиться к окружающей среде. Основным механизмом эволюции – это естественный отбор. Его суть состоит в том, что более приспособленные особи имеют больше возможностей для выживания и размножения и, следовательно, приносят больше потомства, чем плохо приспособленные особи популяции. При этом благодаря передаче генетической информации потомки наследуют от родителей основные их качества. Таким образом, потомки сильных индивидуумов также будут относительно хорошо приспособленными, а их доля в общей массе особей будет возрастать. После смены нескольких десятков или сотен поколений средняя приспособленность особей в популяции заметно возрастет. Используя знания о генетике и об эволюции, люди давно стали применять и распространять их принципы в вычислительной технике, экономике и в других областях, где конечной целью является получение оптимального решения в сложной комбинаторной задаче.

Основные понятия

Прежде всего, необходимо расшифровать некоторые термины, относящиеся к теории эволюции и генетике в применении к задаче поиска и оптимизации решений:

- популяция – случайное множество случайных решений;
- особь популяции – одно из решений множества;
- размер популяции – число решений в множестве;
- число поколений – количество итераций, в течение которых обрабатывается множество (время генетического поиска);
- хромосома – наследственная информация, генотип особи: закодированная структура данных S , определяющая, описывающая решение (строка или строка кодов);
- ген – элементарный код в структуре S ;
- фенотип (фитнес) – оценка решения, позволяющая судить о его качестве (целевая функция);
- кроссинговер – генетический оператор, при использовании которого происходит генерация новых решений (потомков) путем переконфигурации двух решений из популяции (родителей);
- мутация – генетический оператор генерации нового решения на основе старого путем перестройки кода его структуры или самой структуры.

Генетический алгоритм (ГА) – это простая модель эволюции в природе, реализованная в виде компьютерной программы. В нем используются как аналог механизма генетического наследования, так и аналог естественного отбора. При этом сохраняется биологическая терминология в упрощенном виде. С помощью ГА осуществляется поиск и нахождение новых субоптимальных решений за счет объединения как уже имевшихся в начальной популяции, так и полученных в предыдущих циклах решений. ГА

как итерационная процедура работает до тех пор, пока не пройдет заданное число циклов, либо не будет получено решение, удовлетворяющее заданным критериям [1].

Основные отличия ГА от других оптимизационных и поисковых процедур заключаются в следующем.

- Поиск промежуточного решения происходит выбором из уже имеющегося случайного набора решений, а также их оптимизацией по заданным критериям, т.е. свойства промежуточных субоптимальных решений улучшаются.
- Решение задачи представляет собой некоторую закодированную структуру, а не набор параметров, что в некоторых случаях позволяет ускорить процесс поиска-оптимизации.
- Для оценки пригодности решения, наряду с использованием целевой функции, дополнительно моделируются правила выживания в исследуемом множестве. Эти правила повышают разнообразие множества решений, которое необходимо для выполнения первого пункта.
- При генерации, преобразовании и других видах обработки решений широко используются вероятностные и случайностные методы и правила, тем самым, позволяя избежать образования набора одинаковых решений.

Существует несколько вариантов ГА. Рассмотрим основные этапы так называемого простого генетического алгоритма.

1. Генерация начальной популяции. На этом этапе создается первоначальное множество решений исходной задачи, причем стратегии их создания могут быть различны [2]. Наибольшее распространение получили следующие стратегии.
 - a. Стратегия «одеяла», т.е. полный охват всевозможных решений. Не всегда возможно осуществить полный перебор, что является недостатком данного пути. С точки зрения адаптивного развития не предоставляет интерес, так как начальная популяция уже включает все решения, и происходит только выбор оптимального. Кроме того, этот путь может занимать ощутимо большее время.
 - b. Стратегия «фокусировки» – стартовое множество включает разновидности одного решения. Применяется в тех случаях, когда есть предположение, что некоторое решение является разновидностью известного промежуточного. Тогда путем постепенных малозначительных изменений имеющегося решения можно получить более качественное. Недостатком можно назвать отсутствие четкой зависимости качества решения от задаваемых параметров.
 - c. Стратегия «дробовика» – генерируется случайное множество решений, достаточно большое, но не исчерпывающее всех возможных вариантов.
2. Случайный выбор родительских особей.
3. Копирование родительских хромосом для применения генетических операторов.
4. Применение оператора кроссинговера (ОК) с вероятностью $P(OK)$.
5. Применение оператора мутации (ОМ) с вероятностью $P(OM)$.
6. Отбор и последующее удаление хромосом из популяции для восстановления первоначального ее размера.
7. Если не прошло заданное число поколений и не выполнилось какое-либо дополнительное условие, то переход к шагу 2.
8. Вывод индивида с наилучшим значением целевой функции.

Фенотип, целевая функция. Как уже было сказано, критерием выбора и оценки решений служит целевая функция, зависящая от нескольких переменных. Такими переменными, применительно к САПР, могут быть: суммарная длина трасс, количество переходных отверстий и т.д.

Оператор кроссинговера. Существует несколько разновидностей оператора кроссинговера: стандартный, универсальный, упорядочивающий, частично соответствующий и некоторые другие. Стандартный кроссинговер в зависимости от количества

точек скрещивания делится на одноточечный, двухточечный и многоточечный. Рассмотрим вариант одноточечного кроссинговера [3].

Допустим, родитель 1 имеет стринг «011011001», а родитель 2 – «101101101»; тогда, выбрав случайным образом точку скрещивания (см. табл.), получим 2 потомка, являющихся совершенно новыми решениями.

	Точка скрещивания->								
Родитель1	0	1	1	0	1	1	0	0	1
Родитель2	1	0	1	1	0	1	1	0	1
Потомок1	0	1	1	0	1	1	1	0	1
Потомок2	1	0	1	1	0	1	0	0	1

Таблица. Пример одноточечного кроссинговера

Оператор мутации. Под влиянием внешней среды в хромосомах индивида может произойти разрыв. В большинстве случаев фрагменты снова воссоединяются по месту разрыва. Если такого восстановления не происходит, фрагменты остаются либо открытыми, либо могут воссоединиться другим способом, что приводит к хромосомной перестройке. Это приводит к изменению наследственной информации. Изменение наследственной информации в течение жизни одного индивида называется мутацией, а сам индивид – мутантом.

Мутация – это генетическое изменение, приводящее к качественно новому проявлению основных свойств генетического материала: дискретности, непрерывности или линейности. Генные мутации на молекулярном уровне – результат различных повреждений в молекулах ДНК. В пределах одного гена за одно поколение обычно встречаются единичные повреждения, редко – большие. Результатом этого могут быть гибель клетки, изменения характера индивидуального развития, изменение признаков и т.д. В зависимости от характера изменений, возникающих в генетическом материале, выделяют следующие типы мутаций:

- точечные;
- хромосомные aberrации, или перестройки;
- инсерции;
- полиплоидизацию.

Примером точечной мутации является замена одного из генов хромосомы «011100» на противоположный: «011000» [3].

Волновая процедура и работа программы

Волновая процедура в генетическом алгоритме применяется на стадиях формирования начальной популяции, а также при формировании новых решений после применения генетических операторов и является одним из самых уникальных алгоритмов трассировки. Она позволяет построить трассу (путь) между двумя элементами в любом лабиринте, при условии, что такая трасса вообще существует. Она дает возможность легко учитывать технологическую специфику печатного монтажа со своей совокупностью конструктивных ограничений. Такой принцип трассировки применяется в основном в однослойных платах с нерегулярным размещением компонентов. Недостатком его является значительное число длинных, параллельных проводников, вызывающее большую взаимоиндуктивность, а также малое быстродействие программ на его основе на большом монтажном пространстве, при большом количестве проводников и контактных площадок.

Все ячейки монтажного поля подразделяют на занятые и свободные. Занятыми считаются ячейки, в которых уже расположены проводники, построенные на предыдущих шагах, или находятся монтажные выводы элементов, а также ячейки, соответствующие границе платы и запрещенным для прокладывания проводников участкам. Каждый раз при проведении новой трассы можно использовать лишь свободные ячейки, число которых по мере проведения трасс сокращается. На множестве свободных ячеек моделируют волну от одной ячейки к другой. Точка, испускающая волну – источник, вторая точка – приемник. Для оценки эффективности проводки каждой ячейке фронта волны присваивают определенный вес:

$$P_k = P_{k-1} + \varphi(f_1, f_2, \dots, f_g),$$

где P_k и P_{k-1} – веса k -го и $(k-1)$ -го фронтов; $\varphi(f_1, f_2, \dots, f_g)$ – весовая функция, являющаяся показателем качества проведения пути, каждый параметр которой $f_i (i=1, 2, \dots, g)$ характеризует путь с точки зрения одного или нескольких критериев качества (длины пути, числа пересечений и т.п.). На P_k накладывают одно ограничение – веса ячеек предыдущих фронтов не должны быть больше весов ячеек последующих фронтов. Фронт распространяется только на соседние ячейки, которые имеют с ячейками предыдущего фронта общую сторону. Процесс распространения волны (см. рис.) продолжается до тех пор, пока ее расширяющийся фронт не достигнет приемника или на Θ -ом шаге не найдется ни одной свободной ячейки, которая могла бы быть включена в очередной фронт, что соответствует случаю невозможности проведения трассы при заданных ограничениях.

10	11	10	9	8	7	8	9	10	11
9	8	9	10	8	6	7	8	9	10
8	8	8	9	8	5	6	7	8	9
7	6	7	8	8	4	5	6	7	8
6	5	6	8	2	3	4	8	6	7
5	4	8	2	1	2	3	8	5	6
4	3	2	1	A	1	2	3	4	5
5	4	3	2	1	2	3	4	5	6
6	5	4	3	2	3	4	5	6	7

Рис. Пример распространенной волны

Если в результате распространения волна достигла приемника, то осуществляют «проведение пути», которое заключается в движении от приемника к источнику по пройденным на этапе распространения волны ячейкам, следя за тем, чтобы значения P_k монотонно убывали. В результате получают путь, соединяющий эти две точки. Из описания алгоритма следует, что все условия, необходимые для проведения пути, закладываются в правила приписывания веса ячейкам.

Описание программы

Программа написана на языке Delphi. Причинами выбора этого языка были: легкость создания графической части (интерфейса), простота написания кода и создания информационной модели рабочих массивов, а также несложность прочтения кода без сопроводительной документации.

Исходные данные. Задается массив, содержащий множество выводов компонентов в дискретном монтажном пространстве, а также массив, содержащий множество связей, соединяющих эти выводы между собой. Происходит создание матрицы смежности.

Вывод данных. Происходит выбор и вывод наиболее оптимального решения при заданных начальных условиях. Кроме этого, программа осуществляет вывод всех про-

межуточных результатов: начальную популяцию, выбранных родителей и отображает популяцию решений после применения операторов кроссинговера и мутации.

Данная программа является моделью для изучения работы генетического алгоритма и волновой процедуры. Она также может помочь в учебном процессе для подготовки к работе со специализированными программными пакетами, такими как PCAD, AUTOCAD и др. так как наглядно иллюстрирует в упрощенном виде часть конструкторского этапа автоматизированного проектирования, а именно трассировку.

Как известно, работа с моделями упрощает понимание исследуемого процесса, обычно не требуя больших денежных затрат на ее организацию, потому можно порекомендовать включать подобный вид деятельности в образовательный процесс.

Заключение

Как уже было отмечено, генетические алгоритмы – достаточно универсальное средство для поиска оптимальных решений, потому в настоящее время они используются в большинстве программных продуктов САПР. В работе изложены основные принципы работы простого генетического алгоритма волновой трассировки печатных плат с использованием обучающей программы, иллюстрирующей его работу.

Литература

1. Аналитические технологии для прогнозирования и анализа данных. / <http://www.neuroproject.ru/genealg.htm>
2. Курейчик В.М. Генетические алгоритмы. Таганрог: изд-во ТРТУ, 1998. 242 с.
3. Курейчик В.М., Родзин С.И. Эволюционные алгоритмы: генетическое программирование. Обзор. // Известия РАН. ТиСУ. 2002. №1. С. 127–137.

ИСПОЛЬЗОВАНИЕ ФУНДАМЕНТАЛЬНЫХ АЛГОРИТМОВ ПОИСКА И ОБЪЕДИНЕНИЯ В РАЗРАБОТКЕ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ЭЛЕКТРОННЫХ АРХИВОВ САПР

А.А. Малинин, Ю.Б. Таяновская

Научный руководитель – доцент Н.Ю. Иванова

В статье описаны алгоритмы объединения-поиска в разрезе использования в современных системах автоматизированного проектирования.

Введение

Неотъемлемой частью современных систем автоматизированного проектирования (САПР) является подсистема документационного сопровождения. Под термином «документ» подразумевается отражение любой операции на всех этапах проектирования в определенной, как правило, предустановленной форме.

Документы обладают некоторыми общими свойствами с остальными объектами САПР, как-то: название; дата создания, модификации, доступа; список авторов; родители и наследники; история изменений и так далее. В то же время можно выделить ряд уникальных особенностей, например: срок действия документа; тематика; ссылки на схожие документы; наличие подчинённых объектов и т.д. – всего двенадцать различных свойств.

В результате работы современной САПР архив документации является сложной организованной многосвязной структурой с количеством объектов порядка 10^4 . В рамках разработки структуры подобного хранилища данных и процессов его управления разработчикам приходится решать фундаментальные вопросы: задачу связности, реализацию алгоритмов объединения и поиска.

В статье будет показано, что важно затратить большее время на разработку структуры и алгоритмов. При повышении эффективности работы программы, используемое машинное время уменьшается до 10^2 – 10^6 раз, в то время как аппаратное быстроедействие с каждым поколением увеличивается на порядок.

Задача связности

Предположим, что связь двух документов по какому-либо свойству представляется парой чисел $a-b$, т.е. пара $a-b$ интерпретируется как « a связано с b ». Пусть это отношение является транзитивным.

В процессе работы системы документооборота имеем набор существующих связей. Допустим, возникает ситуация, когда необходимо создать новую связь $a-b$. Задача состоит в разработке алгоритма, в результате работы которого новая связь добавлялась бы только тогда, когда добавленные ранее пары не связывают a и b .

Алгоритмы объединения-поиска

В процессе работы системы часто возникает потребность в нахождении связанных документов: при создании нового, просмотре и редактировании существующих документов, во время поиска.

Во время поиска существующих связей необходимо найти все объекты, связанные с заданным. При добавлении новой связи каждый раз, когда алгоритм получает новую пару чисел, необходимо определить, представляет ли она новое соединение, а затем добавить информацию об обнаруженном соединении в общую картину о связности объектов для дальнейших проверок.

Обычно эти две задачи инкапсулируются в виде абстрактных операций поиска и объединения, считая, что вводимые целочисленные значения представляют элементы в неких абстрактных наборах.

Работа прикладной программы

Как правило, в процессе разработки эффективного алгоритма решения поставленной задачи первым шагом является реализация простого алгоритма решения. Если нужно решить несколько вариантов конкретной задачи, то на простой реализации можно и остановиться. Если требуется более сложный алгоритм, то простая реализация позволяет проверить правильность работы алгоритма для простых случаев и послужит отправной точкой для оценки характеристик производительности.

Программа 1 – реализация простого алгоритма, который решает задачу связности. В основе этого алгоритма лежит использование массива целых чисел, обладающего тем свойством, что a и b связаны тогда, когда a -й и b -й элементы массива равны. Вначале i -й элемент массива инициализируется значением i при $0 \leq i \leq N$. Более простые варианты, например, с построением массива, подобного матрице смежности, или с изменением двух элементов массива, в этой статье рассмотрены не будут, так как они не являются столь показательными, а их эффективность не выше рассматриваемого алгоритма.

Программа 1.

```
function QuickF($N, $in) {
  $id = array($N);
  for ($i = 0; $i < $N; $i++) $id[$i] = $i;
  for each $line in $in {
    $temp = split ($line, " ");
    $a = $temp[0];
    $b = $temp[1];
    $temp = $id[$a];
    if ($temp == $id[$b] continue;
    for ($i = 0; $i < $N; $i++)
      if ($id[$i] == $temp) $id[$i] = $id[$b];
    echo ("$a $b");
  }
}
```

Как видно из кода, для реализации поиска достаточно проверить указанные элементы массива на равенство. Поэтому подобные алгоритмы называют алгоритмами быстрого поиска. Ниже будет проанализировано быстродействие этой реализации и предложено другое решение.

Анализ алгоритмов

Анализ – это ключ к пониманию алгоритмов в степени, достаточной для их эффективного применения при решении прикладных задач. Время выполнения большинства алгоритмов зависит от входных данных. При анализе важно использовать различные исходные данные – реальные, случайные и необычные. К сожалению, не всегда в результате анализа можно дать гарантии быстродействия, зачастую находятся ограничения или предсказания о затрачиваемых ресурсах на выполнение алгоритма.

Алгоритму, показанному выше, для решения задачи связности N объектов, требующей выполнения M операций объединения, требуется выполнение MN инструкций.

При разработке пакета прикладных программ очень важно оценить эффективность используемых алгоритмов до начала их реализации в виде программного кода. Анализ алгоритмов может оказаться очень трудоемким, например, когда анализ сложного алгоритма ведет к нерешенным математическим проблемам, или характеристики

входных данных не поддаются точному описанию. Но без этого этапа разработчик не может гарантировать эффективного использования системных ресурсов.

Рассмотрим другой вариант реализации задачи связности. В этой программе используется дополнительный массив для хранения количества узлов, чтобы присоединять меньшее из двух деревьев к большему. Подобные реализации называют взвешенными версиями быстрого объединения.

Программа 2.

```
function QuickUW($N, $in) {
    $id = array($N);
    $sz = array($N);
    for ($i = 0; $i < $N; $i++) {
        $id[$i] = $i;
        $sz[$i] = 1;
    }
    for each $line in $in {
        $temp = split ($line, " ");
        $a = $temp[0];
        $b = $temp[1];
        $temp = $id[$a];
        for ($i = $a; $i != $id[$i]; $i = $id[$i]);
        for ($j = $b; $j != $id[$j]; $j = $id[$j]);
        if ($temp == $id[$b] continue;
        if ($sz[$i] < $sz[$j]) {
            $id[$i] = $j;
            $sz[$j] += $sz[$i];
        }
        else {
            $id[$j] = $i;
            $sz[$i] += $sz[$j];
        }
        echo ("$a $b");
    }
}
```

В результате анализа быстродействия этой реализации приходим к выводу, что количество инструкций для обработки M ребер между N объектами не превышает $M \lg N$, умноженного на некоторую константу. Этот вывод резко отличается от вывода о быстродействии программы 1. На практике при количестве объектов, равном 10 000, время, затраченное этим алгоритмом, меньше предыдущего примерно в 40 раз. Таким образом, при использовании взвешенного быстрого объединения можно гарантировать решение очень больших задач за приемлемое время.

Заключение

Фундаментальные алгоритмы поиска, сортировки и объединения, базовые структуры данных, а также основные подходы к анализу быстродействия алгоритмов в большинстве случаев были описаны еще в 1975 году (например, в книге А. Эйхо, Дж. Хопкрофта и Дж. Ульмана «Разработка и анализ алгоритмов», Addison-Wesley, Reading, MA, 1975). Значимость этих разработок трудно переоценить даже сегодня, ведь как было сказано в начале статьи, быстродействие алгоритма можно повысить в 10^4 раз – пожалуй, этому способу математической оптимизации уступают возможности оптимизации машинного кода, настройки аппаратного обеспечения и сетевого оборудования, вместе взятые.

**ИССЛЕДОВАНИЕ ПОСТРАДИАЦИОННОГО ЭФФЕКТА
В МОП-СТРУКТУРАХ, ОБЛУЧЕННЫХ ПОД НАПРЯЖЕНИЕМ****Ю.В. Смирнов, Р.А. Халецкий****Научный руководитель – доктор технических наук, профессор А.М. Скворцов**

Приведены результаты исследования процессов изменения электрофизических параметров МОП-структур при их выдержке при комнатной температуре после радиационного воздействия под напряжением смещения на затворе различной амплитуды и полярности. Показано, что пострадиационные процессы связаны с наличием в окисле отжигаемых и неотжигаемых при комнатной температуре дефектов, характеризующихся различным энергетическим положением в запрещенной зоне.

Введение

Стабильность электрических параметров интегральных схем (ИС) типа металл-окисел-полупроводник (МОП) при воздействии ионизирующего излучения в значительной мере определяется свойствами системы затвор-подзатворный окисел-подложка МОП-структуры, составляющей основу конструкции МОП ИС. Ионизирующее излучение приводит к изменению электрофизических параметров МОП-структуры, которое заключается в наведении радиационного встроенного заряда в окисле и увеличении плотности поверхностных состояний на границе раздела кремний-диоксид кремния [1, 2].

После радиационного воздействия в окисле продолжают протекать процессы, обуславливающие дрейф электрофизических параметров МОП-структуры. Это явление носит название пострадиационного эффекта (ПРЭ), сущность которого заключается в термическом выбросе дырок с одних дефектов в окисле и их захвате другими дефектами, которые характеризуются более глубоким уровнем залегания в запрещенной зоне [3]. Захват дырок глубокими дефектами приводит к остаточному сдвигу порогового напряжения транзистора.

В процессе пострадиационного изменения электрофизических параметров возможно несколько вариантов проявления ПРЭ. Первый вариант заключается в значительном увеличении порогового напряжения МОП-транзистора, второй – в возвращении к исходному значению, которое было до облучения, и третий вариант – в уменьшении порогового напряжения относительно начального значения. Какой именно вариант ПРЭ будет реализован, зависит от структурных свойств окисла [4, 5] и внешних условий, в которых осуществляется выдержка МОП-транзистора (температура среды, напряжение на затворе). Однако механизм процесса изменения электрофизических свойств подзатворного окисла после радиационного воздействия остается все еще не до конца изученным. Целью настоящей работы явилось исследование пострадиационного эффекта в МОП-структурах с поликремниевым затвором, подвергавшихся облучению под смещением на затворе.

Постановка эксперимента

Экспериментальные образцы представляли структуры Si-SiO₂-Si^{*}-Al. Окисел образцов был сформирован термическим окислением кремниевой подложки марки КЭФ 7,5 (100) в режимах, представленных в таблице. Осажденный при 640°C из SiH₄ слой Si^{*} (0,6 мкм) был легирован фосфором методом термической диффузии в течение 1 ч при температуре 860°C. После получения рисунка контактных площадок в слое Si^{*} методом фотолитографии был проведен окончательный отжиг поликремния при 860°C в течение 30 мин.

Для создания омических контактов термовакуумным напылением на обе стороны пластины был нанесен слой алюминия, в котором с помощью фотолитографии были сформированы контактные площадки. Далее в тексте экспериментальные структуры будут называться тестовыми.

№ образца	Толщина окисла, нм	Продолжительность термического окисления, мин				Температура окисления, °C
		В сухом O ₂	В парах воды	В сухом O ₂	Суммарное время	
1	80	10	100	10	120	850
2	120	10	18	10	38	1000
3	120	90	–	–	90	1050
4	150	120	–	–	120	1050

Таблица. Режимы термического окисления при получении окислов экспериментальных МОП-структур

Тестовые структуры подвергались воздействию гамма-излучения, источником которого являлся Co⁶⁰ с энергией гамма-квантов 1,3 МэВ и интенсивностью излучения 0,24 Мрад/час. Поглощенная доза составила 0,24 Мрад. При облучении образцы находились под напряжением смещения различной амплитуды и полярности. Знак полярности напряжения смещения соответствует знаку полярности затвора.

Непосредственно после облучения регистрировались высокочастотные вольт-фарадные характеристики (ВФХ) экспериментальных образцов, после чего МОП-структуры выдерживались при комнатной температуре в течение 100 ч. После выдержки проводилось измерение ВФХ.

На основе полученных ВФХ производился расчет изменения плотности встроенного заряда в окисле ΔQ_f и плотности поверхностных состояний N_{SS} , вызванных воздействием гамма-излучения и выдержкой при комнатной температуре. Изменение плотности встроенного заряда рассчитывалось с использованием выражения

$$\Delta Q_f = \Delta U_{MG} C_0,$$

где ΔU_{MG} – сдвиг ВФХ относительно исходной на уровне емкости C_{MG} , соответствующей моменту пересечения уровнем Ферми середины запрещенной зоны кремния, В; C_0 – удельная емкость окисла, нФ/см².

Результаты и их обсуждение

На рис. 1 приведены экспериментальные графики, демонстрирующие изменение встроенного заряда в окисле при облучении и выдержке тестовых структур при комнатной температуре в течение 100 ч после облучения. Из рис. 1 видно, что характер из-

менения радиационно-индуцированного заряда существенно различается для положительной и отрицательной полярностей напряжений на затворе.

При увеличении абсолютного значения напряжения смещения отрицательной полярности наблюдается монотонное увеличение встроенного заряда в окисле. При этом наблюдается корреляция между величиной радиационного изменения встроенного заряда и технологическими режимами формирования подзатворного окисла. В структурах с окислом, полученным комбинированным окислением, наблюдается более интенсивное увеличение положительного встроенного заряда при облучении по сравнению с тестовыми структурами с сухим SiO₂.

Выдержка тестовых структур приводит к уменьшению радиационно-индуцированного заряда. Необходимо отметить, что уменьшение плотности заряда, вызванного облучением при положительном напряжении на затворе, является более интенсивным по сравнению с уменьшением заряда, индуцированного гамма-излучением при отрицательной полярности на затворе.

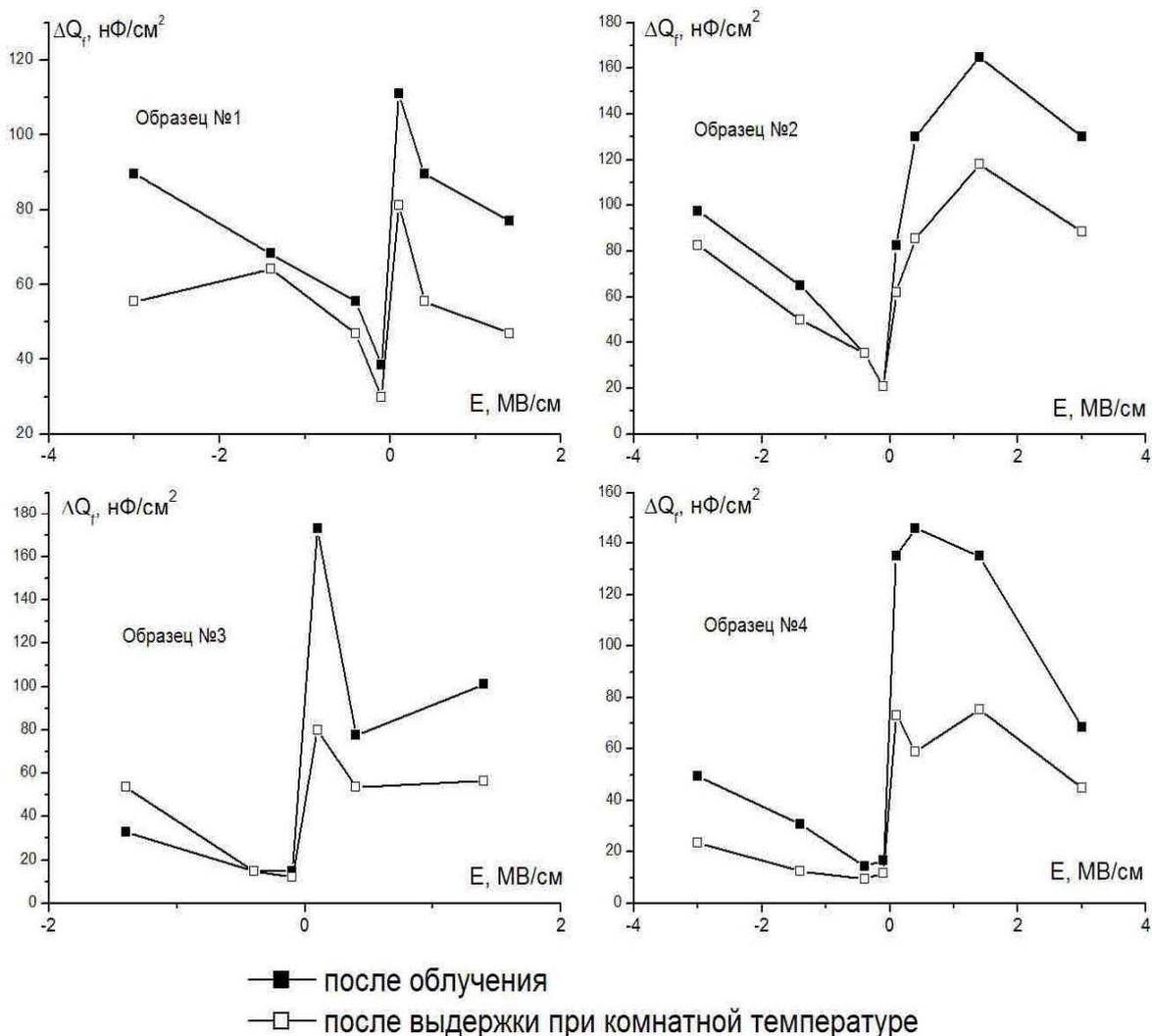


Рис. 1. Графики зависимости изменения встроенного заряда от напряженности поля в окисле для образцов после облучения и выдержки при комнатной температуре

На рис. 2 показаны подобные зависимости для плотности поверхностных состояний. В этом случае при отрицательных смещениях на затворе плотность поверхностных состояний почти не изменяется после выдержки при комнатной температуре, а для образца № 4 с сухим окислом толщиной 0,15 мкм наблюдается значительное увеличение

плотности поверхностных состояний при выдержке тестовых структур при комнатной температуре.

Полученные результаты можно объяснить следующим образом. При гамма-облучении происходит образование неравновесной электронно-дырочной пары, обусловленное переходом электрона из валентной зоны диэлектрика в зону проводимости.

Под действием электрического поля неравновесные носители заряда начинают разделяться в объеме окисла и двигаться в соответствии с полярностью смещения. При положительной полярности на затворе образованные дырки перемещаются к границе раздела, где интенсивно захватываются на напряженные связи, что приводит к увеличению положительного встроенного заряда в окисле и плотности поверхностных состояний. Тот факт, что для образцов, находившихся при положительных смещениях во время облучения, после выдержки при комнатной температуре наблюдается существенное изменение встроенного заряда, свидетельствует о том, что за изменение заряда при облучении отвечают дефекты, характеризующиеся энергетически неглубоким расположением в запрещенной зоне. Эти дефекты легко отжигаются уже при комнатной температуре, а неотжигаемые дефекты приводят к стационарному сдвигу ВФХ.

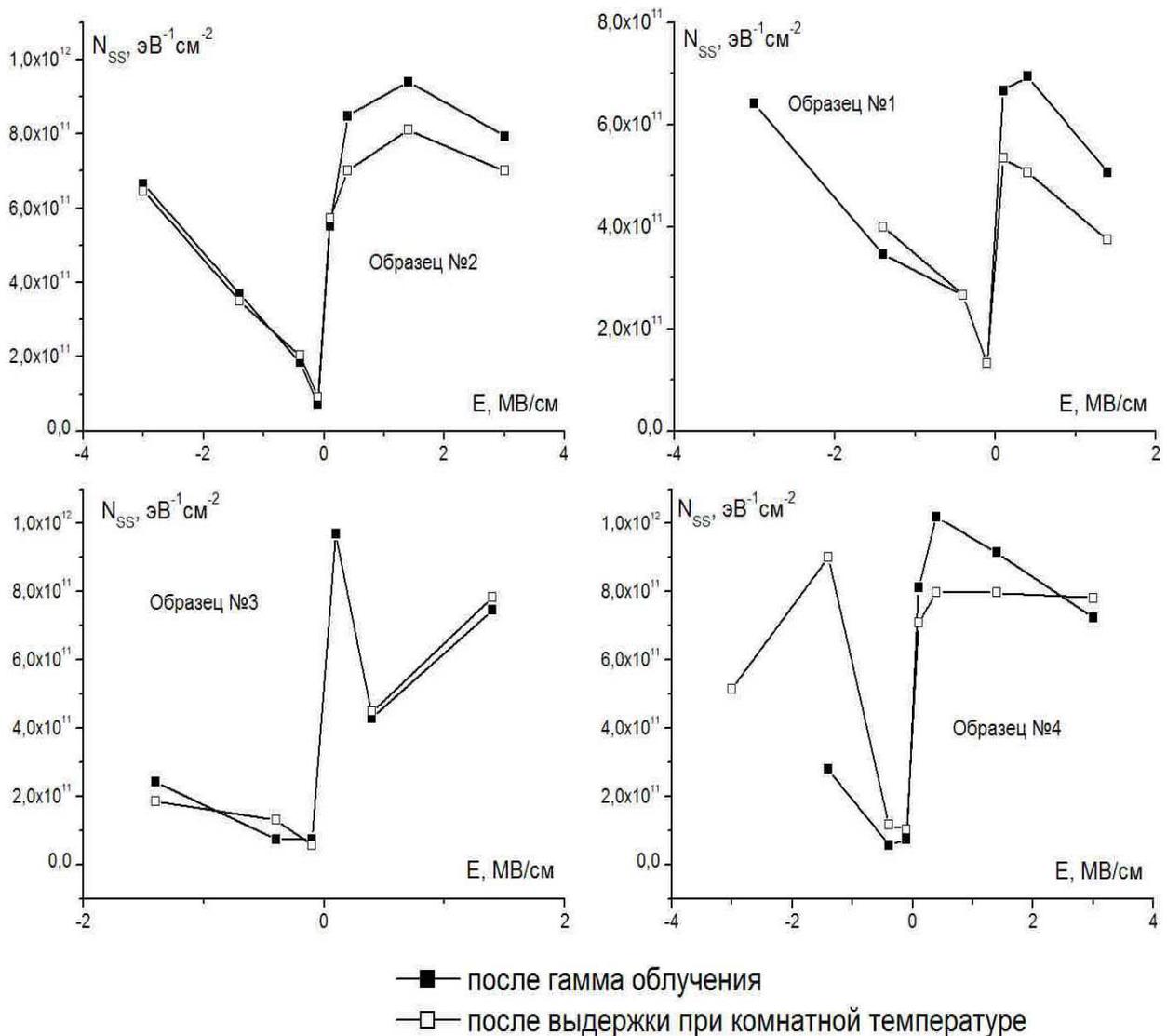


Рис. 2. Графики зависимости изменения плотности поверхностных состояний от напряженности поля в окисле при облучении и выдержке при комнатной температуре

При отрицательном напряжении на затворе неравновесные дырки уходят на затвор-катод. Поэтому в процессах дефектообразования, по-видимому, участвуют дырки из инверсионной области кремниевой подложки, которые поставляются в окисел посредством туннелирования. При этом образовавшиеся дефекты характеризуются энергетически глубоким положением в запрещенной зоне окисла, что определяется стабильностью радиационного заряда, за появление которого эти дефекты отвечают.

Заключение

Таким образом, полярность напряжения на затворе МОП-структур при облучении влияет на энергетическое положение дефектов, ответственных за изменение заряда в диоксиде кремния. В свою очередь, характер пострадиационного изменения заряда в окисле МОП-структур определяется энергетическим положением радиационных дефектов, что обусловлено вероятностью термического выброса дырок с этих дефектов в валентную зону кремния.

Литература

1. Гуртов В.А. Радиационные процессы в структурах металл-диэлектрик-полупроводник: Учебное пособие. Петрозаводск, 1988. 96 с.
2. Ringel H., Knoll M., Braunig D., Fahrner W.R. Charges in metal-oxide-semiconductor samples of various technologies induced by ^{60}Co - γ and x-ray quanta. // J. Appl. Phys. 1985. V. 57. № 2. P. 393–399.
3. Попов В.Д. Пострадиационный эффект в ИС, неразрушающий контроль качества ИС. // Электроника: Наука, технология, бизнес. 2002. Вып. 4. С. 36–39.
4. Preffer R.L. Molecular diffusion in α - SiO_2 : its role in annealing radiation-induced defect centres. // Structure and Bonding in Non-Crystalline Solids. New York: Plenum Press., 1986. P. 169–176.
5. Shanfield Z., Moriwaki M.M. Characteristics of hole traps in dry and pyrogenic gate oxides. // IEEE Transactions on Nuclear Science. 1984. V. 31. P. 1242–1247.

ГИСТЕРЕЗИС ВОЛЬТ-ФАРАДНЫХ ХАРАКТЕРИСТИК МОП-СТРУКТУР С ПОЛИКРЕМНИЕВЫМ ЗАТВОРОМ

Р.А. Халецкий

Научный руководитель – доктор технических наук, профессор А.М. Скворцов

В работе приведены результаты исследования гистерезиса вольт-фарадных характеристик (ВФХ) МОП-структур с легированным поликремниевым затвором. Показано, что при формировании поликремниевого затвора появляется характерный гистерезис ВФХ структур кремний-окисел-поликремний, который связан с процессами релаксационной поляризации, обусловленными движением дефектов в окисле.

Введение

Метод вольт-фарадных характеристик (ВФХ) является наиболее распространенным методом исследования электрофизических свойств структур вида металл-окисел-полупроводник. С помощью метода ВФХ можно определить концентрацию различных видов заряда в подзатворном диэлектрике, плотность поверхностных состояний, концентрацию примеси в подложке [1]. При измерении ВФХ в некоторых случаях может наблюдаться дисторсионный эффект или гистерезис ВФХ, который заключается в появлении разности хода между прямой и обратной ветвями ВФХ [2].

Этому явлению почти не уделяется внимания в литературе, а его механизм остается плохо изученным. Целью настоящей работы является исследование гистерезиса ВФХ МОП-структур с поликремниевым затвором на различных этапах ее формирования структуры.

Постановка эксперимента

Экспериментальные образцы представляли собой структуры вида алюминий–поликремний-окисел-кремний, которые формировались следующим образом. Кремниевые пластины марки КЭФ 7,5 (100) подвергались термическому окислению в различных режимах, представленных в таблице. На окисленную поверхность пластин методом разложения моносилана в вакууме при 640°C осаждался слой поликристаллического кремния (Si^*) толщиной 0,6 мкм. После легирования поликремниевой пленки фосфором методом термической диффузии в течение 1 ч при температуре 860°C проводились фотолитография в слое Si^* и окончательный термический отжиг при 860°C в течение 30 мин.

Для создания омических контактов термовакuumным напылением на обе стороны пластины был нанесен слой алюминия, в котором с помощью фотолитографии были сформированы контактные площадки. На рис. 1 показано сечение полученных структур, и далее в тексте они будут называться тестовыми структурами.

№ образца	Толщина окисла, нм	Продолжительность термического окисления, мин				Температура окисления, °C
		в сухом O ₂	в парах воды	в сухом O ₂	суммарное время	
1	80	10	100	10	120	850
2	120	10	18	10	38	1000
3	120	90	–	–	90	1050
4	150	120	–	–	120	1050

Таблица. Режимы термического окисления при получении окислов образцов

После окисления кремниевых пластин и после получения готовых тестовых структур проводилось измерение высокочастотных ВФХ. ВФХ измерялись в двух направлениях: прямом – из режима аккумуляции в режим инверсии (прямая ветвь ВФХ) и обратном – из режима инверсии в режим аккумуляции (обратная ветвь ВФХ).

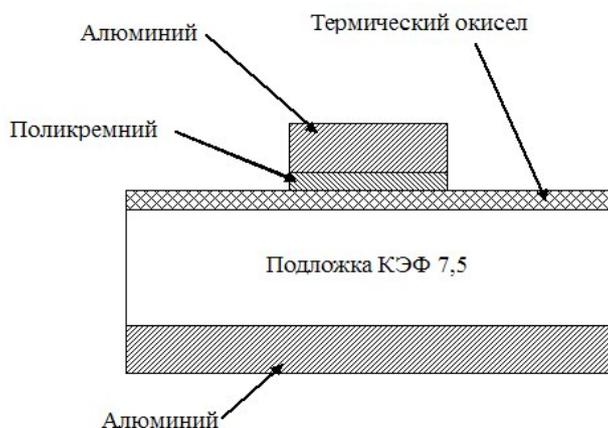


Рис. 1. Сечение тестовых структур

На основе полученных ВФХ был определен сдвиг (гистерезис) ΔU прямой ветви характеристики относительно обратной. Так как любой сдвиг ВФХ по оси напряжений относительно исходной ВФХ обусловлен изменением заряда в окисле, то сдвиг ΔU также можно интерпретировать как появление заряда, обуславливающего гистерезис ВФХ.

Сдвиг ΔU определялся при различных значениях емкости МОП-структур. Так как каждому значению высокочастотной емкости однозначно соответствует определенное значение поверхностного потенциала, то можно получить энергетическое распределение заряда Q , обуславливающего гистерезис, в запрещенной зоне кремния. При этом значение заряда Q было рассчитано по формуле

$$Q = C_{\text{ок}} \Delta U,$$

где $C_{\text{ок}}$ – удельная емкость диэлектрика, Ф/см²; ΔU – величина гистерезиса, В.

Результаты и их обсуждение

На рис. 2 показано типичное энергетическое распределение заряда $Q(E)$, обуславливающего гистерезис ВФХ, в запрещенной зоне кремния для тестовых структур после окисления подложек (Si-SiO_2) и после формирования затворной композиции ($\text{Si-SiO}_2\text{-Si}^*\text{-Al}$). Характер и амплитуда изменения $Q(E)$ мало отличались ($\pm 20\%$) для образцов с одним типом структуры. Из представленных на рис. 2 графиков видно, что после получения затворной композиции происходит существенное изменение характера зависимости $Q(E)$, которое заключается в значительном снижении чувствительности заряда, обуславливающего гистерезис ВФХ, к изменению поверхностного потенциала подложки ϕ_s ($\phi_s \sim E$), при этом знак этого заряда становится отрицательным по всей ширине запрещенной зоны кремния. Важно отметить, что вид зависимости $Q(E)$, характерной для участков структуры $\text{Si-SiO}_2\text{-Si}^*\text{-Al}$ со слоем $\text{Si}^*\text{-Al}$ на SiO_2 , является характерным и для участков без слоя $\text{Si}^*\text{-Al}$.

С целью выявления причины, которая привела к изменению характера гистерезиса ВФХ, изображенного на рис. 2, был поставлен следующий эксперимент. Структуры Si-SiO_2 , полученные комбинированным окислением при 1000°C , были разделены на 3 группы, которые затем подвергались различным обработкам, соответствующим тем, что использовались при получении пленки Si^* в тестовых структурах $\text{Si-SiO}_2\text{-Si}^*\text{-Al}$.

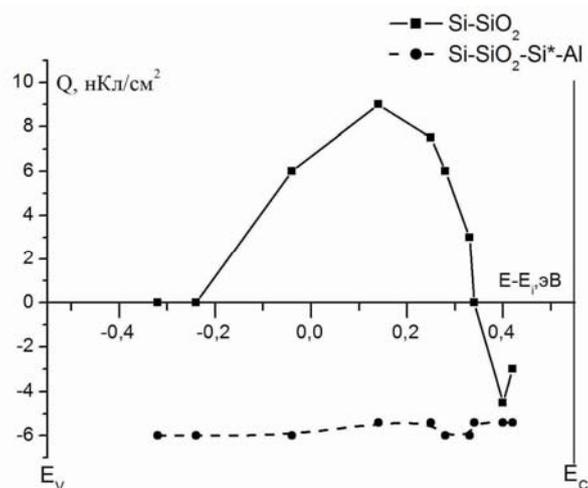


Рис. 2. Типичное энергетическое распределение заряда Q , обуславливающего гистерезис ВФХ, на различных этапах изготовления тестовых структур

1-я группа подвергалась термообработке в режиме $640^{\circ}C$ (1 ч) + $860^{\circ}C$ (1,5 ч) в среде азота. В этом случае предполагалось исследование влияния только температурной составляющей процесса формирования поликремниевого затвора. Так, в работе [3] показано, что дополнительное длительное термическое воздействие при высокой температуре может приводить к изменению структурно-примесных свойств системы кремний-окисел.

2-я группа подвергалась осаждению пленки Si^* и последующему ее удалению.

3-я группа подвергалась осаждению пленки Si^* , легированию фосфором при $860^{\circ}C$ в течение 1 ч и последующему травлению слоя Si^* .

Для обработанных структур $Si-SiO_2$ до и после обработки производилось измерение высокочастотных ВФХ, на основе которых строились графики зависимости заряда, обуславливающего гистерезис, от энергии.

В результате эксперимента было обнаружено, что изменение характера зависимости $Q(E)$, подобное приведенному на рис. 2, происходит при нанесении пленки поликремния, ее легировании при $860^{\circ}C$ в течение 1 ч и последующем травлении (образцы 3-й группы).

При анализе причины влияния легирования поликремния на изменение $Q(E)$ можно выделить два фактора.

Первый фактор – влияние фосфора. Оно заключается в том, что фосфор может накапливаться на границе раздела диоксид кремния-поликремний, приводя к образованию новой фазы $2SiO_2 \cdot P_2O_5$ и изменению электрофизических свойств окисла [4, 5]. С другой стороны, гистерезис ВФХ связывается с некими ловушечными центрами вблизи внутренней границы раздела кремний-окисел [2], куда, по нашему мнению, проникновение фосфора (диффузия при температуре $860^{\circ}C$ в течение 1 ч) маловероятно.

Второй фактор – влияние высокотемпературной обработки Si^* на электрофизические свойства окисла. В этом случае свойства окисла могут изменяться как вследствие изменения размеров зерен поликристаллической пленки [6], так и из-за проникновения водорода вглубь пленки диоксида кремния, приводя к образованию заряженных центров [7]. Вторая причина является наиболее вероятной для изменения характера гистерезиса.

Так как процессы диффузии водорода и изменения структуры поликристаллической пленки связаны с длительностью термической обработки, то был проведен следующий эксперимент. На структуры $Si-SiO_2$, полученные комбинированным окислением при $1000^{\circ}C$, был осажден поликристаллический кремний толщиной 0,6 мкм. После осаждения поликремниевой пленки образцы были разделены на группы, которые затем

подвергались отжигу при температуре 860°C в среде азота в течение разного времени: 0, 20, 40, 60, 90 мин. После отжига пленки поликремния стравливались.

В результате исследования экспериментальных образцов было обнаружено:

- операции нанесения пленки поликремния и ее последующего стравливания без дополнительных термообработок не приводят к изменению характера зависимости $Q(E)$, свойственной окислу непосредственно после окисления;
- после отжига при 860°C в течение 20 мин наблюдается появление характерной для Si-SiO₂-Si^{*}-Al структуры энергетического распределения заряда, обуславливающего гистерезис (рис. 2);
- однако после 1,5–2 часов отжига наблюдается возврат к зависимости $Q(E)$, свойственной для Si-SiO₂, которая показана на рис. 2, только в 2–2,5 раза меньшей амплитуды.

Из рис. 2 видно, что заряд, который определяет гистерезис, может быть как положительным, так и отрицательным. Знак полярности этого заряда в нашем случае определяется знаком выражения $\Delta U = U_{\text{прям}} - U_{\text{обр}}$, который в равной мере может быть рассчитан как $\Delta U = U_{\text{обр}} - U_{\text{прям}}$. То обстоятельство, что гистерезис ΔU был рассчитан относительно прямой ветви ВФХ, обусловлено тем, что при дискретном приложении различных напряжений смещения к затвору тестовых структур, уровни измеряемой емкости соответствовали значениям прямой ветви ВФХ, а обратная ветвь формировалась только при развертке напряжения смещения в обратном направлении.

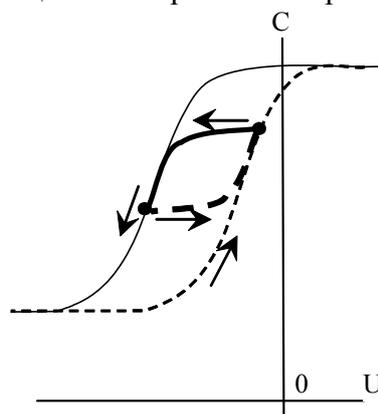


Рис. 3. Схематичное представление перехода между ветвями ВФХ при изменении направления развертки напряжения смещения

На рис. 3 схематично показаны типичная ВФХ тестовых структур, измеренная в двух направлениях, и характер перехода от прямой ветви ВФХ к обратной (и наоборот) при смене направления развертки напряжения смещения на затворе в интервале наклонного участка ВФХ. Точками на рис. 3 отмечены моменты, в которых выключалась развертка напряжения смещения и затем снова включалась в обратном направлении. Стрелками на рис. 3 показаны направления изменения напряжения смещения. Из рис. 3 видно, что изменение направления развертки напряжения на затворе приводит к появлению гистерезиса. Следовательно, гистерезис или сдвиг прямой ветви ВФХ относительно обратной зависит не от величины и знака приложенного напряжения, а от знака его изменения, т.е. знака производной dU/dt .

Производная dU/dt прямо пропорциональна производной электрического поля по времени dE/dt . В свою очередь, согласно [8], поляризация диэлектрика P связана с электрическим полем соотношением:

$$P = k\epsilon_0 E,$$

где k – относительная диэлектрическая восприимчивость, ϵ_0 – электрическая постоянная, E – напряженность электрического поля. Если производная dP/dt , а, следовательно, и dE/dt ,

существует и не равна нулю, то это означает, что в диэлектрике протекает ток поляризации, обусловленный некоторым упорядоченным движением связанных зарядов.

Таким образом, одновременно с основным процессом поляризации окисла происходит релаксационная поляризация, обусловленная, по-видимому, движением заряженных дефектов в диоксиде кремния. При изменении направления вектора напряженности происходит деформация электронных оболочек дефектов (электронная поляризация), а затем перемещение дефектов в направлении границы раздела Si-SiO₂ или от нее, что создает релаксационный ток поляризации или релаксационную поляризацию, вызывая гистерезис ВФХ тестовых структур. Изменение поверхностного потенциала кремния $\Delta\phi_s$ определяется полем E , наводимым зарядом Q на дефектах, а также их перемещением Δr , вызванным процессом латентной поляризации в окисле, согласно уравнению: $\Delta\phi_s = E/\Delta r$. При этом зависимость $Q(E)$ (рис. 2) следует интерпретировать как соответствующее этому заряду индуцированное в полупроводнике электрическое поле.

Необходимо отметить, что гистерезис ВФХ тестовых структур, наблюдаемый после формирования подзатворного окисла, отличается по характеру появления от гистерезиса ВФХ готовых тестовых структур. Это отличие заключается в том, что обратная ветвь ВФХ тестовых структур после этапа термического окисления формируется только при изменении направления развертки напряжения смещения на пологих участках ВФХ. Такое свойство гистерезиса ВФХ обусловлено способностью ловушечных центров в приграничной к кремнию области окисла обмениваться носителями заряда с подложкой [2].

Заключение

Процесс формирования поликремниевого затвора вызывает появление характерного гистерезиса ВФХ, отличающегося по своей природе от гистерезиса, обусловленного наличием ловушечных центров. При этом природа этого гистерезиса связана с релаксационными эффектами процесса диэлектрической поляризации диоксида кремния, которые возникают вследствие термической обработки структуры кремний-диоксид кремния-поликремний. Для более полного понимания природы гистерезисных эффектов и ее связи со структурой пленки подзатворного окисла требуются дополнительные исследования.

Литература

1. Зи С. Физика полупроводниковых приборов. В 2-х кн. Кн. 1. М.: Мир, 1984. 455 с.
2. Fleetwood D.M., Reber Jr. R.A., Riewe L.C., Winokur P.S. Thermally stimulated current in SiO₂. // *Microelectronics Reliability*, 1999, vol. 39, P. 1323–1336.
3. Красников Г.Я., Зайцев Н.А. Система кремний-диоксид кремния субмикронных СБИС. М.: Техносфера, 2003. 384 с.
4. Bravman J.C., Sinclair R. The Morphology of the polysilicon-SiO₂ interface. // *Thin films and interface. 2 Symp.* Boston, Mass., 14–18 Nov. 1983. New York. 1984. P. 311–316.
5. Howe K.T., Muller R.S. Stress in polycrystalline and amorphous silicon thin films. // *J. Appl. Phys.* 1983. V. 54. № 8. P. 4674–4675.
6. Глебов А.С., Зайцев Н.А., Манжа Н.М. Снижение величины эффективного заряда в МДП-структурах. // *Электронная промышленность*. 1994. № 1. С. 71–72.
7. Урицкий В.Я. Формирование и электрофизические характеристики многослойных структур на основе системы кремний–диоксид кремния: Диссертация на соискание ученой степени д-ра физ.-мат. наук. Санкт-Петербург, 1995. 341 с.
8. Геворкян Р.Г. Курс физики: учебное пособие. М.: Высш. школа, 1979. 656 с.

ВЛИЯНИЕ ЭЛЕКТРОФИЗИЧЕСКИХ ПАРАМЕТРОВ КРЕМНИЕВОЙ ПОДЛОЖКИ НА ГИСТЕРЕЗИС ВОЛЬТ-ФАРАДНЫХ ХАРАКТЕРИСТИК МОП-СТРУКТУР

Р.А. Халецкий

Научный руководитель – доктор технических наук, профессор А.М. Скворцов

В работе приведены результаты исследования гистерезиса ВФХ структур металл-окисел-полупроводник. Показано, что свойства гистерезиса ВФХ, возникающего при облучении МОП-структур, определяются типом проводимости подложки и его удельным сопротивлением и обусловлены процессами захвата носителей заряда из разрешенных энергетических зон кремния на ловушечные центры в окисле.

Введение

Диоксид кремния является наиболее распространенным материалом подзатворного диэлектрика интегральных микросхем (ИС) на основе системы металл-диэлектрик-полупроводник, а его электрофизические свойства в значительной мере определяют электрические параметры и стабильность работы МОП ИС. Заключение о качестве сформированного подзатворного окисла делают, как правило, на основе анализа вольт-фарадной характеристики (ВФХ). При этом из ВФХ рассчитывают плотность встроенного заряда в окисле и плотность поверхностных состояний на границе раздела кремний-окисел.

При исследовании ВФХ МОП-структур, подвергнутых воздействию ионизирующего излучения, может наблюдаться гистерезис ВФХ, сущность которого заключается в появлении сдвига прямой ветви ВФХ относительно обратной ветви [1, 2]. Гистерезис ВФХ может привести к сбою в работе МОП ИС во время переключения МОП-транзистора из одного режима в другой.

При исследовании ВФХ МОП-структур гистерезису почти не уделяется внимания, поэтому его свойства и природа возникновения остаются плохо изученными. Целью настоящей работы явилось исследование влияния типа проводимости полупроводниковой подложки МОП-структур на гистерезис ВФХ.

Постановка эксперимента

Экспериментальные образцы представляли собой МОП-структуры, полученные термическим окислением кремниевых пластин двух марок, отличающихся типом проводимости: КДБ 1 (100) и КЭФ 7,5 (100). Термическое окисление проводилось комбинированным методом сухой O_2 -влажный O_2 -сухой O_2 при температуре $1000^\circ C$, толщина слоев сформированных окислов составляла $0,12 \text{ мкм}$. После термического окисления на поверхность SiO_2 был нанесен алюминий. Для получения контактных площадок в слое Al была проведена фотолитография, после чего сформированные структуры отжигались при температуре $450^\circ C$ в течение 10 мин в среде N_2 .

Экспериментальные структуры были подвергнуты воздействию гамма-излучения, в качестве источника которого использовался Co^{60} с энергией генерируемых гамма-квантов $1,3 \text{ МэВ}$ и мощностью $0,24 \text{ Мрад/ч}$. До и после облучения проводилось измерение высокочастотных ВФХ.

Результаты и их обсуждение

На рисунке схематично показаны характерные для экспериментальных структур ВФХ с наблюдаемым гистерезисом. На рис. также приведены кривые, характеризующие гистерезис ВФХ, которые представляют собой разность между прямой и обратной

ветвями. Необходимо отметить, что у необлученных образцов также наблюдался гистерезис ВФХ. Из рис. видно, что для необлученных МОП-структур на n-кремнии прямая ветвь ВФХ сдвинута левее обратной, а в случае МОП-структур на основе p-Si прямая ветвь сдвинута правее обратной.

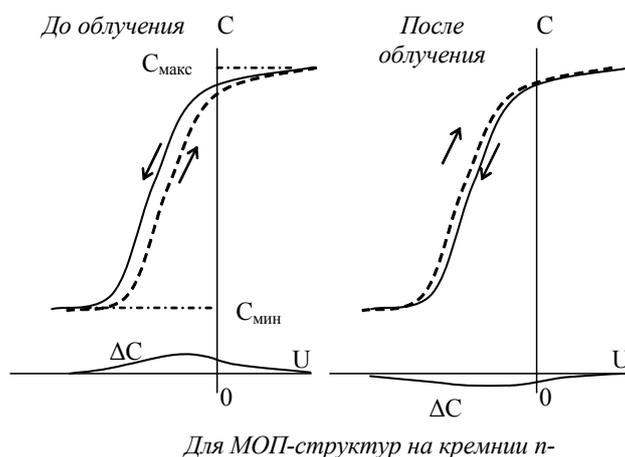
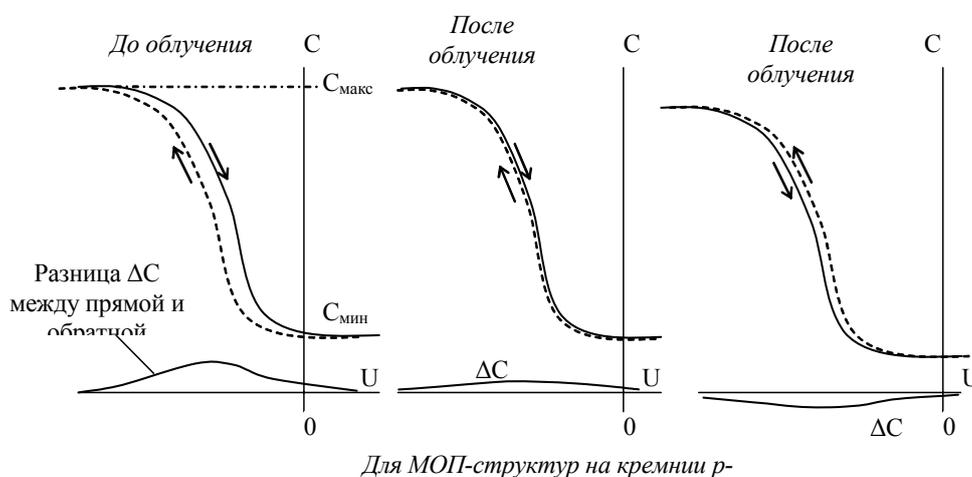


Рис. Характерные ВФХ экспериментальных МОП-структур до и после гамма-облучения

Под действием гамма-излучения происходит изменение ВФХ МОП-структур. При этом после облучения дозой 0,5 Мрад МОП-структур на основе p-Si наблюдается уменьшение сдвига между прямой и обратной ветвями ВФХ. Увеличение дозы облучения МОП-структур на основе дырочного кремния до 25 Мрад приводит к изменению в расположении прямой и обратной ветвей ВФХ относительно друг друга. В то же время для структур на основе n-Si уже после облучения дозой 0,5 Мрад наблюдается изменение в расположении прямой и обратной ветвей относительно друг друга. Такая различная чувствительность гистерезиса к дозе облучения для МОП-структур на основе дырочного и электронного кремния обусловлена, по-видимому, различием в значениях удельного сопротивления подложки. Это связано с тем, что увеличение удельного сопротивления в кремнии приводит к увеличению влияния зарядов в диоксиде кремния на изгиб энергетических зон в подложке. Таким образом, заряд на ловушках, образованный при облучении и обуславливающий гистерезис ВФХ [1], вызывает наибольшие изменения гистерезиса ВФХ в МОП-структурах с наибольшим значением удельного сопротивления подложки.

Сдвиг обратной ветви ВФХ относительно прямой можно интерпретировать как появление дополнительного заряда в окисле, вызвавшего этот сдвиг. Направление и величина сдвига соответствуют знаку и значению этого заряда в окисле, соответственно.

Таким образом, знак заряда, вызывающего появление гистерезиса, в случае ВФХ до облучения соответствует полярности основных носителей заряда, а в случае ВФХ, снятых после радиационной обработки, – полярности неосновных носителей заряда. Это обстоятельство позволяет сделать предположение о различной природе гистерезиса ВФХ, который наблюдается при измерении ВФХ МОП-структур до и после облучения.

Объяснить гистерезис ВФХ, вызванный гамма-излучением, можно следующим образом. При облучении в пленках двуоксида кремния в приграничной к границе раздела кремний-окисел генерируются структурные дефекты [1]. Эти дефекты играют роль ловушечных центров захвата носителей заряда, энергетические уровни которых располагаются в диоксиде кремния в интервале, соответствующем интервалу энергий вблизи середины запрещенной зоны кремния. Такое энергетическое положение определяется способностью генерируемых ловушечных центров захватывать носители заряда обоих знаков.

Для объяснения механизма появления гистерезиса ВФХ, возникающего при облучении МОП-структур, была предложена следующая схема заряжения и разряжения ловушечных центров, ответственных за гистерезис. На начальном этапе измерения ВФХ МОП-структуры на основе n-Si, что соответствует режиму аккумуляции основных носителей заряда – электронов, ловушечные центры захватывают электроны вследствие их туннелирования из зоны проводимости (ЗП) кремния. В режиме инверсии ловушечные центры выбрасывают электроны в валентную зону (ВЗ) полупроводника, что равносильно захвату дырок из ВЗ. Поэтому при изменении направления развертки напряжения смещения обратная ветвь идет левее прямой. При достижении режима обогащения эти центры вновь захватывают электроны из ЗП. Подобная схема действует и для МОП-структуры на кремнии р-типа, однако в этом случае в режиме обогащения источником дырок служит ВЗ, а ЗП в режиме инверсии поставляет электроны.

Заключение

Таким образом, гистерезис ВФХ, который появляется при облучении МОП-структур, обусловлен появлением ловушечных центров в диоксиде кремния в приграничной к кремнию области. Ловушечные центры обладают способностью обмениваться носителями заряда с подложкой. При этом степень влияния заряда ловушечных центров на изгиб зон кремния определяется удельным сопротивлением кремния, а знак этого заряда определяется полярностью неосновных носителей заряда в подложке. Следует отметить, что гистерезис, который наблюдается до облучения, отличается по своим свойствам от «радиационного» гистерезиса, и для выявления механизма его образования необходимы дополнительные исследования.

Литература

1. Fleetwood D.M., Reber Jr. R.A., Riewe L.C., Winokur P.S. Thermally stimulated current in SiO₂. // *Microelectronics Reliability*. 1999. V. 39. P. 1323–1336.
2. Скворцов А.М., Соколов В.И., Халецкий Р.А., Шамарин П.А. Особенности радиационной деградации системы кремний-окисел, полученной в различных технологических условиях с участием хлора. // *Микроэлектроника и информатика – 2004*. 11-ая Всероссийская межвузовская научно-техническая конференция студентов и аспирантов: Тезисы докладов. М.: МИЭТ, 2004. С. 26.

ЭЛЕКТРОХИМИЧЕСКИЙ МЕТОД МИКРОСТРУКТУРИРОВАНИЯ МОНОКРИСТАЛЛОВ КРЕМНИЯ (КРАТКИЙ ОБЗОР)

В.В. Власов

Научный руководитель – доктор технических наук, профессор А.М. Скворцов

Рассматривается получение пористого кремния методом электрохимического травления в растворе плавиковой кислоты. Описаны химические процессы, происходящие во время травления, механизм образования пор и зависимость структуры и свойств полученного пористого слоя от условий, при которых проводилось травление

Введение

Химический элемент кремний (Si) – второй после кислорода по распространенности на Земле. Его содержание в земной коре оценивается в 27,6 мас. %. Важнейшее достоинство кремния заключается в том, что во многих его разновидностях – монокристаллической, поликристаллической или аморфной – проявляются его полупроводниковые свойства. В настоящее время монокристаллический кремний (mono-Si) представляет собой основной материал микроэлектронной техники. На базе mono-Si выпускаются разнообразные полупроводниковые приборы – от дискретных диодов и транзисторов до сверхсложных интегральных схем и процессоров.

Кроме того, широко используется высокая фоточувствительность кремния (изменение электропроводности при освещении), что позволяет преобразовывать световую энергию в электрическую. Этот эффект находит применение в кремниевых фотоприемниках и солнечных батареях. Однако обратный процесс, т.е. достаточно эффективное преобразование электрической энергии в видимый свет, в кремниевых устройствах осуществить пока не удастся. Последнее связано с особенностями электронных свойств кремния.

Электронные свойства кремния можно изменить посредством формирования на его основе наноструктур – пространственно разделенных кремниевых участков с минимальными размерами в несколько нанометров ($1 \text{ нм} = 10^{-9} \text{ м}$). В этом случае носители заряда (электроны и дырки) приобретают дополнительную энергию вследствие квантоворазмерного эффекта. Рассматриваемый эффект ведет к увеличению ширины запрещенной зоны наноструктуры по сравнению с монокристаллом этого материала.

Обычно технология создания наноструктур в виде так называемых квантовых ям, нитей и точек довольно сложна. Кроме того, полученные результаты пока не позволяют говорить о практическом использовании подобных кремниевых структур в светоизлучающих полупроводниковых устройствах. Однако еще в 1956 г. А. Улиром (A. Uhlig) [1] был получен так называемый пористый кремний (por-Si). Этот материал представлял собой mono-Si, в котором в результате электрохимического травления образовалось огромное количество мельчайших пор. Плотность пор в некоторых образцах была столь большой, что происходило их перекрытие, и непротравленные участки кремния имели вид кораллоподобной системы нитей переменного сечения (рис. 1). Минимальные размеры сечения кремниевых нитей и их изолированных участков (кластеров) в пористом слое составляют, по данным электронной микроскопии, единицы нанометров. Por-Si обычно формируется как слой на поверхности пластины mono-Si, что очень важно для использования в микроэлектронике.

В конце 1990 г. в науке о полупроводниках произошла настоящая сенсация. Доктор Л. Кэнхэм (L. Canham) из Великобритании [2] сообщил о наблюдении эффективной красно-оранжевой фотолюминесценции из por-Si при комнатной температуре; величина h составляла 1–10%. Практически одновременно о сходных результатах информировала группа исследователей из Франции. [1] Авторы этих работ прямо связывали наблюдаемую люминесценцию с квантоворазмерным эффектом в наноструктурах порис-

того слоя. Результаты заинтересовали специалистов, и вскоре было получено сообщение о наблюдении электролюминесценции (ЭЛ) в данном материале. При ЭЛ излучение света происходит в результате прохождения электрического тока в полупроводнике, приводящего к возбуждению электронов и дырок с последующей их рекомбинацией. Количество публикаций в ведущих физических научных журналах по por-Si (огромную долю их составляет проблема люминесценции por-Si) стало лавинообразно расти и в настоящее время приближается к тысяче статей в год.



Рис. 1. Структуры пор слоев пористого кремния, приготовленного в различных условиях [3]

Такой резкий подъем интереса к por-Si был обусловлен принципиальной возможностью применения por-Si в качестве излучающего в видимом диапазоне света материала оптоэлектроники. За короткий период 1991–1996 гг. были проведены исследования квантоворазмерного эффекта в por-Si . Первые работы в этой области относятся к 1991–1992 гг. [4]. Кроме этого, были сделаны первые попытки реализации светодиодов на por-Si , а также получены образцы por-Si с сине-зеленой фотолюминесценцией. К сожалению, все созданные лабораторные варианты светоизлучающих приборов обладали низким квантовым выходом и быстро (в течение нескольких минут) деградировали. Таким образом, наиболее актуальными стали задачи повышения внешнего квантового выхода и стабилизации люминесценции por-Si . Последнее позволило бы совершить качественный скачок в развитии полупроводниковой технологии.

Исследования люминесценции por-Si в период 1995–2002 гг. посвящены вопросам поиска путей стабилизации свечения por-Si за счет модификации исходного пористого слоя или за счет изменения условий формирования por-Si , а также исследованию механизмов деградации фото- и электролюминесценции [1].

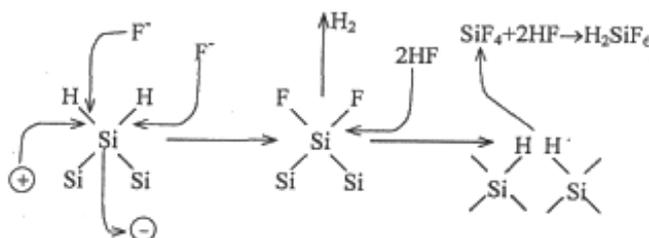
Получение пористого кремния

Первоначально por-Si предполагалось использовать в качестве изолятора активных приборов микроэлектроники. С этой точки зрения наиболее интересными аспектами por-Si были проводимость, морфология пор, контакт mono-Si – пористый слой и некоторые другие. Именно поэтому до середины 1990-х гг. было достаточно много работ, посвященных вопросам формирования por-Si . Проведенные исследования позволили выявить электрохимические реакции, протекающие в процессе образования пор, структуру por-Si и влияние условий травления на размер частиц в пористом слое. Особенно важным приложением результатов этих работ стало то, что теперь, задавая определенные режимы электрохимического травления кремния, можно получать необходимую структуру и размеры пор в образце [5].

Наиболее распространенным в современной литературе методом формирования por-Si является метод электрохимического травления пластины mono-Si в растворе, содержащем плавиковую кислоту. При этом порообразование наиболее эффективно происходит только при анодном травлении, когда на пластину mono-Si подается положи-

тельный потенциал относительно электролита, и порообразования практически не происходит в случае обратной полярности [6]. Для mono-Si p-типа проводимости еще одним дополнительным условием формирования por-Si является подсветка поверхности пластины Si для генерации свободных дырок в приповерхностном слое образца [5]. Вторым электродом (катодом) обычно служит платиновая пластина. При подходящем выборе плотности электрического тока на поверхности mono-Si происходит формирование пористого слоя. Относительно деталей механизма порообразования в литературе до сих пор нет единого мнения.

Наличие положительных носителей заряда является необходимым условием протекания следующей химической реакции растворения кремния:



В тех же условиях, что и при травлении кремния, но при большем приложенном потенциале, возможно протекание другой электрохимической реакции, отвечающей за его полировку, когда в результате образуется лишь матовая поверхность без формирования самого пористого слоя. Процесс электрополировки начинается лишь при превышении анодным потенциалом некоторого критического значения, величина которого зависит от концентрации молекул HF в травителе, типа проводимости пластины кремния, концентрации в ней свободных носителей и освещенности.

Предполагается, что при анодной обработке в электролитах, содержащих плавиковую кислоту, на поверхности электрода протекают следующие реакции:

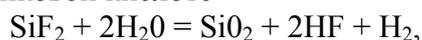
электрохимическая реакция образования бифторида кремния



химическое восстановление кремния из бифторида кремния



химическое окисление бифторида кремния до двуокиси кремния и его растворение в плавиковой кислоте



В зависимости от условий обработки одна из реакций (2 или 3) преобладает, в результате чего происходит либо образование por-Si (реакции 1 и 2), либо электрополировка (реакции 1 и 3) [7]. Варьируя в процессе травления величину анодного потенциала, можно переходить из области травления в область полировки. Данный прием применяется для образования свободных слоев por-Si, когда после образования пористого слоя через образец пропускается большой ток, в результате чего происходит локальная электрополировка на остриях образовавшихся пор и отсоединение пористого слоя от подложки. Подобные слои свободного por-Si применяют для исследования оптических спектров пропускания [8].

Однако полученные пленки неоднородны. Чтобы избавиться от этого недостатка, в приповерхностный слой монокристаллической кремниевой подложки имплантируют ионы кремния, фтора и хлора, т.е. ионы тех элементов, которые не обладают электрической активностью в кремнии. Результатом этого процесса является то, что травление происходит на всей поверхности пластины, и образующиеся пленки por-Si отличаются от анод-

ных пленок значительно большей однородностью. Но и эти пленки не лишены недостатка. Время травления в зависимости от дозы имплантации достигает 10 часов. Интенсификация процесса была достигнута применением более сложного травителя, составленного на основе плавиковой и азотной кислот с добавкой фтористого аммония [9]:

8 об. ч. (40% HF) + 5 об. ч. (65% HNO₃) + 9 об. ч. (40% (NH₄)₂F₆).

Время формирования пленок por-Si в этом травителе не превышает 15 мин. Выбранный состав травителя обеспечивает также высокую латеральную локальность формирования пленок por-Si.

Анализ реакций, протекающих при анодной обработке кремния в электролитах на основе плавиковой кислоты, показывает, что процесс образования пористого материала определяется, главным образом, двумя факторами [7]:

- процессом доставки ионов фтора в зону реакции и образованием бифторида кремния (фактор, связанный с электролитом и режимом анодной обработки);
- наличием подвижных носителей заряда положительного знака в приповерхностном слое кремниевого анода (фактор, связанный с электрофизическими свойствами кремния).

В связи с этим процессы анодной обработки кремния дырочного и электронного типов проводимости из-за различий в концентрации подвижных носителей положительного заряда, представленных дырками, будут существенно отличаться.

Дальнейшая интенсификация процесса формирования пленок наноструктурированного кремния обеспечивается травителем, в котором идет процесс формирования пористой структуры кремния без специального создания дефектов. Травитель представляет собой смесь концентрированных фторсодержащих и азотной кислот [10]. Фторсодержащие кислоты выбраны в качестве основного компонента травителя. Константа диссоциации этих кислот на несколько порядков выше, чем у плавиковой кислоты, что обеспечивает более полное насыщение «оборванных» химических связей на поверхности кристаллического кремния. Травление в этом растворе происходит следующим образом. На первой стадии азотная кислота стравливает естественный окисел кремния и затем вступает в реакцию с атомами на поверхности кремния. Результатом является оксид кремния, который стравливается фторсодержащей кислотой. Травление кремния происходит обычным образом, но с той лишь разницей, что концентрация молекул азотной кислоты подобрана так, чтобы молекулы находились друг от друга на достаточном расстоянии и окисление происходило не по всей плоскости пластины, а точечно. Полученные образцы por-Si характеризуются исключительно высокой стабильностью люминесцентных характеристик.

Механизм образования микроstructures пор при анодном травлении кремния заключается в следующем. В самом начале травления на дефектах и примесях происходит зарождение пор за счет локального увеличения потенциала в этом месте и, соответственно, большой скорости химической реакции. Далее в процессе роста пор продолжается выщелачивание примесей, но этот процесс не является определяющим в образовании самой структуры пор. Интересен тот факт, что основным направлением формирования пор в кремнии является направление $\langle 100 \rangle$ [11]. Просвечивающей электронной микроскопией, а также математическим моделированием данного процесса показано, что конечная структура пор близка к фрактальной. Этот результат логично вытекает из предположения простой диффузии свободных дырок в кремнии и их определяющего участия в реакции травления. Установлено, что толщина пленки [7] por-Si возрастает с увеличением продолжительности процесса анодной обработки. Вначале это соотношение линейно, причем наклон прямых увеличивается с ростом плотности анодного тока; при более длительной обработке линейность соотношения нарушается.

Скорость образования пленки [7 и 11] por-Si линейно растет с увеличением концентрации электролита. Температура электролита слабо влияет на скорость образования por-

Si. Основное влияние на скорость роста пленок por-Si оказывает плотность анодного тока: для кремния с дырочной проводимостью она прямо пропорциональна плотности тока, для кремния с электронной проводимостью зависимости сложнее. При слабой интенсивности освещения скорость роста мало зависит от плотности тока, при повышении интенсивности освещения наблюдается увеличение скорости роста. Эти зависимости объясняются, по-видимому, тем, что при слабой интенсивности освещения недостаток дырок ограничивает протекание анодной реакции (1), несмотря на увеличение плотности тока, вызывающее повышение концентрации ионов фтора в зоне реакции.

Структура пористого слоя определяется плотностью тока, концентрацией HF в электролите и характером легирования кремниевой подложки. Так, например, для кремния с электронным типом проводимости (n-Si) или сильно легированного дырочного кремния (p+-Si) поры имеют вид перпендикулярных поверхности каналов диаметром в десятки нанометров с более мелкими боковыми ответвлениями. Для образцов слабо легированного дырочного кремния (p-Si) или при освещении n-Si формируется структура в виде губки или коралла. Размеры пор и непротравленных участков при этом очень малы и составляют всего несколько нанометров.

На основании анализа экспериментальных данных по анодной обработке можно отметить области изменения параметров этого процесса и реально достижимые толщины и скорости роста пленок por-Si [7].

Концентрация плавиковой кислоты в электролите, %.	10–50
Плотность тока, мА/см ²	5–300
Температура электролита, К	293–343
Интенсивность освещения (при обработке кремния n-типа), мВт/см ²	20–60
Скорость роста пленок пористого кремния, мкм/с	0,01– 1
Толщина формируемых пленок пористого кремния, мкм	100–200

Заключение

В данной статье только упомянуто об одном из самых замечательных свойств пористого кремния – фото- и электролюминесценции, потому что описание этого явление требует более глубокого изучения как процессов люминесценции, так и свойств por-Si.

Очевидно, дальнейшее изучение электронных и атомных процессов в этом материале представляет несомненный интерес как с чисто научной, так и с прикладной точек зрения. Это обусловлено возможностью создавать слои с различными размерами наноструктур, легко изменять состав внешнего слоя кремниевого скелета, тем самым модифицируя характер указанных процессов. Результаты, полученные учеными при исследовании пористого кремния, позволяют построить детальную картину электроники квантовых нитей и точек и использовать разработанные модели для создания элементов наноэлектроники, которая, очевидно, будет определять развитие информационных систем XXI века. По-видимому, найдет практическое применение и такое свойство пористого кремния, как чувствительность его электронных свойств к окружающей среде. Появляется принципиальная возможность создания сенсоров на различные молекулы, причем для считывания данных удобно использовать люминесцентное излучение. Указанные приборы могут быть использованы в экологии, медицине и т.д.

Литература

1. Ligeon M., Muller F., Herino R. Analysis of the electroluminescence observed during the anodic oxidation of porous layers formed on lightly p-doped silicon. // J. Appl. Phys. 1993. V. 74. №2. P.1265–1271
2. Canham L.T. // Appl. Phys. Lett. 1990. 57. 1146.

3. Gullis A.G., Canham L.T. The structural and luminescence properties of porous silicon. // *Appl. Phys. Rev.* 1997. V. 82(3). № 1. P. 909–965.
4. Brandt M.S., Fuchs H.D., Stutzmann M., Weber J. and Cardonna M. // *Solid State Commun.* 1992. V. 81. P. 300–342.
5. Бучин Э.Ю., Проказников А.В. Управление морфологией пористого кремния n-типа. // *Письма в ЖТФ.* 1997. Т. 23. В. 6. С. 80–84.
6. Зимин С.П. Классификация электрических свойств пористого кремния. // *ФТП.* 2000. Т. 34. В. 3. С. 359–363
7. Лабунов В.А., Бондаренко В.П. Пористый кремний в полупроводниковой электронике. // *Зарубежная электронная техника.* 1978. № 15 (185).
8. Белогоров А.И., Белогорохова А.И., Караванский В.А. Инфракрасная спектроскопия и фотолюминесцентные свойства пленок пористого кремния: влияние режимов формирования. // *ФТП.* 1994. Т. 28. В. 8. С. 1424–1430.
9. Симченко А.В., Юзова В.А. / Конференция «Современные проблемы радиоэлектроники. Красноярский ГТУ. / <http://www.rtf.kgtu.runnet.ru>
10. Мордкович В.Н. «Кремний на изоляторе» – перспективный материал микроэлектроники. // *Материалы электронной техники.* 1998. В. 2. С. 4–7.
11. Smith R.L., Collins S.D. Porous silicon formation mechanisms. // *J. Appl. Phys.* 1992. V. 71. № 8. P. 1–22.

ПРИМЕНЕНИЕ РЕДКОЗЕМЕЛЬНЫХ ЭЛЕМЕНТОВ В КВАНТОВОЙ ЭЛЕКТРОНИКЕ

И.И. Стройков

Научный руководитель – доктор технических наук, профессор А.М. Скворцов

Проведен обзор использования редкоземельных элементов в современной квантовой электронике. Рассмотрены вопросы поглощения энергии атомами эрбия и последующего вынужденного излучения энергии в эрбиевых волоконных усилителях. Приведены наиболее перспективные направления исследований применения эрбия и иттербия, связанных, прежде всего, с передачей информации и лазерной техникой.

Введение

Полупроводниковые материалы, легированные редкоземельными элементами, в частности эрбием и иттербием, представляют большой интерес. На сегодняшний день существует несколько наиболее перспективных направлений исследований использования этих элементов в электронной технике. Свойства редкоземельных материалов дают возможность создать эффективные излучатели на длине волны 1.54 мкм, лежащей в области максимальной прозрачности кварцевых волоконных световодов. В данной статье будет рассмотрено применение эрбия и иттербия в лазерной и оптоволоконной технике.

Основная часть

Хорошо известно, что через каждые 50–100 км волоконно-оптического тракта происходит ослабление оптического сигнала на 10–20 дБ, что требует его восстановления. В настоящее время для волоконно-оптических систем связи разработаны три типа оптических усилителей: полупроводниковые оптические усилители, эрбиевые волоконные усилители и рамановские волоконные усилители.

Самым распространенным в настоящее время являются эрбиевые волоконные усилители. Главным образом это определяется спектром люминесценции ионов эрбия, лежащим в области длин волн $\lambda=1.54$ мкм – области минимальных потерь современных кварцевых световодов.

Эффективность оптоэлектронных приборов на основе полупроводниковой матрицы, легированной эрбием, будет определяться, в первую очередь, эффективностью передачи энергии от носителей заряда матрицы во внутреннюю 4f-оболочку иона эрбия. Такая передача энергии в полупроводнике происходит за счет кулоновского взаимодействия между свободными носителями заряда (электронами и дырками) и сильно локализованными 4f-электронами, основное состояние которых находится ниже валентной зоны примерно на 10 эВ [1].

Эрбиевый волоконный усилитель характеризуется следующими основными параметрами:

- коэффициентом линейного усиления (усиления при малом входном сигнале);
- мощностью насыщения;
- спектральной полосой насыщения;
- рабочей длиной волны;
- эффективностью оптического преобразования и мощностью накачки.

Оптическая накачка эрбиевых волоконных усилителей осуществляется, как правило, в высокоэффективные полосы поглощения эрбия на длинах волн $\lambda=980$ нм и $\lambda=1480$ нм. Для накачки используются полупроводниковые лазеры, излучающие на соответствующих длинах волн мощности порядка нескольких ватт. При этом эффективность оптического преобразования может достигнуть 50–60% [2].

Принцип работы эрбиевых волоконных усилителей основан на явлении усиления света при вынужденном излучении. Это то же явление, которое обеспечивает возникновение генерации в лазерах. Возможность усиления света в световодах, легированных ионами эрбия, обуславливается схемой уровней энергии данного редкоземельного элемента, представленной в упрощенном виде на рис. 1.

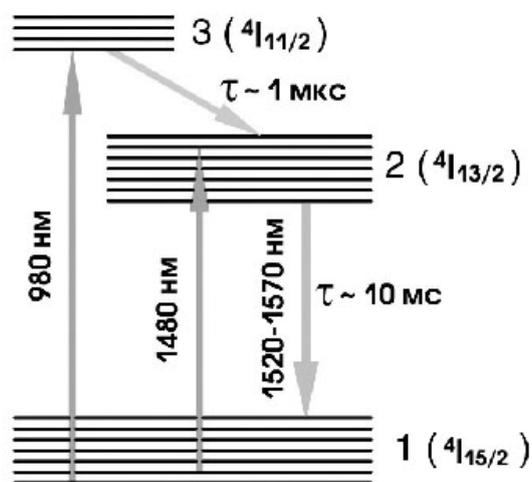


Рис. 1. Упрощенная схема уровней энергии ионов эрбия (Er^{3+}) в кварцевом стекле [3]

Усиление света в эрбиевом усилителе происходит благодаря переходу между уровнями 2–1 (${}^4I_{13/2}$ – ${}^4I_{15/2}$). Каждый из этих уровней расщеплен на ряд подуровней из-за взаимодействия ионов эрбия с внутрикристаллическим полем кварцевого стекла. Под действием накачки за счет поглощения фотонов накачки ионы эрбия переходят из основного состояния (уровень 1) в верхнее возбужденное состояние (уровень 3), которое является короткоживущим (время жизни $\tau_3 = 1$ мкс), и за счет процессов релаксации переходят в долгоживущее состояние (на метастабильный уровень 2). Термин «метастабильный» означает, что время пребывания иона эрбия на этом уровне энергии относительно велико. Поэтому число ионов, находящихся на уровне 2, при соответствующей мощности накачки может превышать число ионов на уровне 1. Уровень 1 называется основным состоянием, так как в отсутствие накачки практически все ионы эрбия находятся на этом энергетическом уровне. Доля частиц, находящихся на остальных уровнях, в отсутствие накачки мала.

Число ионов в единице объема, находящихся на некотором уровне энергии, называется населенностью этого уровня энергии. В отсутствие накачки населенность основного уровня энергии вещества максимальна, населенности всех остальных уровней энергии быстро уменьшаются с увеличением энергии уровня. Состояние среды, при котором населенность некоторого более высокого уровня энергии иона превышает населенность некоторого нижележащего уровня, является очень необычным и получило название состояния с инверсией населенностей.

Если в среду с инверсией населенности попадает излучение с энергией фотона, совпадающей с энергией перехода из метастабильного состояния в основное, то с большой вероятностью происходит переход иона с метастабильного уровня 2 на основной уровень 1 с одновременным рождением еще одного фотона. Увеличение числа фотонов при их взаимодействии с ионами эрбия означает, что происходит усиление света, распространяющегося в среде с инверсией населенности.

Не все ионы эрбия находятся в метастабильном состоянии и обеспечивают усиление. Часть ионов находится на уровне 1, и эти ионы, взаимодействуя с фотонами, энергия которых совпадает с энергией перехода, эффективно их поглощают, переходя на

уровень 2. При этом спектр усиления ионов эрбия практически совпадает со спектром поглощения. Если количество ионов, находящихся на уровне 2, меньше числа ионов, находящихся на основном уровне 1, то наблюдается поглощение. Именно поэтому необходимым условием усиления света является создание инверсии населенностей между двумя рабочими уровнями энергии 2 и 1. Для создания инверсии населенностей в эрбиевом усилителе необходимо перевести примерно половину ионов эрбия на метастабильный уровень 2. Мощность накачки оптического усилителя, при которой населенность уровней 1 и 2 равны, называется пороговой мощностью.

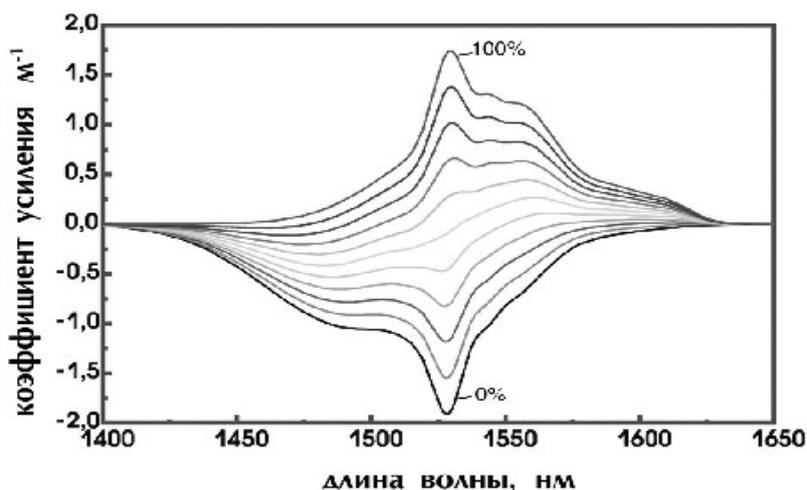


Рис. 2. Спектральная зависимость усиления поглощения эрбиевого волокна при разных значениях относительной населенности метастабильного уровня энергии. Нижняя кривая – населенность 0%, верхняя кривая – населенность 100%. Кривые проведены для населенности, изменяющейся с шагом 10% [3]

При мощности накачки ниже пороговой наблюдается не усиление, а поглощение светового сигнала. На рис. 2 представлены спектры поглощения/усиления при различных значениях относительной населенности уровня 2, определяемой уровнем мощности накачки. Нижняя кривая, наблюдающаяся в отсутствии накачки (все частицы находятся в основном состоянии, населенность уровня 2 равно 0%), соответствует «отрицательному усилению», т.е. поглощению во всем рабочем спектральном диапазоне. По мере увеличения мощности накачки все большее число активных ионов переходит в возбужденное состояние. Это приводит, как видно из рисунка, сначала к уменьшению коэффициента поглощения, а затем к усилению света. Отметим также, что спектр усиления несколько сдвинут в длинноволновую область относительно спектра поглощения. Следовательно, для усиления в длинноволновой части спектра требуется меньшее значение инверсии.

При отсутствии усиливаемого сигнала ионы эрбия переходят в основное состояние самопроизвольно, излучая фотоны с энергией, соответствующей данному переходу, т.е. появляется спонтанное излучение.

В рабочем режиме при наличии усиливаемого сигнала часть возбужденных ионов переходит в основное состояние спонтанно, приводя к появлению усиленного спонтанного излучения. Усиленное спонтанное излучение является основным источником шумов, а также ограничивает коэффициент усиления, особенно в случае слабого сигнала.

Изготовление эрбиевых световодов производится теми же методами, что и световодов для передачи информации, с добавлением промежуточной операции пропитки непроплавленного материала сердцевины раствором солей эрбия либо операции легирования ионами эрбия из газовой фазы непосредственно в процессе осаждения сердцевины.

Принципиальным является выбор легирующих добавок, формирующих сердцевину активного световода, а также подбор концентрации ионов эрбия. Различные добавки в кварцевое стекло изменяют характер штарковского расщепления уровней энергии ионов эрбия. В свою очередь, это приводит к изменению спектров поглощения и излучения. На рис. 3 представлены спектры излучения ионов эрбия в кварцевом стекле, легированном наиболее часто применяемыми в технологии волоконных световодов добавками.

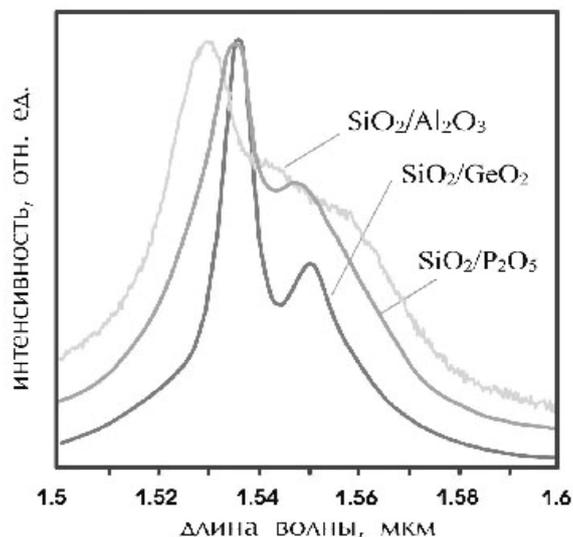


Рис. 3. Спектры излучения ионов эрбия в кварцевом стекле с различными добавками [3]

Из представленных данных видно, что наиболее широкий спектр излучения (а, значит, и спектр усиления), составляющий около 40 нм по полувысоте, достигается при использовании в качестве добавки алюминия.

Концентрация ионов эрбия в сердцевине оптического волокна определяет длину волновода, используемую в усилителе при заданных уровнях сигнала и накачки. На практике она составляет 10^{18} – 10^{19} см⁻³, что обеспечивает длину используемого активного световода от нескольких единиц до нескольких десятков метров [3].

Чтобы получить структуру с высокой концентрацией ионов эрбия и одновременно с высокой квантовой эффективностью усиления, необходимо каким-то образом преодолеть термодинамические ограничения на взаимную растворимость окислов. Выход из положения был найден благодаря применению плазмохимии для синтеза стекол. В этом методе молекулы окислов образуются в газовой фазе в результате плазмохимических реакций, а стекло формируется осаждением молекул из газовой фазы на сравнительно холодную твердую поверхность, так что возможность перемещения ионов эрбия в объеме стекла затруднена. При отсутствии жидкой фазы затруднена возможность образования кластеров, наличие которых соответствует минимуму свободной энергии системы и поэтому является неизбежным в термодинамически равновесных условиях синтеза [4].

В связи с тем, что интенсивность люминесценции зависит от концентрации ионов редкоземельного элемента, представляется интересным совместная имплантация ионов эрбия и электрически неактивных примесей (С, О, N, Р), приводящая к увеличению интенсивности фотолюминесценции [5].

Установлено, что эрбиевая люминесценция значительно увеличивается, если одновременно с эрбием в кремний имплантируется кислород с концентрацией, превышающей концентрацию эрбия на порядок. В этом случае образуется комплекс эрбия, окруженного кислородом, которому соответствует донорный уровень с энергией связи $E_D = 150$ – 200 МэВ.

При поглощении света (в относительно слабо легированном эрбием (10^{17} см^{-3}) и кислородом кремнии) образуются свободные экситоны, время жизни которых в кристаллическом кремнии порядка 10 мкс. Свободный экситон имеет большую вероятность быть захваченным на нейтральный донорный центр, образованный эрбиевым комплексом. Оже-возбуждение 4f-оболочки иона эрбия Er^{3+} происходит при рекомбинации связанного экситона; избыток энергии передается электрону, уже находящемуся на нейтральном донорном центре, в результате чего он выбрасывается в зону проводимости (рис. 4, а, где FE – свободный экситон, BE – связанный экситон, D_{Er} – донорный уровень эрбиевого комплекса).

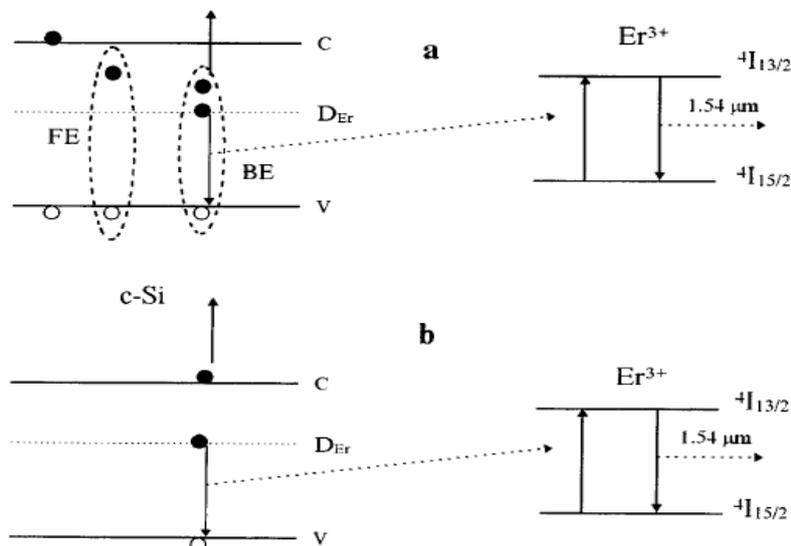


Рис. 4. а – схема экситонного механизма возбуждения иона эрбия. FE – свободный экситон, BE – связанный экситон, D_{Er} – донорный уровень эрбиевого комплекса; б – схема оже-процесса возбуждения иона эрбия при рекомбинации электрона, находящегося на донорном уровне эрбиевого комплекса, с дыркой из валентной зоны. Избыточная энергия передается либо третьему телу (на схеме – электрону из зоны проводимости), либо локальным фононам эрбиевого комплекса [1]

В более сильном легированном эрбием ($> 10^{18} \text{ см}^{-3}$) и кислородом кремнии возбуждение эрбия происходит при оже-рекомбинации электрона, захваченного на донорный уровень эрбиевого комплекса, с дыркой, находящейся в валентной зоне. В этом оже-процессе избыточная энергия передается локальным фотонам эрбиевого центра или третьему телу (электрону или дырке), если оно находится на небольшом расстоянии от эрбиевого центра (рис. 4, б).

Основным препятствием к созданию высокоэффективных светодиодов на длину волны $1.54 \mu\text{m}$ на основе кремния, легированного эрбием, является температурное гашение люминесценции. Оно может быть обусловлено как ослаблением эффективности возбуждения, так и безызлучательным девозбуждением ионов эрбия. Различие между этими процессами может быть выявлено по исследованию температурной зависимости времени жизни иона эрбия в возбужденном состоянии: в первом случае время жизни не будет зависеть от температуры, во втором оно должно уменьшаться с ростом температуры.

Основной процесс девозбуждения иона эрбия – это передача энергии 4f – электрона, находящегося на возбужденном уровне, свободному носителю тока (рис. 5, а). Этот процесс не имеет энергии активации и должен приводить к температурно-независимому девозбуждению.

Девозбуждение иона эрбия может происходить за счет рождения электронно-дырочной пары (электрон может рождаться как на донорном уровне эрбиевого комплекса, так и в зоне проводимости) (рис. 5, б). Этот процесс обратен оже-рекомбинации

эрбия при рекомбинации электрона, связанного на донорном уровне эрбия, с дыркой валентной зоны [1].

Также хотелось бы отметить, что иттербий-эрбиевые волоконные усилители имеют ряд достоинств по сравнению с обычными эрбиевыми усилителями. Концентрация ионов эрбия в жиле Yb-Er световода может быть порядка 10^{19} см^{-3} и выше без заметного снижения энергетической эффективности работы усилителя. Кроме того, иттербий за счет безызлучательного переноса возбуждения эффективно передает энергию эрбию. Ионы иттербия имеют широкую линию поглощения от 850 до 1100 нм. Для накачки таких усилителей используются многомодовые полупроводниковые лазерные диоды (InGaAs) с выходной мощностью более 1В.

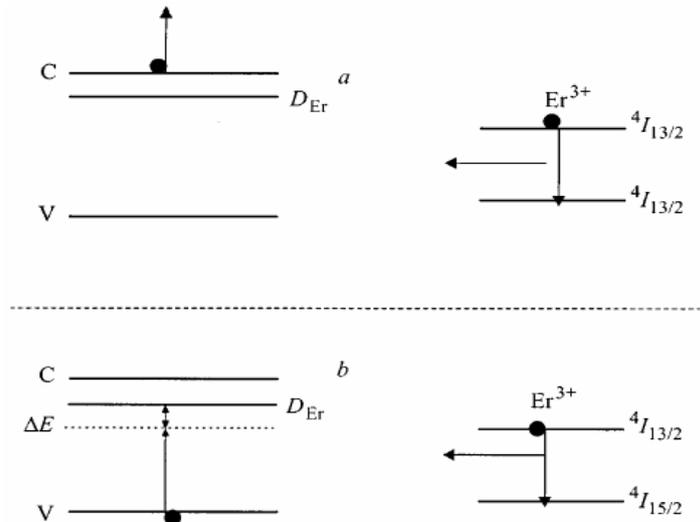


Рис. 5. а – схема девозбуждения иона эрбия при его взаимодействии со свободным носителем тока (электроном из зоны проводимости); б – схема девозбуждения иона эрбия при захвате электрона из валентной зоны на донорный уровень эрбиевого комплекса, ΔE – энергия, заимствуемая от решетки и определяющая энергию активации процесса девозбуждения [1]

Другой, наиболее перспективный путь применения редкоземельных элементов – это непосредственно использование их в лазерах. Лазеры на эрбиевом стекле благодаря своей способности излучать в безопасном для зрения спектральном диапазоне (1.54 мкм) находят все более широкое применение.

Импульсно-периодический режим работы лазера предполагает возможность реализации высокой мощности генерации и, как следствие, значительное тепловыделение в активном элементе, которое приводит к ряду негативных последствий, например, к возникновению наведенной тепловой линзы, а также к механическим напряжениям, вызывающим ограничение предельной мощности накачки из-за разрушения активного элемента. Это заставляет искать компромисс при выборе концентраций активаторов.

Основной сенсбилизатор люминесценции ионов эрбия – ионы иттербия. Для полного излучения лампы накачки в стекло вводят ионы хрома, которые имеют широкие полосы поглощения с максимумами на 450 и 660 нм (рис. 6). С учетом этого был разработано хром-иттербий-эрбиевое стекло ЛГС-ХЧ с пониженной концентрацией ионов хрома ($7 \cdot 10^{18} \text{ см}^{-3}$), предназначенное для использования в импульсно-периодических лазерных системах. Термооптическая константа стекла W была минимизирована и составляла $1.2 \cdot 10^{-6} \text{ K}^{-1}$. Для компенсации потерь из-за снижения концентрации ионов хрома была увеличена концентрация ионов иттербия. Спектр поглощения стекла при концентрации ионов эрбия 10^{19} см^{-3} представлен на рис. 6 [6].

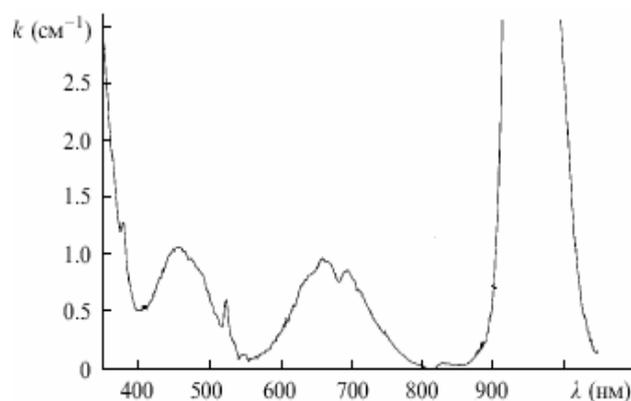


Рис. 6. Спектр поглощения стекла ЛГС-ХЧ [6]

Также необходимо отметить, что возможно формирование режима стабильного периодического излучения волоконных лазеров, а именно, что колебания волоконного иттербий-эрбиевого лазера могут быть синхронизированы внешним импульсно-периодическим излучением полупроводникового лазера, длина волны которого близка к волне излучения волоконного лазера в режиме свободной генерации [7].

Заключение

В работе на примере эрбия и иттербия были рассмотрены наиболее интересные и перспективные направления в использовании редкоземельных элементов в квантовой электронике. Выделены два направления – применение Er и Yb для усиления оптических сигналов в волоконных усилителях, а также изготовление стекол на основе эрбия и иттербия для лазеров. Из приведенного материала видно, что применение эрбия основано на его особенности иметь спектр люминесценции ионов в области длин волн $\lambda=1.54$ мкм. Использование же иттербия вместе с эрбием объясняется его свойством передавать энергию эрбию за счет безызлучательного переноса возбуждения.

Литература

1. Бреслер М.С., Грегоркевич Т., Гусев О.Б., Соболев Н.А., Теруков Е.И., Захарченя Б.П., Ясиевич И.Н. Механизмы возбуждения и температурного гашения люминесценции ионов эрбия в кристаллическом и аморфном кремнии. // ФТТ. 1999. Т. 41. В. 5. С. 851–855.
2. Егоров Ф.А., Потапов В.Т. Усилители оптических сигналов в ВОЛС. // Фотон–Экспресс. 2000. № 21. С. 9–12.
3. Куков А.С., Наний О.Е. Эрбиевые волоконно-оптические усилители. // Lightwave RE. 2003. № 1. С. 14–19.
4. Голант К.М. Плазмохимическая технология решает новые задачи волоконной оптики и оптоэлектроники. // Фотон–Экспресс. 2004. № 6. С. 41–42.
5. Александров О.В., Захарьин А.О., Николаев Ю.А., Соболев Н.А. Влияние электрически неактивных примесей на образование донорных центров в слоях кремния, имплантированных эрбием. // ФТТ. 2000. Т 34. В. 5. С. 526–529.
6. Азынев А.А., Бышевская-Конопко Л.О., Воробьев И.Л., Садовский П.И., Сергеев С.Н. Оптимизация параметров накачки импульсно-периодического эрбиевого лазера. // Квантовая электроника. 2001. № 10. С. 861–863.
7. Белотицкий В.И., Петров М.П. Кольцевой волоконный лазер с синхронизированными релаксационными колебаниями. // ЖТФ, 1999. Т 25. № 2. С. 24–28.

ВЛИЯНИЕ ЛАЗЕРНОГО МИКРОСТРУКТУРИРОВАНИЯ ПОВЕРХНОСТИ КРЕМНИЯ НА МОРФОЛОГИИ И ОПТИЧЕСКИЕ СВОЙСТВА ПОЛУЧЕННЫХ СТРУКТУР

С.С. Дышловенко, М.В. Погумирский, Фам Куанг Тунг

Научный руководитель – доктор технических наук, профессор А.М. Скворцов

Введение

В настоящее время процесс лазерного микроstructuring поверхности широко применяется в полупроводниковой микроэлектронике. Изменяя энергию лазерного импульса и время воздействия, можно реализовать широкий диапазон режимов воздействия – от локального нагрева до строго дозированного удаления материала. Из основных способов лазерного микроstructuring поверхности в настоящей работе используется способ, основанный на локальном нагревании и испарении вещества, который позволяет модифицировать структуру поверхности кремния и его свойства.

Методика эксперимента

В работе используется CO_2 -лазер с длиной волны в инфракрасном диапазоне $\lambda=10,6$ мкм; длительность импульса $\tau=0,88 \times 10^{-4}$ с, частота следования импульсов $f=80$ Гц, мощность лазерного излучения $P=0,3$ Вт; скорость сканирования $V=2,5$ мм/с. В качестве образцов использовались пластины монокристаллического кремния толщины 0,38 мм с ориентацией (100). Для активирования поглощения кремнием лазерного излучения пластина кремния была помещена на стеклянную подложку. Процесс сканирования производился в воздухе в разных направлениях относительно базового среза пластины: параллельно, перпендикулярно и под углом 45° .

Относительное перемещение лазерного пучка по поверхности обрабатываемого изделия схематически представлено на рис. 1. Из рисунка видно, что размер облученной зоны d зависит от ее расположения относительно центра рабочего поля y . Увеличение размера облученной зоны при отклонении пучка происходит по двум причинам: вследствие расфокусировки пучка и вследствие наклонного нападения пучка на обрабатываемую поверхность.

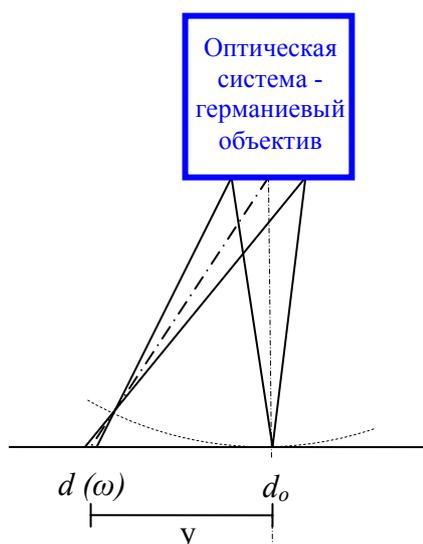


Рис. 1. Изменение облученной области при сканировании пучка

Пропускательная способность T является одним из важных оптических свойств материала. Зная T , можно вычислить другие оптические характеристики, такие как поглощательная способность, коэффициент преломления и т.д. Для измерения пропускательной способности кремния в инфракрасной области спектра используют спектрофотометр ИКС-29. Для оценки изменения пропускательной способности полученных образцов в работе были проведены измерения и сравнения пропускательной способности для неокисленного и окисленного образцов до и после облучения.

Результаты исследований и их обсуждение

На рис. 2а–2г представлен вид структур, взятых из различных частей обработанной зоны, которые схематически изображены на рис. 2д. Фотографии выполнены на растровом электронном микроскопе. В этих структурах (рис. 2а–2б) можно выделить три части: центральная часть в виде выпуклости, которую окружают волнообразные структуры (вторая часть структуры), и третья часть, имеющая вид светлого кольца. Структуры, взятые из периферийных областей, имеют менее сложное строение, в них отсутствует центральная выпуклая часть (рис. 2в–2г).

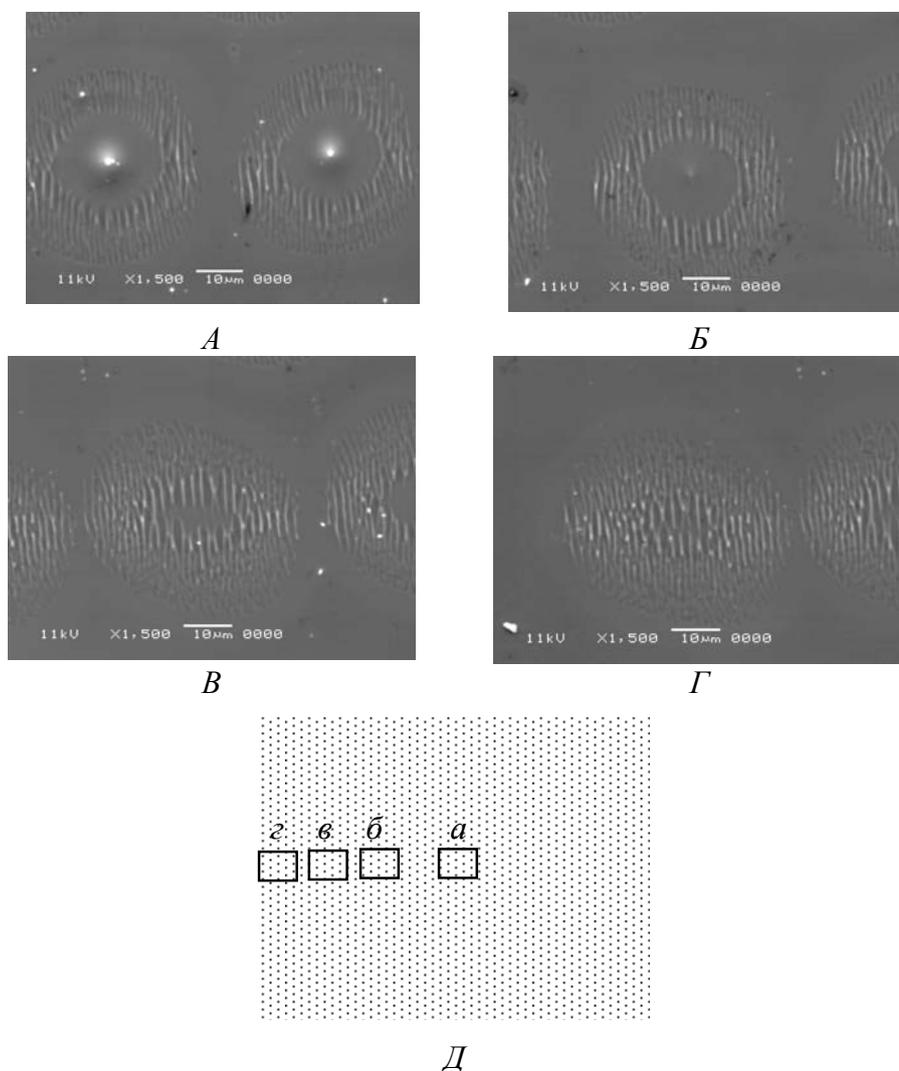


Рис. 2. Фотографии структур различной морфологии, образованных при лазерном воздействии в различных частях области обработки (а–г); схематическое изображение области обработки и расположения структур а–г (д)

Поскольку кремний прозрачен для длин волн в инфракрасном диапазоне, он не поглощает излучение CO₂-лазера. Этим объясняется тот факт, что на образцах монокристаллического кремния, облученного CO₂-лазером, не было обнаружено каких-либо структурных изменений. Поэтому для активации поглощения и взаимодействия кремния с лазерным излучением под пластины образцов кремния были помещены стеклянные пластины, хорошо поглощающие излучение CO₂-лазера. При облучении лазерным пучком $\lambda=10,6$ мкм кремний пропускает лазерное излучение. Стекло под кремниевой пластиной поглощает излучение, нагревается и частично испаряется. За счет процесса теплообмена пластина кремния также нагревается, ее температура повышается. Вследствие этого изменяются оптические свойства кремния: его коэффициент поглощения повышается. Монокристалл кремния начинает поглощать лазерное излучение. Феномен изменения оптических свойств кремния в данном случае может быть объяснен двумя причинами.

1) При повышении температуры за счет термической энергии ширина запрещенной зоны кремния уменьшается. Это приводит к тому, что электроны на низких уровнях энергии поглощают излучение CO₂-лазера. В результате активируется внутрizonное поглощение и увеличивается скорость термической генерации свободных носителей, особенно в зоне контакта кремниевой пластины со стеклом, так что в итоге процесс поглощения нарастает лавинообразно.

2) Свободные горячие электроны с высокой кинетической энергией, генерируемые на обратной стороне кремниевой пластины, диффундируют в верхние области пластины и обеспечивают дальнейшее увеличение концентрации свободных носителей в облученной зоне пластины. Эти свободные электроны сталкиваются с атомами и приводят к возбуждению электронов и дырок, которые содействуют дальнейшему поглощению лазерного излучения [1, 2].

Таким образом, энергия лазерного излучения поглощается кремнием и за счет повышения коэффициента поглощения кремниевой пластины преобразуется в тепловую энергию. При температуре порядка 800°C и выше начинается пластическая деформация кремния. Когда температура поверхности достигает $T=1420^\circ\text{C}$, начинается плавление материала. В дальнейшем, при охлаждении, происходит рекристаллизация кремния и образование сложных структур.

В центре области сканирования лазерный пучок сфокусирован, что обеспечивает наибольшую плотность мощности в области обработки (см. рис. 1). Чем больше плотность мощности, тем больше температура, тем больше площадь расплава и соответственно – площадь рекристаллизованного кремния. С отклонением от центра области обработки увеличивается расфокусировка лазерного пучка, что производит к уменьшению плотности мощности лазерного излучения в пятне. В результате уменьшается площадь расплава и, соответственно, площадь рекристаллизованного кремния.

Исследование этих структур на растровом электронном микроскопе показало, что они имеют сложную морфологию и состоят из нескольких частей (рис. 2). Возникновение подобных сложных структур может быть связано с неравномерным, псевдогаусовым распределением энергии в самом лазерном пятне.

Наиболее сложные по строению структуры наблюдаются в центральной зоне области обработки (рис. 2а). Они состоят из следующих частей. Центральная часть структуры, где энергия лазерного излучения и температура поверхности максимальны, представляет собой рекристаллизованную область расплава. Далее следует деформированная часть структуры, которая, по-видимому, имеет дислокационную структуру (результат пластической деформации, возникающей при более низкой температуре). И, наконец, третья часть, обнаруженная только на электронном микроскопе, имеет вид светлого кольца и представляет собой монокристалл кремния с деформированной решеткой.

С отклонением пучка от центра облученной области и, следовательно, его расфокусировкой снижается плотность мощности и температура в центре лазерного пятна, что приводит к уменьшению площади расплава (рис. 2б). По мере удаления от центра области обработки плотности мощности уменьшается настолько, что температура не превышает температуры плавления, и центральная часть полностью исчезает, остаются только зоны с дислокационной структурой и с деформированной решеткой (рис. 2в–г). Графики зависимости пропускательной способности T от длин волн λ неокисленного и окисленного образцов представлены на рис. 3, 4. Для сравнения на каждом графике представляются две кривых зависимости пропускательной способности от длины волн: первая – для необлученного и вторая – для облученного лазером образцов.

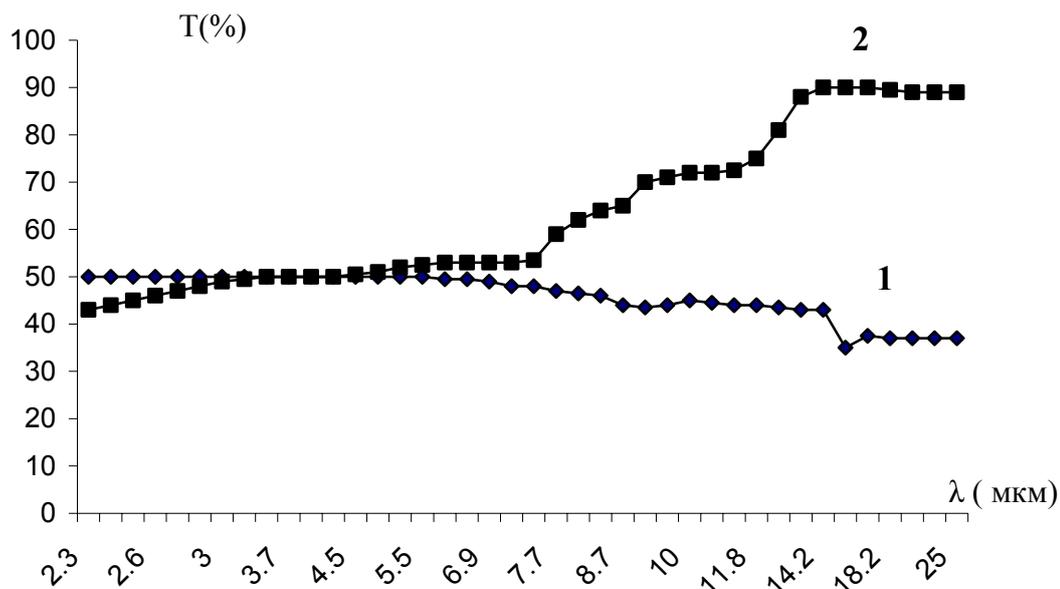


Рис. 3. График зависимости пропускательной способности T чистого монокристаллического кремния (1) и неокисленного облученного лазером образца (2) от длин волн λ в ИК-диапазоне

Чистый кремний можно считать полупрозрачной средой для длин волн в дальней ИК-области: он пропускает порядка 40–50% потока излучения. По мере увеличения длин волн в дальней ИК-области пропускательная способность кремния постепенно снижается, но незначительно: во всем диапазоне она уменьшается приблизительно на 10%. Это хорошо видно на графике (рис. 3).

Другая картина наблюдается при изменении пропускательной способности образца чистого монокристаллического кремния с лазерным микроstructuringированием поверхности. На первом участке дальней ИК-области ($\lambda=2-8$ мкм) этот образец также является полупрозрачным (пропускает 40–50% потока излучения), но его пропускательная способность постепенно возрастает (из расчета 1% на 1 мкм увеличения длины волн). При длине волны 8 мкм пропускательная способность начинает резко увеличиваться. Во втором диапазоне дальней ИК-области ($\lambda=10-25$ мкм) образец является почти прозрачным. Его пропускательная способность поднимается до 70% при $\lambda=10$ мкм и достигает 90% при $\lambda>14,2$ мкм. Другими словами, после лазерного микроstructuringирования поверхности кремния его пропускательная способности для длин волн в дальней ИК-области значительно изменяется: она может быть на 10–50% больше.

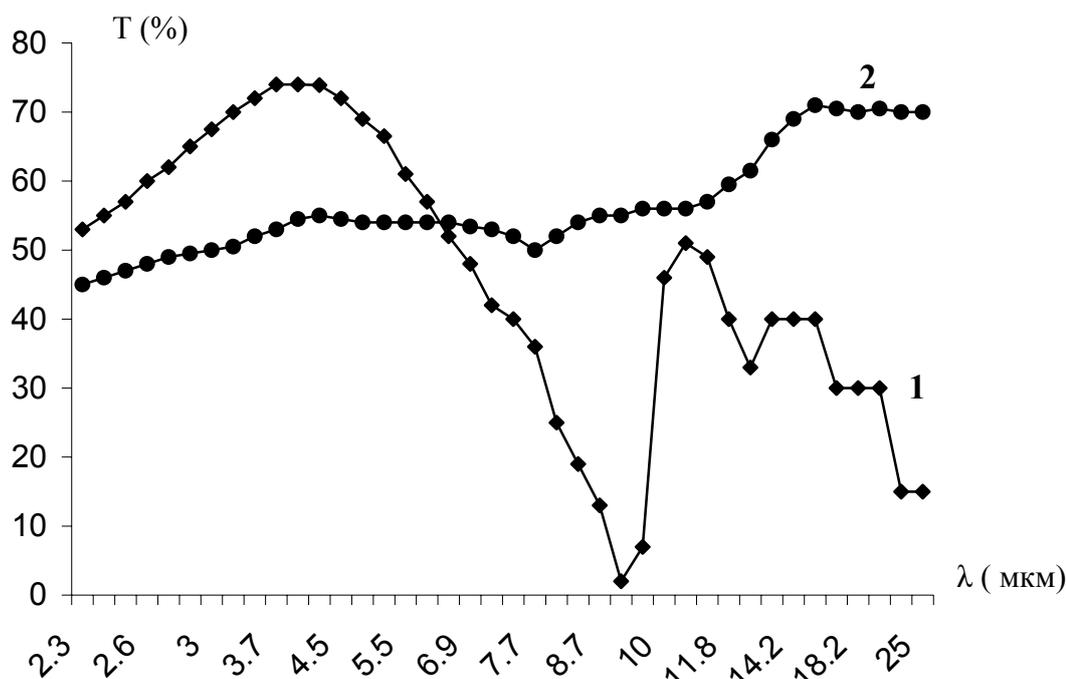


Рис. 4. График зависимости пропускательной способности T монокристаллического кремния с окислом SiO_2 (1) и окисленного образца, облученного лазером (2), от длин волн λ в ИК-диапазоне

На рис. 4 представлен график зависимости пропускательной способности T монокристаллического кремния с окислом SiO_2 и образца 2 – кремний с окислом SiO_2 , облученный лазером, от длин волны λ в ИК-диапазоне. За счет изменения показателя преломления тонкого слоя окисла SiO_2 от длины волны его график зависимости пропускательной способности от длины волны имеет вид затухающего колебания [3, 4].

После облучения лазером окисленного образца его пропускательная способность T очень сильно изменилась. Наблюдается уменьшение зависимости оптических параметров тонкого слоя окисла образца от длин волн. Вместо линии затухающего колебания наблюдается монотонный рост пропускательной способности, который достигает значения 70% в дальнем ИК-диапазоне (до 28 мкм).

Заключение

С помощью лазерной обработки (CO_2 -лазера с длиной волны $\lambda=10,6$ мкм) получены микроструктуры с разной мощностью облучения по площади сканирования. На этих образцах проведены исследования топологии, морфологии, оптических и электрофизических свойств.

Были обнаружены следующие особенности микроструктуры.

- Свойства микроструктурных областей не зависят от направления сканирования лазерного пучка относительно главных кристаллографических плоскостей кристаллической решетки кремния.
- Структуры облученных областей зависят от расположений этих областей на площади области сканирования.
- Область облучения состоит из следующих частей: центральная часть – рекристаллизованная область расплава; далее – сильно деформированная часть; третья – деформированная часть, обнаруженная только на электронном микроскопе – представляет собой монокристалл кремния с деформированной решеткой.

- Пропускательная способность облученной области существенно изменяется по отношению к исходному монокристаллическому кремнию.
- Наблюдается уменьшение зависимости пропускательной способности участков облученных лазером окисленных образцов от длины волн ИК-излучения, в отличие от обычного окисленного кремния.

Литература

1. Wang W.J., Lu Y.F., An C.W., Hong M.H., Chong T.C. Controllable periodic structure on silicon wafer by CO₂-laser irradiation. // Applied Surface Science. 2002. V. 186. P. 594–598 (www.elsevier.com/locate/apsusc).
2. Вейко В.П. Лазерные технологии в микроэлектронике. София: Издательство Академии наук Болгарии.
3. Павлова В.Т. Оптика тонких пленок и технология их нанесения. Минск: БПИ, 1990.
4. Майсера Л., Глэнта Р., Елинсона М.И., Смольно Г.Г. Технология тонких пленок. / Справочник. М.: Советское радио, 1977.

ВЛИЯНИЕ ЛАЗЕРНОГО МИКРОСТРУКТУРИРОВАНИЯ ПОВЕРХНОСТИ КРЕМНИЯ НА ЭЛЕКТРОФИЗИЧЕСКИЕ СВОЙСТВА ПОЛУЧЕННЫХ СТРУКТУР

С.С. Дышловенко, И.И. Стройков, Фам Куанг Тунг

Научный руководитель – доктор технических наук, профессор А.М. Скворцов

Введение

Кремний – один из самых широко используемых материалов в современной полупроводниковой микроэлектронике и интегральной оптике. Оптические, электрические, физические и фотолюминесцентные свойства кремния определяются морфологией и структурой поверхности кремния. Изменяя морфологию и структуру поверхности кремния, можно изменить и его свойства. В предлагаемой работе проведен эксперимент по лазерному микроstructuringу поверхностей кремния, позволяющий модифицировать не только поверхность кремния, но и его физические, электрические и оптические свойства и, соответственно, значительно расширить область применения данного материала.

Методика эксперимента

В работе используется CO₂-лазер с длиной волны в инфракрасном диапазоне $\lambda=10,6$ мкм, длительностью импульса $\tau=0,88 \times 10^{-4}$ с, частотой следования импульсов $f=80$ Гц и мощностью излучения $P=0,3$ Вт; скорость сканирования составляла $V=2,5$ мм/с. В качестве образцов использовались пластины монокристаллического кремния толщиной 0,38 мм с ориентацией (100). Для активирования поглощения кремнием лазерного излучения пластина кремния была помещена на стеклянную подложку. Процесс сканирования производился в воздухе в разных направлениях относительно базового среза пластины: параллельно, перпендикулярно и под углом 45°.

Относительное перемещение лазерного пучка по поверхности обрабатываемого изделия схематически представлено на рис. 1. Из рисунка видно, что размер облученной зоны d зависит от ее расположения относительно центра рабочего поля y . Увеличение размера облученной зоны при отклонении пучка происходит по двум причинам: вследствие расфокусировки пучка и вследствие наклонного нападения пучка на обрабатываемую поверхность.

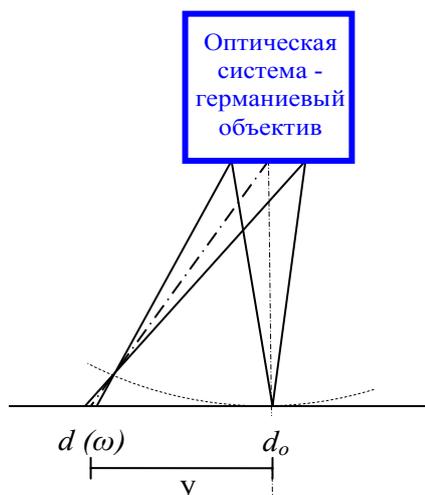


Рис. 1. Изменение облученной области при сканировании пучка

Для исследования электрофизических и структурных свойств образцов, изготовленных описанным способом, использовались методы катодолюминесценции и измерения вольтамперных характеристик (ВАХ).

Люминесцентные методы исследования являются одними из наиболее информативных при изучении структуры двуокиси кремния. Метод электролюминесценции наиболее часто используется при исследовании генерационно-рекомбинационных процессов в МОП-структурах с тонким диэлектриком [1, 2], а также для исследования возможности токовой накачки светоизлучающих наноконструкций, содержащих нанокристаллы кремния.

Метод катодолюминесценции используется сравнительно редко, хотя и обладает рядом преимуществ перед методом фотолюминесценции. В их числе можно отметить возможность регистрации спектров с разрешением по глубине окисной пленки, а также высокую энергию ионизирующего излучения, позволяющие исследовать люминесцентные центры с высокими энергиями возбуждения. В целом нужно отметить, что спектры катодо- и фотолюминесценции весьма похожи.

Исследование катодолюминесцентных свойств системы Si/SiO₂ проводилось на электронно-зондовом микроанализаторе «Camebax» фирмы «Cameca» (Франция), снабженной оптическим спектрометром, регистрирующим катодолюминесцентное излучение образцов в видимой области спектра (350–750 нм).

Вольт-амперные характеристики (ВАХ) полученных образцов были сняты с помощью характеристикиста Л2-56 и зондовой головки. Для эксперимента использовались два облученных лазером образца: чистый монокристаллический кремний и предварительно окисленный монокристаллический кремний. Исследуемые образцы размещались на стеклянной подложке. Для обеспечения омического контакта, а также для исключения выпрямления в контакте на очищенный от окисла SiO₂ участок поверхности был нанесен сплав InGa.

Экспериментальные результаты

В спектрах люминесценции различных силикатных систем можно выделить два четких пика, обусловленных структурными дефектами – немостиковым атомом кислорода и двухкоординированным атомом кремния [2, 3].

Рассмотрим структуру этих дефектов и их спектральные характеристики более подробно. На рис. 2 представлен спектр катодолюминесценции кварцевого стекла марки КИ, характеризующегося правильной стехиометрией и высокой химической чистотой, в частности, малым содержанием гидроксильных групп.

Полоса люминесценции 1,94 эВ на представленном спектре обусловлена наличием оборванных Si-O-связей в структуре стекла, что приводит к появлению комплементарной пары дефектов: немостикового атома кислорода $\equiv\text{SiO}\cdot$, ответственного за люминесценцию в данной области, и трехкоординированного атома кремния $\equiv\text{Si}\cdot$, представляющий собой E'-центр, возбуждение которого не сопровождается люминесценцией. Здесь символ “-” означает химическую связь, а “•” – неспаренный электрон.

Полоса люминесценции, лежащая в области 1,9 эВ, обусловлена электронными переходами между расщепленными 2p-состояниями немостикового атома кислорода. Одно из этих состояний представляет собой заполненную 2p_x-орбиталь, а второе – 2p_y-орбиталь с неспаренным электроном.

Причины расщепления 2p-состояний объясняются на основе динамического эффекта Яна-Теллера, когда немостиковый атом кислорода прецессирует на вершине кремний-кислородного тетраэдра, образуя химические связи с другими кислородными атомами в том же тетраэдре [1].

Полоса люминесценции 2,7 эВ, согласно [3], обусловлена синглет-триплетным расщеплением в молекулярном комплексе двухкоординированного атома кремния $=\text{Si}\cdot$.

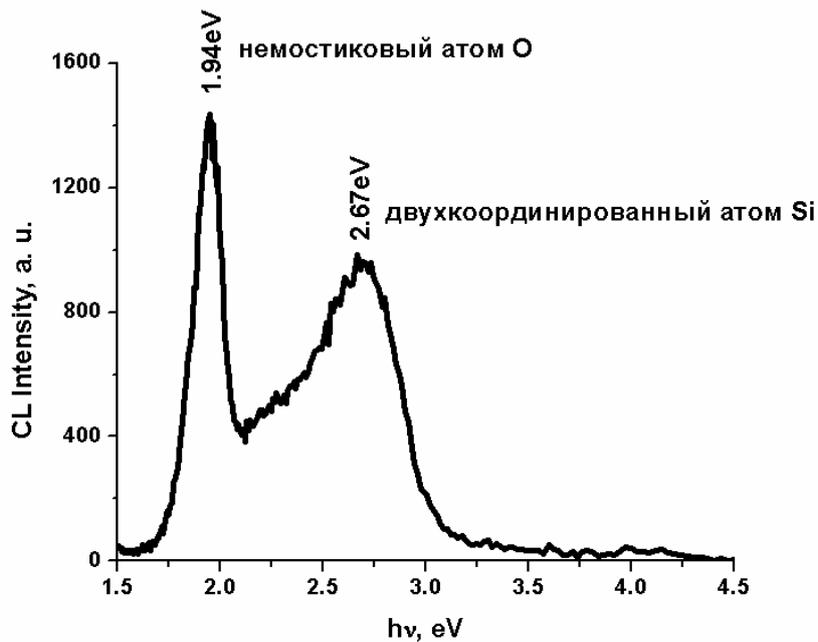


Рис. 2. Спектр катодолуминесценции чистого кварцевого стекла марки КИ

Таким образом, можно утверждать, что полосы 1,9 эВ и 2,7 эВ в спектрах силикатных систем свидетельствуют о наличии кислородо-обогащенных и кремний-обогащенных дефектов, соответственно [1–3]. В спектрах люминесценции термического окисла на кремнии, кроме того, часто наблюдаются пики в зеленой области (2,2–2,5 эВ, рис. 3), отсутствующие в спектрах кварцевого стекла, но характерные для нанокомпозитов опал-кремний, что свидетельствует о наличии нанокластеров кремния в окисле. Точное положение и интенсивность таких пиков сильно зависят от температуры и среды окисления, типа проводимости подложки, режима термической обработки, а также от распределения центров люминесценции по толщине пленки [1–5].

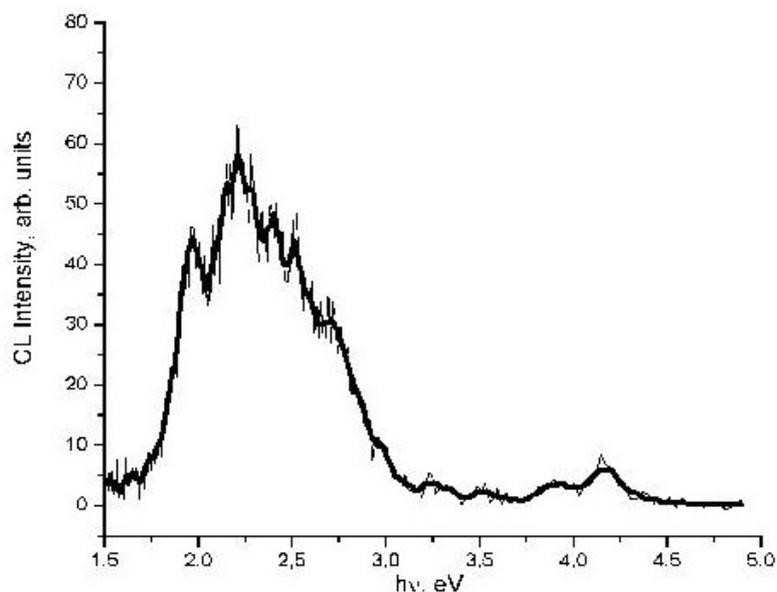


Рис. 3. Спектр катодолуминесценции монокристаллического кремния с окислом SiO_2

В настоящей работе были сняты спектры катодолюминесценции для двух образцов: образец 1 – чистый монокристаллический кремний, облученный лазером и образец 2 – кремний с окислом, облученный лазером. Для каждого из двух образцов катодолюминесцентные характеристики были сняты в облученных областях. Для образца 1 дополнительно был снят спектр люминесценции вне области облучения. Люминесценция возбуждалась пучком электроном с энергией 5 кэВ, величина тока через образец составляла 20 нА, время экспозиции 0,3 с для всех образцов. Спектры катодолюминесценции полученных образцов представлены на рис. 4–6.

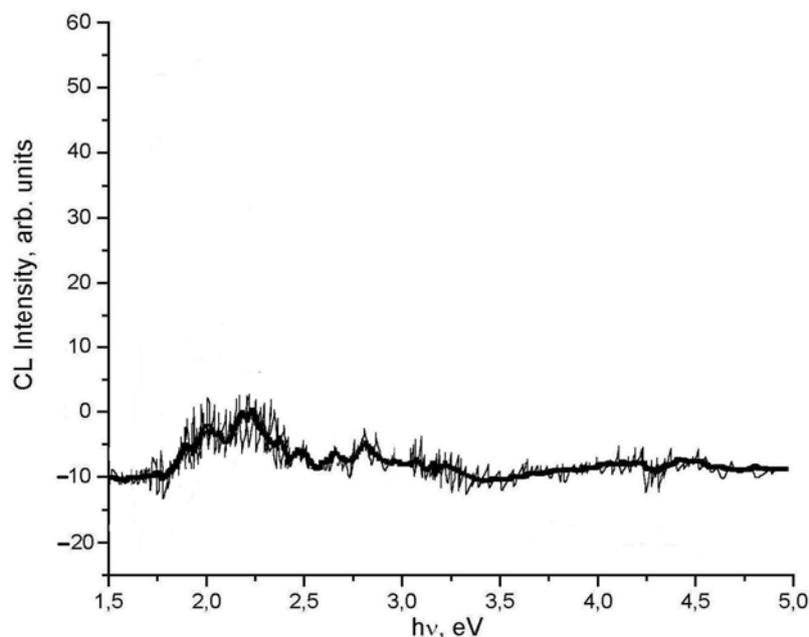


Рис. 4. Спектр катодолюминесценции, снятый вне области облучения монокристаллического кремния (образец 1)

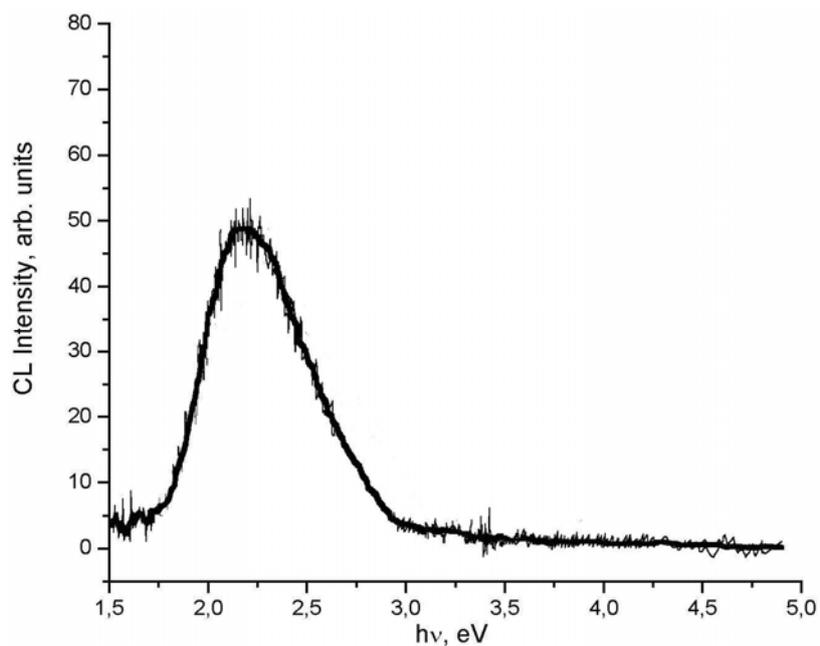


Рис. 5. Спектр катодолюминесценции, снятый в облученных областях предварительного окисленного монокристаллического кремния после облучения лазером (образец 2)

Рассмотрим эти спектры люминесценции более подробно. На поверхности образца 1 вне области облучения существует только тонкий естественный слой окисла. На его спектре люминесценции можно выделить два разных пика, соответствующие энергиям 1,94 эВ, 2,25 эВ и 2,65 эВ. Из-за малой толщины слоя естественного окисла эти пики выражены нерезко (рис. 4).

На спектре люминесценции образца 1, снятом в области облучения, четко выделяется одна яркая полоса, соответствующая энергии 2,25 эВ (рис. 5). Другие полосы, соответствующие энергиям 1,97 эВ и 2,67 эВ, отсутствуют, т.е., как было сказано выше, отсутствуют кислородо-обогащенные и кремний-обогащенные дефекты.

Следует отметить большую интенсивность пика, соответствующего энергии 2,25 эВ. Можно предположить, что после облучения на структурированной поверхности сформировались нанокластеры монокристаллического кремния в пересыщенном атомами кремния окисле.

На спектре люминесценции предварительно окисленного монокристаллического кремния, снятом в облученных областях, четко выделяется только одна яркая полоса, соответствующая энергии 2,25 эВ (рис. 6), аналогичная случаю монокристаллического кремния (рис. 5). Отличие заключается в том, что при энергии 1,97 эВ образовался выступ, характеризующий появление нового вида эффектов окисла. Корректное объяснение этого пика в настоящее время дать затруднительно.

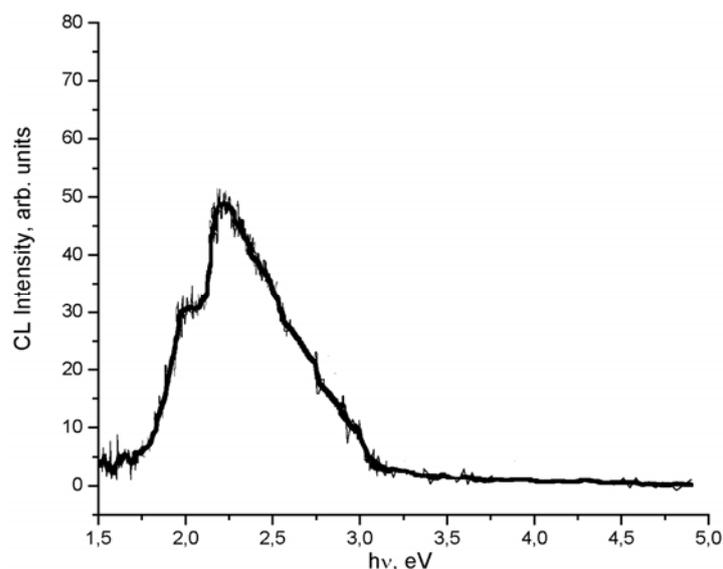


Рис. 6. Спектр катодолюминесценции, снятый в области облучения образца 2 – кремний с окислом SiO_2 , облученный лазером

Для неокисленного образца наблюдается ВАХ близкая к нормальной, аналогично характеристике обычного точечного диода, образованного контактом металл–полупроводник (рис. 7).

На образцах, окисленных до облучения лазером, пробой слоя SiO_2 происходит при напряжении порядка 60 В. При измерении ВАХ после пробоя наблюдается anomальное поведение ВАХ (рис. 8). Из рис. 8 видно, что на обратной характеристике наблюдается резкое увеличение обратного тока, сопровождаемое релаксацией тока в диапазоне 30–50 В. Релаксация заключается в том, что после выдержки измеряемой структуры при напряжении 60 В наблюдается изменение обратной ветви ВАХ по кривой, обозначаемой пунктиром. Возможно, ход обратных ветвей ВАХ, приведенных выше, связан с наличием в окисле структурных дефектов, способных перестраиваться в результате протекания тока по этим дефектам.

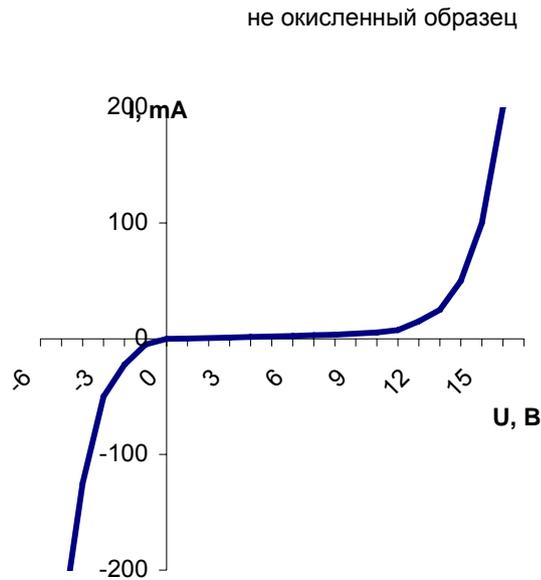


Рис. 7. ВАХ неокисленного монокристаллического кремниевого образца, облученного лазером

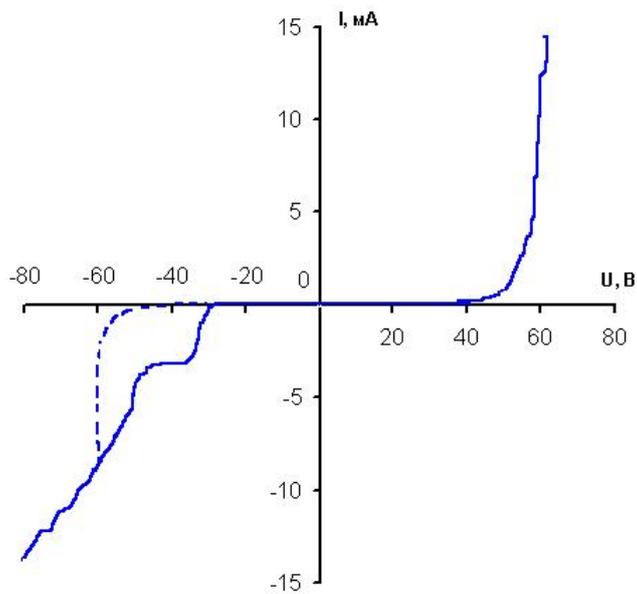


Рис. 8. ВАХ окисленного монокристаллического кремниевого образца, облученного лазером

Заключение

В результате проведенной работы разработан технологический процесс микроstructuring поверхности кремния, обеспечивающий формирования областей с разной энергией облучения.

В результате облучения неокисленного монокристаллического кремния пучком лазера возникает полоса катодоллюминесценции, соответствующая энергии 2,25 эВ, кото-

рая характеризует наличие нанокластеров кремния в структуре кремний – «естественный окисел».

В катодолюминесценции предварительного окисленного образца появился дополнительный выступ, который характеризует появление нового вида эффектов окисла. Корректное объяснение этого пика в настоящее время дать затруднительно.

В результате исследования ВАХ облученных участков была обнаружена аномальная зависимость обратного ВАХ на образцах с предварительным окислением поверхности до облучения.

Литература

1. Kim K., Suh M.S., Kim T.S., Youn C.J., Suh E.K., Shin Y.J., Lee K.B. and Lee H.J. Room-temperature visible photoluminescence from silicon-rich oxide layers deposited by an electron cyclotron resonance plasma source. // *Appl. Phys. Lett.* 1996. V. 69. P. 3908–3910.
2. Bai G.F., Wang Y.Q., Ma Z.C., Zong W.H. and Qin G.G. Electroluminescence from Au/native silicon oxide layer/pC-Si and Au/native silicon oxide layer/nC-Si structures under reverse biases. // *J. Phys.: Condens. Matter.* 1998. V. 10. P. 717–721.
3. Lin C.-F., Liu C.W., Chen M.-J., Lee M.H. and Lin I.C. Infrared electroluminescence from metal-oxide-semiconductor structures on silicon. // *J. Phys.: Condens. Matter.* 2000. V. 12. P. 205–210.
4. Torchinska T., Aguilar-Hernandez J., Morales-Rodriguez M., Mejia-Garcia C., Contreras-Puente G., Becerril Espinosa F.G., Bulakh B.M., Schverbina L.V., Goldstein Y., Many A. and Jedrzejewski J. Comparative investigation of photoluminescence of silicon wire structures and silicon oxide films. // *Journal of Physics and Chemistry of Solids.* 2002. V. 63. P. 561–568.
5. Torchynska T.V., Bulakh B.M., Polupan G.P., Palacios Gomez J., Flores Gonzalez H.A., Bacarril Espinoza F.G., Ita Torre A. and Scherbina L.V. Comparative investigation of surface structure, photoluminescence and its excitation in silicon wires. // *Journal of Electron Spectroscopy and Related Phenomena.* 2001. V. 114–116. P. 235–241.

**ЭКСПЕРТНЫЕ СИСТЕМЫ ДЛЯ КЛАССИФИКАЦИИ БОЛЕЗНЕЙ
В МЕДИЦИНСКОЙ ДИАГНОСТИКЕ****П.А. Головин, В.А. Нечаев, Д.А. Нечаев****Научный руководитель – доктор технических наук, профессор К.Г. Коротков**

Произведен анализ методов разработки экспертных систем для классификации болезней в медицинской диагностике. Изложены достоинства и недостатки каждого из методов. Приведен обзор существующих медицинских экспертных систем.

Введение

Развитие новых методов диагностики, а также расширение возможностей уже существующих методов было и остается актуальной задачей в медицине. Появление новейших диагностических и лечебных технологий требуют использования методов искусственного интеллекта для обработки и интерпретации данных с возможностью накопления, хранения и многократного использования медицинских данных [1]. Одним из наиболее эффективных средств в данной области являются экспертные системы. Они позволяют автоматизировать процесс принятия решения при освидетельствовании пациентов, повышая уровень квалификации пользователя до уровня опытных экспертов [3]. Поэтому необходимо, чтобы экспертные системы обладали возможностью гибкой постановки задач, были применимы ко всем областям биологии и медицины, обладали большой информационной емкостью и помехоустойчивостью, не нуждались в длительном времени для разработки [1]. Непрерывное развитие средств вычислительной техники расширяет потенциальные возможности подобных систем, в связи с чем необходимо постоянно пополнять знания в данной области.

Экспертные системы в течение длительного времени применяются в медицине для диагностики заболеваний. Каждая система при этом обладает ограниченной областью применения, в связи с изначальной направленностью разработки. Применение этих экспертных систем в областях, для которых они изначально не были предназначены, затруднено и зачастую просто невозможно, в том числе потому, что они ограничены встроенными способами вывода [3].

В настоящее время медицинские диагностические системы строятся на основе нескольких методов обработки информации:

- построение дерева решений;
- статистическая обработка данных;
- использование элементов искусственного интеллекта [3].

Цель данной работы – провести анализ существующих методов разработки экспертных систем и выявить достоинства и недостатки каждого из методов.

Основная часть

Решение задач разработки экспертных систем можно подразделить на две большие категории:

- задачи, решаемые по явным алгоритмам;
- задачи принятия решений на основе опыта и знаний.

Соответственно, можно выделить 2 способа решения задачи – логический и интуитивный. Логический метод оперирует набором формальных правил, интуитивный –

накопленным опытом. При решении задачи первым способом задача обычно разбивается на подзадачи, каждая из которых, в свою очередь, разбивается на несколько элементарных функций с известным алгоритмом вычислений. Зная алгоритм каждой элементарной функции, можно решать сложнейшие задачи, соединяя элементарные функции в нужной последовательности. При попытке применить такой метод к решению интуитивных задач программист может столкнуться с невозможностью выделить алгоритм элементарной функции или вообще разбить задачу на такие функции [3].

Существующие традиционные системы принятия решений, основанные на явных правилах вывода, создаются, как правило, группой специалистов, в числе которых – математики, программисты и предметные специалисты, ставящие задачи. Возможности настройки таких систем на конечного потребителя часто недостаточны.

Самообучающиеся медицинские экспертные системы (ЭС) принятия решений должны удовлетворять следующим требованиям:

- индивидуализация (настройка на традиции клинических школ, геосоциальные особенности региона применения, наборы медико-биологических данных, особенности лечебно-диагностических технологий, индивидуальный опыт и знания специалиста);
- динамическое развитие (накопление опыта системы в процессе функционирования, следуя изменениям в пунктах, перечисленных в предыдущем требовании);
- возможность перенастройки при резком изменении условий, например, при перенесении в другой регион;
- способность к экстраполяции результата – требование, обратное индивидуальности. Система не должна резко терять качество работы при изменении условий;
- возможность конструирования с нуля конечным пользователем (специалист должен иметь возможность придумать совершенно новую ЭС и иметь возможность просто и быстро создать ее);
- «нечеткий» характер результата. Решение, выдаваемое системой, не должно быть окончательным. Оно может быть вероятностным или предлагать сразу несколько вариантов на выбор. Это дает возможность специалисту критически оценивать решение системы и не лишает его инициативы в принятии окончательного решения;
- ЭС является только советчиком специалиста, не претендуя на абсолютную точность решения. Она должна накапливать опыт и знания и значительно ускорять доступ к ним, моделировать результат при изменении условий задачи. Ответственность за решение всегда лежит на специалисте [1].

Большинство разработанных экспертных систем удовлетворяют этим требованиям. Перечислим наиболее известные из них:

- PUFF – экспертная система, осуществляющая диагностику легочных заболеваний на основе легочных функциональных тестов;
- SPE – производит диагностику состояний при воспалительных процессах;
- ABEL – осуществляет диагностику кислотных и электролитных заболеваний;
- AI/RHEUM – диагностика заболеваний соединительных тканей;
- CADUCEOS – диагностика внутренних заболеваний общего профиля;
- BLUE FOX – диагностика и лечение депрессивных состояний;
- CASNET/GLACOMA – диагностика и лечение глазных заболеваний, связанных с глаукомой;
- MYCIN – диагностика и лечение инфекционных заболеваний;
- ONCOCIN – лечение больных раком химиотерапией и наблюдение за ними;
- PIP – диагностика заболеваний почек;
- МОДИС-2 – диагностика симптоматической гипертензии;

- GUIDON – обучающая система диагностики и лечения инфекционных заболеваний [2].

Каждая из этих систем обладает узкой областью применения и может использоваться по назначению только в ней.

Медицинские экспертные системы строятся на основе нескольких подходов к обработке информации. Используются следующие подходы.

Построение дерева решений. При использовании этого подхода в программе протоколируется последовательность вопросов, задаваемых врачом при решении диагностической проблемы. Такой протокол структурируется в виде дерева. Каждая вершина такого дерева представляет собой определенный вопрос, задаваемый больному, а ветвления, исходящие из вершины, соответствуют альтернативным ответам на вопрос и ведут, в свою очередь, к новым вопросам. Программа осуществляет переход от вопроса к вопросу до тех пор, пока не будет найдено решение или исчерпаны возможные переходы. Недостатком такого подхода можно считать следующее:

- при попытке построить подобные деревья для решения сложных диагностических задач количество вершин и ветвлений становится настолько большим, что анализ логического дерева оказывается крайне сложным;
- малейшие изменения, вносимые в логику программы, приводят к необходимости строить дерево заново и перепрограммировать всю задачу.

В то же время такой подход крайне удобен, так как позволяет представить в программе логику составления последовательности вопросов врачом при решении диагностической задачи в клинических условиях. Этот подход позволяет имитировать процесс принятия решения врачом при постановке диагноза.

Статистическая обработка данных. Этот подход заключается в применении методов математической статистики. Он основывается на обработке больших массивов информации, собранных по заболеваниям, подлежащим машинной диагностике. Обработанная информация может использоваться различным образом.

Часть медицинских экспертных систем базируется на использовании теории распознавания образов. При этом необходимо иметь некоторое множество конкретных историй болезней с известными диагнозами. Такие множества анализируются с целью определения статистически «типичной» для каждого заболевания картины – образа. Определяются те признаки заболевания, которые наиболее характерны для него, исходя из собранной информации, подвергнутой статистической обработке. «Типичные» картины заболеваний используются при анализе истории каждого конкретного пациента, для определения того, насколько «похож» рассматриваемый случай на «типичный». Оценивая «расстояние» между сравниваемыми картинами, программа формирует решение о диагнозе.

Ряд методов компьютерной диагностики основан на применении теоремы Байеса. Такой подход позволяет определить вероятности заболевания у человека и базируется на установлении частоты появления признаков при заболевании. Например, предположим, что D_i – одно из n заболеваний, подлежащих диагностике, а E – признаки, наблюдаемые у больного. Тогда если $P(D_i)$ – вероятность появления i -го заболевания, то вероятность i -го заболевания при условии наличия признаков E будет

$$P(D_i / E) = \frac{P(D_i)P(E / D_i)}{\sum_{j=1}^n P(D_j)P(E / D_j)},$$

где $P(D_i / E)$ – условная (апостериорная) вероятность появления признаков E при заболевании D_i .

Несомненными достоинствами статистических методов является то, что с их помощью делается попытка объективизировать имеющуюся информацию о заболеваниях. Присутствуют у них также и существенные недостатки.

Первый из них связан с тем, что информация, необходимая для построения статистических моделей, часто отсутствует. Для ее накопления, хранения и обработки требуются специальные средства вычислительной техники, создание банков данных по заболеваниям. Количество систематизированной информации по отдельным областям медицины еще мало. Да и само построение статистических моделей является непростым делом.

Использование теоремы Байеса имеет некоторые ограничения, связанные с тем, что оно основывается на использовании некоторых предположений. Прежде всего, предполагается, что каждое заболевание имеет свои непересекающиеся наборы симптомов, в то время как на практике часто встречается ситуация, когда одинаковые симптомы встречаются у разных заболеваний. Другое допущение используется при определении условных вероятностей появления признаков при заболевании и основано на независимости симптомов между собой. Такие зависимости, тем не менее, имеются и налагают существенные ограничения на применимость теоремы.

Еще один существенный недостаток этих методов состоит в том, что внесение в модели новой информации представляет собой значительную трудность, так как ведет к изменению самой модели и пересчету всех вероятностей.

Также недостатком является невозможность объяснения врачу и больному того, каким именно образом получен данный результат. Причиной этого является то, что данный результат является следствием математических операций и не имеет ничего общего с обычным процессом принятия решения врачом при обследовании пациента.

Обработка данных, построенная на основе использования элементов искусственного интеллекта (ИИ). В таких системах делается попытка моделирования способности человека к рассмотрению предметной области и совершения умозаключений с отбрасыванием наименее перспективных направлений поиска. Для этого используются наборы правил, задаваемые априорно и являющиеся, по сути, знаниями экспертов в определенной проблемной области. Именно качество знаний определяет «компетентность» проблемно-ориентированной экспертной системы. Знания о проблемной области, формализованные определенным образом, хранятся отдельно от остальных программ, что позволяет модифицировать работу системы ИИ, меняя набор знаний и не затрагивая остальных компонентов системы. Совокупность таких правил принято называть базой знаний, в то время как совокупность данных о пациентах – базой данных.

Существует два основных подхода к организации баз знаний, выражающиеся в разной структуре соответствующих систем:

- описание отношений между описываемыми объектами;
- представление обработки знаний в виде процедур.

Среди наиболее известных способов представления знаний упомянем продукционный. При данном подходе реализуется четкое разделение между данными, операциями и управлением. Можно выделить некую глобальную базу данных (не следует ее путать с базой данных в традиционном понимании), над которой выполняются некоторые действия, описываемые совокупностью правил продукции. Управление этим процессом происходит в соответствии с некоторой глобальной стратегией управления.

Глобальная база данных – это структура данных, анализируемая и преобразуемая системой ИИ. Вид базы зависит от решаемой задачи и может быть простым (как, например, списочная структура) или сложным.

Правила продукции состоят из двух частей: условие (левая часть) и действие (правая часть). Условие устанавливает применимость правила к базе данных. Действие заключается в изменении информации в глобальной базе данных.

Система управления выбирает, какое из нескольких применимых правил должно быть использовано, а также определяет момент прекращения дальнейших действий путем проверки терминальных условий на базе данных.

Поиск решения осуществляется на основе двух основных подходов. Метод «от данных к цели» заключается в последовательном изменении базы данных всеми правилами, которые могут быть применены, и в поиске после изменения новых применимых правил. Поиск заканчивается при наличии утверждения, являющегося решением, или при выполнении условия прекращения поиска. Метод «от цели к данным» сводится к проверке некоторых утверждений, которые могут выступать в качестве наиболее вероятных решений. При этом проводится поиск правых частей правил с целью обнаружения искомого отрицания и проверка левых частей соответствующих правил на базе данных – при подтверждении либо гипотеза считается истинной, либо левая часть рассматривается в качестве новой гипотезы.

Заключение

В работе рассмотрены основные концепции и принципы, лежащие в основе методов разработки экспертных систем. Уровень биомедицинского анализа и диагностики болезней зависит также от внедрения новых биомедицинских приборов. В работе [4] показано, что ГРВ-камера с помощью соответствующего программного обеспечения позволяет оценить функциональное состояние систем и органов человека. Имея эти данные, эксперт может сделать заключение о наличии заболевания. Создание экспертной системы для ГРВ-диагностики является задачей дальнейшей работы. Разрабатываемая экспертная система должна обладать необходимой гибкостью и простотой в использовании. Анализ методов разработки экспертных систем показал, что данным требованиям удовлетворяют статистические методы в сочетании с методами искусственного интеллекта.

Литература

1. Kononenko W. Machine learning for medical diagnosis: history, state of the art and perspective. // *Artificial Intelligence in Medicine*. Ljubjana. 2001. P. 89–109.
2. Россиев Д.А. Нейросетевые самообучающиеся экспертные системы в медицине. / Молодые ученые – практическому здравоохранению. Красноярск, 1994. С. 17.
3. Смолин Д.В. Введение в искусственный интеллект. М.: ФИЗМАТЛИТ, 2004. 208 с.
4. Коротков К.Г. Основы ГРВ биоэлектрографии. СПб: СПбГИТМО (ТУ), 2001. 360 с.

РАЗРАБОТКА РАСПРЕДЕЛЕННОЙ БАЗЫ ДАННЫХ НА ПРИМЕРЕ БАЗЫ ДАННЫХ СПОРТСМЕНОВ ОЛИМПИЙСКОГО РЕЗЕРВА

П.А. Головин, Д.В. Денисов, В.А. Нечаев, Д.А. Нечаев

Научный руководитель – доктор технических наук, профессор К. Г. Коротков

Статья посвящена описанию распределенной базы данных. Приводятся основные возможности и требования к таким базам данных, показывается реализация возможностей системы в конкретном случае.

Россия, в силу своего географического положения и размеров, «обречена» на преимущественное использование распределенных систем и распределенных баз данных (БД). Поэтому актуальность разработки таких систем не вызывает сомнения. Целью данной статьи является сжатое описание схемы реализации распределенной БД.

Под распределенной подразумевается БД, включающая фрагменты из нескольких БД, которые располагаются на различных узлах сети, и, возможно, управляются различными СУБД. Распределенная база данных выглядит для пользователя и прикладных программ как обычная локальная база данных.

Дэйт (С.Д. Date) предложил 12 качеств идеальной распределенной базы данных [1]:

- локальная автономия (local autonomy);
- независимость узлов (no reliance on central site);
- непрерывные операции (continuous operation);
- прозрачность расположения (location independence);
- прозрачная фрагментация (fragmentation independence);
- прозрачное тиражирование (replication independence);
- обработка распределенных запросов (distributed query processing);
- обработка распределенных транзакций (distributed transaction processing);
- независимость от оборудования (hardware independence);
- независимость от операционных систем (operationg system independence);
- прозрачность сети (network independence);
- независимость от баз данных (database independence).

Исходя из требований конфиденциальности, в БД спортсменов олимпийского резерва были реализованы основные качества распределенной БД. Ее схематичная топология изображена на рис.

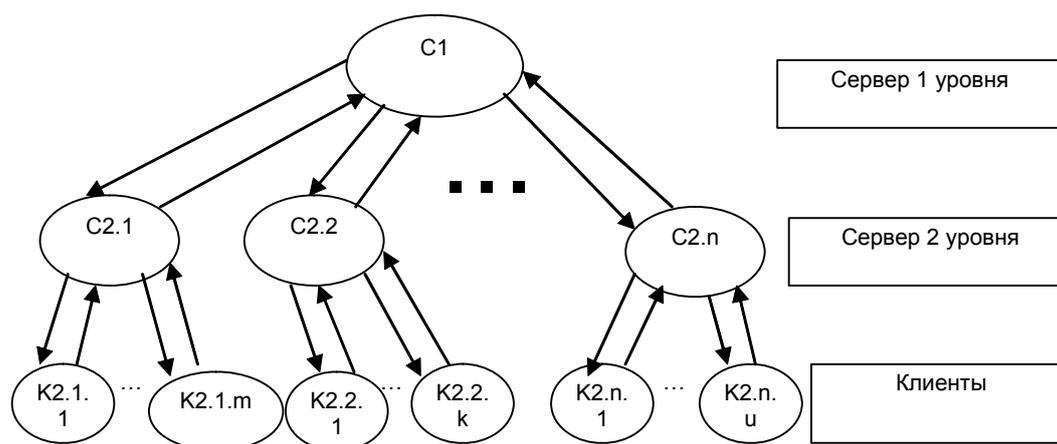


Рис. Физическая организация базы данных

Качество «локальной автономии» было реализовано путем установки на клиентских машинах и серверах 2-го уровня локальных БД (в качестве СУБД используется MySQL). Строго говоря, клиентские БД не являются частью распределенной БД, так как данные клиентов содержат, среди прочего, и конфиденциальную информацию. Пе-

риодически с БД клиентов формируется отчет, данные из которого добавляются в БД сервера 2-го уровня, которые составляют кластеры распределенной БД. Таким образом, реализовано и качество «независимости узлов», так как каждый клиент и сервер 2-го уровня являются самостоятельными локальными серверами БД.

Качество «непрерывных операций» возможно реализовать только на уровне серверов 2-го уровня, так как клиентские узлы не могут работать круглые сутки, а доступ к серверам 2-го уровня имеется всегда.

Клиент, работающий с базой данных, может не знать о ее реальной структуре, для него работа выглядит так же, как работа с локальной БД. Качество «прозрачности расположения» реализуется путем отправки всех запросов клиента к БД на сервер 2-го уровня, который уже сам занимается дальнейшей обработкой пакета. В результате клиент получает ответ с этого же сервера независимо от пути прохождения запроса.

Из-за специфики информации, ее сбора и обработки нет необходимости в качествах «прозрачная фрагментация» и «прозрачное тиражирование», которые не реализованы в рамках проекта.

Качество «выполнение распределенных запросов» реализуется путем переправления запроса пользователя серверу 1-го уровня, который организует дальнейшие запросы к другим кластерам БД. Ответы с других кластеров обрабатываются сервером 1-го уровня, объединяются в один общий ответ и отправляются серверу, с которого пришел запрос. Сервер обрабатывает ответ и объединяет со своим ответом, и все результаты запроса передаются пользователю.

Качества «независимости от оборудования» и «независимости от операционной системы» реализованы путем использования СУБД, реализованной под большое количество операционных и компьютерных систем. Клиентская часть написана на РНР, что позволяет запускать ее на различных системах.

В связи с проблемами защиты системы качество «прозрачность сети» реализовано только в рамках протокола НТТР, что не сужает круг клиентов данной базы.

Несмотря на то, что в качестве СУБД используется MySQL, есть возможность реализовать различные кластеры БД под управлением различных СУБД, что в дальнейшем и планируется сделать.

Приведенная система имеет все основные качества распределенной базы данных, некоторые из них требуют более детальной проработки, что не позволяет назвать эту БД законченной. Текущая ее реализация работает через web-интерфейс и имеет достаточно широкие возможности.

Литература

1. Date C.J. What is a Distributed Database System. // The Relational Journal. Издание 1. 1987.

МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЯ ПЛОТНОСТИ ТОКА В СЛОЖНОМ НЕОДНОРОДНОМ ПРОВОДНИКЕ.

ЧАСТЬ 1

А.Ю. Гришенцев

Научный руководитель – доктор технических наук, профессор К.Г. Коротков

В статье рассмотрена математическая основа моделирования распределения плотности тока в сложном, неоднородном проводнике, имеющем произвольную форму поперечного сечения.

Введение

Получивший в последнее время широкое распространение во всем мире процесс газоразрядной визуализации (ГРВ) [1], непосредственно связан с протеканием токов высокой частоты через исследуемые объекты. При этом наблюдается неравномерное распределение плотности тока по сечению, называемое поверхностным эффектом. Моделирование поверхностного эффекта в однородных проводниках подробно рассмотрено в различной литературе по электротехнике. В данной статье построена математическая модель поверхностного эффекта для случая неоднородных проводников. В этом случае вклад в неравномерность распределения тока по сечению вносит неоднородность среды, например, биологических объектов, исследуемых на ГРВ оборудовании.

Моделирование электропроводящих свойств, сложных биологических тканей, рассмотрено в ряде современной литературы (например: [2–4] и др.). К сожалению не одна из предложенных моделей не позволяет оценить плотность распределения тока по биологическим тканям в процессе ГРВ, ИПЧ исследований.

Автором статьи предложен возможный вариант оценки плотности распределения тока, по биологическим тканям применимый к ГРВ и ИПЧ измерениям. Моделирование производилось на основе законов классической электродинамики ([5–9]), с учетом особенностей протекания токов по биологическим тканям ([3, 4, 10–14]).

Модель плотности распределения высокочастотного тока по сечению во время исследования на ГРВ оборудовании позволяет:

- рассмотреть различия распределения плотности тока в зависимости от электрических параметров исследуемого объекта;
- определить части объекта, по которым протекает преобладающая часть тока;
- предсказать тенденцию изменения картины ГРВ-граммы при изменении электрических параметров исследуемого объекта;
- предсказать тенденцию изменения интегральной суммы тока для приборно-измерителей, интегральной суммы токов поверхностных частот (ИПЧ), при изменении электрических параметров исследуемого объекта;
- с высокой точностью оценить значения протекающих по различным тканям токов (при исследовании человека) с целью выяснения безопасности обследования на ГРВ, ИПЧ оборудовании.

Построение математической модели

Рассмотрим картину магнитного поля в плоскости поперечного сечения проводника с током, выполненного из однородного материала (рис. 1). Представим этот провод в виде совокупности нитей, параллельных его оси. Чем ближе нить расположена к оси проводника, тем с большим числом магнитных линий она сцеплена.

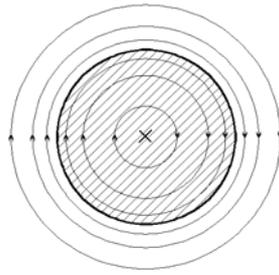


Рис. 1. Магнитное поле в плоскости поперечного сечения проводника с током

При периодическом изменении тока изменяется магнитное поле, и в нитях проводника наводится э.д.с., противодействующие изменениям тока. Это противодействие тем значительнее, чем больше э.д.с. (чем больше магнитных линий сцеплено с нитью), т.е. чем ближе нить проводника расположена к оси провода. В результате плотность тока в различных точках поперечного сечения получается не одинаковой: наибольшая на периферии и наименьшая на его оси.

Рассмотренное явление концентрации переменного тока в поверхностном слое проводника называют поверхностным эффектом. Резкость проявления его возрастает с увеличением частоты f , диаметра проводника d , магнитной проницаемости μ и удельной проводимости γ материала проводника. Следствием поверхностного эффекта является некоторое уменьшение собственной индуктивности проводника ввиду ослабления магнитного поля во внутренней части проводника. В предельном случае, когда ток концентрируется на поверхности проводника в бесконечно тонком слое, магнитное поле внутри проводника отсутствует.

Другим следствием поверхностного эффекта является возрастание тепловых потерь при одинаковых значениях переменного и постоянного тока (равенстве значения постоянного тока и действующего значения переменного тока): тепловые потери больше при переменном токе. Поэтому сопротивление проводника переменному току (активное сопротивление) выше, чем сопротивление проводника постоянному току.

Степень неравномерности распределения тока по сечению проводника благодаря поверхностному эффекту в каждом конкретном случае можно оценить, сравнивая линейные размеры поперечного сечения провода с длиной волны электромагнитных колебаний в проводнике λ или с эквивалентной глубиной проникновения электромагнитной волны ([8], стр. 6):

$$\delta = \sqrt{\frac{2}{\omega \mu \gamma}} = \frac{\lambda}{2\pi}, \quad (1)$$

где ω – круговая частота, μ – абсолютная магнитная проницаемость вещества, γ – удельная объемная электропроводность вещества.

На распределение переменного тока в проводе также оказывают влияние токи соседних проводников. Это явление называют эффектом близости ([6, 8]).

Следует отметить, что в поверхностном слое плотность тока не только убывает по величине, но и изменяется по фазе. На поверхности проводника плотность тока опережает по фазе полный ток на 45° : плотность тока на поверхности пропорциональна комплексной напряженности электрического поля \dot{E} , полный ток пропорционален комплексной напряженности магнитного поля \dot{H} , \dot{E} опережает \dot{H} по фазе на 45° . По мере перехода к более глубоким слоям плотность тока запаздывает относительно поверхностного тока, и на глубине, равной половине длины волны, ток изменяет направление на противоположное. Однако этот обратный ток не играет существенной роли – в поверхностном слое толщиной в половину длины волны $\lambda/2$ протекает свыше 95% общего тока [6].

Если проводники имеют сложную форму поперечного сечения, то из-за эффектов поверхностного и близости по отдельным частям проводников протекают токи различной плотности.

Допустим, что имеется проводник длиной l , произвольного поперечного сечения рис. 2 с током:

$$i = I_m \sin(\omega t), \quad (2)$$

где I_m – амплитуда тока [А]; $\omega = 2\pi f$ – угловая частота тока [рад/сек]; f – частота тока [Гц]; t – текущее время [с]. Разобьем поперечное сечение проводника на элементарные площадки. В проводнике площадки могут отличаться по размерам одна от другой. Однако, для удобства расчетов, желательно иметь одинаковые по размерам площадки. Диаметр площадки рекомендуется выбирать примерно вдвое меньше, чем глубина проникновения δ . Будем рассматривать образованные разбиением проводники как элементарные со своим током i_k , где принимает значения $0-n$. Заменим получившийся контур эквивалентной электрической схемой рис. 3, состоящей из $(n+1)$ ветвей. Если найти токи в каждой ветви, а затем плотность тока в каждом элементарном проводнике, то задача по определению поля плотности тока будет решена.

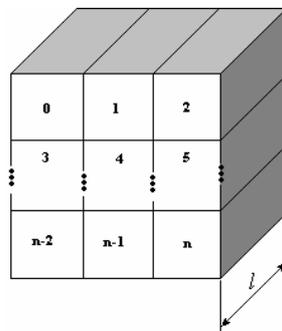


Рис. 2. Проводник произвольного поперечного сечения

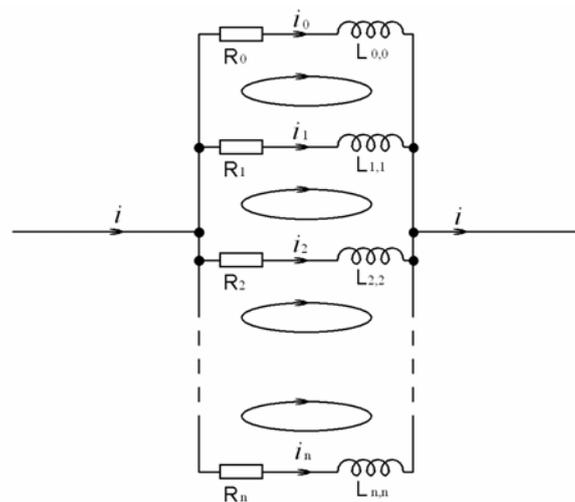


Рис. 3. Эквивалентная схема

На эквивалентной схеме (рис. 3) обозначены: I – мгновенное значение суммарного тока [А]; i_0, i_1, \dots, i_n – мгновенные значения токов в элементарных проводниках [А]; $L_{0,0}, L_{1,1}, \dots, L_{n,n}$ – собственные индуктивности элементарных проводников [Гн]; R_0, R_1, \dots, R_n – активные сопротивления элементарных проводников [Ом].

Для расчета собственных индуктивностей ветвей с 0-й по n -ю рассмотрим произвольный k -й элемент (рис. 4). Координаты центра сечения элементарного проводника x_k, y_k ; размеры сечения b_k, c_k [м]. При моделировании удобно проводить разбиение на элементарные проводники квадратного сечения, поэтому $b_k=c_k$. Если размеры сечения элементарного проводника составляют половину или меньше от глубины проникновения δ , формула (1) (случай низкой частоты), можно считать плотность тока в пределах сечения единичного проводника постоянной. В этом случае индуктивность и активное сопротивление единичного проводника можно находить, как при постоянном токе. В случае высокой частоты, когда размеры сечения больше половины глубины проникновения, надо учитывать неравномерность распределения тока по сечению. В дальнейшей реализации модели есть возможность разбиения проводника на равные элементарные площадки квадратного (со стороной ребра b_k) и кругового (диаметр d_k) сечений, поэтому будем рассматривать квадратное и круговое сечения. В приведенных ниже формулах: γ_k – удельная проводимость, δ_k – глубина проникновения для k -го элементарного проводника.

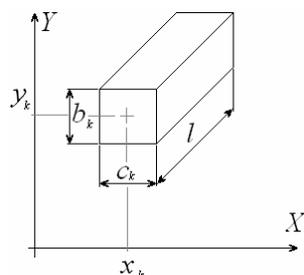


Рис. 4. Произвольный элемент проводника

В случае низкой частоты имеем:

- для проводников квадратного сечения собственную индуктивность будем вычислять по следующей формуле ($b_k \leq 2\delta$) ([8], с. 101):

$$L_{k,k} = \frac{\mu_0 l}{2\pi} \left(\ln \frac{l}{b_k} + \frac{1}{2} - \frac{(\omega \mu_0 \gamma_k)^2 (0.2b_k)^4}{24} \right), \quad (3)$$

где $\mu_0 = 4\pi \cdot 10^{-7}$ Гн/м – магнитная проницаемость воздуха;

- для расчета индуктивности проводников кругового сечения ($d_k \leq \delta$) ([8], с. 96):

$$L_{k,k} = \frac{\mu_0 l}{2\pi} \left(\ln \frac{4l}{d_k} - \frac{3}{4} \right), \quad (4)$$

где $d_k = b_k$ – диаметр поперечного сечения элементарного проводника [м];

- для прямоугольного сечения, активное сопротивление ($b_k \leq 2\delta$):

$$R_k = \frac{1}{\gamma_k} \frac{l}{b_k^2}; \quad (5)$$

- для кругового сечения, активное сопротивление ($d_k \leq 2\delta$):

$$R_k = \frac{1}{\gamma_k} \frac{4l}{\pi d_k^2}. \quad (6)$$

В случае высокой частоты имеем:

- для проводников квадратного сечения собственную индуктивность будем вычислять по следующей формуле ($b_k > 2\delta$) ([8], с. 101):

$$L_{k,k} = \frac{l}{\pi} \left(\frac{\mu_0}{2} \left(\ln \frac{2l}{0.5902b_k} - 1 \right) + \frac{1}{b_k} \sqrt{\frac{\mu_0}{2\omega\gamma_k}} \right); \quad (7)$$

- для расчета индуктивности проводников кругового сечения ($d_k > 2\delta$) ([8], с. 96):

$$L_{k,k} = \frac{l}{2\pi} \left(\mu_0 \left(\ln \frac{4l}{d_k} - 1 \right) - \frac{2}{d_k} \sqrt{\frac{\mu_0}{2\omega\gamma_k}} \right); \quad (8)$$

- для прямоугольного сечения с учетом глубины проникновения выведем формулу расчета активного сопротивления ($b_k > 2\delta$):

$$R_k^{\approx} = \frac{1}{\gamma_k} \frac{l}{\left(b_k^2 - \frac{\pi(b_k - 2\delta_k)^2}{4} \right)}; \quad (9)$$

- для кругового сечения с учетом глубины проникновения выведем формулу расчета активного сопротивления ($d_k > 2\delta$):

$$R_k^{\approx} = \frac{1}{\gamma_k} \frac{l}{\pi \delta_k (d_k - \delta_k)}. \quad (10)$$

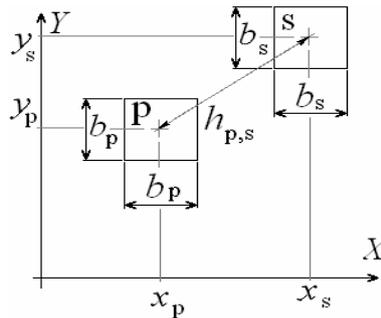


Рис. 5. К расчету взаимной индуктивности элементарных проводников

Рассчитаем взаимные индуктивности M элементарных проводников. На рис. 5 изображены сечения двух проводников с произвольными номерами p, s , соответственно координаты центров данных проводников $(x_p; y_p)$ и $(x_s; y_s)$. Расстояние между центрами проводников

$$h_{p,s} = \sqrt{(x_s - x_p)^2 + (y_s - y_p)^2}. \quad (11)$$

Далее будем обозначать взаимные и собственные индуктивности символами $L_{p,s}$, $L_{k,k}$, соответственно, так как собственную индуктивность можно рассматривать как взаимную саму на себя. Такое обозначение принято в связи с тем, что численные значения всех индуктивностей удобно при вычислениях хранить в одном массиве данных.

Рассмотрим возможные случаи для расчета взаимной индуктивности элементарных проводников p, s при различных соотношениях расстояния $h_{p,s}$ между осями проводников и их длины l ([8], с. 112):

- при $h_{p,s} \leq l$:

$$L_{p,s} = \frac{\mu_0 l}{2\pi} \left(\ln \frac{2l}{h_{p,s}} - 1 + \frac{h_{p,s}}{l} - \frac{1}{4} \frac{h_{p,s}^2}{l^2} + \frac{1}{32} \frac{h_{p,s}^4}{l^4} \right); \quad (12)$$

- при $h_{p,s} > l$:

$$L_{p,s} = \frac{\mu_0 l^2}{4\pi h_{p,s}} \left(1 - \frac{1}{12} \frac{l^2}{h_{p,s}^2} + \frac{1}{40} \frac{l^4}{h_{p,s}^4} \right). \quad (13)$$

При расчетах для уменьшения вычислений необходимо учесть, что $L_{p,s} = L_{s,p}$. Для схемы замещения рис. 3 можно составить следующие уравнения. Поочередно обходя по часовой стрелке контуры, образованные ветвями 0 и 1, 1 и 2, ..., $n-1$ и n , запишем в соответствии с законами Кирхгофа:

$$\left\{ \begin{array}{l} \frac{d\Psi_{0,1}}{dt} + i_0 R_0 - i_1 R_1 = 0 \\ \frac{d\Psi_{1,2}}{dt} + i_1 R_1 - i_2 R_2 = 0 \\ \dots \\ \dots \\ \frac{d\Psi_{n-1,n}}{dt} + i_{n-1} R_{n-1} - i_n R_n = 0 \\ i = \sum_{k=0}^n i_k \end{array} \right. , \quad (14)$$

где $\Psi_{p,s}$ – потокосцепление контуров, образованных соответствующими ветвями:

$$\left\{ \begin{array}{l} \Psi_{0,1} = \Psi_0 - \Psi_1 \\ \Psi_{1,2} = \Psi_1 - \Psi_2 \\ \dots \\ \dots \\ \Psi_{n-1,n} = \Psi_{n-1} - \Psi_n \end{array} \right. . \quad (15)$$

Потокосцепления Ψ_k , $k=0,1,\dots,n$, можно рассчитать следующим образом:

$$\left\{ \begin{array}{l} \Psi_0 = L_{0,0}i_0 + L_{0,1}i_1 + \dots + L_{0,n}i_n \\ \Psi_1 = L_{1,0}i_0 + L_{1,1}i_1 + \dots + L_{1,n}i_n \\ \dots \\ \dots \\ \Psi_n = L_{n,0}i_0 + L_{n,1}i_1 + \dots + L_{n,n}i_n \end{array} \right. , \quad (16)$$

где $L_{k,k}$ – собственная индуктивность элементарного проводника; $L_{s,p}=L_{p,s}$ – взаимная индуктивность p и s элементарных проводников, соответственно.

Запишем уравнение для расчета $\Psi_{p,s}$, используя (15) и (16):

$$\Psi_{p-1,p} = \sum_{k=0}^n (L_{p-1,k} - L_{p,k})i_k, \quad (17)$$

где $p=1,2,\dots,n$. Подставив (17) в (14), получим следующую систему уравнений:

$$\left\{ \begin{array}{l} \sum_{k=0}^n \frac{d}{dt} (L_{p-1,k} - L_{p,k})i_k + i_{p-1}R_{p-1} - i_p R_p = 0 \\ \sum_{k=0}^n i_k = i \end{array} \right. , \quad (18)$$

где $p=1,2,\dots,n$; i – суммарный ток.

При синусоидальном токе в проводнике потокосцепление Ψ_k и токи i_0, i_1, \dots, i_n в элементарных проводниках также изменяются по гармоническому закону. Поэтому можно воспользоваться символическим методом расчета токов в элементарных проводниках. Запишем систему (18) в символическом виде:

$$\left\{ \begin{array}{l} j\omega \sum_{k=0}^n (L_{p-1,k} - L_{p,k}) \dot{I}_k + I_{p-1} R_{p-1} - I_p R_p = 0 \\ \sum_{k=0}^n \dot{I}_k = \dot{I} \end{array} \right. , \quad (19)$$

где j – мнимая единица, $p=1,2,\dots,n$; \dot{I} – суммарный ток. Система уравнений (19) состоит из $n+1$ уравнений и содержит $n+1$ неизвестных токов \dot{I}_k . По вещественным и мнимым составляющим токов в элементарных проводниках можно вычислить модуль тока в них:

$$|\dot{I}_k| = \sqrt{[\operatorname{Re}(\dot{I}_k)]^2 + [\operatorname{Im}(\dot{I}_k)]^2}. \quad (20)$$

При решении системы (19) в качестве суммарного тока \dot{I} можно принимать амплитудное значение. Тогда найденные модули токов $|\dot{I}_k|$ будут иметь амплитудные значения. Если же понимать суммарный ток как действующий, то и найденные модули токов будут действующими.

Плотность тока в каждом элементарном проводнике можно определить по формулам:

- комплексная плотность токов:

$$\dot{\sigma}_k = \frac{\dot{I}_k}{S_k}; \quad (21)$$

- модуль плотности токов:

$$\sigma_k = \left| \dot{\sigma}_k \right| = \frac{|\dot{I}_k|}{S_k}, \quad (22)$$

где S_k – площадь сечения k -го элементарного проводника. В зависимости от задания суммарного тока \dot{I} в качестве амплитудного или действующего, плотность тока σ_k также будет амплитудной или действующей.

Заключение

Предложенная модель актуальна для моделирования распределения плотности тока по сечению исследуемого объекта в ГРВ-процессах. Новизной модели является учет неоднородности проводящего материала при расчете распределения плотности тока по сечению. В результате проведения расчетов в соответствии с предложенной моделью можно определить распределение комплексной плотности токов по сечению неоднородного проводника произвольной формы.

Литература

1. Коротков К.Г. Основы ГРВ биоэлектрографии. СПб: СПбГИТМО, 2001.
2. Ордабаев Б.Б., Исаков К.М., Ры. Алма-Ата: Казахский научно исследовательский институт кардиологии, 2000.
3. Рубин А.Б. Теоретическая биофизика. В 2 т. М.: Наука, 2004.
4. Гнездецкий В.В. Обратная задача ЭЭГ и клиническая электроэнцефалография (картирование и локализация источников электрической активности мозга). М.: МЕД-пресс-информ, 2004.
5. Зевеке Г.В., Ионкин П.А., Нетушил А.В. Основы теории цепей. М.: Энергия, 1975.
6. Новгородцев А.Б., Теория электромагнитного поля, конспект лекций. СПб: СПбГТУ, 1994.
7. Мансуров Н.Н., Попов В.С. Теоретическая электротехника. М.: Энергия, 1966.
8. Калантаров П.Л., Цейтлин Л.А. Расчет индуктивностей. Справочная книга. Л.: Энергоатомиздат, 1986.
9. Свешников А.Г., Тихонов А.Н. Теория функции комплексной переменной. М.: Физматлит, 1999.

10. Коган А.Б., Косицкий Г.И. Физиология человека и животных 1,2. М.: Высшая школа, 1984.
11. Новгородцев А.Б. Теория электромагнитного поля, конспект лекций. СПб: СПбГТУ, 1994.
12. Иванов Г.Г., Николаев Д.В. Возможности оценки общей воды и внеклеточной жидкости методом биоимпедансной спектроскопии: современные подходы к решению актуальной проблемы. // Вестник Российского университета дружбы народов. Кафедра госпитальной терапии. РУДН. Серия «Медицина». 1998. № 1. С. 213–226.
13. Иванов Г.Г., Сыркин А.Л., Дворников В.Е. Мультичастотный сегментарный биоимпедансный анализ в оценке изменений объёма водных секторов организма. М.: Московская медицинская академия им. И.М. Сеченова, Российский университет дружбы народов, НТЦ «Медасс» , 2000.
14. Афонин П.Н., Афонин Д.Н., Бегун П.И., Пахарьков Г.Н. Исследование проницаемости кожных капилляров импедансометрическим методом. // Тезисы докладов II Съезда биофизиков России, Москва, 23-27 августа 1999. М., 1999. Т. 2. С. 642–643.

МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЯ ПЛОТНОСТИ ТОКА В СЛОЖНОМ НЕОДНОРОДНОМ ПРОВОДНИКЕ.

ЧАСТЬ 2

А.Ю. Гришенцев

Научный руководитель – доктор технических наук, профессор К.Г. Коротков

В статье рассмотрен вариант реализации математической модели распределения плотности тока по сечению неоднородного проводника произвольного сечения.

Введение

Задачей данной работы является численное моделирование распределения плотности тока по сечению неоднородного проводника произвольного сечения на основании математической модели, изложенной в части 1 [см. настоящий сборник], а также создание удобного для пользователя интерфейса с возможностью графического ввода данных.

Реализация алгоритма



Рис. 1. Блок-схема алгоритма

На рис. 1 приведен краткий алгоритм программы. В блок-схеме алгоритма использованы ссылки на формулы, приведенные в первой части статьи. Математическая модель реализована как приложение Windows NT в среде программирования Builder C++6 [1, 2], рабочее название DC (Density current). На рис. 2 приведен возможный вариант вида программы для моделирования распределения плотности тока. Проверка программы на правильность проводилась при помощи нахождения суммы токов по сечению и сопоставления с исходным. Также проводился анализ полученной картины плотности тока по сечению для различных проводников.

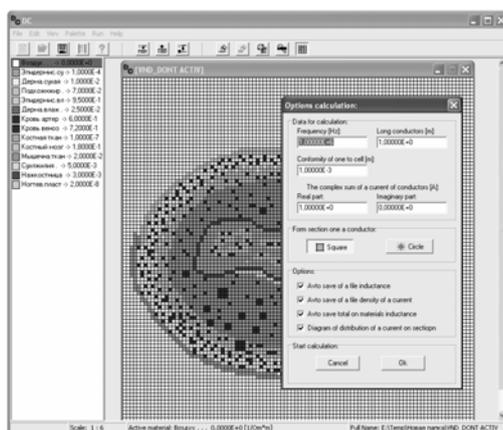


Рис. 2. Интерфейс программы

Пример моделирования плотности распределения тока по сечению пальца человека при исследовании на ГРВ-приборе

Рассмотрим пример работы с программой DC на модели плотности распределения тока по сечению пальца человека, помещенного на высоковольтный электрод ГРВ-прибора [3].

Поставив в данной модели задачу получения картины плотности тока в зависимости от перспирации и капиллярного просвета [4, 5], не будем ее усложнять расчетом параметров внешней цепи (ГРВ-разрядная схема). Примем суммарный ток $I=1\text{A}$ полностью вещественным, без мнимой части. Частота тока $f=1\text{МГц}$. Элементарные площадки разбиения выберем квадратные с длиной стороны $b=0.17\text{ мм}$. Электрические параметры тканей человека, входящих в модель, указаны в табл. 1.

При активизации раздражителем высшей нервной деятельности (ВНД) у человека нарастает давление крови за счет роста активности сердечных сокращений, как следствие, увеличивается просвет капилляров, кровенаполнение некоторых органов и тканей, преимущественно тех, которые непосредственно связаны с раздражителем. Также увеличивается кожная перспирация. Эти факторы вызывают увлажнение и набухание верхних слоев кожи, особенно эпидермиса. Благодаря поверхностному эффекту плотность тока при высоких частотах сконцентрирована именно в кожном слое тканей. Увлажнение и набухание кожи влечет за собой рост удельной объемной электропроводности, уменьшение глубины проникновения δ (1). На основании вышесказанного можно сделать вывод о том, что при активизации ВНД происходит увеличение поверхностного эффекта для токов внешнего воздействия на человека.

На рис. 3 изображено поперечное сечение четвертого (безымянного) пальца [6]. Данное изображение выполнено в графическом интерфейсе программы DC, имеет в своем составе 5053 элементарных проводника квадратного сечения. Эта модель используется для расчетов распределения плотности тока по сечению пальца при ВНД в спокойном и возбужденном состоянии.

Наименование тканей		Примерная толщина слоя, мм	Проводимость 1/(Ом·м)	Источники
Сухая кожа	Эпидермис	0,07–2,5	1,00Е-04	[4, 5, 7–9]
	Дерма	0,5–5,0	1,00Е-02	
	Подкожная жировая клетчатка		7,00Е-02	
Влажная кожа	Эпидермис	0,07–2,5	9,50Е-01	
	Дерма	0,5–5,0	1,50Е-02	
	Подкожная жировая клетчатка		7,00Е-02	
Кровь	Артериальная		6,00Е-01	
	Венозная		7,20Е-01	
Кость	Костная ткань		1,00Е-07	
	Костный мозг		1,80Е-01	
Прочие ткани	Мышцы		2,00Е-02	
	Сухожилия		5,00Е-03	
	Надкостница	0,3–4,5	3,00Е-03	
	Ногтевая пластинка	0,2–0,5	2,00Е-08	

Таблица 1. Электрические параметры тканей человека

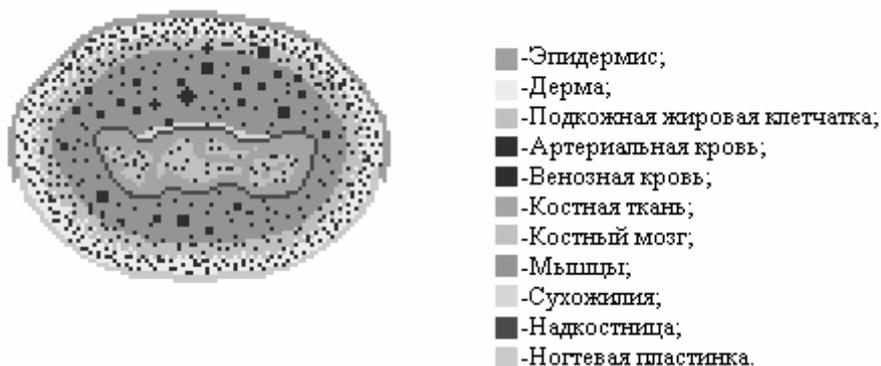


Рис. 3. Поперечное сечение безымянного пальца человека

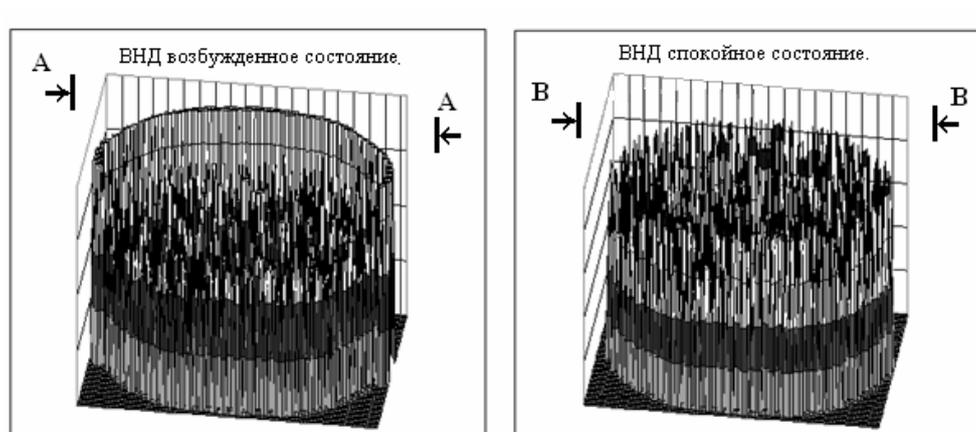


Рис. 4. Вытеснение тока при активизации ВНД на кожную поверхность

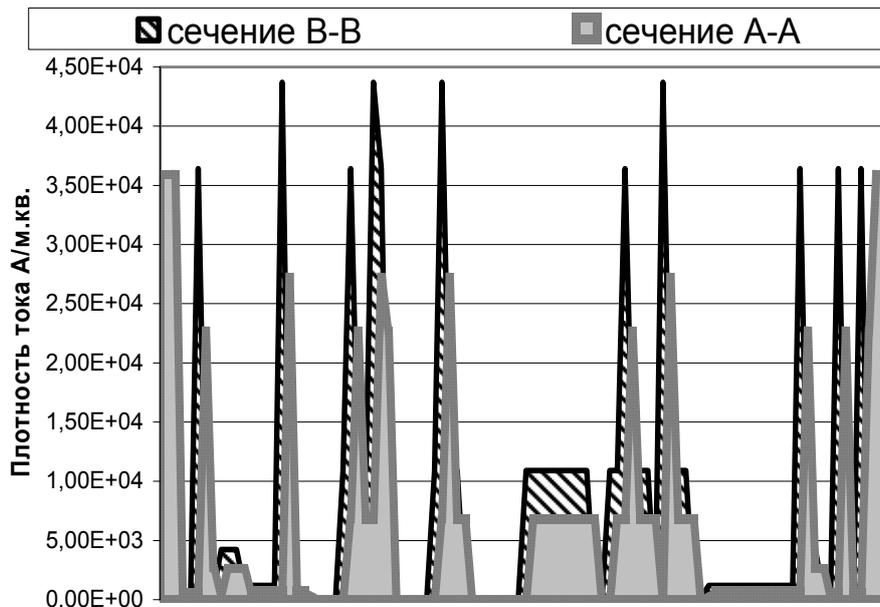


Рис. 5. Распределение плотности токов по сечениям

Результаты моделирования распределения модуля плотности тока по сечению представлены на рис. 4, 5. Рис. 4 наглядно отображает вытеснение тока при активизации ВНД на кожную поверхность. Такое перераспределение тока вызывает изменение картины скользящего разряда на поверхности ГРВ электрода. Отдельно рассмотрена плотность токов по сечениям А-А и В-В (рис. 5). Очевидно, что для сечения А-А, когда ВНД находится в возбужденном состоянии, наибольшая плотность тока сосредоточена по краям сечения, в отличие от случая ВНД в спокойном состоянии. Сильные выбросы соответствуют сосудам и капиллярам, а также тканям костного мозга.

Возможные области применения и перспективы программного продукта DC

Программа DC может применяться не только к ГРВ-оборудованию, но и для моделирования плотности распределения тока в любых проводниках, относительная магнитная проницаемость которых близка к единице ($\mu \approx 1$). DC позволяет подбирать оптимальное, с точки зрения использования, сечение проводников для переменных токов, например, волноводов или длинных линий.

В перспективе можно дополнить DC модулями, расширяющими область применения:

- расчета эффекта близости между произвольным числом проводников с различными по фазе токами;
- расчета полной индуктивности проводника, взаимной индуктивности между проводниками и любыми частями сечения проводников;
- построения картины плоскопараллельного поля, нахождения значения вектора напряженности электрического поля E и вектора B индукции магнитного поля в окрестностях проводников;
- теплового расчета модели с учетом поверхностного эффекта и эффекта близости для установившихся режимов;
- расчета электродинамических сил между проводниками;
- графической интерпретации полученных результатов.

Заключение

В результате работы была создана программа, позволяющая моделировать плотность распределения тока по сечению неоднородного проводника произвольного сечения. Программу можно использовать для моделирования распределения плотности тока по биологическим объектам.

Литература

1. Архангельский А.Я. Программирование в C++Builder 6, 2-е изд. М.: ООО «Бином-Пресс», 2005.
2. Архангельский А.Я., Тагин М.А. Приемы программирования в C++Builder 6 механизмы Windows, сети. М.: ООО «Бином-Пресс», 2005.
3. Коротков К.Г. Основы ГРВ биоэлектрографии. СПб: СПбГИТМО, 2001.
4. Иванов Г.Г., Сыркин А.Л., Дворников В.Е. Мультичастотный сегментарный биоимпедансный анализ в оценке изменений объема водных секторов организма. М.: Московская медицинская академия им. И.М. Сеченова, Российский университет дружбы народов, НТЦ «Медасс», 2000.
5. Коган А.Б., Косицкий Г.И. Физиология человека и животных 1,2. М.: Высшая школа, 1984.
6. Электронный атлас анатомии человека. М.: ИД «Равновесие», 2005.
7. Иванов Г.Г., Николаев Д.В. Возможности оценки общей воды и внеклеточной жидкости методом биоимпедансной спектроскопии: современные подходы к решению актуальной проблемы. // Вестник Российского университета дружбы народов. Кафедра госпитальной терапии. РУДН. Серия "Медицина". 1998. № 1. С. 213–226.
8. Афонин П.Н., Афонин Д.Н., Бегун П.И., Пахарьков Г.Н. Исследование проницаемости кожных капилляров импедансометрическим методом. // Тезисы докладов II Съезда биофизиков России, Москва, 23-27 августа 1999. М., 1999. Т.2. С. 642–643.
9. Гнездецкий В.В. Обратная задача ЭЭГ и клиническая электроэнцефалография (картирование и локализация источников электрической активности мозга). М.: МЕДпресс-информ, 2004.

ПРЕДВАРИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ ПО ИЗМЕНЕНИЮ ХАРАКТЕРИСТИК ГАЗОРАЗРЯДНОГО СВЕЧЕНИЯ ВОЛОС ПРИ РАЗЛИЧНЫХ УСЛОВИЯХ

А.Ю. Гришенцев, Е.Н. Петрова

Научный руководитель – доктор технических наук, профессор К.Г. Коротков

Изучены характеристики свечения волос в зависимости от конфигурации исследуемого пучка волос и их изменение при воздействии на волосы ультрафиолетового облучения. Выдвинута гипотеза о прыжковой электронной проводимости внутренних белковых структур волоса.

Введение

Задача исследования волос и анализа изменения их состояния при различных воздействиях является актуальной для разных областей науки [1, 2]. Техника динамической газоразрядной визуализации (ГРВ) позволяет проводить экспресс-анализ состояния биологических объектов [3] и выявлять изменения электронной проводимости исследуемого объекта [4, 5].

Проведенные ранее исследования волос методом ГРВ позволили выдвинуть гипотезу, что волосы обладают свойствами электропроводности [4, 6]. Для понимания процессов, происходящих в волосах и установления факторов, влияющих на их свойства, необходимы дальнейшие исследования и проведение дополнительных экспериментов по обоснованию выдвинутых гипотез.

Задачи и методика исследований

Задачами исследования являлись определение влияния конфигурации исследуемого пучка волос на параметры свечения, а также выявление различий характеристик газоразрядного свечения волос при воздействии на волосы ультрафиолетового облучения.

Исследования волос проводились методом ГРВ. В ходе проведения экспериментов было замечено, что характер свечения волос меняется в зависимости от характеристик верхней части пучка волос, выступающей в разрез тefлоновой трубки. На рис. 1 приведены два варианта расположения волос в тefлоновой трубке: размещение верхней части волос в форме «кисточки» и в форме «петли».

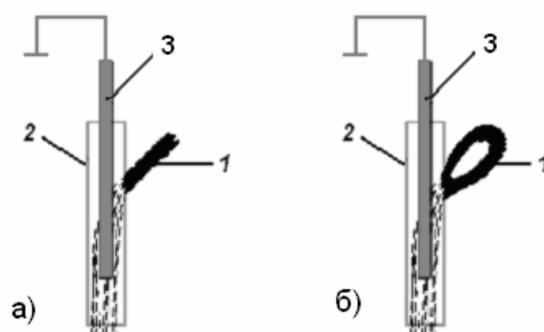


Рис. 1. Варианты размещения пучка волос в тefлоновой трубке: а – размещение пучка волос в виде «кисточки», б – размещение пучка волос в виде «петли»: 1 – образец волос, 2 – тefлоновая трубка, 3 – заземляющий электрод

При проведении экспериментов по влиянию формы выступающей части пучка волос на параметры свечения образец волос длиной 120 мм помещался в тefлоновую трубку в виде «петли» и устанавливался на электрод ГРВ-камеры. При данной конфи-

гурации пучка проводились съемки десяти серий газоразрядного свечения волос (по 5 секунд каждая). Затем «петля» разрезалась, и проводились съемки 10 серий при выступающей части пучка в виде «кисточки». В некоторых экспериментах проводилось несколько измерений при каждой форме выступающей части пучка волос.

Второй этап экспериментов включал в себя исследование влияния длины выступающей части пучка в виде «кисточки» на параметры свечения волос. Для этого брался образец волос длиной 50–70 мм и помещался в тефлоновую трубку в виде «кисточки». При исходной длине пучка волос снималось 10 avi-серий, по которым при обработке данных проводилось усреднение параметров свечения. Затем пучок волос постепенно укорачивался со стороны выступающей части до длины 35 мм. При каждом изменении длины образца волос снималось по 10 avi-серий.

Также в ходе работы были проведены предварительные эксперименты по выявлению влияния на характеристики свечения волос ультрафиолетового облучения. Для этого пучок волос стандартного размера (длина 55 мм и вес 0,45 г) помещался в тефлоновую трубку, закреплялся на заземляющий электрод и помещался на оптическое окно ГРВ-камеры. Область распространения разряда и нижняя часть пучка волос (приблизительно 7–9 мм) затемнялись светонепроницаемой пластиковой крышкой и плотной черной материей. Воздействие ультрафиолета на верхнюю выступающую часть пучка волос производилось ультрафиолетовой лампой с длиной волны 254 нм.

В ходе эксперимента снималось по 5 avi-серий (100 кадров в каждой) с исходными волосами, под действием ультрафиолета и после воздействия ультрафиолета.

Результаты исследований

В ходе проведенной работы были построены зависимости параметров свечения волос от конфигурации выступающей части пучка. Гистограмма изменения интенсивности свечения волос при разрезании «петли» приведена на рис. 2. В данном эксперименте при каждой форме выступающей части пучка было проведено по три измерения. Как видно из рисунка, при разрезании «петли» значения интенсивности свечения волос уменьшаются. Снижение параметров свечения при разрезании «петли» наблюдалось для большинства исследованных образцов волос, и в некоторых случаях уменьшение интенсивности достигало 50%.

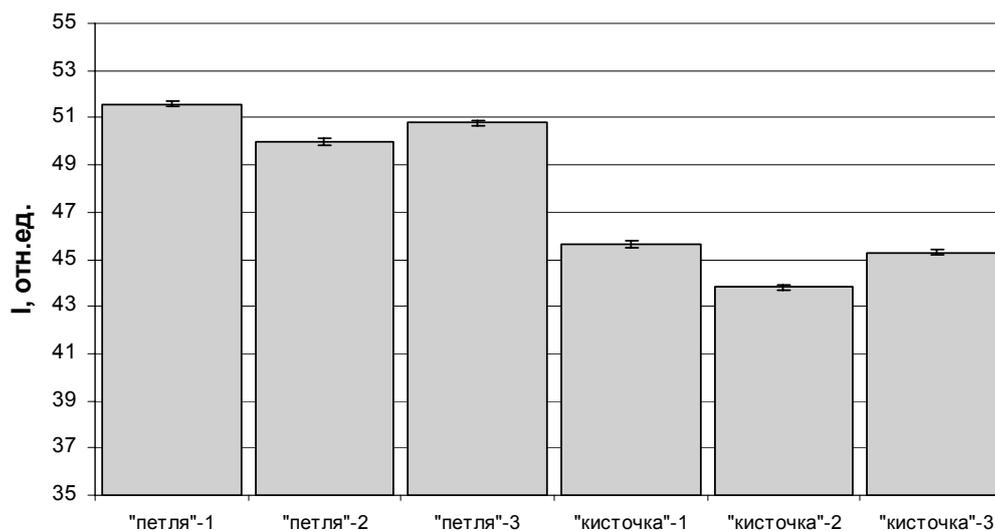


Рис. 2. Интенсивность свечения волос при двух вариантах размещения выступающей части пучка волос

Результаты одного из экспериментов по определению влияния длины выступающей части пучка волос в форме «кисточки» на параметры свечения волос приведены на рис. 3. Результаты представлены в виде гистограммы зависимости интегральной интенсивности свечения от длины пучка волос. При проведении данного эксперимента было снято три серии снимков (по 10 avi-фильмов каждая) при исходной длине пучка волос, затем волосы были сняты с заземляющего электрода и установлены вновь. На рис. 3 сериям снимков, произведенных после перестановки волос, соответствуют данные, помеченные*. Перестановка образца волос проводилась, чтобы убедиться, что незначительные различия в положении волос на электроде ГРВ-камеры (например, при шевелении пучка волос при уменьшении длины «хвоста») не вносят существенных изменений в характеристики свечения волос. После перестановки волос было произведено две серии снимков, после чего с выступающей части пучка было срезано 30 мм волос, и длина образца волос уменьшилась до 35 мм. На коротком пучке волос было снято три серии снимков. Как видно из представленных результатов, интенсивность свечения волос практически не меняется при перестановке пучка и существенно возрастает при уменьшении длины выступающей части пучка волос.

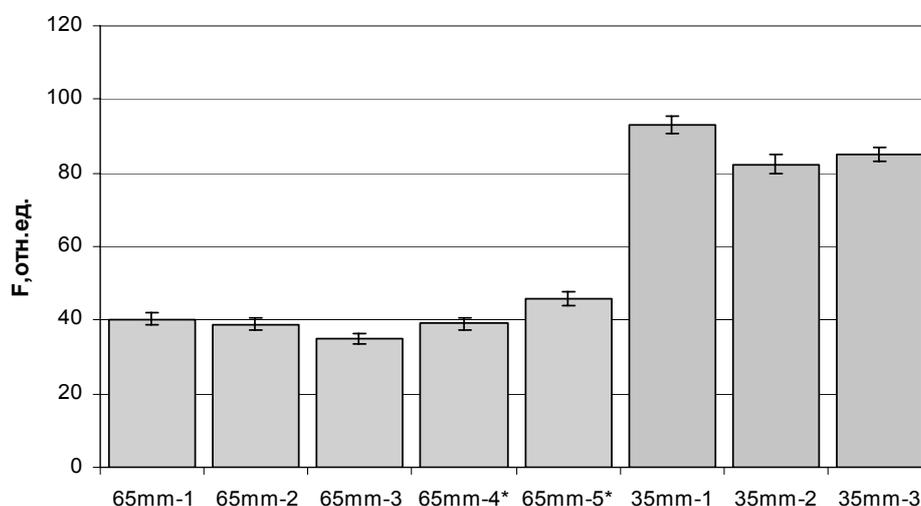


Рис. 3. Гистограмма интегральной интенсивности свечения волос в зависимости от длины исследуемого пучка (* соответствует сериям эксперимента, снятым после перестановки образца волос на электроде ГРВ камеры)

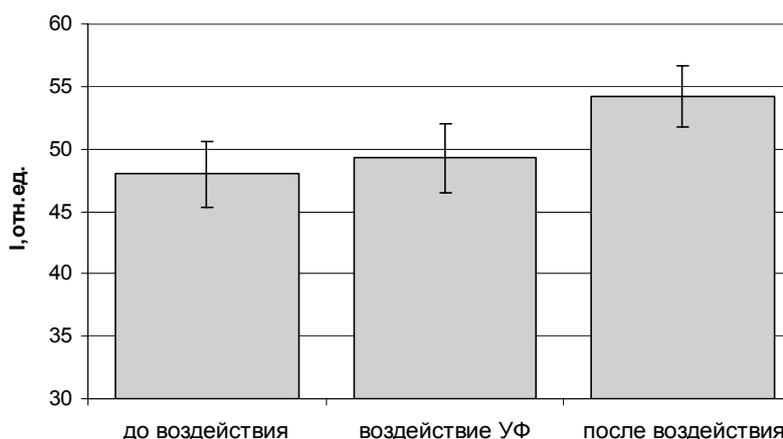


Рис. 4. Гистограмма интенсивности свечения волос до, во время и после воздействия ультрафиолетового излучения

Результаты одного из предварительных экспериментов по выявлению влияния ультрафиолетового облучения волос на параметры газоразрядного свечения приведены на рис. 4. При исследовании свечения образца «свежих» волос, срезанных за 18 часов

до проведения эксперимента, наблюдалось незначительное увеличение интенсивности свечения при воздействии ультрафиолета и дальнейшее увеличение сигнала после прекращения воздействия.

Обсуждение

Приведенные экспериментальные данные по изменению величины регистрируемого ГРВ-сигнала в зависимости от длины и конфигурации верхней части исследуемого пучка волос позволяет выдвинуть гипотезу о превалирующей роли емкостной проводимости в протекании тока по биологическим тканям волоса. Полученные результаты могут быть объяснены перераспределением емкостных токов во внутренних структурах волоса при изменении параметров пучка относительно точки заземления (рис. 1). Увеличение же интенсивности свечения после облучения волос ультрафиолетом может быть обусловлено переходом электронов в молекулах на более высокий энергетический уровень [7, 8], что увеличивает электронную проводимость волос. Перечисленные факторы позволяют выдвинуть гипотезу о прыжковой проводимости [7] внутренних структур волоса, что требует дальнейшей проверки. Кроме того, влияние характеристик внешней части пучка волос на ГРВ-сигнал открывает новые перспективы изучения влияния на волосы различных факторов (путем воздействия на внешнюю часть пучка).

Заключение

Изучено изменение параметров газоразрядного свечения волос в зависимости от конфигурации исследуемого пучка волос. Получены предварительные результаты по увеличению интенсивности свечения волос, и, следовательно, их проводимости под воздействием ультрафиолетового излучения. Выдвинута гипотеза о прыжковой проводимости внутренних молекулярных структур волоса. Разработанный подход открывает перспективы изучения влияния на волосы различных процедур и препаратов.

Литература

1. Clarence R. Robbins. Chemical and Physical Behavior of Human Hair. Springer-Verlag, New York, 1994. 391 p.
2. Скальный А.В., Быков А.Т. Эколого-физиологические аспекты применения макро- и микроэлементов в восстановительной медицине. Оренбург: РИК ГОУ ОГУ, 2003. 198 с.
3. Муромцев Д.И. Практика применения динамической ГРВ-графии в медико-биологических исследованиях. // Наука. Информация. Сознание: тезисы VIII международного конгресса по ГРВ биоэлектрографии. СПб, 2004. С. 140–142.
4. Vainshelboim A.L., Hayes M.T., Korotkov K., Momoh K.S. Observing the behavioral response of human hair to a specific external stimulus using dynamic gas discharge visualization. *J. Cosm. Science*. Princeton, New Jersey. 2004. Vol. 55. P. 91–104.
5. Vainshelboim A.L., Hayes M.T., Korotkov K., Momoh K.S. GDV Technology Applications for Cosmetic Sciences. IEEE 18th Symposium on Computer-Based Medical Systems (CBMS 2005). Dublin, Ireland, June 2005.
6. Vainshelboim A.L., Hayes M.T., Korotkov K., Momoh K.S. Electric and magnetic fields and electron channelling in human hair. Podium Proceedings. IFSCC International Conference. Florence: CEC Editor. September 2005. P. 137–171.
7. Рубин А.Б. Биофизика. М: Наука. 2004. Том 1. 463 с.
8. Самойлов В.О. Элементы квантовой биофизики. СПб: Изд-во СПбГТУ. 2001. 44 с.

**РАЗРАБОТКА ПОЛЬЗОВАТЕЛЬСКОЙ КОНСОЛИ
ДЛЯ ПРОГРАММНОГО КОМПЛЕКСА МОНИТОРИНГА
И АУДИТА БАЗ ДАННЫХ QDAS****С.В. Антонов****Научный руководитель – кандидат технических наук, доцент Д.И. Муромцев**

В рамках работы произведена разработка клиентской части для программного комплекса мониторинга и аудита баз данных QDAS, предоставляющей возможности по настройке политик безопасности и конфигурации сервера QDAS. Помимо этого, разработанная консоль предоставляет удобный пользовательский интерфейс для просмотра и анализа накопленных в системе данных.

Введение

В настоящее время практически на всех крупных предприятиях внедрены системы автоматизации производственных и бизнес-процессов. Это системы автоматизированного проектирования, поддержки жизненного цикла продуктов, автоматизированного документооборота и т.д. Подобные комплексы в процессе своей работы накапливают и манипулируют огромными объемами данных, причем с данными одновременно может работать множество пользователей. По этой причине в таких комплексах для хранения информации применяются специальные хранилища, называемые базами данных, а для доступа к ним – системы управления базами данных (СУБД).

Современные СУБД, помимо структурированного хранения данных, предоставляют множество дополнительных сервисов, таких как параллельная обработка запросов, поддержка целостности данных, исполнение программного кода и т.д. Одним из наиболее важных сервисов является контроль наличия прав на доступ к данным и выполнение операций. В настоящее время большинство СУБД позволяют настроить права доступа определенного пользователя или группы пользователей для всех объектов базы данных. Однако, учитывая темпы роста и усложнения современных систем, а также то, что все больше стратегически важной информации переводится в электронный вид, становится актуальным вопрос о контроле над санкционированным доступом к информации, а также сборе статистических сведений об обращении к объектам баз данных.

В связи с этим в последние несколько лет активно разрабатываются комплексы для мониторинга и аудита различных типов баз данных. Одним из таких комплексов, поддерживающим такие СУБД, как Oracle, MS SQL и Sybase, является Quest Database Auditing and Security (QDAS). Центральным узлом данного комплекса является QDAS-сервер. Сервер представляет собой набор автономных модулей, поддерживающих функционирование на платформах Sun Solaris, Linux Red Hat, MS Windows NT-XP. Он предназначен исключительно для перехвата, обработки и сохранения сетевого трафика с максимальным быстродействием и не содержит пользовательского интерфейса, или иных средств администрирования и постобработки. В связи с этим в рамках работы была спроектирована и разработана пользовательская консоль, предоставляющая следующие функциональности:

- настройку параметров работы сервера анализа и политик сбора информации о транзакциях, осуществляемых в организации;
- вывод накопленных в системе данных в удобном для пользователей виде;
- одновременную работу нескольких пользователей;
- предотвращение несанкционированного доступа к хранимым в системе данным;

- одновременную работу пользователя с несколькими серверами анализа;
- «дружелюбный» пользовательский интерфейс.

Анализ средств сопряжения консоли с комплексом

К основной функциональности, которую должна обеспечивать консоль, можно отнести вывод накопленных в системе данных, настройку сервера анализа и своевременное реагирование на изменения в системе. Последнее требование обусловлено тем, что разработанная консоль является многопользовательской системой, и при изменении данных одним пользователем эти изменения должны быть отображены остальным пользователям системы.

В соответствии с указанными требованиями консоль должна обеспечивать удобные в использовании и надежные механизмы для настройки системы и вывода и анализа накопленной в системе информации. На рис. 1 показана общая схема комплекса QDAS в рабочей среде.

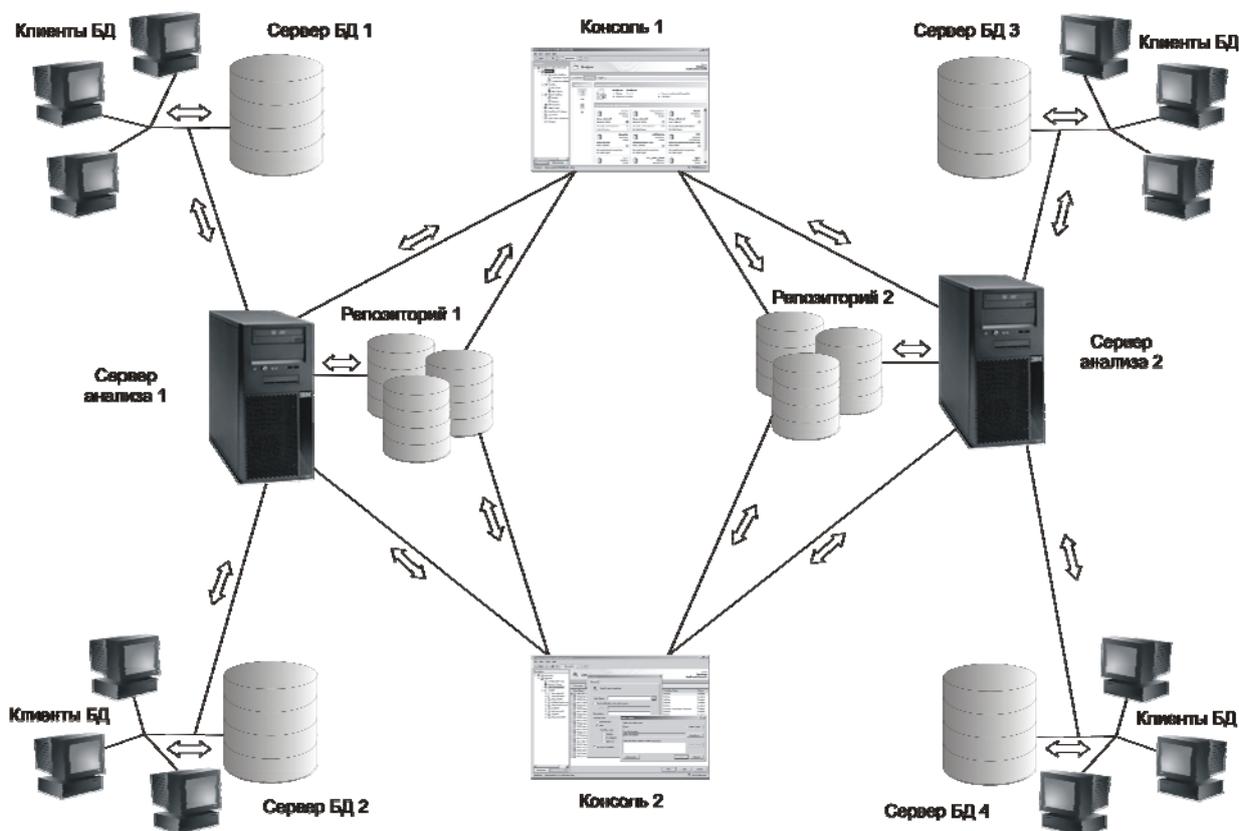


Рис. 1. Структурная схема комплекса QDAS

В корпоративной сети имеется несколько серверов баз данных (Сервер БД1 – Сервер БД4), к каждому из которых подключено множество клиентов. Сервер анализа подключен к каналу передачи данных между ними и перехватывает весь идущий трафик. Попадая на сервер анализа, данные подвергаются разбору в соответствии с политиками, установленными на сервера анализа. После разбора и анализа нужные данные передаются в репозитории. Следующей задачей является извлечение данных из репозитория, их отображение и анализ. Эти задачи возлагаются на пользовательскую консоль. Как видно из схемы, к одному серверу анализа может быть одновременно подключено несколько консолей, и в то же время одна консоль позволяет настраивать и анализировать данные с нескольких серверов анализа.

В схеме возможно построение стандартной трехзвенной архитектуры, в которой клиент (консоль в данном случае) взаимодействует с базой данных (репозиторий в данном случае) только через сервер приложений, роль которого в системе может выполнять сервер анализа. Однако такой подход неприемлем, поскольку сервер анализа не должен быть нагружен задачами, не относящимися к обработке трафика, так как эта задача является крайне ресурсоемкой и критической с точки зрения скорости обработки. В отдельные моменты времени объемы обрабатываемого трафика могут быть крайне велики, и сервер анализа будет не в состоянии обработать их «на лету». В такой ситуации данные будут сохранены в памяти и обработаны при первой возможности. Если же ресурсов недостаточно, то сервер перезаписывает старые пакеты данных новыми. В случае же, когда сервер анализа отвечает за передачу данных клиентам, вероятность потери данных, вследствие их перезаписи, существенно выше.

Также возможен вариант выделения отдельного сервера для обработки запросов клиентов и реализация консоли в виде тонкого клиента, т.е. без необходимости установки дополнительного программного обеспечения на клиентские машины. Однако такой вариант нецелесообразен, так как в системе не подразумевается работа большого количества пользователей. В соответствии с этим в консоли реализован модуль доступа к серверу анализа и модуль доступа к репозиторию, абстрагирующие остальные модули от реализации механизмов связи с сервером анализа и репозиторием, соответственно.

Для взаимодействия с сервером анализа были рассмотрены следующие технологии: ремоутинг, веб-сервисы, сокет, COM+, IPC, CORBA. При выборе технологии взаимодействия консоли и сервера анализа, ограничивающими факторами являются:

- совместимость с технологиями, использованными при разработке комплекса:
 - консоль разработана на платформе .NET;
 - сервер анализа разработан на основе стандартных библиотек языка C и функций операционной системы;
- кроссплатформенность сервера анализа;
- время, затраченное на реализацию взаимодействия на основе выбранной технологии.

В результате по общей совокупности характеристик была выбрана технология сокетов, так как она реализована на всех платформах [1], а в среде .NET для нее есть удобная оболочка, полностью реализующая функциональность сокетов и предоставляющая дополнительные методы для асинхронной работы. Помимо этого, важным фактором является то, что в сервере анализа на данный момент уже реализована функциональность, обеспечивающая передачу сообщения по сокетам.

Что касается репозитория, то он представляет собой РСУБД (реляционную систему управления базами данных) Oracle версии 10i. Система Oracle была выбрана в первую очередь по соображениям безопасности и производительности. Более подробную информацию по этому вопросу можно найти в [2]. Помимо этого, Oracle является наиболее гибкой системой в настоящее время и предоставляет множество механизмов для реализации программной логики прямо внутри базы данных [3, 4].

Для взаимодействия с Oracle, как и с другими РСУБД, требуется клиентское программное обеспечение, реализующее интерфейс с данным сервером баз данных. Сегодня под платформу .NET предлагается множество Oracle-клиентов, из которых наиболее сильные позиции занимают Oracle Data Provider for .Net (ODP.NET), Microsoft .Net Data Provider for Oracle и OraDirect .Net Data Provider. В наиболее полной мере возможности Oracle покрывает ODP.NET, однако он требует дополнительной установки и занимает около 200 Мб на диске. Что касается OraDirect .Net Data Provider, то он компактен (11 Мб) и в режиме ограниченной функциональности поддерживает все версии Oracle, начиная с версии 8.

В итоге для взаимодействия с сервером репозитория по общей совокупности характеристик был выбран Microsoft .Net Data Provider for Oracle – провайдер, в достаточной мере покрывающий функциональность сервера БД, а также обеспечивающий высокую производительность и имеющий поддержку со стороны производителя платформы разработки.

Разработка программной модели

Разработанная консоль представляет собой Windows-приложение, реализованное на платформе .NET. Программная структура консоли представлена на рис. 2.

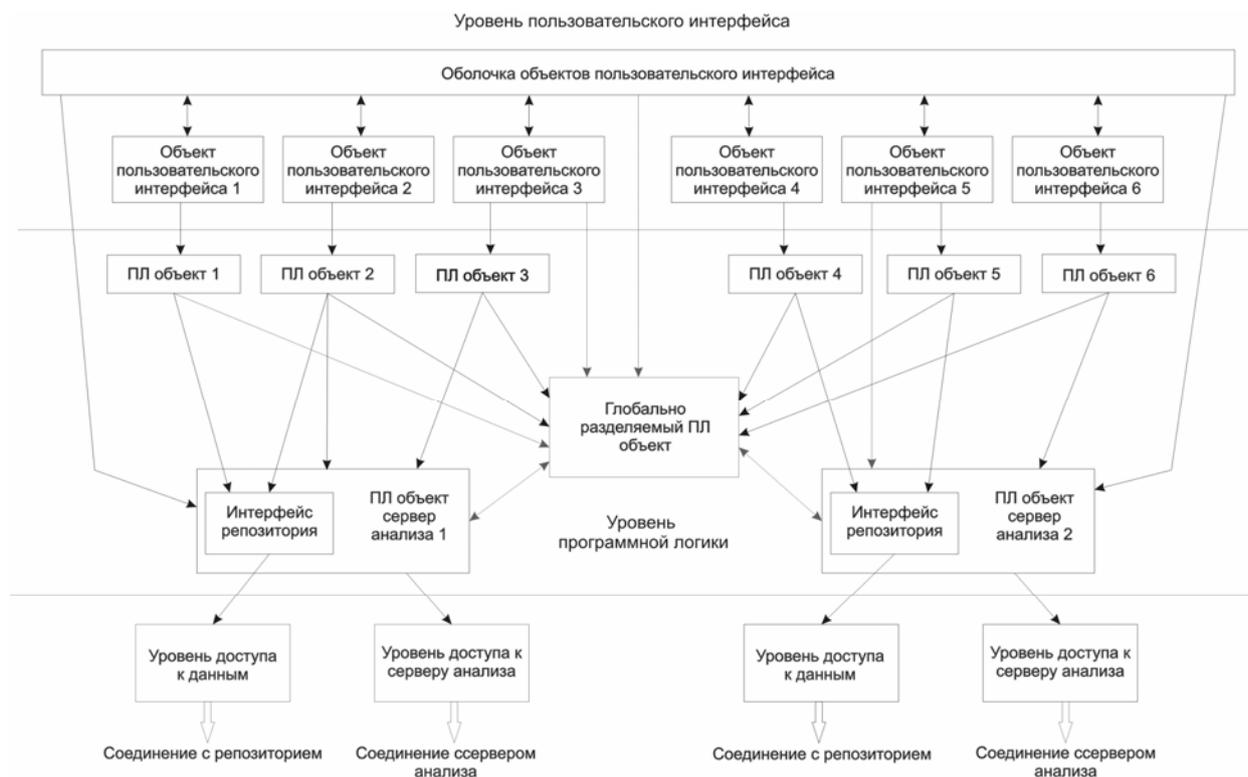


Рис. 2. Программная структура консоли

Логически консоль можно подразделить на четыре уровня: уровень пользовательского интерфейса, уровень программной логики, уровень доступа к данным (или уровень абстрагирования данных) и уровень доступа к серверу анализа. Такое деление обусловлено необходимостью такого выделения модулей системы, чтобы при глобальных изменениях системы как можно больше ее частей оставалось в исходном состоянии.

Уровень пользовательского интерфейса отвечает за прорисовку элементов управления консоли, отображение окон, вывод диалогов. На этом уровне также происходит создание объектов программной логики. Это обусловлено тем, что программный код данного уровня является стартовым, и, соответственно, при запуске программы управление передается именно ему. Уровень пользовательского интерфейса также отвечает за реакцию на действия пользователя, тем самым обеспечивая интерактивность работы. Помимо этого, он содержит функциональность, необходимую для оповещения пользователя о событиях, происходящих в системе. На этом уровне также выполняются основные операции, связанные с синхронизацией потоков [5] консоли.

Уровень программной логики отвечает за выполнение операций, связанных с бизнес-логикой системы, а также логикой работы самой консоли. Он является центральным

звеном консоли, так как связывает все остальные в единое целое и определяет четкие интерфейсы взаимодействия компонентов консоли. Также важной функцией уровня является инкапсуляция программой логики, не зависящей от взаимодействия консоли с окружающей средой. Такая инкапсуляция требуется, в первую очередь, для того, чтобы не переписывать программную логику при изменении в одном из модулей системы. Уровень является своего рода скелетом, на который крепятся остальные модули системы.

Уровень доступа к данным осуществляет передачу данных между консолью и репозиторием. В его обязанности входит преобразование типов данных, заполнение коллекций из курсоров, открытие/закрытие соединений, управление транзакциями и блокировками. Реализация уровня зависит от используемой системы управления базами данных в качестве репозитория. Основной же его задачей является абстрагирование остальных уровней от деталей реализации репозитория.

Уровень доступа к серверу анализа абстрагирует консоль от методов взаимодействия с сервером анализа. Он содержит коды всех команд, механизмы асинхронного запуска операций, методы для декодирования и распознавания сообщений, приходящих с сервера анализа. На уровне доступа к серверу анализа также содержится функциональность для протоколирования операций обмена сообщениями.

Проектирование структуры консоли

С точки зрения предметной области консоль можно разделить на две основные части – пользовательскую и административную.

Пользовательская часть предоставляет возможности настройки встроенных политик мониторинга, а также конфигурирования произвольных пользовательских политик. Помимо этого, пользовательская консоль содержит множество отчетов, отображающих все данные, накопленные в системе, в структурированном виде.

Административная консоль предоставляет функциональность по конфигурированию сервера анализа, администрированию учетных записей пользователей комплекса, указанию параметров целевых серверов и др.

По реализуемой функциональности в консоли можно выделить следующие основные части: мониторинг соединений с базами данных, мониторинг по встроенным политикам, мониторинг по пользовательским политикам, аутентификация пользователей.

Мониторинг соединений с базами данных

Раздел мониторинга соединений с базами данных позволяет настраивать политику, в соответствии с которой определяется, какие соединения с базой данных считать допустимыми, а какие – нет. При настройке политики пользователь может указать сохранение всех команд перехваченных сессий в репозиторий. В дальнейшем сохраненные команды могут быть просмотрены и проанализированы администратором безопасности. Помимо этого, пользователь может указать получать извещение о сессии, не входящей в список допустимых сессий. Также важной особенностью является возможность указать автоматически прекращать все недопустимые сессии.

Помимо подраздела настроек, в данном разделе содержится два отчета по всем сессиям в системе. Один отчет содержит линейный список сессий, что позволяет производить фильтрацию и сортировку по всем параметрам сессий. Другой отчет (рис. 3) содержит все сессии в сгруппированном виде.

Группировка осуществляется по параметрам, однозначно определяющим сущность, породившую данную сессию. При выборе одной из строк в нижней таблице отображается список всех сессий с соответствующими параметрами. При желании пользователь может либо завершить выбранную сессию, либо сделать параметры, идентифи-

цирующие ее, допустимыми. Также у пользователя есть возможность просмотреть все команды, выполненные этой сессией, а также данные, переданные в соответствующую команду, и ошибки, вызванные ее выполнением.

Connection Report Quest Database Auditing and Security

Summary: All Connections

Select period: Last Year From: 1/12/2005 3:21:49 PM To: 1/12/2006 3:21:49 PM

Server Name: All Program: All

Client IP: All RDBMS User: All

Total	Server Name	Client IP	Program	RDBMS User
4	sbase	10.30.34.186	sqlplus	QDA5
1	sbase	10.30.34.186	sqlplus	SCOTT
10	sbase	10.30.33.11	sqlplus	SCOTT
1	sbase	10.30.33.13	TOAD	SCOTT
1	sbase	10.30.35.49	sqlplus	SCOTT
5	sbase	10.30.34.140	EXP	SCOTT
15	sbase	10.30.34.140	sqlplus	SCOTT
6	sbase	10.30.34.186	sqlplus	SCOTT

Start Time	Finish Time	Duration (hr:m:s)	Client OS User	Host Name	Response (sec)	Rows Processed	Rows Returned
1/12/2006 3:19...	1/12/2006 3:19...	00:00:32	SAntonov	PRODDMG-AN...	0.335	3	2
1/12/2006 12:1...	1/12/2006 12:2...	00:07:08	SAntonov	PRODDMG-AN...	0.130	3	2
1/12/2006 12:0...	1/12/2006 12:1...	00:06:56	SAntonov	PRODDMG-AN...	0.549	3	2
1/12/2006 11:4...	1/12/2006 12:0...	00:21:55	SAntonov	PRODDMG-AN...	0.321	4	2
1/11/2006 5:13...	1/11/2006 6:01...	00:47:54	SAntonov	PRODDMG-AN...	0.523	5	2
1/11/2006 5:08...	1/11/2006 5:12...	00:03:51	SAntonov	PRODDMG-AN...	0.353	3	2
1/11/2006 4:58...	1/11/2006 5:08...	00:10:40	SAntonov	PRODDMG-AN...	0.307	4	2
1/11/2006 1:46...	1/11/2006 4:58...	03:12:02	SAntonov	PRODDMG-AN...	0.332	3	2
1/11/2006 1:42...	1/11/2006 1:46...	00:03:37	SAntonov	PRODDMG-AN...	0.376	3	2

Рис. 3. Отчет по соединениям с базами данных

Мониторинг по встроенным политикам

Встроенные политики представлены следующими разделами: мониторинг привилегированных сессий, мониторинг ошибок авторизации, мониторинг отсутствия привилегий доступа. Каждый из разделов содержит средства для настройки соответствующей политики, а также подробные отчеты по данным, собранным соответствующей политикой. На рис. 4 представлен отчет по политике отсутствия привилегий доступа.

Данный отчет содержит записи обо всех отказах доступа за выбранный период. В каждой записи, помимо параметров, идентифицирующих сущность, которой было отказано в доступе, отображается запрос, породивший ошибку доступа, код и расшифровка ошибки, запрашиваемые объекты.

Политика мониторинга отсутствия привилегий доступа ведет отчетность по все случаям отказа какой-либо сессии в доступе к объекту базы данных. При настройке политики пользователь может указать, посылать ли извещение при отказе доступа, настроить порог, при котором посылается извещение, а также указать, сохранять ли записи обо всех отказах доступа или же сохранять запись только при превышении порога.

Политика мониторинга ошибок авторизации записывает информацию обо всех неудачных попытках регистрации на сервере баз данных. Параметры настройки и отчеты по большей части схожи с разделом политики мониторинга отсутствия привилегий доступа.

Политика мониторинга привилегированных сессий предоставляет информацию о том, какая сущность была подключена к базе данных в привилегированном режиме, когда и какие операции были выполнены в рамках соответствующей сессии. Параметры настройки и отчеты по большей части схожи с разделом политики мониторинга от-

сутствия привилегий доступа, однако пользователю предоставляется возможность выбирать, какие именно сессии считать привилегированными.

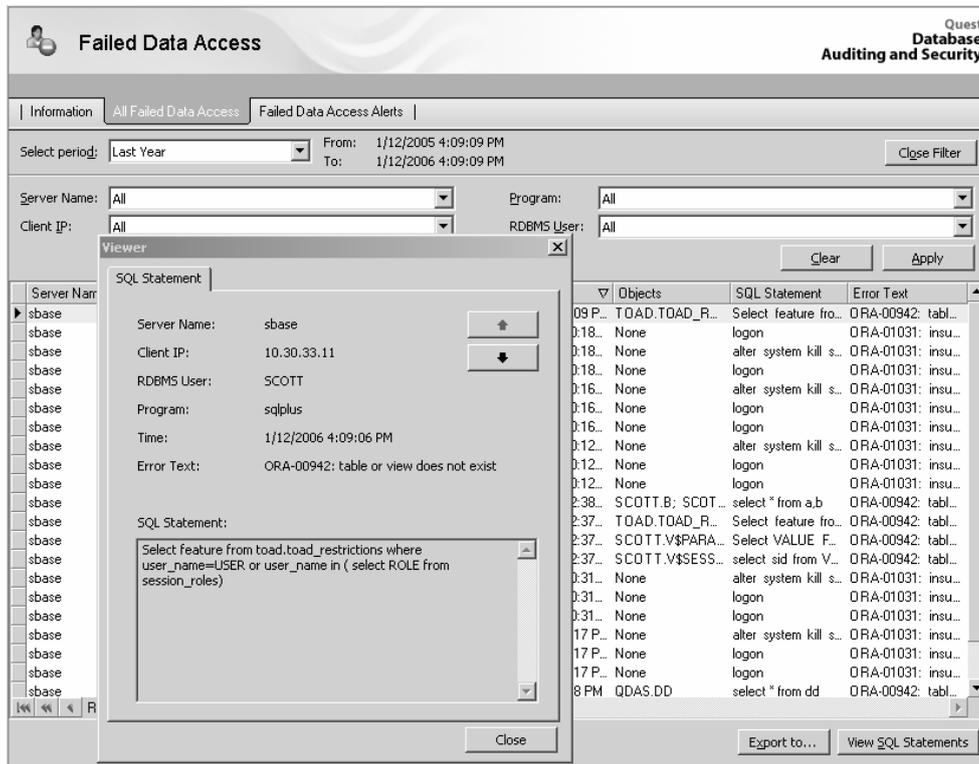
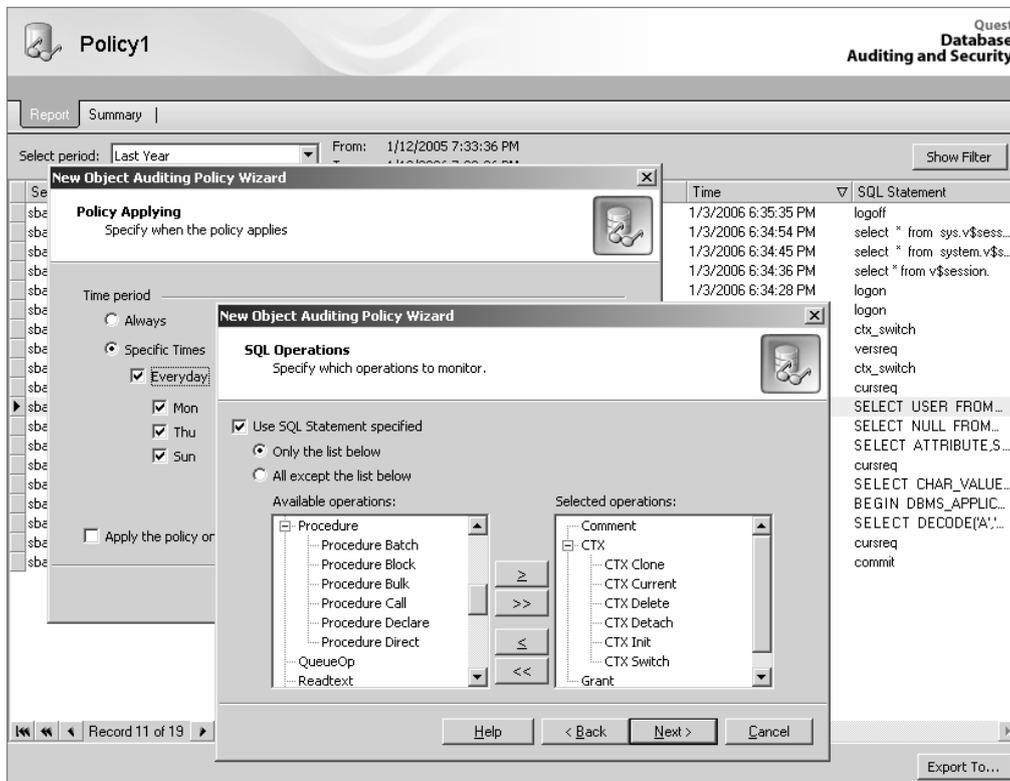


Рис. 4. Отчет по политике мониторинга отсутствия привилегий доступа

Мониторинг по пользовательским политикам



дения новых учетных записей применяется встроенная подсистема поиска пользователей, позволяющая находить как локальных, так и доменных пользователей.

Заключение

Разработанная консоль предоставляет пользователям удобный доступ ко всей функциональности комплекса QDAS, включая возможности конфигурирования режимов работы сервера, настройки политик сбора информации и вывода всех накопленных в системе данных в удобной для восприятия форме. Помимо этого, консоль предоставляет возможности по фильтрации и анализу выводимых данных, тем самым позволяя производить планирование политики безопасности в организации.

Литература

1. Гантер Д., Барнет С., Гантер Л. Интерация Windows NT и UNIX в подлиннике. СПб: БХВ, 1998. 464 с., ил.
2. Кайт Т. Oracle для профессионалов. Книга 1. Архитектура и основные особенности. / Второе издание. СПб: ДиаСофтЮП, 2004. 672 с.
3. Кайт Т. Oracle для профессионалов. Книга 2. Расширение возможностей и защита. / Второе издание. СПб: ДиаСофтЮП, 2004. 848 с.
4. Ферштейн С., Прибыл Б. Oracle PL/SQL для профессионалов. 3-е изд. СПб: Питер, 2004. 941 с.; ил.
5. Рихтер Д. Windows для профессионалов: создание эффективных Win32 приложений с учетом спецификации 64-разрядной версии Windows. М.: Русская Редакция, 2001. 752 с.; ил.

ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ СИСТЕМ АВТОМАТИЗИРОВАННОГО УЧЕТА НА ПРЕДПРИЯТИИ

Г.С. Александров, О.В. Елисеев, Н.С. Григорьева
(ОАО «Российский институт радионавигации и времени»), Д.В. Соловьев
Научный руководитель – кандидат технических наук, доцент Н.С. Кармановский

В статье представлен обзор существующих баз данных. Рассматриваются их достоинства и недостатки. Представляется пример реально разработанной и введенной в опытную эксплуатацию базы данных на конкретном предприятии.

Введение

Современная жизнь немыслима без управления. Для эффективного управления необходимо анализировать массу информации и в зависимости от анализа принимать то или иное решение. В современном мире высоких технологий существует большое количество систем обработки и анализа информации, созданных для оптимизации и упрощения управления, от которых во многом зависит эффективность работы любого предприятия или учреждения. Такая система должна обеспечивать получение общих и/или детализированных отчетов по итогам работы, позволять легко определять тенденции изменения важнейших показателей, обеспечивать получение информации, критической по времени, без существенных задержек, выполнять точный и полный анализ данных.

Несомненно, существенную по значимости среди совокупности всех этих систем нишу занимают всевозможные базы данных (БД).

Современные системы управления БД (СУБД) в основном являются приложениями Windows и прочих операционных систем и используют все возможности персональных компьютеров. Снижение стоимости высокопроизводительных ПК обусловило не только переход большинства предприятий на их использование, но также сделало программное обеспечение ПК в целом и СУБД, в частности, менее критичными к аппаратным ресурсам ПК и очень удобными в качестве систем обработки информации [1, 2].

Среди наиболее ярких представителей СУБД можно отметить Lotus Approach, Microsoft Access, Borland dBase, Borland Paradox, Microsoft Visual FoxPro, а также Microsoft SQL Server и Oracle, используемые в приложениях, построенных по технологии «клиент-сервер». Фактически у любой современной СУБД существует аналог, выпускаемый другой компанией, имеющий аналогичную область применения и возможности, любое приложение способно работать со многими форматами представления данных, осуществлять экспорт и импорт данных благодаря наличию большого числа конвертеров. Общепринятыми также являются технологии, позволяющие использовать возможности других приложений, например, текстовых процессоров, пакетов построения графиков и т.п., и встроенные версии языков высокого уровня (чаще – диалекты SQL и/или VBA) и средства визуального программирования интерфейсов разрабатываемых приложений. Поэтому уже не имеет существенного значения, на каком языке и на основе какого пакета написано конкретное приложение и, какой формат данных в нем используется. В одном ряду с «классическими» СУБД все чаще упоминаются языки программирования Delphi и Visual C++, которые позволяют быстро создавать необходимые компоненты приложений, критичные по скорости работы, которые трудно, а иногда невозможно разработать средствами «классических» СУБД. Современный подход к управлению базами данных подразумевает также широкое использование технологии «клиент-сервер».

Таким образом, на сегодняшний день разработчик не связан рамками какого-либо конкретного пакета, а в зависимости от поставленной задачи может использовать самые разные приложения.

Авторы статьи разработали систему автоматизированного учета измерительного и контрольно-поверочного оборудования, находящегося на балансе ОАО «РИРВ». Эта система представляет собой уникальную базу данных и систему управления этой базой. Цель

разработки – эффективный контроль за измерительным и контрольно-поверочным оборудованием, своевременный анализ информации о состоянии этого оборудования [3, 4].

Краткая характеристика программного обеспечения, используемого при создании СУБД

Рассмотрим более подробно программные продукты для работы с БД, а именно Visual FoxPro, Visual Basic, Visual C++, Access, SQL Server 2000, Delphi 7. Наиболее интересной чертой этих пакетов являются их большие возможности интеграции, совместной работы и использования данных, так как данные пакеты являются продуктами одного производителя, а также используют сходные технологии обмена данными.

Visual FoxPro отличается высокой скоростью, имеет встроенный объектно-ориентированный язык программирования с использованием xBase и SQL, диалекты которых встроены во многие СУБД. Имеет высокий уровень объектной модели. При использовании в вычислительных сетях обеспечивает как монопольный, так и отдельный доступ пользователей к данным. Применяется для приложений масштаба предприятия для работы на различных платформах.

Access входит в состав популярного пакета Microsoft Office. Основные преимущества: знаком многим конечным пользователям и обладает высокой устойчивостью данных, прост в освоении, может использоваться непрофессиональным программистом, позволяет готовить отчеты из БД различных форматов. Предназначен для создания отчетов произвольной формы на основании различных данных и разработки некоммерческих приложений.

Visual Basic – это универсальный объектно-ориентированный язык программирования, диалекты которого встроены в Access (да и вообще в весь пакет Microsoft Office), Visual FoxPro. Преимущества: универсальность, возможность создания компонентов OLE, невысокие требования к аппаратным ресурсам ЭВМ. Применяется для создания приложений средней мощности, не связанных с большой интенсивностью обработки данных, разработки компонентов OLE, интеграция компонентов Microsoft Office.

Visual C++ – один из наиболее мощных объектно-ориентированных языков программирования. Он обладает неограниченной функциональностью. Предназначен для создания компонентов приложений для выполнения операций, критичных по скорости.

Delphi7 – пожалуй, самый удобный и универсальный в работе на сегодняшний день объектно-ориентированный язык программирования. Также как и Visual C++, обладает безграничной функциональностью. Имеет поддержку обращений-запросов к SQL Server'у.

SQL Server – сервер баз данных, реализует подход «клиент-сервер» и взаимодействует с указанными пакетами. Главные достоинства: высокая степень защиты данных, мощные средства для обработки данных, высокая производительность. Область применения: хранение больших объемов данных, хранение высокоценных данных или данных, требующих соблюдения режима секретности. Минимальные ресурсы ПК: процессор 468DX–33МГц, Windows NT, объем оперативной памяти 16 (32) Мб, занимаемый объем на ЖМД 80 Мб.

Указанные программные продукты имеют возможности визуального проектирования интерфейса пользователя, т.е. разработчик из готовых фрагментов создает элементы интерфейса, программирует только их изменения в ответ на какие-либо события [5].

Принципы организации данных, лежащие в основе современных СУБД

Современные СУБД являются объектно-ориентированными и реляционными. Основными единицами являются объекты, имеющие свойства, и связи между объектами. СУБД используют несколько моделей данных: иерархическую и сетевую (с 60-х гг.) и реляционную (с 70-х). Основное различие данных моделей в представлении взаимосвязей между объектами.

Иерархическая модель данных строится по принципу иерархии объектов, т.е. один тип объекта является главным, все нижележащие-подчиненными. Устанавливается связь «один ко многим», т.е. для некоторого главного типа существует несколько подчиненных типов объектов. Иначе главный тип именуется исходным типом, а подчиненные – порожденными. У подчиненных типов могут быть, в свою очередь, подчиненные типы. Наивысший в иерархии узел (совокупность атрибутов) называют корневым.

Сетевая модель данных строится по принципу «главный и подчиненный тип одновременно», т.е. любой тип данных одновременно может одновременно порождать несколько подчиненных типов (быть владельцем набора) и быть подчиненным для нескольких главных (быть членом набора).

Реляционная модель данных: объекты и связи между ними представляются в виде таблиц, при этом связи тоже рассматриваются как объекты. Все строки, составляющие таблицу в реляционной БД, должны иметь *первичный ключ*. Все современные средства СУБД поддерживают реляционную модель данных.

Объект (сущность) – элемент какой-либо системы, информация о котором сохраняется. Объект может быть как реальным (например, человек), так и абстрактным (например, событие – поступление человека в стационар).

Атрибут – информационное отображение свойств объекта. Каждый объект характеризуется набором атрибутов.

Таблица – упорядоченная структура, состоящая из конечного набора однотипных записей.

Первичный ключ – атрибут (или группа атрибутов), позволяющий однозначным образом определить каждую строку в таблице.

Альтернативный ключ – атрибут (или группа атрибутов), не совпадающий с первичным ключом и однозначным образом определяющий каждую строку в таблице [1].

Современные технологии, используемые в работе с данными

Технология «Клиент-сервер» – технология, разделяющая приложение – СУБД – на две части: клиентскую часть (интерактивный графический интерфейс, расположенный на компьютере пользователя) и сервер, собственно осуществляющий управление данными, разделение информации, администрирование и безопасность, находящийся на выделенном компьютере. Взаимодействие «клиент-сервер» осуществляется следующим образом: клиентская часть приложения формирует запрос к серверу баз данных, на котором выполняются все команды, а результат исполнения запроса отправляется клиенту для просмотра и использования. Данная технология применяется, когда размеры баз данных велики, велики размеры вычислительной сети и требуется высокая производительность при обработке данных, хранящихся не на компьютере пользователя (в крупном учреждении обычно имеет место именно такая ситуация). Если технология «клиент-сервер» не применяется, то для обработки даже нескольких записей весь файл копируется на компьютер пользователя, а только затем обрабатывается. При этом резко возрастает нагрузка сети и снижается производительность труда многих сотрудников.

Microsoft Access, Microsoft Visual FoxPro, Microsoft Visual Basic и др. обеспечивают средства для создания клиентских частей в приложениях «клиент-сервер», которые сочетают в себе средства просмотра, графический интерфейс и средства построения запросов, а Microsoft SQL Server является на сегодняшний день одним из самых мощных серверов баз данных.

OLE 2.0 (Object Linking and Embedding – связывание и внедрение объектов) – стандарт, описывающий правила интеграции прикладных программ. Применяется для использования возможностей других приложений. OLE 2.0 используется для определения и совместного использования объектов несколькими приложениями, которые поддерживают данную технологию. Например, использование в среде Access таблиц Excel и его

мощных средств построения диаграмм или использование данных, подготовленных Access, в отчетах, составленных в редакторе текстов Word.

OLE Automation (Автоматизация OLE) – компонент OLE, позволяющий программным путем устанавливать свойства и задавать команды для объектов другого приложения. Позволяет без необходимости выхода или перехода в другое окно использовать возможности нужного приложения. Приложение, позволяющее другим прикладным программам использовать свои объекты, называется *OLE сервером*. Приложение, которое может управлять объектами OLE серверов, называется *OLE контроллером* или *OLE клиентом*. Из рассмотренных программных средств в качестве OLE серверов могут выступать Microsoft Access, а также Microsoft Excel, Word и Graph. Microsoft Visual FoxPro 3.0 и 5.0 может выступать только в виде OLE клиента.

RAD (Rapid Application Development – Быстрая разработка приложений) – подход к разработке приложений, предусматривающий широкое использование готовых компонентов и/или приложений и пакетов (в том числе от разных производителей).

ODBC (Open Database Connectivity – открытый доступ к базам данных) – технология, позволяющая использовать базы данных, созданные другим приложением при помощи SQL.

SQL (Structured Query Language – язык структурированных запросов) – универсальный язык, предназначенный для создания и выполнения запросов, обработки данных как в собственной базе данных приложения, так и с базами данных, созданных другими приложениями, поддерживающими SQL. Также SQL применяется для управления реляционными базами данных.

VBA (Visual Basic for Applications – Visual Basic для Приложений) – разновидность (диалект) объектно-ориентированного языка программирования Visual Basic, встраиваемая в программные пакеты [6, 7].

Разработка структуры и основных форм БД структуры

В программный пакет Delphi 7 входят следующие утилиты: BDE Administrator; Database Desktop; Delphi7; Datapump; Image Editor; SQL Explorer; SQL Monitor; WinSight32; XML Mapper.

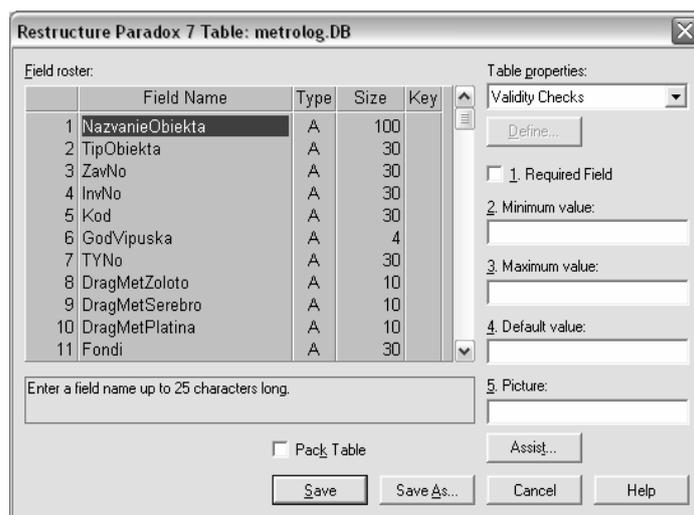


Рис. 1. Database Desktop

Database Desktop является удобной утилитой для создания структур баз данных. Database Desktop предлагает на выбор список форматов возможных структур БД. При выборе типа будущей таблицы следует учитывать то, какого типа данные будут в ней храниться. Так как разрабатываемая структура содержит каких-либо особенных данных, была выбрана таблица Paradox 7, поддерживающая все необходимые типы. После выбо-

ра таблицы создаем в ней необходимые поля и указываем тип данных, которые будут храниться в них. В разрабатываемой системе большинство входных данных состоит из набора букв и цифр, следовательно, таблицы должны «понимать» такую информацию. На рис. 1 приведен общий вид утилиты Database Desktop в процессе создания первой таблицы Paradox. Таким образом, реализуется поставленный в техническом задании пункт о создании структуры БД, содержащей общую и техническую информацию по измерительному и контрольно-поверочному оборудованию ОАО «РИРВ».

Разработка основных форм

Для создания основных форм запроса, добавления и отображения базы данных используется язык Delphi 7. На рис. 2 представлен внешний вид графической оболочки Delphi 7.

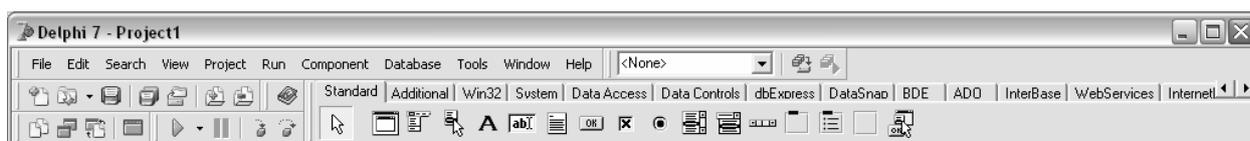


Рис. 2. Delphi 7

Создаем внешний вид СУБД. Работа в Delphi основана на визуально-осязаемых компонентах, таких как кнопки, диалоговые окна, текстовые вставки и многое другое. Сначала создаем форму. Форма – это заготовка стандартного исполняемого окна в Windows. Помещаем на нее из строки компонентов Delphi необходимые элементы. После утверждения желаемого графического интерфейса создаем основную форму [5, 7–9].

Аналогично, оперируя стандартными компонентами Delphi, создается форма добавления нового элемента БД (рис. 3).

Рис. 3. Добавление нового элемента

Форма добавления нового элемента содержит все необходимые компоненты для добавления новой записи в таблицу с общими характеристиками (год выпуска прибора, № свойства, код прибора, содержание драгметаллов и т.д.). Но необходимо занести еще и техническую информацию по прибору. В зависимости от названия прибора, необходимо создать двенадцать форм, одноименных с названием приборов – анализаторы спектра; антенны; ваттметры; вольтметры; генераторы; измерители АЧХ; измерители КСВН; измерители сопротивления; источники питания; мосты; осциллографы; частотомеры – так как для каждого прибора существуют свои индивидуальные технические характеристики. Все эти формы внешне похожи, хоть и позволяют внести разные технические характеристики. На рис. 4 и рис. 5 приведены в качестве примера две типовые формы ввода технических характеристик.

Рис. 4. Вольтметр

Рис. 5. Источник питания

Рис. 6. Запрос

Рассмотрим подробнее форму запроса, утвержденную заказчиком. Запрос – это выборка из БД отдельных записей по следующим условиям: по названию прибора, по типу прибора, по коду прибора, по году выпуска прибора. Например, если указать в графе формы запроса «Выберите название прибора» вольтметр, то программа выведет только те записи, которые относятся к вольтметрам. На рис. 6 приведено графическое изображение формы запроса.

После создания формы запроса необходимо вывести информацию на экран. Основная форма вывода информации представляет собой таблицу типа таблиц Microsoft Excel с рядом функциональных кнопок. Эта форма является стартовой для большинства алго-

ритмов, реализованных в системе. Об этом можно судить по наименованиям и количеству кнопок на форме, представленной на рис. 7.

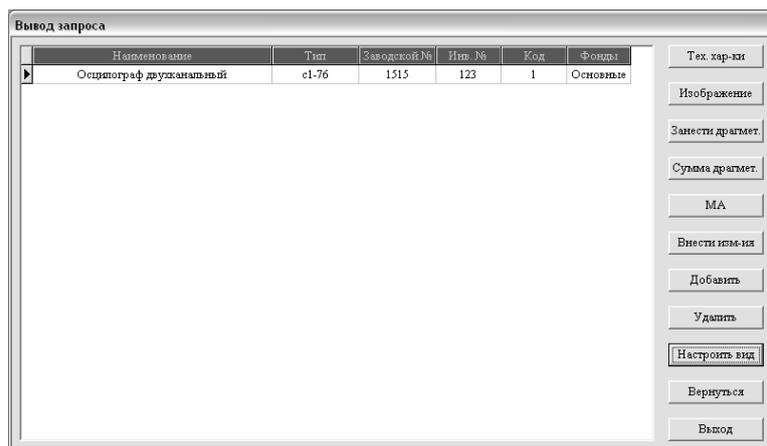


Рис. 7. Вывод запроса

Как уже было отмечено выше, технические характеристики многих приборов довольно сильно различаются между собой, у некоторых приборов присутствует параметр, которого нет у других, и наоборот. Поэтому под различные приборы созданы различные формы отображения технических характеристик, значения которых составляет вторую таблицу. На рис. 8 приведена одна из таких типовых форм – форма вывода данных по осциллографам).

Рис. 8. Осциллографы

Структура программы

Все алгоритмы и пользовательский интерфейс были разработаны и реализованы с помощью объектно-ориентированного языка программирования Delphi 7. Структура БД была разработана с помощью утилиты Database Desktop, входящей в стандартный состав утилит программного пакета Delphi 7. На рис. 9 представлен основной алгоритм работы разработанной БД.

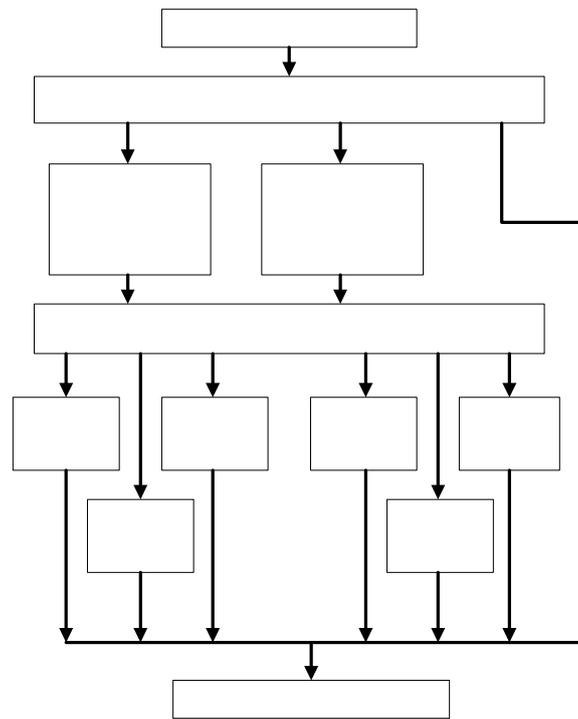


Рис. 9. Алгоритм работы БД

Заключение

В статье рассмотрены системы управления базами данных, используемые в приложениях, построенных по технологии «клиент-сервер». Сегодня уже не имеет значения, на каком языке и на основе какого пакета написано конкретное приложение и какой формат данных в нем используется. Более того, стандартом «де-факто» стала «быстрая разработка приложений» – RAD (Rapid Application Development), основанная на широко декларируемом в литературе «открытом подходе». Поэтому в одном ряду с «классическими» СУБД все чаще упоминаются языки программирования Delphi и Visual C++, которые позволяют быстро создавать необходимые компоненты приложений, критичные по скорости работы, которые трудно, а иногда невозможно разработать методами «классических» СУБД. В статье рассмотрена введенная в эксплуатацию конкретная база данных по контролю и учету контрольно-поверочного оборудования, находящегося на балансе ООО «РИРВ». Исследованы и применены на практике алгоритмы программирования БД.

Литература

1. Вирт Н. Алгоритмы и структуры данных / Пер. с англ. М.: Мир, 1989. 360 с., ил.
2. Кандзюба С.П., Громов В.Н. Delphi 6/7. Базы данных и приложения. Лекции и упражнения. М: ДиаСофт, 2002.
3. Нечаев В.М. Microsoft Excel. Электронные таблицы и базы данных в задачах. М: Интеллект-Центр, 2001.
4. Яргер Р.Дж., Риз Дж., Кинг Т. MySQL и mSQL. Базы данных для небольших предприятий и Интернета. СПб: Символ-Плюс, 2000.
5. Архангельский А.Я. 100 компонентов общего назначения библиотеки Delphi5. М: Бином, 1999.
6. Послед Б. Access 2000. Базы данных и приложения. Лекции и упражнения. М: ДиаСофт, 2000.
7. Фаронов В.В. Delphi 5. Руководство программиста. СПб: Питер, 2001.
8. Архангельский А.Я. Delphi 6. Справочное пособие. М: Бином, 2001.
9. Кэнту М. Delphi 7. Для профессионалов. СПб: Питер, 2004.

Запуск пр

Начало р
заглавн

Добавление
НОВОЙ
записи

Основное окно: вы
запр

Настройка
внешнего
вида

Вывод
изображе
ния

Поддержка
драг-
метал-
лов

Выход из

УСТАНОВКА И КОНФИГУРИРОВАНИЕ WEB-СЕРВЕРА АРАШЕ, ИНТЕРПРЕТАТОРА PHP И СЕРВЕРА БАЗ ДАННЫХ MYSQL

Ф.Г. Петров, Ю.И. Рядчин

Научный руководитель – доктор технических наук, профессор А.Г. Коробейников

Статья посвящена установке и настройке популярной связки web-сервера Apache + PHP + сервера баз данных MySQL. Рассмотрены вопросы установки и конфигурирования перечисленных программ, а также проверка правильности их установки.

Введение

В связи со стремительным ростом числа сайтов в сети InterNet и возможностями серверов растут и требования, предъявляемые к сайтам, и уже практически невозможно сделать хороший динамический сайт, который легок в управлении и способен обрабатывает запросы, а также оповещать администратора об определенных событиях, не используя механизмы обработки на стороне сервера. Все эти требования и многие другие можно достаточно просто выполнить, используя соответствующие инструменты. Наиболее гибким и удобным инструментом является персональный гипертекстовый пре-процессор PHP. Лучше всего использовать в связке с web-сервером Apache и сервером баз данных MySQL.

В данной статье описывается, как сделать максимально просто установку и конфигурирование очень популярной связки web-сервера Apache + PHP + сервера баз данных MySQL. Статья не может претендовать на полный обзор настройки таких серьезных вещей, как web-сервер Apache, сервер баз данных MySQL и PHP, но она предназначена для общего ознакомления с установкой, настройкой и возможностями этих программ, а также позволит начинающим программистам и веб-дизайнерам установить и настроить эти приложения в достаточном объеме для разработки своих приложений. Это позволит им видеть работу своих приложений глазами пользователя еще до выкладывания сайта у хост-провайдера и правильно отладить приложения.

Web-сервер Apache – это программа, которая исполняет функции http-сервера (web-сервера), т.е. обрабатывает запросы, пришедшие по протоколу http. Apache обладает широчайшими конфигурационными возможностями. Существуют тысячи возможных комбинаций значений для сотни конфигурационных переменных. Без сомнения, тот, кто работал с ОС Unix или DOS, знаком с концепцией опций, задаваемых в командной строке. Основной принцип, заложенный в концепцию директивы, аналогичен концепции опции, задаваемой в командной строке. Но здесь, вместо того чтобы задавать при каждом запуске Web-сервера дюжину или сотню конфигурационных переменных Apache, параметры собраны в одном конфигурационном файле httpd.conf, который автоматически считывается сервером во время запуска.

По умолчанию конфигурационные файлы хранятся в каталоге conf, который находится в главном каталоге сервера Apache. Как и все остальное окружение web-сервера Apache, размещение и имена этих файлов легко меняется, но для простоты изложения материала здесь и далее воспользуемся стандартными именами файлов и каталогов.

Как подчеркивалось выше, значения конфигурационных переменных хранятся в конфигурационных файлах. Эти переменные и есть директивы. Основное время администрирования сервера Apache (до 90%) затрачивается на определение того, значения каких директив будут изменяться, и значений, которые эти директивы должны задавать.

Следует обратить внимание на то, что далеко не все директивы автоматически распознаются сервером Apache. Существует достаточно большое подмножество директив, которые называются основными. Эти директивы устанавливаются по умолчанию.

Другие директивы распознаются в зависимости от того, какой набор модулей скомпилирован при построении данного конкретного сервера Apache. Поэтому указание директивы в конфигурационном файле совсем не означает, что она вызовет какое-либо действие: работающий вариант сервера вполне может не иметь модуля, который взаимодействует с данной директивой.

Далеко не все директивы применяются одновременно. Иногда очень полезно иметь один набор директив для одного варианта сервера, а другой – для другого варианта. Наглядным примером может служить случай, когда один и тот же сервер обслуживает более одного web-узла. В таком случае для каждого узла задаются различные директивы, и, как минимум, у них должны быть разные имена.

Диапазон конкретной директивы может быть ограничен тремя способами:

- по каталогу с помощью директив `<Directory>`, `<DirectoryMatch>` или файла `.htaccess`;
- по URL (Unified Resource Locator) с помощью директив `<Location>` и `<LocationMatch>`;
- по файлу с помощью директив `<Files>` и `<FilesMatch>`.

Наиболее интересно применение файла `.htaccess`. Директивы, указанные в этом файле, работают в каталоге, в котором содержится этот конкретный файл, а также в его подкаталогах.

Сервер Apache имеет ядро, гарантирующее выполнение основных функций. Ядро обеспечивает работу директив, возможность чтения конфигурационных файлов, усеченную возможность управления доступом, возможность дополнения функциональных возможностей и полдесятка других основных функциональных возможностей.

Кроме того, сервер Apache разработан таким образом, что всегда существует возможность варьирования основных функциональных возможностей. Функциональные части могут быть и не подключены к первоначальной исполняемой программе. Эти субсекции называются модулями, причем значительная их часть поставляется по умолчанию в стандартном дистрибутиве. Если модуль включен, он становится составной частью исполняемого процесса с тем же идентификатором процесса и доступом к тем же системным ресурсам. Задача состоит в том, как наиболее удобно и эффективно выполнить связку Apache + PHP + MySQL, которая бы позволила решить задачи по обработке данных на стороне сервера. MySQL – это сервер баз данных, исполняющий запросы на языке SQL.

PHP является персональным гипертекстовым препроцессором, выполняющим сценарии, написанные на соответствующем языке программирования.

Так как данная статья предназначена для разработчиков веб-приложений, а не администраторов, то будет рассмотрена установка программ на ПК с популярной ОС Windows XP.

Установка сервера баз данных MySQL

Загрузите дистрибутив MySQL для Windows с <http://mysql.com/downloads/>. Разархивируйте его в подходящее место, после этого запустите программу `setup`. После стандартной вводной части и экрана `ReadMe` выберите каталог, в который будет устанавливаться MySQL. Если нет каких-либо веских причин для иного, выберите установку в каталог, предложенный по умолчанию. На следующем экране выберите тип установки `Typical`, если только вы не испытываете крайнего недостатка свободного места на диске. MySQL надо установить как службу.

После установки MySQL следует создать пользователя `root` и задать ему пароль. Для этого надо открыть сеанс MS DOS, зайти в папку `C:/mysql/bin` и выполнить там команду `mysql -u root -p test`. Флаг `-u` задает пользователя, а `-p` приводит к выдаче при-

глашения ввода на пароль. Также в этой папке можно выполнить команду *mysqlshow*, которая выведет список существующих баз данных.

Службу MySQL можно запускать как в ручном, так и в автоматическом режиме. Подробно об этом написано в [1].

Установка web-сервера Apache

Загрузите мастер установки Apache с <http://www.apache.org/dist/httpd/>. Дважды щелкните по программе установки и нажмите кнопку Next в окне мастера установки, после чего появится окно, в котором надо ввести в качестве Network Domain и Server Name *localhost* или *127.0.0.1*. В любом случае в Administrator's Email Address следует указать свой реальный адрес электронной почты или тот, с которого происходит переадресация.

В нижней части экрана можно указать, как следует установить Apache: в виде службы или в виде приложения. Следует выбрать службу, если нет какой-то особой причины этого не делать. Служба всегда невидимо работает, независимо от того, кто зарегистрирован в системе, и даже если не зарегистрирован никто, а приложение может быть запущено только пользователем, зарегистрированным в системе. При завершении пользователем работы завершаются все запущенные им программы.

Следующий экран предлагает сделать полную или выборочную установку. Большинству пользователей подойдет полная установка, но если жесткий диск заполнен, а вы обычно подключены к Интернету, то можно остановиться на выборочной установке и сбросить флажки для Documentation и Source Code, электронную документацию и исходный код можно взять на <http://apache.org/>. При выборочной установке следует оставить предлагаемый по умолчанию для размещения программы каталог C:\Program Files\Apache Group\, если нет достаточных оснований изменить его.

После запуска Apache в браузере по адресу <http://localhost/> или <http://127.0.0.1/> должна открыться страница «Test Page for Apache Installation». Если она открылась, значит, все сделано правильно.

Установка PHP

Сначала загрузите дистрибутив PHP для Windows с <http://php.net/downloads.php>.

Затем остановите web-сервер Apache и сервер баз данных MySQL, если они работают в данный момент. Начните установку, выполнив двойной щелчок по программе setup. После стандартного экрана представления и лицензионного соглашения запрашивается вид установки – Standard или Advanced. Рекомендуется выбрать Standard – по этому пути мы и пойдём. Опытные пользователи могут попробовать установку Advanced. В любой момент можно вернуться к предыдущему экрану, нажав кнопку Back, или даже нажать Cancel и начать все с самого начала.

Если выбран стандартный тип установки, то следующим появляется диалоговое окно, в котором предлагается задать каталог, в который требуется установить PHP. Лучше всего оставить предложенный по умолчанию *C:\php*, если нет веских оснований выбрать другой каталог.

Затем запрашиваются настройки, необходимые PHP для отправки электронной почты с помощью встроенной почтовой функции. В этом диалоге надо задать действующий сервер SMTP и адрес From: по умолчанию. Если вы знаете, какой сервер SMTP обрабатывает исходящие сообщения, и знаете, что для авторизации отправки почты он использует только адрес From:, можно задать эти значения. Если имя сервера SMTP вам неизвестно или авторизация на этом сервере требует не только правильного адреса

From:, но и указания имени пользователя/пароля, укажите *localhost* в качестве сервера SMTP и свой обычный адрес электронной почты в поле From:.

Поскольку в большинстве версий Windows сервер SMTP не устанавливается, имя *localhost* не будет реально работать, пока вы не установите на своей машине сервер SMTP. К счастью, эти настройки можно потом легко изменить, а для выполнения функций, которые могут потребоваться, есть бесплатно доступные инструменты.

Наконец, все решения приняты, и все готово к установке. Щелкните по последней кнопке Next, и программа установки начнет работу.

Мастер установки в настоящее время не настраивает автоматически файл *httpd.conf* для Apache, о чем может быть сообщено в диалоговом окне.

После завершения установки отображается окно, объявляющее об успешной инсталляции, с кнопкой ОК. Нажимайте ее и переходите к настройке Apache.

Конфигурирование

Сейчас на компьютере установлены все необходимые компоненты, теперь осталось их только настроить.

Перед внесением любых изменений в рабочие конфигурационные файлы рекомендуется сделать их резервные копии.

Настоящим конфигурационный файл web-сервера Apache *httpd.conf*. Если вы не меняли пути при установке, то, скорее всего, файл *httpd.conf* находится в папке *C:\Program Files\Apache Group\Apache2\conf*.

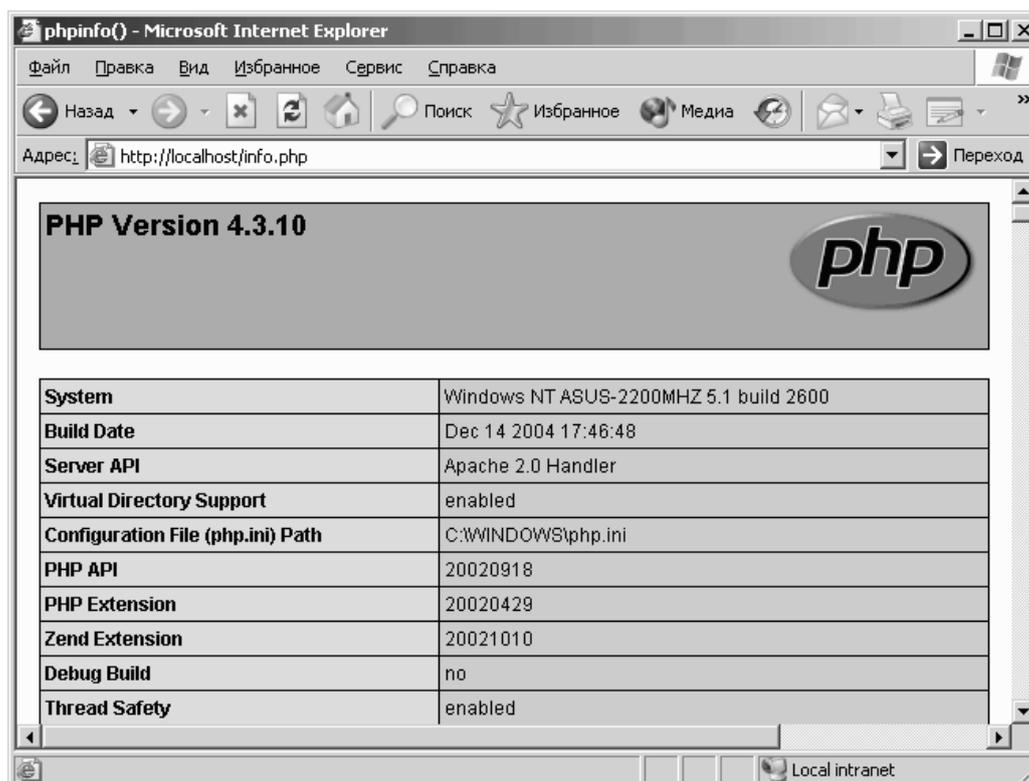


Рис. Сведения о конфигурации web-сервера Apache

Далее идет описание минимально-необходимых настроек.

1. Указывается корневой каталог вашего сайта, это делается с помощью директивы *DocumentRoot*. По умолчанию директива выглядит так: *DocumentRoot "C:/Program Files/Apache Group/Apache2/htdocs"*. Вы можете задать какой угодно путь, но лучше не задавать каталог, в котором содержится что-либо еще, помимо файлов вашего сайта.

2. После того как Apache связал заданный URL с каталогом, он возвращает файл или файлы, которые находятся в этом каталоге. По соглашению это файл `index.html`. Директивой `DirectoryIndex` задается список имен файлов, которые будет пытаться вернуть Apache при обращении к каталогу. Apache вернет первый элемент списка, который будет найден в каталоге. По умолчанию эта директива выглядит так:

```
DirectoryIndex index.html index.html.var.
```

3. Директива `AddDefaultCharset` задает кодировку, в которой страницы будет выдаваться клиенту. По умолчанию эта директива выглядит так:

```
AddDefaultCharset ISO-8859-1.
```

Эту строку можно закомментировать, поставив знак `#` в начале строки, и вставить другую, например: `AddDefaultCharset WINDOWS-1251`.

Как оказалось, настройка web-сервера Apache достаточно легка. Огромное количество остальных директив и настроек не рассматривается, так как достаточно этих, чтобы начать работать. Если необходимо больше информации по директивам, для начала следует обратиться к ресурсу <http://www.apache.ru/docs/fishki.html>, а также к [2].

Если вы хотите задать свои собственные страницы ошибок, это можно сделать с помощью директивы `ErrorDocument`. Пример записи: `ErrorDocument 404 "/missing.html"` – директива, потом номер ошибки и страница, которая будет отображаться при возникновении этой ошибки.

Конфигурирование PHP

PHP с web-сервером Apache может работать как CGI или модуль. Подключение PHP в качестве модуля более предпочтительно, так как в данной конфигурации PHP будет частью ядра Apache и работать будет, соответственно, быстрее [3]. Чтобы PHP подключить в качестве модуля, надо в файл `httpd.conf` надо сделать 3 изменения.

1. Добавить в соответствующий раздел строку

```
LoadModule php4_module "c:/php/sapi/php4apache2.dll"
```

Директива `LoadModule` загружает указанный модуль.

2. Добавить в соответствующий раздел строку

```
AddType application/x-httpd-php .php
```

Директива `AddType` ассоциирует MIME-тип с расширением файла. Благодаря этой записи Apache понимает, что если запрошен файл с расширением `.php`, то перед тем, как его отправить клиенту, файл должен быть предварительно обработан соответствующим модулем, в нашем случае PHP.

3. В директиву `DirectoryIndex` добавить `index.php`. Это позволит сделать файл `.php` индексным.

Помимо этих изменений, необходимо:

- скопировать файл `php.ini` из папки `C:/php` в папку `C:/Windows`.
- скопировать файл `php4ts.dll` из папки `C:/php` в папку `C:/Windows/System32` и `C:/Program Files/Apache Group/Apache2/bin`.
- изменить в файле `php.ini` в папке `C:/Windows` значение параметра `error_reporting` на `E_ALL`. Это позволит вам отслеживать все ошибки и предупреждения при программировании приложений.

В результате всех действий должна получиться рабочая связка Apache+PHP+MySQL. Проверить работоспособность связки Apache+PHP можно с помощью файла `info.php`, который состоит из 3 строк:

```
<? php  
phpinfo();  
?>
```

Этот файл надо поместить в корневую директорию сайта (`C:/Program Files/Apache Group/Apache2/htdocs`) и открыть в браузере страницу по адресу <http://localhost/info.php>

или *http://127.0.0.1/info.php*. Должна будет открыться страница с заголовком `phpinfo()`, содержащая сведения о конфигурации web-сервера Apache, как показано на рисунке.

Проверить работу всей связки Apache+PHP+MySQL можно точно так же, как и в предыдущем случае, но файл должен состоять из следующих строк:

```
<? php
// Используется имя пользователя root и пароль password
if(!mysql_connect("localhost","root","password"))
{
echo "Не могу соединиться с сервером\n";
echo mysql_error();
exit;
}
echo "Работает!"
?>
```

При открытии этого файла в окне браузера появится слово «Работает!».

Серверы Apache и MySQL устанавливаются как службы, поэтому они могут запускаться вручную или автоматически, это зависит от того, как вы их настроите в разделе «Управление компьютером».

Если возникли какие-либо проблемы при установке или конфигурировании, следует сначала обратиться к электронной документации, поставляемой в комплекте с соответствующим ПО.

Как видим, несмотря на широчайшие конфигурационные возможности установленных нами пакетов, поставленная цель может быть достигнута относительно просто и быстро. В результате всех выполненных действий получена рабочая связка web-сервера Apache, интерпретатора PHP и сервера баз данных MySQL, готовая для решения большинства задач и отладки приложений.

Литература

1. Дюбуа П. MySQL.: Пер. с англ. М.: Издательский дом «Вильямс», 2004. 1056 с.: ил.
2. Хокинс С. Администрирование Web-сервера Apache и руководство по электронной коммерции.: Пер. с англ. М.: Издательский дом «Вильямс», 2001. 336 с.: ил.
3. Аргерих Л. и др. Профессиональное PHP программирование, 2-е издание. Пер. с англ. СПб: Символ-Плюс, 2005. 1048 с.: ил.

ГРУППА СТАНДАРТОВ WiMAX

А.Ю. Прокопенко

Научный руководитель – кандидат технических наук, доцент Б.А. Крылов

Статья посвящена краткому обзору группы стандартов беспроводных сетей WiMAX.

Введение

Стандарты WiMAX (Worldwide Interoperability for Microwave Access) на сегодняшний день – одни из перспективных технологий в мире беспроводных сетей (рис. 1).

Сети, построенные на базе WiMAX (IEEE 802.16), ориентированы на использование в городской среде (Metropolitan Area Network (MAN)).

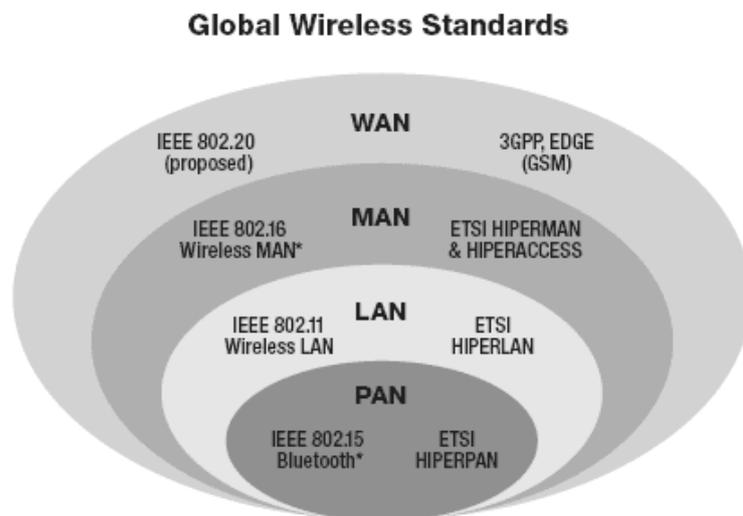


Рис. 1. Типы беспроводных сетей. PAN (Private Area Network) – частные сети; LAN (Local Area Network) – локальные сети; MAN (Metropolitan Area Network) – городские сети; WAN (WorldWide Area Network) – мировые сети

Группа стандартов WiMAX

WiMAX – группа стандартов IEEE 802.16, созданная с целью обеспечения широкополосного беспроводного доступа «последней мили». Понятие доступа «последней мили» на сегодняшний день все больше используется в описаниях беспроводных сетей. Значение этого термина объясняется экономическим фактором: легче и дешевле поставить одну коллективную точку беспроводного доступа в густо населенном районе, чем тянуть сеть в каждый дом, офис и так далее [1, 2].

В 2001 г. фирмами Nokia, Harris Corp., Ensemble и Crosspan была создана организация «WiMAX forum» с целью поддержки и продвижения стандарта IEEE 802.16, а также обеспечения взаимодействия с другими стандартами (HIPERMAN (Европа) и WiBro (Южная Корея)).

На сегодняшний день пока не существует стандартизированных решений для оборудования WiMAX, однако достаточно большое количество оборудования уже доступно на рынке. Оно выполнено в соответствии со стандартами IEEE 802.16 и внутренними разработками выпускающих компаний. Такое оборудование называется pre-WiMAX продуктами. Официальная сертификация оборудования началась в июне 2005 г. На данный момент только Cetecom labs имеет право сертифицировать продукты на соответствие стандартам IEEE 802.16 [3].

Рабочая группа, занимающаяся разработкой стандартов IEEE 802.16, была организована в 1999 г. для работы над WLL (Wireless Local Loop). Целью группы стало построение стандарта широкополосного беспроводного доступа со следующими характеристиками:

- IP-ориентированный стандарт;
- обеспечение качества обслуживания QoS (Quality of Service) для каждого соединения в отдельности;
- работа в нелицензируемом и лицензируемом спектре с ориентацией на США;
- увеличение дистанции между передатчиком и приемником.

В табл. 1 представлены стандарты IEEE 802.16 по мере их разработки, начиная с 2001 г.

IEEE	802.16	802.16a	802.16REVd	802.16e
Одобрено	декабрь 2001	январь 2003	июль 2004	июль 2005
Частотный диапазон	от 10 до 66 ГГц	менее 11 ГГц	менее 11 ГГц	от 2 до 6 ГГц
Видимость	прямая, LOS	непрямая, NLOS	непрямая, NLOS	непрямая, NLOS
Модуляция	QPSK, 16QAM и 64QAM	OFDM 256, OFDMA + 802.16	OFDM 256, OFDMA + 802.16	OFDM 256, OFDMA + 802.16
Скорость	32 - 134 Мбит/с	1 – 75 Мбит/с	см. 802.16a	до 15 Мбит/с
Мобильность	Нет	Нет	Нет	да, с возможностью регионального роуминга
Ширина канала	20, 25 и 28 МГц	Изменяемая от 1,25 до 20 МГц с 16 логическими каналами	см. 802.16a	Более 5 МГц
Радиус ячейки	От 1 до 5 км	От 5 до 8 км; максимум 50 км с соотв. антенной и максимальной мощностью передачи	см. 802.16a	от 1 до 5 км
Терминальное оборудование		внешнее с выносной антенной	внешнее с встроенной антенной	PC – карта

Таблица 1. Семейство стандартов IEEE 802.16

IEEE 802.16 (IEEE 802.16-2001), одобренный в декабре 2001 г., стал первым стандартом «точка–многоточка» в области WMAN. Стандарт был ориентирован на работу в частотном диапазоне от 10 до 66 ГГц и, как следствие, требовал нахождения передатчика и приемника в зоне прямой видимости (LOS, Line of Sight). Это существенный недостаток при использовании технологии в условиях города. IEEE 802.16 не был хорошо проработан, и даже после его утверждения осталось большое количество открытых вопросов [4].

IEEE 802.16a стал первым действующим стандартом, где было устранено большинство недостатков предыдущего. Главное отличие – понижение рабочей частоты и реализация функционирования в зоне не прямой видимости (NLOS, Non Line of Sight).

IEEE 802.16REVd (IEEE 802.16-2004) – исправленная версия стандарта IEEE 802.16a, с поддержкой фиксированного офисного или домашнего оконечного терминального оборудования в соответствии с NLOS посредством таких дополнительных функций, как формирование антенной направленного сигнала и использование поднесущих OFDM (ортогональное частотное разделение канала).

IEEE 802.16e будет следующей ступенью развития идей предыдущих стандартов с ориентацией на мобильность оконечного пользователя.

Операторы, предоставляющие услуги связи, имеют возможность достаточно гибко настраивать сети WiMAX в соответствии со своими нуждами:

- для удаленных на десятки километров от базовой станции пользователей:
 - обеспечение телефонной связью, доступом в Интернет;
 - необходимо наличие внешней направленной антенны на стороне пользователя в зоне LOS, причем антенна может быть использована как точка коллективного доступа;
- для удаленных на 5–20 км от базовой станции пользователей:
 - обеспечение широкополосного доступа к услугам связи;
 - необходимо наличие внешней антенны на стороне пользователя в зоне LOS, причем антенна может быть использована как точка коллективного доступа;
- для удаленных на 5 км от базовой станции пользователей:
 - обеспечение широкополосного доступа к услугам связи с использованием внутренней антенны;
 - пользователь может быть ограниченно мобильным – в зависимости от уровня сигнала.

Различные стандарты семейства IEEE 802.16 ориентированы на различные сегменты рынка. Например, стандарт IEEE 802.16-2004 обеспечивает решение «последней мили» для *фиксированных* пользователей. IEEE 802.16e ориентирован на обслуживание *мобильных* пользователей (рис. 2).

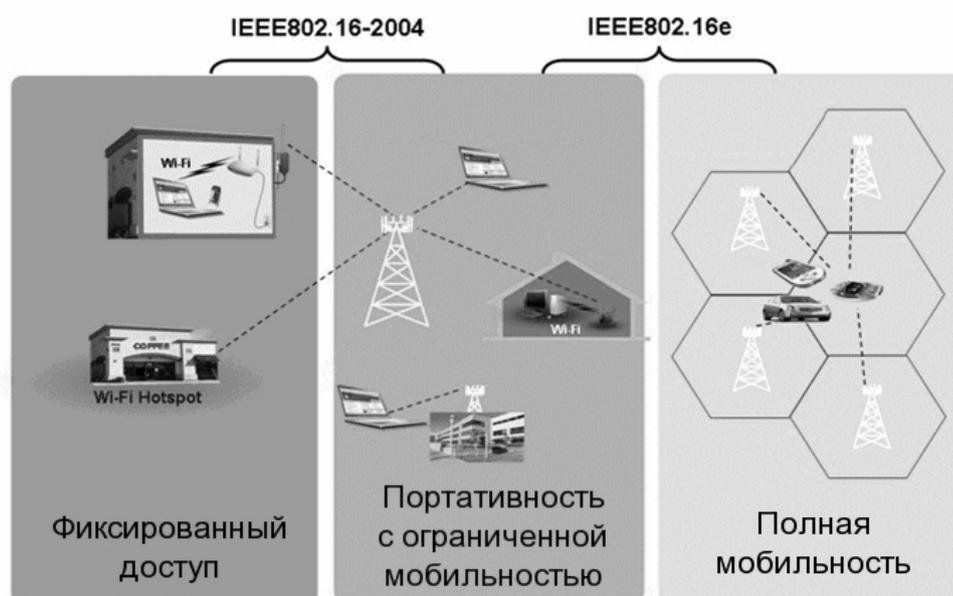


Рис. 2. Стандарты IEEE 802.16 в различных областях применения

В табл. 2 отражены главные отличия WiMAX от популярного на сегодняшний день Wi-Fi (IEEE 802.11). Последний не является прямым конкурентом WiMAX, так

как создавался для других целей. Однако это не мешает им взаимодействовать друг с другом, образуя комплексы беспроводных сетей (рис. 2).

IEEE 802.11 (Wi-Fi)	IEEE 802.16 (WiMAX)
Был разработан для доступа к Интернет, в основном ориентируясь на услугу Best Effort	Разрабатывается для предоставления мультимедийных и голосовых услуг с заданным качеством обслуживания
Не ориентирован на соединение	Полностью ориентирован на соединение
Ориентирован на нелицензируемые частоты, количество пользователей сравнительно низкое	Ориентирован как на лицензируемые, так и нелицензируемые частоты, предполагаемое количество пользователей должно быть высоко
Дистанция передачи данных – метры	Дистанция передачи данных – километры

Таблица 2. Основные отличия Wi-Fi от WiMAX

На рис. 3 представлено применение стандартов IEEE 802.16 в общей концепции сетевой архитектуры, а на рис. 4 – стек стандарта IEEE 802.16 [4].

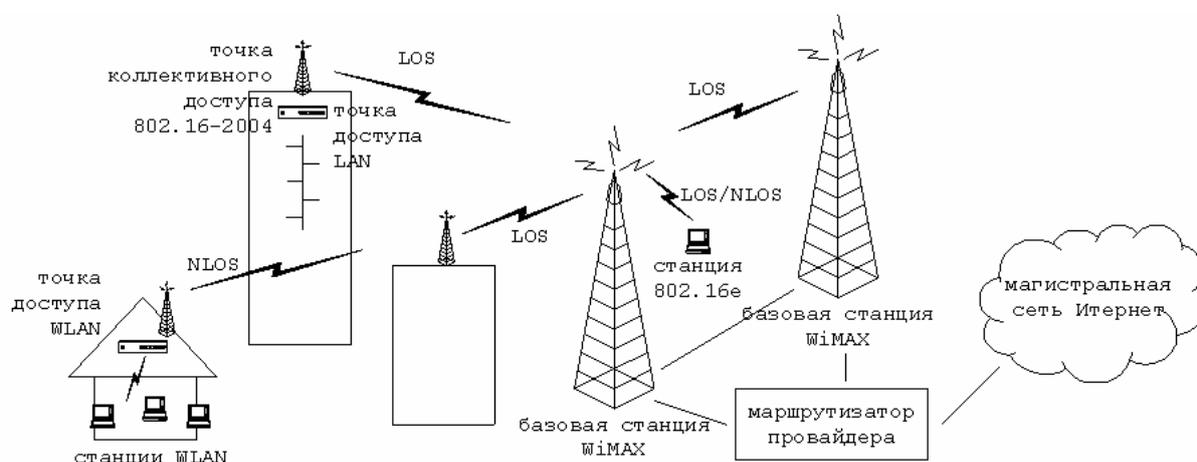


Рис. 3. Архитектура WiMAX (IEEE 802.16)

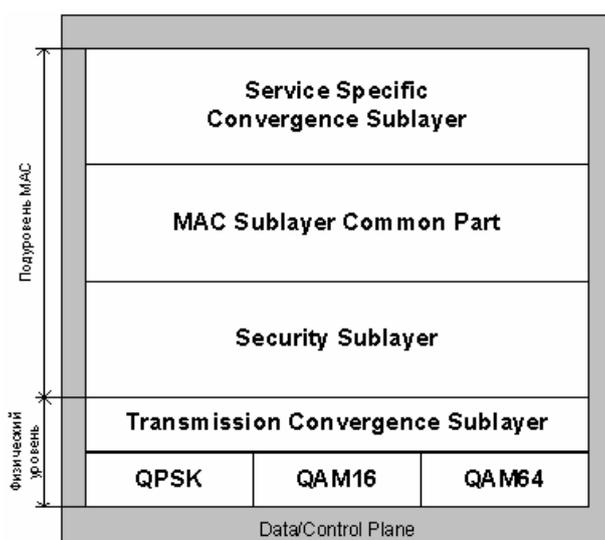


Рис. 4. Стек стандарта IEEE 802.16

Физический уровень IEEE 802.16 использует модуляцию OFDM (ортогональное частотное разделение канала) и поддерживает различные уровни модуляции, включая BPSK (двоичная фазовая модуляция), QPSK (квадратурная фазовая модуляция), 16QAM (16-позиционная квадратурная амплитудная модуляция) и 64QAM (64-позиционная квадратурная амплитудная модуляция) [4].

MAC уровень содержит подуровень безопасности (security sublayer), осуществляющий шифрование и управление ключами, и подуровень конвергенции (service specific convergence sublayer), поддерживающий протоколы более высокого уровня, такие как PPP, IP, ATM, Ethernet и т.п.

С выходом на рынок продукты WiMAX при правильной организации теоретически могут создать конкуренцию сотовым сетям в области голосовых услуг.

Заключение

Стандарт WiMAX постоянно развивается, создаются модификации для различных областей применения, и уже сейчас можно предвидеть успешное будущее данной технологии в мире беспроводных сетей.

Литература

1. www.intel.com/netcomms/technologies/wimax/index.htm – представление технологии WiMAX компанией Intel.
2. www.softco.ru/wimax.htm – Общие сведения о технологии WiMAX.
3. Григорьев В.А., Лагутенко О.И., Распаев Ю.А. Сети и системы радиодоступа. М. 2005.
4. Портной С.Л. Теоретико-информационные основы систем широкополосного беспроводного доступа. М. 2004.

РЕКУРРЕНТНЫЙ АНАЛИЗ – ТЕОРИЯ И ПРАКТИКА

В.Б. Киселев

Научный руководитель – кандидат технических наук, доцент Б.А. Крылов

Кратко рассмотрен перспективный метод анализа характера процессов – (кросс)-рекуррентный анализ. Уделено внимание вопросам реконструкции фазового пространства из одномерного временного ряда. Проведен обзор возможных областей применения метода. Предложены информационные модели систем с использованием (кросс)-рекуррентного анализа.

Введение

Развитие науки и техники ставит перед человечеством все увеличивающееся число вопросов для всестороннего изучения, хотя большинство из уже стоящих перед современной наукой проблем требуют более глубокого понимания Природы. Исследования сложных систем, как природных, так и искусственных, показали, что в их основе лежат нелинейные процессы, тщательное изучение которых необходимо для понимания и моделирования сложных систем.

В последние десятилетия набор традиционных (линейных) методик исследования был существенно расширен нелинейными методами, полученными из теории нелинейной динамики и хаоса; многие исследования были посвящены оценке нелинейных характеристик и свойств (скейлинг, фрактальная размерность) процессов, протекающих в Природе. Однако большинство методов нелинейного анализа требуют либо достаточно длинных, либо стационарных рядов данных, которые довольно трудно получить из Природы. Более того, было показано, что данные методы дают удовлетворительные результаты для идеализированных моделей реальных систем [1]. Эти факторы потребовали разработки новых методик нелинейного анализа данных.

Состояние природных или искусственных систем, как правило, изменяется во времени. Изучение этих, зачастую сложных, процессов – важная задача во многих дисциплинах, позволяющая понять и описать их сущность, например, для прогнозирования состояния на некоторое время в будущее. Целью таких исследований является нахождение математических моделей, которые бы достаточно соответствовали реальным процессам и могли бы быть использованы для решения поставленных задач.

В данной статье будет вкратце рассмотрена идея и теория рекуррентного анализа, приведены некоторые примеры, рассмотрены его возможные области применения, предложены возможные модели информационных систем с его использованием.

Фазовое пространство, его реконструкция

Состояние системы описывается ее переменными состояниями

$$x^1(t), x^2(t), \dots, x^d(t),$$

где верхний индекс – номер переменной. Набор из d переменных состояния во время t составляет вектор состояния $\ddot{x}(t)$ в d -мерном фазовом пространстве. Данный вектор перемещается во времени в направлении, определяемом его вектором скорости:

$$\dot{\ddot{x}}(t) = \partial_t \ddot{x}(t) = \ddot{F}(x). \quad (1)$$

Последовательность векторов $\ddot{x}(t)$ образует траекторию в фазовом пространстве, причем поле скорости $\ddot{F}(x)$ касательно к этой траектории. Эволюция траектории описывает динамику системы и ее аттрактор. Зная $\ddot{F}(x)$, можно получить информацию о состоянии системы в момент t путем интегрирования выражения (1). Так как форма траектории позволяет судить о характере процесса (периодические или хаотические процессы имеют характерные фазовые портреты), то для определения состояния систе-

мы не обязательно проводить интегрирование, достаточно построить графическое изображение траектории [2, 3].

При исследовании сложных систем зачастую нет информации обо всех переменных состояниях, либо не все из них возможно измерить. Как правило, имеется единственное наблюдение $u(t)$, проведенное через дискретный временной интервал Δt . Таким образом, измерения записываются в виде ряда $u_i(t)$, где $t = i \cdot \Delta t$.

Интервал Δt может быть постоянным, однако это не всегда возможно и создает проблемы для применения стандартных методов анализа данных, требующих равномерной шкалы наблюдений.

Взаимодействия и их количество в сложных системах таковы, что даже по одной переменной состояния можно судить о динамике всей системы в целом (данный факт был установлен группой американских ученых при изучении турбулентности). Таким образом, эквивалентная фазовая траектория, сохраняющая структуры оригинальной фазовой траектории, может быть восстановлена из одного наблюдения или временного ряда по теореме Такенса (Takens) методом временных задержек:

$$\hat{x}(t) = (u_i, u_{i+\tau}, \dots, u_{i+(m-1)\tau}),$$

где m – размерность вложения, τ – временная задержка (реальная временная задержка определяется как $\tau \cdot \Delta t$). Топологические структуры восстановленной траектории сохраняются, если $m \geq 2 \cdot d + 1$, где d – размерность аттрактора [4]. На практике оказывается, что в большинстве случаев аттрактор может быть восстановлен и при $m \leq 2d$ [5] (см. рис. 1). Задержка τ , как правило, выбираются априорно.

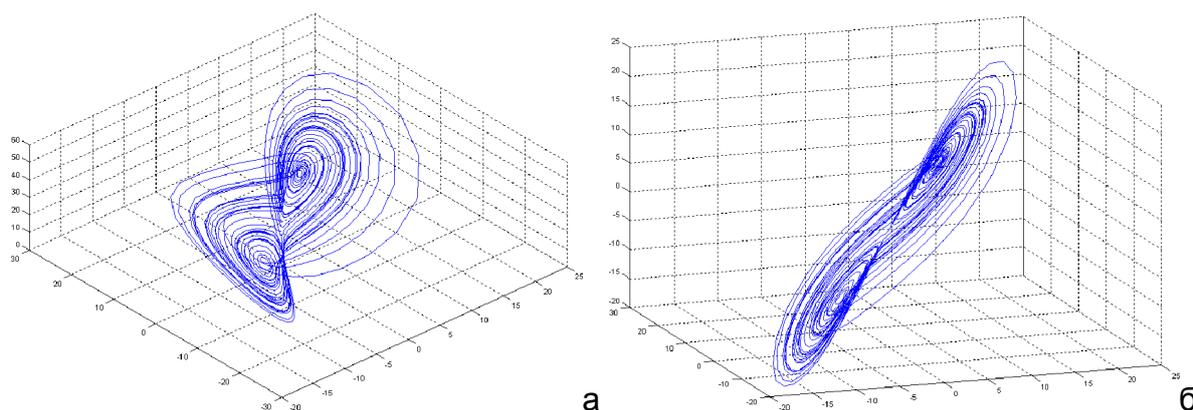


Рис. 1. Аттрактор системы Лоренца в трехмерном фазовом пространстве (проекция подобрана так, чтобы было возможно оценить подобие топологий двух фазовых портретов, длина временного ряда 2000 отсчетов):

а – построен из трех компонент; б – построен из реконструкции x -компоненты в трехмерное фазовое пространство ($m = 3$)

Существует несколько подходов к выбору минимально достаточной размерности m , помимо аналитического. Высокую эффективность показали методы, основанные на концепции ложных ближайших точек (false nearest neighbours, FNN). Суть ее заключается в том, что при уменьшении размерности вложения происходит увеличение количества ложных точек, попадающих в окрестность любой точки фазового пространства. Вытекающий отсюда простейший метод – определение количества FNN как функции от размерности, в пределе равной 0. Существуют и другие методы, основанные на этой концепции – например, определение отношений расстояний между одними и теми же соседними точками при разных m [6, 7]. Размерность аттрактора также может быть определена при помощи кросс-корреляционных сумм [8].

Так как низкая точность измерений и случайные ошибки могут привести к появлению линейной зависимости между последовательными векторами $\ddot{x}(t)$, то задержку τ следует выбирать так, чтобы минимизировать эту зависимость.

Рекуррентный анализ

Процессы в Природе обладают ярко выраженным рекуррентным поведением, таким, как периодичность или иррегулярная цикличность. Более того, рекуррентность (повторяемость) состояний в смысле прохождения последующей траектории достаточно близко к предыдущей является фундаментальным свойством диссипативных динамических систем. Это свойство было отмечено еще в 80-х гг. XIX века французским математиком Пуанкаре (Poincaré) и впоследствии сформулировано в виде «теоремы рекуррентности», опубликованной в 1890 г.:

Если система сводит свою динамику к ограниченному подмножеству фазового пространства, то система почти наверняка, т.е. с вероятностью, практически равной 1, сколь угодно близко возвращается к какому-либо изначально заданному режиму.

Суть этого фундаментального свойства в том, что, несмотря на то, что даже самое малое возмущение в сложной динамической системе может привести систему к экспоненциальному отклонению от ее состояния, через некоторое время система стремится вернуться к состоянию, некоторым образом близкому к предыдущему, и проходит при этом подобные этапы эволюции.

Убедиться в этом можно при помощи графического изображения траектории системы в фазовом пространстве. Однако возможности такого анализа сильно ограничены. Как правило, размерность фазового пространства сложной динамической системы более 3, что делает практически неудобным его рассмотрение напрямую; единственная возможность – проекции в двух- и трехмерные пространства, что зачастую не дает верного представления о фазовом портрете.

В 1987 г. Экман (Eckmann) и соавторы [9] предложили способ отображения m -мерной фазовой траектории состояний системы $\ddot{x}(t)$ длиной N на двумерную квадратную двоичную матрицу размером $N \times N$, в которой 1 (черная точка) соответствует повторению состояния при некотором времени i в некоторое другое время j , а обе координатные оси являются осями времени. Такое представление было названо *рекуррентной диаграммой* (recurrence plot, RP), так как оно фиксирует информацию о рекуррентном поведении системы.

Математически вышесказанное описывается как

$$\mathbf{R}_{i,j}^{m,\varepsilon_i} = \Theta(\varepsilon_i - \|\ddot{x}_i - \ddot{x}_j\|), \quad \ddot{x} \in \mathfrak{R}^m, \quad i, j = 1 \setminus N, \quad (2)$$

где N – количество рассматриваемых состояний x_i , ε_i – размер окрестности точки \ddot{x} в момент i , $\|\cdot\|$ – норма и $\Theta(\cdot)$ – функция Хэвисайда.

Непрактично и, как правило, невозможно обнаружить полную рекуррентность в смысле $\ddot{x}_i \equiv \ddot{x}_j$ (состояние динамической, а особенно – хаотической системы не повторяется полностью эквивалентно начальному состоянию, а подходит к нему сколь угодно близко). Таким образом, рекуррентность определяется как достаточная близость состояния \ddot{x}_j состоянию \ddot{x}_i . Иными словами, рекуррентными являются состояния \ddot{x}_j , попадающие в m -мерную окрестность с радиусом ε_i и центром в \ddot{x}_i . Эти точки \ddot{x}_j называются *рекуррентными точками* (recurrence points).

Поскольку $R_{i,i} = 1 (i = 1 \setminus N)$ по определению, то рекуррентная диаграмма всегда содержит черную диагональную линию – *линию идентичности* (line of identity, LOI), под углом $\pi/4$ к осям координат. Произвольно взятая рекуррентная точка (i, j) не несет какой-либо полезной информации о состояниях во времена i и j . Только вся совокупность рекуррентных точек позволяет восстановить свойства системы.

Внешний вид рекуррентной диаграммы позволяет судить о характере протекающих в системе процессов, наличии и влиянии шума, наличии состояний повторения и замирания (ламинарность), совершении в ходе эволюции системы резких изменений состояния (экстремальных событий) – см. примеры диаграмм на рис. 2.

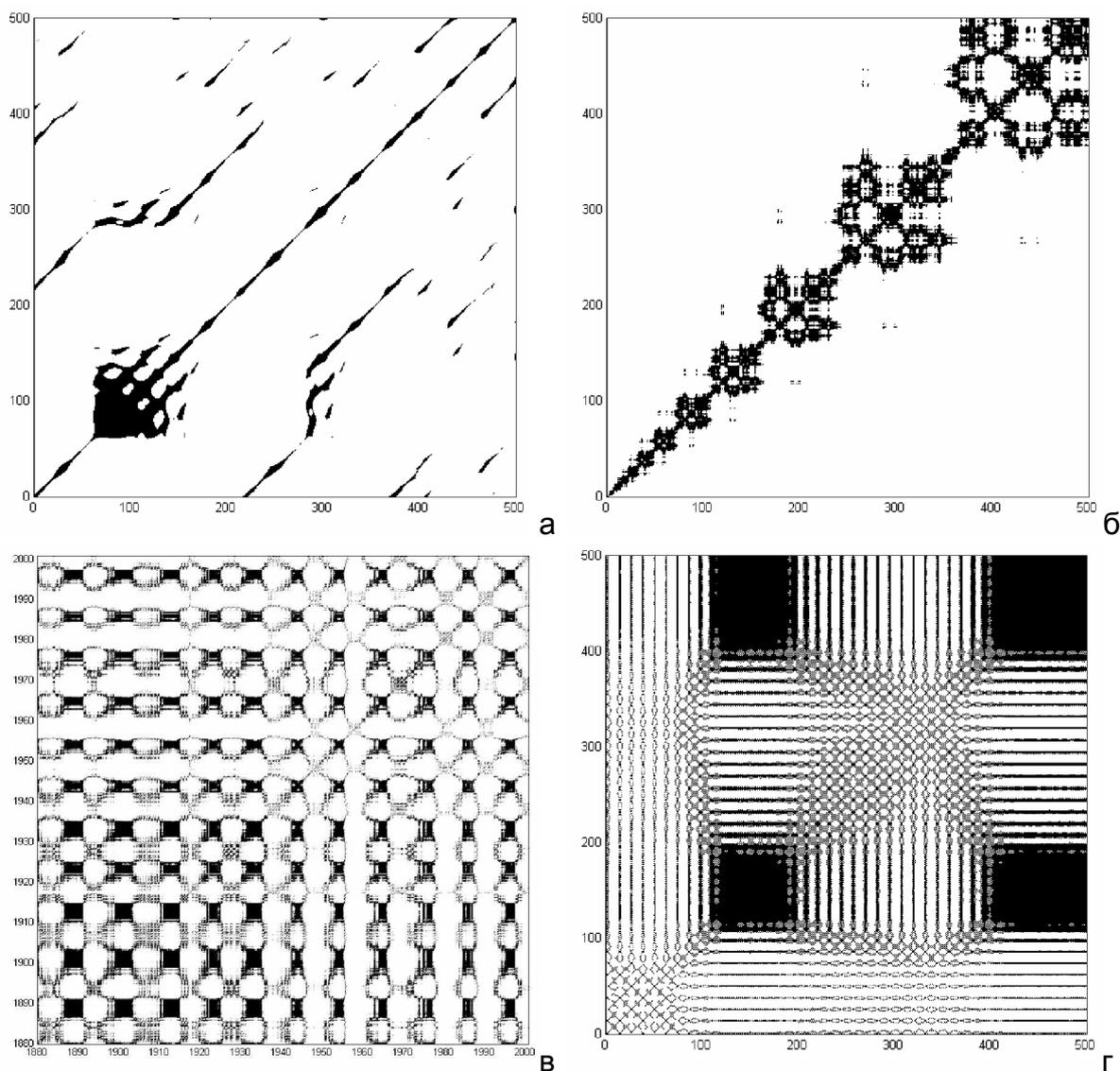


Рис. 2. Примеры рекуррентных диаграмм: а – три компонента системы Лоренца при $\varepsilon = 0,1$; б – множество Мандельброта при $\varepsilon = 0,2$; в – числа Вольфа с 01.1880 года по 12.2000 года (отчетливо виден ~11-ти летний цикл солнечных пятен) при $\varepsilon = 0,1$; г – смоделированный из 12 главных частот временной ряд геомагнитных пульсаций типа Pс-1 при $\varepsilon = 0,15$

Пороговое расстояние ε_i может выбираться как непосредственно для каждой точки, исходя из получения в такой окрестности некоторого заранее определенного количества рекуррентных данной точек (при этом мы получаем несимметричную диаграмму

– $\mathbf{R}_{i,j} \neq \mathbf{R}_{j,i}$), так и оставаться все время постоянным, что дает симметричную диаграмму $\mathbf{R}_{i,j} = \mathbf{R}_{j,i}$ и является наиболее используемым вариантом.

Тип нормы также оказывает влияние на вид диаграммы. Как правило, используются нормы L_1 , L_2 (евклидова норма), L_∞ (максимальная норма). Границы этих норм имеют разные фигуры, что применительно к подходу с постоянным ϵ означает, что норма L_∞ обеспечивает нахождение наибольшего, L_2 – среднего, а L_1 – минимального количества соседних точек (см. рис. 3). Как правило, при построении рекуррентных диаграмм используется норма L_∞ , так как, во-первых, она независима от размерности фазового пространства, во-вторых, наиболее просто вычисляема и, в-третьих, позволяет изучать диаграммы теоретически, поскольку аналитические выражения для L_∞ решаются проще, нежели для любой другой нормы [10]. Независимость от размерности находит применение при анализе реконструированных временных рядов с различными размерностями вложения, т.е. рекуррентные диаграммы разных вложений могут сравниваться напрямую, в то время как для остальных норм требуется масштабирование.

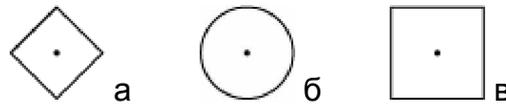


Рис. 3. Основные нормы с одинаковым радиусом вокруг точки фазового пространства (черная точка), изображенные для двумерного фазового пространства:

а – L_1 , б – L_2 , в – L_∞

Следует упомянуть вариант рекуррентной диаграммы – матрицу, на которой отмечаются непосредственно расстояния между двумя точками фазового пространства:

$$\mathbf{D}_{i,j}^m = \|\ddot{x}_i - \ddot{x}_j\|.$$

Данная матрица, будучи отображенной на некоторую цветовую палитру, может служить некоторым аналогом топологической карты – цвет наглядно характеризует расстояние между точками. Такое изображение называется *диаграммой расстояний* (distance plot) и предложено в [11, 12]. Следует отметить, что такие диаграммы могут служить не только прозаическим инструментом исследователя, но и привлекать внимание получаемым орнаментом так же, как и фрактальные рисунки (рис. 4).

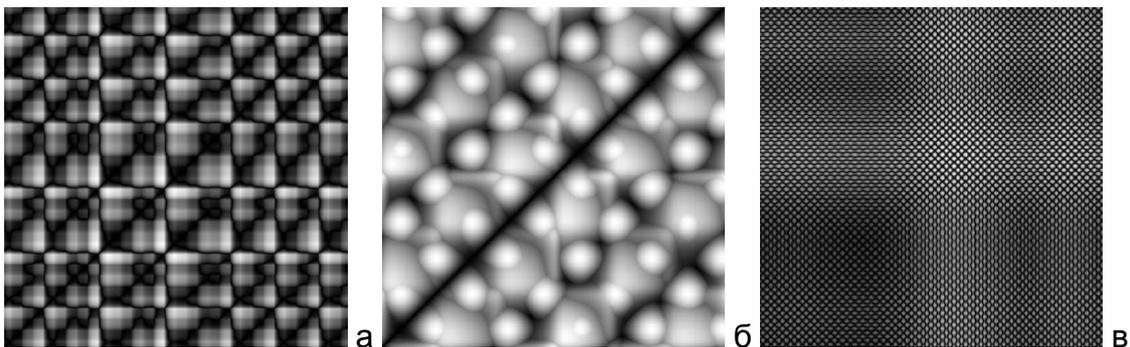


Рис. 4. Примеры диаграмм расстояний (длины временных рядов – тысяча отсчетов):

а – уравнение Меки-Гласса; б – три компонента генератора Ван дер Поля;

в – геомагнитные пульсации типа Pc-1

Существует еще некоторое количество разновидностей рекуррентных диаграмм, разработанных с целью получения дополнительных сведений об объектах исследования, например [11, 13, 14].

Анализ диаграмм

Очевидно, что процессы различного поведения будут давать рекуррентные диаграммы с различным рисунком. Таким образом, зрительная оценка диаграмм может дать представление об эволюции исследуемой траектории. Выделяют два основных класса структуры изображения: *топология* (typology) [9], представляемая крупномасштабными структурами, и *текстура* (texture), формирующаяся мелкомасштабными структурами.

Топология дает общее представление о характере процесса; выделяют четыре различных класса (см. рис. 5):

- *однородные* рекуррентные диаграммы типичны для стационарных и автономных систем, в которых время релаксации мало по сравнению с длиной ряда;
- *периодические* повторяющиеся структуры (диагональные линии, узоры в шахматном порядке) соответствуют различным осциллирующим системам с периодичностью в динамике;
- *дрейф* соответствует системам с медленно изменяющимися параметрами, что делает белыми левый верхний и правый нижний углы рекуррентной диаграммы;
- резкие изменения в динамике системы, равно как и экстремальные ситуации, обуславливают появление *белых областей или полос*. Рекуррентные диаграммы упроощают выявление экстремальных и редких событий.

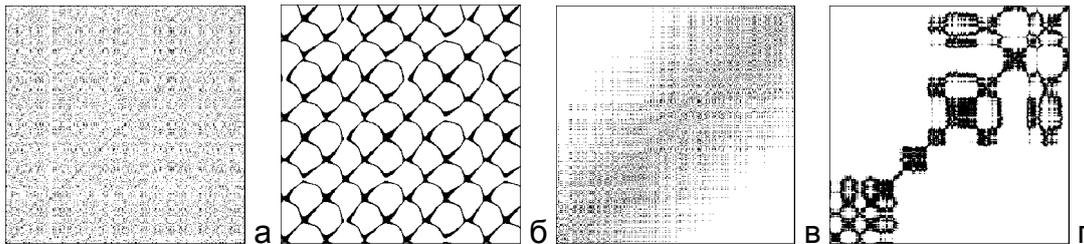


Рис. 5. Характерные топологии рекуррентных диаграмм: а – однородная (нормально распределенный шум); б – периодическая (генератор Ван дер Поля); в – дрейф (отображение Икеды с наложенной линейно растущей последовательностью); г – контрастные области или полосы (обобщенное броуновское движение)

Подробное рассмотрение рекуррентных диаграмм позволяет выявить мелкомасштабные структуры – текстуру, которая составляется из простых точек, диагональных, горизонтальных и вертикальных линий. Комбинации вертикальных и горизонтальных линий формируют прямоугольные кластеры точек.

- *одинокие, отдельно стоящие рекуррентные точки* появляются в том случае, когда соответствующие состояния редки, либо неустойчивы во времени, либо вызваны сильной флуктуацией. При этом они не являются признаками случайности или шума;
- *диагональные линии* $\mathbf{R}_{i+k, j+k} = 1$ (при $k = 1 \dots l$, где l – длина диагональной линии) появляются в случае, когда сегмент траектории в фазовом пространстве пролегает параллельно другому сегменту, т.е. траектория повторяет саму себя, возвращаясь в одну и ту же область фазового пространства в разное время. Длина таких линий определяется временем, в течение которого сегменты траектории остаются параллельными; направление (угол наклона) линий характеризует внутренне время подпроцессов, соответствующих данным сегментам траектории. Прохождение линий параллельно линии идентичности (под углом $\pi/4$ к осям координат) свидетельствует об одинаковом направлении сегментов траектории, перпендикулярно – о противоположном («отраженные» сегменты), что может

также являться признаком реконструкции фазового пространства с несоответствующей размерностью вложения. Нерегулярное появление диагональных линий является признаком хаотического процесса;

- *вертикальные (горизонтальные) линии* $\mathbf{R}_{i,j+k} = 1$ (при $k = 1 \setminus \nu$, где ν – длина вертикальной или горизонтальной линии) выделяют промежутки времени, в которые состояние системы не изменяется или изменяется незначительно (система как бы «заморожена» на это время), что является признаком «ламинарных» состояний.

Перечисленные мелкомасштабные структуры используются для вычисления мер *количественного анализа* рекуррентных диаграмм (recurrence quantification analysis, RQA). Збилут (Zbilut) и Веббер (Webber) разработали инструмент [15] вычисления ряда мер на основе подсчета плотности рекуррентных точек и построения частотного распределения длин диагональных линий: *recurrence rate* (RR, коэффициент самоподобия), *determinism* (DET, предсказуемость), *divergence* (DIV, обратная величина максимальной длины диагональной линии), *entropy* (ENTR, энтропия) и *trend* (TREND, тренд). Вычисление этих мер в подматрицах рекуррентной диаграммы вдоль линии идентичности показывает поведение этих мер во времени. Некоторые исследования данных мер показали, что их применение может содействовать обнаружению точек бифуркации, особенно переходов хаос–порядок [16].

Позднее были предложены меры, использующие плотность вертикальных (или горизонтальных) структур [17]: *laminarity* (LAM, замирение), *trapping time* (TT, показатель задержки), позволяющие выявлять переходы хаос–хаос. Следует отметить также нечувствительность этих мер к стационарности и длине исследуемых рядов.

Мы не будем в настоящей работе подробно рассматривать эти меры и формулы для их вычисления. При необходимости они могут быть обнаружены в упомянутой литературе.

Кросс-рекуррентные диаграммы

При построении рекуррентной диаграммы мы анализируем взаимную близость точек одной траектории x_i длиной N_x . Очевидно, что аналогичное сравнение можно провести для двух временных рядов. Добавим в то же фазовое пространство траекторию y_j длиной N_y . Графическое отображение сравнения близости точек первой траектории с точками второй траектории называется кросс-рекуррентной диаграммой (cross-recurrence plot, CRP):

$$\mathbf{CR}_{i,j}^{m,\varepsilon_i} = \Theta(\varepsilon_i - \|\ddot{x}_i - \ddot{y}_j\|), \quad \ddot{x}, \ddot{y} \in \mathfrak{R}^m, \quad i = 1 \setminus N_x, \quad j = 1 \setminus N_y.$$

Это выражение полностью аналогично выражению (2) для рекуррентной диаграммы. Если состояние второй траектории во время j близко состоянию первой траектории во время i , то 1 (черная точка) будет установлена в матрице \mathbf{CR} в положении (i, j) . Важно, что такая ситуация не означает «повторение» состояния, и таким образом, матрица \mathbf{CR} отображает не повторяемость, а *соответствие* траекторий друг другу. Таким образом, получаемая диаграмма не является «кросс-рекуррентной диаграммой» в буквальном смысле. Однако это название сложилось исторически, используется исследователями и учеными, так как удобно с целью обобщения понятия рекуррентных диаграмм. Аналогичным образом обобщается понятие количественного анализа к cross-recurrence quantification analysis (CRQA).

Траектории x_i и y_j не обязательно должны быть одинаковой длины – таким образом, кросс-рекуррентная диаграмма может быть произвольной прямоугольной формы. Для упрощения изложения примем, что мы имеем дело с рядами одинакового масштаба

(т.е. одинаковой длины N_x и интервала измерения Δt), таким образом, кросс-рекуррентная диаграмма принимает вид массива размера $N \times N$.

Траектории для построения кросс-рекуррентных диаграмм должны отображать значения переменных состояния или компонент подобных динамических систем. Для устранения проблемы единиц измерения и масштаба значений ряды следует нормировать на максимальное по модулю значение. Бессмысленно использовать кросс-рекуррентные диаграммы для анализа данных, полученных от принципиально разных динамических систем, хотя зачастую диаграммы расстояний, получаемые от такого «смешивания», могут быть весьма оригинальны и занимательны.

Так как значения $CR_{i,i}$ ($i = 1 \setminus N$) не обязательно отличны от нуля, то, как правило, главная диагональ (LOI) размывается. Сказанное выше относительно толкования топологии рекуррентных диаграмм справедливо и для кросс-рекуррентных диаграмм, причем диагональные линии в данном случае представляют куда больший интерес, так как указывают на проходящие близко участки траекторий двух *разных* процессов. Частота и длины этих линий очевидным образом указывают на подобие динамики исследуемых систем. Увеличивающееся со временем подобие рассматриваемых траекторий приводит к увеличению плотности рекуррентных точек вдоль главной диагонали $CR_{i,i}$ ($i = 1 \setminus N$) до появления на ней сплошной черной линии (которая, фактически, является линией идентичности, а кросс-рекуррентная диаграмма в области старших индексов превращается в обычную рекуррентную). Аналогичные рассуждения, только в обратном порядке, справедливы и для двух расходящихся траекторий.

Данное расширение рекуррентного анализа было предложено в работе [18].

Области применения

На первый взгляд применение рекуррентного анализа может показаться ограниченным. Однако это не так. Приведем некоторые возможные области применения рекуррентного анализа.

1. Изучение:
 - а. самой методики;
 - б. динамических систем, например, сложных физико-химических процессов, на предмет выявления хаотического поведения, природных явлений, динамики экономических индексов и др.
2. Мониторинг состояния систем, протекания процессов (в том числе в реальном времени после каждого рабочего цикла). В процессе работы системы производятся измерения каких-либо ее параметров, характеризующих протекающий процесс. После окончания каждого рабочего цикла производится построение рекуррентной (или кросс-рекуррентной) диаграммы и вычисление мер. Если меры выходят за определенные пределы, то это может свидетельствовать о какой-либо неисправности.
3. Распознавание (отнесение к определенному классу):
 - а. определение неисправностей сложных систем (в том числе заболеваний человека и животных). В отличие от мониторинга, данные испытания проводятся в индивидуальном порядке, что позволяет строить вывод о неисправности на основе сопоставления полученного результата с базой данных (см. далее);
 - б. фильтрация, верификация входных или выходных данных. Предполагает использование методики в составе более сложного комплекса. Задача может состоять в проверке, удовлетворяет ли временной ряд, подаваемый на вход комплекса, некоторым обязательным критериям. Другой ва-

риант задачи – проверка соответствия результата некоторым обязательным критериям (также может являться детектором неисправностей).

Рассмотрим последовательность обработки сигнала рекуррентным анализом (рис. 6). На вход системы анализа подается d_x -мерный временной ряд $X_{N_x}^{d_x}$ длиной N_x отсчетов. Если $d = 1$, то предварительно может быть проведена реконструкция фазового пространства (секция RPS). Затем производится построение рекуррентной диаграммы и вычисление ее мер. Полученные меры в совокупности с самой диаграммой формируют *характеристический вектор* (CV) временного ряда, который далее может быть использован не только для вывода о характере процесса, но и сохранен в базу данных для последующего использования.

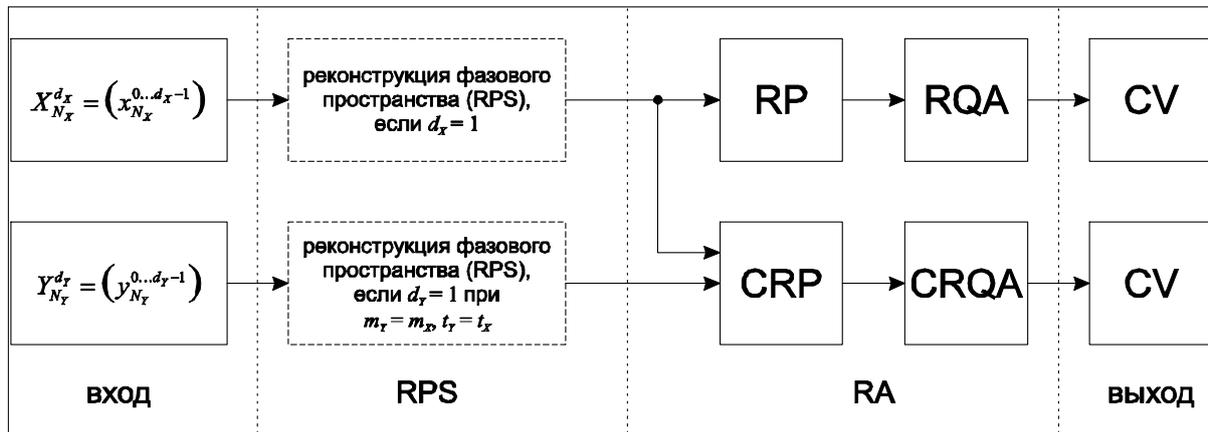


Рис. 6. Последовательность обработки сигнала при рекуррентном анализе

Аналогично строится обработка двух сигналов с использованием кросс-рекуррентных диаграмм. Теперь на входе системы анализа имеется еще один d_y -мерный временной ряд $Y_{N_y}^{d_y}$ длиной N_y отсчетов (как правило, данный ряд является эталонным, и стоит задача выявить подобие ему ряда X). Если применяется реконструкция фазового пространства обоих рядов, то размерности вложения должны совпадать ($m_x = m_y$). Если реконструируется только один ряд, то его следует приводить к размерности второго. Построенная кросс-рекуррентная диаграмма и вычисленные на ее основе меры формируют *характеристический вектор подобия*, который также может быть сохранен в базу данных для последующего использования.

Заключение

В данной статье был кратко рассмотрен один из перспективных методов анализа динамических систем – метод (кросс-)рекуррентных диаграмм. Были описаны основные критерии визуальной оценки диаграмм. Упомянут вопрос количественного анализа диаграмм, но без конкретных формул вычисления мер (при необходимости их можно найти в упомянутой литературе).

Увеличивающееся количество работ, посвященных практическому применению данного метода, свидетельствует о росте его популярности среди исследователей. Метод применяется в геофизике, палеонтологии, медицине и даже для анализа динамики макроэкономических индексов. Преимущества данного метода – одинаковая работа с данными любой размерности, не критичность к происхождению и качеству данных, маргинальность (метод не вносит дополнительную информацию «от себя»), наглядность результатов – позволяют применять его достаточно широко при анализе и изучении сложных динамических систем совершенно различной природы.

К сожалению, сравнение результатов разных исследований затруднено из-за практически полного отсутствия общих критериев для анализа систем, например, выбора размерности вложения и задержки при реконструкции фазового пространства. Хотя такая несогласованность, как правило, не влияет на результат визуальной (качественной) оценки, она затрудняет систематизацию и обобщение результатов.

Была дана оценка возможных областей применения метода.

Предложенные информационные модели могут найти применение в реальных аналитических и контролируемых системах – как индивидуально, так и в сочетании с другими методами анализа.

Литература

1. Manuca R., Savit R. Stationarity and nonstationarity in time series analysis. // *Physica D* 1996. 99 (2–3). P. 134–161.
2. Берже П., Помо И., Видадь К. Порядок в хаосе. О детерминированном подходе к турбулентности. М.: Мир, 1991. 368 с.
3. Мун Ф. Хаотические колебания: Вводный курс для научных работников и инженеров. М.: Мир, 1990. 312 с.
4. Takens F. Detecting Strange Attractors in Turbulence. Berlin, Springer, 1981. P. 366–381.
5. Киселев Б.В., Волобуев Д.М. Реконструкция аттракторов в трехмерном фазовом пространстве и моделирование геомагнитных пульсаций. // *Вопросы геофизики*. СПб: Изд-во С.-Петербургского ун-та, 1998. Вып. 35. С. 338–348.
6. Kennel M.B., Brown R., Abarbanel H.D.I. Determining embedding dimension for phase-space reconstruction using a geometrical construction. // *Physical Review*. 1992. A 45 (6). P. 3403–3411.
7. Cao L. Practical method for determining the minimum embedding dimension of a scalar time series. // *Physica D*. 1997. 110 (1–2). P. 43–50.
8. Grassberger P., Procaccia I. Measuring the strangeness of strange attractors. // *Physica D*. 1983. 9 (1–2). P. 189–208.
9. Eckmann J.-P., Kamphorst S.O., Ruelle D. Recurrence Plots of Dynamical Systems. // *Europhysics Letters*. 1987. 5. P. 973–977.
10. Thiel M. et al. Influence of observational noise on the recurrence quantification analysis. // *Physica D*. 2002. 171 (3). P. 138–152.
11. Iwanski J.S., Bradley E. Recurrence plots of experimental data: To embed or not to embed? // *Chaos*. 1998. 8 (4). P. 861–871.
12. Webber Jr.C.L. Recurrence Quantification Analysis, 2003. homepages.luc.edu/~cwebber
13. Casdagli M.C. Recurrence plots revisited. // *Physica D*. 1997. 108 (1–2). P. 12–44.
14. Manuca R., Savit R. Stationarity and nonstationarity in time series analysis. // *Physica D*. 1996. 99(2–3). P. 134–161.
15. Zbilut J.P., Webber Jr.C.L. Embeddings and delays as derived from quantification of recurrence plots. // *Physics Letters*. 1992 A 171 (3–4). P. 199–203.
16. Trulla L.L., Guilianì A., Zbilut J.P., Webber Jr.C.L.: Recurrence quantification analysis of the logistic equation with transients. // *Physics Letters*. 1996. A 233 (4). P. 255–260.
17. Marwan N., Wessel N., Meyerfeldt U., Schirdewan A., Kurths J. Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. // *Physical Review*. 2002. E66. 026702.
18. Zbilut J.P., Giuliani A., Webber Jr.C.L. Detecting deterministic signals in exceptionally noisy environments using cross-recurrence quantification. // *Physics Letters*. 1998. A 246 (1–2). P. 122–128.

МЕТОДЫ И АЛГОРИТМЫ СЖАТИЯ ГРАФИЧЕСКОЙ ИНФОРМАЦИИ

М.С. Тулякова

Научный руководитель – кандидат технических наук, доцент Б.А. Крылов

К настоящему времени разработано множество алгоритмов сжатия, ориентированных на различные типы изображений. В статье описаны основные методы и алгоритмы сжатия графической информации.

Введение

Сжатие – способ более эффективного представления информации. Эффективность сжатия изображения определяется уровнем его избыточности. В цифровых изображениях существуют различные типы избыточности. Первый тип обусловливается повторением в файле изображения пикселей одинаковых цветов, преобладанием какого-либо цвета (статистическая избыточность). Такая избыточность характерна для деловой графики (диаграммы, графики). Другой тип избыточности является результатом корреляции пикселей. Такая избыточность называется пространственной и чаще всего присутствует в фотореалистичных изображениях. Также для фотореалистичных изображений характерна психофизическая избыточность. Она связана с тем, что не все детали изображения одинаково воспринимаются зрителем. Ряд деталей, имеющих одновременно малые размеры и малый контраст, не видны на изображении, поэтому могут быть удалены без вреда для качества восприятия изображения.

Невозможно в общем виде сказать, имеет алгоритм сжатия высокую или низкую эффективность. Эффективность компрессии зависит от типа сжимаемого изображения. Можно выделить следующие типы:

- *Двухуровневое (монокроматическое) изображение.* Каждый пиксель изображения может принимать только одно из двух значений. Это самый простой тип изображения, где каждый пиксель изображения представлен одним битом.
- *Полутонное изображение.* Значение каждого пикселя в этом случае представляет собой яркость соответствующей точки. Полутонное изображение может иметь 2^n ($n = 4, 8, 12, 16, 24$) градаций серого.
- *Полноцветное изображение.* Каждый пиксель такого изображения кодируется 24 битами, которые делятся на три 8-битовые группы. Эти группы содержат значения, которые управляют информацией о красном, зеленом и голубом цвете и их интенсивности. Типы данных истинного цвета RGB содержат 16.7 миллионов цветов.
- *Изображение с непрерывным тоном.* Этот тип изображения характеризуется тем, что может иметь много похожих цветов. Изображения такого типа чаще всего являются фотореалистичными.
- *Изображения с небольшим количеством цветов.* Примерами таких изображений могут быть объекты деловой графики – гистограммы, диаграммы, графики и т.п.

Методы сжатия можно разделить на две основные группы: с потерями и без потерь. Методы сжатия без потерь сжимают и восстанавливают изображение без искажений, т.е. исходное и восстановленное изображения идентичны. Но в связи с тем, что эти методы имеют очень маленький коэффициент сжатия, они используются только в том случае, когда искажение информации недопустимо. Эти методы сжатия устраняют статистическую избыточность.

Сжатие без потерь

RLE. Одним из самых простых алгоритмов сжатия без потерь является кодирование серий последовательностей – алгоритм RLE. На первом этапе реализации этого ал-

горитма происходит «вытягивание» двумерного файла изображения в одномерную цепочку бит. Дальнейшее кодирование осуществляется за счет наличия в этой цепочке повторяющихся блоков бит. В закодированный файл записываются пары «счетчик повторений – значение повторяющегося байта» [1].

Этот алгоритм рассчитан на изображения с большими областями постоянной яркости или цвета и маленьким количеством цветов. Примерами таких изображений может служить деловая графика – диаграммы, графики. К минусам данного алгоритма можно отнести его неэффективность применительно к фотореалистичным изображениям. В этом случае возможно даже увеличение размера заархивированного файла относительно исходного.

Групповое кодирование используется в таких форматах bitmap файлов, как BMP, PCX, TIFF.

Метод Хаффмана. Этот алгоритм практически не используется в чистом виде, а применяется как один из этапов компрессии в более сложных схемах. В его основе лежит учет частоты появления одинаковых пикселей в изображении. При этом пикселям исходного изображения, которые встречаются часто, сопоставляется код меньшей длины, а встречающимся редко – код большей длины [2]. Кодирование по Хаффману нуждается в точной статистике, касающейся того, как часто каждая величина появляется в исходном файле. Таким образом, для обеспечения правильной работы метод Хаффмана реализуется в два прохода. В первый проход собирается статистика частоты появления символов; во второй – кодируются сами символы. В этом случае в дополнение к закодированному изображению в файле необходимо хранить таблицу с соответствиями кодируемых пикселей и кодирующих цепочек битов. Существует вариант реализации этого алгоритма, при котором такая таблица строится адаптивно, в процессе архивации.

Рассмотрим пример сжатия по Хаффману (табл. 1). Пусть к, с, ч, з, ж, ф – цвета пикселей (красный, синий, черный, зеленый, желтый, фиолетовый). Входной поток: «ксссччззжжжжжжжжжжф». Сначала для каждого символа подсчитывается частота его появления в этом потоке. Далее строится двоичное дерево. Наиболее редко используемые значения в этом примере – «к» и «ф», так что они становятся первой парой; «к» присваивается 0-я ветвь, а «ф» – 1-я. Это означает, что 0 и 1 будут младшими битами кодов для «к» и «ф», соответственно. Более старшие биты будут получены из дерева по мере того, как оно будет построено. Затем частоты этих первых 2-х суммируются, что дает в итоге 2. Поскольку теперь самая низкая частота – 2, эта пара объединяется с «з» (которое также имеет частоту 2). Исходной паре присваивается нулевая ветвь этого дерева, а «з» – первая ветвь. Таким образом, код для «к» заканчивается на 00; «ф» – на 01; «з» – на 1 и будет на один бит короче по сравнению с кодами для «к» и «ф».

Символ	Частота	Выходная последовательность
К	1	0000
Ф	1	0001
З	2	001
С	3	00
Ч	3	01
Ж	9	1

Таблица 1. Сжатие Хаффмана для входного потока пикселей

Алгоритм Хаффмана эффективен при сжатии изображений с ограниченным количеством цветов пикселей. Сжатие с фиксированной таблицей применяется в качестве последнего этапа архивации в JPEG. Этот алгоритм реализован также в формате TIFF.

LZW. Это более современный алгоритм сжатия без потерь, чем алгоритм Хаффмана. В отличие от метода Хаффмана, в этом алгоритме не требуется предварительного создания таблицы кодов, этот алгоритм является адаптивным, и оптимальная таблица кодов создается в процессе сжатия. Сжатие происходит за счет наличия в сжимаемом файле повторяющихся цепочек пикселей.

Первый шаг этого алгоритма – создание таблицы с одним элементом кода для каждой возможной величины данных; например, 256 элементов для 8-битных данных. Далее алгоритм «просматривает» входной поток, и в эту таблицу добавляется каждое уникальное сочетание пикселей в сжимаемом изображении. Если на вход поступает уже имеющаяся комбинация пикселей, то в выходной поток записывается индекс ячейки в таблице, в которой эта комбинация содержится. Таким образом, алгоритм преобразует поток символов на входе в поток индексов ячеек таблицы на выходе. При переполнении таблицы алгоритм либо прекращает ее заполнение, либо очищает (полностью или частично) [1].

Любая ячейка таблицы, кроме самых первых, содержащих односимвольные цепочки, хранит копию некоторой другой ячейки, к которой в конец приписан один символ. Вследствие этого часто при сжатии используют «хэш-таблицу», в каждом элементе которой содержится код предыдущей подстроки, добавленный символ, код этой строки.

Алгоритм LZW эффективен при сжатии 8-битных изображений, имеющих небольшое количество цветов. Наличие шума в изображении значительно снижает эффективность сжатия. Существует огромное количество различных реализаций алгоритма сжатия LZW. Метод LZW используется для сжатия изображений при записи их в форматы TIFF и GIF.

Арифметическое сжатие. Алгоритм арифметического сжатия является наиболее оптимальным, так как достигается теоретическая граница степени сжатия. Этот метод позволяет упаковывать значения пикселей входного потока без потерь при условии, что известно распределение частот этих пикселей.

Сжимаемая последовательность значений пикселей представляется как некоторая двоичная дробь из интервала $[0, 1)$. Для сжатия цепочки пикселей определяют частоту появления каждого символа во входном потоке. Интервал $[0, 1)$ разбивается на подинтервалы с длинами, пропорциональными этим частотам. Далее за основу разбиения берется подинтервал, соответствующий частоте появления первого пикселя из последовательности, и также разбивается на подинтервалы, пропорциональные вероятностям появления пикселей. Далее берется подинтервал, соответствующий частоте появления следующего пикселя из входной последовательности, и также разбивается. Процедура повторяется. В качестве выходной последовательности берется любая дробь из окончательно определенного интервала [1].

Рассмотрим пример арифметического сжатия. Пусть k, c, z – цвета пикселей (красный, синий, зеленый). Входной поток: «кзкк». Вычисляется частота появления символа, и в соответствии с ней интервал $[0, 1)$ делится на подинтервалы. Вычисленные значения указаны в табл. 2. Первый пиксель – k , поэтому для дальнейшего разбиения выбираем интервал $[0.0; 0.5)$. Разбиваем этот интервал на подинтервалы и выбираем тот, который соответствует c – $[0.25; 0.375)$. Разбиваем этот интервал на подинтервалы, выбираем соответствующие пикселю z – $[0.34375; 0.375)$. Разбиваем этот интервал на подинтервалы еще раз и выбираем подинтервал, соответствующий пикселю k – $[0.34375; 0.359375)$. Таким образом цепочку «кзкк» можно представить любым числом из этого подинтервала, например 0,35.

Система арифметического кодирования лучше сжимает изображения, которые в большинстве своем состоят из повторенной одной и той же последовательности пикселей. Этот алгоритм может сжимать довольно короткие цепочки пикселей из-за ограничений разрядности переменных.

Символ	Частота	Первое деление на подинтервалы	Второе деление на подинтервалы
K	0.5	[0.0; 0.5)	[0; 0.25)
C	0.25	[0.5; 0.75)	[0,25; 0.375)
З	0.25	[0.75; 1.0)	[0.375; 1)
	Третье деление на подинтервалы	Четвертое деление на подинтервалы	
	[0.25; 0.3125)	[0.34375; 0,359375)	
	[0.3125; 0.34375)	[0,359375; 0,3671875)	
	[0.34375; 0.375)	[0,3671875; 0.375)	

Таблица 2. Вычисление интервалов при арифметическом сжатии

Сжатие с потерями

Методы сжатия с потерями – это методы, при которых некоторая часть исходных данных отбрасывается, поскольку считается необязательной для конечного приложения. Методы сжатия с потерями применяются в основном к фотореалистичным изображениям. Это связано с тем, что, во-первых, такие изображения плохо поддаются методам сжатия без потерь, а во-вторых, фотореалистичные изображения обладают большой пространственной и психофизической избыточностью. Эти типы избыточности как раз и устраняются методами сжатия с потерями.

JPEG. Особенностью сжатия алгоритмом JPEG является то, что он сжимает изображения, опираясь на то, что видит человеческий глаз, и удаляет информацию, которая несущественна для восприятия изображения. В результате, сжимая изображение алгоритмом JPEG, можно достигнуть высокого уровня сжатия без заметных потерь в качестве изображения.

Сжатие JPEG происходит в несколько этапов. Первый этап заключается в преобразовании цветовой схемы RGB в YUV (Y-компонента – это яркость, а две остальные – это голубая и красная цветовые компоненты). Цветное изображение в системе RGB представляется как результат сложения трех компонент

$$X_C = a_1 X_R + a_2 X_G + a_3 X_B.$$

В этом выражении a_1, a_2, a_3 – калориметрические коэффициенты. Переход от цветовой RGB-палитры к компонентам YUV выполняется по формуле (1):

$$Y = \sqrt{a_1 [X_R]^2 + a_2 [X_G]^2 + a_3 [X_B]^2}, \quad (1)$$

$$U = X_B,$$

$$V = X_R.$$

После такого преобразования получается исходное изображение в градациях серого плюс два канала информации о голубом и красном цветах. Хотя этот этап не является обязательным для JPEG-сжатия, большинство реализаций производят эту операцию, потому что она позволит повысить эффективность сжатия. Это связано с тем, что человеческий глаз гораздо сильнее реагирует на незначительно изменение яркости, чем на изменение цветового тона, а, разделив яркостную и цветовую компоненты, можно сжать каждую отдельно с разными потерями.

Следующим этапом алгоритма JPEG является дискретизация. Исходное изображение разбивается на блоки, размером 8×8 пикселей. На этом этапе используется то, что яркостная и цветовая составляющие изображения хранятся отдельно. Для увеличения степени сжатия дискретизация проводится для разных составляющих с разной частотой – изображение делится по компоненте Y, а для компонент U и V цветовая ин-

формация берется через строку и через столбец. Таким образом, сразу получаем степень сжатия в два раза, а на качестве изображения это сказывается не сильно [2].

Основным этапом работы алгоритма является дискретное косинусное преобразование (ДКП), позволяющее переходить от пространственного представления изображения к спектральному. Обработка каждого блока выполняется независимо и заключается в применении к каждой рабочей матрице формулы

$$F(u, v) = \frac{1}{4} C(u) C(v) \left[\sum_{x=0}^7 \sum_{y=0}^7 f(x, y) \cos \frac{(2x+1) u \pi}{16} \cos \frac{(2y+1) v \pi}{16} \right], \quad (2)$$

где

$$C(z) = \begin{cases} \frac{1}{\sqrt{2}}, & \text{если } z = 0 \\ 1, & \text{если } z \neq 0 \end{cases}.$$

В получившейся матрице коэффициентов низкочастотные компоненты расположены ближе к левому верхнему углу, а высокочастотные – справа и внизу. Для восстановления сжатого изображения гораздо более важны значения низких частот, чем значения высоких, потому что медленные изменения цвета в изображении более заметны, чем быстрые.

На этом основывается следующий этап сжатия методом JPEG – квантование. Именно на этом этапе происходит самая значительная потеря информации. Квантование представляет собой деление каждого полученного после ДКП блока коэффициентов на матрицу квантования поэлементно. Для каждой компоненты матрица квантования своя. Значение коэффициентов этой матрицы определяет степень сжатия изображения. Поскольку для U и V компонентов квантование может быть более грубым, чем для Y компонента, на последнем этапе процесса сжатие U и V компонентов происходит в большей степени. После квантования большинство значений в блоке становятся нулями, сохраняются только небольшое количество наиболее значимых коэффициентов, к тому же эти значения относительно невелики. Квантование также обеспечивает возможность последующего эффективного сжатия данных при помощи любого способа сжатия без потерь. Для этого блок 8×8 обходится при помощи зигзаг-сканирования, все коэффициенты вытягиваются в цепочку. В начале этой цепочки находятся коэффициенты, соответствующие низким частотам, а в конце – высоким. Далее к получившейся цепочке коэффициентов применяется RLE-кодирование, получившиеся в результате пары сжимаются по методу Хаффмана [5].

JPEG – один из самых популярных алгоритмов сжатия. Он обеспечивает высокую эффективность сжатия при приемлемом уровне потерь. JPEG применяется, в основном, к фотореалистичным изображениям. Основной плюс этого алгоритма – это высокие коэффициенты сжатия изображений. К минусам можно отнести высокую вычислительную сложность и специфические искажения – эффект Гиббса. Этот эффект проявляется в том, что линии JPEG изображений с четкими контурами начинают заметно «дрожать». Так, например, если изображение содержит какие-либо подписи, то подобный эффект может возникнуть вокруг символов.

Фрактальное сжатие. Различные методы сжатия изображений основываются на устранении тех или иных форм избыточности, в частности, фрактальные методы рассматривают самоподобие как источник избыточности. Считается, что самоподобие является свойством почти всех природных объектов и их изображений, и, следовательно, устранение этой формы избыточности может значительно уменьшить объем данных, необходимых для описания природного объекта или его изображения.

Фрактал – это структура, выделенная при анализе изображения и обладающая схожей формой независимо от ее размеров. Например, в изображении кроны дерева фрактал – изображение листа. Фрактальное сжатие изображений основано на гипотезе,

согласно которой в любом изображении можно обнаружить локальное самоподобие различных его частей.

При фрактальном сжатии изображение представляется в более компактной форме – с помощью коэффициентов системы итерируемых функций (IFS). IFS представляет собой набор трехмерных аффинных преобразований, переводящих одно изображение в другое. Преобразованию подвергаются точки в трехмерном пространстве (x -координата, y -координата, яркость). Фактически фрактальная компрессия – это поиск самоподобных областей в изображении и определение для них параметров аффинных преобразований.

Функция аффинного преобразования описывается следующим образом:

$$f(x, y) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \cdot \begin{vmatrix} x \\ y \end{vmatrix} + \begin{vmatrix} l \\ f \end{vmatrix},$$

$$f(x) = ax + by + l,$$

$$f(y) = cx + dy + f.$$

Здесь (x, y) – координата пикселя исходного изображения. Первая матрица определяет масштабирование и трансформацию исходного изображения (a, b, c и d – аффинные коэффициенты вращения, деформации, расширения и сжатия). Вторая матрица определяет величину параллельного переноса (l и f – коэффициенты перемещения).

Кодируемое изображение разбивается на множество неперекрывающихся блоков (ранговых областей), для каждого из которых, в пределах этого же изображения, ищется блок большего размера (домен), пиксели которого путем некоторого преобразования, задаваемого несколькими коэффициентами, переводились бы в пиксели ранговой области. Из преобразований, переводящих домены в ранговые области, формируется отображение, переводящее изображение в изображение. Если для какого-либо домена не находится подобного среди ранговых областей, то этот домен разбивают на части и ищут подобные ранговые области уже для этих частей. Кодом изображения будут являться местоположение и размеры ранговых областей, а также коэффициенты преобразований, описывающих самоподобие внутри изображения [5]. Количество бит, необходимых для описания кода, будет существенно меньше количества бит, необходимых для описания исходного изображения. Параметры аффинного преобразования кодируются по методу Хаффмана и сохраняются. Файл со сжатым изображением содержит две части: заголовок, содержащий информацию о расположении доменов и ранговых областей, и упакованную таблицу аффинных коэффициентов для каждого домена.

Алгоритм фрактального сжатия несимметричен по времени – на сжатие изображения требуется примерно в 1000–10000 раз больше времени, чем на распаковку. Это объясняется большими затратами времени на поиск в изображении подобных областей.

Алгоритм ориентирован на полноцветные изображения и изображения в градациях серого цвета, без резких переходов цветов. Фрактальное сжатие реализовано в формате FIF. К плюсам этого метода сжатия можно отнести высокие коэффициенты сжатия и возможность свободно масштабировать изображение при разжатии, увеличивая его в 2–4 раза. К минусам можно отнести зависимость результатов работы алгоритма от принципов отбора базовых элементов и доменов и то, что коэффициент сжатия зависит от повторяемости базовых элементов.

JPEG2000. Несмотря на большую популярность алгоритма JPEG, по прошествии нескольких лет был разработан новый алгоритм JPEG2000. Это связано с тем, что были разработаны новые, более эффективные, методы сжатия изображений, в частности wavelet-преобразование.

Базовая схема JPEG2000 похожа на схему JPEG, но есть ряд отличий. Основным является то, что вместо дискретного косинусного преобразования (ДКП) используется дискретное wavelet-преобразование. Применение wavelet-преобразования имеет ряд преимуществ.

шеств по сравнению с ДКП, таких как увеличение коэффициента сжатия и избавление от блочной структуры изображения, возникающей в сильно сжатых JPEG файлах.

Первым этапом сжатия в алгоритме JPEG2000, также как и в JPEG, является преобразование цветовой схемы из RGB в YUV. После этого к каждой компоненте исходного изображения применяется wavelet-преобразование. Вначале исходное изображение подвергается фильтрации посредством четырех цифровых фильтров, которые делят спектр исходного изображения на четыре равные неперекрывающиеся области. Низкочастотный фильтр пропускает информацию, передаваемую на низких частотах (т.е. малое количество деталей), блокируя информацию, передаваемую на высоких частотах (т.е. содержащую большое количество деталей), а высокочастотный фильтр – наоборот. В итоге получается четыре области: первая представляет собой «размытую» версию исходного изображения (низкочастотный фильтр), три оставшихся – высокочастотные компоненты изображения, которые имеют вид горизонтальных, вертикальных и диагональных границ. Все четыре изображения подвергаются прореживанию путем устранения каждой второй строки и каждого второго отсчета в строке, а затем размещаются на той же площади, что и исходное изображение. Далее в пределах первой области, которая была отфильтрована низкочастотным фильтром, повторяют эту операцию. Количество повторений соответствует количеству уровней вейвлет-декомпозиции.

Следующим шагом wavelet-преобразования, так же как и JPEG-сжатия, является квантование. Низкочастотные компоненты квантуются на меньшее число уровней, высокочастотные – на большее. Процесс квантования осуществляется путем деления матрицы отсчетов на матрицу квантования [4]. При квантовании отсчетов компонент, полученных в результате wavelet-преобразования, на пониженное число уровней в них вносится шум квантования, но если сжатие не слишком велико, то шумы в восстановленном изображении будут не очень заметны.

Далее для сжатия получившегося массива данных используется арифметическое сжатие. Кодироваться не все изображение и даже не отдельные субполосы, а более мелкие объекты – кодируемые блоки (КБ). Размер кодируемого блока может быть не более 4096 пикселей, высота не менее 4 пикселей. Такое разбиение, хотя и снижает несколько коэффициент сжатия, но повышает устойчивость сжатого потока к ошибкам канала связи: ошибка испортит лишь небольшой блок. Кодирование блоков ведется в три этапа, битовыми плоскостями [3].

Алгоритм сжатия JPEG2000 ориентирован на полноцветные 24-битные изображения без резких переходов цветов. Стандарт JPEG2000 превосходит по эффективности стандарт JPEG примерно в 2 раза при сжатии с потерями и на 5–20% при сжатии без потерь. Представление изображения, полученное в результате wavelet-преобразования, очень удобно тем, что оно обеспечивает возможность получения приблизительных копий изображения без осуществления полного обратного преобразования, это имеет большое значение при размещении изображения в Интернете.

Заключение

В статье описаны основные применяемые сегодня методы сжатия изображений. Выбор того или иного метода зависит от области, где будут применяться сжатые изображения, и от характеристик этих изображений. Каждый из методов имеет множество вариаций, но в статье приведены только основные принципы работы алгоритмов.

Литература

1. Ватолин Д., Ратушняк А., Смирнов М., Юкин В. Методы сжатия данных. М.: ДИАЛОГ-МИФИ, 2003. 384 с.

2. Сэлмон Д. Сжатие данных, изображений и звука. М.: Техносфера, 2004. 368 с.
3. www.compression.ru
4. Красильников Н.Н. Цифровая обработка изображений. М.: Вузовская книга, 2001. 320 с.
5. Уэлсид С. (S. Welstead). Фракталы и вейвлеты для сжатия изображений. М: Триумф, 2003. 320 с.

ИССЛЕДОВАНИЕ СРЕД И СРЕДСТВ ДЛЯ ЗАПИСИ СИНТЕЗИРОВАННЫХ КОМПЬЮТЕРОМ ГОЛОГРАММ

Г.С. Александров, О.В. Елисеев, Д.В. Соловьев, А.А. Федоров

Научный руководитель – старший преподаватель К.О. Ткачев

Настоящая статья посвящена методу отображения синтезированных компьютером голограмм на поверхности CD-R диска. Были исследованы различные марки CD-R дисков и приводов для записи дисков и выбраны наиболее подходящие для данного метода.

Введение

Для многих приложений голографии, в том числе и для голографической фотолинтографии, весьма удобно синтезировать голограммы несуществующих в реальности объектов в виртуальном пространстве компьютера и затем «нарезать» их с помощью специальных машин. Но серьезным неудобством при реализации данного метода является высокая стоимость этих машин. При создании голограмм высокого разрешения точность позиционирования штрихов при нанесении должно быть порядка 0.1 мкм на довольно значительных площадях (порядка нескольких см²). Подобные машины стоят порядка 100 тыс. \$. Поэтому весьма актуальным является поиск альтернативных методов отображения синтезированных компьютером голограмм [1–4].

Основная часть

Год назад возникла идея использовать для создания таких голограмм дешевое устройство записи CD или DVD дисков. Благодаря наличию штампованной спирали на диске CD-R позиционирование по радиусу диска обеспечивается с высокой точностью в 1,6 мкм (шаг спирали), что нельзя сказать о координате угла [5, 6]. На рис. 1 показаны этапы формирования голографического изображения на CD-R диске.

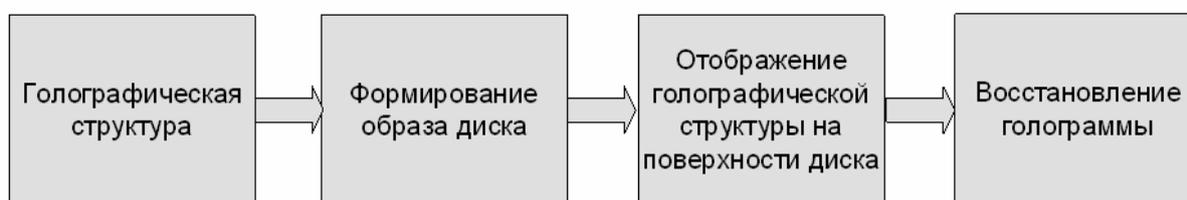


Рис. 1. Этапы формирования голографического изображения на CD-R диске

Нами был произведен обзор имеющейся литературы, и обнаружилось, что две группы специалистов (из Англии и Японии) начали разрабатывать эту идею раньше нас. Причем специалисты компании Yamaha использовали привод собственной разработки CRW F1, в котором реализована функция нанесения изображений на CD-R диск DiskT@2, по их заявлению, максимальное разрешение полученных голограммных структур составило порядка 40 мкм [7, 8]. Однако практически полное отсутствие в доступной нам литературе технических подробностей программного обеспечения для записи дисков обусловили невозможность повторения их результатов. В результате этого на начальном этапе было решено провести исследование переноса изображения голограммы на CD-R диск. Выяснилось, что серьезными проблемами на пути реализации данного метода станут (рис. 2):

- низкая контрастность изображений, полученных на дисках;
- наличие ошибок геометрических форм получаемых элементов рисунка, в том числе муара;

- наличие растростности в получаемых изображениях решеток.

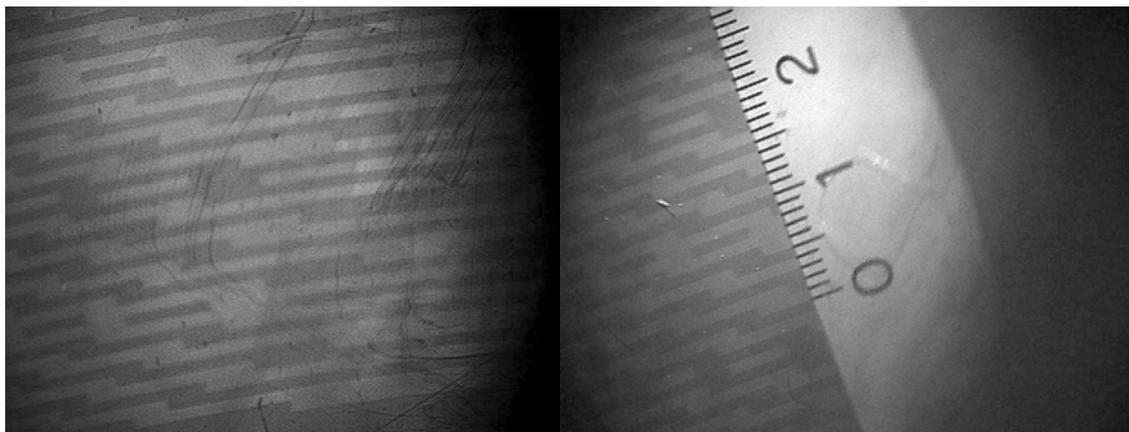


Рис. 2. Структура решеток, полученных на CD-R дисках

В результате были определены задачи дальнейших исследований, решение которых актуально при реализации метода отображения синтезированных компьютером голограмм на поверхностях CD-R дисков:

- исследование оптических и механических свойств CD-дисков как носителей голограммной структуры;
- обеспечения стабильности угловой скорости вращения оптического диска;
- поиск другого способа отображения элементов рисунка.

Обзор практически всей номенклатуры присутствующих на рынке CD-R дисков показал, что их условно можно разделить на 3 типа (рис. 3):

1. Цианин – ярко-синие диски;
2. Цианин Азо – схожи по цвету с цианиновыми;
3. Фталоцианин – светлые диски [9].

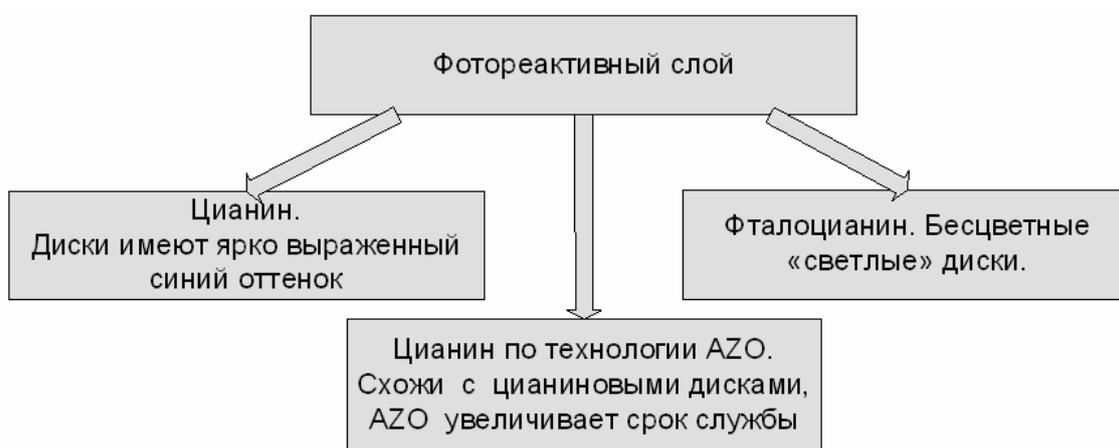


Рис. 3. Классификация CD-R дисков по типу фотореактивного слоя

Для исследований были отобраны 7 марок CD-R дисков, среди них присутствовали все 3 типа используемых в настоящий момент типов фотореактивных слоев. Исследование CD-R дисков проходило по следующей методике:

1. на диске прожигалось изображение решетки;
2. проводилась визуальная оценка контрастности полученного изображения;
3. проводилось снятие отражающего слоя;
4. полученная структура наблюдалась с помощью микроскопа.

Понятно, что чем выше контраст изображения голограммы на CD-R диске, тем выше ее эффективность и выше качество восстановленного изображения.

В результате исследований выяснилось, что наиболее контрастные изображения получились на дисках Verbatim Cyan (код 97m24s01f) и Verbatim Azo (код 97m34s23f), однако только диски Verbatim Cyan показали хорошую «податливость» при снятии верхней отражающей пленки: проводился небольшой надрез скальпелем, воздух проникал под пленку, и ее дальнейшее отделение не вызывало проблем, что нельзя сказать о других «болванках». Следует отметить, что оценка возможности снятия защитной пленки проводилась из-за того, что при работе на отражение существенно увеличивается влияние некачественной подложки диска. Так как ошибки форм поверхностей подложки голограммы существенно влияют на качество голограммных структур, проводились измерения ошибок геометрической формы рабочей поверхности CD-R дисков с помощью голографического пробного стекла [10, 11]. Было определено, что поверхность рабочей стороны диска имеет плавные неровности с амплитудой порядка 50 мкм на масштабе радиуса диска (3.5 см). Однако это не является существенной проблемой, так как она может быть решена помещением диска в иммерсионную жидкость.

Для исследования приводов для записи дисков была собрана установка (рис. 4). Решено было во время измерений записывать на диск аудиоданные сплошным потоком на скорости 4x. С микросхемы-драйвера двигателей привода снимался сигнал «FG», символизирующий частоту оборотов диска при записи [12]. С помощью подключенного осциллографа и частотомера проводилась оценка стабильности скорости при записи.



Рис. 4. Блок-схема установки для исследования пишущих приводов

Следует отметить особенность схемотехнического решения, реализованного в приводе Yamaha CRW F1 (рис. 5). Сигнал с датчиков Холла в цифровой форме поступает на компаратор микросхемы-драйвера двигателя, где он совмещается и преобразуется в частотный сигнал, символизирующий скорость оборота диска, далее он поступает в головную микросхему контроллера привода («чипсет»), где сравнивается с эталонной частотой генератора для данной скорости вращения. В случае несовпадения частот контроллер управляет микросхемой двигателей с помощью поправочного сигнала, а именно, путем ускорения или торможения мотора шпинделя, причем сигнал этот уже

имеет аналоговую форму, что указывает на наличие высококачественных ЦАП/АЦП преобразователей в микросхемах.

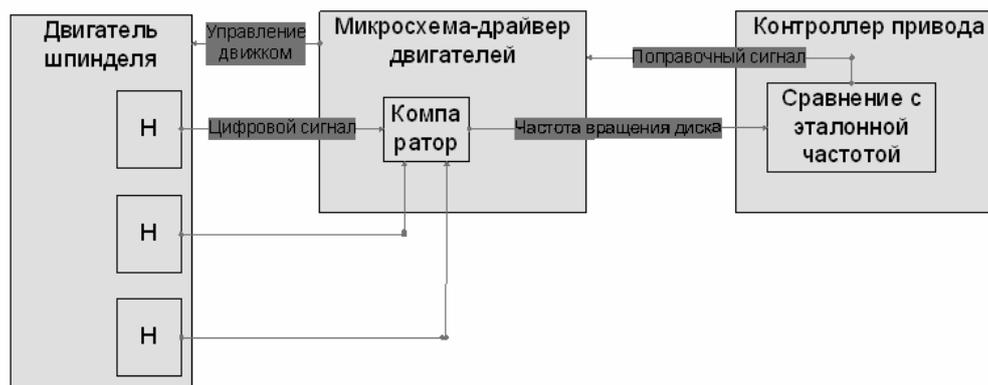


Рис. 5. Схематическое решение привода Yamaha CRW-F1

Следует отметить, что реализовать такое включение позволяют все контроллеры двигателей исследованных пишущих приводов, однако эта функция, как оказалось, задействована только в приводе F1. Понятно, что наиболее подходящем для отображения синтезированных компьютером голограмм на диске CD-R является привод, обеспечивающий наиболее стабильный режим записи диска. Это, в том числе, обуславливает и точность позиционирования точек изображения голограммы по угловой координате диска.

Все приводы, кроме Yamaha, характеризуются снижением скорости вращения диска от центра к краю, т.е. пишут диск с постоянной линейной скоростью (CLV), а привод Yamaha записывает аудиодиск с включенной опцией Audio Master – фирменной технологией, обеспечивающей хорошее качество звука записанного диска [13]. Также следует отметить, что в приводе Yamaha уровень неравномерности, нестабильности скорости вращения был намного меньше, чем в других исследованных приводах.

Заключение

Исследования показали, что наиболее подходящей средой для отображения синтезированных компьютером голограмм являются диски Verbatim Cyan (код 97m24s01f), а наиболее подходящем приводом – привод Yamaha CRW F1. Результаты настоящих исследований имеют важное значение для дальнейшей реализации метода отображения синтезированных компьютером голограмм на дисках CD-R.

Литература

1. Clube F., Gray S. et al. Holographic microlithography. // Opt. Eng. 1995. 34(9). P. 2724–2730.
2. Корешев С.Н., Ратушный В.П. Использование метода голографии для получения изображения двумерных объектов при решении задач коротковолновой фотолитографии. // Оптический журнал. 2004. № 10. С. 32–39.
3. Бедж Дж., Коронкевич В.П., Корольков В.П. и др. Синтез голограмм для контроля оптических систем. / Материалы конференции стран СНГ и Прибалтики «Голография-96». СПб, 1996. С. 49.
4. Лукин А.В. Голография в оптической технологии и приборостроении. / Тезисы докладов 6 Всесоюзной конференции по голографии, 18–20 сентября 1990. С. 12.

5. Data interchange on read-only 120mm optical data disks (CD-ROM), standard ECMA-130, 2nd edition june 1996. / <http://www.ecma-international.org/publications/files/ECMA-ST/Есma-130.pdf>
6. How CDs work. / <http://electronics.howstuffworks.com/cd.htm>
7. Sakamoto Y., Morishima M., Usui A. Computer generated holograms o a CD-R disk. // SPIE. Vol. 5290. P. 42–29.
8. Cable Mash P., Wilkinson T. Production of computer-generated holograms on recordable compact disk media using a compact disk writer. // Opt. Eng. September 2003. 42(9). P. 2514–2520.
9. Как делаются CD диски. / http://cdcopy.com.ua/ru/articles/cd_production.html
10. Кольер Р., Берхарт К., Лин Л. Оптическая голография. М.: Мир, 1973.
11. Корешев С.Н. Голографическое пробное стекло и интерферометры на его основе. // Оптика и спектроскопия. 1997. Т. 82. №4. С. 683–687.
12. 6ch power driver for CD-ROM, DVD-ROM, optical disk ICs, datasheet. / www.rohm.com
13. CRW-F1 and DiskT@2. / <http://www.nix.ru/news/articles/reviews/storage/crw-f1/i01.htm>

АВТОМАТИЗИРОВАННАЯ СИСТЕМА СБОРА И ОБРАБОТКИ ГЕОФИЗИЧЕСКОЙ ИНФОРМАЦИИ В РЕАЛЬНОМ МАСШТАБЕ ВРЕМЕНИ ДЛЯ ВЫЯВЛЕНИЯ ЭЛЕКТРОМАГНИТНЫХ КРАТКОСРОЧНЫХ ПРЕДВЕСТНИКОВ СИЛЬНЫХ ЗЕМЛЕТРЯСЕНИЙ

Д.Ю. Сарычев

Научный руководитель – доктор технических наук, профессор А.Г. Коробейников

Начиная с 1998 г., две группы трехкомпонентных магнитных станций производства СПбФ ИЗМИРАН (три станции в каждой группе) были установлены юго-западнее и юго-восточнее Токио. Расстояние между группами станций примерно 150 км, станции в каждой группе расположены на расстоянии 4–7 км друг от друга. Такая конфигурация дает возможность на каждой группе определять вектора градиентов и фазовых скоростей УНЧ геомагнитных возмущений ($F < 1$ Гц) вдоль земной поверхности. Данные со всех магнитных станций через модем передаются в центр сбора и обработки информации. Предлагается создать автоматизированную систему сбора информации, чтобы в реальном масштабе времени определять аномальные градиенты и фазовые скорости УНЧ геомагнитных возмущений с целью краткосрочного прогноза землетрясений.

Введение

Наблюдения УНЧ электромагнитных волн ($F = 10 - 0.001$ Гц) в сейсмоактивных зонах проводятся в течение последних 15–20 лет, после появления высокочувствительных магнитометров. Результаты измерений в различных сейсмически активных районах Земли показали, что как на подготовительной стадии развития очаговых зон средних и сильных (магнитуда $M > 5$) землетрясений, так и в период афтершоковой активности на больших расстояниях от будущего эпицентра могут наблюдаться аномальные УНЧ электромагнитные излучения большой интенсивности. Как правило, эти эмиссии имели шумовой характер, возникали задолго до главного сейсмического толчка и продолжались в течение всей афтершоковой активности. Впервые УНЧ электромагнитные эмиссии наблюдались перед сильными землетрясениями в Армении (Спитак, 1988) [1, 2] и США (Лома Приетта, 1989) [3, 4]. Увеличение интенсивности УНЧ эмиссий наблюдалось и за несколько месяцев перед разрушительным землетрясением в Гуаме (1993, $M = 8$) [5, 6], которое произошло внезапно, без форшоков.

Эксперимент

С 1998 г. в Японии, южнее Токио, ведется регистрация вариаций магнитного поля и теллурических токов шестью высокочувствительными цифровыми трехкомпонентными магнитными станциями MVC-3DS, разработанными в СПбФ ИЗМИРАН [7]. Три станции расположены на полуострове Изу, и три – на полуострове Босо. Каждая группа из трех станций представляет собой магнитный градиентометр. Данные на каждой магнитовариационной станции регистрируются на жесткий диск ПК с дискретностью 50 Гц, а для синхронизации одновременной работы магнитных датчиков используется система GPS на каждой станции. Данные с дискретностью 1 Гц с каждой станции через модем ежесуточно передаются в центр сбора информации, расположенный в университете г. Чибо (Япония). На рис. 1 показано расположение магнитных станций (треугольники) и эпицентров землетрясений (звездочки). Цифры возле эпицентров землетрясений означают их магнитуды.

Работы [8–11] посвящены исследованиям градиентов и фазовых скоростей УНЧ геомагнитных вариаций перед землетрясениями 2000 и 2002 гг. возле полуострова Изу и на полуострове Босо (юго-западнее и юго-восточнее Токио). В этих работах было показано, что примерно за 3–6 месяцев до сильных землетрясений ($M > 5$) начиналось аномальное увеличение величин градиентов в вертикальной и полной горизонтальной

компоненте магнитного поля и уменьшение величин фазовой скорости в этих же компонентах. Авторами высказывается предположение о том, что аномальное изменение градиентов и фазовых скоростей связано с двумя процессами в области очага будущего землетрясения – образуется аномалия повышенной проводимости, и возникают широкополосные литосферные УНЧ электромагнитные излучения. Аномалия повышенной проводимости может возникнуть как вследствие тектонических движений, так и вследствие подъема магмы к поверхности земной коры. Один из возможных механизмов возникновения электромагнитных излучений связан с активизацией процесса образования микротрещин в области очага будущего землетрясения [12]. Высокочастотные электромагнитные излучения сильно затухают в земной коре, и на поверхности мы наблюдаем в основном ультранизкочастотные излучения ($F < 1$ Гц).

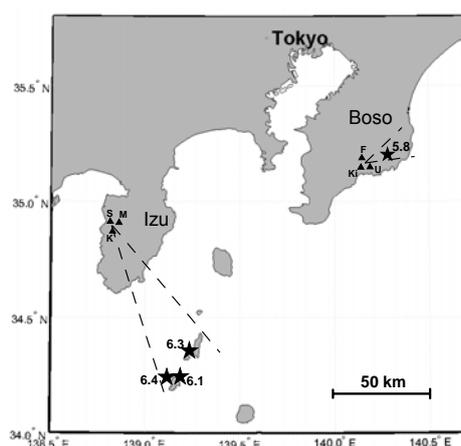


Рис.1. Расположение магнитных станций (черные треугольники) и эпицентров сильных землетрясений (звездочки) в Японии в 2000 (Изу) и 2002 (Босо) гг.

Особенности автоматизированной системы сбора и обработки геофизической информации в реальном масштабе времени

Градиенты и фазовые скорости геомагнитных вариаций можно находить двумя способами – чисто экспериментальным путем и в рамках модели плоской электромагнитной волны. В первом случае необходимо определять фазовые задержки и разности величин амплитуд вариаций между двумя любыми парами станций магнитного градиентометра, состоящего из трех разнесенных станций. Поскольку координаты магнитных станций и расстояние между ними известно, то можно определить градиенты и фазовые скорости для двух пар станций, выбранных из трех станций магнитного градиентометра, и затем построить вектор фазовой скорости и градиента пульсаций в соответствии с формулами, приведенными в [8] и [11]. Поскольку фазовые скорости геомагнитных волн вдоль земной поверхности для УНЧ геомагнитных вариаций составляют десятки км/с, а фазовые задержки, соответственно, – доли секунды, то необходимо использовать данные с дискретностью 50 Гц. В рамках модели плоской электромагнитной волны величина фазовой скорости между двумя точками на земной поверхности определяется через амплитуды соответствующих компонент вариаций магнитного поля с учетом фазовой задержки [10, 11] следующим образом:

$$V_{12} = \frac{2\pi d_{12}}{T \ln \frac{B_1(t)}{B_2(t+\tau)}} \quad (1)$$

В выражении (1) для геомагнитных вариаций с периодом T величины B_1 и B_2 определяются в момент времени t на первой станции и в момент времени $t+\tau$ на второй станции (τ – фазовая задержка при прохождении геомагнитной волны расстояния d_{12} между двумя станциями). Используя три магнитные станции, расположенные треугольником на земной поверхности, можно определить по (1) фазовые скорости V_{12} (между станциями 1 и 2) и V_{13} (между станциями 1 и 3) и затем найти направление и величину вектора фазовой скорости геомагнитных волн вдоль земной поверхности. При применении этого метода нет необходимости в высокой дискретности регистрируемых данных, поскольку в качестве величин B_1 и B_2 могут быть использованы среднеквадратичные значения амплитуд УНЧ геомагнитных пульсаций.

Имеющаяся в Японии система сбора геофизической информации может быть использована для обработки данных в реальном масштабе времени, используя данные с дискретностью 1 Гц, передаваемые в центр сбора информации ежедневно.

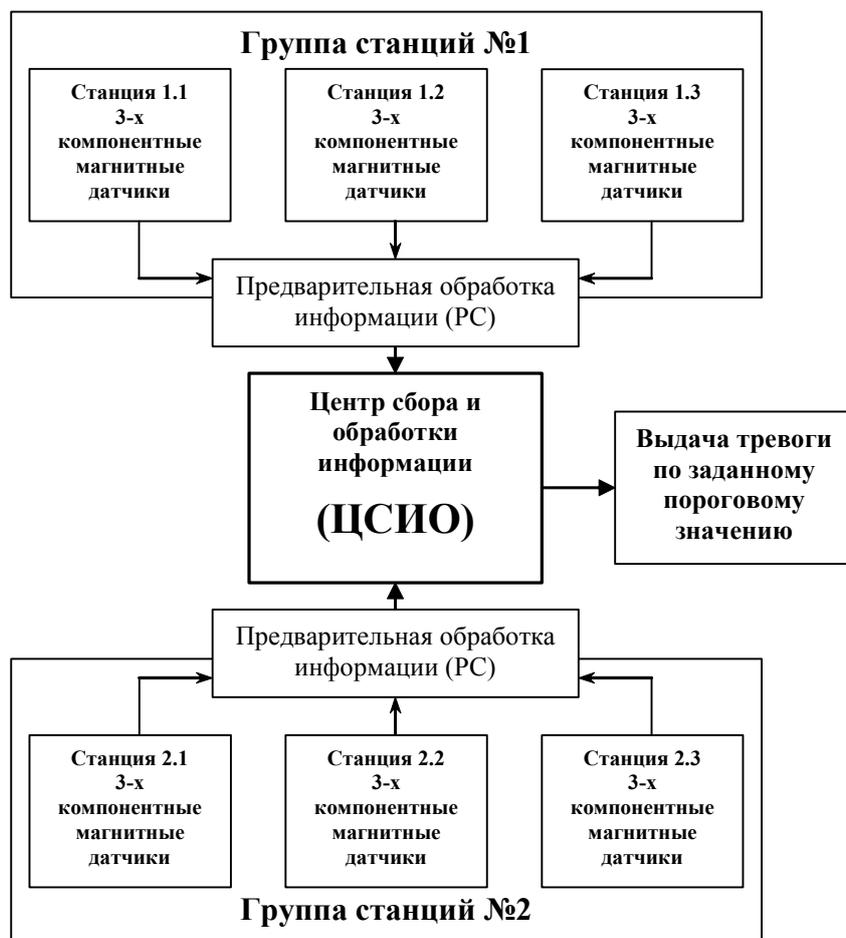


Рис. 2. Блок-схема автоматизированной системы сбора и обработки геофизической информации в реальном масштабе времени

На рис. 2 показана блок-схема предлагаемой автоматизированной системы сбора и обработки геофизической информации в реальном масштабе времени. Группы станций 1 и 2 состоят каждая из трех трехкомпонентных магнитовариационных станций, расположенных на земной поверхности в вершинах треугольника на расстоянии ~ 5 км друг от друга. Расстояние между двумя группами составляет ~ 100 км. При такой конфигурации установленных датчиков будет контролироваться район 200×200 км. Данные с каждой магнитной станции обрабатываются процессором и отправляются в центр сбора и обработки информации (дискретность данных – 1 Гц) через модем каждые три

часа. Данные с дискретностью 50 Гц накапливаются на жестких накопителях и изымаются по мере необходимости.

В центре сбора и обработки информации геофизические данные немедленно обрабатываются, и каждые три часа на экран компьютера выдается следующая информация:

- среднеквадратические значения геомагнитных вариаций в полосе частот 0.001–1 Гц, разбитой на 10 поддиапазонов;
- величины векторов градиентов и фазовых скоростей геомагнитных вариаций в той же полосе частот;
- направления векторов градиентов и фазовых скоростей геомагнитных вариаций в той же полосе частот.

В случае появления устойчивых новых направлений векторов градиентов и фазовых скоростей, а также аномального изменения величин градиентов и фазовых скоростей производится расчет местоположения будущего эпицентра землетрясения и выдается сигнал тревоги.

Литература

1. Копытенко Yu.A. Discovering of ultra-low-frequency emissions connected with Spitak earthquake and his aftershock activity on data of geomagnetic pulsations observations at Dusheti and Vardzija. / Moscow, IZMIRAN. Preprint № 3 (888), 1990.
2. Копытенко Yu.A. Detection of Ultra-Low Frequency emissions connected with the Spitak Earthquake and its aftershock activity based on geomagnetic pulsations data at Dusheti and Vardzia observatories. // *Phys. Earth and Planet. Inter.* 1993. V. 77.
3. Bernardi A. ULF magnetic field measurements near the epicenter of the M_s 7.1 Loma Prieta earthquake. // *Phys. Earth Planet. Interiors.* 1991. V. 68.
4. Molchanov O.A. Results of ULF magnetic field measurements near the epicenters of the Spitak ($M_S=6.9$) and the Loma-Prieta ($M_S=7.1$) earthquakes: Comparative analysis. // *Geophys. Res. Lett.* 1992. № 19.
5. Hayakawa M. Results of Ultra-low- frequency magnetic field measurements during the Guam earthquake of 8 August 1993// *Geophys. Res. Lett.* 1996. V. 23.
6. Kawate R. Ultra-low-frequency magnetic fields during the Guam earthquake of 8 August 1993 and their interpretation. // *Phys. Earth Planet. Interiors.* 1998. V. 105.
7. Goto T.-N. Calibration and running test of torsion magnetometer made in Russia. // *Rep. of Japan Marin Sci. and Tech. Center (JAMSTEC).* 2002. 45 (March, 2002).
8. Копытенко Yu.A. Investigation of the ULF electromagnetic phenomena related to earthquakes: contemporary achievements and the perspectives. // *Annali di Geofisika.* 2001. V. 44. № 2.
9. Копытенко Yu.A. Monitoring of the ULF electromagnetic disturbances at the station network before EQ in seismic zones of Izu and Chiba peninsulas. / In: «Seismo Electromagnetics: Litosphere-Atmosphere-Ionosphere Coupling». Eds. M. Hayakawa and O.A. Molchanov. Tokyo: TERRAPUB, 2002.
10. Ismaguilov V.S. ULF Magnetic Emissions Connected with Under Sea Bottom Earthquakes. // *Natural Hazards and Earth Sys. Sci.* 2001. V. 1.
11. Ismaguilov V.S. Variations of phase velocity and gradient values of ULF geomagnetic disturbances connected with the Izu strong earthquakes. // *Natural Hazards and Earth Sys. Sci.* 2002. V. 20.
12. Molchanov O.A. Generation of ULF electromagnetic emissions by microfracturing. // *Geoph. Res. Lett.* 1995. V. 22.

МОДЕЛИРОВАНИЕ УЛЬТРАНИЗКОЧАСТОТНЫХ ЭЛЕКТРОМАГНИТНЫХ ЭМИССИЙ, ВОЗНИКАЮЩИХ ПЕРЕД И ВО ВРЕМЯ СИЛЬНЫХ ЗЕМЛЕТРЯСЕНИЙ

Н.А. Семенов

Научный руководитель – доктор технических наук, профессор А.Г. Коробейников

Перед и в течение сейсмически активного периода на земной поверхности наблюдается ультранизкочастотное (УНЧ) электромагнитное излучение ($F=0.001-10$ Гц). Одним из источников этого излучения служит процесс образования микротрещин, возникающих в области очага будущего землетрясения при усилении тектонических движений. Моделированию этого процесса и посвящена данная работа.

Введение

Землетрясения – это явления, при которых земная кора и верхняя мантия постепенно подвергаются воздействию обширного поля сил [1]. В ослабленных участках внезапно образуются разрывы, земная кора как бы вспарывается на какой-то глубине, при этом единым импульсом высвобождаются напряжения и деформации и возбуждаются сейсмические волны. Как правило, в течение сейсмоактивного периода происходит много сейсмических толчков различной магнитуды. Собственно землетрясением обычно называется самый сильный сейсмический толчок в сейсмическом рою. В сейсмоактивных зонах ведутся многочисленные наблюдения и съемки с целью слежения за движениями коры, сейсмичностью, геомагнетизмом, грунтовыми водами, геохимическими изменениями, выделяющимися газами и т.д., однако проблема предсказания разрушительных землетрясений еще далека от своего решения. Как указывается в литературе, очень много сильных землетрясений (~50%) происходит без форшоковой активности, поэтому исследования предвестников землетрясений несейсмического характера, имеют большое значение для решения проблемы предсказания землетрясений [1].

Среди выделенных многочисленных среднесрочных и краткосрочных прогностических характеристик особое место занимают электромагнитные предвестники землетрясений, особенно в ультранизкочастотном (УНЧ) диапазоне ($F=10-0.001$ Гц) [2]. Одним из источников этого УНЧ излучения служит процесс образования микротрещин размером ~0.5–10 мм, возникающих в области очага будущего землетрясения при усилении тектонических движений в земной коре [3]. В момент раскрытия микротрещин в лабораторных экспериментах наблюдалось возникновение коротких электромагнитных импульсов длительностью ~1 мкс [4, 5]. Данная работа посвящена моделированию этого процесса.

Описание математической модели

Для решения поставленной задачи была выбрана математическая модель, основанная на том, что участок земной среды, электромагнитное излучение которого моделируется, берется в форме параллелепипеда. Задаются координаты расположенной на земной поверхности точки наблюдения (той точки, в которой мы хотим наблюдать излучение), обозначаемые на рис. 1 как (X, Y, Z) .

Сам параллелепипед разбивается на параллелепипеды меньшего размера, каждый из которых считается в дальнейшем местом расположения источника электромагнитного излучения. Задается также некоторый период времени, в течение которого происходит тектонический процесс, вызывающий образование микротрещин размером 0.5–10 мм в породах земной коры. В момент своего образования микротрещины являются источниками, генерирующими электромагнитные импульсы очень малой длительности (~1 мкс) [1]. Этот период разбивается на две части:

- первая часть (1), в которой происходит нарастание плотности излучения перед землетрясением;
- вторая часть (2), в которой происходит снижение плотности излучения после землетрясения.

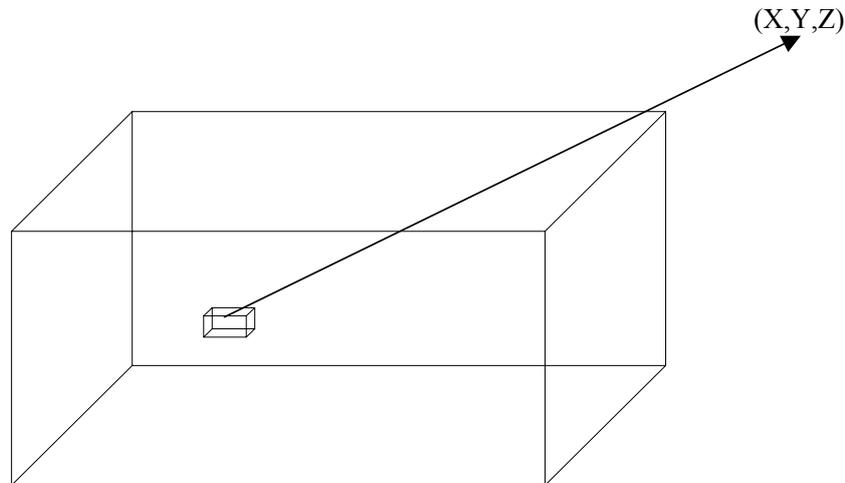


Рис. 1. Геометрическая схема модели

Плотность временных отметок (моментов возникновения микротрещин) в первой части (ρ_1) рассчитывается по формуле

$$\rho_1 = a_1 e^{b_1 t + t_1},$$

где a_1 и b_1 – некоторые положительные коэффициенты. Плотность временных отметок во второй части (ρ_2) рассчитывается по формуле

$$\rho_2 = a_2 e^{-b_2 t + t_2},$$

где a_2 и b_2 – некоторые положительные коэффициенты, t – время, а t_1 и t_2 – начальные отступы. Коэффициенты a_1 , b_1 , a_2 и b_2 выбираются таким образом, что плотности временных отметок в точке пересечения первой и второй частей совпадают. При этом спад плотности излучения происходит значительно быстрее, чем возрастание [5].

Сам процесс излучения представляется таким образом, что в каждый момент времени из тех, которые были рассчитаны заранее, одна из точек в изначальном параллелепипеде может стать источником направленного излучения (излучения, максимум амплитуды которого приходится на направление, перпендикулярное большей оси микротрещины). Значение амплитуды излучаемого сигнала в точке излучения представляется формулой

$$B(t) = a e^{-k(t-t_0)}, \quad (1)$$

где a и k – некоторые положительные коэффициенты, а t_0 – момент начала излучения. При этом коэффициенты берутся таким образом, что через время, равное некоторому заранее заданному Δt , значение амплитуды сигнала можно принять равным нулю.

Поскольку амплитуда сигнала с расстоянием меняется в зависимости от частоты и проводимости земной коры, необходимо разложить сигнал из формулы (1) на частотные составляющие. Для разложения используется быстрое преобразование Фурье [6].

Рассмотрим значение амплитуды магнитного поля в точке, отдаленной от источника излучения на некоторое расстояние r в момент времени t :

$$B(t) = c(r) B_0(w) e^{-\beta r} \cos(wt - kr) * \cos(\alpha),$$

где w – частота сигнала, k – волновой вектор, B_0 – амплитуда сигнала, $c(r)$ – некоторый положительный коэффициент, зависящий от расстояния, β – некоторый положи-

тельный коэффициент, α – угол между направлением распространения сигнала и направлением на точку наблюдения. Применяя приближение плоской волны, окончательно получим:

$$B(t) = B_0(w)e^{-kr} \cos(\omega t - kr) * \cos(\alpha). \quad (2)$$

Для плоской волны волновой вектор k рассчитывается по формуле

$$k = \frac{2\pi}{\lambda},$$

где $\lambda = \sqrt{10^7 \rho_k T}$ – длина волны, $T = \frac{2\pi}{\omega}$ – период сигнала, ρ_k – удельное электрическое сопротивление среды [2, 7].

Амплитуда сигнала $B_0(f)$ зависит от частоты и частотной плотности в соответствии со следующей формулой:

$$B_0(\omega) = \sqrt{pf},$$

где p – частотная плотность, а f – частота, причем $\omega = 2\pi f$.

Следует также заметить, что, поскольку скорость распространения волны конечна, то сигнал достигнет точки наблюдения не мгновенно, а через некоторое время

$$t_p = \frac{r}{V}, \quad (3)$$

где r – расстояние до точки наблюдения, а V – фазовая скорость. При этом

$$V = \frac{\lambda}{T} = \sqrt{\frac{10^7 \rho_k}{T}}. \quad (4)$$

Подставляя значение V из формулы (4) в формулу (3), окончательно получаем:

$$t_p = \frac{r\sqrt{T}}{\sqrt{10^7 \rho_k}}.$$

Каждый источник излучения (рис. 1) имеет свои собственные координаты (x_i, y_i, z_i) . Соответственно, координаты точки наблюдения относительно источника с этими координатами будут $(X - x_i, Y - y_i, Z - z_i)$, а расстояние до источника будет вычисляться по формуле

$$r_i = \sqrt{(X - x_i)^2 + (Y - y_i)^2 + (Z - z_i)^2}. \quad (5)$$

Поскольку каждый из источников имеет свое собственное направление распространения, необходимо отдельно рассчитывать горизонтальные и вертикальную составляющую магнитного поля. Вертикальная составляющая B_z характеризует составляющую магнитного поля, перпендикулярную земной поверхности и направленную вниз, а B_G – составляющую магнитного поля, лежащую в плоскости земной поверхности. Горизонтальная составляющая B_G , в свою очередь, раскладывается на компоненты по широте и по долготе B_H и B_D , как показано на рис. 2. Здесь B_V – проекция вектора магнитного поля на плоскость, перпендикулярную земной поверхности, O – проекция направления распространения волны на земную поверхность, H – ось координат, соответствующая широте, α_1 – угол между B и B_V , α_2 – угол между B_V и B_z , β_1 – угол между B_G и O , β_2 – угол между O и H .

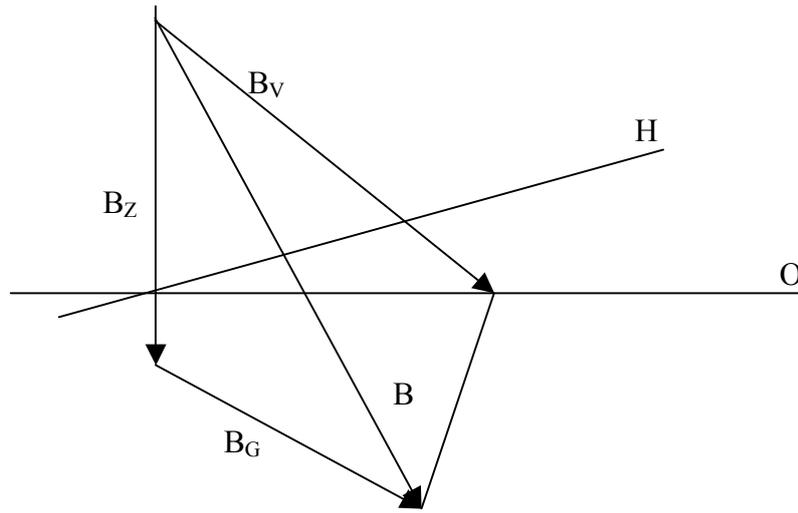


Рис. 2. Схема проекции вектора магнитного поля на оси координат

Как следует из рис. 2:

$$B_z = B \cos(\alpha_1) \cos(\alpha_2), \quad (6)$$

$$B_G = B \sqrt{1 - \cos^2(\alpha_1) \cos^2(\alpha_2)}, \quad (7)$$

где B – величина магнитного поля в точке наблюдения. Соответственно,

$$B_H = B_G \cos(\beta_1 + \beta_2), \quad (8)$$

$$B_D = B_G \sin(\beta_1 + \beta_2), \quad (9)$$

$$\cos(\beta_1 + \beta_2) = \cos(\beta_1) \cos(\beta_2) - \sin(\beta_1) \sin(\beta_2) =$$

$$= \sqrt{1 - \frac{\sin^2(\alpha_1)}{1 - \cos^2(\alpha_1) \cos^2(\alpha_2)}} \cos(\beta_2) - \frac{\sin(\alpha_1)}{1 - \cos(\alpha_1) \cos(\alpha_2)} \sin(\beta_2);$$

$$\sin(\beta_1 + \beta_2) = \sin(\beta_1) \cos(\beta_2) + \cos(\beta_1) \sin(\beta_2) =$$

$$= \frac{\sin(\alpha_1)}{1 - \cos(\alpha_1) \cos(\alpha_2)} \cos(\beta_2) + \sqrt{1 - \frac{\sin^2(\alpha_1)}{1 - \cos^2(\alpha_1) \cos^2(\alpha_2)}} \sin(\beta_2).$$

Для получения графика компонент магнитного поля в точке наблюдения весь период времени наблюдения разбивается на равные части с шагом, величина которого такова, что значение компонент магнитного поля для каждого из отрезков можно считать постоянным. Тогда

$$B_z^j = \sum_{i \in t_j} B_z^i, \quad B_D^j = \sum_{i \in t_j} B_D^i, \quad B_H^j = \sum_{i \in t_j} B_H^i,$$

где t_j – отрезок времени с индексом j . При этом компоненты магнитного поля берутся из формул (6)–(9), а величина вектора магнитного поля – из формулы

$$B_i(t_j) = B_0(w) e^{-k_i t_j} \cos(w(t - t_p^j) - k_i r_i) * \cos(\alpha_i). \quad (10)$$

Реализация математической модели

Реализация математической модели строилась исходя из минимальных системных требований и требований высокой производительности. Соответственно, исключались все компоненты, использование которых могло привести к снижению скорости расчетов. Также проводился предварительный расчет некоторых величин, постоянных для всех частот. Это несколько усложняло структуру основного цикла, но давало ощутимые преимущества в скорости.

Структуру основного расчета можно представить следующим образом.

1. Разбиение исходного экспоненциального излучения на некоторое заранее заданное число частот. Поскольку для получения спектра использовалось быстрое преобразование Фурье, то число это должно было равняться степени двойки.
2. Расчет некоторого числа точек во времени, в которое происходило излучение. Расчет этот делался по схеме, описанной в предыдущей части. При этом учитывалось, что излучение происходит в случайные моменты времени, и, соответственно, были внесены случайные сдвиги от экспоненциальной схемы.
3. Период времени, за который велся расчет, разбивался на отрезки фиксированной длины. Число этих отрезков было достаточно велико (обычно бралось 10^3 – 10^5), но при этом на несколько порядков меньше, чем число источников излучения.
4. Основной цикл стартовал по заранее рассчитанным точкам времени, при этом имелся внутренний цикл по частотам. Происходило постепенное заполнение выходного потока. Следует отметить, что такая схема полностью аналогична схеме, непосредственно вытекающей из формул

$$B_Z^j = \sum_{i \in t_j} B_Z^i, \quad B_D^j = \sum_{i \in t_j} B_D^i, \quad B_H^j = \sum_{i \in t_j} B_H^i,$$

где t_j – отрезок времени с индексом j . Но такая схема требует индексирования в массиве по данным (косвенного индексирования). Выбранная же схема позволяет использовать прямое индексирование, хотя она и менее очевидна.

5. При вычислении компонент магнитного поля следовало учитывать, что для каждого источника излучения имеются свои параметры:

- $r_i = \sqrt{(X - x_i)^2 + (Y - y_i)^2 + (Z - z_i)^2}$ – расстояние до точки наблюдения;
- t_p – время, за которое сигнал достигает точки наблюдения;
- α_i – угол между направлением на точку наблюдения и основным направлением распространения сигнала.

Перед расчетом задаются удельное электрическое сопротивление среды, начальная амплитуда магнитного поля в микрофарадах, а также координаты точки, для которой производится расчет.

Результаты

На верхней панели рис. 3 показан результат проведенного моделирования (представлена вертикальная компонента магнитного поля). Момент землетрясения совпадает с максимумом амплитуды вариаций магнитного поля. Точка наблюдения располагалась на расстоянии 50 км от эпицентра

Очаг землетрясения (параллелепипед из раздела 2) задавался на глубине 15 км. На нижней панели рис. 3 представлен результат узкополосной фильтрации полученной реализации модельных данных ($F=0.004$ – 0.005 Гц). Из рисунка видно, что, несмотря на то, что модельное поле возникло как результат действия большого количества коротких электромагнитных импульсов длительностью 1 мкс, полученная реализация модельных данных имеет весьма широкий спектр частот.

На рис. 4 построены зависимости фазовой скорости, градиента и амплитуды полученной реализации модельных данных от периода вариаций. Координатные оси представлены в логарифмическом масштабе. Подобные зависимости этих величин соответствуют экспериментальным данным [8].

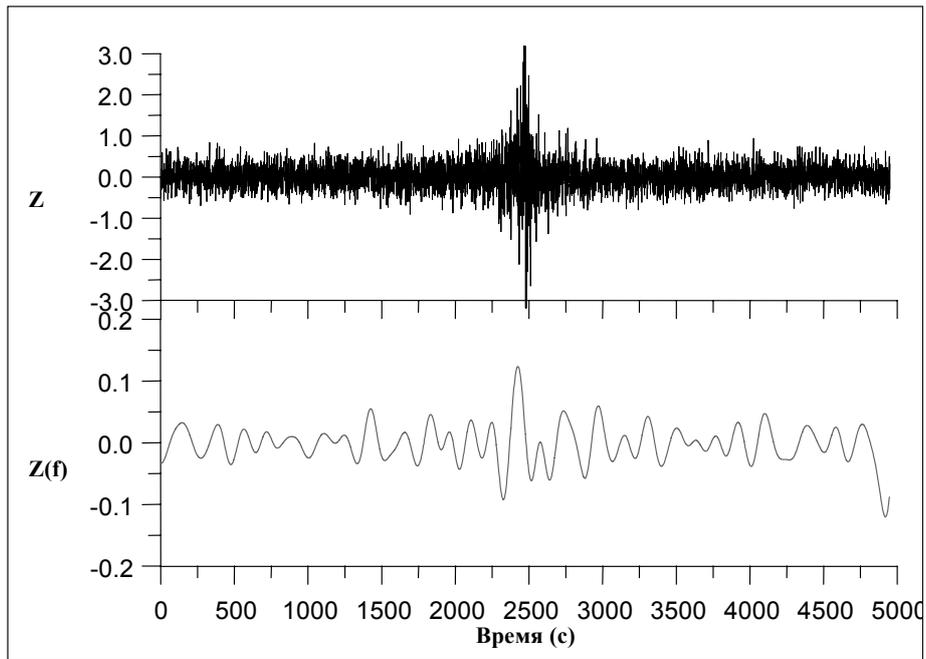


Рис. 3. Реализация нефiltroванных (сверху) и фильтрованных (снизу) модельных данных землетрясения

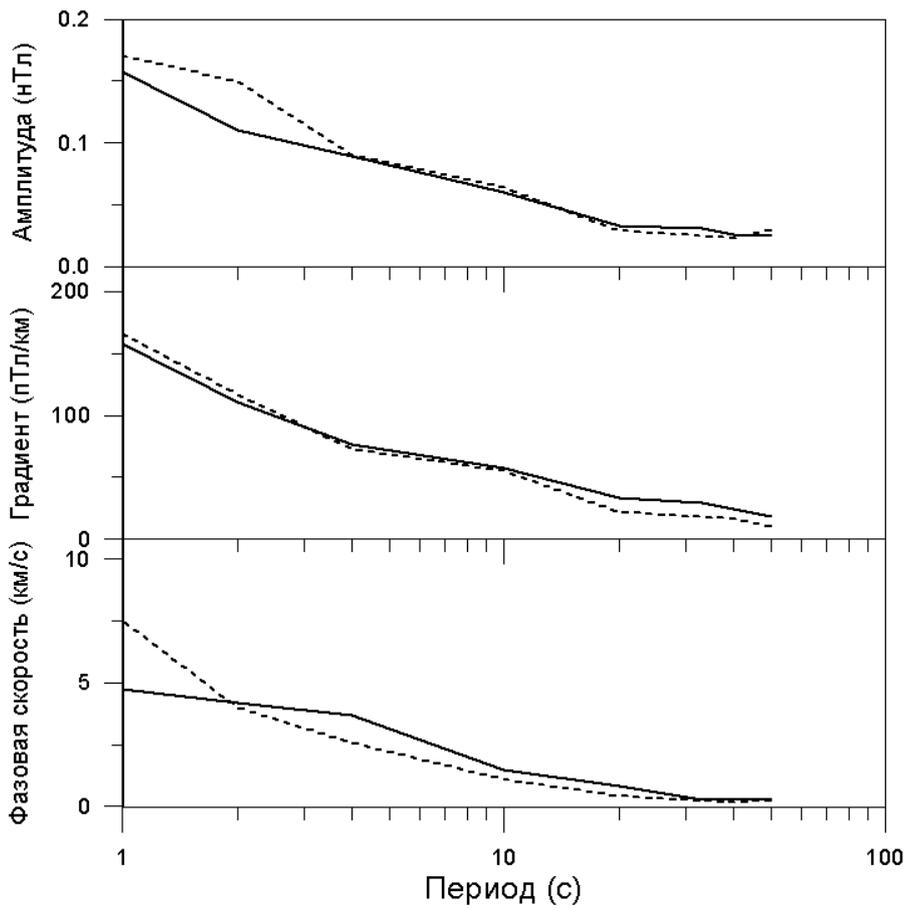


Рис. 4. Зависимости фазовой скорости, градиента и амплитуды реализации модельных данных от периода вариаций

Заключение

В работе показано, что большое количество излучателей, генерирующих короткие электромагнитные импульсы длительностью ~ 1 мкс, могут создать на земной поверхности магнитное поле, имеющее весьма широкий спектр; в том числе возникают ультранизкочастотные магнитные вариации, наблюдаемые в большом числе экспериментов. По-видимому, теория образования микротрещин, возникающих в земной коре перед и в течение сейсмоактивного периода, вполне адекватно описывает процесс генерации УНЧ электромагнитных возмущений в земной коре.

Литература

1. Mogi K. Earthquake predictions. Academic Press Japan, 1985, 166 p.
2. Ismaguilov V.S., Kopytenko Yu.A., Hattori K., Voronov P.M., Molchanov O.A., Hayakawa M. ULF Magnetic Emissions Connected with Under Sea Bottom Earthquakes // Natural Hazards and Earth Sys. Sci. 2001. V.1. P. 1–9.
3. Scholz C.H., Molnar P., Johnson T. Detailed studies of frictional sliding of granite and implications for earthquake mechanism. // J. Geophys. Res. 1972. V.77. P. 6392–6406, Cress G.O., Brady B.T., Rowell G.A. Sources of electromagnetic radiation from fracture of rock samples in laboratory. // Geophys. Res. Lett. 1987. V.14. P. 331–334,
4. Molchanov O.A., Hayakawa M. Generation of ULF electromagnetic emissions by microfracturing. // Geoph. Res. Lett. 1995. V. 22. P. 3091–3094.
5. Ковтун А.А. Использование естественного электромагнитного поля при изучении электропроводности Земли. Л.: Изд. ЛГУ, 1980. 195 с.
6. Семенов А.А. Теория электромагнитных волн. М.: Изд. МГУ, 1968. 316 с.
7. Ismaguilov V.S., Kopytenko Yu.A., Hattori K., Hayakawa M. Variations of phase velocity and gradient values of ULF geomagnetic disturbances connected with the Izu strong earthquakes. // Natural Hazards and Earth Sys. Sci. 2002. V.20. P. 1–5.

РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ УПРАВЛЕНИЯ СИСТЕМАМИ АВТОМАТИКИ ХОЛОДИЛЬНЫХ УСТАНОВОК

А.В. Пазухин

Научный руководитель – доктор технических наук, профессор А.Г. Коробейников

Статья затрагивает вопросы создания отечественного программного обеспечения для контроля и управления холодильными системами на базе оборудования западных производителей.

Введение

Современные тенденции развития требуют все шире и шире применять электронные системы для управления холодильными установками. Без этого сегодня уже невозможно создавать качественные и конкурентоспособные системы. Более того, потребность в таких системах будет постоянно расти, ведь даже самое современное холодильное оборудование, несмотря на высокий уровень надежности, не может работать безотказно неограниченно долгое время, и рано или поздно в его работе случается сбой [1]. Специально для этих целей создавались так называемые системы мониторинга и управления.

На сегодняшний день в мире существует несколько крупных западных компаний, производящих автоматику для холодильной техники, систем вентиляции и кондиционирования воздуха, теплоснабжения. Среди них на рынке холодильной автоматики наиболее ярко представлены такие компании, как «Danfoss» (Дания), «Carel» (Испания), «Alco Controls» (США) и другие. Каждая из этих компаний, помимо разработки и выпуска приборов холодильной автоматики, занимается также разработкой программного обеспечения для систем мониторинга и управления холодильными установками.

На российском рынке холодильной автоматики на данный момент не существует программного продукта, который по своим характеристикам был бы конкурентоспособным аналогом продуктов, производимых указанными выше компаниями. Такой продукт должен отличаться простотой и удобством в эксплуатации и монтаже, быть менее дорогим по сравнению с программным обеспечением западных производителей.

Основные понятия

Под мониторингом подразумевается автоматическая регистрация и визуальное отображение измеряемых температур и системных значений, характерных для холодильных систем. Это понятие также охватывает автоматическую регистрацию системных событий и случаев возникновения сигналов об аварии [2].

Системы мониторинга и управления предоставляют, помимо вышеперечисленных функций, возможность удаленного управления холодильным оборудованием. Специально для этих целей разработаны системы удаленного мониторинга и управления холодильным оборудованием. Система имеет различные варианты исполнения – только мониторинг или мониторинг и управление. Управление всем спектром холодильного оборудования на одном или нескольких объектах при этом возможно с одного персонального компьютера, расположенного в произвольном месте и подключенного к системе по коммуникационным каналам. Диспетчер видит на экране компьютера полную информацию о работе холодильного оборудования на объектах и может:

- изменять его настройки в зависимости от ситуации;
- вести непрерывную запись рабочих параметров холодильных установок (рис. 1) (давление всасывания, давление нагнетания, температура в камерах, перегревы, степень открытия ТРВ и т.д.) и хранить их в памяти длительное время (до 1 года) с возможностью последующего просмотра;

- сокращение стоимости производства программного обеспечения при условии использования российских финансовых и трудовых ресурсов;
- сокращение стоимости программного обеспечения для конечного потребителя вследствие его разработки, установки и обслуживания отечественными специалистами;
- поддержание отечественного производителя программных средств;
- создание программного обеспечения с учетом использования дорогостоящего импортного оборудования в сложных климатических условиях России, особенностей монтажа и эксплуатации оборудования;
- простота использования программного обеспечения российского производства для эксплуатирующих организаций;
- возможность всесторонней помощи эксплуатирующим оборудованием службам со стороны разработчиков программного обеспечения.

Возможности создания

В настоящее время на территории Российской Федерации практически не производится приборов автоматики для холодильной техники. Естественно, что разработка такого программного обеспечения чрезвычайно сложна, а порой и невозможна без создания производственной базы для изготовления приборов автоматики. Тем не менее, в связи с ростом в настоящее время промышленного производства представляется возможным в ближайшем будущем появление высокотехнологичных производств в области исследований, разработки и выпуска приборов автоматики для холодильной техники, систем кондиционирования воздуха и теплоснабжения и систем мониторинга и управления к ним. Перед такими предприятиями для увеличения конкурентоспособности своей продукции на рынке встанет вопрос о необходимости вложения определенного количества финансовых средств в создание целого комплекса программных средств и, тем самым, обеспечения полного комплекса услуг по автоматизации холодильного оборудования.

Представляется возможным создание таких производств на базе бывших и ныне существующих приборостроительных предприятий при условии привлечения достаточного финансирования для модернизации предприятий и привлечения необходимых трудовых ресурсов. На начальном этапе предлагается выпуск приборов автоматики для холодильного оборудования западных производителей как наиболее надежного и простого в обслуживании и эксплуатации. В качестве основы для разработки программного обеспечения наиболее вероятным видится базирование на существующих программных продуктах западных производителей и создание аналогий им.

Заключение

На сегодняшний день не существует широко распространенного отечественного программного обеспечения для приборов холодильной автоматики, отвечающего потребностям современного рынка холодильной техники. Тем не менее, в настоящий момент разработка и выпуск такой продукции видится одним из наиболее перспективных направлений в области высокотехнологичных производств, что объясняется очевидными преимуществами такого продукта.

Литература

1. Ужанский В.С., Каплан Л.Г., Вольская Л.С. Холодильная автоматика. М.: Пищевая промышленность, 1971.
2. Курылев Е.С., Оносовский В.В., Румянцев Ю.Д. Холодильные установки. СПб: Политехника, 2002.

АНАЛИЗ И ОПТИМИЗАЦИЯ БИЗНЕС-ПРОЦЕССОВ НА ПРЕДПРИЯТИИ С ПОМОЩЬЮ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

А.А. Дегтярев, Е.В. Овечкина, А.В. Смелова

Научный руководитель – доктор технических наук, профессор В.Л. Ткалич

Описание и оптимизация бизнесов-процессов на предприятии – очень важный процесс. От правильности и своевременности проведения оптимизации во многом зависит прибыль, получаемая предприятием.

Введение

Бизнес-процесс – это совокупность производственных операций, при выполнении которой получают определенный конечный или промежуточный результат [1]. Все действия и операции на предприятии можно описать в виде последовательности операций и привести к виду бизнес-процесса. Всю схему работы предприятия можно разделить на несколько основных и вспомогательных процессов, связанных между собой.

Анализ такой схемы может помочь предприятию улучшить свою работу: снизить затраты, сократить производственный цикл, уменьшить количество управленческих ошибок или принять неотложные меры по выходу из кризиса [2].

Описание и оптимизация бизнес-процессов

Чтобы получить общую схему бизнес-процессов, надо рассмотреть все операции, проводящиеся на данном предприятии. Для анализа и оптимизации бизнес-процессов за короткие сроки чаще всего используется экспресс-метод. Задача анализа схемы работы предприятия изначально сводится к выделению ключевых бизнес-процессов, их описанию и определению взаимосвязи между ними.

После разработки схемы проводится ABC-анализ, в основе которого лежит правило Парето [3]. Этот анализ позволяет выделить небольшое число наиболее значимых бизнес-процессов, которым и нужно уделить основное внимание. Для этого все бизнес-процессы делятся на три группы – А, В и С.

К группе А относятся процессы, требующие наибольших затрат. Как правило, это 5–10% от всех процессов, происходящих на предприятии, на которые приходится 75–80% всех затрат. В группу В входят примерно 20% бизнес-процессов, к которым относятся примерно 10–20% затрат, и, наконец, группу С составляют остальные 60–75% процессов, «съедающие» всего 5–10% затрат предприятия [3].

После проведения анализа можно приступать к оптимизации бизнес-процессов, т.е. непосредственно к этапу улучшений. Прежде всего, необходимо оптимизировать процессы группы А.

Одним из основных процессов, оптимизация которого проводилась при написании статьи, является система планирования закупок. Рационализация именно этого процесса позволит уменьшить количество сырья и материалов на складе, снизить затраты на их транспортировку и хранение.

В системе планирования закупок все происходящие операции также были разделены на процессы группы А, В и С. К группе А отнесли закупку товаров, составляющих 70% единиц от общего объема продаж, к группе В – 20%, к группе С – 10%. Товары группы А составили 5% находящихся на складе, группы В – 15% и группы С – 80%.

После анализа было выявлено, что 80% товара, находящего на складе, залеживается. Соответственно, на этот товар нет спроса или завышена продажная стоимость.

Для оптимизации работы системы закупок было решено усилить контроль за наличием товара группы А путем установления цифр минимального объема и страхового

запаса товара на складе. Минимальный объем наличия товара был рассчитан исходя из количества проданных единиц товара в месяц и времени доставки товара на склад. Цифра страхового запаса на складе была выведена статистическим путем, с учетом задержек доставки и сезонного спроса на товар. Минимальный объем является точкой заказа товара. Страховой объем составляет 30% минимального объема, и в идеале он не должен использоваться регулярно – только в экстренных случаях. Страховой объем не должен составлять слишком большой процент от минимального объема и всегда должен корректироваться. Страховой объем – это неиспользуемые денежные средства, а потому эта цифра должна быть выбрана очень аккуратно. Все эти принципы можно пояснить графической схемой (рис. 1).

Те же принципы были применены к группе В.

В группе С были выделены товары, лежащие на складе уже более 8 месяцев, их цена была скорректирована и стала равна себестоимости (распродажа). У остальных товаров группы была проанализирована цена и частично скорректирована.

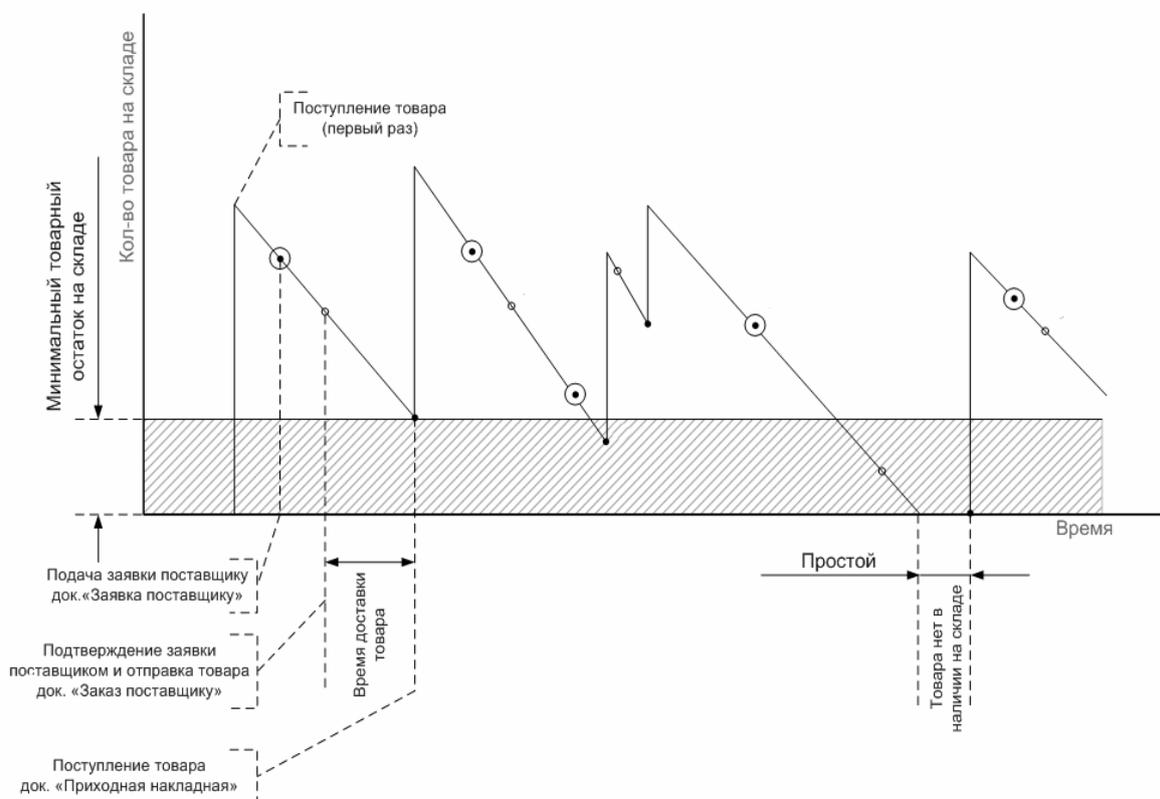


Рис. 1. График расхода и пополнения товарного запаса

Оптимизация описанных бизнес-процессов проводилась на предприятии, занимающемся оптовой и розничной продажей автозапчастей, а также ремонтом микроавтобусов и легковых автомобилей. Ассортимент запчастей на предприятии составляет более 30 000 позиций. Количество товара различных позиций, продаваемого в месяц, составляет более 100 000 штук. Поэтому так важно правильно оптимизировать все бизнес-процессы, происходящие на данном предприятии, особенно процессы, касающиеся цикла закупка – доставка – хранение – реализация товара.

Оптимизация была реализована на платформе 1С: Предприятие 7.7, конфигурация «Комплексная». Бизнес-процессы были описаны с использованием программного продукта Business Studio 1.4.

Заключение

Таким образом, удалось оптимизировать некоторые процессы, происходящие на предприятии, что привело к увеличению прибыли, снижению затоваривания склада, позволило анализировать возможную прибыль.

Литература

1. Система поддержки регулярного управления Business Studio 1.4. Руководство пользователя. Самара: Группа компаний «Современные технологии управления», 2004–2005.
2. Робсон М., Уллах Ф. Практическое руководство по реинжинирингу бизнес-процессов. / <http://business.rin.ru>
3. О бизнес-процессах и их реинжиниринге... / <http://www.md-marketing.ru>

АНАЛИЗ РИСКОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ В БАНКЕ

М.В. Лекомцева, Н.А. Нестерова, В.А. Семенов

Научный руководитель – доктор технических наук, профессор А.Г. Коробейников

В статье рассматривается метод анализа рисков в банке, предлагается способ их качественной оценки.

Введение

Риск – возможность возникновения у банка финансовых потерь (убытков), незапланированных расходов или снижения планируемых доходов, возникающий в результате недостатков в организации системы безопасности банка, используемых технологиях, функционировании информационных систем, неадекватных действий или ошибок сотрудников, а также в результате внешних событий.

Анализ рисков состоит в том, чтобы выявить существующие риски и предложить способ их оценки.

В самом широком смысле мера риска может рассматриваться в дальнейшем как описание видов неблагоприятных действий, воздействию которых может подвергнуться система или организация, и вероятностей того, что эти действия могут произойти. Результат этого процесса должен указать банку степень риска, связанного с определенными ценностями. Этот результат важен, потому что он является основой для дальнейшего выбора средств защиты и решений по минимизации риска.

Потенциальный нарушитель

Любое лицо, имеющее логический или физический доступ к информационным активам и компонентам соответствующих информационных технологий (программному обеспечению и данным, средствам вычислительной техники, коммуникационному оборудованию и каналам связи), может являться потенциальным злоумышленником. При этом предполагается возможность сговора сотрудника банка с внешним злоумышленником, но не сговор двух и более сотрудников банка.

Целью злоумышленника является получение контроля над информационным активом, приводящего к нарушению его доступности, целостности или конфиденциальности. Для достижения целей злоумышленник может использовать все экономически соизмеримые с потенциальным ущербом способы проведения атак на всех уровнях архитектуры информационных систем.

Идентификация рисков

Источниками угроз информационным активам банка являются: внешние и внутренние злоумышленники, ошибочные действия персонала банка, вирусные атаки, отказы и сбои оборудования и программного обеспечения, техногенные и природные катастрофы, террористические угрозы.

В зависимости от сферы возникновения источников угроз в банке существенными следующие являются следующие типы рисков.

1. Внутренние:

- риск бизнес-процессов (технологический риск) – риск неадекватности/неэффективности внутренних процессов и процедур, включая внедренче-

ский риск, бухгалтерский риск, риск неэффективности системы контроля за совершением операций и т.п.;

- риск автоматизации процессов – риск, связанный с использованием информационных технологий, неудовлетворительным состоянием автоматизированных систем банка (риск ошибок и сбоев в программном обеспечении, электронных технологиях и системах коммуникации и т.п.);
- риск персонала – риск ошибок, превышения установленных полномочий, операционных лимитов и ограничений, некорректного исполнения операций (нарушения действующих регламентов и процедур), недобросовестного исполнения сотрудниками своих служебных обязанностей, злоупотреблений и противоправных действий (хищений, несанкционированного доступа, утечки информации и т.п.) со стороны сотрудников, риск недостаточной квалификации персонала.

2. Внешние:

- риск противоправных действий, включая преступные посягательства (риск мошенничества¹) – риск противоправных действий извне (хищения (кражи, ограбления, разбойные нападения, акты вандализма и т.п.), несанкционированное проникновение в электронные системы Банка), риск отрицательной репутации контрагентов;
- риск утраты или повреждения имущества Банка вследствие возникновения форс-мажорных обстоятельств.

3. Комбинированный:

- правовой риск – риск прямых или косвенных потерь вследствие правовых ошибок при осуществлении банковской деятельности из-за ошибочной интерпретации или нарушения законодательства и иных нормативных правовых актов; несоответствия им внутренних нормативных документов банка; нечеткого формулирования взаимных прав и обязанностей сторон в договорных отношениях; неверно составленной документации, в том числе при решении спорных вопросов в судебных органах; несовершенства законодательства; нарушения нормативных правовых актов и условий заключенных договоров контрагентами.

Расчет информационных рисков

Имеется большое количество способов измерения и представления риска. Качественные подходы часто связаны с измерением риска в качественных терминах, заданных с помощью шкалы или ранжирования. Одномерные подходы рассматривают только ограниченные компоненты. Многомерные подходы рассматривают дополнительные компоненты в измерении риска, такие как надежность, безопасность, производительность. Одним из наиболее важных аспектов меры риска является то, что представление должно быть понятным и логичным для тех, кто должен выбирать средства защиты и решать вопросы минимизации риска. Формула, используемая при расчете рисков, представляет собой произведение трех параметров.

1. Стоимость ресурса (*Asset Value, AV*). Указанная величина характеризует ценность ресурса. При качественной оценке рисков стоимость ресурса чаще всего ранжируется в диапазоне от 1 до 3, где 1 – минимальная стоимость ресурса, 2 – средняя стоимость ресурса и 3 – максимальная стоимость ресурса. К примеру, сервер автоматизированной банковской системы имеет $AV = 3$, тогда как отдельный информационный киоск, предназначенный для обслуживания клиента, имеет $AV = 1$ по отношению к информационной банковской системе.
2. Мера уязвимости ресурса к угрозе (*Exposure Factor, EF*). Этот параметр показывает, в какой степени тот или иной ресурс уязвим по отношению к рассматриваемой угрозе.

¹ в терминологии Нового соглашения по оценке достаточности капитала Базельского комитета по банковскому надзору

Например, с точки зрения банка ресурс автоматизированной банковской системы имеет наибольшую доступность. Таким образом, атаки с целью реализации отказа в обслуживании (Denial of Service, DoS) представляют для него максимальную угрозу. При качественной оценке рисков данная величина также ранжируется в диапазоне от 1 до 3, где 1 – минимальная мера уязвимости (слабое воздействие), 2 – средняя (ресурс подлежит восстановлению), 3 – максимальная (ресурс требует полной замены после реализации угрозы).

3. Оценка вероятности реализации угрозы (Annual Rate of Occurrence, *ARO*) демонстрирует, насколько вероятна реализация определенной угрозы за определенный период времени (как правило, в течение года), и также ранжируется по шкале от 1 до 3 (низкая, средняя, высокая).

На основании полученных данных выводится оценка ожидаемых потерь (уровень риска):

1. оценка ожидаемого возможного ущерба от единичной реализации определенной угрозы (Single Loss Exposure, *SLE*) рассчитывается по формуле: $SLE = AV \times EF$;
2. итоговые ожидаемые потери от конкретной угрозы за определенный период времени (Annual Loss Exposure, *ALE*) характеризуют величину риска и рассчитывается по формуле: $ALE = SLE \times ARO$.

Четкой методики количественного расчета величин рисков предложить на данном этапе не представляется возможным. Это связано, в первую очередь, с отсутствием достаточного объема статистических данных о вероятности реализации той или иной угрозы в банке, так как данная информация конфиденциальна и для каждого конкретного банка сугубо индивидуальна [3].

Заключение

Оценка и прогнозирование уровня консолидированного (совокупного) риска банка должна осуществляться на основе использования статистической информации, объединенной в базу данных о фактах реализованных рисков. Банком должна проводиться работа по сбору и систематизации соответствующей информации, формированию базы данных для последующего анализа, оценки и прогнозирования с использованием современных математических методов.

На первоначальном этапе одновременно с формированием базы данных должна производиться комплексная оценка, ранжирование по уровню потерь и определение степени значимости идентифицированных рисков банка.

Управление риском нацелено на максимально возможное его предотвращение и вследствие этого основано на применении качественных и количественных методов анализа. Необходимо всегда помнить, что состояние безопасности информации определяется отсутствием недопустимых рисков.

Литература

1. Стандарт Банка России «Обеспечение информационной безопасности организаций банковской системы Российской Федерации. Общие положения» (утвержден распоряжением Банка России от 18 ноября 2004 № Р-609).
2. ISO/IEC IS 17799-2000. Information Technology. Code of practice for information security management.
3. <http://www.ot.ru/press20041106.html>

ОПИСАНИЕ КЛАССИФИКАЦИИ УГРОЗ DSECCT

Н.Ю. Куканова

Научный руководитель – кандидат технических наук И.Д. Медведовский
(ООО «Диджитал Секьюрити»)

В статье представлено описание классификации угроз информационной безопасности DSECCT.

Введение

При разработке классификации угроз информационной безопасности были рассмотрены и проанализированы различные существующие классификации угроз информационной безопасности (метод OCTAVE [1], классификация угроз и уязвимостей Microsoft [2], классификация угроз, описанная в Стандарте Банка России [3]). Попытки использования данных классификаций для описания по возможности большего количества угроз показали, что во многих случаях реальные угрозы либо не подходили ни под один из классификационных признаков, либо, наоборот, удовлетворяли нескольким.

Перечислим основные недостатки рассмотренных классификаций.

1. Недостаточная детализация. Классификационные признаки являются слишком общими, что затрудняет описание конкретных угроз.
2. Слишком глубокая детализация. Классификационные признаки, по сути, являются конечными угрозами. Избыточная детализация увеличивает вероятность неучета всех возможных классификационных признаков.
3. Дублирование классификационных признаков. Классификационные признаки сформулированы таким образом, что возможно подключение одной конечной угрозы к нескольким признакам.
4. Отсутствие иерархической структуры классификации. Классификация иерархически не структурирована. Классификационные признаки не являются логическим продолжением друг друга. Разные классификационные признаки находятся на одном структурном уровне.

Таким образом, основная цель создания классификации угроз – наиболее полная, детальная классификация, которая описывает все существующие угрозы информационной безопасности, по которой каждая из угроз попадает только под один классификационный признак, и которая, таким образом, наиболее применима для анализа рисков реальных информационных систем.

Основная часть

Классификация угроз основывается на ряде классификационных признаков:

- характер угрозы;
- вид воздействия угрозы;
- источник угрозы;
- объект угрозы.

По характеру угрозы информационной безопасности можно разделить на *технологические* и *организационные*. Соответственно, получим верхний уровень классификации:

- угрозы технологического характера;
- угрозы организационного характера.

Рассмотрим *технологические* угрозы информационной безопасности, которые по виду воздействия делятся на:

- 1.1. физические;

1.2. программные (логические).

Следующая ступень классификации – источник угрозы. Источниками *физических* угроз могут быть:

- 1.1.1. действия нарушителя (человека);
- 1.1.2. форс-мажорные обстоятельства;
- 1.1.3. отказ оборудования и внутренних систем жизнеобеспечения.

Независимо от причины *физические* угрозы воздействуют:

- 1.1.1.1. на ресурс;
- 1.1.1.2. на канал связи.

Далее перейдем к рассмотрению *программных* угроз. Источниками *программных* угроз могут быть локальный и удаленный нарушитель, т.е.:

- 1.2.1. угрозы, исходящие от локального нарушителя;
- 1.2.2. угрозы, исходящие от удаленного нарушителя.

Объектом *локального нарушителя* может быть только ресурс. При этом на ресурсе *локальный нарушитель* может реализовать угрозы, направленные на:

- 1.2.1.1.1. операционную систему;
- 1.2.1.1.2. прикладное программное обеспечение;
- 1.2.1.1.3. информацию.

Угрозы, исходящие от *удаленного нарушителя*, могут воздействовать:

- 1.2.2.1. на ресурс.
- 1.2.2.2. на канал связи.

При доступе к *ресурсу* удаленный нарушитель может воздействовать на:

- 1.2.2.1.1. операционную систему;
- 1.2.2.1.2. сетевые службы;
- 1.2.2.1.3. информацию.

При воздействии на *канал связи* удаленный нарушитель может реализовать угрозы, направленные на:

- 1.2.2.2.1. сетевое оборудование;
- 1.2.2.2.2. протоколы связи.

Рассмотрим *организационные* угрозы. Организационные угрозы по характеру воздействия разделим на:

- 2.1. воздействие на персонал;
- 2.2. действия персонала.

Воздействие на персонал может быть:

- 2.1.1. физическим.
- 2.1.1. психологическим.

Как физическое, так и психологическое воздействие на персонал направлено на сотрудников компании с целью:

- 2.1.1.1. получения информации;
- 2.1.1.2. нарушения непрерывности ведения бизнеса.

Источником *действий персонала*, способных вызвать угрозы информационной безопасности, могут быть:

- 2.2.1. умышленные действия;
- 2.2.2. неумышленные действия.

Угрозы, вызванные *умышленными* и *неумышленными действиями* персонала, могут быть направлены на:

- 2.2.2.1. информацию;
- 2.2.2.2. непрерывность ведения бизнеса.

Пример классификации угроз представлен на рис. 1.

Классификация угроз Digital Security (Digital Security Classification of Threats)

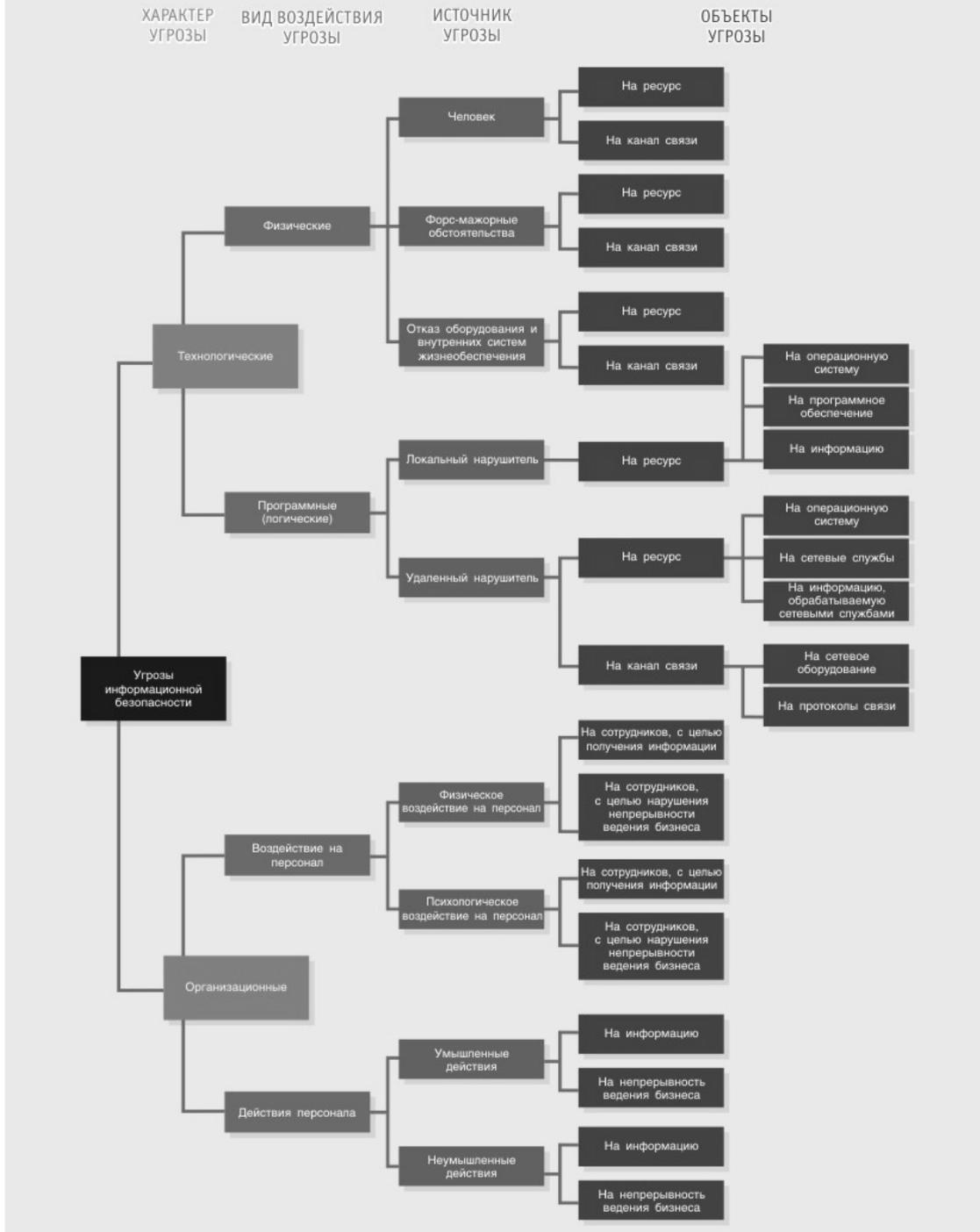


Рис. 1. Классификация угроз Digital Security

Заключение

При создании классификации необходимо было учесть все недостатки рассмотренных классификаций и выработать новую структуру классификационных признаков, удобных для описания реальной информационной системы.

Разработанная классификация основана на четырех классификационных признаках, каждый из которых четко структурирован и логически вытекает из предыдущего. Данная классификация учитывает не только технологические, но и организационные угрозы информационной безопасности, которые часто являются наиболее актуальными.

Учитывая тот факт, что угрозы и уязвимости информационной безопасности являются неразделимыми сущностями, уязвимости также можно классифицировать, используя разработанную классификацию.

Данная классификация наиболее применима к реальной информационной системе, так как была создана на основе практического опыта группы экспертов. Представленная классификация удобна благодаря тому, что каждая угроза имеет однозначное определение, т.е. относится к одному четко описанному классификационному признаку.

Литература

1. Christopher Alberts, Audrey Dorofee «OCTAVE Threat Profiles»; Software Engineering Institute, Carnegie Mellon University.
2. The Security Risk Management Guide v1.0, 2004, Microsoft Corporation.
3. Стандарт Банка России «Обеспечение информационной безопасности организаций банковской системы Российской Федерации. Общие положения»; дата введения: 2004-12-01.

ОЦЕНКА И УПРАВЛЕНИЕ РИСКАМИ ИТ-БЕЗОПАСНОСТИ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

**А.Л. Липатов (ЗАО «Лаборатория противодействия промышленному шпионажу»),
А.Н. Русинов, Е.А. Салмин**

Научный руководитель – доктор технических наук, профессор Ю.А. Гатчин

В статье даются практические рекомендации по созданию в компании системы управления рисками ИТ-безопасности, базирующейся на оценке рисков. Организация работ по управлению рисками даст возможность компании осуществлять экономически обоснованный выбор мер и средств защиты ИТ-активов с учетом имеющихся и потенциально возможных угроз безопасности информационной системе.

Введение

За последние годы информационные технологии (ИТ) значительно расширили возможности многих компаний для ведения бизнеса. Трудно представить работу современного предприятия без использования автоматизированных систем управления внутренними и внешними бизнес-процессами, систем бюджетирования и делопроизводства, электронной почты и Интернета. Однако повсеместное внедрение ИТ сделало компании более уязвимыми. В связи с этим руководством абсолютного большинства компаний необходимость защиты своих информационных систем (ИС) как от внешних, так и от внутренних угроз не подвергается сомнению.

В настоящей статье даются практические рекомендации по созданию системы управления рисками ИТ-безопасности. Внедрение такой системы в компании позволит обеспечить дифференцированный подход к выбору необходимых и достаточных мер и средств обеспечения ИТ-безопасности в зависимости от уровня ущерба, который может быть нанесен предприятию вследствие реализации угроз безопасности.

Оценка рисков ИТ-безопасности

Работы по анализу и управлению рисками ИТ-безопасности должны начинаться с проведения оценки рисков ИТ-безопасности. Оценку рисков предлагается осуществлять в следующем порядке:

- анализ рисков:
 - определение среды, границ и идентификация ИТ-активов;
 - идентификация угроз ИТ-безопасности;
 - анализ мер и средств обеспечения ИТ-безопасности и идентификация уязвимостей ИС;
 - определение вероятности реализации угроз;
 - оценка уровня ущерба;
- оценивание риска.

Анализ рисков. Определение среды, границ и идентификация ИТ-активов

При проведении данных работ должны быть определены:

- границы контролируемой зоны объекта информатизации, ИС;
- используемые меры и средства обеспечения ИТ-безопасности;
- состав пользователей ИС, их полномочия в системе и по доступу к защищаемым ИТ-активам. При этом в состав защищаемых ИТ-активов включаются аппаратные средства, программное обеспечение, информация, содержащая сведения, составляющие служебную, коммерческую, банковскую и т.п. тайну, средства обеспечения ИТ-безопасности, а также пользователи, имеющие доступ в ИС и обслуживающий персонал;
- внешние подключения ИС, потоки информации.

Анализ рисков. Идентификация угроз ИТ-безопасности

При идентификации угроз ИТ-безопасности должны быть выявлены все имеющиеся и потенциально возможные угрозы. Для каждой угрозы ИТ-безопасности должно быть произведено:

- описание источника угрозы:
 - тип источника угрозы (объективная или субъективная, внутренняя или внешняя, случайная или преднамеренная);
 - мотивация;
 - потенциал нападения;
- описание способа реализации угрозы;
- описание используемой уязвимости;
- описание вида защищаемых ИТ-активов, на которые воздействует угроза;
- описание вида воздействия на активы, порождаемые угрозой;
- описание нарушаемого свойства безопасности ИТ-активов [1].

Анализ рисков. Анализ мер и средств обеспечения ИТ-безопасности и идентификация уязвимостей ИС

При проведении анализа мер и средств обеспечения ИТ-безопасности и идентификации уязвимостей ИС должны использоваться:

- результаты анализа соответствия используемых мер и средств обеспечения безопасности установленным в компании требованиям. Требования по обеспечению ИТ-безопасности устанавливаются в следующих организационно-распорядительных и эксплуатационных документах: политике безопасности ИС, стандартах, регламентах, инструкциях и положениях по обеспечению и управлению ИТ-безопасностью;
- печатные и электронные источники, содержащие известные уязвимости средств обеспечения ИТ-безопасности;
- результаты работы автоматизированных средств выявления уязвимостей (сканеров уязвимостей, программ анализа рисков и т.п.);
- результаты тестирования средств обеспечения ИТ-безопасности (серверных и клиентских операционных систем, межсетевых экранов, средств построения виртуальных частных сетей (VPN), систем обнаружения/предотвращения вторжений (IDS/IPS) и т.п.).

Анализ рисков. Определение вероятности реализации угроз

При определении вероятности реализации угроз должны быть учтены:

- мотивация и потенциал источника угрозы;
- наличие и эффективность мер и средств обеспечения ИТ-безопасности.

Анализ рисков. Оценка уровня ущерба

Уровень ущерба от реализации угроз должен определяться как максимальный уровень ущерба для нарушаемых в результате реализации угроз свойств безопасности защищаемых ИТ-активов.

Оценивание рисков

Значение риска от реализации уязвимости должно определяться как функция вероятности возникновения ущерба, который может быть нанесен в результате реализации угрозы ИТ-безопасности.

Управление рисками ИТ-безопасности

Управление рисками ИТ-безопасности предлагается проводить следующим образом. В зависимости от значения риска по отношению к риску в соответствии со стратегией управления рисками, в компании должны предприниматься действия по управлению рисками, приведенные в табл.

Действия, предпринимаемые в случаях, когда значение риска является недопустимым	Действия, предпринимаемые в случаях, когда значение риска является допустимым
Перенос риска Снижение риска	Сохранение риска Принятие риска

Таблица. Возможные действия по управлению рисками в зависимости от значения рисков

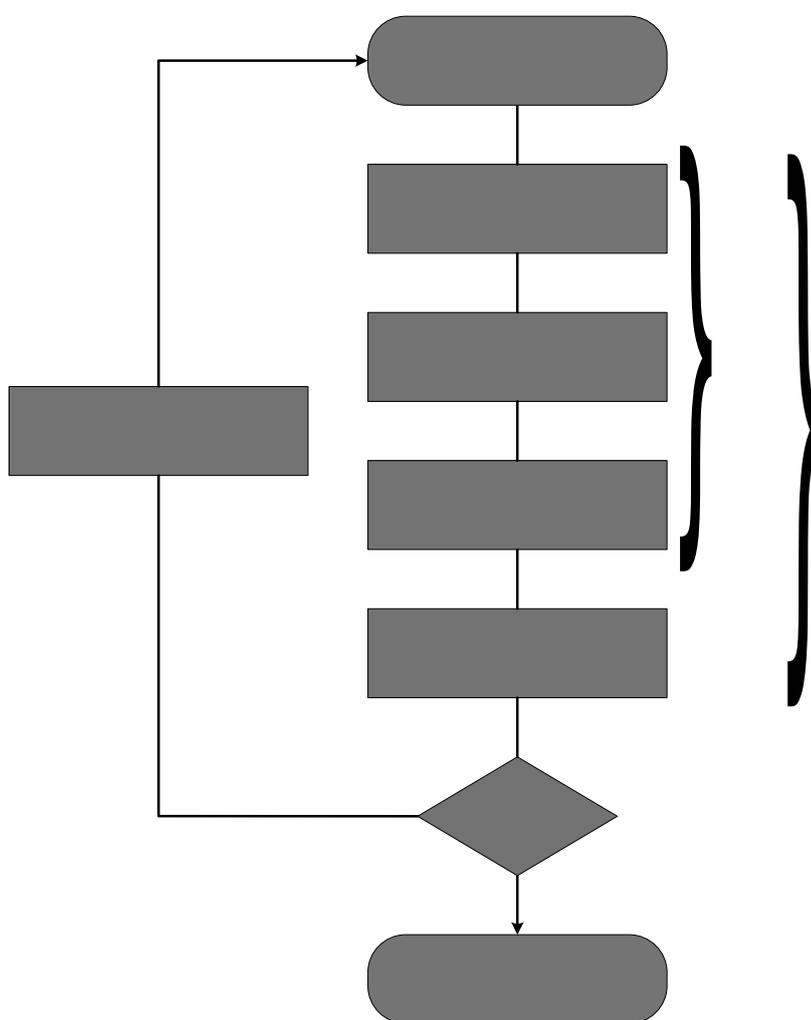


Рис. Интерактивный процесс оценки и уменьшения рисков ИТ-безопасности

Для рисков нарушения ИТ-безопасности со значениями, превышающими максимально допустимые, должно обеспечиваться снижение уровней этих рисков до допустимых значений (допустимый риск).

Допустимый риск представляет собой оптимальный баланс между фактическим уровнем защищенности ИС и установленным требованиями ИТ-безопасности, которым

**Определ
безопаснос
отношению**

должна соответствовать ИС, а также эффективностью затрат, удобством для пользователей и т.п. Допустимый риск должен достигаться с помощью интерактивного процесса оценки и уменьшения рисков ИТ-безопасности, приведенного на рисунке. Уменьшение рисков должно обеспечиваться путем применения соответствующих контрмер [2].

Заключение

Принятые в компании подходы к анализу и управлению рисками должны быть отражены в политике безопасности ИС. Результаты оценки рисков ИТ-безопасности и принятые в рамках менеджмента рисков по отношению к рискам действия по управлению рисками должны пересматриваться не реже одного раза в год или после внесения существенных изменений в ИС, которые могут повлиять на ИТ-безопасность.

Организация работ по анализу и управлению рисками в компании в соответствии с описанными выше рекомендациями позволит осуществлять выбор необходимых и достаточных контрмер, направленных на защиту ИТ-активов, с учетом имеющихся и потенциальных угроз ИС.

Литература

1. Руководящий документ «Безопасность информационных технологий. Критерии оценки безопасности информационных технологий. Часть 1. Введение и общая модель» (Гостехкомиссия России, 2002). С. 20–22.
2. BS ISO/IEC 17799:2005 «Information technology. Security techniques. Code of practice for information security management». Р. 54–55.

ПРИМЕНЕНИЕ СТАНДАРТА 802.1X ДЛЯ КОНТРОЛЯ ДОСТУПА В ЛОКАЛЬНУЮ ВЫЧИСЛИТЕЛЬНУЮ СЕТЬ

**А.Л. Липатов (ЗАО «Лаборатория противодействия промышленному шпионажу»),
А.Н. Русинов, Е.А. Салмин**

Научный руководитель – доктор технических наук, профессор А.Г. Коробейников

В статье рассмотрены возможности протокола 802.1x для контроля доступа в локальную вычислительную сеть и примеры реализации данного протокола. Также рассмотрены недостатки протокола и пути их решения.

Введение

Одним из наиболее незащищенных компонентов в локальной вычислительной сети является коммутационная розетка. Любой, имеющий физический доступ к коммутационной розетке, может подключиться в ЛВС и произвести DoS атаку, получить несанкционированный доступ к информационным ресурсам организации, используя многочисленные уязвимости программного обеспечения, либо использовать сниффер для прослушивания трафика сети организации. В данной статье рассматриваются имеющиеся решения данной проблемы на базе оборудования и программного обеспечения Cisco и Microsoft. Перечислим их.

Если известно, что коммутационная розетка не будет использоваться, то соответствующий порт на коммутаторе отключается командой shutdown. Если кто-либо захочет использовать розетку, ему необходимо будет обратиться к сетевому администратору для включения соответствующего порта коммутатора. Данное решение связано с увеличением стоимости владения ЛВС, особенно при большом количестве мобильных пользователей. Кроме того, оно не мешает злоумышленнику отключить от работающего порта оборудования и получить доступ в ЛВС.

Другим подходом к решению вышеописанной проблемы является использование функции port security на всех портах коммутатора. При данном подходе происходит регистрация на коммутаторе MAC-адреса устройства, который будет использовать порт. При возникновении трафика от устройств с другим MAC-адресом возможно временное отключение соответствующего порта коммутатора или полная блокировка последующего трафика. Данное решение кажется приемлемым до тех пор, пока вы не начнете прописывать для каждого порта коммутатора соответствующие ему валидные MAC-адреса. Возможно динамическое обучение MAC-адресов на порте коммутатора, однако и в этом случае при перемещении устройства потребуются переконфигурация коммутатора, что нельзя считать масштабируемым подходом.

В настоящее время перспективным решением по контролю доступа в ЛВС является внедрение в проводных и беспроводных сетях протокола 802.1x. Этот подход и будет рассмотрен в статье.

Стандарт 802.1x

Стандарт 802.1x описывает процедуру передачи EAP-сообщений сервером доступа (например, коммутатором или беспроводной точкой доступа) в проводных или беспроводных Ethernet-сетях. При этом стандарт 802.1x напрямую упаковывает EAP и сообщения в Ethernet-кадры, не применяя для их передачи протокол PPP. Это вызвано тем, что использовать протокол PPP во многих случаях не обязательно – например, при подключении Ethernet-рабочей станции, не поддерживающей протокол TCP/IP, или в том случае, когда использование протокола PPP является избыточным.

В стандарте 802.1x определяется три основных элемента:

- аппликant – пользователь, который нуждается в сетевой аутентификации;
- сервер аутентификации – обычно RADIUS-сервер, который производит фактическую аутентификацию;
- аутентификатор – сетевое устройство, находящееся между аппликантом и сервером аутентификации и предоставляющее доступ в сеть, например, точка доступа или Ethernet-коммутатор.

Ключевым моментом здесь является то, что сетевые устройства – аутентификаторы – могут быть достаточно простыми, поскольку для реализации функций 802.1x в них требуются минимальные аппаратные затраты, в то время как весь интеллект концентрируется в RADIUS-сервере. Такая схема имеет дополнительные выгоды и позволяет организовать тесную интеграцию управления сетевым оборудованием и сетевым ПО, что значительно облегчает управление информационной системой большого предприятия в целом. Протокол передачи EAP-сообщений в стандарте 802.1x называется EAPOL (EAP encapsulation over LAN) и в настоящее время определен для Ethernet ЛВС, а также беспроводных сетей стандартов серии IEEE 802.11 и ЛВС, использующих технологии token ring и FDDI.

Схема работы протокола EAPOL достаточно проста. Можно выделить следующие основные режимы работы.

1. Аутентификатор посылает запрос на аутентификацию (EAP-Request/Identity) аппликанту, как только он определит, что какой-то из его Ethernet-портов перешел в активное состояние (link active), т.е. к нему подключен сетевой адаптер. Таким образом, если отключить клиентскую станцию, которая уже прошла аутентификацию, и снова подключить к сетевому порту, то потребуются пройти аутентификацию еще раз.
2. Аппликant посылает сообщение/ответ (EAPResponse/Identity) аутентификатору, которое затем передается им на сервер аутентификации (RADIUS).
3. Сервер аутентификации в ответ отсылает пакет-запрос (challenge) аутентификатору, который затем переупаковывает его из IP-транспорта в EAPOL и передает аппликанту. В различных схемах аутентификации число таких сообщений может изменяться. В EAP поддерживается как аутентификация клиентской стороны, так и взаимная «сильная» аутентификация клиента и сервера, но только последний вариант считается приемлемым для использования в беспроводных сетях.
4. Аппликant отвечает на запрос соответственно выбранному алгоритму и передает его аутентификатору, который пересылает его на сервер аутентификации.
5. Если аппликant предоставляет правильный ответ на запрос, сервер посылает сообщение об успешной аутентификации аппликанту. В этой ситуации аутентификатор открывает клиенту доступ к ЛВС, который может зависеть от дополнительных параметров, передаваемых ему RADIUS-сервером, например, от номера VLAN или определенного уровня качества обслуживания.

Таким образом, использование сетевой аутентификации позволяет предоставлять пользователю определенный номер VLAN или уровень качества обслуживания вне зависимости от точки подключения в корпоративную ЛВС. Это обеспечивает как мобильность пользователей, так и постоянное соблюдение профиля безопасности сети – если даже сетевые кабели будут случайно перепутаны, пользователь не сможет войти в VLAN, доступ к которой ему запрещен [1].

Протокол RADIUS

Протокол RADIUS часто используется в различных сетевых устройствах (маршрутизаторы, модемные стойки, коммутаторы и т.д.) для аутентификации пользовате-

лей. Основной причиной этого является то, что сетевые устройства имеют обычно очень ограниченные аппаратные ресурсы и не могут хранить в памяти информацию о большом числе пользователей.

Протокол RADIUS обеспечивает централизованное управление пользователями, что очень важно в целом ряде случаев. Например, Интернет-провайдеры могут иметь десятки и даже сотни тысяч пользователей, и разместить такой объем информации в памяти любого сетевого устройства просто невозможно. При этом число пользователей может постоянно варьироваться в течение суток, дня или часа. Именно поэтому необходимо иметь централизованную базу данных, где хранится информация обо всех пользователях. Следует отметить, что протокол RADIUS поддерживается практически всеми производителями сетевого оборудования, в то время как другие протоколы аутентификации удаленных пользователей не получили массовой поддержки со стороны производителей.

Протокол RADIUS также имеет встроенные механизмы защиты от целого ряда сетевых атак, включая использование сетевых сниферов для получения паролей пользователей. Основными соперниками RADIUS на поле удаленной аутентификации являются протоколы TACACS+ и LDAP. Протокол LDAP изначально не имеет никаких средств защиты от снифинга паролей, и хотя в протоколе TACACS+ (в отличие от RADIUS) шифруется весь трафик, а не только пользовательские пароли, он также не лишен ряда слабых сторон.

Структура сообщения протокола RADIUS представлена на рисунке (RFC 2138), а значения и расшифровка поля Code – в таблице под рисунком.

Поле Identifier длиной один байт устанавливается RADIUS-клиентом в ответ на запрос RADIUS-сервера. Поле атрибутов содержит имя пользователя и пароль и также позволяет передавать дополнительные данные о клиенте от RADIUS-сервера сетевым устройствам, к которым непосредственно подключены пользователи.

Для прохождения аутентификации на сервере клиент создает запрос доступа (Access-Request) и передает его RADIUS-серверу, поле атрибутов данного сообщения должно включать как минимум имя пользователя и пароль. Поле идентификации запроса доступа также создается клиентом. Этот процесс не регламентируется в самом протоколе RADIUS, но обычно поле реализуется как простой счетчик, который увеличивается на 1 при каждом новом запросе. Запрос доступа содержит 16-байтное поле запроса аутентификатора (Request Authenticator), которое генерируется случайным образом. Данное сообщение в целом не защищено, шифруются только поля атрибутов, содержащие имя пользователя и пароль. Для этого клиент и сервер имеют общий секрет. Общий секрет совместно с полем запроса аутентификатора используется для вычисления 16-байтного значения (с помощью хэш-функции MD5), которое затем объединяется с паролем пользователя.

После получения сообщения запроса доступа RADIUS-сервер проверяет, обладает ли он общим секретом с клиентом, и если нет, то сообщение просто сбрасывается без уведомления клиента. Поскольку сервер также обладает общим секретом с клиентом, он может вычислить незашифрованное имя и пароль клиента (через процедуру, обратную описанной выше). Затем имя и пароль сверяются с пользовательской базой данных.

В случае успешной проверки имени и пароля пользователя сервер создает сообщение разрешения доступа и передает его пользователю, в обратном случае он получает сообщение об отказе в доступе. Оба сообщения имеют одинаковые номера идентификаторов, равные номеру идентификатора в запросе доступа клиента. Поле ответа аутентификатора (Response Authenticator) вычисляется с помощью применения хэш-функции MD5 над полями запроса аутентификатора и полями пакета разрешения доступа. Когда клиент получает сообщение-ответ от сервера, он проверяет, отсылал ли ра-

нее запрос с номером идентификатора, который указан в сообщении, и если нет, то оно просто сбрасывается. Далее клиент декодирует поле ответа аутентификатора с помощью процедуры, обратной вышеописанной, и сравнивает полученный результат с полем аутентификатора в поле запроса. Это гарантирует взаимную проверку клиента и сервера и делает практически невозможными хакерские атаки, основанные на подмене сервера [2].

Существующие недостатки в протоколе 802.1x и возможные пути их решения

Стандарт 802.1x является идеальным фундаментом для обеспечения безопасности беспроводных сетей, что уже неоднократно было продемонстрировано в материалах конференций и документации на веб-узле корпорации Майкрософт ([http://www.microsoft.com/wifi\(EN\)](http://www.microsoft.com/wifi(EN))). Но развертывание 802.1x без каких-либо дополнительных действий по защите проводных сетей от вредоносных компьютеров имеет серьезные недостатки. Одним из таких недостатков является сложность работы с устройствами, не поддерживающими этот стандарт. Другой недостаток – отсутствие управляемости: групповая политика AD предусматривает несколько объектов для управления 802.1x в беспроводных сетях, а для проводных интерфейсов таких объектов групповой политики не существует, как не существует и опубликованных интерфейсов прикладных программ (API) для управления клиентскими компьютерами в проводных сетях 802.1x. Отсутствие таких объектов групповой политики для проводных сетей 802.1x в Windows 2000 и Windows XP обусловлено рядом причин архитектурного свойства. Ввиду отсутствия централизованного управления крупномасштабное развертывание 802.1x в среде Windows в настоящий момент неосуществимо и повлечет большие расходы на поддержку.

Наконец, в протоколе есть одно чрезвычайно важное слабое место: проверка подлинности производится только при установлении подключения (в некоторых реализациях предусмотрена периодическая повторная проверка, но она не снижает риск атаки, описываемой ниже). После того, как подлинность просителя проверена и порт коммутатора открыт, дальнейший обмен данными между просителем и коммутатором осуществляется без проверки подлинности. Это создает ситуацию, в которой злоумышленник может подключиться к сети. (Хотелось бы поблагодарить Святослава Пидгорного, опытного специалиста по безопасности корпорации Майкрософт, который указал на эту уязвимость).

Подготовка этой атаки требует физического доступа к сети, поэтому в некоторых отношениях она является эзотерической. Атакующему требуется отключить компьютер (который мы будем называть «жертвой») от порта коммутатора сети, защищенной протоколом 802.1x, подключить к порту концентратор, подключить «жертву» к концентратору, а затем к тому же концентратору подключить атакующий компьютер (который мы будем называть «тенью»). Это тривиальная задача, если атакующий физически находится на вашей территории и имеет доступ к вашим разъемам Ethernet. Как вариант, атакующий может подключить к концентратору неконтролируемую точку беспроводного доступа, а затем произвести атаку с автостоянки вашего предприятия. (Разумеется, атакующий попытается скрыть свое присутствие, отключив трансляцию SSID этой точки доступа.)

Кратковременное отключение «жертвы» от сети не мешает успеху атаки. Когда компьютер-«жертва» снова подключится к сети, он снова успешно пройдет проверку подлинности на коммутаторе. Теперь наличие на пути коммутатора не будет играть роли, поскольку в основе своей коммутатор – всего лишь удлинитель с несколькими портами. Электрическое соединение «жертвы» с коммутатором по-прежнему присутствует. Концентраторы невидимы для протокола 802.1x.

Проделав все вышеописанное, атакующий настраивает MAC-адрес и IP-адрес своего компьютера так, чтобы они совпадали с адресами компьютера-«жертвы». Быстро определить эти адреса поможет анализ пакетов. Атакующий также настраивает на своем компьютере брандмауэр, который игнорирует весь входящий трафик, не являющийся ответом на инициированные им запросы.

Теперь рассмотрим принцип работы этой атаки. После успешной проверки подлинности компьютера-«жертвы» и открытия порта коммутатора атакующий получает доступ к ресурсам в защищенной сети. Это происходит из-за отсутствия по пакетной проверке подлинности трафика после того, как порт был открыт. Поскольку компьютер-«тень» имеет те же MAC- и IP-адреса, что и компьютер-«жертва», с точки зрения коммутатора они представляют собой один компьютер, подключенный к порту. Итак, отсутствие последующей по пакетной проверке подлинности в 802.1x создает условия для только что описанной атаки. Протокол 802.1x обеспечивает проверку подлинности только для самого подключения, предполагая, что весь трафик, идущий через это подключение, является санкционированным.

Обратим внимание, что обмен данными возможен только по протоколам без сохранения состояния – например, ICMP или UDP. Таким образом, атакующий может посылать тестовые запросы командой ping на компьютеры в сети и получить временный IP-адрес от DHCP-сервера (тот же IP-адрес, что и у «жертвы»), однако он не может вести обмен данными в сети по протоколу TCP: компьютер-«жертва» будет сбрасывать любые подключения, инициированные «тенью». Вот как это происходит.

- Компьютер-«тень» посылает пакет SYN на сервер в защищенной сети.
- Сервер возвращает пакет ACK-SYN, который принимается как «тенью», так и «жертвой».
- Компьютер-«жертва» не ожидает этого пакета ACK-SYN, поэтому отвечает серверу пакетом RST. В это время «тень» отвечает серверу пакетом ACK.
- Если пакет RST «жертвы» достигает сервера первым, сервер закрывает подключение и не обрабатывает пакет ACK от «тени». Если же первым приходит пакет ACK, посланный «тенью», то соединение устанавливается, но тотчас же разрывается приходящим вслед пакетом RST от «жертвы».

У этого правила есть одно интересное исключение. Если на компьютере-«жертве» работает межсетевой экран, отбрасывающий входящие пакеты ACK-SYN, на которые не было запроса (а большинство межсетевых экранов настроено именно так), то «жертва» не обработает принятый на шаге 2 пакет ACK-SYN, а, следовательно, и не отправит на сервер пакет RST. Оставшаяся часть описанного выше процесса выполнена не будет, и компьютер-«тень» сможет получить полный доступ в защищенную сеть. Это единственный известный авторам случай, когда наличие индивидуального меж сетевого экрана на компьютере может уменьшить безопасность остальной части сети! Разумеется, это не является поводом для отказа от развертывания межсетевых экранов; выгоды от их использования значительно перевешивают вероятность фактической реализации этой атаки.

Вышеописанная ситуация возможна вследствие того, что протокол 802.1x проверяет валидность прохождения только первого пакета, но не проверяет каждый пакет в отдельности. В случае повышенных требований к информационной безопасности ЛВС, возможно, стоит рассмотреть вариант одновременного развертывания 802.1x и IPsec. Так, протокол AH обеспечивает аутентификацию каждого пакета.

Также достаточно интересным является предложение ввести аутентификацию каждого фрейма на коммутаторе. С первого взгляда данное предложение может показаться в принципе неприемлемым в связи с возрастающей нагрузкой на коммутатор. Но проверка каждого поступающего фрейма и так выполняется на каждом современном коммутаторе – но не на валидность отправителя фрейма, а на отсутствие ошибок при

передаче фрейма (CRC). Но что если CRC будет представлять функцию не только от содержимого фрейма, но и от ключа, согласованного при обработке протокола 802.1x? В результате мы получаем проверку каждого фрейма коммутатором на отсутствие ошибок и на валидность отправителя, при этом, если реализовать данную функцию аппаратно, не произойдет увеличения задержки и нагрузки на коммутатор при прохождении фрейма через коммутатор. Существующие реализации протокола 802.1x поддерживают периодическую реаутентификацию аппликанта. При этом можно менять и ключ, используемый при расчете CRC, что обеспечивает стойкость к взлому ключа злоумышленником.

Также следует подчеркнуть, что имеющиеся решения для внедрения протокола 802.1x реализуют только проверку подлинности компьютера/пользователя. Между тем компьютер, имеющий валидные аутентификационные данные, также может представлять угрозу для ЛВС, например, в случае устаревшей антивирусной базы данных, отсутствия обновлений безопасности программного обеспечения, установленного нелицензионного программного обеспечения. Теоретически на базе протокола 802.1x можно реализовать проверку любого рода. Для этого необходимо внести изменения в программное обеспечение сервера аутентификации и аппликанта.

Конфигурация 802.1x на коммутаторе

Следующие коммутаторы Cisco могут быть использованы в качестве аутентификатора:

- Catalyst 6500/6000;
- Catalyst 5500/5000;
- Catalyst 4500/4000;
- Catalyst 3750/3550;
- Catalyst 2950;

со следующими установленными операционными системами:

- CAT 6.2 или выше;
- Cisco IOS 12.1(6) EA2 или выше.

При конфигурации контроля доступа на коммутаторе прежде всего необходимо указать IP адрес сервера RADIUS:

```
Switch#conf t
Switch (config)#aaa new-model
Switch (config)#radius-server host 192.168.100.100
Switch (config)#radius-server key 123qweASD
```

Далее необходимо включить процесс аутентификации 802.1x. В результате коммутатор становится аутентификатором. Для конфигурации коммутатора в качестве аутентификатора необходимо ввести следующие команды:

```
Switch (config)#aaa authentication dot1x default group radius
```

Последним шагом является конфигурация портов коммутатора. Порт коммутатора может быть в трех режимах:

- **force-authorized** – режим по умолчанию. В данном режиме порт всегда авторизован. Этот режим используется, если вы не хотите использовать 802.1x на определенном порту (например, при подключении к маршрутизатору или коммутатору);
- **auto** – нормальный режим для 802.1x. Порт в режиме авто посылает EAP пакеты серверу аутентификации и не переходит в авторизованное состояние без получения положительного ответа от сервера аутентификации;
- **force-unauthorized** – в данном режиме порт не перейдет в авторизованное состояние, даже если пользователь представит валидные учетные данные.

Для конфигурации режима функционирования порта коммутатора необходимо использовать следующие команды:

```
Switch#conf t
Switch (config)#interface fastethernet mod/port
Switch (config-if)#dot1x port-control [auto | force-authorized | force-
unauthorized]
```

После конфигурации коммутатора в режиме auto пользователи, подключенные к данному порту, не смогут передавать трафик через порт до тех пор, пока не будут авторизованы сервером аутентификации.

Важным моментом в данной схеме является то, что при использовании RADIUS-сервера появляется возможность назначать номер VLAN в зависимости от аутентификации пользователя [3].

Заключение

В настоящее время протокол 802.1x только начинает внедряться в проводных локальных вычислительных сетях, и внедрение это никак нельзя назвать массовым. Данный факт связан во многом с пренебрежением решениями по информационной безопасности в организациях, тем более, если соответствующие решения связаны с начальными расходами и увеличением стоимости владения сетевой инфраструктурой. Также к причинам медленного внедрения данного протокола можно отнести и его недостатки, а также то, что его теоретические возможности использованы не полностью. В статье выдвинут ряд предложений по усовершенствованию протокола, которые требуют внесения изменений в имеющееся программное и аппаратное обеспечение, в связи с чем в рамках статьи предложения были обозначены, но не реализованы.

Литература

1. 802.1X – Port Based Network Access Control. Institute of Electrical and Electronics Engineers, Inc. (<http://www.ieee802.org/1/pages/802.1x.html>).
2. RFC3580. IEEE 802.1X Remote Authentication Dial In User Service (RADIUS) Usage Guidelines. Author: P. Congdon, B. Aboba, A. Smith, G. Zorn, J. Roesse.
3. Catalyst 3750 Switch Software Configuration Guide, 12.2(25)SEE (http://www.cisco.com/en/US/products/hw/switches/ps5023/products_installation_and_configuration_guides_list.html).

СОВРЕМЕННАЯ СПЕЦИФИКА ОБЕСПЕЧЕНИЯ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ ЛОКАЛЬНЫХ ВЫЧИСЛИТЕЛЬНЫХ СЕТЕЙ СРЕДНИХ И КРУПНЫХ ПРЕДПРИЯТИЙ РОССИИ

К.В. Строганов

Научный руководитель – кандидат технических наук, доцент Б.А. Крылов

Цель статьи – осветить комплексный подход при обеспечении информационной безопасности ЛВС крупных предприятий России, а также рассмотреть проблемы управления затратами на информационную безопасность.

Введение

Развитие информационных технологий и систем неразрывно связано с проблемой обеспечения информационной безопасности, что является системообразующим фактором и воздействует на различные сферы деятельности предприятия, и в ходе технического прогресса это влияние будет расти. Закончился тот период, когда обеспечение безопасности локальных вычислительных сетей (ЛВС) в большей степени сводилось к локализации и компенсации уже произошедших инцидентов нарушения информационной безопасности, а службы, занимающиеся защитой информации, не могли угнаться за стремительным развитием средств и методов, использующих обширные уязвимости ЛВС. На данном этапе на крупных российских предприятиях все большее внимание уделяется комплексному анализу уязвимостей ЛВС и организации предупредительных мероприятий по предотвращению возможных реализаций угроз, что, несомненно, означает прогресс в сфере информационной безопасности.

Комплексный подход по обеспечению информационной безопасности предприятия может быть нецелесообразным без решения вопросов оценки эффективности затрат в этой области. Возросшее внимание руководителей служб информационной безопасности российских предприятий к оценке эффективности затрат определяется, в основном, двумя факторами: поворотом руководства к вопросам информационной безопасности предприятий и недостаточным уровнем информационной безопасности в существующих информационных системах. По мнению экспертов, в отечественных информационных системах с повышенными требованиями в области информационной безопасности (банковские системы, ответственные производства и т.д.) затраты на обеспечение режима информационной безопасности составляют до 30% всех затрат на информационные системы, и владельцы информационных ресурсов серьезно рассматривают экономические аспекты обеспечения информационной безопасности. Даже в тех информационных системах, уровень информационной безопасности которых явно недостаточен, у технических специалистов зачастую возникают проблемы обоснования перед руководством (владельцами информационных ресурсов) затрат на повышение этого уровня [2].

Возникают вопросы:

- как представить систему информационной безопасности для удобного подробного анализа?
- как проанализировать затраты на информационную безопасность, чтобы выработать соответствующие шаги по финансированию данной области?

В данной статье делается попытка осветить перечисленные вопросы.

Составляющие системы информационной безопасности ЛВС

Для ясного и четкого видения системы обеспечения информационной безопасности ЛВС предприятия в целом целесообразно условно выделить составляющие ее под-

системы. Как правило, на крупных и средних предприятиях присутствуют следующие подсистемы.

1. Подсистема хранения и резервирования данных. Задача данной подсистемы – безопасное хранение данных и обеспечение своевременной доступности информации для пользователей и приложений. Централизация и данных позволяет обеспечить необходимый уровень защиты путем уменьшения воздействующих на информацию угроз. Резервирование данных предназначено для защиты критической информации от уничтожения и модификации.

2. Подсистема межсетевого экранирования. Ее задачи:

- защита ЛВС от внешних, по отношению к данным, программно-технических воздействий (сетевых атак) через точку доступа;
- фильтрация пакетов данных;
- разграничение доступа и контролем информационного обмена между сегментами сети;
- защитное преобразование данных, передаваемых по незащищенным каналам (VPN соединения).

3. Подсистема обеспечения целостности, предназначена для защиты информации от модификации путем осуществления контроля целостности (неизменности) файлов системного и прикладного программного обеспечения (в том числе файлов специального ПО СЗИ) на серверах и АРМ, а также аппаратных объектов. Контроль реализуется различными способами:

- используется цифровая подпись;
- для критических файлов и директорий определены контрольные суммы и ведется их проверка;
- осуществляется физическая защита объектов ЛВС.

4. Подсистема антивирусной защиты, предназначена для защиты программ и данных на серверах и АРМ от компьютерных вирусов и иных вредоносных программ. Антивирусная защита информационных ресурсов осуществляется путем:

- проведения регулярных антивирусных проверок файлов на жестких дисках АРМ и серверов;
- мониторинга выполняющихся процессов;
- мониторинга сообщений электронной почты;
- проверки исполняемых файлов перед их запуском на выполнение.

5. Подсистема документации, регламентирующей информационную безопасность ЛВС. Нормативная база системы информационной безопасности предприятия регулирует отношения, возникающие при использовании средств защиты ЛВС предприятия.

Применение нормативной базы системы безопасности направлено на:

- повышение эффективности использования инженерно-технических, информационно-технологических и других средств защиты объектов и ресурсов предприятия;
- создание условий для повышения достоверности и полноты информации, используемой в процессе производства основного продукта или услуги предприятия;
- сокращение сроков выполнения работ с использованием информации ограниченного доступа;
- повышение оперативности обработки и анализа результатов контрольно-профилактических мероприятий;
- снижение финансовых затрат на обеспечение безопасности функционирования предприятия.

6. Другие подсистемы. В основном в этом пункте стоит упомянуть подсистемы информационной безопасности, которые обеспечивают бесперебойную работу ЛВС и

предупреждают реализацию угроз, влияющих на доступность сетевых ресурсов. В целях предотвращения отключения основных устройств ЛВС из-за перебоев с подачей электроэнергии используется подсистема бесперебойного питания. Подсистема резервирования средств ВТ обеспечивает доступность ресурсов при отказе критичного программно-аппаратного устройства (кластеризация).

Перечисленные подсистемы тесно взаимосвязаны между собой и образуют комплекс методов для защиты корпоративной ЛВС. При анализе угроз любой крупной системы необходимо выделить ее составляющие и их взаимосвязи друг с другом, а также большое внимание следует уделить документации, регламентирующей информационную безопасность.

Основой реализации комплексного подхода является разработка организационно-распорядительной документации ЛВС, которая регламентирует порядок обработки и хранения информации на уровне каждой из подсистем обеспечения информационной безопасности. На многих российских предприятиях главенствующую роль в подсистеме документации, регламентирующей информационную безопасность ЛВС, занимает руководящий документ «Концепция информационной безопасности» [1]. В «Концепции» изложены основные принципы, требования и задачи, описание направлений, функций, средств и способов обеспечения защиты, вытекающие из интересов предприятия в обеспечении информационной безопасности. Этот документ служит основой для разработки целевых программ по обеспечению информационной безопасности (ИБ) и подготовки предложений по их совершенствованию. Остальную документацию, регламентирующую информационную безопасность ЛВС предприятия, можно разделить на три вида.

1. Положения о подразделениях и должностные инструкции:

- положение об отделе информационных технологий (или отдела ИБ);
- должностные инструкции сотрудников отдела информационных технологий (или отдела ИБ);
- положения об иных подразделениях;
- должностные инструкции сотрудников иных подразделений;
- и др.

2. Технологические инструкции по защите информации:

- инструкция пользователя электронной почты предприятия и сети Интернет;
- инструкция по обеспечению антивирусной защиты;
- инструкция по ведению списка пользователей и наделению их правами доступа к ресурсам АС;
- инструкция по установке и модификации программного обеспечения АС;
- инструкция по организации резервного копирования;
- и др.

3. Приказы, стандарты предприятия, распоряжения по защите информации:

- распоряжение об упорядочивании использования ресурсов сети Интернет и электронной почты;
- положение по реализации механизмов защиты от НСД при проектировании, разработке и вводе в эксплуатацию программного обеспечения;
- положение по организации антивирусной защиты;
- и др.

Последовательный и глубокий анализ всех составляющих по каждой из подсистем, анализ соответствующей документации позволяет в достаточной мере выявить недостатки системы в целом и перейти к ликвидации основных уязвимостей.

Анализ затрат на информационную безопасность

Анализ затрат необходим для грамотного управления системой информационной безопасности и дает возможность целесообразно расходовать выделенные на информа-

ционную безопасность денежные средства, руководствуясь финансовой отчетностью. Представляется целесообразным применить уже существующую методiku компании Gather Group, при которой общие затраты на информационную безопасность делятся на три категории [2].

1. Затраты на контроль обеспечения информационной безопасности.

Основной объем затрат на контроль составляет оплата труда персонала службы безопасности и прочего персонала предприятия, занятого проверками и испытаниями. Эти затраты могут быть определены весьма точно. Оставшиеся затраты в основном связаны со стоимостью конкретных специальных работ и услуг внешних организаций и материально-техническим обеспечением системы безопасности.

2. Затраты на компенсацию потерь, связанных с реализацией какой-либо угрозы.

В основном это затраты на приобретение вышедшего из строя оборудования и восстановление данных, а также на оплату сверхурочных работ. Один из элементов внешних потерь, который невозможно точно вычислить – это потери, связанные с подрывом имиджа предприятия, снижением доверия потребителя к продукции и услугам предприятия (упущенная выгода).

3. Затраты на предупреждение реализации угроз.

Эти затраты сложно оценить, поскольку предупредительные мероприятия проводятся в разных отделах и затрагивают многие службы. Эти затраты могут появляться на всех этапах жизненного цикла ресурсов информационной среды предприятия:

- планирования и организации;
- приобретения и ввода в действие программных и аппаратных средств;
- доставки и поддержки средств обеспечения информационной безопасности;
- мониторинга процессов, составляющих информационную технологию.

В связи с этим необходимо провести анализ затрат (счет, договоров, ведомостей, накладных, записей в складской документации и т.п.) отдела информационной безопасности (информационных технологий) и выделить те данные, которые каким-либо образом определяют информационную безопасность ресурсов ЛВС на предприятии.

Результаты анализа затрат на безопасность и итоговый отчет должны показать объективную картину в отношении безопасности. В зависимости от увеличения затрат той или иной категории можно проанализировать, на какие меры обеспечения информационной безопасности были выделены средства, и понять, как в дальнейшем распределить финансирование. Все эти данные необходимы для лучшего осмысления экономики безопасности и выбора дальнейшей стратегии по улучшению защищенности информационной среды.

Заключение

На сегодняшний день одним из главных условий бесперебойной работы предприятия становится обеспечение безопасности его деятельности. Для решения этой задачи требуются определенные затраты. Понесенные затраты окупаются, если принятые меры сокращают или предотвращают ущерб от угроз безопасности функционирования предприятия. Таким образом, комплексный подход к обеспечению информационной безопасности и анализ затрат позволяют выбрать оптимальные методы и средства не только в области защиты информации, но и в области финансирования данной сферы.

Литература

1. Петренко С.А., Петренко А.А. Аудит безопасности Intranet. М.: ДМК Пресс, 2002.
2. Петренко С., Симонов С., Кислов Р. Информационная безопасность: экономические аспекты. / <http://www.jetinfo.ru>

ПРЕДОТВРАЩЕНИЕ УГРОЗЫ ВИРУСНЫХ АТАК В АВТОМАТИЗИРОВАННЫХ СИСТЕМАХ

С.В. Егоров, М.С. Иванов, Ю.И. Рядчин

Научный руководитель – доктор технических наук, профессор Ю.А. Гатчин

Антивирусные программы являются неотъемлемой частью любой системы защиты информации. В данной статье рассказывается про основные функции антивирусного программного обеспечения, выявляются типы этих программ, приводится аналитическое сравнение наиболее надежных и популярных программ, лучшая из которых была интегрирована в систему защиты информации автоматизированной системы ООО «Ленспецпроизводство».

Введение

За последние годы локальные вычислительные сети (ЛВС), совершив неоспоримый технологический прорыв, стали неотъемлемой составляющей производственно-технологической базы любой организации как крупного, так среднего и малого бизнеса не только в России, но и за рубежом. Корпоративные информационные системы становятся сегодня одним из главных инструментов управления бизнесом и фактически важнейшим средством производства любой компании. Однако новые информационные технологии, применяемые в ЛВС, не только значительно расширили возможности организаций, но и сделали их более уязвимыми для вероятных нарушителей информационной безопасности. В сложившейся обстановке одна из важных и сложных научно-технических задач на ближайшую перспективу состоит в создании и внедрении современных систем защиты информации (СЗИ) ЛВС и АС, призванных обеспечить безопасность информации, обрабатываемой, передаваемой и хранимой в ЛВС и АС.

При этом, решая данную задачу, необходимо четко осознавать, что для создания СЗИ недостаточно одной лишь установки различных средств защиты информации. Средства защиты, вне зависимости от их технических характеристик, сами по себе не способны обеспечить требуемый уровень информационной безопасности ЛВС. Во-первых, эти средства необходимо квалифицированно настроить. Во-вторых, в организации необходимо назначить персонал, который отвечал бы за администрирование этих средств защиты, в частности, и за защиту информацию в ЛВС в целом. В-третьих, необходимо провести обучение пользователей ЛВС вопросам обеспечения безопасности информации. И, наконец, необходимо разработать комплект документов, образующий информационную систему защиты информации ЛВС, который бы четко и однозначно определял порядок установки, настройки и эксплуатации используемых в ЛВС средств защиты, состав объектов и субъектов доступа и их полномочий в системе, перечень защищаемых информационных ресурсов ЛВС, схему технологического процесса обработки защищаемой информации, а также права и обязанности пользователей и администраторов ЛВС. Без наличия этих документов в организации говорить о созданной СЗИ ЛВС и/или АС невозможно.

В данной статье речь пойдет об антивирусной защите ЛВС и АС. На сегодняшний день существует множество различных антивирусных программ. В статье рассмотрены основные принципы работы такого программного обеспечения (ПО), приведен список наиболее распространенных на сегодняшний день программных средств по антивирусной ЗИ и обоснован выбор тех из них, которые были использованы для создания СЗИ АС ООО «Ленспецпроизводство».

Функции антивирусного программного обеспечения

На сегодняшний день перечень доступных антивирусных программ весьма обширен. Они различаются как по цене (от весьма дорогих до абсолютно бесплатных), так

и по своим функциональным возможностям. Наиболее мощные (и, как правило, наиболее дорогие) антивирусные программы представляют собой на самом деле пакеты специализированных утилит, способных при совместном их использовании поставить заслон практически любому виду зловредных программ.

Вот типовой (но, возможно, неполный) перечень тех функций, которые способны выполнять такие антивирусные пакеты:

- сканирование памяти и содержимого дисков по расписанию;
- сканирование памяти компьютера, а также записываемых и читаемых файлов в реальном режиме времени с помощью резидентного модуля;
- выборочное сканирование файлов с измененными атрибутами;
- распознавание поведения, характерного для компьютерных вирусов;
- блокировка и/или удаление выявленных вирусов;
- восстановление зараженных информационных объектов;
- принудительная проверка подключенных к корпоративной сети компьютеров;
- удаленное обновление антивирусного программного обеспечения и баз данных с информацией о вирусах, в том числе автоматическое обновление баз данных по вирусам через Интернет;
- фильтрация трафика (внешний – Интернет, внутренний – сетевой) на предмет выявления вирусов в передаваемых программах и документах;
- выявление потенциально опасных Java-апплетов и модулей ActiveX;
- ведение протоколов, содержащих информацию о событиях, касающихся антивирусной защиты.

К наиболее мощным и популярным на сегодняшний день (в России) антивирусным пакетам относятся:

- Doctor Web (в документации часто именуется более кратко – Dr. Web) – программа российской компании «ДиалогНаука»;
- Антивирус Касперского (в документации именуется более кратко – AVP) – разработка еще одной российской фирмы «Лаборатория Касперского»;
- Norton AntiVirus корпорации Symantec – сайт компании имеет русскоязычный раздел;
- McAfee VirusScan компании Network Associates.

Популярность перечисленных пакетов обусловлена, прежде всего, тем, что в них реализован комплексный подход к борьбе с вредоносными программами. Установив такой пакет на своем компьютере, вы избавляетесь от необходимости использовать какие-либо дополнительные антивирусные средства.

Последние версии антивирусных пакетов содержат в своем составе также и средства борьбы с вредоносными программами, проникающими из сети. Тем не менее, специализированные инструменты отражения «сетевых» угроз будут рассмотрены отдельно.

Технологии выявления и нейтрализации вредоносных программ

Так какие же, собственно, существуют технологии выявления и нейтрализации компьютерных вирусов?

Специалисты в области антивирусной защиты (в частности, Е. Касперский) выделяют пять типов антивирусов, реализующих соответствующие технологии: сканеры, мониторы, ревизоры изменений, иммунизаторы и поведенческие блокираторы.

Принцип работы антивирусного сканера состоит в том, что он просматривает файлы, оперативную память и загрузочные секторы дисков на предмет наличия вирусных масок, т.е. уникального программного кода вируса. Вирусные маски (описания) известных вирусов содержатся в антивирусной базе данных сканера, и если он встреча-

ет программный код, совпадающий с одним из этих описаний, то он выдает сообщение об обнаружении соответствующего вируса. В качестве маски вируса обычно используется так называемая «сигнатура», т.е. характерная для данного вируса последовательность байтов.

Недостаток любого сканера заключается в том, что он не способен обнаруживать новые (неизвестные) вирусы, о которых отсутствует информация в базе данных сканера. Кроме того, сканер практически бессилен против полиморфных вирусов, каждая новая копия которых отличается от предыдущей.

Мониторы являются разновидностью сканеров. Монитор постоянно находится в памяти компьютера и осуществляет автоматическую проверку всех используемых файлов в масштабе реального времени. Современные мониторы проверяют программы в момент их открытия и закрытия. Благодаря этому исключается возможность запуска ранее инфицированных файлов и заражения новых файлов резидентным вирусом. Для включения антивирусной защиты пользователю достаточно загрузить монитор при запуске операционной системы или приложения. Как правило, это делает сам антивирусный пакет в процессе его установки. В случае обнаружения вредоносной программы монитор, в зависимости от настроек, вылечит файл, заблокирует его выполнение или изолирует, переместив в специальную «карантинную» папку для дальнейшего исследования.

В настоящее время реально используются мониторы трех типов: файловые мониторы, мониторы для почтовых программ и мониторы для специальных приложений.

Файловые мониторы работают как часть операционной системы, в масштабе реального времени проверяя все используемые объекты, вне зависимости от их происхождения и принадлежности какому-либо приложению.

Мониторы для почтовых программ интегрируются в программы обработки электронной почты (как серверные, так и клиентские) и при поступлении нового письма автоматически проверяют его.

Мониторы для специальных приложений также обеспечивают фоновую проверку объектов, но только в рамках приложения, для которого они предназначены. Типичный пример – мониторы для MS Office. Подобно своим «почтовым коллегам», они интегрируются в программу и находятся в памяти компьютера во время ее работы.

Принцип работы ревизоров изменений основан на вычислении для файлов, системных секторов и системного реестра контрольных сверток (CRC). Они сохраняются в базе данных ревизора, и при следующем запуске ревизор сверяет «отпечатки» с их оригиналами и сообщает пользователю о выявленных отклонениях, обращая особое внимание на вирусоподобные изменения. При использовании ревизоров лечение зараженных объектов основывается не на опознании конкретного вируса, а на знании того, как должен выглядеть «чистый» файл или сектор: любые отклонения от эталона регистрируются ревизором, который способен вернуть объект к исходному состоянию.

Однако у ревизоров изменений имеются свои недостатки. Во-первых, они не способны поймать вирус в момент его появления в системе, а делают это лишь через некоторое время – после того, как вирус заразит хотя бы один объект. Во-вторых, они не могут определить вирус в новых файлах, поскольку для таких файлов в базе данных ревизора отсутствует эталонное значение CRC.

Иммунизаторы, которые иногда называют также вакцинами, действуют подобно медицинским вакцинам: они помещают в тело информационного объекта специальные метки, препятствующие заражению вирусом.

Иммунизаторы бывают двух типов: информирующие и блокирующие. Первые обычно записываются в конец файлов (по принципу файлового вируса) и каждый раз при запуске файла проверяют его на изменение. Такие иммунизаторы имеют один существенный недостаток: они не способны обнаружить заражение вирусами-

невидимками. Иммунизаторы второго типа защищают систему от заражения каким-либо определенным вирусом. Блокирующий иммунизатор помечает файлы таким же образом, как и нейтрализуемый вирус, благодаря чему тот считает их уже зараженными. Например, чтобы предотвратить заражение COM-файла вирусом Jerusalem, достаточно дописать в конец файла строку MSDos. Для защиты от резидентного вируса в память компьютера заносится программа, имитирующая копию вируса. При запуске вирус обнаруживает ее и считает, что система уже заражена.

Иммунизаторы не получили большого распространения и в настоящее время практически не используются.

Поведенческие блокираторы, в отличие от других антивирусных программ, не просто «смотрят» и пытаются «узнавать» вирусы, они еще и «думают». Принцип действия таких антивирусов основан на эвристическом анализе поведения активных программ и «подозрительных» событий, происходящих в системе. Другими словами, поведенческие блокираторы опираются в своей работе не на базу данных, а на базу знаний о том, что может происходить в системе в случае попадания в нее вредоносного программного обеспечения. Это делает поведенческие блокираторы универсальным оружием, которое может использоваться в борьбе не только с вирусами, но и, например, с программами-шпионами.

Считается, что разработка поведенческих блокираторов, использующих методы искусственного интеллекта, – это наиболее перспективное направление в создании антивирусных средств.

Существующие сегодня поведенческие блокираторы не предназначены для удаления вирусов. Их цель – обнаружение и предотвращение распространения вирусов. Поэтому их необходимо использовать совместно с антивирусным сканером, который способен уничтожить выявленный вирус.

Решения по антивирусной защите в АС ООО «Ленспецпроизводство»

Для предотвращения несанкционированных и непреднамеренных воздействий, утечки защищаемой информации за счет воздействия компьютерных вирусов в рамках проектирования и внедрения СЗИ АС ООО «Ленспецпроизводство» было решено установить и настроить на всех серверах и рабочих станциях АС антивирусную программу. Произведен выбор лучшего антивируса по соотношению «цена – эффективность – надежность». Для этого произведен анализ наиболее популярных и зарекомендовавших себя с точки зрения надежности антивирусных программ, имеющих на российском рынке. Результаты сравнения приведены в табл. 1.

Из данных, приведенных в таблице, видно, что лучшим для использования в АС является антивирус Dr. Web. Для АС целесообразным с точки зрения снижения накладных затрат на администрирование СЗИ АС является использование антивируса Dr. Web версии Enterprise, позволяющего производить централизованное управление всеми антивирусными средствами, установленными в АС. Обязательной установкой подлежит Spider, который должен функционировать на всех рабочих станциях и серверах резидентно.

Заключение

В статье рассмотрена такая важная компонента любой современной системы защиты информации в ЛВС и/или АС, как антивирусная защита. Выявлены основные функции антивирусного ПО: фильтрация трафика (внешний – Интернет, внутренний – сетевой) на предмет выявления вирусов в передаваемых программах и документах;

сканирование памяти и содержимого дисков; блокировка и/или удаление выявленных вирусов; протоколирование событий.

Антивирусное ПО можно разделить на пять типов, реализующих соответствующие технологии: сканеры, мониторы, ревизоры изменений, иммунизаторы и поведенческие блокираторы.

В процессе разработки и внедрения СЗИ АС ООО «Ленспецпроизводство» было решено использовать антивирусное ПО Dr. Web Enterprise. Этот выбор лучшего антивируса по соотношению «цена – эффективность – надежность» мотивируется тщательным анализом существующих сегодня на российском рынке наиболее надежных и популярных средств антивирусной защиты.

Характеристики	Наименование антивируса			
	Dr. Web	Антивирус Касперского Personal Pro	McAfee VirusScan	Norton AntiVirus
Цена (в долларах США)	50	69	49.95	49,95
Язык интерфейса	rus	rus	eng	eng
Условия распространения	share	share	share	share
Тестирование HD (время/кол-во файлов)	7min 20sec/31914	08min 56sec/31582	10min 56sec/11210 /*/*	5min 26sec/32788
Тестирование CD (время/кол-во файлов)	7min 32sec/7377	11min 20sec/15775	5min 50sec/1506 /*/*	4min 42sec/14128
Проверка файлов инсталляции	да	да	да	нет
Проверка архивов	да	да	да	да
Эвристический анализ	да	да	да	да
Проверка входящей/исходящей почты на этапе трафика	да	нет	нет	да
Размер монитора в ОЗУ	697 Кбайт	6,82 Мбайт	2,12 Мбайт	15,8 Мбайт
Плюсы программы	Есть опция проверки только почтовых файлов, компактный быстрый монитор, Dos-сканер	Большое количество настроек, возможность удаления зараженных архивов	Простота в использовании, надежность	Высокая надежность, лечение в zip-архиве
Минусы программы	Неудобство при настройке	Потребляет много системных ресурсов	Малое количество настроек	Небольшое количество настроек
Общая условная оценка	9	9	7	6

Таблица 1. Сравнительный анализ антивирусов

Литература

1. Гультаев А. Восстановление данных. СПб: Питер, 2004. 336 с.
2. <http://www.xakep.ru/post/26299/default.asp>
3. <http://www.securitylab.ru/analytics/216387.php>

ЗАЩИТА ИНФОРМАЦИИ ПРИ МЕЖСЕТЕВОМ ВЗАИМОДЕЙСТВИИ С ПОМОЩЬЮ МЕЖСЕТЕВЫХ ЭКРАНОВ

С.В. Егоров, М.С. Иванов, Ю.И. Рядчин

Научный руководитель – доктор технических наук, профессор А.Г. Коробейников

В статье описаны структура межсетевых экранов и выполняемые ими функции. На основе практического опыта авторов подробно рассмотрены особенности внедрения программного комплекса VipNet Custom как части системы защиты информации автоматизированной системы ООО «Ленспецпроизводство».

Перед авторами стояла практическая задача – создание системы защиты информации автоматизированной системы ООО «Ленспецпроизводство». Неотъемлемой частью современной системы защиты информации автоматизированной системы (АС) являются межсетевые экраны (МСЭ). В данной статье описаны принципы их функционирования, а также практические особенности их выбора и применения для защиты конкретной автоматизированной системы от несанкционированного доступа (НСД).

В рекламе современных производителей МСЭ обычно представляются в виде непроницаемой кирпичной стены, практически непобедимого защитника. Пользователям обещают полную автоматизированную защиту, способную блокировать все опасности еще на стадии их возникновения. Но так ли это?

Чтобы ответить на этот вопрос, необходимо определиться с основными понятиями. Прежде всего, чем является программный МСЭ? Его ядро есть не что иное, как фильтр – то, что находится между вашими приложениями и сетевыми компонентами ОС, и решает, что можно, а что нельзя. Для этого происходит внедрение данного фильтра в ключевые места взаимодействия приложений и сети и анализ всего проходящего трафика на предмет соответствия определенному набору правил. Все, что соответствует этим правилам, пропускается через фильтр, остальной трафик блокируется.

Как это отразится на внутренней архитектуре? По каким критериям производится анализ данных? Два этих вопроса являются основными при создании МСЭ, и результаты их решения становятся двумя основными компонентами архитектуры программного межсетевого экрана.

Первый компонент работает на пакетном уровне (уровни 3 и 4 модели OSI). Его задача заключается в обнаружении подозрительных и некорректных пакетов, опознавании сканирования портов и принятии решения о пропуске пакета в стек протокола. Пакеты могут анализироваться в соответствии со следующими критериями: формальная корректность пакета, направление пакета (входящий или исходящий), хост и порт отправителя, наличие установленных флагов.

Другой компонент МСЭ работает на более высоком уровне, имея дело с конкретными процессами. В его задачи входит определение, можно ли позволить процессу X инициировать соединение с данным хостом по данному порту, можно ли прослушивать данный диапазон портов и т.д. По сути своей, межсетевой экран является совокупностью двух фильтров – на уровне пакетов и на уровне процессов.

Детали реализации изменяются от продукта к продукту, но функциональные принципы ядра очень схожи. Серьезные программные МСЭ используют два драйвера для фильтрации на двух вышеупомянутых уровнях. Также обычно есть графический интерфейс, позволяющий изменять настройки, но основная работа происходит на уровне ядра, с использованием всего двух драйверов [1, 2].

Что касается более низкого уровня, пакетный фильтр обычно реализуется одним из двух способов.

Первый способ основан на использовании драйвера NDIS (Network Driver Interface Specification). Этот драйвер находится между драйвером сетевой карты и драйверами протоколов (TCP/IP и т.д.). В сущности, это виртуальный адаптер, который является

NIC драйвером для драйверов протокола и наоборот. Так как каждый входящий и исходящий пакет проходит через этот промежуточный драйвер, это позволяет осуществить атаку «man-in-the-middle – человек посередине» (в данном случае с благой целью).

Другой стандартный способ реализации пакетного фильтра основан на внедрении в NDIS, путем перехвата части функций библиотеки NDIS, используемых драйверами протоколов. Это означает, что пакетный фильтр находится в самом NDIS, между драйверами протоколов и чем-либо более низкоуровневым. Несмотря на то, что этот способ очень сильно отличается от предыдущего по реализации, функционально они подобны.

Теперь принцип работы пакетного фильтра должен быть понятен. Он анализирует каждый пакет в соответствии с критериями, указанными в наборе правил МСЭ, сохраненными в некоторой внутренней структуре данных. Осуществляется анализ таких параметров, как хост и порт отправителя и получателя, уровень фрагментации, тип протокола, флаги пакета, действительно ли пакет является частью уже открытого соединения и т.д. Например, если протокол – TCP, и пакет имеет SYN флаг (попытка открыть соединение), фильтр в зависимости от правил разрешит или запретит соединение для данного исходного и целевого хоста. Если все же соединение будет разрешено, фильтр добавит его во внутренний список открытых соединений. Таким способом межсетевой экран следит за текущими соединениями, осуществляя непрерывную инспекцию пакетов. Если пакет удовлетворяет правилам или принадлежит соединению из вышеупомянутого списка открытых соединений, он пропускается. Пакетный фильтр передает пакет на следующий уровень – драйвера протоколов, если пакет входящий, или драйвера сетевой карты, если пакет исходящий. Если пакет блокируется правилами, он никогда не будет передан на следующий сетевой уровень. В этом случае опционально, для обратной связи с пользователем, на экране может появиться предупреждение и запись в лог-файле.

Этот драйвер также присутствует во многих программных МСЭ. На уровне рассмотренного выше пакетного фильтра у нас не никакой информации о процессах. Работа происходит только с входящими и исходящими пакетами – вещами, которые с точки зрения прикладного приложения, происходят вне машины. Таким образом, для реализации фильтрации на уровне процессов нужно создавать фильтр, работающий на более высоком уровне. Данный механизм будет работать на уровне ядра, являясь оберткой над TDI (Transport Driver Interface) и перехватывая функции, которые приложения и/или вспомогательные библиотеки (WinSock) используют для передачи данных между собой и драйверами протоколов.

Есть также и другие методы. Например, WinSock API, набор функций, использующийся большинством приложений для доступа к сети, основан на многоуровневой модели, позволяющей вставлять расширения (extensions) третьих лиц между интерфейсом приложений и базовым сетевым протоколом. Такое расширение можно добавить, реализовав Layered Service Provider (LSP) и вставив его в LSP-цепочку. LSP – это стандартная Windows DLL, соответствующая некоторой спецификации и имеющая специальную функцию, которая предназначается для вставки в цепочку WinSock протокола. Согласно модели WinSock, все сетевые данные проходят через эту цепочку, в которой каждый LSP принимает решение о пропуске данных на уровень выше (или ниже, в зависимости от того, являются ли конкретные данные входящими или исходящими), предварительно обработав или изменив данные в соответствии со своей функцией. Quality of Service (QoS) является примером такого расширения, реализованного как LSP. Фильтр процессов может быть реализован в виде LSP, находясь в цепочке протокола и выборочно передавая данные к следующему элементу цепочки или блокируя их, руководствуясь собственными критериями.

Однако описанный метод – не лучшее решение для реализации фильтра, так как он действует только для приложений, использующих WinSock для передачи данных.

Для обхода фильтра процессов, основанного на LSP, нужно всего лишь воспользоваться собственным драйвером для прямой связи через TDI с драйвером протокола, что позволит обойтись без WinSock. Первый способ, основанный на обертке над TDI, является лучшей альтернативой, так как работает на более низком уровне.

Задача фильтра процессов – анализ попыток создания соединений приложениями. Он смотрит на идентификатор (PID) процесса, пытающегося отправлять или получать данные, и анализирует его характеристики на соответствие набору правил, используя более или менее сложный набор критериев, в зависимости от качества реализации фильтра процессов МСЭ. Но основной вопрос обычно выглядит так: разрешено ли приложению, создавшему этот процесс, выполнять те действия, которые оно пытается выполнить? Это задача сводится к проверке целостности файла, создавшего процесс, путем сравнения текущего хэша с известным, поиску файла в наборе правил, а затем проверке разрешения на выполнение соответствующих действий.

Существует несколько проблем, связанных с реализацией фильтров процессов. Например, неплохо было бы проверять, какие DLL присутствуют в адресном пространстве процесса, и действительно ли они должны там находиться. Это имеет смысл, так как код загруженных модулей исполняется в контексте процесса, в адресном пространстве которого они находятся. Таким образом, злонамеренный процесс может загрузить свой DLL в адресное пространство «хорошего» процесса X, и процесс X выполнит код из этого DLL перед своим собственным. Решением этой проблемы будет нахождение всех загруженных модулей, поиск их образов на диске и затем проверка на наличие в таблице модулей, разрешенных для загрузки данным процессом. МСЭ часто осуществляют такую проверку и спрашивают разрешения на загрузку модуля, но очевидно, что это не лучшее решение проблемы. Межсетевой экран не может знать все возможные настоящие и будущие DLL, «плохие» и «хорошие», поэтому пользователю приходится решать, разрешить ли загрузку данного модуля или нет. Другая проблема – был ли код процесса изменен в памяти. Такое также возможно, если процесс был создан злонамеренной программой. Это список можно долго продолжать...

Только некоторые из этих проблем адекватно решаются в современных МСЭ. Существует слишком много путей маскировки злонамеренного кода, которые могут ввести в заблуждение фильтры и простых, и сложных межсетевых экранов. Злонамеренный процесс может изменить в памяти код другого процесса. Можно использовать в своих интересах возможности легитимной программы, например, запустив из командной строки браузер, передав ему в аргументах указание загрузить злонамеренную страницу. Даже если фильтр отслеживает отношение родительский-дочерний процесс, можно использовать много уровней косвенности. Или враждебное приложение может применить агрессивную тактику против МСЭ, вообще его отключив с помощью эмуляции кликов мыши, закрывающих предупреждения. Все, что может сделать пользователь, может сделать и программа.

Все что написано выше, фокусируется на введении в заблуждение фильтра процессов. Совсем другие возможности открываются, если атакующему даже не нужно проходить через этот фильтр. Представьте, что информация о процессах, которую использует фильтр процессов, потеряна. Вся информация, которую в этом случае имеет МСЭ, – это входящий и исходящий трафик, работа с которым ведется на уровне пакетного фильтра. Все, что нужно сделать атакующему, – это использовать драйвер, который произведет инъекцию пакетов на уровень более низкий, чем уровень фильтра процессов, что даст возможность волноваться только о пакетном фильтре. Пока он будет использовать разрешенные порты, пакетный фильтр будет пропускать пакеты атакующего. А еще остается возможность работы под уровнем пакетного фильтра...

МСЭ является неотъемлемой частью созданной авторами системы защиты информации автоматизированной системы ООО «Ленспецпроизводство», включающей в

себя комплекс организационных мер и программно-технических средств защиты информации от несанкционированного доступа ней в АС.

Был проведен аналитический обзор представленных на рынке МСЭ и осуществлен выбор лучшего по соотношению «стоимость – эффективность – защищенность – надежность» решения для защиты серверов и рабочих станций. В табл. 1 представлены основные параметры программных МСЭ, наиболее широко распространенных на российском рынке [3–6].

Производитель, модель	Производительность, Мбит/с	Класс МСЭ	Стоимость (для АС), долл. США	Метод шифрования трафика	Сертификат Гос-техкомиссии России / класс защиты
1	2	3	4	5	6
Check Point, FW-1/VPN-1, ver. 1	100	Enterprise	2995	DES 40, 56, 168, RSA 512 /1024	На партию / 3
Cisco, PIX 520, ver. 4.2	1000	SOHO / Enterprise	5400	DES 56, 112, 186, CISCO PIX Ravlin encr. card	На единичные экземпляры / 4
ОАО «ЭЛВИС+», Застава, ver. 3.3	30–90	Enterprise	4040	DES 56, 112, 186, RSA 512/1024	На производство / 3
ОАО «Инфотекс», VipNet Custom	15–44	SOHO / Enterprise	2410	DES 56, 112, 186, RSA 512/1024, ГОСТ 28147-89	На производство / 3
ЗАО «НИП «Информ-защита», АПКШ Континент	100	SOHO / Enterprise	2900	ГОСТ 28147-89	На производство / 4

Таблица 1. Анализ распространенных на российском рынке МСЭ

На основе проведенного анализа был сделан вывод о том, что самым предпочтительным МСЭ для автоматизированной системы ООО «Ленспецпроизводство» является сертифицированный программный комплекс VipNet Custom (Сертификат Гос-техкомиссии России № 545 от 17.12.01 г.), в состав которого входят средства криптографической защиты информации «Домен-К» версии 2.0 (Сертификаты ФАПСИ №№ СФ/114-0613, СФ/124-0614 от 19.05.03 г.), а также МСЭ VipNet (Сертификат Гос-техкомиссии России № 546 от 17.12.01 г., Сертификат ФАПСИ № СФ/525-0558 от 01.10.01 г.).

Средства защиты семейства VipNet Custom было решено установить в АС согласно табл. 2 и настроить в соответствии с эксплуатационной документацией на средства защиты и разработанной политикой безопасности.

Технология VipNet предназначена для создания целостной системы доверительных отношений и безопасного функционирования главного узла АС и удаленных узлов АС, взаимодействующих как между собой, так и с сетью, и позволяет обеспечить защищенную передачу конфиденциальной информации по линиям связи, выходящим за пределы контролируемой зоны по сети Интернет. Технология VipNet позволяет объединить в единую виртуальную сеть главный узел АС и локальные вычислительные сети удаленных филиалов, а также защитить ее от НСД по классу 1Г для автоматизированных систем и по 3 классу для МСЭ.

В качестве криптографического ядра системы должно использоваться криптоядро «Домен-К», обеспечивающее использование в АС симметричных и асимметричных процедур распределения ключей, электронных цифровых подписей, автоматических процедур ключевого взаимодействия в системе.

№ п/п	Наименование элемента АС	Устанавливаемые компоненты системы VipNet Custom	Кол-во, клиентов	Примечание
1	Главный узел ЛВС АС	VipNet [Администратор]	1	
		VipNet [Клиент]	10	Устанавливается на серверы и рабочие станции, обрабатывающие наиболее критичную информацию
2	Удаленный узел ЛВС АС № 1	VipNet [Координатор]	1	
		VipNet [Клиент]	6	Устанавливается на серверы и рабочие станции, обрабатывающие наиболее критичную информацию
3	Удаленный узел ЛВС АС № 2	VipNet [Координатор]	1	
		VipNet [Клиент]	6	Устанавливается на серверы и рабочие станции, обрабатывающие наиболее критичную информацию
4	Портативные ПЭВМ сотрудников АС	VipNet [Клиент]	2	

Таблица 2. Установка средств защиты семейства VipNet Custom

В состав системы ViPNet CUSTOM, предполагаемой к использованию в АС, включаются следующие программные компоненты.

1. ViPNet [Администратор] устанавливается на один из компьютеров (администратора безопасности) распределенной сети (в ООО «Ленспецпроизводство» – в главном здании). Он позволяет:

- создать топологию VPN-сети АС и сгенерировать ключевую информацию для объектов VPN-сети, задать названия объектам VPN-сети и разрешить или запретить связи между ними;
- модифицировать объекты VPN-сети с последующей рассылкой обновленной справочной и ключевой информации тем объектам VPN-сети, которых коснулось конкретное изменение;
- централизованно обновить установленное ПО (ViPNet [Клиент] и ViPNet [Координатор]) в случае выхода новой версии.

2. ViPNet [Координатор] – многофункциональный модуль, выполняющий следующие функции:

- маршрутизацию почтовых и управляющих защищенных сообщений при взаимодействии объектов сети между собой и ViPNet [Администратором];
- осуществление в реальном времени регистрации и предоставление информации о состоянии объектов сети, их местоположении, значении их IP-адресов;
- обеспечение работы защищенных компьютеров локальной сети в VPN от имени одного адреса (функция проху);
- туннелирование пакетов от обслуживаемой ViPNet [Координатором] группы незащищенных серверов и рабочих станций ЛВС для передачи трафика от них к другим объектам VPN в зашифрованном виде по открытым каналам связи;
- фильтрацию трафика от источников, не входящих в состав VPN, в соответствии с заданной политикой безопасности (функция межсетевое экрана);
- обеспечение возможности работы защищенных по технологии ViPNet серверов и рабочих станций ЛВС через сетевые экраны и прокси-серверы других производителей.

3. ViPNet [Клиент] – модуль, обеспечивающий защиту информации при ее передаче в сеть, защиту от доступа к ресурсам рабочей станции или сервера ЛВС и атак на нее из локальных и глобальных сетей, а также реализующий следующие функции: защищенные службы для организации циркулярного обмена сообщениями, проведения конференций, защищенную почтовую службу и др. При этом ViPNet [Клиент] может быть установлен как на рабочую станцию, так и на все типы серверов с целью обеспечения безопасных режимов их использования.

Основой всех модулей системы ViPNet CUSTOM является ViPNet-драйвер, взаимодействующий непосредственно с драйвером сетевого интерфейса, что обеспечивает максимально возможную независимость программы от операционной системы и ее приложений. Драйвер контролирует весь IP-трафик, поступающий и исходящий из сервера или рабочей станции, и выполняет его фильтрацию по многочисленным параметрам и, при необходимости, шифрование и инкапсуляцию в единый тип IP-пакетов [5].

В статье описано, как фактически работают программные МСЭ, рассмотрены принципы работы пакетного фильтра и фильтра процессов. Выявлены потенциально уязвимые места, а также приемы обхода защиты межсетевых экранов в целом. Это позволяет сказать, что МСЭ, хотя и являются обязательной частью современных систем защиты информации, но панацеей от всех бед их назвать нельзя, как это иногда пытаются представить в рекламных буклетах их разработчики. Только применение МСЭ в комплексе с другими средствами защиты, описание которых выходит за рамки данной статьи, позволяет говорить о надежной и эффективной защите информации автоматизированной системы от несанкционированного доступа.

Литература

1. Голицкий А.В. и др. Защита информации в сети – анализ технологий и синтез решений. М.: ДМК Пресс, 2004. 616 с.
2. Щеглов А.Ю. Защита компьютерной информации от несанкционированного доступа. СПб: Наука и Техника, 2004. 384 с.: ил.
3. Информационно-методический журнал «Защита информации. Конфидент» № 6 (60). Ноябрь – декабрь 2004. СПб: «Келла Принт», 2004.
4. <http://www.infotecs.ru>
5. <http://www.sec.ru>
6. <http://www.iss.net>

ОБЗОР БЕЗОПАСНОСТИ БЕСПРОВОДНЫХ СЕТЕЙ СЕМЕЙСТВА 802.11

Э.В. Белов

Научный руководитель – доктор технических наук, профессор А.Г. Коробейников

В статье рассматриваются основные проблемы безопасности беспроводных сетей Wi-Fi, основанных на семействе стандартов 802.11. Приведены известные уязвимости и возможные методы их парирования.

Введение

За последние несколько лет популярность беспроводных сетей значительно возросла. На рост популярности беспроводного доступа в значительной мере оказывают влияние такие факторы, как интеграция в современные ноутбуки карт беспроводного доступа, появления PDA устройств, телефонов и т.д. Точки беспроводного доступа к различным сетевым сервисам (к примеру, Интернет) организовываются в кафе, ресторанах и других публичных местах. Компании объединяют разрозненные сегменты своих сетей беспроводным оборудованием в случае невозможности или нецелесообразности организации сетевых кабельных систем. Большая зона покрытия сетей на базе стандартов 802.11 – это одна из основных причин возникновения проблем с обеспечением безопасности, поскольку атакующий может находиться на значительном удалении от места физического развертывания сети.

Проблемы безопасности беспроводного доступа

Существуют различные протоколы и методы, применяемые для обеспечения безопасности беспроводных сетей.

1. Отключение широковещательной рассылки ESSID.

Параметр ESSID является идентификатором беспроводной сети. Он применяется для разделения пользователей беспроводной сети на логические группы. ESSID позволяет пользователю подключиться к требуемой беспроводной сети и, при необходимости, может быть сопоставлен с идентификатором виртуальной локальной сети (VLAN). Такое сопоставление необходимо для организации разграничения уровней доступа беспроводных пользователей к ресурсам корпоративной инфраструктуры.

При проектировании беспроводной сети существует заблуждение, что ESSID является одним из средств обеспечения безопасности, и отключение при широковещательной рассылке значения ESSID позволит усилить безопасность сети. На самом деле значение ESSID продолжает присутствовать в управляющих пакетах с запросами на повторную аутентификацию и повторное соединение.

2. Аутентификация с использованием MAC-адреса.

Аутентификация – это процесс определения личности клиента по предоставленной им информации, например, по имени и паролю. Многие производители беспроводного оборудования поддерживают аутентификацию пользовательских устройств по MAC-адресам, однако стандарт IEEE (Institut of Electrical and Electronic Engineers) 802.11 такой тип аутентификации не предусматривает.

Аутентификация по MAC-адресу без использования дополнительных методов обеспечения безопасности малоэффективна. Злоумышленнику достаточно просто получить доступ к беспроводной сети, в которой настроена только аутентификация по MAC-адресу. Для этого необходимо проанализировать радиоканал, на котором точка радиодоступа работает с клиентами, и получить список MAC-адресов устройств, которым открыт доступ к сети. Для получения доступа к сетевым ресурсам по беспровод-

ной сети необходимо заменить MAC-адрес своей беспроводной карты на известный MAC-адрес клиента.

3. Шифрованием с использованием статических ключей WEP.

WEP (Wired Equivalent Privacy) – протокол, который предназначен для шифрования трафика между точкой радиодоступа и ее пользователями. Шифрование WEP основано на недостаточно криптоустойчивом алгоритме шифрования RC4. Длина WEP ключа составляет 40 или 104 бита. К ключу добавляется нешифрованная последовательность символов (вектор инициализации) для успешного декодирования сигнала на обратной стороне размером 24 бита. Таким образом, принято говорить о длинах ключей 64 и 128 бит, однако эффективная часть ключа составляет всего лишь 40 и 104 бита.

Существует пассивные и активные атаки на WEP. К пассивным можно отнести следующие:

- атака методом полного перебора;
- атака FMS – предложена в 2001 году Скотом Флурером, Ицик Мантином и Ади Шамиром [1];
- улучшенная атака FMS [2].

Эти атаки основаны на анализе перехваченных пакетов беспроводной сети, и эффективность их работы зависит от количества собранных данных. Для получения секретного ключа WEP теоретически требуется перехватить порядка 6000000 пакетов, что может занять 3–4 часа [3].

Суть активных атак заключается в воздействии на беспроводную сеть с целью получения данных, после обработки которых будет получен доступ к ресурсам радиосети. К ним относятся:

- повторное использование вектора инициализации. Злоумышленник многократно посылает одну и ту же информацию (заранее известного содержания) пользователю, который работает в атакуемом беспроводном сегменте, через внешнюю сеть. Все время, пока злоумышленник посылает информацию пользователю, он также прослушивает радиоканал (канал между пользователем и атакуемой точкой радиодоступа) и собирает зашифрованные данные, в которых содержится посланная им информация. Затем злоумышленник вычисляет ключевую последовательность, используя полученные зашифрованные данные и известные нешифрованные;
- манипуляция битами. Атака основана на уязвимости вектора контроля целостности. Например, злоумышленник манипулирует битами пользовательских данных внутри фрейма с целью искажения информации третьего уровня. Фрейм не претерпел изменений на канальном уровне, проверка целостности на точке радиодоступа проходит успешно, и фрейм передается дальше. Маршрутизатор, получив фрейм от точки радиодоступа, распаковывает его и проверяет контрольную сумму пакета сетевого уровня, контрольная сумма пакета оказывается неверной. Маршрутизатор генерирует сообщение об ошибке и отправляет фрейм обратно на точку радиодоступа. Точка радиодоступа шифрует пакет и отправляет клиенту. Злоумышленник захватывает зашифрованный пакет с заранее известным сообщением об ошибке, после чего вычисляет ключевую последовательность.

4. Применение протокола WPA (Wi-Fi Protected Access).

В связи с множеством уязвимостей протокола WEP был разработан и принят в 2003 году новый протокол безопасного беспроводного доступа WPA [4]. Для шифрования WPA использует Temporal Key Integrity Protocol (TKIP) с динамически генерируемыми ключами длиной 128 бит. Для усиления аутентификации применяется стандарт IEEE 802.1x и Extensible Authentication Protocol (EAP).

В случае если аутентификация не производится средствами протокола 802.1x и сервером RADIUS, то применяется предварительно разделенный ключ (Preshared Key – PSK). Хотя на каждом клиентском компьютере может быть свой PSK, но в настоящее

время во всех реализациях используется один PSK на каждый ESSID точно так же, как в протоколе WEP. Поэтому в случае использования PSK возможны успешные атаки на TKIP путем получения и анализа перехваченных пакетов во время процедуры аутентификации.

В 2004 году был представлен стандарт беспроводного доступа второго поколения WPA2 [5]. В отличие от его предшественника WPA, он полностью базируется на заключительной редакции стандарта IEEE 802.11i и в качестве протокола шифрования использует AES (Advanced Encryption Protocol) с длиной ключа 128 бит.

Следует отметить, что беспроводные сети также могут подвергаться атакам DoS (Deny of Service), результатом которых становится отказ в обслуживании клиентов беспроводной сети. Суть этих атак заключается в том, чтобы парализовать работу беспроводной сети.

Специалистами Квинслендского технологического университета была опубликована информация об обнаруженной уязвимости, связанной с оценкой доступности радиоканала в технологии с прямой последовательностью распространения спектра (DSSS). На базе этой технологии реализован широко распространенный стандарт 802.11b. Злоумышленник, используя уязвимость, имитирует постоянную занятость беспроводной сети. В результате такой атаки все пользователи, работающие с точкой радиодоступа, по отношению к которой произошла атака, будут отключены.

Необходимо заметить, что данная атака может быть применима не только к оборудованию, работающему в стандарте 802.11b, но и к оборудованию стандарта 802.11g, хотя он не использует технологию DSSS. Это возможно тогда, когда точка радиодоступа, работающая в стандарте 802.11g, поддерживает обратную совместимость со стандартом 802.11b.

На сегодняшний день защиты от DoS атак для оборудования стандарта 802.11b не существует, но для избежания такой атаки целесообразно использовать оборудование стандарта 802.11g (без обратной совместимости с 802.11b).

Обсуждение и рекомендации

При проектировании и построении беспроводной сети необходимо уделить основное внимание безопасности, надежности, а также максимально упростить эксплуатационный процесс.

Перед построением сети беспроводного доступа необходимо произвести изучение местности, т.е., вооружившись точкой радиодоступа и портативным компьютером, выехать на объект предполагаемой инсталляции. Это позволит определить места наиболее удачного расположения точек радиодоступа, позволяя добиться максимальной зоны покрытия местности.

При построении системы безопасности беспроводного доступа необходимо помнить о трех составляющих:

- 1) архитектура аутентификации;
- 2) механизм аутентификации;
- 3) механизм обеспечения конфиденциальности и целостности данных.

В качестве архитектуры аутентификации используется стандарт IEEE 802.1x. Он описывает единую архитектуру контроля доступа к портам устройств с использованием различных методов аутентификации абонентов.

В качестве механизма аутентификации целесообразно использовать протокол EAP (Extensible Authentication Protocol). Протокол EAP позволяет осуществлять аутентификацию на основе имени пользователя и пароля, а также поддерживает возможность динамической смены ключа шифрования. Имена пользователей и пароли необходимо хранить на сервере RADIUS. Схема проектируемой сети представлена на рис.

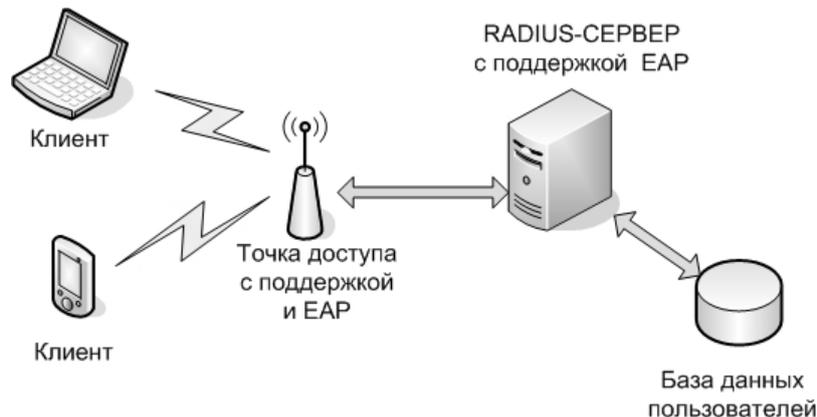


Рис. Схема безопасной беспроводной сети

В качестве механизма, обеспечивающего конфиденциальность и целостность данных, целесообразно использовать протокол TKIP (Temporal Key Integrity Protocol). Протокол TKIP позволяет усилить защиту WEP шифрования за счет таких механизмов, как MIC и PPK. Рассмотрим более подробно их предназначение.

MIC (Message Integrity Check) повышает эффективность функции контроля целостности в стандарте IEEE 802.11 за счет добавления во фрейм полей SEC (sequence number) и MIC, что позволяет предотвращать атаки, связанные с повторным использованием вектора инициализации и манипуляции битами.

PPK (Per-Packet Keying) – по пакетной смене ключа шифрования. Она позволяет снизить вероятность успешных атак, целью которых является определение WEP-ключа, но не гарантирует полную защиту.

Чтобы избежать атак типа «отказ в обслуживании», основанных на уязвимости технологии DSSS, беспроводная сеть будет построена на базе оборудования нового стандарта 802.11g (при этом стандарт 802.11g не должен быть обратно совместимым со стандартом 802.11b). Стандарт 802.11g основан на использовании технологии OFDM (Orthogonal Frequency Division Multiplexing), данная технология позволяет добиться скорости до 54Mbps.

Литература

1. Weakness in Key Scheduling Algorithm of RC4. / www.cs.umd.edu/~waa/class-pubs/rc4_ksaproc.ps
2. Practical Exploration of RC4 Weaknesses in WEP Environments. / www.dachboden.com/projects/bsd-air-tools/wepexp.txt
3. Владимирив А.А., Гавриленко К.В., Михайловский А.А. Wi –фу: «боевые» приемы взлома и защиты беспроводных сетей. М.: ИТ Пресс, 2005. 463 с.
4. Deploying Wi-Fi Protected Access (WPA™) and WPA2™ in the Enterprise. // Wi-Fi Alliance. March, 2005.
5. Wi-Fi Protected Access: Strong, standards-based, interoperable security for today's Wi-Fi networks. // Wi-Fi Alliance. April 29, 2003.

ОЦЕНКА МЕТРОЛОГИЧЕСКИХ ХАРАКТЕРИСТИК КОРРЕЛЯЦИОННОГО МЕТОДА ЧАСТОТНО-ФАЗОВЫХ ИЗМЕРЕНИЙ

В.Н. Ваняев

Научный руководитель – доктор технических наук, профессор В.Л. Ткалич

В статье приведено описание корреляционного метода частотно-фазовых измерений, описание макета измерителя, реализующего данный метод, и результаты оценки метрологических характеристик корреляционного метода на макете измерителя.

Введение

Область применения частотно-измерительной аппаратуры непрерывно расширяется. Одной из таких областей является радионавигация, в рамках которой стоит вопрос единой синхронизации с высокой точностью. Таким образом, задача построения высокоточных частотно-измерительных приборов является одной из главных в данном направлении.

Цифровые методы обработки информации, преимущества которых хорошо известны, широко распространены в настоящее время. Однако для реализации того или иного устройства необходимо грамотно выбрать метод измерения для конкретных условий его применения, оценить погрешность, произвести расчет и оценить метрологические характеристики выполненного прибора.

В настоящее время для высокоточного сличения частот сигналов стандартов частоты (СЧ) используется умножительно-преобразовательный метод, по которому осуществляется гетеродинирование сличаемых сигналов с одновременным умножением разности их частот. Реализация указанного метода приводит к увеличению массогабаритных и стоимостных характеристик измерителей. Кроме того, упомянутые измерители имеют дело с преобразованными сигналами и часто дают оптимистические оценки флуктуаций фазы и частоты [1].

Реализация оптимальных следящих дискретных алгоритмов частотно-фазовых измерений [2], в которых априорный интервал возможных значений частоты сигнала f_x задан: $f_x \in (f) - F, f) + F$, а начальная фаза φ_x полагается случайной равномерно распределенной величиной, приводит к различным схемам фазовой автоподстройки частоты (ФАПЧ). При этом при оценке метрологических характеристик частот сличаемых сигналов вторично возникает задача повышения точности получения цифровых отсчетов фазы или частоты подстраиваемого генератора.

Одним из перспективных путей повышения точности сличения близких частот является использование оптимальных измерителей начальной фазы с ортогональной корреляционной обработкой [3], в которых оценка разности частот сличаемых сигналов дается с помощью цифрового дифференцирующего устройства (ЦДУ).

Оценка начальной фазы по методу максимального правдоподобия может быть найдена в соответствии с выражением

$$\varphi_m = \arctg \frac{\int_{t_1}^{t_2} \sin((\omega_0 + \Delta\omega) \cdot t + \varphi_0) \cdot \sin \omega_0 t dt}{\int_{t_1}^{t_2} \sin((\omega_0 + \Delta\omega) \cdot t + \varphi_0) \cdot \cos \omega_0 t dt} \quad (1)$$

Проведя тригонометрические преобразования в (1), можно получить выражение

$$\Delta\varphi = \varphi_{m+1} - \varphi_m = \frac{\Delta\omega(t_3 + t_4)}{2} - \frac{\Delta\omega(t_1 + t_2)}{2} = \Delta\omega \cdot \tau, \quad (2)$$

где $\tau = t_3 - t_1 = t_4 - t_2$. Отсюда оценка разности частот $\Delta\omega$ сличаемых сигналов получится в виде $\Delta\omega = \frac{\Delta\varphi}{\tau}$.

Непосредственное использование алгоритма (1) сдерживается достижимой точностью выполнения математических операций в аналоговой форме. Использование дискретной версии упомянутого алгоритма ограничено из-за сложности реализации, приводящей к существенному увеличению массогабаритных и стоимостных характеристик [4]. Попытка упростить ортогональный измеритель использованием бинарного квантования [5], [6], хотя и приводит к положительным результатам в случаях, где не требуется высокая точность и быстродействие, но не обеспечивает работоспособность в условиях сильного сигнала, т.е. при больших отношениях сигнал/шум.

Свободной от указанных недостатков является структура корреляционного измерителя, использующая многоуровневое квантование выборок ортогональных компонент опорного сигнала, взятых в моменты времени, совпадающие с фронтами бинарно-квантованного измеряемого сигнала [7].

Далее приведено описание корреляционного измерителя, а также проведена оценка его метрологических характеристик.

Описание корреляционного измерителя

Структурная схема корреляционного измерителя представлена на рис. 1. Входной сигнал U_{fx} измеряемой частоты подвергается бинарному квантованию с помощью предельного ограничителя и далее поступает на входы синхронизации многоуровневых аналого-цифровых преобразователей (АЦП), имеющих устройства выборки и хранения (УВХ), в обоих квадратурных каналах. АЦП формируют цифровые эквиваленты напряжений $a_1 = A \sin \omega_0 t_1$ и $b_1 = A \cos \omega_0 t_1$.

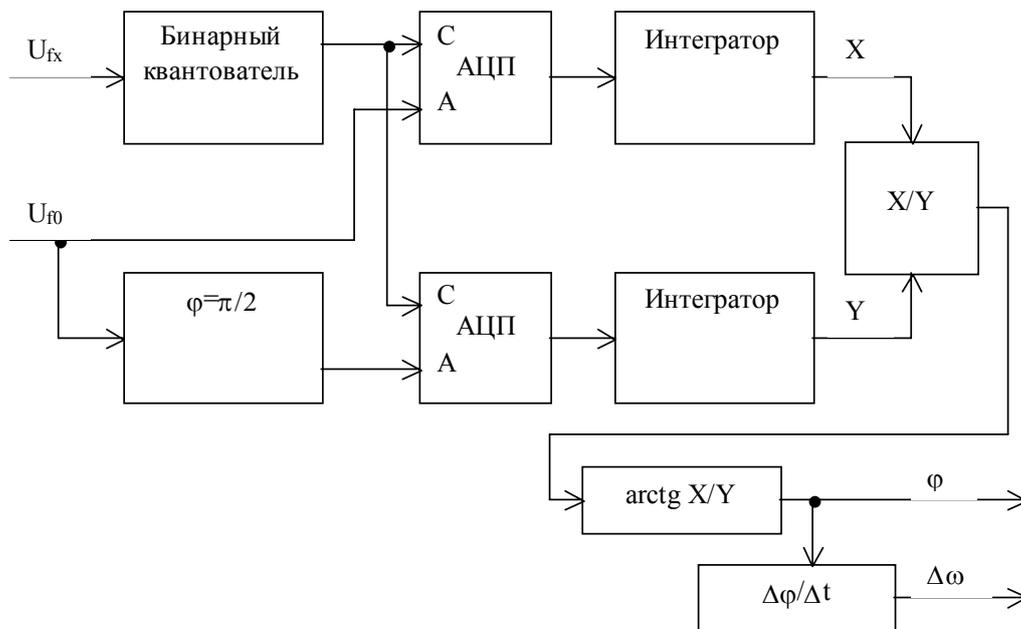


Рис. 1. Структурная схема корреляционного измерителя

Цифровые интеграторы осуществляют накопление кодов квадратурных компонент a_i и b_i в течение времени $T = \sum_i^n (t_{i+1} - t_i)$. Оценка фазового сдвига формируется в соответствии с выражением (1).

Анализ потенциальных точностных метрологических характеристик приведенной структуры проделан в [7], где показано, что среднее квадратичное значение погрешности измерения частоты (СКП) из-за шума квантования равно $\sigma(\Delta) = \frac{1}{\sqrt{3} \cdot 2^K \cdot n \cdot \tau}$, где K – разрядность АЦП, n – число накоплений в квадратурных каналах, τ – время измерения. Например, для получения СКП $= (2-3) \cdot 10^{-13}$ при $\tau=1$ с $n=100$, т.е. частота выборок квадратурных компонент должна быть 100 Гц, что не вызывает технических трудностей при реализации.

Описание макета корреляционного измерителя и экспериментальной установки

Структурные схемы макета измерителя и экспериментальной установки приведены на рис. 2 и 3, соответственно.

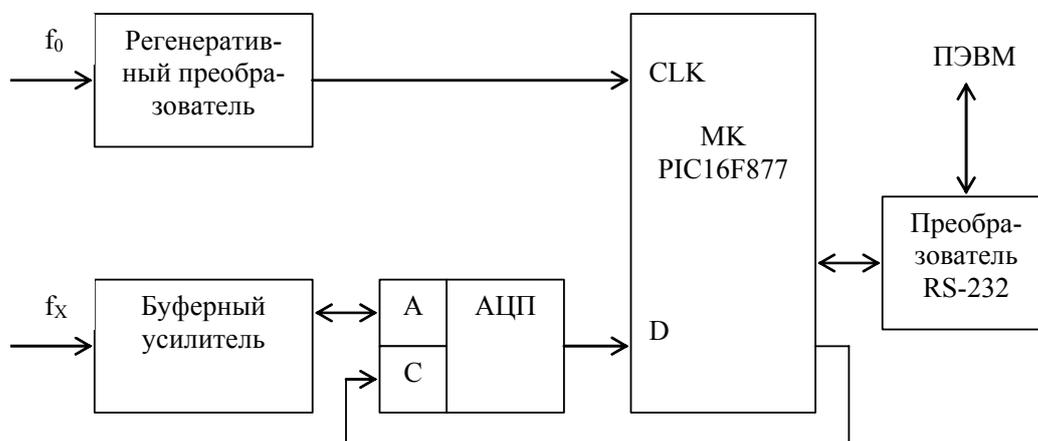


Рис. 2. Структурная схема макета измерителя

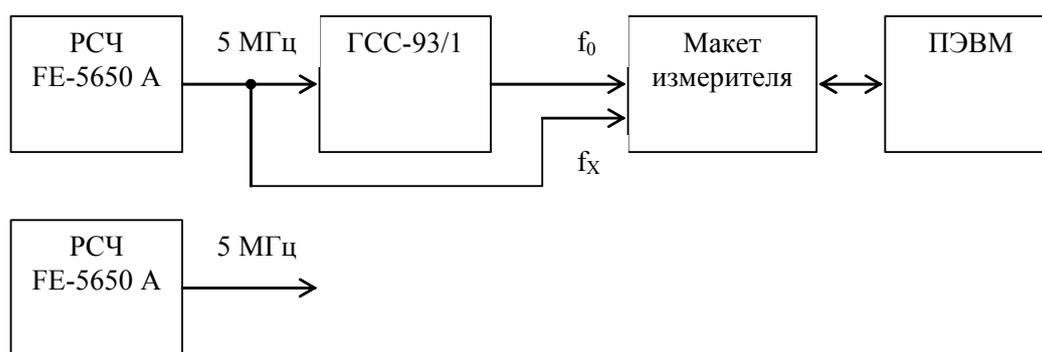


Рис. 3. Схема экспериментальной установки для проверки характеристик макета измерителя

Регенеративный преобразователь в макете измерителя осуществляет преобразование сигнала номинальной частоты 5 МГц в сигнал частотой 16 МГц для получения тактовой частоты микроконтроллера 4 МГц (период 250 нс). Это позволяет при номиналах частот сличаемых сигналов, равных 5 МГц (период 200 нс), осуществлять выборки квадратурных компонент последовательно во времени, что, в свою очередь, обеспечи-

вает использование одного многоуровневого АЦП для обоих каналов и исключает технические проблемы, связанные с неидентичностью амплитуд сигнала в квадратурных каналах.

В экспериментальной установке использовался генератор сигналов специальной формы ГСС-93/1 для оценки несмещенности измерения действительного значения частоты (ДЗЧ) макетом измерителя. ГСС-93/1 позволяет сдвигать частоту выходного сигнала с дискретностью 10^{-4} Гц.

Реальные сличения частот проводились путем измерения сигналов двух рубидиевых стандартов частоты (РСЧ) типа FE-5650 А фирмы FEI Communications Inc, США, имеющих стабильность $\frac{(1,4-3) \cdot 10^{-11}}{\sqrt{\tau}}$.

Алгоритм квадратурного корреляционного измерителя реализуется с помощью двух типов вычислительных устройств:

- микроконтроллер PIC 16F877;
- ПЭВМ.

Программное обеспечение (ПО) микроконтроллера обеспечивает:

- прием по интерфейсу RS-232 команд управления от ПЭВМ;
- формирование кодов выборок квадратурных каналов a_i и b_i ;
- передачу кодов a_i и b_i в ПЭВМ по интерфейсу RS-232.

ПО для ПЭВМ обеспечивает:

- прием кодов отсчетов a_i и b_i по интерфейсу RS-232;
- расчет и индикацию текущей фазы в соответствии с формулой (1) с устранением нелинейности накопленной фазы;
- расчет и индикацию приращений фазы в соответствии с выражением (2);
- расчет и индикацию среднего значения приращений фазы;
- расчет и индикацию двухвыборочной вариации (вариации Аллана) $\sigma_y(\tau)$ в соответствии с выражением:

$$\sigma_y(\tau) = \left\langle \frac{1}{2 \cdot \tau^2} \cdot [\varphi_{i+2} - 2 \cdot \varphi_{i+1} + \varphi_i]^2 \right\rangle^{1/2};$$

- возможность оперативного выбора числа n накапливаемых значений отсчетов в квадратурных каналах (время интегрирования);
- возможность оперативного выбора объема N массива обрабатываемых данных;
- возможность оперативного выбора интервала τ времени измерения.

Результаты измерений

Данные измерения ДЗЧ при отстройках с помощью ГСС-93/1 приведены в табл. 1.

Значение отстройки выходного сигнала ГСС-93/1, Гц	Измеренное значение отклонения частоты, Гц	Относительная погрешность измерения ДЗЧ
+0,000025	+0,000026	$2 \cdot 10^{-13}$
+0,0001	+0,000102	$4 \cdot 10^{-13}$
+0,001	+0,001001	$2 \cdot 10^{-13}$
-0,001	-0,001002	$4 \cdot 10^{-13}$

Таблица 1. Данные измерения ДЗЧ

Результаты измерений в режиме калибровки приведены в табл. 2, которая содержит результаты различных режимов измерения с использованием ГСС-93/1 и регенера-

тивного преобразователя в качестве преобразователя 5 МГц в 16 МГц. Также использовалось накопление в квадратурных каналах.

Время измерения, с		Преобразователь 5МГц → 16МГц	Относительная погрешность измерения	Вариация Аллана
1	n = 1	ГСС-93/1	$5,8 \cdot 10^{-12}$	$3,1 \cdot 10^{-10}$
	n = 6		$1,0 \cdot 10^{-12}$	$6,3 \cdot 10^{-11}$
		Регенеративный	$7,4 \cdot 10^{-13}$	$2,6 \cdot 10^{-11}$
100		ГСС-93/1	$5,0 \cdot 10^{-14}$	$7,0 \cdot 10^{-13}$
		Регенеративный	$4,5 \cdot 10^{-15}$	$1,2 \cdot 10^{-13}$
10		Регенеративный	$1,2 \cdot 10^{-15}$	$1,3 \cdot 10^{-12}$
1000		Регенеративный	$6,5 \cdot 10^{-15}$	$4,5 \cdot 10^{-14}$

Таблица 2. Результаты измерения в режиме калибровки

В табл. 3 приведены результаты сличения частот РСЧ типа FE-5650 А. Сличение проводилось макетом измерителя с ГСС-93/1 и регенеративным преобразователем в качестве преобразователя 5 МГц в 16 МГц, а также, для сравнения, анализатором частоты и временных интервалов ТСЮИ.468169.016 (АЧВИ), имеющего сертификат средства измерений (СИ) военного назначения.

Средство измерения	Время измерения, с	Относительная погрешность измерения	Вариация Аллана
АЧВИ	100	$1,0 \cdot 10^{-11}$	$1,7 \cdot 10^{-12}$
Макет измерителя с ГСС-93/1 в качестве преобразователя 5МГц → 16МГц	10	$1,0 \cdot 10^{-11}$	$2,3 \cdot 10^{-12}$
Макет измерителя с регенеративным преобразователем 5МГц → 16МГц	10	$1,9 \cdot 10^{-11}$	$9,0 \cdot 10^{-12}$
	100	$1,7 \cdot 10^{-11}$	$2,1 \cdot 10^{-12}$

Таблица 3. Результаты сличения частот РСЧ типа FE-5650 А

Заключение

В работе была проведена оценка метрологических характеристик корреляционного метода частотно-фазовых измерений. С помощью макета корреляционного измерителя был произведен ряд измерений в различных режимах и с использованием различного оборудования. На основе полученных данных можно сделать нижеследующие заключения.

Данные табл. 1 показывают, что при измерении ДЗЧ с помощью макета измерителя получаются несмещенные оценки. Погрешность измерения, находящаяся в диапазоне $(2-4) \cdot 10^{-13}$, определяется в основной части погрешностью установки частоты выходного сигнала ГСС-93/1. Этот вывод подтверждается измерениями в режиме калибровки (см. табл. 2).

Относительно высокая погрешность измерения в случае отсутствия накопления в квадратурных каналах вызвана следующими причинами:

- конструктивно-технологическое изготовление макета измерителя не позволило выполнить весь объем требований для исключения межканальных помех и аппаратных шумов;
- динамические погрешности многоуровневого АЦП типа AD9071, технические данные которого определяют минимальную частоту преобразования 40 МГц;
- использование ГСС-93/1 в качестве преобразователя 5 МГц в 16 МГц.

Увеличение частоты выборок АЦП в 6 раз привело к снижению погрешности измерения в 6 раз (табл. 2). Таким образом, подтверждается сделанный ранее вывод о возможности достижения относительной разрешающей способности ниже, чем $2 \cdot 10^{-13}$, при числе n выборок отсчетов АЦП более 100 за 1 с.

Использование регенеративного преобразователя 5 МГц в 16 МГц позволяет снизить погрешность измерения более чем на порядок даже при отсутствии накопителей в квадратурных каналах.

Полученные результаты (см. табл. 2) показывают, что измеритель позволяет получить приемлемые метрологические характеристики при времени измерения более 10 с даже при наличии конструктивно-технологических недостатков макетного исполнения. Этот же вывод можно сделать, исходя из данных табл. 3.

Оценка метрологических характеристик корреляционного метода частотно-фазовых измерений показала, что на базе корреляционного измерителя с квадратурной обработкой можно строить измерители времени и частоты на уровне не ниже образцовых средств измерения.

Литература

1. Stein S.R. Characterizing Precision Oscillators. // Measurement Techniques/ iming Solutions Corporation, 2004.
2. Фалькович С.Е., Хомяков Э.Н. Статистическая теория измерительных радиосистем. М.: Радио и связь, 1981.
3. Пестряков В.Б. Фазовые радиотехнические системы. М.: Советское радио, 1968.
4. Ткач В.И., Фисков М.М., Чмых М.К. Помехоустойчивый цифровой фазометр полигармонических сигналов. / Современные проблемы фазоизмерительной техники. Всесоюзная научно-техническая конференция. Красноярск, 1989.
5. Krajewsky and Bogus. Sequential Phasometer. // IEEE. Transactions in Instrumentation and measurement. March, 1974.
6. Заявка N 1-46835 Япония. Способ и устройство для цифрового измерения фазы сигнала в радионавигационной системе «Омега».
7. Пат. RU 2041469.C1. Российская Федерация. Измеритель характеристик фазовых флуктуаций. / Карелин В.А.

ОЦЕНКА ТОЧНОСТИ ДАННЫХ НАВИГАЦИОННЫХ СООБЩЕНИЙ КОСМИЧЕСКИХ АППАРАТОВ СПУТНИКОВЫХ РАДИОНАВИГАЦИОННЫХ СИСТЕМ

А.С. Бандура

Научный руководитель – доктор технических наук, профессор В.Л. Ткалич

Статья посвящена вопросам оценки точности данных, используемых в спутниковых радионавигационных системах. В работе произведен анализ характера возникающих погрешностей, намечены направления исследований точности координатно-временного обеспечения пользователей.

Введение

На нынешнем этапе развития спутниковые радионавигационные системы (СРНС) способны оказать существенную помощь в таких сферах хозяйственной деятельности человека, как навигация, телекоммуникации и др. Основной операцией, выполняемой СРНС, является решение навигационно-временной задачи. Эта задача решается в современных СРНС, таких как ГЛОНАСС и GPS, в соответствии с концепцией независимой навигации, предусматривающей вычисление искомым навигационных параметров непосредственно в аппаратуре потребителя. В рамках этой концепции в СРНС применяется метод беззапросных дальномерных измерений по сигналам нескольких навигационных космических аппаратов (НКА) с известными координатами [1]. Результатом решения навигационно-временной задачи является определение пространственных координат потребителя, скорости их изменения, а также точного времени.

Решение широкого круга задач с помощью СРНС невозможно без обеспечения привязки шкал времени аппаратуры, участвующей в вычислениях, причем недостаточная точность привязки может резко понизить точность решения этих задач. Одним из основных требований является обеспечение привязки ШВ НКА и приемника к ШВС.

1. Постановка задачи

1.1. Общая методика решения временной задачи

Общую методику определения расхождения БШВ КА относительно ШВ приемника, в соответствии с которой происходит решение задачи, можно представить в виде выражения:

$$\Delta \hat{T}_{ПРМ-КА} = S^i - T_{ГЕОМ} - \hat{\tau}_{ион}^i - \hat{\tau}_{трон}^i - \hat{\tau}_{прив}^i - \hat{\tau}_{рел}^i - \hat{\tau}_{прм}^i, \quad (1)$$

где S^i – измеренная псевдодальность между фазовым центром передающей антенны i -го НКА и фазовым центром приемной антенны приемника. Измерение псевдодальности есть измерение расхождения сигнала 1 Гц ШВ местного эталона времени и частоты (ЭВЧ) относительно сигнала 1 Гц, принимаемого с НКА, выраженное в секундах;

$$T_{ГЕОМ} = \frac{D^i}{c}, \quad (2)$$

где D^i – геометрическая дальность от КА до измерительного пункта,

$$D^i = \sqrt{[X_{П} - X_{КА}(t_K)]^2 + [Y_{П} - Y_{КА}(t_K)]^2 + [Z_{П} - Z_{КА}(t_K)]^2},$$

$X_{П}$, $Y_{П}$, $Z_{П}$ – прямоугольные геоцентрические координаты приемника в системе координат WGS-90; $X_{КА}(t_K)$, $Y_{КА}(t_K)$, $Z_{КА}(t_K)$ – прямоугольные геоцентрические координаты i -го НКА в системе координат на момент времени t_K ; c – скорость света; $\hat{\tau}_{ион}^i$ – временная задержка сигнала, обусловленная влиянием ионосферы; $\hat{\tau}_{трон}^i$ – временная

задержка сигнала, обусловленная влиянием тропосферы; $\hat{\tau}_{прив}^i$ – временная поправка, обусловленная тем, что псевдодальность измеряется от фазового центра антенны КА ГЛОНАСС, а соответствующие этому измерению эфемериды привязаны к центру масс КА; $\hat{\tau}_{рел}^i$ – временная задержка, обусловленная влиянием релятивистских эффектов взаимного движения ШВ; $\hat{\tau}_{прм}^i$ – временная задержка в приемнике и в соединительных кабелях между приемной антенной и приемником [2].

1.2. Источники ошибок, характерные для спутниковых измерений

Исходя из анализа измерительного процесса, характерного для СРНС, все основные источники ошибок можно условно разбить на три основные группы:

(1) ошибки, связанные с неточностью знания исходных данных, из которых определяющая роль принадлежит погрешностям знания эфемерид (перечня местоположений или расположений небесного объекта как функция времени [3]) КА;

(2) ошибки, обусловленные неточностями учета влияний внешней среды, среди которых выделяют такие источники, как воздействие атмосферы (ионосферы и тропосферы), а также отражение радиосигналов от окружающих объектов (многолучевость);

(3) инструментальные источники ошибок [3].

В зависимости от характера воздействия отмеченных источников возникающие ошибки подразделяются на две основные группы: систематические погрешности, которые применительно к спутниковым измерениям получили названия смещений, и погрешности случайного характера, которые часто отождествляются с названием «шум».

В данной работе проанализированы особенности влияния ошибок, связанных с неточностью эфемеридных данных.

Неточность расчета данных навигационного сообщения КА обуславливает соответствующие погрешности определения значений координат определяемых точек и, соответственно, величины геометрической дальности.

Передаваемая в составе навигационного сообщения информация «закладывается» на борт для последующей передачи пользователям заранее. Бортовые эфемериды также называют оперативными. Неточность бортовых эфемерид связана, прежде всего, с наблюдаемым на практике расхождением между предсказываемой (невозмущенной) и реальной (подверженной влиянию возмущающих сил) орбитами.

При использовании КНС GPS потребителями для решения широкого круга научных и прикладных задач точность определения расхождения шкал времени на основе информации, передаваемой в составе навигационного сообщения КА КНС GPS, в ряде случаев оказывается недостаточной. В таких случаях можно прибегнуть к использованию апостериорных эфемерид вместо бортовых эфемерид. Апостериорные эфемеридные данные поступают от специально организованных служб, которые накапливают реальные, а не предсказанные значения эфемерид. Декларируемая такими службами точность эфемеридной информации на порядок выше точности оперативных эфемерид, однако для пользователя, вынужденного работать в реальном времени, этот метод, естественно, неприменим.

На основании изложенного можно сделать вывод о важности оценки погрешностей, вносимых неточностью бортовых эфемеридных данных и поиске методов их понижения.

2. Оценка точности определения геометрической дальности с использованием оперативных эфемеридных данных

В качестве одной из методик оценки точности бортовых эфемерид можно предложить их сравнение с уточненными эфемеридами. Расчет координат КА при этом производится на одни и те же временные точки, затем по соотношению (1.2) производится расчет величины геометрической дальности.

2.1. Схема расчета координат КА на с использованием апостериорных эфемерид

Значения апостериорных эфемерид есть таблично заданная функция, поэтому при их размножении (вычислении значений координат на заданные временные точки) применяется аппроксимации полиномами, что обеспечивает минимальные погрешности расчета размножаемых величин. При размножении апостериорных эфемерид применяется интерполяция полиномом Лагранжа.

Исходными данными являются $X_{КА}(t_K), Y_{КА}(t_K), Z_{КА}(t_K)$ – прямоугольные геоцентрические координаты НКА.

Интерполяционный полином Лагранжа имеет вид:

$$L_n(x) = \sum_{i=0}^n f(x_i) \prod_{i \neq j} \frac{x - x_j}{x_i - x_j} = \sum_{i=0}^n \frac{(x - x_0)(x - x_1) \dots (x - x_{i-1})(x - x_{i+1}) \dots (x - x_n)}{(x_i - x_0)(x_i - x_1) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_n)} f(x_i), \quad (3)$$

где x_0, x_1, \dots, x_n – равноотстоящие временные точки, в которых известны значения координат, x – значение временной точки, в которой необходимо определить координаты, $f(x_i)$ – значения координаты в i -ой точке, $L_n(x)$ – значение координаты в искомой точке [4]. Вычисления проводятся последовательно для всех трех прямоугольных координат; выходными данными алгоритма являются $X_{КА}(t_K), Y_{КА}(t_K), Z_{КА}(t_K)$ – прямоугольные геоцентрические координаты НКА.

2.2. Схема расчета координат КА на с использованием оперативных эфемерид

Исходными данными являются эфемеридная информация, содержащаяся в кадре навигационного сигнала НКА GPS:

\sqrt{A} – корень квадратный из большой полуоси, [м^{1/2}];

Δn – возмущенное значение среднего движения в полуциклах за секунду, [пц/с];

e – эксцентриситет орбиты;

C_{rs} – синусная составляющая возмущения орбиты по геоцентрическому радиусу, [рад];

C_{rc} – косинусная составляющая возмущения орбиты по геоцентрическому радиусу, [рад];

C_{is} – синусная составляющая возмущения орбиты по боковому уклонению, [рад];

C_{ic} – косинусная составляющая возмущения орбиты по боковому уклонению, [рад];

C_{us} – синусная составляющая возмущения орбиты вдоль орбиты, [рад];

C_{uc} – косинусная составляющая возмущения орбиты вдоль орбиты, [рад];

ω – аргумент перигея, [пц];

M_0 – средняя аномалия, [пц/с];

Ω_0 – долгота восходящего узла орбиты, [пц];

Ω_0' – скорость ухода восходящего узла орбиты, [пц/с];

i_0 – наклонение орбиты, [пц];
 i' – скорость ухода наклона орбиты, [пц/с];
 t_{oe} – опорное время представления эфемерид, [с];
 t_k – текущее время (временная точка, на которую необходимо определить координаты КА), [с].

Используемые константы:

$\mu = 3,986005 * 10^{14}$ – геодезическая константа гравитационного поля Земли, [м³/с²];

$\Omega_3 = 7,2921151467 * 10^{-5}$ – угловая скорость вращения Земли, [рад/с];

$c = 2,99792458 * 10^8$ – скорость света, [м/с].

Ниже приведен алгоритм расчета координат КА с использованием бортовых эфемерид [5].

- Рассчитываются вспомогательные параметры:

$$A = (\sqrt{A})^2;$$

$$n_0 = \sqrt{\mu / A^3};$$

$$\Delta t_k = \begin{cases} t_k - t_{oe}, & \text{если } (t_k - t_{oe}) \leq |302400|, \\ t_k - t_{oe} - 604800, & \text{если } (t_k - t_{oe}) > 302400 \text{ с}, \\ t_k - t_{oe} + 604800, & \text{если } (t_k - t_{oe}) < -302400 \text{ с} \end{cases}$$

$$n = n_0 + \Delta n;$$

$$M = M_0 + n * \Delta t_k.$$

- Определяется эксцентрическая аномалия E_k путем решения уравнения Кеплера $M = E_k - e * \sin E_k$ методом последовательных приближений:

$$E_0 = M, E_j = M + e * \sin E_{j-1}, \text{ где } j = 1, 2, 3 \dots, \text{ до тех пор, пока } |E_j - E_{j-1}| < 10^{-11}.$$

- Рассчитываются вспомогательные поправки:

$$\sin V_k = \frac{(1 - e^2)^{1/2} \sin E_k}{(1 - e * \cos E_k)};$$

$$\cos V_k = \frac{(\cos E_k - e)}{(1 - e * \cos E_k)};$$

$$V_k = \arctg\left(\frac{\sin V_k}{\cos V_k}\right);$$

$$\Phi_k = \omega + V_k;$$

$$\delta_{uk} = C_{us} * \sin 2\Phi_k + C_{uc} * \cos 2\Phi_k;$$

$$\delta_{rk} = C_{rs} * \sin 2\Phi_k + C_{rc} * \cos 2\Phi_k;$$

$$\delta_{ik} = C_{is} * \sin 2\Phi_k + C_{ic} * \cos 2\Phi_k;$$

$$U_k = \Phi_k + \delta_{uk};$$

$$r_k = A(1 - e * \cos E_k) + \delta_{rk};$$

$$i_k = i_0 + \delta_{ik} + i' * \Delta t_k;$$

$$X'_k = r_k * \cos U_k;$$

$$Y'_k = r_k * \sin U_k;$$

$$\Omega_k = \Omega_0 + (\Omega'_0 - \Omega_0 * e) * \Delta t_k - \Omega'_0 * t_{oe}.$$

- Координаты НКА GPS определяются по формулам:

$$X_{KA}(t_k) = X'_K * \cos \Omega_k - Y'_K * \cos i_k * \sin \Omega_k;$$

$$Y_{KA}(t_k) = X'_K * \sin \Omega_k - Y'_K * \cos i_k * \cos \Omega_k;$$

$$Z_{KA}(t_k) = Y'_K * \sin i_k.$$

2.3. Оценка погрешностей

Был разработан пакет прикладных программ, на основании результатов работы которого можно оценить уровень погрешностей, вносимых неточностью бортовых эфемерид.

Были проведены исследования для нескольких НКА GPS. В качестве примера приведены результаты расчетов по эфемеридам одного из аппаратов (PRN 23). Результаты расчетов для других НКА и характер их изменения во времени сходны с приведенным в данной работе примером, и приводить их здесь нет необходимости. Для различных НКА величины погрешностей составляли до 8–10 нс.

Относительно низкие (менее 2 нс) погрешности при вычислениях геометрической дальности можно объяснить тем, что для последних запущенных спутников (среди которых и рассматриваемый аппарат) наземный сегмент СРНС GPS производит расчет бортовых эфемерид с учетом составляющей поправки $\hat{\tau}_{прив}^i$, в то время как для остальных эту поправку приходится учитывать в аппаратуре пользователя, что не всегда удается сделать с высокой точностью.

На рис. 1 изображен график зависимости разности между величинами геометрической дальности, рассчитанной по оперативным и апостериорным эфемеридам, от времени для КА PRN 23. Интервал времени, на котором проводились расчеты, составляет 7 суток.

На рис. 2 приведен график зависимости разности между величинами геометрической дальности, рассчитанной по оперативным и апостериорным эфемеридам, от времени для КА PRN 23. Интервал времени, на котором проводились расчеты, составляет 4 суток.

На основании анализа рисунков явно видна систематическая составляющая погрешности с периодичностью в двенадцать часов, что соответствует периоду обращения НКА GPS.

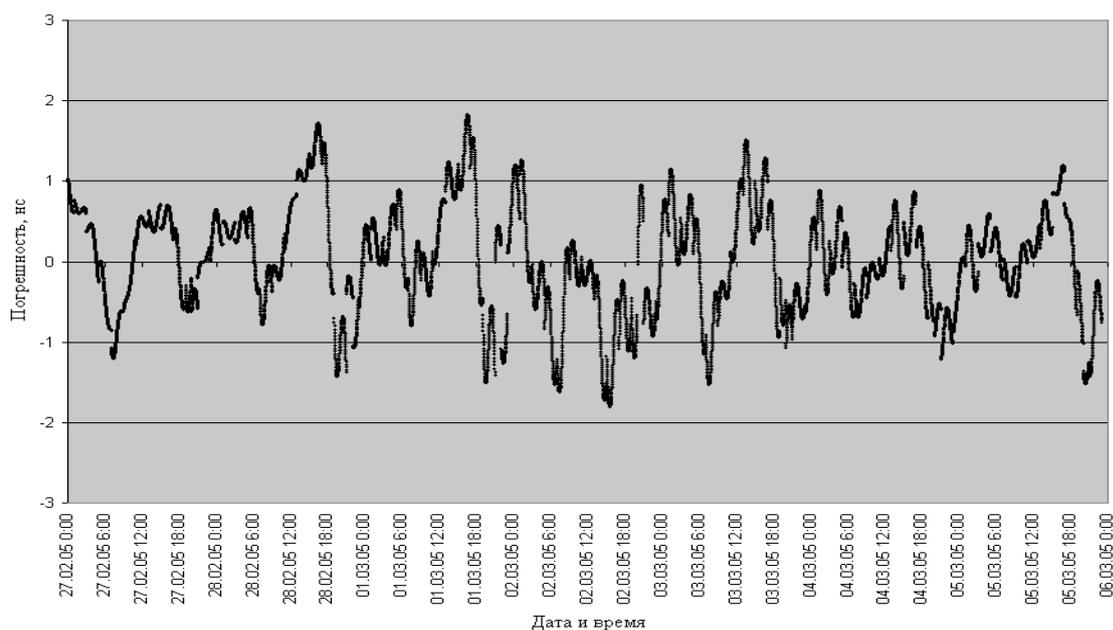


Рис. 1. Характер зависимости погрешности вычисления геометрической дальности от времени с использованием оперативных эфемерид для КА PRN 23

Каждый набор передаваемых в составе навигационного сообщения НКА GPS данных действителен на протяжении двух часов, соответственно, каждые два часа для расчетов начинает использоваться новый набор эфемерид. На рис. 3 четко прослеживается скачок при переходе между используемыми наборами эфемеридных данных, что также негативно сказывается на точности решения временной задачи.

Характер изменения погрешности можно объяснить тем, что оперативные эфемериды рассчитываются на основании закона движения тела по околоземной орбите, и, в соответствии с этим законом, реальное движение НКА по орбите аппроксимируется степенным полиномом.

В результате рассмотрения графиков рассчитанной погрешности определения величины геометрической дальности можно сделать вывод, что случайной составляющей погрешности определения координат КА по сравнению с систематическими погрешностями можно пренебречь.

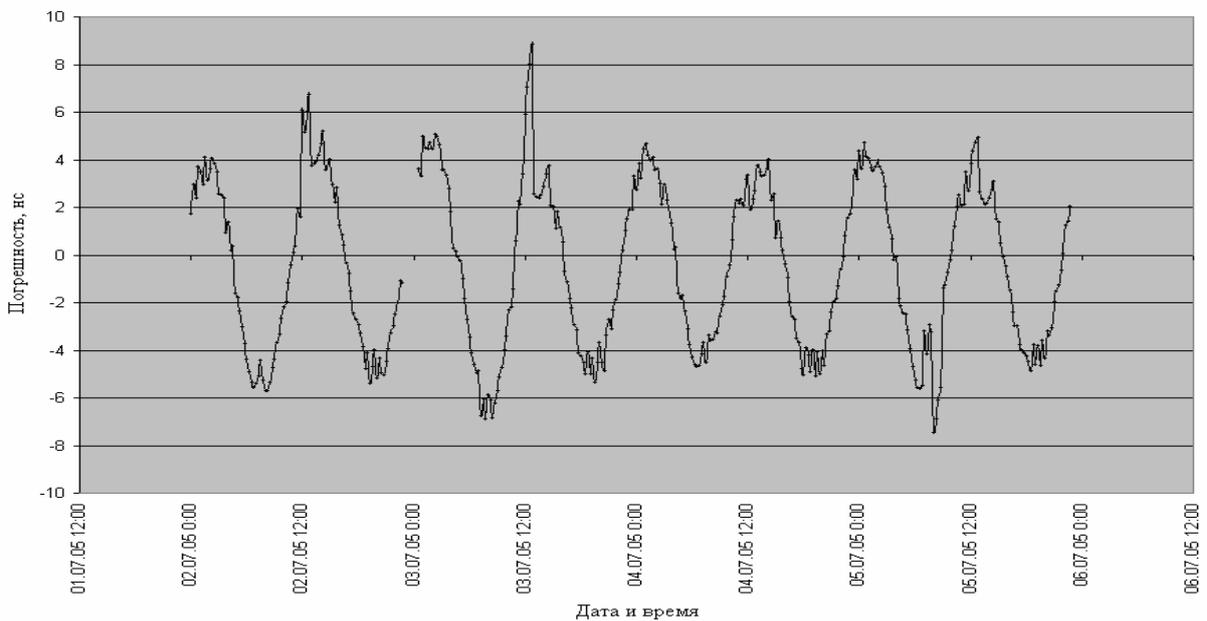


Рис. 2. Характер зависимости погрешности вычисления геометрической дальности от времени с использованием оперативных эфемерид для КА PRN 25

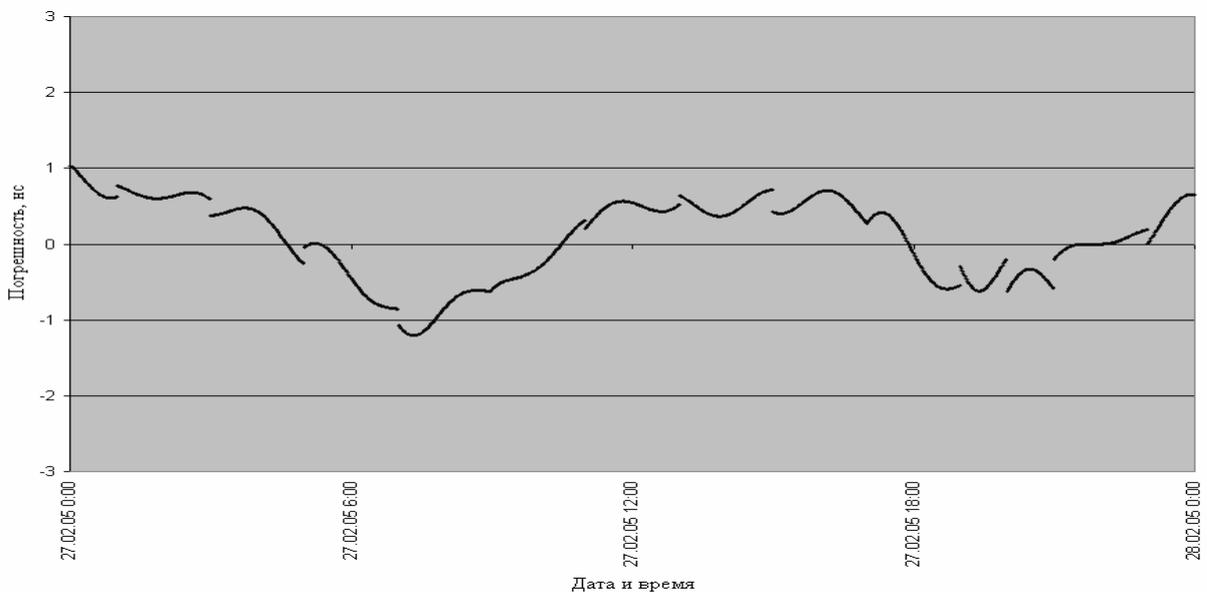


Рис. 3. Характер зависимости погрешности вычисления геометрической дальности с использованием оперативных эфемерид на интервале одних суток для КА PRN 23

Стандартный режим работы СРНС GPS времени предусматривает обеспечение точности передачи времени потребителю на уровне 40 нс. Таким образом, относительная погрешность, вносимая неточностями бортовых эфемерид, может составлять от двух до нескольких десятков процентов в различных случаях. В целях повышения точности временного обеспечения необходимо разработать методы учета или компенсации отмеченных систематических погрешностей.

Можно сделать вывод о том, что при отсутствии необходимости решать временную задачу в реальном времени (когда невозможно использование апостериорных эфемерид) целесообразно использовать апостериорные эфемеридные данные в силу их более высокой точности.

Заключение

На основании результатов работы разработанного пакета прикладных программ были оценены величины и характер изменения погрешностей данных, используемых в аппаратуре потребителя координатно-временного обеспечения СРНС.

Кроме того, были сделаны выводы о выборе исходных данных в различных ситуациях, а также рассмотрены перспективные направления дальнейшего исследования методов повышения точности расчетов на основании этих данных.

Литература

1. Глобальная спутниковая навигационная система ГЛОНАСС. / Под ред. Харисова В.Н., Перова А.И., Болдина В.А. М.: ИПРЖР, 1999.
2. Соловьев Ю.А. Системы спутниковой навигации. М.: Эко-Трейдз, 2000.
3. Генике А.А., Побединский Г.Г. Глобальные спутниковые системы определения местоположения и их применение в геодезии. М.: Картгеоцентр, 2004. С. 147–148, 328.
4. Гловацкая А.П. Методы и алгоритмы вычислительной математики. Учебное пособие для вузов. М.: Радио и связь, 1999.
5. Navstar GPS Space Segment / Navigation User Interfaces. Interface control document. ICD-GPS-200c, 25 SEP 1997.

ОПРЕДЕЛЕНИЕ ПОПРАВОК НА РАСПРОСТРАНЕНИЕ СВЕРХДЛИННЫХ ЭЛЕКТРОМАГНИТНЫХ ВОЛН

А.А. Скобелин

Научный руководитель – доктор технических наук, профессор В.Л. Ткалич

В статье приведена математическая модель расчета поправок на распространение сверхдлинных электромагнитных волн. На основе этой модели составлена программа расчета поправок и проанализированы полученные результаты.

Введение

Единство измерений времени в масштабах страны является необходимым условием для решения ряда важных научных и инженерных задач. Такое единство достигается с помощью системы единого измерения времени (СЕВ), представляющей собой совокупность разнесенных в пространстве часов и средств синхронизации, с помощью которых обеспечивается синхронность шкал времени. Государственная система единого времени и эталонных частот (ГСЕВЭЧ) была создана в 1978 г. ГСЕВЭЧ предназначена для обеспечения сигналами точного времени, эталонными частотами и информацией о параметрах вращения Земли комплексов и систем оборонного, народнохозяйственного и научного назначения, а также метрологического частотно-временного обеспечения Вооруженных сил и народного хозяйства страны [1].

Особое место в СЕВ занимают средства синхронизации, обеспечивающие потребителей точными сигналами времени и высокостабильными по частоте сигналами. На данный момент времени в составе системы существуют следующие средства синхронизации:

- радиоканалы на базе передающих сверхдлинноволновых (СДВ), длинноволновых (ДВ) и коротковолновых (КВ) радиостанций;
- телевизионные каналы;
- специальные радиоканалы с использованием отражений от метеорных потоков;
- радиоканалы с использованием искусственных спутников Земли (ИСЗ).

Так как синхронизация шкал времени потребителей со шкалой системы по сигналам СДВ и ДВ радиостанций основана на фазовых измерениях, основную погрешность синхронизации вносит задержка распространения радиосигналов в атмосфере. Эта задержка учитывается при синхронизации с помощью введения в расчет поправок на распространение.

В настоящее время поправки на распространение рассчитываются по математическим моделям (вариантам поправок), разработанным для конкретного региона, времени года и времени суток [2]. Недостаток этого метода – в том, что он имеет частный характер и может применяться только при соблюдении определенных условий.

Перспективным является расчет поправок на распространение в реальном масштабе времени по оперативной информации, которая может быть получена с помощью метеорологических служб, обсерваторий и т.д. Решение этой задачи позволит приблизиться к теоретической точности синхронизации шкал времени потребителей со шкалой времени системы.

В данной работе рассматривается метод определения поправок на распространение сверхдлинных электромагнитных волн.

Определение поправок на распространение

Сверхдлинные волны распространяются в волноводном канале, образованном поверхностью Земли и нижней границей ионосферы. Благодаря канализации энергии сигналы СДВ распространяются с малым затуханием на большие расстояния [3].

Электромагнитные свойства параметров стенок волновода определяют коэффициент возбуждения, скорость затухания и скорость распространения волн СДВ диапазона. Существенное влияние на распространение СДВ оказывает также магнитное поле Земли, в присутствии которого ионосфера становится анизотропной.

Электрические свойства подстилающей поверхности. Электрические свойства поверхности Земли (относительная диэлектрическая проницаемость ϵ_3 и удельная проводимость σ_3) имеют сложный закон распределения по Земному шару.

В таблице приведены значения ϵ_3 и σ_3 для различных видов поверхности Земли [3]. Из табл. видно, что эти параметры меняются в широких пределах, т.е. значительно изменяется и механизм распространения.

Вид поверхности	σ_3 , См/м	ϵ_3
Морская вода	3–5	80
Сильноувлажненная почва	10^{-2} – 3×10^{-2}	15–30
Сухая песчаная почва	10^{-3}	10–20
Каменистая почва	10^{-4} – 3×10^{-3}	5–10
Пресная вода	10^{-3} – 3×10^{-3}	80
Ледниковый лед (0°C, 20 кГц)	4×10^{-5}	10–20
Лед Гренландии	2×10^{-5}	4–8

Таблица. Электрические свойства поверхности Земли

Электрические параметры ионосферы. Ионосфера представляет собой разреженную многокомпонентную слабоионизированную плазму, в которой непрерывно протекают процессы ионизации и рекомбинации. Основным источником ионизации днем является коротковолновое излучение Солнца с длиной волны $\lambda < 103,8$ нм.

Ионосфера простирается вниз от высоты ≈ 1000 км до высоты 50 км и состоит из ряда слоев (F, E, D). Экспериментальные и теоретические исследования показывают, что основным слоем, определяющим распространение СДВ, является D-слой.

Под действием переменного электрического поля радиоволны в ионосфере вызывают вынужденные колебания электронов и ионов. Поскольку масса иона значительно больше массы электрона, то влиянием ионов пренебрегают, и ионосферу интерпретируют как упругую среду с эффективной электропроводностью (под эффективным параметром понимается параметр однородной среды, оказывающий на распространение волны такое же влияние, как и параметр реальной среды), рассчитываемой по формуле $\sigma_u = Ne^2\nu[m(\omega^2 + \nu^2)]$, где ω – круговая частота; N – концентрация электронов; e – заряд электрона; m – масса электрона; ν – число столкновений в секунду.

Для оценки вышеуказанных параметров ионосферы используются формулы $\omega = 2,5 \times 10^5 \exp(-h_0\beta)$, $N = N_0 \exp(-\eta_N h_0)$, $\nu = \nu_0$, $\beta = \eta_N + \eta_\nu$,

где η_N и η_ν – параметры модели, зависящие от освещенности и широты; h_0 – высота D-слоя ионосферы. Формулы приведены для случая, когда источник и приемник радиоволн находятся на поверхности Земли.

Элементы магнитного поля Земли. Основными элементами, которые описывают магнитное поле, являются проекции напряженности магнитного поля H_0 на оси выбранной системы координат.

Для описания геомагнитного поля в глобальном масштабе его представляют полем диполя, расположенного в центре Земли и направленного вдоль линии, соединяющей геомагнитные полюса. Составляющие геомагнитного поля в сферической системе координат (r, ϑ, λ) определяются выражением

$$V(r, \theta, \lambda) = a \sum_{n=1}^{\infty} \sum_{k=0}^n \left(\frac{a}{r}\right)^{n+1} P_n^R(\cos \theta) (g_n^R \cos k\lambda + h_n^R \sin k\lambda),$$

где a – радиус Земли; $P_n^R(\cos \theta)$ – присоединенный полином Лежандра степени n и порядка k ; g_n^R, h_n^R – коэффициенты Гаусса-Шмидта, характеризующие распределение магнитных масс относительно географической системы координат, жестко связанной с Землей; $r = a + z$ – расстояние от центра Земли до точки наблюдения; θ – географическое дополнение к широте точки наблюдения; λ – долгота точки наблюдения. Точность вычисления составляющих определяется степенью полинома n [3].

Математическая модель распространения сверхдлинных волн

Приведенные выше зависимости показывают, что моделирование распространения СДВ – достаточно сложный процесс. Однако с привлечением средств современной вычислительной техники эта задача может быть решена.

Особый практический интерес в подобных исследованиях представляет дополнительная фаза – величина, характеризующая отличие фазы волны, распространяющейся в реальном волноводе, от фазы волны, распространяющейся вдоль плоской бесконечно проводящей Земли. Эта величина, в основном, и определяет поправку на распространение сигнала. В данной работе используется модель, основанная на определении вертикальной составляющей электрического поля E_z в точке приема, расположенной на удалении d и на высоте z_R от источника, находящегося на высоте z_T . При использовании метода нормальных волн [3] эта величина записывается в виде:

$$E_z(d, z_R, z_T) = E_0 \exp(-ikd) \left[a \sin\left(\frac{d}{a}\right) \right]^{\frac{1}{2}} W(d, z_R, z_T), \quad (1)$$

где $E_0 = 300\sqrt{P}$; P – излучаемая мощность; $k = \frac{\omega}{c}$, ω – круговая частота; c – скорость света; a – радиус Земли; $W(d, z_R, z_T)$ – функция ослабления,

$$W(d, z_R, z_T) = \lambda^2 h^{-1} \sum_{m=1}^{\infty} \Lambda_m G_m(z_T) G_m(z_R) \exp\left[i\left(\nu_m d - \frac{\pi}{4}\right)\right],$$

Λ_m – коэффициент возбуждения m -й моды; G_m – высотные множители, описывающие изменение поля при подъеме приемника и передатчика над поверхностью Земли; ν_m – собственное число m -й моды, определяющее ее коэффициент затухания и фазовую скорость.

Аргумент функции ослабления определяет дополнительную фазу

$$\varphi_{\text{доп}}(d, z_T, z_R) = \arg W(d, z_T, z_R) = \arg E_z(d, z_T, z_R) - kd.$$

Поправка на распространение определяется выражением $P = -\varphi_{\text{доп}}(d, z_T, z_R)$ [4].

Для систем синхронизации наиболее распространен случай, когда передатчик и приемник находятся на поверхности Земли, т.е. $z_T = z_R = 0$; $G_m(0) = 1$.

Из выражения (1) видно, что данная модель учитывает условия освещенности (высоты ионосферы, которая изменяется в зависимости от времени суток), длины волны (частоты), коэффициентов возбуждения, затухания, фазовых скоростей и числа учитываемых мод. Кроме того, коэффициенты возбуждения, затухания, а также фазовые скорости являются величинами, зависящими от геомагнитного поля. Это позволяет считать данную модель достаточно точной.

Экспериментальным путем установлено, что в дневное время уже на небольшом удалении от передатчика коэффициент затухания второй моды значительно превышает

коэффициент затухания первой моды, что позволяет считать характер распространения СДВ одномодовым [4].

Таким образом, для поправки на распространение СДВ при одномодовом характере распространения справедливо выражение $p = -[(\frac{\omega c}{v_\sigma(\sigma) \times v_f(f)} - \frac{\omega}{c})d - \frac{\pi}{4} + \varphi_\Lambda(f)]$,

где $v_\sigma(\sigma) = \sum_{s=0}^j k_{\sigma s} \sigma^s$, $k_{\sigma s}$ – коэффициенты аппроксимации полиномиальной функцией

экспериментальной зависимости относительной фазовой скорости первой моды от проводимости подстилающей поверхности [5]; σ – проводимость подстилающей по-

верхности, См/м; $v_f(f) = \sum_{i=0}^n k_{vi} f^i$, k_{vi} – коэффициенты аппроксимации полиномиальной функцией экспериментальной зависимости относительной фазовой скорости пер-

вой моды от частоты [5]; $\varphi_\Lambda(f) = \sum_{l=0}^m k_{\Lambda l} f^l$, $k_{\Lambda l}$ – коэффициенты аппроксимации поли-

номиальной функцией экспериментальной зависимости фазы коэффициента возбуждения первой моды от частоты [5]; n, m, j – степени аппроксимирующих полиномов; f – частота, Гц; $\omega = 2\pi f$ – круговая частота, с⁻¹; c – скорость света, м/с; d – расстояние от источника до приемника, м.

Конечное выражение для поправки имеет вид:

$$p = -[(\frac{\omega c}{\sum_{s=0}^j k_{\sigma s} \sigma^s \times \sum_{i=0}^n k_{vi} f^i} - \frac{\omega}{c})d - \frac{\pi}{4} + \sum_{l=0}^m k_{\Lambda l} f^l]. \quad (2)$$

Приведенная модель с целью упрощения разработана для распространения СДВ в дневное время и не учитывает геомагнитное поле. Для определения поправки на трассе с резкими изменениями проводимости подстилающей поверхности необходимо разбить трассу на участки со слабо изменяющейся проводимостью и рассчитать поправку на каждом из них, после чего сложить полученные результаты, т.е. проинтегрировать выражение (2) по трассе распространения.

Результаты расчета

В работе рассчитывается поправка на распространение СДВ. Расчетная программа, основанная на приведенной выше математической модели, написана на языке C++.

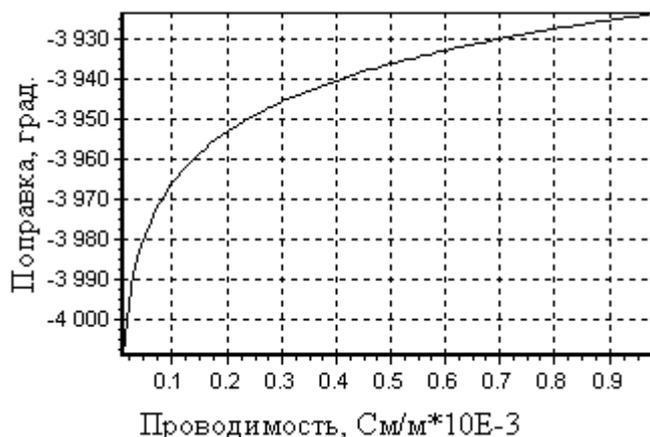


Рис. 1. График зависимости поправки от проводимости подстилающей поверхности ($f = 10$ кГц)

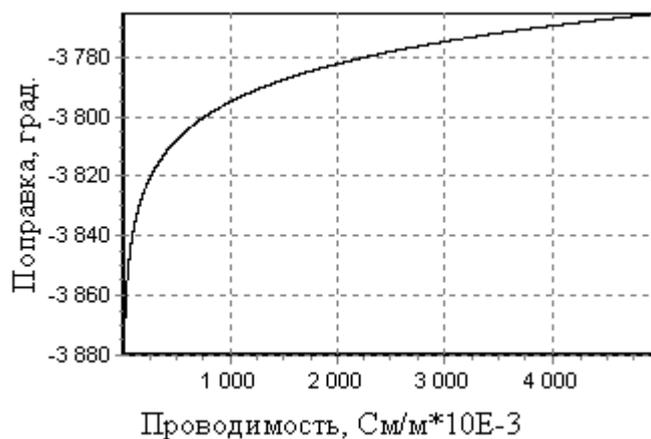


Рис. 2. График зависимости поправки от проводимости подстилающей поверхности ($f = 10$ кГц)

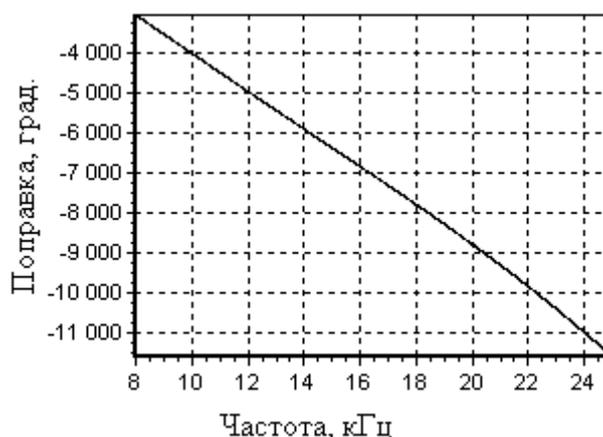


Рис. 3. График зависимости поправки от частоты ($\sigma = 10^{-5}$ См/м)

На рис. 1–3 в виде графиков представлены результаты расчета. Все расчеты проводились при длине трассы распространения $d = 3000$ км.

На основе полученных результатов можно сделать следующие выводы.

С увеличением проводимости поправка на распространение растет и асимптотически стремится к нулю, т.е. наблюдается идеализированный случай плоской бесконечно проводящей Земли. Этот результат говорит также о высокой стабильности морских трасс распространения, на которых, согласно графику рис. 2, наблюдается низкая скорость изменения поправки, что подтверждается экспериментальными данными [6].

С увеличением частоты поправка практически линейно возрастает, что позволяет с достаточной степенью достоверности прогнозировать ее значение.

Точность приведенной модели без учета влияния времени суток и геомагнитного поля оценивается степенями полиномов, аппроксимирующих экспериментальные кривые [7].

Заключение

В работе была разработана математическая модель определения поправок на распространение сверхдлинных электромагнитных волн.

На основе математической модели распространения СДВ была разработана программа, с помощью которой были рассчитаны и проанализированы основные зависимости изменения поправки на распространение от внешних факторов.

Соответствие между экспериментальными и расчетными данными позволяет считать данную модель удовлетворительной для прогнозирования распространения СДВ в зависимости от проводимости подстилающей поверхности и частоты сигналов.

Разработанная математическая модель позволяет, с учетом указанных допущений, которые будут устранены в будущем, приблизиться к теоретической точности синхронизации шкалы времени потребителя со шкалой времени системы.

Литература

1. Акулов И.И., Кузнецов В.П., Найдеров В.З., Притычин А.Г., Хомяков Э.Н., Шур Л.М. Основы теории единого измерения времени. М.: МО СССР, 1975. 384 с.
2. Morris P.B., Milton Y.Ch. OMEGA propagation corrections: Background and computational algorithm. Department of Transportation. Dec. Rep. ONSOD-01-74, 1974, 225 p.
3. Альперт Я.Л. Распространение электромагнитных волн и ионосфера. М.: Наука, 1972. 564 с.
4. Болошин С.Б., Семенов Г.А., Гузман А.С. и др.; Под ред. Олянюка П.В., Головушкина Г.В. Радионавигационные системы сверхдлинноволнового диапазона. М.: Радио и связь, 1985. 264 с.
5. Wait J.R., Spies K.P. Influence of finite ground conductivity on the propagation of VLF radio waves. // J. Res. NBS. 1965. V. 69D. № 10. P. 1359–1375.
6. Tolstoy A.I. OMEGA navigation system operations detail. // HQ-USCG (G-ONSOD/43). Washington. 1967. 49 p.
7. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы: Учебное пособие. М.: Наука. Гл. ред. физ.-мат. лит., 1987. 600 с.

АНАЛИЗ ЭКСПЛУАТАЦИОННЫХ СВОЙСТВ ПЛОСКИХ УПРУГИХ ЧУВСТВИТЕЛЬНЫХ ЭЛЕМЕНТОВ ЭЛЕКТРОМЕХАНИЧЕСКИХ ДАТЧИКОВ С РЕГУЛЯРНЫМ МИКРОРЕЛЬЕФОМ НА РАБОЧИХ ПОВЕРХНОСТЯХ

А.Ю. Буданова, Д.Н. Кокшаров, Д.А. Русин

Научный руководитель – доктор технических наук, профессор В.Л. Ткалич

В статье проведена качественная оценка параметров упругих чувствительных элементов датчиков с регулярным микрорельефом на рабочих поверхностях.

Способ вибронакатывания позволяет получить оптимальную регулярный микрорельеф (РМР) и упрочненный приповерхностный слой материала упругих чувствительных элементов (УЧЭ), что ведет к совокупному улучшению эксплуатационных свойств этих элементов. Необходимость создания РМР на поверхности УЧЭ возникла в связи с тем, что ряд задач, связанных с обеспечением целого комплекса эксплуатационных свойств, не мог быть решен традиционными способами обработки (прокатом, волочением, штамповкой, вытяжкой, полировкой, и т.д.). Эти способы обработки создают не только хаотический и трудноуправляемый микрорельеф, но и весьма неблагоприятный, далекий от оптимального в отношении формы и расположения неровностей (чрезмерно заостренные на выступах и во впадинах неровности с недопустимо малым шагом). Основным недостатком традиционных способов обработки являются неоднородность пластической деформации и примитивность кинематики [1].

Предлагаемый способ изготовления УЧЭ дает возможность не только аналитически рассчитывать значения всех параметров микрогеометрии как функции режимов вибронакатывания, но и устанавливать оптимальный вид РМР и значения его высотных, шаговых и площадных параметров. Это обеспечивает требуемые эксплуатационные свойства УЧЭ. В табл. 1 представлена взаимосвязь эксплуатационных свойств УЧЭ с основными параметрами РМР их рабочих поверхностей.

Комплексное исследование циклопрочности, износо- и коррозионной стойкости УЧЭ проводилось в агрессивных газовых и жидких средах при различных режимах нагружений на базе лабораторных установок кафедры деталей и машин БГТУ.

Свойства	Параметры РМР				Примечания
	R/h	N/Sk	$\gamma, \beta/\alpha$	Tr/Fн	
1	2	3	4	5	6
Повышение качества УЧЭ	+	+	+	+	
Повышение ресурса работы (износостойкости)	+	+			Повышение износостойкости электроконтактов и ламелей в 2 раза
Повышение сопротивления усталости	+	+			Повышение сопротивления усталости в 1.2–2.5 раза
Улучшение прирабатываемости контактов		+		+	

Таблица 1. Взаимосвязь эксплуатационных свойств УЧЭ с параметрами РМР

1	2	3	4	5	6
Стабилизация толщины слоя смазки (ртутносмазываемый КС)		+	+		
Повышение гидроплотности		+			Повышение гидроплотности резиномеханических уплотнителей сильфонных клапанов в 2–2.5 раза
Улучшение качества спая УЧЭ со стеклом	+		+		Отсутствие растрескивания стекла в зоне спая баллона КС
Увеличение коррозионной стойкости	+				В том числе повышение сопротивления фреттинг-коррозии
Повышение темпа очистки поверхности УЧЭ при вакуумировании	+			+	Темп очистки увеличивается в 4–5 раз
Уменьшение механического гистерезиса		+			Актуально для мембранных устройств
Уменьшение толщины покрытия	+	+			Для ламелей, покрытых серебром, улучшения в 1.8–2 раза
Устранение эффекта залипания КС за счет микросварки	+				Повышение токов коммутации МК
Увеличение адгезионных свойств				+	
Уменьшение веса УЧЭ за счет снижения толщины стенки (при сохранении требуемой жесткости)				+	Актуально для сильфонов и витых пружин
Уменьшение и стабилизация электрического сопротивления контактного перехода	+		+		В герконах в 1.5–2 раза
Увеличение плавности хода		+	+		Актуально для скользящих контактов
Увеличение теплопроводности				+	Увеличение рассеиваемой мощности в коммутац.устройствах

Таблица 1. Продолжение

Результаты экспериментальных исследований позволили установить вид зависимостей между рядом эксплуатационных свойств и параметрами РМР. Некоторые из этих зависимостей представлены на рис. 2. Из графиков следует, что все эти зависимости имеют нелинейный характер с явно выраженными оптимумами. Поэтому при проведении оптимизации микрогеометрии УЧЭ необходимо, задавшись критериями оптимизации, принимать компромиссные решения по выбору конкретных величин основных параметров РМР, а, соответственно, и режимов процесса вибронакатки. Было проведено подробное исследование эксплуатационных свойств плоских пружин с регулярным микрорельефом рабочих поверхностей [2].

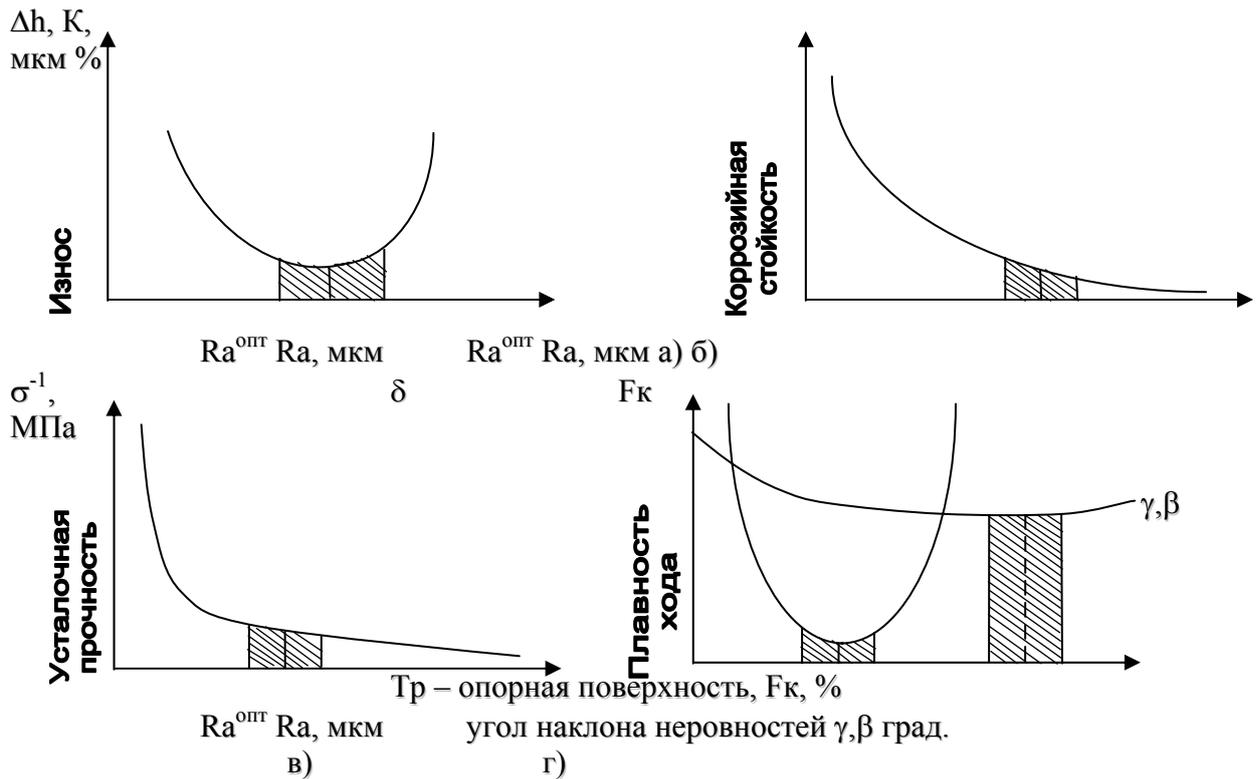


Рис. 2. Экспериментальные зависимости ряда эксплуатационных свойств УЧЭ от параметров РМР

С помощью способа вибронакатывания на поверхности КС геркона может быть создан РМР, существенно отличающийся по всем характеристикам от всех создававшихся до сих пор микрорельефов при обработке КС давлением (прокат, волочение, штамповка). Проведенные экспериментальные исследования выявили [2], что РМР рабочей поверхности КС обеспечивает улучшение многих эксплуатационных характеристик МК:

- (1) повышается качество и надежность МК;
- (2) повышается срок службы КС (износостойкость контактного покрытия) и геркона в целом на 50–60%;
- (3) исключается залипание (микросварка) контактов;
- (4) исключается необходимость проведения электротренировки герконов, целью которой является удаление игл на поверхности КС и их прирабатывание при наличии нерегулярного микрорельефа;
- (5) повышается качество спая КС с герметизированным баллоном;
- (6) повышается усталостная прочность;
- (7) повышается коррозионная стойкость;
- (8) улучшается герметичность спая КС со стеклом герметизированного баллона при длительном хранении, транспортировке и эксплуатации геркона;
- (9) улучшается рассеивание мощности с поверхности нагреваемых коммутируемым током КС, что является большой проблемой в случае КС с нерегулярным микрорельефом;
- (10) уменьшается в 1.5–2 раза и стабилизируется электрическое сопротивление КС;
- (11) исчезает необходимость в термообработке КС;
- (12) экономится металл, в том числе драгоценный, так как можно обходиться без покрытия на КС;

- (13) уменьшается дребезг и динамический шум КС;
- (14) повышается сопротивляемость фреттинг-коррозии.

Помимо этого, способ вибронакатывания является экономичным, так как сокращается технологический цикл обработки (требуемые габаритные размеры КС с РМР поверхности получаются за одну операцию прокатки в валках), не требуется проведения отжига КС после прокатки их в валках.

Проведем краткое рассмотрение основных эксплуатационных свойств КС с регулярным микрорельефом, подробно описанных в работе [2].

Если КС имеет регулярную микрогеометрию поверхности, то контакт осуществляется каждый раз через постоянное число «пятен контакта», которое определяется числом выступов поверхности КС в зоне перекрытия КС. При этом переходное сопротивление КС будет являться стабильной величиной, значение которой может быть меньше величины переходного сопротивления при нерегулярном микрорельефе в 1,5–2 раза. На рис. 3 приведены тепловая и электрическая модели контакта поверхностей с РМР. Если поверхности КС приработаны, то вершины выступов плоские, однако величина, на которую они усечены, у всех выступов одинакова, что обеспечивает стабильность электрического сопротивления геркона. При образовании РМР поверхности КС полностью исключаются залипания, вызываемые микросваркой КС, так как даже в случае приработки контактных поверхностей геркона стирание вершин выступов неровностей происходит на одинаковую величину, и на поверхности КС не образуется зон с локальным разогревом, приводящим к микросварке.

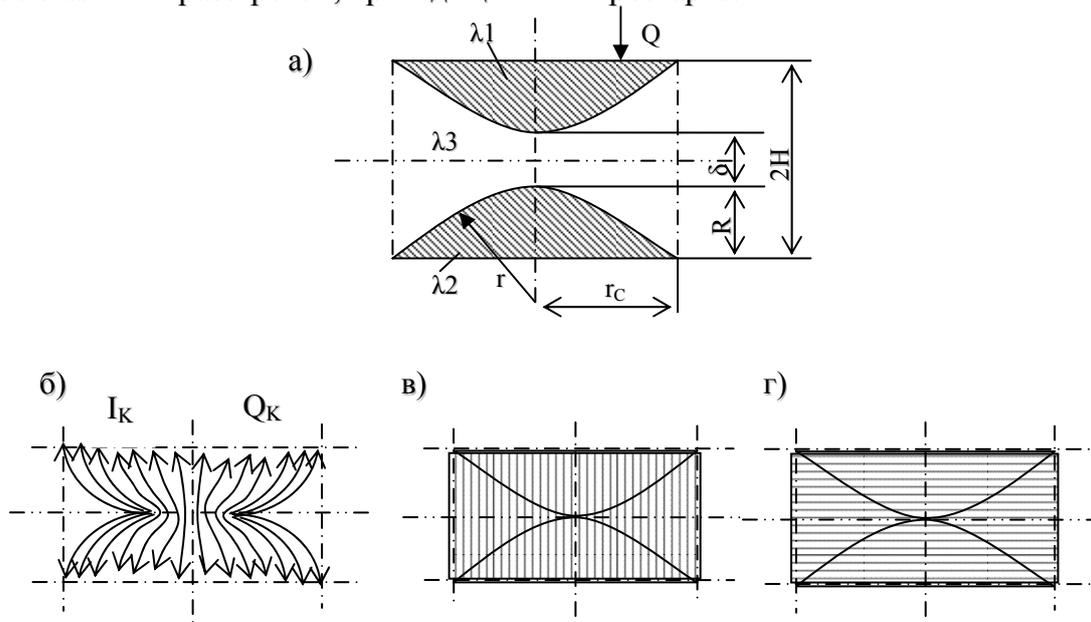


Рис. 3. Тепловая и электрическая модели контакта поверхностей КС с РМР: а – между КС имеется воздушный зазор; б – реальное расположение линий тока и тепловых потоков в элементарной ячейке контакта; в – дробление модели системой адиабатических поверхностей; г – дробление модели системой изотермических поверхностей

Вибронакатывание не влияет на магнитные свойства материала КС [2], поскольку его действие (упрочнение и наклеп) сказывается на тонком поверхностном слое КС и быстро убывает с удалением от поверхности при правильно подобранной величине контактного давления деформирующего инструмента. Поэтому доменная структура внутренних слоев КС не нарушается и магнитные свойства материала КС не меняются.

Повышение срока службы КС геркона при РМР рабочей поверхности КС достигается за счет увеличения их износостойкости. В процессе работы геркона про-

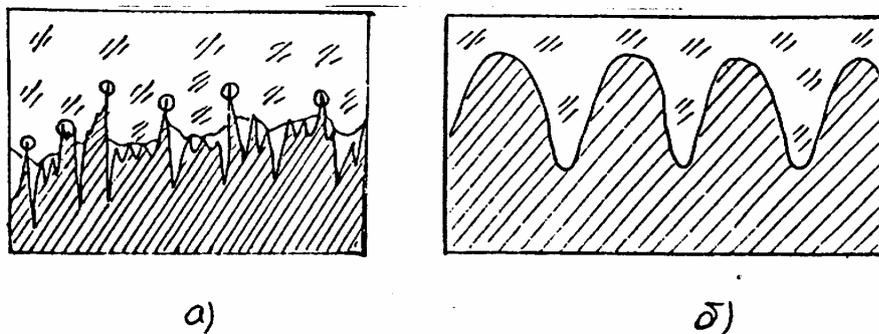
исходит непрерывный износ контактного покрытия, а также изменение фактической площади контактирования под воздействием механических, электрических, электрохимических и других явлений [2]. При отсутствии покрытий изнашивается материал КС. Микрогеометрия контактного покрытия полностью повторяет характер и вид микрорельефа исходной поверхности.

Произведенные расчеты показали, что чисто механический износ контактного покрытия геркона также оказывает значительное влияние на длительность срока их службы, и у КС с РМР ресурс работы контактного покрытия на 50–60% превышает ресурс работы контактного покрытия с нерегулярным микрорельефом [2].

В большинстве современных герконов спай формируется из стекла С95-3 и пермаллоя 50Н-ВИ или 52Н-ВИ. Поэтому для оценки качества спае необходимо иметь данные о технологических свойствах стекла и пермаллоя; краевом угле смачивания, поверхностном натяжении расплавленного стекла, термическом расширении стекла и пермаллоя, температурной зависимости вязкости стекла и о влиянии микрогеометрии поверхности КС на напряжения, возникающие в спае.

Одной из основных предпосылок достижения прочного сцепления стекла с пермаллоем, как показано в [2], является хорошее смачивание расплавом стекла поверхности пермаллоя, которое во многом зависит и от характера микрогеометрии поверхности КС (рис. 4).

Напряжения в спае можно разделить на два класса; временные и остаточные. В свою очередь, временные напряжения включают в себя термоупругие напряжения, напряжения из-за разности хода кривых КТР, напряжения, возникающие вследствие механических воздействий в герконе, обусловленные электромагнитными процессами. К остаточным напряжениям относятся напряжения из-за несогласованности спае, закалочные напряжения и напряжения из-за неоднородности стекла и наличия базовых включений в месте спае стекла с пермаллоем [2].



О - очаги с повышенным уровнем термомеханических напряжений

Рис. 4. Контакт расплава стекла с поверхностью КС: а – при нерегулярной микрогеометрии; б – при РМР

Чтобы спай геркона обладал высокой надежностью, необходимо, чтобы напряжения не превышали предела прочности стекла на растяжение. Уровень этих напряжений существенно зависит от микрогеометрии поверхности, так как все пики и впадины РМР поверхности КС в местах спае являются источниками этих напряжений. У РМР поверхности, создаваемого вибронакатыванием, отсутствуют острые вершины и впадины (он характеризуется пологими выступами, одинаковыми по форме и размерам), а, следовательно, не создается больших концентраций локальных напряжений в зоне спае. Прочность связи стекла с пермаллоем обуславливается силами химического и электростатического взаимодействия. Так, адгезионные свойства стекла зависят от краевого угла смачивания, который, в свою очередь, связан с содержанием в

стекле окислов BaO, CaO, MgO, Fe₂O₃, FeO, K₂O, Li₂O. Эти окислы улучшают смачивание пермаллоя и способствуют растеканию жидкого стекла [3, 4]. Герметичность геркона и долговечность спая, как показали проведенные исследования [5], на 60–70% зависят от микрогеометрии поверхности КС, от формы и расположения микронеровностей поверхности КС. КС с вибронакатанной поверхностью при ПРМР надежно обеспечивают хорошее качество спая геркона.

На сопротивление коррозии КС геркона с вибронакатанной поверхностью воздействуют два конкурирующих фактора. С одной стороны, происходит сглаживание неровностей исходной поверхности и образование микрорельефа со впадинами, радиус которым значительно больше, чем у обработанной любым способом финишной доводки поверхности, что определяет меньшую концентрацию в них продуктов, вызывающих коррозию; в результате коррозионная стойкость повышается. С другой стороны, неоднородный характер пластической деформации приводит к возникновению разности потенциалов между неодинаково деформированными кристаллами, т.е. к образованию множества гальванических пар, являющих причиной коррозии, и коррозионная стойкость снижается.

На рис. 5 показан типовой график зависимости коррозионной стойкости от усилия обкатывания и, соответственно, наклепа. Как видно из рис. 5, коррозионная стойкость повышается с увеличением усилия вибронакатывания и, соответственно, степени сглаживания неровностей исходной поверхности лишь до определенного значения (точки А). Эта оптимальная в отношении коррозионной стойкости величина усилия соответствует при вибронакатывании переходу от частично РМР с пересекающимися канавками (вид три по ГОСТ 24773-81) к ПРМР (четвертый вид), когда образуется новый микрорельеф, но наклеп еще относительно мал.

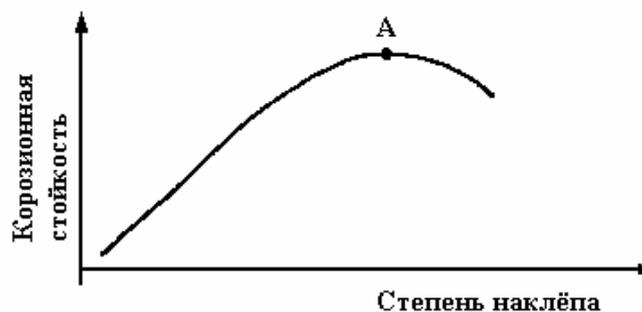


Рис. 5. Типовой график зависимости коррозионной стойкости от наклепа при вибронакатывании

При этом положительное влияние сглаживания и образования микрорельефа поверхности с пологими, большого радиуса впадинами значительно весомее, чем негативное влияние наклепа. Дальнейшее увеличение усилия вибронакатывания, даже в том случае, если шероховатость несколько улучшается, приводит к резкому увеличению степени наклепа и снижению коррозионной стойкости.

Заключение

Таким образом, как показали результаты исследований, подтвержденные практикой эксплуатации деталей, можно не только избежать снижения коррозионной стойкости, но даже заметно ее повысить, а тем самым повысить и срок службы устройства. При использовании ПРМР на поверхности КС надежно обеспечивается качество спая геркона и стекла.

Применение метода виброобработки на поверхности УЧЭ приводит к комплексному улучшению качества коммутационных устройств электромеханических

датчиков систем управления. Кроме того, данный метод является экономичным, так как сокращается технологический цикл обработки и не требуется проведения отжига КС после прокатки их в валках.

Литература

1. Шнейдер Ю.Г. Регуляризация мекрогеометрии поверхностей. Л.: Знание, РСФСР, ЛО ЛДНТП, 1991. 240 с.
2. Ткалич В.Л. Разработка и исследование методов повышения надёжности герконов и иреле на их основе. //Автореферат дис. на соискание степендии кандидата технических наук. ЛИТМО, Санкт-Петербург, 1994. 20 с.
3. Ткалич В.Л., Беккер Я.М., Фролкова Е.Г., Потапов А.И. Магнитоуправляемый контакт. Патент РФ № 2020624, Бюл. № 18, 30.09.94, кл. Н 01 Н 1/66.
4. Ткалич В.Л., Беккер Я.М., Фролкова Е.Г., Шнейдер Ю.Г. Потапов А.И. Магнитоуправляемый контакт. Патент РФ № 2024981, Бюл. № 23, 15.12.94, кл. Н 01 Н 1/66.
5. Иванова В.А. Внутреведение. ЯМР-томография. Ленинград: Знание,1989. 30 с.
6. Шоффа В.Н. Создание теории электрических аппаратов с магнитоуправляемыми контактами: Автореферат дис. на соискание ученой степени доктора технических наук. М.: МЭИ, 1983. 38 с.

МЕТОДЫ ПОВЫШЕНИЯ НАДЕЖНОСТИ ИНТЕГРАЛЬНЫХ СХЕМ НА ОСНОВЕ КОМПОЗИТНЫХ МАТЕРИАЛОВ

А.А. Дегтярев, Е.В. Овечкина, А.В. Смелова

Научный руководитель – доктор технических наук, профессор В.Л. Ткалич

В работе рассмотрены основные причины ухудшения параметров интегральных схем и методы повышения надежности схем, как физические, так и технологические.

Введение

Диагностирование интегральных схем, т.е. выявление причин, ответственных за ухудшение параметров, несомненно, является ключевым моментом в повышении надежности работы электронной аппаратуры. Естественно, что прямые методы и средства диагностирования, опирающиеся на совокупность исследований параметров самой интегральной схемы (ИС), более привлекательны, поскольку непосредственно позволяют судить о надежности этой схемы. Однако далеко не все причины могут быть однозначно установлены на базе только прямых методов, поэтому не менее важным является поиск, развитие и совершенствование косвенных методов с привлечением дополнительных средств и методов диагностирования материалов, технологических процессов и операций [1, 2].

Причины ухудшения параметров ИС

При изготовлении, испытаниях и эксплуатации ИС подвергаются различным температурным обработкам, влияющим на металлическую структуру, механические и электрические свойства микросварных соединений внутренних выводов. Приборы повышенной надежности рассчитаны на длительную эксплуатацию при температуре 125°C. При оценке надежности ИС необходимо учитывать влияние условий изготовления и эксплуатации на структуру и свойства микроконтактов, особенно в случае биметаллических пар. Результаты исследований показывают, что основной причиной отказов микросоединений является образование микропустот на границе раздела интерметаллических фаз и чистых металлов вследствие эффекта Киркендалла. Рекомендации по улучшению надежности микросварных соединений сводятся к повышению чистоты материала покрытия или легированию его микропроволоки примесями, тормозящими рост интерметаллических фаз. Выбирая соотношения толщин контактирующих материалов, можно также замедлить образование фаз из-за недостатка одного из материалов. Слой такой фазы уменьшает скопление пор и микротрещин в сварных соединениях. Наилучшими характеристиками и термостабильностью обладают соединения алюминиевой проволоки с пленками алюминия и никеля. Данные соединения незначительно претерпевают изменения в температурном диапазоне технологических обработок при сборке, а также при работе приборов [2, 3].

Методы повышения надежности ИС

Надежность является основной потребительской характеристикой любого прибора. Тем не менее, до настоящего момента потребитель не в состоянии определить, сколько будет работать конкретный прибор в соответствующей аппаратуре, пользуясь для необходимых оценок лишь вероятностными характеристиками, даваемыми изготовителем или определенными самостоятельно. Это предопределяет необходимость проведения исследований, целью которых, в конечном счете, является разработка методов

повышения надежности, основанных на изучении кинетики эволюции дефектной структуры в ИС.

Для формирования равновесной дефектной структуры в кристалле необходимо, чтобы время, затрачиваемое системой на переход из одного состояния в другое, было существенно больше времени обегания – характеристического времени, которое необходимо точечному дефекту, чтобы побывать во всех узлах решетки. Рассматривая идеальный кристалл, можно оценить, например, зависимость необходимой скорости охлаждения от температуры проведения технологического процесса.

Основной причиной эволюции дефектной структуры ИС являются процессы релаксации внутренних механических напряжений. Внутренние механические напряжения вводятся в ИС на технологических операциях как результат нагрева, легирования, нанесения и травления различных функциональных слоев. В поле механических напряжений резко ускоряется подвижность точечных дефектов, изменяется их равновесная концентрация, и дефектная структура эволюционирует в направлении формирования макроскопических дефектов, наличие которых в ИС резко ухудшает ее надежность. При анализе процессов, протекающих в дефектной структуре ИС при различных условиях эксплуатации, был сделан вывод, что согласованность топологических размеров ИС с характеристиками параметров дефектной структуры позволит существенно повысить надежность характеристики ИС.

Проблема качества и надежности ИС связывается с успехами технологии их изготовления. Однако вопросы прогнозирования надежности с помощью неразрушающих методов, управления параметрами готовых изделий, изменения их свойств в нужном направлении практически не исследовались. Физические методы и процессы, обуславливающие повышение качества и надежности ИС, основаны на:

- теоретических и экспериментальных исследованиях механизмов токопереноса, установлении корреляций и прямых доказательств их взаимосвязи с качеством материала;
- изучении физических механизмов деградации;
- изучении физических процессов, позволяющих целенаправленно изменять свойства структур ИС;
- применении физических методов неразрушающего контроля для отбраковки потенциально ненадежных структур [1, 2].

Заключение

Знание причин возникновения отказов и ухудшения параметров ИС позволяет лучше отслеживать процесс отбраковки некачественных схем, а также разработать новые методы повышения надежности ИС, опираясь на существующие физические и технологические методы. При этом совершенно необходимо учитывать физические процессы, происходящие в ИС, особенно на основе композитных материалов.

Литература

1. Физические основы надежности и деградации полупроводниковых приборов. / Сборник статей Второй всесоюзной конференции. Кишинев, 1986.
2. Физические аспекты надежности, методы и средства диагностирования ИС. / Сборник статей Международной научно-технической конференции. Воронеж, 1998.
3. Шевченко В.Я. О терминологии: наночастицы, наносистемы, нанокompозиты, нанотехнологии. // Микросистемная техника. 2004. № 9.

ОПТИЧЕСКИЙ АВТОКОЛЛИМАЦИОННЫЙ МЕТОД ИССЛЕДОВАНИЯ ПАРАМЕТРОВ ПРОФИЛЯ УПРУГИХ ЧУВСТВИТЕЛЬНЫХ ЭЛЕМЕНТОВ ПРИ СТАТИЧЕСКОМ НАПРЯЖЕННО-ДЕФОРМИРОВАННОМ СОСТОЯНИИ

Д.Н. Кокшаров, Д.А. Русин

Научный руководитель – доктор технических наук, профессор В.Л. Ткалич

Разработана новая модификация оптико-геометрических методов измерения деформаций и перемещений – автоколлимационный оптический бесконтактный метод. Разработана методика для вычисления наружного радиуса гоффра.

Введение

Оптико-геометрические методы применяют для бесконтактного исследования напряженно-деформированного состояния на поверхности плоских и объемных деталей и их моделей. Общий классификационный признак оптико-геометрических методов – наличие оптической бесконтактной связи объекта исследования с регистратором искажения (деформации) базового геометрического элемента, нанесенного на объект, спроецированного или зеркально отраженного в нем. Каждый из оптико-геометрических методов эффективен для определенной группы задач, имеет, как правило, несколько разновидностей и областей применения, различается видом базовых элементов, типом регистрации и обработки [1, 2].

Существующие в настоящее время методы испытания сильфонов с целью проверки уровня надежности в большинстве своем относятся к разрушающим методам исследования (испытания повышенным давлением, испытания на циклопрочность, трещиностойкость и т.д.). Такой путь исследования крайне неэкономичен, так как стоимость сильфонного компенсатора среднего уровня качества превышает 150 \$. Требования экономики и безопасности делают актуальной задачу разработки неразрушающих методов исследования косвенных параметров надежности.

К наиболее распространенным оптико-геометрическим методам относятся следующие [1, 2]: делительных сеток, зеркально-оптический, муаровых полос, стереофотограмметрический, хрупких тензочувствительных покрытий.

В данной работе для исследования геометрии сильфонных и трубчатых УЧЭ разработана новая модификация оптико-геометрических методов измерения деформаций и перемещений – автоколлимационный оптический бесконтактный метод, предусматривающий создание специального отражающего слоя на рабочей поверхности исследуемых упругих чувствительных элементов (УЧЭ). Этот слой целесообразно наносить на регулярный микрорельеф (РМР) рабочей поверхности УЧЭ.

Основная часть

Экспериментально установлено [3–6], что на надежность сильфонных образцов существенное влияние оказывает такой нестандартизованный геометрический параметр как величина наружного радиуса закругления гофра. Взаимосвязь данного геометрического параметра с рядом стандартизованных в настоящее время геометрических характеристик определяется формулой:

$$r_{\text{вн}} = \frac{\Delta x E h^3 [(R_H - R_e - t) - t(R_H - R_e)]}{A Q R_H^2} - (R_H - R_e - L + 2l),$$

где $r_{\text{вн}}$ – внутренний радиус гофра сильфона; h – толщина стенки; t – шаг гофрировки; Q – осевая нагрузка; E – модуль упругости; Δ – линейное перемещение; R_H , R_e – наружный и внутренний диаметры гофрированной поверхности; L , $2l$ – длина гофриро-

ванной поверхности. Пример установки для измерения наружного радиуса гофра автоколлимационным методом изображен на рис. 1.

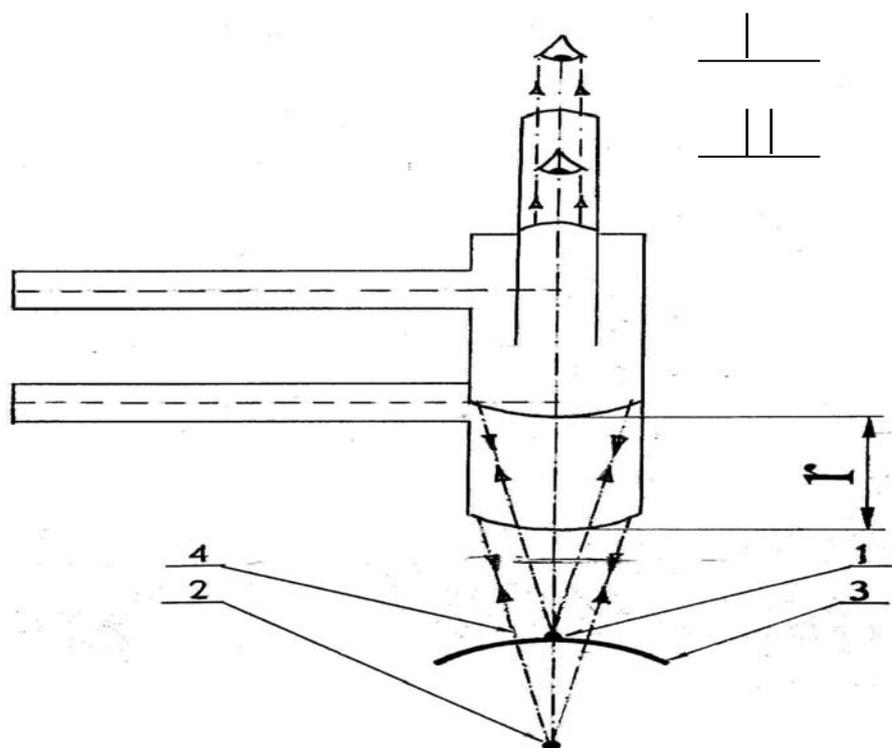


Рис. 1. Измерение наружного радиуса гофра автоколлимационным методом

Изображение объекта в установке при помощи отражающей поверхности проектируется в плоскость самого объекта. При этом лучи, падающие на отражающую поверхность и отраженные от нее, не являются коллинеарными.

Измерение наружного радиуса гофра сильфона выполняется по следующей методике.

1. Осуществляется наведение автоколлимационного микроскопа на вершину гофра (1) до получения четкого изображения тест-объекта.
2. При помощи микрометрического устройства производится перемещение автоколлимационного микроскопа в сторону сильфона до получения второго четкого изображения тест-объекта (положение II). Автоколлимационное изображение тест-объекта образуется вследствие того, что лучи (4) падают на поверхность гофра сильфона (3) под углом 90° и отражаются под тем же углом. Величина перемещения автоколлимационного микроскопа и будет равна значению наружного радиуса гофра. Для повышения точности эксперимента на рабочей поверхности УЧЭ необходимо создать покрытие с высокой отражающей способностью. Такое покрытие можно получить хромированием сильфонного образца и полировкой.
3. Оценка погрешности измерения проводится по формуле:

$$\sigma = \sqrt{(\sigma_1^2 + \sigma_2^2 + \sigma_3^2)}/3$$

где $\sigma_1 = 0.005$ мм – погрешность перемещения каретки предметного столика, $\sigma_2 = 0.005$ мм – погрешность снятия показаний шкалы каретки, σ_3 – погрешность наведения микроскопа, $\sigma_3 = \lambda \cdot (2A) = 0.00055 \cdot (2 \cdot 0.11) = 0.0025$ мм, $\lambda = 550$ нм – длина волны света; $A = 0.11$ – апертура микроскопа.

$$\sigma = \sqrt{(0.005^2 + 0.005^2 + 0.0025^2)}/3 = 0.012 \text{ мм.}$$

Заключение

В работе для исследования геометрии сифонных и трубчатых УЧЭ разработана новая модификация оптико-геометрических методов измерения деформаций и перемещений – автоколлимационный оптический бесконтактный метод. Разработанный метод является неразрушающим методом исследования косвенных параметров надежности и экономически выгодным.

Литература

1. Сухарев И.П. Экспериментальные методы исследования деформаций и прочности. М.: Машиностроение, 1987. 212 с.
2. Ткалич В.Л., Терентьев О.И., Маракулин Д.И. Методика оценки технического уровня и качества амортизирующих конструкций. / Депонировано в ВИНТИ 24.06.98, № 1944-В98. 7 с.
3. Ткалич В.Л., Землянская О.А., Гвоздев С.С. Проблемы расчета и проектирования сифонов. / Депонировано в ВИНТИ 06.03.98, № 640-В98. 9 с.
4. Ткалич В.Л., Гвоздев С.С., Суровый И.С. Оптический метод и устройство для контроля геометрических параметров сифонов. // Тез. докл. Международной конференции «Прикладная оптика-98», 16–18 дек. 1998. ГОИ им.С.И. Вавилова СПб: ГОИ им. С.И.Вавилова, 1998. С. 48.
5. Ткалич В.Л., Суровый И.С., Гвоздев С.С. Влияние изменения геометрических характеристик сифона на надежность при осевом нагружении. / Тез. докл. Международной молодежной конференции «XXУ Гагаринские чтения», Москва, 6–10 апреля 1999. 1 с.
6. Ткалич В.Л., Чугунов М.Ю., Рыбакова Н.А. Разработка и верификация методики определения геометрических параметров сифонов. / Тез. докл. IV Всероссийской научно-технической конференции «Методы и средства измерений физических величин». Нижегородский ГТУ, 16–17 июня 1999. 1 с.

ОСОБЕННОСТИ ФОРМИРОВАНИЯ РЕГУЛЯРНОГО МИКРОРЕЛЬЕФА НА РАБОЧИХ ПОВЕРХНОСТЯХ ПЛОСКИХ УПРУГИХ ЧУВСТВИТЕЛЬНЫХ ЭЛЕМЕНТОВ ЭЛЕКТРОМЕХАНИЧЕСКИХ ДАТЧИКОВ

А.Ю. Буданова, Д.Н. Кокшаров

Научный руководитель – доктор технических наук, профессор В.Л. Ткалич

Показана особенность формирования микрогеометрии поверхностного слоя плоских упругих чувствительных элементов, что ведет к совокупному улучшению эксплуатационных свойств этих устройств.

Введение

Технологии изготовления упругих элементов (УЭ) разнообразны и определяются конструкцией, назначением и материалом УЭ, а также техническими требованиями [1]. Производство современных конструкций определяется технологиями изготовления данных изделий. Поэтому важной задачей производства на сегодняшний день является этап изготовления УЭ. Ниже приведен ряд современных технологических процессов изготовления УЭ:

- технология прецизионного изготовления упругих элементов балочного типа методом электрохимико-механического полирования в деталях из труднообрабатываемых материалов;
- технология высокопрочного неразъемного соединения деталей из щелочных стекол с деталями из кремния в электростатическом поле высокого напряжения;
- технология прецизионной электроискровой обработки непрофилированным электродом – инструментом сложных закрытых профилей в деталях из труднообрабатываемых материалов;
- технология прецизионного размерного формообразования методом химического травления сложноконтурных сквозных пазов и упругих перемычек чувствительных элементов из кварцевого стекла;
- технология размерной обработки кремния для создания чувствительных элементов датчиков механических величин.

С целью повышения надежности, сокращения длительности приработки, замены дорогостоящих материалов на конструкционные элементы, сокращения трудоемкости изготовления деталей повышенной точности и долговечности и т.п. на поверхности контакт-деталей магнитных контактов (МК) предлагается наносить регулярный микро-рельеф (РМР). Наиболее совершенным методом образования РМР является метод вибрационного накатывания профессора Ю.Г. Шнейдера.

Способ образования регулярных микро-рельефов вибронакатыванием и его особенности

Регулярность формы, размеров и взаиморасположения неровностей поверхностей с РМР предоставляет возможность практически впервые аналитически их рассчитывать как функцию различных эксплуатационных свойств, т.е. перейти к расчетному нормированию геометрических параметров микро-рельефа поверхностей УЧЭ [1–3].

Принципиальная схема прокатки и параметры режима вибронакатывания цилиндрических прокатных валков приведены на рис. 1. Деформирующим элементом при вибронакатывании валков может быть шар, алмазный наконечник или валик с нанесенным на нем негативным (выпуклым) РМР [3].

Особенности регулярных микро-рельефов, создаваемых способом вибронакатывания, заключаются в следующем [3]:

- микрорельеф вибронакатанной поверхности образуется как след движения участка деформирующего элемента, контактирующего с обрабатываемой поверхностью;
- благоприятная практически для всех условий эксплуатации форма неровностей регулярных микрорельефов, характеризуемая пологой формой выступов и впадин с радиусами на 1–2 порядка большими, чем при других видах обработки, и соответственно, большей величиной опорной поверхности как исходной после вибронакатывания, так и в приработочный период;
- усложненная кинематика процесса вибронакатывания и большое число регулируемых параметров его режима позволяют весьма тонко и в больших пределах варьировать значения большого числа параметров создаваемого регулярного микрорельефа, в том числе таких, как число выступов и впадин на единице площади, фактическая площадь поверхности и длина профиля, радиусы закруглений впадин и выступов и т.д.;
- возможность независимого регулирования параметров режима вибронакатывания и отсутствие жесткой их связи с параметрами регулярного микрорельефа позволяют управлять формой и взаиморасположением его элементов;
- способ позволяет управлять такими параметрами поверхности, а следовательно, и контакта поверхностей, как фактическая его площадь и фактическая площадь контакта, и создавать высокочистые поверхности достаточной смачиваемости, а также исключать явления молекулярного сцепления, адгезии, фреттинг-коррозии;

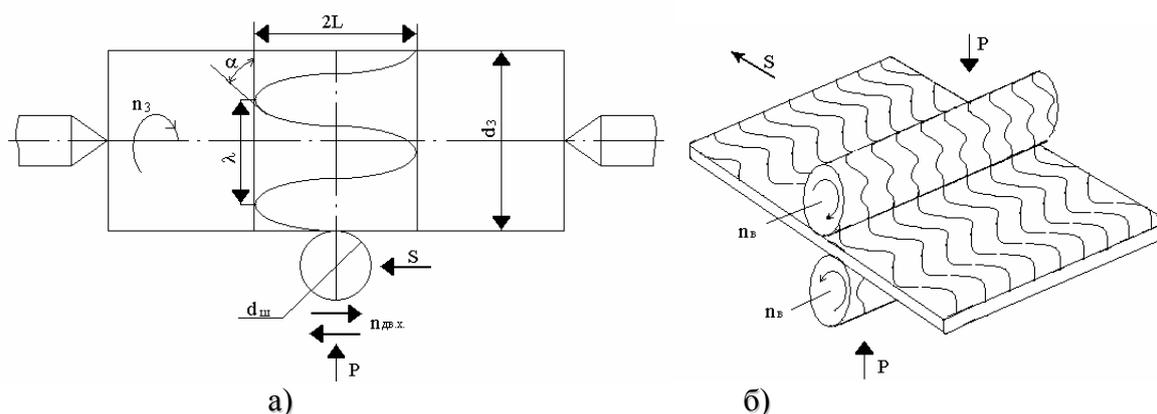


Рис. 1. Схемы вибронакатывания поверхности прокатных валков (а) и рабочей поверхности плоских УЧЭ (б): n_3 – число оборотов заготовки; $n_{дв.х.}$ – число осцилляций инструмента (шарика); $S_{пр}$ – продольная подача; $S_{поп}$ – поперечная подача; P – усилие вдавливания деформирующего инструмента; $n_в$ – число оборотов прокатных валков

- строгая кинематическая связь между параметрами режима процесса вибронакатывания и параметрами регулярного микрорельефа определяет возможность расчетного нормирования, технологического обеспечения и безаппаратного контроля поверхностей с регулярным микрорельефом;
- универсальность и возможность управления фактически всеми параметрами и характеристиками поверхностей с регулярным микрорельефом характеризуют способ вибронакатывания как инструмент для исследования и выявления взаимосвязей огромного многообразия эксплуатационных свойств поверхностей с их микрогеометрией.

Анализ сущности и особенностей регулярных микрорельефов свидетельствует о том, что практически все недостатки, свойственные шероховатым поверхностям деталей УЧЭ, при создании регулярных микрорельефов исключаются.

Регулярные микрорельефы могут быть разделены на две группы (рис. 2) [1,3]:

- с системой регулярно расположенных канавок, между которыми остаются участки исходной поверхности (рис. 2а, б, в);
- с полностью новым регулярным микрорельефом (ПРМР) поверхности (рис. 2г, д), ГОСТ 24775-81.

ПРМР могут иметь гексагональные и тетрагональные выпуклости. Для смены вида РМР рабочей поверхности УЧЭ необходимо менять сами валки, образуя на них различный РМР с помощью сферической виброголовки, за счет смены параметров режима вибронакатывания.

В работе [4] приведены профилограммы и фотографии поверхности КС с ПРМР для геркона типа КЭМ-1. Оптимальным для геркона является создание на поверхности УЧЭ ПРМР четвертого вида. Как показали расчеты, наилучшего качества спая УЧЭ с герметизированным баллоном можно добиться при формировании гексагонального выпуклого ПРМР поверхности УЧЭ [4].

Достоинством проведенного нормирования качества поверхности КС герконов на основании экспериментальных исследований является выявление одновременно с оптимальными значениями параметров микрорельефа условий и режима технологического его обеспечения.

В результате экспериментального исследования и расчетов были установлены следующие оптимальные значения параметров микрорельефа плоских КС герконов: радиус выступов 280 мкм, число выступов на 1 мм² – 10. Этим параметрам соответствуют следующие параметры режима вибронакатывания: при вибронакатывании валков шаром, в целях создания вогнутого ПРМР: число осцилляций инструмента 1400 1/мин, амплитуда осцилляций 2 мм, скорость поперечной подачи 120 мм/с, диаметр шара 4 мм; при вибронакатывании прокатным валком: скорость прокатки 20 м/мин, диаметр валков 40 мм, ширина бочки 60 мм.

Конструкторско-технологическое решение, найденное в результате этих исследований, защищено свидетельством [5].

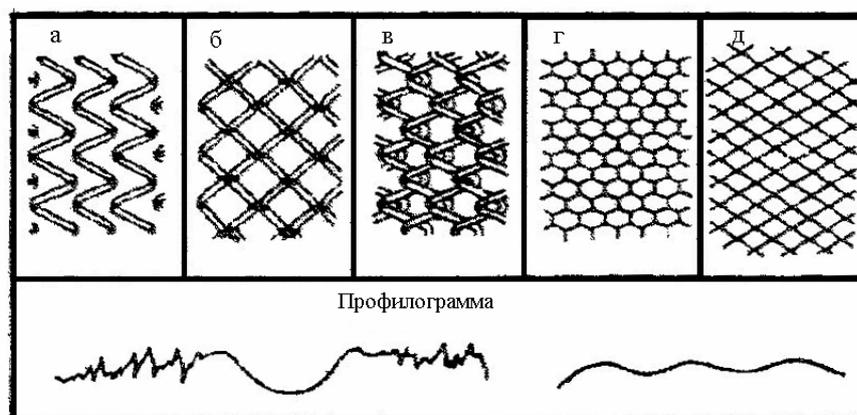


Рис. 2. Схемы и профилограммы поверхностей с системами канавок (а–в) и полностью новым регулярным микрорельефом (г, д)

Заключение

Способ вибронакатывания позволяет получить оптимальную регулярную микрогеометрию и упрочненный приповерхностный слой материала УЧЭ, что ведет к совокупному улучшению эксплуатационных свойств этих элементов. Кроме того, предложенная установка позволяет за одну технологическую операцию вибропрокатки в валках формировать макро- и микрогеометрию УЧЭ коммутационного устройства.

Литература

1. ГОСТ 24773-81. Поверхности с регулярным микрорельефом.
2. Ткалич В.Л., Михеева О.Д. Построение математических моделей присоединенных масс упругих элементов устройств автоматики. / Материалы Всероссийской научной конференции (Computer-Based Conference) «Методы и средства измерений», раздел «Мат. модели и численное моделирование измерительных приборов и датчиков», Нижний Новгород, 2000. 15 с.
3. Ткалич В.Л., Михеева О.Д., Железков В.В. Исследование динамики плоских УЧЭ устройств АСУ. / Материалы Всероссийской научной конференции (Computer-Based Conference) «Методы и средства измерений», раздел «Мат. модели и численное моделирование измерительных приборов и датчиков», Нижний Новгород, 2000. 14 с.
4. Ткалич В.Л. Разработка и исследование методов повышения надежности герконов и иреле на их основе. / Автореферат дис. на соискание степендии кандидата технических наук. ЛИТМО, Санкт-Петербург, 1994. 20 с.
5. Ткалич В.Л., Беккер Я.М., Фролкова Е.Г., Шнейдер Ю.Г. Потапов А.И. Магнитоуправляемый контакт. Положительное решение на выдачу патента РСФСР по заявке № 5054669/07, 27.07.93. 2 с.

РАЗРАБОТКА УСОВЕРШЕНСТВОВАННЫХ КОНСТРУКЦИЙ ГЕРКОНОВ С УЛУЧШЕННЫМИ ПОКАЗАТЕЛЯМИ КАЧЕСТВА

Д.Н. Кокшаров, Д.А. Русин

Научный руководитель – доктор технических наук, профессор В.Л. Ткалич

В статье рассматриваются два варианта конструкций магнитоуправляемых контактов. Предлагаемые авторами конструктивно-технологические решения позволяют устранить дребезг, повысить виброустойчивость, а, следовательно, и увеличить надёжность герконов.

Введение

Надёжность и качество герконов во многом определяется их конструктивными особенностями. Если контактными сердечниками (КС) придать S-образный изгиб у зоны спая с баллоном, на первой трети своей длины, то при замыкании таких сердечников за счет участков с S-образным профилем у места спая создается дополнительная вибрация, такая, что пластины в зоне контактирования будут скользить по поверхности друг друга, при этом оставаясь в замкнутом состоянии. Это исключает дребезг при замыкании контактов. Также улучшается надёжность работы в условиях вибронрузок и увеличивается ударопрочность конструкции. Дребезг при размыкании такой конструкции геркона устраняется за счет сил отталкивания электромагнитной природы.

Во втором случае свободные концы КС имеют форму, совпадающую с кривой прогиба основной гармоники свободно колеблющейся балки. Герконы с такой конструкцией будут обладать повышенной виброустойчивостью и отсутствием дребезга, если соотношение собственных частот колебаний КС равно двум.

Разработка конструкции герконов

В работе [1] проведен анализ динамических характеристик ряда разработанных конструкций МК с модифицированной микро- и макрогеометрией КС в целях повышения их качества. Эскизы этих МК приведены на рис. 1 и 2.

На рис. 3 представлен магнитоуправляемый контакт (МК) с развернутой на 90° (относительно участка заделки) плоскостью контактирования. Устранить дребезг и повысить виброустойчивость геркона можно за счет снижения жесткости колебательной системы в месте спая КС с баллоном и введения магнитов на свободных концах КС.

Рассмотрим конструкцию геркона (рис. 1) [2]. Он представляет собой герметизированный стеклянный баллон, в противоположных торцах которого запаяны (одним концом) КС, имеющие S-образный изгиб, у зоны спая с баллоном, на первой трети своей длины (считая от места спая), а два других свободных конца КС образуют зону перекрытия.

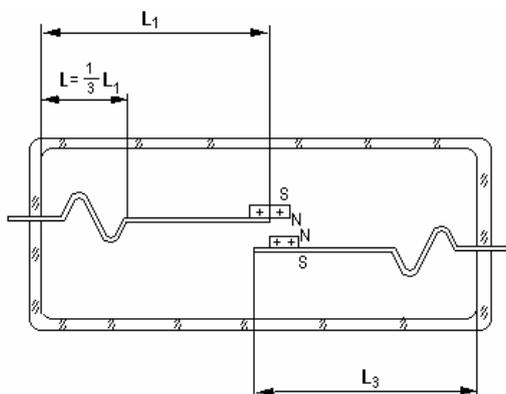


Рис. 1. Бездребезговая конструкция МК с S-образным изгибом у зоны спая КС и постоянными магнитами на свободных концах

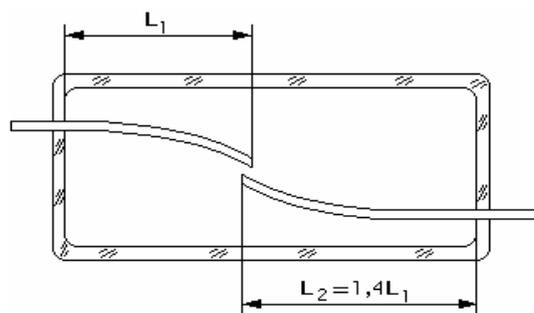


Рис. 2. Бездрезбовая конструкция асимметричного МК с КС в форме кривой прогиба основной гармоники свободно колеблющейся балки

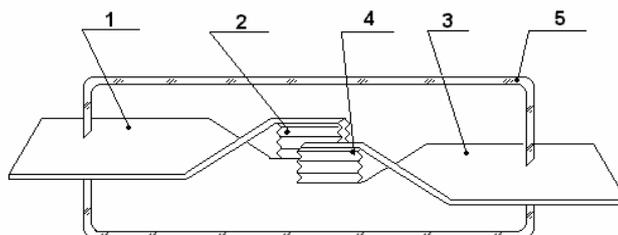


Рис. 3. Геркон

При замыкании таких сердечников, за счет участков с S-образным профилем у места спая, создается дополнительная вибрация, такая, что пластины в зоне контактирования будут скользить по поверхности друг друга, при этом оставаясь в замкнутом состоянии. Это исключает дребезг при замыкании контактов. Также улучшается надежность работы в условиях вибронрузок и увеличивается ударопрочность конструкции. КС в замкнутом состоянии будут совершать совместные изгибные колебания и надежно сохранять замкнутое состояние, когда это необходимо.

Положительным эффектом, достигаемым в такой конструкции, является и существенное снижение механических напряжений в зоне спая КС со стеклом, за счет снижения жесткости контактной системы. Это, в свою очередь, предохраняет стекло в месте спая от растрескивания и чрезмерных механических напряжений, а, следовательно, гарантирует герметичность и долговечность конструкции предлагаемого геркона. Дребезг при размыкании такой конструкции геркона устраняется за счет сил отталкивания электромагнитной природы, появление которых связано с взаимодействием магнитных потоков, расположенных на свободных концах КС. Действие этих сил усиливается при сближении одноименных полюсов постоянных магнитов. Таким образом, силы отталкивания, имеющиеся между постоянными магнитами, препятствуют повторному замыканию КС после размыкания геркона и надежно удерживают КС в разомкнутом состоянии при отсутствии стягивающей силы, создаваемой электромагнитным полем катушки управления.

Принимая конструктивные меры для максимального снижения доли колебаний второй главной формы, представляющей собой основную причину повторных замыканий контактов [1], а также выбирая КС таким образом, чтобы соотношение собственных частот колебаний равнялось четному числу, можно добиться того, чтобы дребезг контактов при коммутации был бы минимальным, а виброустойчивость – максимальной.

Рассмотрим конструкцию такого геркона. Разработанный геркон содержит: герметизированный баллон, два КС, расположенных в противоположных торцах баллона, свободный конец которых имеет форму, совпадающую с кривой прогиба основной гармоники свободно колеблющейся балки (т.е. имеем изгиб с направлением выпук-

лости, противоположным действием распределенной нагрузки), а соотношение длин КС равно 1,415 (см. рис. 2).

Разработанная конструкция геркона является асимметричной, позволяющей получить соотношение собственных частот указанных КС, равное четному числу. Рассмотрим, что дает такая конструкция в сравнении с аналогичной симметричной. При одинаковых длинах КС (когда остальные параметры их тоже одинаковы) их собственные частоты совпадают. Соударение КС происходит при одинаково направленных скоростях, что способствует минимальному рассеиванию энергии в процессе удара. Силы, действующие в зоне взаимодействия КС, в процессе соударения меняются по величине с частотой, равной частоте собственных колебаний КС. Частоты свободных ударных колебаний близки к частотам их свободных изгибных колебаний. При этом переходный колебательный процесс продолжителен.

При соударении КС, имеющих соотношение собственных частот, равное четному числу (т.е. в случае асимметричной конструкции), они начинают перемещаться в противоположном направлении. Ввиду того, что частота собственных колебаний одного КС в два раза (это оптимальное соотношение частот) меньше частоты свободных колебаний другого КС, фазы движения КС в моменты их соударения противоположны, в результате чего рассеивается значительная часть энергии в системе и, кроме того, возбуждаются высшие гармоники колебаний, что вызывает увеличение рассеивания энергии. После нескольких соударений КС переходят к колебательному движению в замкнутом состоянии. Переходный колебательный процесс быстро затухает.

В самом простом случае можно получить соотношение собственных частот, равное двум, при соотношении длин КС, равном 1,415.

Если придать КС изогнутую форму с направлением выпуклости, то можно максимально снизить долю колебаний второй главной формы, что ведет к снижению дребезга. При этом после достижения контакта между КС существенно уменьшаются дальнейшие изгибные деформации стержней, а, значит, и составляющая второй главной формы колебаний в упругой линии.

Таким образом, совокупность двух признаков – соотношение собственных частот колебаний КС, равное двум (или, что то же самое, соотношение длин КС, равное 1,415) и изогнутость формы свободного конца сердечника, с направлением выпуклости, противоположным действию распределенной нагрузки – позволяет получить конструкцию с повышенной виброустойчивостью и отсутствием дребезга, что существенно повышает надежность работы геркоонона.

Заключение

Предлагаемое в работе конструктивно-технологическое решение позволяет улучшить показатели надежности герконов: понизить уровень дребезга, устранить залипание контактов, снизить уровень динамических шумов, увеличить срок службы.

Литература

1. Ткалич В.Л. Разработка и исследование методов повышения надежности герконов и реле на их основе. / Автореферат дис. на соиск. степ. кандидата технических наук. Спб: ЛИТМО, 1994. 20 с.
2. Ткалич В.Л., Беккер Я.М., Фролкова Е.Г. Магнитоуправляемый контакт. Положительное решение на выдачу патента РФСР по заявке № 4917333, 30.08.93. 2 с.
3. Ткалич В.Л. Надежность магнитоуправляемых контактов в системах управления. Монография. СПб, 2000. 100 с.

ПАРАМЕТРЫ КАЧЕСТВА РАБОЧИХ ПОВЕРХНОСТЕЙ УПРУГИХ ЧУВСТВИТЕЛЬНЫХ ЭЛЕМЕНТОВ С РЕГУЛЯРНЫМ МИКРОРЕЛЬЕФОМ

Н.Н. Валентик, Н.В. Когай, Д.Н. Кокшаров

Научный руководитель – доктор технических наук, профессор В.Л. Ткалич

Рассмотрены аспекты взаимосвязи работы датчиков на основе магнитоуправляемых контактов и акселерометров с параметрами качества и надежности упругих чувствительных элементов с регулярным микрорельефом

Создание на поверхности УЧЭ регулярного микрорельефа того или иного класса, группы, вида, типа с различными значениями его параметров достигается за счет варьирования значений параметров режима вибронакатывания. Сложность кинематики процесса вибронакатывания и большое число регулируемых параметров режима, с одной стороны, облегчают решение этой задачи, обеспечивая возможность в больших пределах варьировать их сочетание по номенклатуре и численным значениям, с другой стороны, существенно усложняют технологическое обеспечение заданного конструктором РМР.

Так как РМР при вибронакатывании образован регулярно чередующимися одинаковыми по форме и размерам выступами и впадинами, представляется возможность выделения единичного элемента поверхности и его достаточно точного моделирования. Число выступов и впадин, приходящихся на единицу площади вибронакатанной поверхности, определяется строго аналитически как функция режима обработки. Поэтому возможен аналитический расчет таких важных параметров, как фактическая площадь поверхности, число пятен фактического контакта, фактическая площадь контакта.

Таким образом, поверхность с РМР может рассматриваться как набор идентичных элементарных ячеек, каждая из которых представляет собой сферический сегмент (рис. 1). Такая модель особенно эффективна при решении задач, связанных с теплопередачей, электропроводностью, с контактом твердых тел (контактных сердечников) с жидкой фазой (расплавом стекла). Эта модель позволяет использовать ПК при аналитических расчетах параметров РМР и параметров режима вибронакатывания. Переход на расчетный метод нормирования микрогеометрии является в настоящее время основным эффективным направлением нормирования микрогеометрии технических поверхностей.

Можно также пользоваться при нормировании микрогеометрии поверхности УЧЭ методом нормирования по прототипу, однако этот метод имеет много недостатков [3].

Поэтому основным, наиболее надежным методом нормирования качества поверхности УЧЭ является проведение экспериментальных исследований с целью выявления оптимального вида РМР и значений его параметров.

Решение проблемы качества рабочей поверхности УЧЭ представлено на рис. 2.

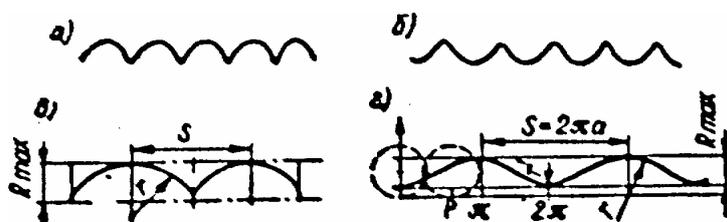


Рис. 1. Моделирование выступов поверхности УЧЭ с РМР: а, б, в – сферическим сегментом; г – укороченной трахоидой



Рис. 2. Проблема качества рабочей поверхности УЧЭ

Применение РМР обеспечивает [1, 2, 4]:

- (1) повышение качества УЧЭ;
- (2) повышение ресурса работы УЧЭ;
- (3) снижение потерь на трение (скользящие контакты);
- (4) исключение залипания КС МК вследствие микросварки;
- (5) исключение приработки контактов;
- (6) повышение контактной жесткости;
- (7) повышение сопротивления ползучести;
- (8) повышение усталостной прочности;
- (9) повышение коррозионной стойкости;
- (10) решение проблемы трения в вакууме;
- (11) повышение гидроплотности;
- (12) улучшение качества спая КС со стеклом баллона МК;
- (13) сохранение эксплуатационных свойств соединений при длительном хранении и транспортировке;
- (14) улучшение теплоотвода от рабочей поверхности УЧЭ;
- (15) ускорение очистки при вакуумировании;
- (16) стабилизацию электрического сопротивления контактного перехода КС МК;
- (17) экономию металлов, в том числе драгоценных;
- (18) исключение термообработки (термотренировки) УЧЭ;
- (19) экологическую чистоту технологического процесса обработки поверхности УЧЭ.

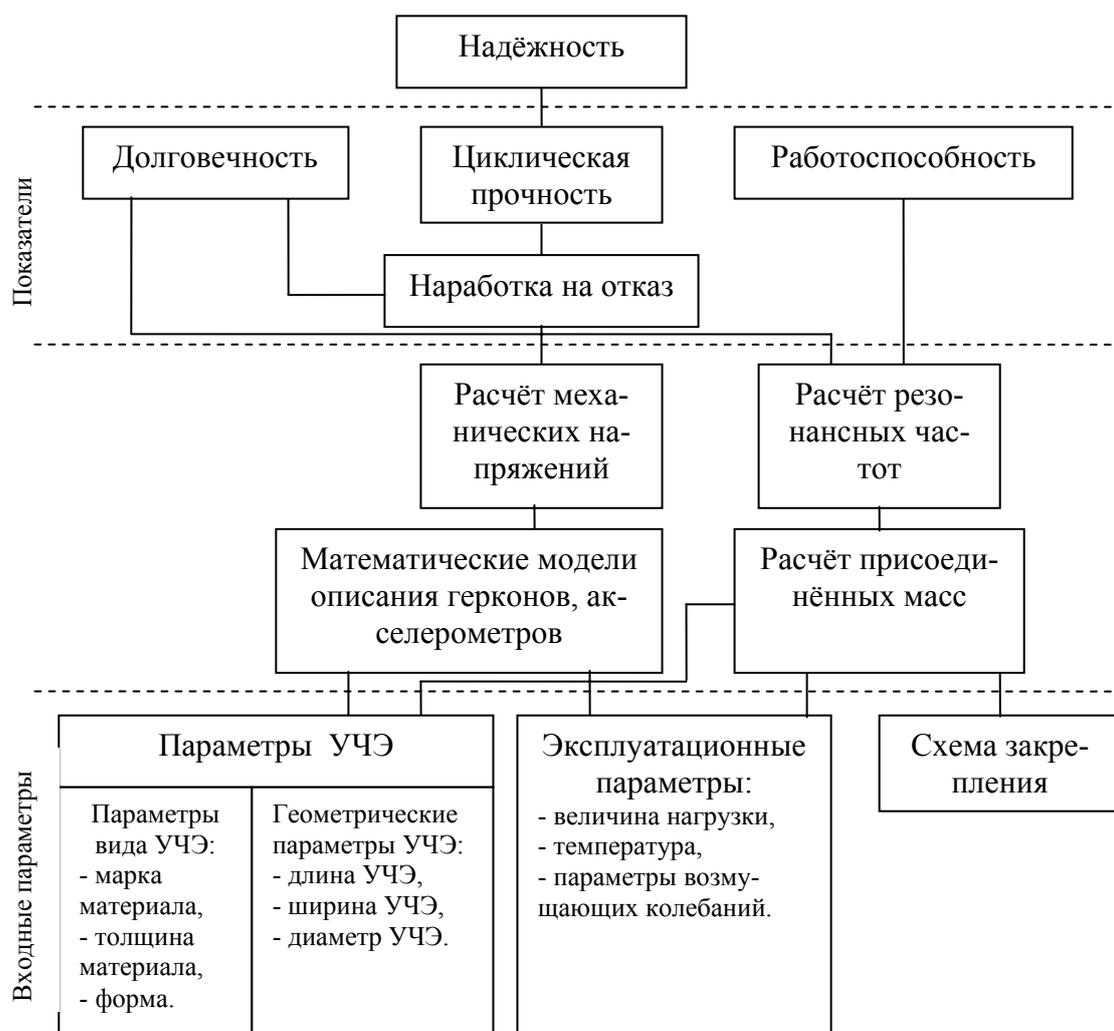


Рис. 3. Взаимосвязь параметров УЧЭ с показателями надежности

Контактные явления при поверхностном взаимодействии с различными фазами (газ, жидкость, твердое тело)	Геометрические параметры рабочей поверхности УЧЭ
Адсорбция	Ra, S, r', l, J
Теплопередача	$R_{max}, J, \omega, \alpha, R_z, r, t_p$
Взаимодействие с ионными и электронными пучками	J, l, r', r
Эмиссия электронов	r, J, S, α
Адгезия	t_p, r, R_z, Ra, α
Статическое и динамическое нагружение	r, R_z, ω, t_p
Электропроводность	r, R_z, ω, t_p
Смачивание	J, α, r', R_z

Таблица 1. Геометрические параметры поверхности УЧЭ и различные виды контактных явлений коммутирующих устройств и датчиков

Качество УЧЭ определяет дальнейшие параметры датчиков в целом. В свою очередь, одним из важнейших параметров работы устройства является надежность. При динамических условиях эксплуатации УЧЭ надежность характеризуется работоспособностью герконов, акселерометров. На рис. 3 приведена трехуровневая иерархическая структура взаимосвязи параметров датчиков с показателями надежности УЧЭ. Нижний уровень, названный входными параметрами, учитывает геометрию, вид и условия экс-

плуатации УЧЭ. Второй уровень представляет собой расчет механических напряжений по математическому описанию герконов, акселерометров и расчет резонансных частот по моделям расчета присоединенных масс. В результате на третьем уровне, используя результаты вычислений второго уровня, может быть осуществлен переход к прогнозированию ресурсных показателей надежности: долговечности, циклической прочности и работоспособности.

Влияние геометрических параметров поверхности УЧЭ на различные виды контактных явлений коммутирующих устройств и датчиков представлено в табл. 1.

Заключение

Таким образом, проведенный анализ зависимости работы датчиков от параметров УЧЭ в совокупности позволит создать современную методику, позволяющую проводить моделирование стереометрических параметров УЧЭ, что является одним из важных аспектов проблемы создания высокоточных и надежных датчиков систем управления.

Литература

1. Ткалич В.Л. Разработка и исследование методов повышения надежности герконов и реле на их основе. // Автореферат дисс. на соискание степени кандидата технических наук. ЛИТМО, Санкт-Петербург, 1994. 20 с.
2. Ткалич В.Л., Добрусин А.М., Фролов Н.Д. Повышение надежности пружинных ртутных и мембранных герконов. / Материалы Всероссийской научной конференции (Computer-Based Conference) «Методы и средства измерений», раздел 11. «Мат. модели и численное моделирование измерительных приборов и датчиков». Нижний Новгород, 2000. 16 с.
3. Ткалич В.Л., Михеева О.Д., Железков В.В. Исследование динамики плоских УЧЭ устройств АСУ. / Материалы Всероссийской научной конференции (Computer-Based Conference) «Методы и средства измерений», раздел «Мат. модели и численное моделирование измерительных приборов и датчиков». Нижний Новгород, 2000. 14 с.
4. Ткалич В.Л., Гвоздев С.С., Фролов Н.Д., Сологубов Д.Н. Экспериментальное исследование эксплуатационных свойств УЧЭ с РМР рабочих поверхностей. / Материалы Юбилейной научно-технической конференции профессорско-преподавательского состава, посвященной 100-летию университета, СПб ГИТМО (ТУ), 29–31 марта 2000 г. 21 с.

ИНФОРМАЦИЯ И ИНФОРМАЦИОННЫЕ РЕСУРСЫ В ТЕХНИЧЕСКОМ ТВОРЧЕСТВЕ

Д.Н. Розилайнен

Научный руководитель – кандидат технических наук, доцент З.Г. Симоненко

Введение

Понятие «информация» – первичное и неопределяемое понятие. Оно предполагает наличие материального носителя информации, источника информации, передатчика информации, приемника и канала связи между передатчиком и приемником. Понятие информации используется во всех сферах человеческой жизни, и любая деятельность человека представляет собой процесс сбора и переработки информации, принятия на ее основе решений и их выполнения. С появлением современных средств вычислительной техники информация стала выступать в качестве одного из важнейших ресурсов научно-технического прогресса. В настоящее время существует два подхода к определению информации:

- информация техническая, количество которой может быть точно вычислено, и процессы, происходящие с такой информацией, подчиняются физическим законам;
- информация семантическая, т.е. смысловая. Для такой информации предлагаются различные количественные оценки и даже строятся математические теории.

Определенным образом и в определенных условиях информация описывает как процессы, происходящие в естественных физических системах, так и процессы в системах, искусственно созданных.

Сторонники третьего подхода считают, что информация едина, но вот количественные оценки должны быть разными. Термин «информация» происходит от латинского слова "informatio", что означает сведения, разъяснения, изложение. Информация – сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые воспринимают информационные системы (живые организмы, управляющие машины и др.) в процессе жизнедеятельности и работы.

Основная часть

Применительно к компьютерной обработке данных под информацией понимают некоторую последовательность символических обозначений (букв, цифр, закодированных графических образов и звуков и т.п.), несущую смысловую нагрузку и представленную в понятном компьютеру виде. Информация может существовать в виде текстов, рисунков, чертежей, фотографий; в виде световых или звуковых сигналов; в виде радиоволн; в виде электрических и нервных импульсов; в виде магнитных записей; в виде жестов и мимики; в виде запахов и вкусовых ощущений; в виде хромосом, посредством которых передаются по наследству признаки и свойства

Информация обладает свойствами, характеризующими ее качественные признаки: достоверность, своевременность, новизна, полнота, ценность, понятность, доступность, краткость и др. Информацию можно: создавать, передавать, воспринимать, использовать, запоминать, принимать, копировать, формализовать, распространять, преобразовывать, комбинировать, обрабатывать, делить на части, упрощать, собирать, хранить, искать, измерять, разрушать, и т.д.

Информационные ресурсы – это идеи человечества и указания по их реализации, накопленные в форме, позволяющей их воспроизводство. Это книги, статьи, патенты, диссертации, научно-исследовательская и опытно-конструкторская документация, технические переводы, данные о передовом производственном опыте и др. Информацион-

ные ресурсы (в отличие от всех других видов ресурсов – трудовых, энергетических, минеральных и т.д.) тем быстрее растут, чем больше их расходуют.

В понятие «информационный потенциал общества» включается не только весь индустриально-технологический комплекс средств информатики той или иной страны, но также и сеть научно-исследовательских, учебных, административных, коммерческих и других организаций и социальных институтов, деятельность которых содействует эффективному использованию информационных ресурсов, а также подготовке для этих целей необходимого количества специалистов соответствующего профиля.

Информационный потенциал становится таким же важным экономическим и социальным фактором развития, как энергетический, промышленный и оборонный потенциал, численность населения и образовательный потенциал. В связи с этим особую актуальность приобретает проблема повышения информационной культуры общества, т.е. степени его подготовленности к эффективному использованию информационных ресурсов и продуцированию новых знаний.

Суть этой проблемы заключается в том, что остро ощущается тенденция отставания темпов и уровня образования в области информатики и информационных технологий от соответствующих темпов и уровня развития инструментальных средств. Эта тенденция носит глобальный характер. В России эта проблема стоит особенно остро по причине низкого уровня оснащенности современными средствами информатики учебных заведений.

Еще одна крупная проблема данного направления заключается в том, что для обеспечения активизации и эффективного использования информационных ресурсов общества необходимо осуществить «электронизацию» информационных фондов. По мнению академика А. Ершова, именно «в загрузке и активизации информационного фонда человечества в глобальной компьютерной сети, собственно, и заключается задача информатизации в ее техническом содержании». Активными информационными ресурсами является та часть национальных ресурсов, которую составляет информация, доступная для автоматизированного поиска, хранения и обработки: формализованные и законсервированные на машинных носителях в виде работающих программ профессиональные знания и навыки, текстовые и графические документы, а также любые другие содержательные данные, потенциально доступные на коммерческой основе пользователям национального парка компьютеров. Есть основания предполагать, что отношение объема активных информационных ресурсов к общему объему национальных информационных ресурсов становится одним из существенных экономических показателей, характеризующих эффективность использования этих важнейших информационных ресурсов. Информационная емкость машиностроительного проекта объекта средней сложности (самолет, энергетический блок и т.д.) составляет 25–100 Мбайт. Если принять минимальную стоимость информации (5 коп.) за 1 байт, то информационная стоимость такого объекта будет равна 12,5 млн. денонмированных рублей.

Уникальность современных компьютерных технологий состоит в том, что они позволяют сохранить так называемые «неосознанные знания». Из всего объема информации, которой владеет человек, вербально (т.е. в словах) может быть выражена лишь малая доля, и специалисты расходятся в ее количественной оценке – от долей процента до пяти процентов. Все остальное – интуиция, «золотые руки», профессиональное чутье – то, что принято называть неосознанным знанием. Большую часть своих знаний любой талантливый мастер уносит с собой. Каждое последующее поколение теряет значительную часть знаний, наработанных предыдущим поколением. Задача компьютерных систем состоит в том, чтобы извлечь и представить в явном виде то, что и сам эксперт не мог для себя объяснить. Воплощенное в компьютерной системе, это знание будет сохранено навсегда; оно станет всеобщим достоянием и начнет автономно действовать. Разработанные сегодня технологии вербализации неосознанного знания позволяют ус-

пешно справляться с этой задачей. Компьютер как новое средство хранения и обработки информации готов сделать общим достоянием целые пласты знаний, ранее терявшихся безвозвратно, т.е. продлить социальную память.

Область человеческой деятельности, связанная с процессами преобразования информации с помощью компьютеров и их взаимодействием со средой применения, изучается информатикой. Она появилась благодаря развитию компьютерной техники, базируется на свойствах информации и аппаратно-программных средствах обработки информации. Информатика в широком смысле представляет собой единство разнообразных отраслей науки, техники и производства, связанных с переработкой информации, главным образом, с помощью компьютеров и телекоммуникационных средств связи во всех сферах человеческой деятельности. Как информатику в целом, так и каждую ее часть обычно рассматривают с разных позиций – как отрасль народного хозяйства, как фундаментальную науку, как прикладную дисциплину.

Информатика как отрасль народного хозяйства состоит из однородной совокупности предприятий разных форм хозяйствования, где занимаются производством компьютерной техники, программных продуктов и разработкой современной технологии переработки информации. Специфика и значение информатики как отрасли производства состоят в том, что от нее во многом зависит рост производительности труда в других отраслях народного хозяйства. Для нормального развития этих отраслей производительность труда в самой информатике должна возрастать более высокими темпами, так как в современном обществе информация все чаще выступает как предмет конечного потребления: людям необходима информация о событиях, происходящих в мире, о предметах и явлениях.

Дальнейший рост производительности труда и уровня благосостояния возможен лишь на основе использования новых интеллектуальных средств и человеко-машинных интерфейсов, ориентированных на прием и обработку больших объемов мультимедийной информации (текст, графика, видеоизображение, звук, анимация). При отсутствии достаточных темпов увеличения производительности труда в информатике может произойти существенное замедление роста производительности труда во всем народном хозяйстве. В настоящее время около 50% всех рабочих мест в мире поддерживается средствами обработки информации.

Информатика как фундаментальная наука занимается разработкой методологии создания информационного обеспечения процессов управления любыми объектами на базе компьютерных информационных систем.

В Европе можно выделить следующие основные научные направления в области информатики: разработка сетевой структуры, компьютерно-интегрированные производства, экономическая и медицинская информатика, информатика социального страхования и окружающей среды, профессиональные информационные системы.

Информатика занимается изучением закономерностей в информационных процессах (накопление, переработка, распространение); созданием информационных моделей коммуникаций в различных областях человеческой деятельности; разработкой информационных систем и технологий в конкретных областях и выработкой рекомендаций относительно их жизненного цикла: для этапов проектирования и разработки систем, их производства, функционирования и т.д.

Главная функция информатики заключается в разработке методов и средств преобразования информации и их использовании в организации технологического процесса переработки информации.

Задачи информатики состоят в исследовании информационных процессов любой природы; в разработке информационной техники и создании новейшей технологии переработки информации на базе полученных результатов исследования информационных процессов; в решении научных и инженерных проблем создания, внедрения и

обеспечения эффективного использования компьютерной техники и технологии во всех сферах общественной жизни.

Информатика является комплексной научно-технической дисциплиной, призванной создавать новые информационные техники и технологии для решения проблем в других областях.

Информационная технология – это совокупность методов и устройств, используемых людьми для обработки информации. Человечество занималось обработкой информации тысячи лет. Первые информационные технологии основывались на использовании счетов и письменности. Около пятидесяти лет назад началось исключительно быстрое развитие этих технологий, что в первую очередь связано с появлением компьютеров. В настоящее время термин «информационная технология» употребляется в связи с использованием компьютеров для обработки информации. Информационные технологии охватывают всю вычислительную технику и технику связи и, отчасти, бытовую электронику, телевидение и радиовещание. Они находят применение в промышленности, торговле, управлении, банковской системе, образовании, здравоохранении, медицине и науке, транспорте и связи, сельском хозяйстве, системе социального обеспечения, служат подспорьем людям различных профессий.

В настоящее время создание крупномасштабных информационно-технологических систем является экономически возможным, хотя совершенствование информационных технологий составляет самую важную, хотя дорогостоящую и трудную задачу, и это обуславливает появление национальных исследовательских и образовательных программ, призванных стимулировать их разработку.

Заключение

Комплекс индустрии информатики станет ведущим в информационном обществе.

Тенденция к обладанию информационными ресурсами в существенной степени зависит от прогресса информатики как единства науки, техники и производства. Однако решение этой грандиозной по масштабам и социальным последствиям проблемы отстоит на много лет вперед.

Литература

1. <http://www.n-t.org/ri/sb/kt07.htm>
2. <http://www.n-t.org/ri/sb/kt.htm>

ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ДЛЯ РЕШЕНИЯ ПРИКЛАДНЫХ ЗАДАЧ ПРИБОРОСТРОЕНИЯ

А.О. Барынин

Научный руководитель – кандидат технических наук, доцент З.Г. Симоненко

В работе рассмотрены особенности использования двух информационных технологий – управленческой и компьютерной технологии обработки данных и АРМ – с целью эффективного внедрения информационных ресурсов в процессы автоматизации производственной или административной деятельности.

Введение

Кроме вещества и энергии, в жизни человека огромную роль играет еще одна составляющая – информация. С появлением ЭВМ появляются новые науки, появляется новый вид переработки информации на базе компьютерных вычислительных систем или информационных технологий. Информационные технологии применяются всюду – в планировании и управлении, научных исследованиях и разработках, экспериментах, проектировании, денежно-кассовых операциях, криминалистике, медицине, образовании и др. Основная цель использования информационных технологий – автоматизация производственной или административной работы, которая наиболее эффективно используется в сочетании с разнообразными средствами связи (коммуникации). Современные средства коммуникации, разработанные для персональных компьютеров, предоставляют в распоряжение пользователей возможности доступа в ЭВМ более высокого класса, подключения их в глобальные и локальные вычислительные сети.

Целью данной работы является рассмотрение компьютерных информационных технологий

Основная часть

Информация является одним из ценнейших ресурсов общества, наряду с такими традиционными материальными видами ресурсов, как нефть, газ, полезные ископаемые и др., а, значит, процесс ее переработки по аналогии с процессами переработки материальных ресурсов можно воспринимать как технологию. Тогда справедливо следующее определение.

Информационная технология – процесс, использующий совокупность средств и методов сбора, обработки и передачи данных (первичной информации) для получения информации нового качества о состоянии объекта, процесса или явления (информационного продукта).

Цель технологии материального производства – выпуск продукции, удовлетворяющей потребности человека или системы.

Цель информационной технологии – производство информации для ее анализа человеком и принятия на его основе решения по выполнению какого-либо действия [1].

Информационная технология является наиболее важной составляющей процесса использования информационных ресурсов общества. Внедрение персонального компьютера в информационную сферу и применение телекоммуникационных средств связи определили новый этап развития информационной технологии.

Можно выделить три основных принципа новой (компьютерной) информационной технологии: интерактивный (диалоговый) режим работы с компьютером; интегрированность (стыковка, взаимосвязь) с другими программными продуктами и гибкость процесса изменения как данных, так и постановок задач.

Техническими средствами производства информации является аппаратное, программное и математическое обеспечение этого процесса. С их помощью производится переработка первичной информации в информацию нового качества. Выделим отдель-

но из этих средств программные продукты и назовем их инструментарием. Для большей четкости можно его конкретизировать, назвав программным инструментарием информационной технологии.

Инструментарий информационной технологии – один или несколько взаимосвязанных программных продуктов для определенного типа компьютера, технология работы в котором позволяет достичь поставленную пользователем цель. В качестве инструментария можно использовать следующие распространенные виды программных продуктов для персонального компьютера: текстовый процессор (редактор), настольные издательские системы, электронные таблицы, системы управления базами данных, электронные записные книжки, электронные календари, информационные системы функционального назначения (финансовые, бухгалтерские, для маркетинга и пр.), экспертные системы и т.д.

Информационная технология тесно связана с информационными системами, которые являются для нее основной средой. На первый взгляд может показаться, что введенные в учебнике определения информационной технологии и системы очень похожи между собой. Однако это не так.

Информационная технология является процессом, состоящим из четко регламентированных правил выполнения операций, действий, этапов разной степени сложности над данными, хранящимися в компьютерах. Основная цель информационной технологии – в результате целенаправленных действий по переработке первичной информации получить необходимую для пользователя информацию.

Информационная система является средой, составляющими элементами которой являются компьютеры, компьютерные сети, программные продукты, базы данных, люди, различного рода технические и программные средства связи и т.д. Основная цель информационной системы – организация хранения и передачи информации, компьютерная система обработки информации. Реализация функций информационной системы невозможна без знания ориентированной на нее информационной технологии. Информационная технология может существовать и вне сферы информационной системы.

Таким образом, информационная технология является более емким понятием, отражающим современное представление о процессах преобразования информации в информационном обществе. В умелом сочетании двух информационных технологий – управленческой и компьютерной – залог успешной работы информационной системы [2].

Информационная технология обработки данных предназначена для решения хорошо структурированных задач, по которым имеются необходимые входные данные и известны алгоритмы и другие стандартные процедуры их обработки. Эта технология применяется на уровне операционной (исполнительской) деятельности персонала невысокой квалификации в целях автоматизации некоторых рутинных, постоянно повторяющихся операций управленческого труда. Поэтому внедрение информационных технологий и систем на этом уровне существенно повысит производительность труда персонала, освободит его от рутинных операций, возможно, даже приведет к необходимости сокращения численности работников.

На уровне операционной деятельности решаются следующие задачи:

- обработка данных об операциях, производимых фирмой;
- создание периодических контрольных отчетов о состоянии дел в фирме;
- получение ответов на всевозможные текущие запросы и оформление их в виде бумажных документов или отчетов.

Пример контрольного отчета: ежедневный отчет о поступлениях и выдачах наличных средств банком, формируемый в целях контроля баланса наличных средств. Пример запроса: запрос к базе данных по кадрам, который позволит получить данные о требованиях, предъявляемых к кандидатам на занятие определенной должности.

Существует несколько особенностей, связанных с обработкой данных, отличающих информационную технологию от всех прочих:

- выполнение необходимых фирме задач по обработке данных. Каждой фирме предписано законом иметь и хранить данные о своей деятельности, которые можно использовать как средство обеспечения и поддержания контроля на фирме, поэтому в фирме обязательно должна быть информационная система обработки данных и разработана соответствующая информационная технология;
- решение только хорошо структурированных задач, для которых можно разработать алгоритм;
- выполнение стандартных процедур обработки. Существующие стандарты определяют типовые процедуры обработки данных и предписывают их соблюдение организациями всех видов;
- выполнение основного объема работ в автоматическом режиме с минимальным участием человека;
- использование детализированных данных. Записи о деятельности фирмы имеют детальный (подробный) характер, допускающий проведение ревизий. В процессе ревизии деятельность фирмы проверяется хронологически от начала периода к его концу и от конца к началу;
- акцент на хронологию событий;
- требование минимальной помощи в решении проблем со стороны специалистов других уровней.

Автоматизированное рабочее место (АРМ) – терминал информационной системы управления, оборудованный средствами, обеспечивающими участие человека в реализации функций автоматизированных систем управления (АСУ). АРМ отличается следующими признаками:

- доступная пользователю совокупность технических, программных, информационных и др. средств;
- размещение вычислительной техники непосредственно (или вблизи) на рабочем месте пользователя;
- возможность создания и совершенствования проектов автоматизированной обработки данных в конкретной сфере деятельности;
- осуществление обработки данных самим пользователем;
- диалоговый режим взаимодействия пользователя с ЭВМ как в процессе решения задач управления, так и в процессе их проектирования.

АРМ в системе управления представляет собой проблемно-ориентированный комплекс технических, программных, лингвистических (языковых) и др. средств, установленный непосредственно на рабочем месте пользователя и предназначенный для автоматизации операций взаимодействия пользователя с ЭВМ в процессе проектирования и реализации задач.

Множество известных АРМ может быть классифицировано на основе следующих обобщенных признаков:

- функциональная сфера использования (научная деятельность, проектирование, производственно-технологические процессы, организационное управление);
- тип используемой ЭВМ (микро-, мини-, макро ЭВМ);
- режим эксплуатации (индивидуальный, групповой, сетевой);
- квалификация пользователей (профессиональные и непрофессиональные).

Внутри каждой из выделенных групп АРМ может быть проведена более детальная классификация. В частности, деловые АРМ сближают пользователя с возможностями современной информатики и ВТ и создают условия для работы без посредника – профессионального программиста. При этом обеспечивается как автономная работа, так и

возможность связи с другими пользователями в пределах организационных структур (с учетом особенностей этих структур).

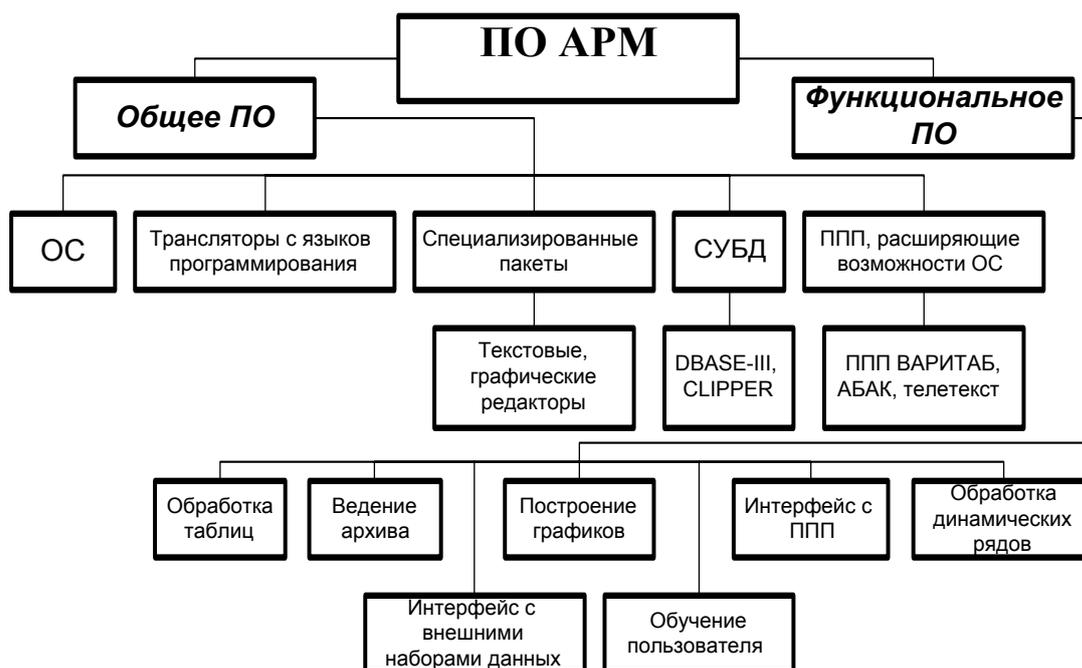


Рис. 1. Программное обеспечение АРМ

Параметрический ряд деловых АРМ позволяет создать единую техническую, организационную и методологическую базу компьютеризации управления. Первоначально информационная технология локализуется в пределах персонального или группового АРМ, а в последующем (при объединении АРМ средствами коммуникации) создаются АРМ сектора, отдела, учреждения и формируется коллективная технология. Тем самым достигается гибкость всей структуры и возможность наращивания информационной мощности.

Можно выделить три класса типовых АРМ: АРМ руководителя; АРМ специалиста; АРМ технического и вспомогательного персонала.

При проектировании ПО АРМ (рис. 1) необходимо соблюдать принцип ориентации разрабатываемых программных средств на конкретного пользователя, что должно обеспечить реализацию функций, соответствующих профессиональной ориентации АРМ. В целом, разрабатываемое ПО АРМ должно обладать свойствами гибкости, адаптивности, модифицируемости и настраиваемости на конкретное применение.

В настоящее время прослеживается тенденция к максимальному приближению информационных и программных ресурсов к пользователю. ПЭВМ, работающие в сети, имеют существенное преимущество перед АРМ, работающими в режиме разделения времени. И, главное, средства интеллектуального интерфейса обеспечивают пользователя простыми и надежными способами решения своих профессиональных задач. Основной результат – это изменение интерфейса конечного пользователя с терминалом. От метода «запоминай (состояние своих ресурсов) → проектируй (необходимую последовательность действий в терминах команд) → набирай (управляющий текст)» происходит переход к методу «смотри (на графическую модель состояния ресурсов) → выбирай (необходимое действие из иерархического меню)». Такой интерфейс поддерживается всеми средствами информационной технологии – составными частями базы знаний, включающей базу данных, прикладное программное обеспечение и опорной технологии, базирующейся на аппаратных средствах, системном и инструментальном программном обеспечении.

В данной работе не рассмотрены информационные этапы разработки технологических процессов и информационная технология обработки данных.

Заключение

В работе рассмотрены понятия «новая информационная технология», «инструментарий», «информационная технология обработки данных» и «АРМ» с целью эффективного внедрения информационных ресурсов в процессы автоматизации в производственной или административной деятельности.

Современное общество вряд ли можно представить без информационных технологий. Перспективы развития вычислительной техники сегодня сложно представить даже специалистам. Однако, ясно, что в будущем нас ждет нечто грандиозное. И если темпы развития информационных технологий не сократятся (а в этом нет никаких сомнений), то это произойдет очень скоро.

Литература

1. Дондуков А.Н., Воронцов В.А. Научно-техническая сфера России на пороге нового века. // Информационное общество. 2000. Вып. 6. С. 3–6.
2. Бодякин В.И. Направленность эволюции информационных систем. / Всероссийский симпозиум «Современные проблемы неравновесной термодинамики и эволюции сложных систем». Тез. докл. М.: МАКС Пресс, 2004. С. 15–16.

РАСЧЕТ ПАРАМЕТРОВ ЭЛЛИПТИЧЕСКИ-ПОЛЯРИЗОВАННОЙ ВОЛНЫ ОТРАЖАЮЩЕЙ СРЕДЫ ДЛЯ ИССЛЕДОВАНИЯ ПАРАМЕТРОВ ОДНОРОДНОГО ВОЛНОВОДА

О.Л. Студеникин

Научный руководитель – кандидат технических наук, доцент З.Г. Симоненко

Работа посвящена выявлению возможностей использования параметров эллиптически-поляризованной волны отражающей среды при исследовании свойств оптико-волоконных датчиков. С помощью программы расчета и параметров эллиптически поляризованной волны отражающей среды найден коэффициент отражения и вычислены потери на отражение в оптическом волокне.

Введение

Современные информационные технологии, посвященные изучению общих закономерностей изменения физических свойств материалов и сред, широко используются в разработке элементной базы электроники и микроэлектроники. Использование тонкопленочных диэлектрических волноводов открывает путь к созданию оптических модуляторов, переключателей, перестраиваемых фильтров, поляризаторов. Потому понятен интерес к технике экспериментальных исследований в области отражающей эллипсометрии, позволяющей получить такие оптические параметры отражающей среды, как интенсивность отраженной волны, угол поляризации, составляющие интенсивности поляризованного света, фаза между s- и p-компонентами отраженной волны

Актуальное поле исследования предоставляется в создаваемых программах, которые весьма расширяют возможности моделирования параметров эллиптически поляризованной волны отражающей среды для целей исследования параметров высокоскоростных волоконно-оптических линий связи, датчиков и систем оптической обработки информации.

Основная часть

Широко распространено применение чувствительных и быстродействующих методов эллипсометрии с лазерными источниками излучения и поляризационными фильтрами для модуляции световых лучей и изменения плоскости поляризации света. Такие оптические параметры отражающей среды, как интенсивность отраженной волны, поляризация светового пучка и ее параметры (угол поляризации, составляющая интенсивности поляризованного пучка, фаза между s- и p-компонентами отраженной волны) используются в микроэлектронике, сенсорной технике, при исследовании явления двулучепреломления в современных опто-волоконных датчиках.

На основе теоретического анализа и последующих разработок [1, 2] в созданных схемах показана принципиальная возможность получения следующих параметров поляризованной волны: интенсивность отраженной волны, угол поляризации, составляющая интенсивности поляризованного света.

При падении света на границу раздела двух сред происходит его частичное преломление и отражение. Отражение света от поверхности диэлектрика описывается формулами Френеля:

$$\rho_p = -\frac{\sin(\alpha - \beta)}{\sin(\alpha + \beta)}; \quad \rho_s = -\frac{\operatorname{tg}(\alpha - \beta)}{\operatorname{tg}(\alpha + \beta)}, \quad (1)$$

где ρ – амплитудный коэффициент отражения, причем ρ_p – коэффициент отражения для волны с электрическим вектором, перпендикулярным плоскости падения, ρ_s – коэффициент отражения для волны с электрическим вектором, расположенным в плоско-

сти падения, а α и β – углы падения и преломления, соответственно, причем углы α и β связаны законами преломления

$$n_2 \sin \beta = n_1 \sin \alpha. \quad (2)$$

Формулы (1) можно переписать в ином виде:

$$\rho_P = -\frac{n_2 \cos \beta - n_1 \cos \alpha}{n_2 \cos \beta + n_1 \cos \alpha}; \quad \rho_S = -\frac{n_2 \cos \alpha - n_1 \cos \beta}{n_2 \cos \alpha + n_1 \cos \beta}. \quad (3)$$

Если $n_1 > n_2$, то при некотором критическом угле падения $\alpha_0 \sin \beta = 1$, т.е. $n_2 = n_1 \sin \alpha_0$, и при дальнейшем увеличении угла падения закон преломления можно формально использовать лишь при условии, что угол $\beta = \frac{\pi}{2} - i\gamma$. При этом закон преломления запишется как

$$n_2 \operatorname{sh} \gamma = n_1 \sin \alpha, \quad (4)$$

а формулы (3) примут вид:

$$\rho_P = -\frac{i n_2 \operatorname{sh} \gamma - n_1 \cos \alpha}{i n_2 \operatorname{sh} \gamma + n_1 \cos \alpha}; \quad \rho_S = -\frac{n_2 \cos \beta - i n_1 \operatorname{sh} \gamma}{n_2 \cos \beta + i n_1 \operatorname{sh} \gamma}. \quad (5)$$

В этом случае ρ_P и ρ_S по абсолютной величине оказываются равными единице, и реализуется полное внутреннее отражение (ПВО). Коэффициенты отражения из формулы (5) при этом можно представить в виде:

$$\rho_P = e^{-i\delta_P}, \quad \rho_S = e^{-i\delta_S}, \quad (6)$$

где

$$\operatorname{tg} \delta_P = \frac{2n_1 n_2 \cos \alpha \cdot \operatorname{sh} \gamma}{n_1^2 \cos^2 \alpha - n_2^2 \operatorname{sh}^2 \gamma}, \quad \operatorname{tg} \delta_S = \frac{2n_1 n_2 \cos \alpha \cdot \operatorname{sh} \gamma}{n_1^2 \operatorname{sh}^2 \gamma - n_2^2 \cos^2 \alpha}. \quad (7)$$

При $\alpha < \alpha_0$ результат взаимодействия падающей волны с границей раздела проявляется в изменении амплитудного коэффициента отражения в соответствии с формулами (2), а при $\alpha > \alpha_0$, т.е. при ПВО – в изменении фазы отраженной волны при коэффициенте отражения, равном единице, в соответствии с формулами (5). В обоих случаях характер взаимодействия зависит от поляризации падающей волны. Амплитуды как преломленной, так и отраженной волны окажутся комплексными, что означает: между компонентами отраженной волны возникает разность фаз, приводящая к эллиптической поляризации отраженной волны. Вычислив или измерив отношение полуосей эллипса этой поляризованной волны, найдем требуемую разность фаз:

$$\frac{a}{b} = \operatorname{ctg} \frac{\delta}{2}. \quad (8)$$

При получении зависимостей коэффициентов отражения от угла падения волны используются формулы (5). При подстановке в них формулы (4) получаем:

$$\rho_P = \frac{\cos \alpha - i \cdot \sin \alpha}{\cos \alpha + i \cdot \sin \alpha}; \quad \rho_S = -\frac{n_2^2 \cos \alpha - i \cdot n_1^2 \sin \alpha}{n_2^2 \cos \alpha + i \cdot n_1^2 \sin \alpha}. \quad (9)$$

Согласно [2], при построении зависимости разности фаз составляющих отраженной волны от угла падения $\delta_P = 2 \cdot \alpha$, т.е.

$$\delta_S = \operatorname{arctg} \frac{\sin 2\alpha}{\left(\frac{n_1}{n_2} \cdot \sin \alpha\right)^2 - \left(\frac{n_2}{n_1} \cdot \cos \alpha\right)^2}. \quad (10)$$

Основным фактором, вносящим вклад в обратное отражение, является френелевское отражение R_F вследствие воздушного зазора S между торцами волокон. Наиболее общее выражение для коэффициента отражения из (12) имеет вид:

$$R_F = \frac{(n_1^2 - n_2^2) \sin^2(2\pi n S / \lambda)}{4n_1^2 n_2^2 + (n_1^2 - n_2^2) \sin^2(2\pi n S / \lambda)}, \quad (11)$$

где λ – длина волны лазерного излучения, n_1 – показатель преломления сердцевинки волокна, n_2 – показатель преломления среды в зазоре.

Непараллельность торцов приводит к разным значениям зазора S для разных участков сердцевинки. В этом случае происходит усреднение по осцилляциям синуса, а потери на отражение определяются как

$$b_y = -10 \lg R_F = -10 \lg \frac{(n_1 - n)^2}{(n_1 + n)^2} [\text{дБ}]. \quad (12)$$

Существует три градации контакта для оптического волокна, отличающиеся уровнем потерь на обратное отражение: РС (–30 дБ), SuperPC (менее –40 дБ), UltraPC (менее –50 дБ). При $\lambda = 0,6328$ нм, $n_1 = 1,5$ и $n = 1,0$ $b = -11$ дБ, а угол $\alpha \leq 8^\circ$.

Расчет требуемых зависимостей программно реализован на языке C++ средствами Turbo C++3.0 и Borland C++5.0 [3]. В результате получены исполняемые файлы Raschet.exe – расчетный модуль; Graph.exe – графический модуль.

Заключение

Выполненный расчет с заданными параметрами позволяет оптимизировать параметры рабочего образца однородного оптического волновода без затрат на его изготовление

Литература

1. Симоненко З.Г. Исследование параметров скорости массопереноса в жидких бинарных системах с границей раздела. // Материалы IV Международной научной конференции «Проблемы пространства, времени и движения», СПбГИТМО (ТУ), Санкт-Петербург, 2000. с. 22.
2. Симоненко З.Г. Численные методы решения задач эллипсометрии при анализе системы подложка-пленка. / В сб.: Оптические методы исследования дефектов и дефектообразования элементной базы микроэлектроники микросенсорной техники. СПб: СПбГУ ИТМО, 2002. С. 109–111.
3. Симоненко З.Г., Ваняев В.Н. Расчет оптических параметров отражающей среды методами эллипсометрии. / В сб.: Диагностика и функциональный контроль качества оптических материалов. СПб: СПбГУ ИТМО, 2004. С. 212–218.

АППАРАТНЫЕ СРЕДСТВА РЕШЕНИЯ ЗАДАЧ МАССОПЕРЕНОСА ЖИДКИХ СРЕД

З.С. Пуликовская

Научный руководитель – кандидат технических наук, доцент З.Г. Симоненко

В работе рассматриваются аппаратные средства решения задач массопереноса жидких сред: с помощью лазерных поляризационных интерферометров с элементами нуль-эллипсометрии и ЯМР-томографа.

Введение

Изучение нетрансляционного массопереноса в анизотропных средах связано с широким кругом теоретических и прикладных задач математической физики. Современные измерительные технологии позволяют проводить исследование параметров массопереноса на качественно новом уровне двумя методами.

Разработаны и созданы новые лазерные методы и аппаратура для исследования явления массопереноса в жидких бинарных средах с границей раздела путем измерения параметров массопереноса, обеспечивающих объективный контроль за измеряемыми величинами в непрерывном и дискретном аппаратных режимах с учетом специфики протекающего динамического процесса переноса [1].

Сущность второго метода – метода ЯМР-томографии – заключается в воздействии на исследуемый объект, помещенный в постоянное поле магнитное поле с градиентом напряженности, каскадом радиочастотных импульсов [2]. Параметры излучаемых ядрами сигналов, представляющих собой высвобожденную поглощенную энергию после прекращения воздействия, зависят от содержания (плотности) исследуемых протонов и от времени релаксации протонов, а точнее, от его составляющих T_1 (продольной или спин-решетчатой релаксации) и T_2 (поперечной или спин-спиновой релаксации).

В работе рассматриваются методы и приборы, требования к ним, особенности проведения экспериментальных исследований, их возможности, область применения.

Основная часть

Первый метод. В основании первого метода лежит высокоточное измерение скорости массопереноса (коэффициента диффузии) в жидкой бинарной системе с границей раздела путем совместного использования оптических методов поляризационной интерферометрии и лазерной нуль-эллипсометрии [3]. Уникальность эксперимента, проведенного с помощью такого эллипсометра, заключается в возможности регистрации параметров массопереноса в реальном масштабе времени на основании зависимости показателя преломления раствора от его концентрации.

На рис. 1 изображена кювета, конструкция которой обеспечивает четкую границу раздела фаз в начальный момент времени, в которой находятся водные растворы KCl различных концентраций – C_1 и C_2 .

Математическое описание эксперимента основано на анализе изменения геометрии изоконцентрационных поверхностей в процессе диффузии. Аналитическое описание геометрии таких поверхностей требует обращения к эллиптическим функциям Якоби $\operatorname{sn}(U, E)$, $\operatorname{cn}(U, E)$, $\operatorname{dn}(U, E)$, рассматриваемым как функции двух переменных – аргумента U и эксцентриситета E . Начальные условия задачи описываются функцией $\operatorname{sn}(U, E)|_E^2=1$. На рис. 2 приведен вид этой функции при различных значениях эксцентриситета. Рисунок иллюстрирует начальные условия задачи – ступеньку концентраций в начальный момент времени – и процесс ее выравнивания в дальнейшем. По оси абсцисс отложено время t , по оси ординат – синусная составляющая интенсивности поляризованного излучения I , проходящего через исследуемую среду. Определяя проме-

жутки времени t_1, t_2, t_3 , которые соответствуют экстремумам кривой, были получены точные значения коэффициента диффузии для различных растворов.

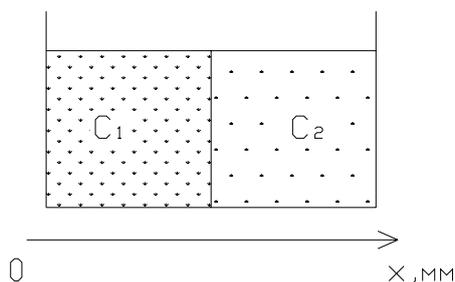


Рис. 1. Диффузионная кювета с растворами различной концентрации

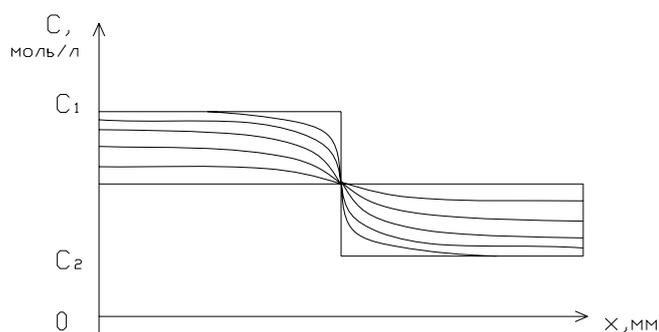


Рис. 2. График функции эллиптического синуса при различных значениях эксцентриситета, отражающий изменение концентрационного профиля системы

В результате эксперимента впервые получена феноменологическая кривая протекания процесса массопереноса бинарной жидкости в реальном масштабе времени, приведенная на рис. 3.

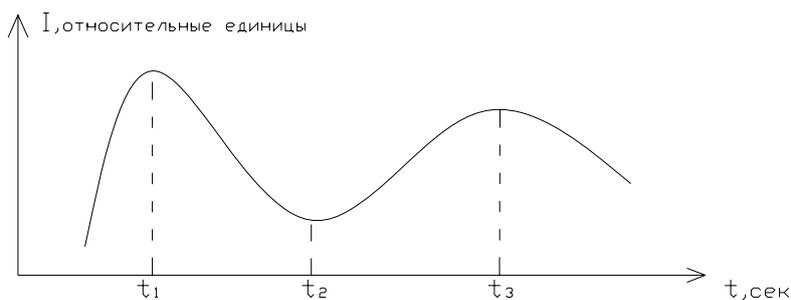


Рис. 3. Кривая хода диффузии

Обширный экспериментальный материал [4] по исследованию влияния на времена релаксации процессов ассоциации, диссоциации, самодиффузии и диффузии воды позволяет проводить точные расчеты для моделирования термодинамики жидких состояний. Показано, что по мере приближения критической смеси бинарных систем к критической температуре встречные молекулярные потоки выравниваются по молекулярному составу независимо от имеющихся концентраций. Метод измерения основывается на влиянии атомов гадолиния на интенсивность ЯМР-сигнала. Процесс диффузии наблюдается с помощью ЯМР-томографа в образце известной формы. При этом регистрировался график зависимости интенсивности ЯМР-сигнала в двух частях U-образной

трубки от времени. Наблюдения завершаются, когда сигнал в различных частях становится одинаковым, т.е. молекулы Gd-DTPA распределены в растворе равномерно.

Процесс диффузии наблюдался с помощью мини ЯМР-томографа, созданного на кафедре измерительных технологий и компьютерной томографии СПбГУ ИТМО [5], с помощью которого доступны для исследования образцы с линейными размерами до 20 мм. Для эксперимента была изготовлена U-образная трубка, представленная на рис. 4.

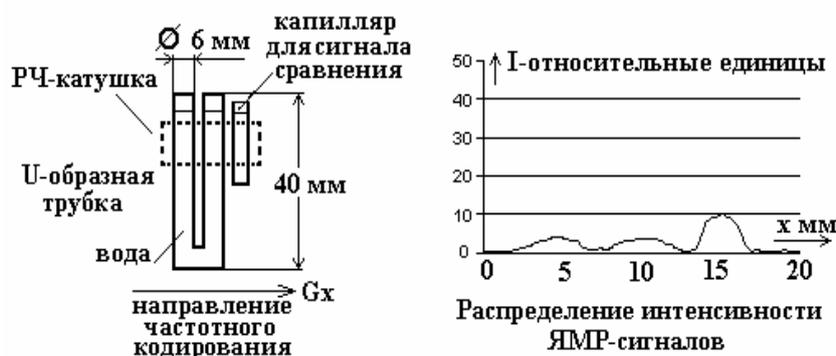


Рис. 4. U-образная трубка и распределение интенсивности ЯМР-сигнала до введения контрастного вещества Gd-DTPA

В данном случае была достаточна одномерная регистрация. При этом используется один градиент магнитного поля и одномерное преобразование Фурье, что увеличивает на порядок скорость регистрации данных. Линейный градиент магнитного поля прикладывался вдоль двух колен U-образной трубки. Величина градиента была подобрана так, чтобы два ЯМР-сигнала от разных колен U-образной трубки не перекрывались.

Так как диффузия – длительный процесс, занимающий несколько дней, то параметры томографа (например, коэффициент усиления, условия резонанса) могут изменяться за время эксперимента. Для учета влияния возможных изменений параметров томографа рядом с исследуемой пробиркой был помещен герметично закрытый капилляр, заполненный водой с небольшим количеством парамагнитной соли CuSO_4 для уменьшения времени релаксации протонов воды. Сигнал от этого капилляра использовался как контрольный для коррекции и обеспечения одинакового масштаба амплитуд ЯМР сигналов при длительной регистрации диффузии гадолиния

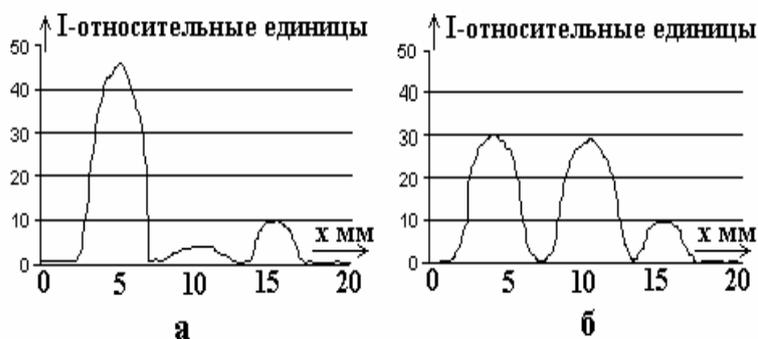


Рис. 5. Распределение интенсивности ЯМР сигналов: а) – в начале и б) – в конце регистрации диффузии атомов гадолиния

По результатам эксперимента были построены графики (рис. 6) зависимости отношения площадей сигналов от левой части пробирки $I_1(t)$ (спадающая кривая) и правой части пробирки $I_2(t)$ к площади сигнала от контрольной ампулы.

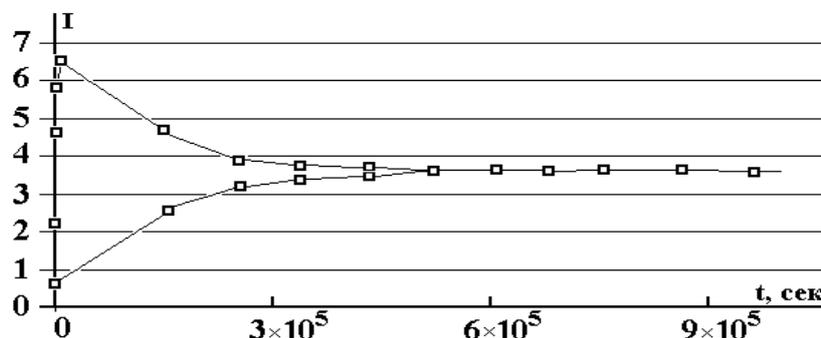


Рис. 6. Изменение интенсивности ЯМР сигналов I/I_0 в относительных единицах в зависимости от времени для левой и правой частей U-образной трубки

Коэффициент диффузии $D(\text{Gd-DTPA}) = (2,4 \pm 0,3) \times 10^{-5} \text{ см}^2/\text{с}$. Для более точного определения коэффициента диффузии потребуется термостатирование образца. Отметим, что полученная величина для $D(\text{Gd-DTPA})$ примерно соответствует коэффициенту самодиффузии протонов в воде: $D(\text{H}_2\text{O}) \approx 2,2 \times 10^{-5} \text{ см}^2/\text{с}$ при 25°C .

Заключение

Рассмотренные методы и аппаратные средства решения задач массопереноса с помощью лазерных поляризационных интерферометров с элементами нуль-эллипсометрии и ЯМР-томографа наглядно демонстрируют возможности расширения диапазона исследования при контроле параметров жидких и твердых сред с границей раздела, что является актуальным для целей микроэлектроники и волоконной оптики.

Литература

1. Симоненко З.Г., Москалев В.А. Способ определения коэффициента молекулярной диффузии в жидкостях и устройство для его реализации. Авт. свид. СССР № 1349452. Бюлл. ОИ ПОТЗ, 1987. № 45.
2. Симоненко З.Г. Исследование параметров скорости массопереноса в жидких бинарных системах с границей раздела. // Материалы IV Международной научной конференции «Проблемы пространства, времени и движения». СПб: СПбГИТМО (ТУ), 2000.
3. Симоненко З.Г. Экспериментальные и теоретические аспекты решения параболического уравнения для случая нетрансляционного массопереноса в жидкостях. // Научно-технический вестник СПбГУ ИТМО. 2001. Выпуск 3.
4. Иванов В.А. Способ определения внутреннего строения материальных объектов. Авт. свид. СССР № 1112266 с приор. от 26.03.1960 г.
5. Иванов В.К., Неронов Ю.И., Иванов В.А. «Мини – ЯМР-томограф кафедры измерительных технологий и компьютерной томографии СПбГУ ИТМО. // Научно-технический вестник СПбГУ ИТМО. 2001. Выпуск 3. С.201–208.

ЗАДАЧА ПРИНЯТИЯ ИНЖЕНЕРНЫХ РЕШЕНИЙ В УСЛОВИЯХ НЕОПРЕДЕЛЕННОСТИ

Н.Г. Шеховцов

Научный руководитель – кандидат технических наук, доцент З.Г. Симоненко

В работе рассматриваются основные этапы процесса инженерного анализа, показана необходимость разработки методов, позволяющих эффективно получать, обрабатывать нечеткую экспертную информацию и выполнять принятие решений в условиях неопределенности.

Введение

Проблема принятия решений или проблема выбора вариантов является одной из самых распространенных задач, возникающих практически во всех сферах деятельности: технической, экономической, социальной и др.

В некоторых случаях принятие решений является исключительно сложным вопросом, который носит субъективный характер и предполагает учет неколичественных человеческих факторов и суждений о ценности. Оптимальный компромисс можно найти, обращаясь к научным методам принятия решений, т.е. используя математические методы оптимизации, теорию вероятностей, математическую статистику или теорию полезности. Однако наиболее часто при принятии решений производится учет как количественных, так и качественных факторов, которые должны рассматриваться одновременно. В настоящее время принятие решений является одновременно и искусством и наукой. Однако появление вычислительной техники и успехи, достигнутые в разработке научных методов принятия решений, привели к изменению этих взглядов. В настоящее время в этой области интенсивно внедряются количественные методы.

В данной работе рассматриваются характеристики принятия решения, а также методика решения задач с детерминированными параметрами их «размытых» значений в условиях неопределенности на основе теории нечетких множеств.

Основная часть

Принятие решений является своего рода решением задачи. Каковы же существенные черты процесса принятия решений?

1. Наличие цели. Необходимость принятия решений диктуется наличием некоторой цели, которую необходимо достичь: например, выполнить задание, выбрать материал, назначить свидание девушке и т.д. Если же цель не поставлена, то и не возникает необходимости принимать какое-либо решение.

2. Наличие альтернативных линий. Решения принимаются в условиях, когда существует более одного способа достижения цели. Очевидно, что если существует лишь одна линия поведения, то выбора нет, и решения принимать не требуется. С различными альтернативами могут быть связаны различные затраты и различные вероятности успеха. Эти затраты и вероятности не всегда могут быть известны. Именно по этим причинам принятие решений часто сопряжено с неясностью и неопределенностью.

3. Учет существенных факторов. Решения принимаются в условиях действия большого числа факторов, которые, однако, различны для различных альтернатив. Это факторы экономического, технического, социального, личного и иного характера.

Итак, задача принятия решений возникает в том и только в том случае, когда существует цель, которую нужно достичь, когда возможны различные способы ее достижения и когда имеется большое число факторов, определяющих ценность различных альтернатив или вероятность успеха каждой из них. Теперь рассмотрим более подробно каждую из

этих характеристик в отдельности применительно к принятию решений при инженерном проектировании.

Внимательное рассмотрение процесса принятия решений с целью его лучшего уяснения приводит к необходимости четкого определения целей и задач. Однако нередко цели или, во всяком случае, непосредственно связанные с ними факторы являются как количественными (объективными), так и качественными (субъективными). В этих случаях для применения научных методов принятия решений нужны зрелость суждений и дальновидность, а также аналитическое и математическое мастерство.

Некоторые из обычных целей принятия решений при инженерном проектировании перечислены ниже, однако этот список нельзя считать исчерпывающим. Следует также помнить, что иногда для достижения поставленной цели требуется установить баланс между двумя или большим числом рассматриваемых факторов, причем в определенных ситуациях некоторые из них будут входить в задачу как ограничения, а не как компоненты поставленной цели. Целями при инженерном проектировании могут быть: начальные затраты, стоимость эксплуатации или обслуживания в течение определенного периода времени, надежность, вес, рабочие характеристики, к.п.д., внешний вид, безопасность, прибыль в течение определенного периода времени и т.д.

Очевидно, что, в конечном счете, может быть выбрана не самая лучшая альтернатива из числа рассматриваемых. В этом смысле качество выбора ограничено качеством альтернатив. Исчерпывающий список имеющихся альтернатив оказывает большую помощь при принятии решений.

Принятие решений есть выбор одной из альтернатив, и составление их списка является неотъемлемой частью этого процесса. В некотором смысле составление списка альтернатив совершенно аналогично определению задачи при инженерном анализе. Когда альтернативы неопределенны, список их неполон или даже не продуман, принять решение невозможно. Однако, когда альтернативы четко перечислены, задача больше не является неосязаемой. Составление списка альтернатив перед принятием решений в основном является творческим этапом.

Имеется одна альтернатива, которая почти всегда, во всяком случае, в самом начале, присутствует в любом списке. Это альтернатива – не принимать решения вообще. Иногда (и только иногда) оптимальным компромиссом будет отложить принятие решения, чтобы иметь больше времени для накопления новых фактов. Если же цель должна быть достигнута немедленно, то, разумеется, обычно нельзя откладывать принятие решений на неопределенно долгий срок.

В любой задаче принятия инженерных решений, по существу, можно выявить бесконечное множество факторов. Любая попытка составить их полный перечень или подробно их анализировать сопряжена с опасностью опустить некоторые реальные и важные факторы. Различные факторы, подлежащие рассмотрению, можно разделить на три группы и затем привести примеры самого общего характера. Основными группами факторов являются: факторы, связанные с ресурсами, технические факторы и чисто человеческие факторы. Под факторами, связанными с ресурсами, понимают время, денежные средства и производственные возможности. Под производственными возможностями здесь подразумеваются такие разнообразные вещи, как наличие материалов, деталей, техническое и научное мастерство, организационные возможности и т.д.

Для инженерных решений характерно, что без специального изучения или исследования информация о существенных сторонах таких факторов может оказаться недостаточно полной. К техническим факторам относятся факторы, которые непосредственно связаны с инженерным анализом или выработкой требований к конструкции. Обычно технические факторы являются конкретными и выражаются количественно. Эти технические факторы часто определяют один из трех видов ограничений: функциональные, областные и экстремальные. Функциональным ограничением является точное задание рабочих

характеристик, входных параметров или других ограничений. Областные ограничения отличаются от функциональных лишь тем, что они выражаются неравенствами. Экстремальные ограничения требуют, чтобы некоторый параметр был как можно больше или как можно меньше. Они требуют, чтобы рассматриваемый параметр в определенном направлении имел наибольшее или оптимальное значение. Очевидно, что экстремальные ограничения приводят к проблемам оптимизации. Примерами экстремальных ограничений являются следующие: «длина должна быть как можно меньше» или «расход должен быть как можно больше».

Кроме ресурсов и технических факторов, в ходе принятия инженерных решений важную роль играют чисто человеческие факторы. Эти факторы выражают не только требования политической или социальной целесообразности осуществления или достижения альтернативы, но и требования человеческой этики и морали. Для принятия правильного решения требуется не только техническая компетентность в оценке ресурсов и технических факторов, но и учет чисто человеческих факторов. Ниже приводится возможный алгоритм подготовки принятия решения.

1. Составляется возможно более полный список альтернатив. (Здесь необходимы творческий подход и изобретательность).
2. Составляется возможно более полный перечень факторов.
3. Список рассматриваемых факторов используется для уменьшения числа альтернатив, при этом обращается внимание на причину исключения каждой альтернативы. На данном этапе можно увидеть, что многие альтернативы нереальны. Другие альтернативы могут оказаться в высшей степени нецелесообразными. Этот процесс может быть крайне субъективным, и в некоторых случаях он строится на догадках. Однако, если нужно принять решение, другого выбора нет. В построении этих субъективных догадок и вынесении суждений о ценности как раз и проявляется искусство лица, принимающего решение. Таким образом, Некоторые факторы могут в одинаковой степени относиться ко всем оставшимся альтернативам, и поэтому их тоже не нужно больше рассматривать.
4. После выполнения перечисленных выше шагов получим один из следующих вариантов:
 - если больше не осталось альтернатив, нужно приложить больше творческих усилий для составления нового списка альтернатив;
 - если исключены все факторы, влияющие на выбор альтернатив, следует воспользоваться случайным выбором, чтобы остановиться на одной альтернативе из числа оставшихся;
 - если осталась только одна альтернатива, то решение будет принято совершенно субъективно;
 - если останется только один фактор, влияющий на выбор, то обычно не представляет труда отыскать наилучшую альтернативу. Если факторы исключались осторожно, то можно считать, что решение принято;
 - если позволяют условия, можно использовать научные методы принятия решений;
 - если ситуация остается все еще слишком сложной и возможен лишь субъективный подход, а применение научных методов принятия решений невозможно, то необходимо вернуться к выполнению п. 2, стараясь четко представить себе причины исключения различных альтернатив.

Одной из наиболее важных особенностей прикладных инженерных задач выбора является нечеткий характер критериев выбора альтернатив, их параметров, ограничений, накладываемых на возможность выбора тех или иных вариантов и т.д. Вследствие этого во многих случаях оказывается невозможным построение адекватной математической модели исследуемой проблемы, что влечет за собой необходимость использования экспер-

ных оценок, которые часто оказываются единственной информацией для принятия решений. Естественно, возникает необходимость разработки методов, позволяющих эффективно получать и обрабатывать нечеткую экспертную информацию.

Для нечетких множеств вводятся операции пересечения, объединения, дополнения, концентрации, размывания. Первые три операции являются обобщениями обычных операций, оставшиеся специфичны для нечетких. Методика решения рассматриваемых задач в условиях неопределенности на основе теории нечетких множеств при этом не обсуждается и остается скрытой от исполнителя, однако создание алгоритма ввода в задачи с детерминированными параметрами их «размытых» значений является непростой задачей.

«Размытость» параметра обычно можно представить интервалом его возможных значений (размахом) и наиболее ожидаемым (средним) значением, т.е. аргументами так называемого треугольного нечеткого числа. Эти аргументы (значимые точки нечеткого числа) естественны для анализа. Треугольные (тернарные) нечеткие числа задаются характеристической функцией (функцией принадлежности) треугольной формы: вершина треугольника при наиболее ожидаемом (среднем) значении равна единице, в точках основания треугольника, отвечающих размаху, принадлежность уменьшается до нуля.

В соответствии с этим значения с наиболее ожидаемой величиной однозначно относятся к множеству рассматриваемого параметра, а интервалы до минимально и максимально возможных значений являются нечеткими, принадлежность к которым линейно уменьшается до нуля. Треугольные нечеткие числа соответствуют высказыванию: «Рассматриваемый параметр приблизительно равен среднему значению и однозначно находится в диапазоне значений от минимально до максимально возможного».

Прикладной смысл нечеткого числа – это число, заданное с погрешностью. Чтобы работать с такими числами, нужно задавать функции принадлежности и погрешностей, но это невозможно сделать для всех значений нечеткого числа из рассматриваемого множества в силу бесконечности самого множества. Один из способов преодоления этой трудности – использование нечетких L-R (left-right) чисел. Если значимые точки нечеткого числа совпадают, то оно вырождается в обычное действительное однозначно число. Фиксированному уровню принадлежности отвечает соответствующий ему интервал достоверности нечеткого числа. (L-R)-нечеткость можно интерпретировать как способ взвешенного учета погрешностей.

Основные операции с нечеткими числами сводятся к операциям с их интервалами достоверности. Операции с интервалами, в свою очередь, выражаются через операции с действительными числами – границами интервалов. Для этого достаточно задаться фиксированным уровнем принадлежности и определить соответствующие ему интервалы достоверности нечетких чисел. Алгебраические операции выполняются по аксиоматическим правилам, называемым интервальными или сегментными. В соответствии с этими правилами детерминированные решения таких задач, модели, производственные функции, функции спроса легко преобразуются к виду, учитывающему нечеткость параметров. В рамках компьютерного вычисления операции по введению в детерминированные формулы нечетких параметров и вычислению нечетких функций реализуются в символьном, численном и графическом виде в программной среде Mathcad [1–7].

В заключение перечислим основные этапы процесса инженерного анализа.

1. Определение задачи в такой форме, для которой возможно получение решения.
2. Построение модели, которая достаточно проста, чтобы получить решение, и в то же время достаточно сложна, чтобы получаемые результаты имели смысл.
3. Везде, где возможно, желательно использовать в анализе основные физические принципы, а не специальные формулы.
4. Проверка работы. Для этого можно воспользоваться предельными значениями величин, проверкой размерности, проверкой исходя из физического смысла и т.д.

5. Оценка полученных результатов и их изучение или обобщение. В каком соотношении с реальной задачей находятся результаты, полученные с помощью модели?
6. Выдача полученных результатов наиболее подходящим и эффективным способом.

Заключение

Отмечена важность понимания количественных факторов в процессе принятия решений. Кратко изложен рациональный порядок принятия решений, который должен служить там, где это окажется возможным, средством приведения сложного процесса принятия решений к задаче, которую можно решать с помощью математических методов.

При инженерном проектировании необходимо принимать много решений. В некоторых случаях инженеру нужно уметь принимать решения в условиях неопределенности и затем продолжать движение дальше. Ответы на эти вопросы в свою очередь связаны с принятием определенных решений, однако в настоящее время они в значительной мере определяются вкусом, склонностями и личными качествами. В определенной мере успех инженера зависит именно от этого.

Литература

1. Диксон Дж. Проектирование систем: изобретательство, анализ и принятие решений. М.: Мир, 1969. 440 с.
2. Вязгин В.А. О некоторых схемах последовательного анализа вариантов в проектировании технических систем. // Изв. АН СССР. Техн. кибернет. 1984. № 6. С. 63–68.
3. Орловский С.А. Проблемы принятия решений при нечеткой исходной информации. М.: Наука, 1981. 208 с.
4. Борисов А.Н., Глушков В.И. Использование нечеткой информации в экспертных системах. // Новости искусственного интеллекта. № 3. 1991. С. 13–41.
5. Горелик А.Л., Абаев Л.Ч. О методе последовательного анализа вариантов в задачах выбора в нечеткой среде. // Кибернетика и системный анализ. 1992. № 4. С. 95–105.
6. Кофман А. Введение в теорию нечетких множеств. М.: Радио и связь, 1982. 432 с.
7. Андриевский Б.Р. Фрадков А.Л. Элементы математического моделирования в программных средах МАТЛАБ 5 и Scila. СПб: Наука, 2001. 286 с.

ОЦЕНКА ВЛИЯНИЯ ВЕЛИЧИНЫ ИСТОЧНИКА ИЗЛУЧЕНИЯ НА ПАРАМЕТРЫ МАССОПЕРЕНОСА

Хоанг Зянг

Научный руководитель – кандидат технических наук, доцент З.Г. Симоненко

В настоящее время основной тенденцией развития науки является повышение информативности исследований, обеспечиваемое разработкой и применением высокоточных быстродействующих комплексных методов.

Экспериментально наблюдаемый с помощью лазерной эллипсометрии процесс массопереноса в жидких средах позволяет провести аналитический анализ аппаратурных и измеряемых характеристик с целью дальнейшего прогнозирования и математического моделирования поведения аналогичных жидких систем в динамическом режиме. Общий классификационный признак рассматриваемого оптического метода – наличие бесконтактной неразрушающей связи источника излучения с объектом исследования и последующей регистрацией и обработкой измеряемой величины.

Измерительный комплекс [1], используемый для исследования, включает в себя источник излучения (лазер), измерительную часть (собственно эллиптического интерферометра) и регистрирующую часть, сопряженную с компьютером (рис. 1).

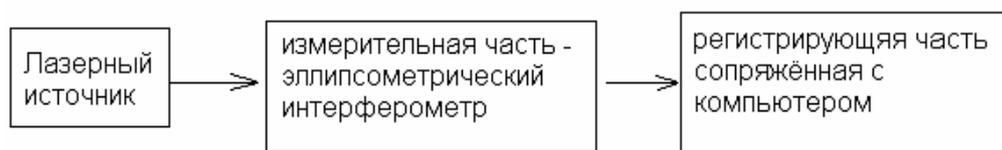


Рис. 1. Схема измерительного комплекса

Нами рассматриваются влияние характеристики лазерного источника на параметры измерительного процесса массопереноса или молекулярной диффузии [2].

Диффузия в жидких системах представляет собой самопроизвольный процесс перераспределения вещества в пространстве и времени вследствие теплового (броуновского) движения частиц, образующих рассматриваемую жидкую систему. Броуновское движение частицы характеризуется ее смещением за определенный промежуток времени, которое, согласно уравнению Смолуховокого-Эйнштейна, равно

$$\langle X^2 \rangle = 2Dt, \quad (1)$$

где X – смещение частицы в направлении оси X за промежуток времени \sqrt{t} ; $\langle X^2 \rangle$ – среднеквадратичное смещение частицы в направлении оси – за равные промежутки времени t , D – коэффициент диффузии. Для двухкомпонентной диффундирующей среды частицы компонентов растворитель–растворенное вещество смещаются, и их среднеквадратичное смещение характеризуется коэффициентом диффузии, величину которого находят из решения задачи о бесконечно длинном цилиндре. При условии, что концентрация в любом горизонтальном сечении этого цилиндра зависит от момента времени t и расстояния X сечения от поверхности раздела, т.е. $C=C(X, t)$, коэффициент D входит в уравнение Фика следующим образом:

$$\frac{dC}{d\tau} = D \frac{d^2 C}{dX^2}. \quad (2)$$

После преобразования получаем:

$$\frac{dn}{dx} = \Delta n \sqrt{\frac{k}{\pi}} l^{-kx^2}, \quad (3)$$

где $\frac{dn}{dx}$ – градиент показателя преломления в слое, отстоящем на расстоянии X от границы раздела, $\Delta n = n_1 - n_2$ – первоначально заданная разность показателей преломления двух жидкостей, $K = \frac{1}{4\pi D}$, D – коэффициент диффузии, τ – время от момента образования границы раздела.

В эксперименте [3] две жидкости заливаются в кювету (рис. 2) длиной h с резко выделенной границей раздела, параллельно которой на расстоянии a друг от друга распространяются два луча.

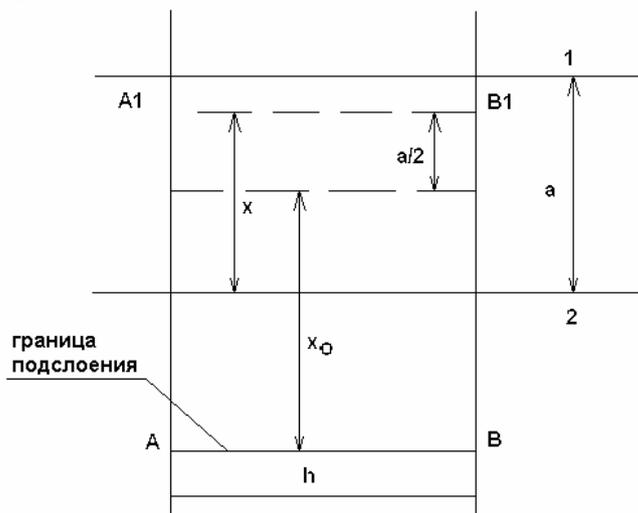


Рис. 2. Оптическая схема эксперимента

Приобретаемая ими разность хода (в единицах длин волн X) или разность фаз δ , приобретаемая этими лучами на пути длиной h в слое dn/dx , выражается следующим соотношением:

$$d\delta = \frac{h}{\lambda} dn. \quad (4)$$

Из (3) и (4) имеем:

$$d\delta = \frac{\Delta n l}{\lambda} \sqrt{\frac{k}{\pi}} \int_{x-a/2}^{x+a/2} l^{-kx^2} dx. \quad (5)$$

При заданных начальных данных h , Δn и a есть функция x и X_0 , где X_0 – расстояние от границы раздела двух жидких сред до прямой, равноудаленной от двух ортогонально поляризованных пучков. При введении новой переменной $y = x\sqrt{k}$ соотношение (5) преобразуется к виду

$$\delta = \frac{h}{2\lambda} L, \quad (6)$$

где $X = X_0 \pm a/2$,

$$L = \operatorname{erf}\left[\frac{1}{\sqrt{4D\tau}}(X_0 + a/2)\right] - \operatorname{erf}\left[\frac{1}{\sqrt{4D\tau}}(X_0 - a/2)\right]. \quad (7)$$

Определение коэффициента диффузии в простейшем случае сводится к вычислению параметров кривой разности фаз δ , зарегистрированной в различные моменты времени τ . С другой стороны, при последовательном прохождении максимумов и минимумов кривой хода диффузии на основании выражений для рассматриваемых значений δ_1 , δ_2 и δ_3 можно записать следующие соотношения:

$$\delta_2 = \delta_1 + \pi, \quad (8)$$

$$\delta_3 = \delta_2 + \pi. \quad (9)$$

Из соотношений (8) и (9) получим выражение

$$\frac{\delta_3 - \delta_2}{\delta_2 - \delta_1} = 1,$$

откуда в общем виде имеем

$$\frac{L_{i+2} - L_{i+1}}{L_{i+1} - L_i} = 1, \quad (10)$$

где L определяется из выражения (7), $i = 1, 2, 3$ – индекс, присвоенный каждому измеренному интервалу времени τ_1 , τ_2 и τ_3 , соответственно, D – коэффициент массопереноса, a – расстояние между двумя ортогонально поляризованными пучками.

В эксперименте [3] измеряются значения величин τ_1 , τ_2 и τ_3 . Тогда, зная величины X_0 и a , можно вычислить коэффициент массопереноса D из следующего уравнения:

$$\frac{\operatorname{erf}\left[\frac{1}{\sqrt{4D\tau_3}}(X_0+a/2)\right] - \operatorname{erf}\left[\frac{1}{\sqrt{4D\tau_3}}(X_0-a/2)\right] - \operatorname{erf}\left[\frac{1}{\sqrt{4D\tau_2}}(X_0+a/2)\right] + \operatorname{erf}\left[\frac{1}{\sqrt{4D\tau_2}}(X_0-a/2)\right]}{\operatorname{erf}\left[\frac{1}{\sqrt{4D\tau_2}}(X_0+a/2)\right] - \operatorname{erf}\left[\frac{1}{\sqrt{4D\tau_2}}(X_0-a/2)\right] - \operatorname{erf}\left[\frac{1}{\sqrt{4D\tau_1}}(X_0+a/2)\right] + \operatorname{erf}\left[\frac{1}{\sqrt{4D\tau_1}}(X_0-a/2)\right]} = 1. \quad (11)$$

Полученное математическое выражение для связи разности фаз δ и коэффициента массопереноса D позволяет получить величину последнего при измеренных значениях τ_1 , τ_2 и τ_3 [4].

Анализ аппаратных погрешностей, проведенный в работе [1], показывает, что изменение величины конечного размера источника излучения, которое обусловлено величиной светового диаметра пучка на выходном зеркале плоского резонатора лазерного излучателя, определяемого диаметром активного элемента лазера, влияет на величину амплитуды светового потока.

Известно [5], что интенсивность излучения от точечного источника есть

$$I = I_0 \{ \cos[2\pi\delta(x_0)] \} \quad (12)$$

в предположении, что этот точечный источник, расположенный в точке $X=X_0$, имеет единичную силу и нулевую фазу. Полная интенсивность излучения от протяженного источника имеет вид:

$$I = I_0 \int_{x_0-b/2}^{x_0+b/2} \{ 1 + \cos[2\pi\delta(x)] \} dx, \quad (13)$$

где b – конечный размер сечения (или ширина) пучка.

В качестве лазерного источника выбран гелий-неоновый многомодовый лазер непрерывного действия со следующими рабочими характеристиками: длина волны лазерного излучения $0,6328 \cdot 10^{-6}$ м, расходимость пучка лазерного источника излучения $2 \cdot 10^{-3}$ рад, мощность лазерного излучения $25 \cdot 10^{-3}$ Вт. Тогда при заданных граничных условиях рассматриваемого эксперимента и соответствующих преобразованиях [6] имеем:

$$I = I_0 \{ 1 + \cos \left[2\pi \sqrt{\frac{\Delta n l}{\lambda}} (\operatorname{erf} [\sqrt{k} (x_0 + a/2)] - \operatorname{erf} [\sqrt{k} (x_0 - a/2)]) \right] \} \frac{\sin \left(\frac{2\pi l}{\lambda} \frac{dn}{dx} \Big|_{x=x_0 b} \right)}{\frac{2\pi l}{\lambda} \frac{dn}{dx} \Big|_{x=x_0 b}}. \quad (14)$$

В эксперименте использованы следующие значения параметров: $\Delta l = 1 \cdot 10^{-5}$, $X_0 = 4,5 \cdot 10^{-3}$ м, $b = 2 \cdot 10^{-3}$ м, $\lambda = 6,326 \cdot 10^{-6}$ м, $l = 4 \cdot 10^{-3}$ м.

Использование численных методов для решения задач эллипсометрии [7] позволяет оценить влияние величины лазерного излучения на измеряемую характеристику интенсивности лазерного излучения. Для этого создана программа, алгоритм которой (рис. 3) реализован с использованием компилятора Borland C++ Builder под операционной системой Windows.

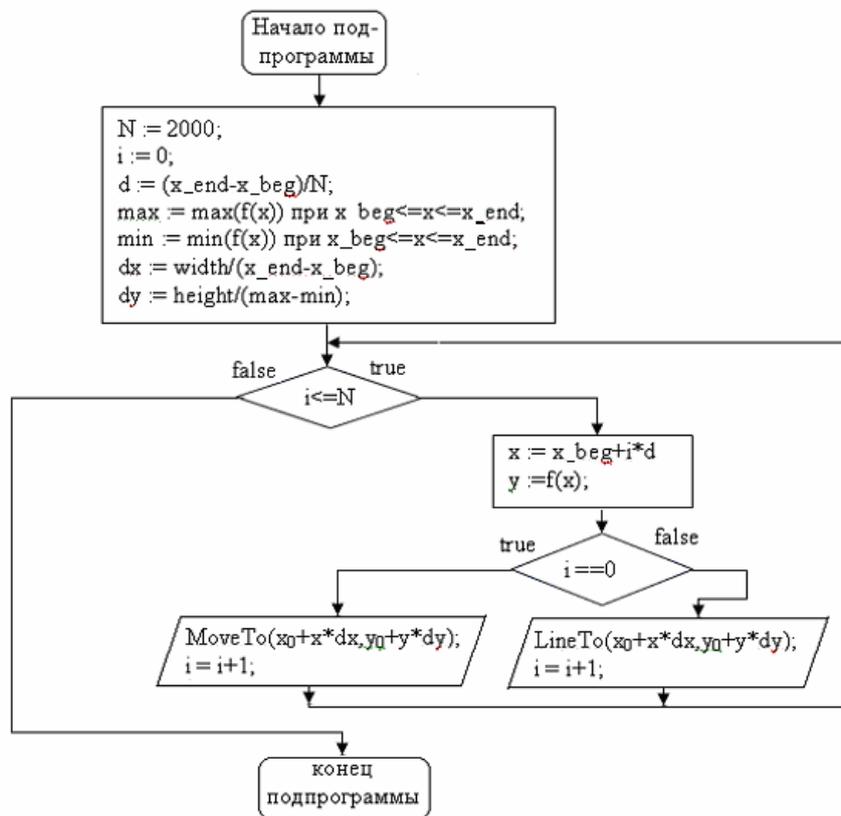


Рис. 3. Алгоритм программы

Программа строит график $I = f(b)$ в рамках прямоугольника, имеющего размеры $width \times height$, где b изменяется от x_beg до x_end . В программе используется подпрограммы min и max для получения минимального и максимального значений функции.

```

DrawGraph(HDC hdc, int x0, int y0, double (*f)(double))
{
    int N=2000, i;
    double x, y, d;
        max = Max(f, x_beg, x_end);
        min = Min(f, x_beg, x_end);
        dx = width / (x_end - x_beg);
    dy = height / (max - min);
    d = (x_end - x_beg) / n;
    for(i=0; i<=N; i++)
    {
        x = x_beg + i * d;
        y = f(x);
        if(i==0) MoveToEx(hdc, x * dx, y * dy, NULL);
        else LineTo(hdc, x * dx, y * dy);
    }
}
  
```

Полученная зависимость интенсивности обучения от размера лазерного источника представлена на рис. 4.

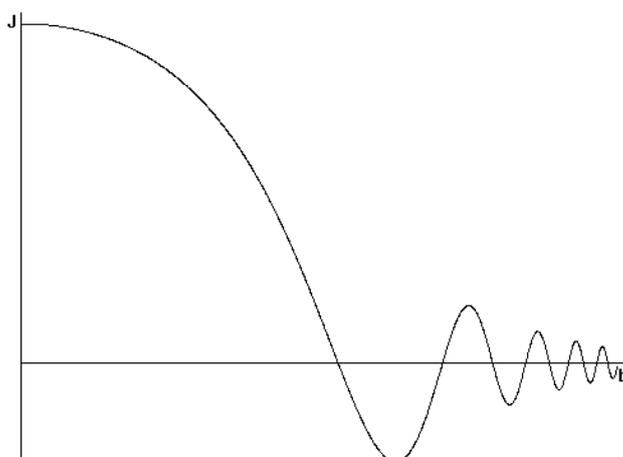


Рис. 4. Зависимость интенсивности обучения от размера лазерного источника

Из рис. 4 видно, что с ростом размера пучка источника излучения величина измеряемой составляющей интенсивности лазерного излучения уменьшается довольно быстро, что, в свою очередь, вносит существенный вклад в погрешность измеряемого параметра исследуемого процесса массопереноса.

Литература

1. Порай-Кошиц А.Б., Симоненко З.Г., Фейгельс В.И. Устройство для определения коэффициента молекулярной диффузии в жидкостях. / АС СССР. № 1178132. Бюллетень ОИ ПОТЗ, 1985.
2. Порай-Кошиц А.Б., Симоненко З.Г., Шмуйлович Г.А. / В сб.: VI Менделеевская дискуссия. Результаты экспериментов и их обсуждение на молекулярном уровне. Тезисы докладов, ч. II, Харьков, 1983, с. 341.
3. Симоненко З.Г. / Материалы IV Международной научной конференции «Проблемы пространства, времени и движения», СПбГИТМО (ТУ), Санкт-Петербург, 25–29 сентября 2000 г.
4. Плотников В.В., Ильина Л.П., Симоненко З.Г., Федоров В.Н. Теорема Остроградского и решение параболических уравнений. / Научная конференция «Петербургская математическая школа в период XIX века». С-Петербург, 24–28 сентября 2001 г.
5. Борн М., Вольф Э. Основы оптики. М.: Наука, 1973. 640 с.
6. Симоненко З.Г. Исследование параметров скорости массопереноса в жидких бинарных системах с границей раздела. / Материалы IV Международной научной конференции «Проблемы пространства, времени и движения». СПбГИТМО (ТУ), Санкт-Петербург, 2000. С 22.
7. Симоненко З.Г. Численные методы решения задач эллипсометрии при анализе системы подложка-плёнка. С. 109–111.

1. Системы автоматизированного проектирования	3
Пантелеев А.Е. Методы и алгоритмы создания системы верхнего (блочного) уровня АСУ ТП АЭС.....	3
Чернокнижная Е.Г. База данных САПР ТП гальванической обработки	9
Барышев М.В., Гатчин И.Ю. Модели представления знаний экспертных систем.....	14
Климанов В.А. Алгоритм функционирования системы автоматизированного проектирования рекурсивных кодовых шкал.....	19
Дрюков Н.Ю. Практические вопросы создания автоматизированных систем в защищенном исполнении	25
Касев К.В. Генетический алгоритм размещения.....	30
Полежаев В.С., Пудов Д.В. Генетический алгоритм трассировки.....	33
Малинин А.А., Таяновская Ю.Б. Использование фундаментальных алгоритмов поиска и объединения в разработке программного обеспечения электронных архивов САПР.....	38
2. Микроэлектроника. Дефектообразование в процессах производства и эксплуатации полупроводниковых интегральных схем.....	41
Смирнов Ю.В., Халецкий Р.А. Исследование пострadiационного эффекта в МОП-структурах, облученных под напряжением.	41
Халецкий Р.А. Гистерезис вольт-фарадных характеристик МОП-структур с поликремниевым затвором.....	46
Халецкий Р.А. Влияние электрофизических параметров кремниевой подложки на гистерезис вольт-фарадных характеристик МОП-структур.....	51
Власов В.В. Электрохимический метод микроструктурирования монокристаллов кремния (краткий обзор).....	54
Стройков И.И. Применение редкоземельных элементов в квантовой электронике.	60
Дышловенко С.С., Погумирский М.В., Фам Куанг Тунг. Влияние лазерного микроструктурирования поверхности кремния на морфологии и оптические свойства полученных структур.....	67
Дышловенко С.С., Стройков И.И., Фам Куанг Тунг. Влияние лазерного микроструктурирования поверхности кремния на электрофизические свойства полученных структур.....	73
3. Биотехнические измерительно-вычислительные системы	80
Головин П.А., Нечаев В.А., Нечаев Д.А. Экспертные системы для классификации болезней в медицинской диагностике.....	80
Головин П.А., Денисов Д.В., Нечаев В.А., Нечаев Д.А. Разработка распределенной базы данных на примере базы данных спортсменов олимпийского резерва.....	85

Гришенцев А.Ю. Моделирование распределения плотности тока в сложном неоднородном проводнике (часть 1).	87
Гришенцев А.Ю. Моделирование распределения плотности тока в сложном неоднородном проводнике (часть 2).	95
Гришенцев А.Ю., Петрова Е.Н. Предварительные эксперименты по изменению характеристик газоразрядного свечения волос при различных условиях.....	100
4. Перспективные информационные технологии.....	104
Антонов С.В. Разработка пользовательской консоли для программного комплекса мониторинга и аудита баз данных QDAS.....	104
Прокопенко А.Ю. Группа стандартов WiMAX.....	113
Киселев В.Б. Рекуррентный анализ – теория и практика.....	118
Тулякова М.С. Методы и алгоритмы сжатия графической информации.....	128
Александров Г.С., Елисеев О.В., Соловьев Д.В., Федоров А.А. Исследование сред и средств для записи синтезированных компьютером голограмм	136
Сарычев Д.Ю. Автоматизированная система сбора и обработки геофизической информации в реальном масштабе времени для выявления электромагнитных краткосрочных предвестников сильных землетрясений.....	141
Семенов Н.А. Моделирование ультранизкочастотных электромагнитных эмиссий, возникающих перед и во время сильных землетрясений.....	145
Пазухин А.В. Разработка программного обеспечения для управления системами автоматизации холодильных установок	152
Дегтярев А.А., Овечкина Е.В., Смелова А.В. Анализ и оптимизация бизнес-процессов на предприятии с помощью программного обеспечения	155
Александров Г.С., Елисеев О.В., Григорьева Н.С. (ОАО «Российский институт радионавигации и времени»), Соловьев Д.В. Практическое применение систем автоматизированного учета на предприятии.	158
Петров Ф.Г., Рядчин Ю.И. Установка и конфигурирование web-сервера Apache, интерпретатора PHP и сервера баз данных MySQL.....	166
5. Методы защиты информации.....	172
Лекомцева М.В., Нестерова Н.А., Семенов В.А. Анализ рисков информационной безопасности в банке	172
Куканова Н.Ю. Описание классификации угроз DSECCT.....	175
Липатов А.Л. (ЗАО «Лаборатория противодействия промышленному шпионажу»), Русинов А.Н., Салмин Е.А. Оценка и управление рисками ИТ-безопасности в информационных системах.	179
Липатов А.Л. (ЗАО «Лаборатория противодействия промышленному шпионажу»), Русинов А.Н., Салмин Е.А. Применение стандарта 802.1x для контроля доступа в локальную вычислительную сеть.....	183
Строганов К.В. Современная специфика обеспечения информационной безопасности локальных вычислительных сетей средних и крупных предприятий России.....	190
Егоров С.В., Иванов М.С., Рядчин Ю.И. Предотвращение угрозы вирусных атак в автоматизированных системах	194
Егоров С.В., Иванов М.С., Рядчин Ю.И. Защита информации при межсетевом взаимодействии с помощью межсетевых экранов	199
Белов Э.В. Обзор безопасности беспроводных сетей семейства 802.11	205

6. Оптические методы контроля качества элементной базы микроэлектроники	209
Ваняев В.Н. Оценка метрологических характеристик корреляционного метода частотно-фазовых измерений.....	209
Бандура А.С. Оценка точности данных навигационных сообщений космических аппаратов спутниковых радионавигационных систем	215
Скобелин А.А. Определение поправок на распространение сверхдлинных электромагнитных волн.....	222
Буданова А.Ю., Кокшаров Д.Н., Русин Д.А. Анализ эксплуатационных свойств плоских упругих чувствительных элементов электромеханических датчиков с регулярным микрорельефом на рабочих поверхностях.	228
Дегтярев А.А., Овечкина Е.В., Смелова А.В. Методы повышения надежности интегральных схем на основе композитных материалов.....	235
Кокшаров Д.Н., Русин Д.А. Оптический автоколлимационный метод исследования параметров профиля упругих чувствительных элементов при статическом напряженно-деформированном состоянии	237
Буданова А.Ю., Кокшаров Д.Н. Особенности формирования регулярного микрорельефа на рабочих поверхностях плоских упругих чувствительных элементов электромеханических датчиков.....	240
Кокшаров Д.Н., Русин Д.А. Разработка усовершенствованных конструкций герконов с улучшенными показателями качества	244
Валентик Н.Н., Когай Н.В., Кокшаров Д.Н. Параметры качества рабочих поверхностей упругих чувствительных элементов с регулярным микрорельефом.....	247
Розилайнен Д.Н. Информация и информационные ресурсы в техническом творчестве	251
Барынин А.О. Использование информационных технологий для решения прикладных задач приборостроения	255
Студеникин О.Л. Расчет параметров эллиптически-поляризованной волны отражающей среды для исследования параметров однородного волновода	260
Пуликовская З.С. Аппаратные средства решения задач массопереноса (диффузии) жидких сред	263
Шеховцов Н.Г. Задача принятия инженерных решений в условиях неопределенности	267
Хоанг Зянг. Оценка влияния величины источника излучения на параметры массопереноса.....	272

Научно-технический вестник СПбГУ ИТМО. Выпуск 29. I сессия научной школы «Информационная безопасность, проектирование, технология элементов и узлов компьютерных систем» / Главный редактор д.т.н., профессор В.Н. Васильев – СПб: СПбГУ ИТМО, 2006. 280 с.

**НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК СПбГУ ИТМО
Выпуск 29**

**I СЕССИЯ НАУЧНОЙ ШКОЛЫ
«Информационная безопасность, проектирование, технология элементов и узлов компьютерных систем»**

Главный редактор
д.т.н., профессор В.Н. Васильев
Научный редактор выпуска
д.т.н., профессор Ю.А. Гатчин
Дизайн обложки М.А. Петров
Редакционно-издательский отдел СПбГУ ИТМО
Зав. РИО Н.Ф. Гусарова
Лицензия ИД № 00408 от 05.11.99.
Подписано в печать 05.05.2006
Заказ 959. Тираж 100 экз.